

ALTO
Internet-Draft
Intended status: Standards Track
Expires: January 4, 2018

S. Randriamasy
Nokia Bell Labs
July 3, 2017

ALTO Contextual Cost Values
draft-randriamasy-alto-cost-context-02

Abstract

The Application-Layer Traffic Optimization (ALTO) Service has defined network and cost maps to provide basic network information, where the cost maps allow only one JSON value for a requested metric.

This document introduces several protocol extensions to allow ALTO clients to support use cases such as context based connection selection in cellular networks and calendaring for unattended data. This document refers to other extension proposals posted in the ALTO WG that can support the present use cases as well. Likewise, some of the proposed extensions may serve other ALTO use cases.

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

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on January 4, 2018.

Copyright Notice

Copyright (c) 2017 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	3
2. Use cases	4
2.1. Use Case 1: conditional RF costs in cellular networks . .	4
2.2. Use case 2: access-aware endpoint selection	5
3. Required ALTO extensions	6
4. Design options and examples	7
4.1. Overview of context features	7
4.1.1. Applicable ALTO services	8
4.2. Example IRD	8
4.3. Use case 1: Example scenario for the FCM Service	10
4.4. Design option: Network Map with cells as PIDs	11
4.4.1. Example: FCM Request for contextual 'ARFcost'	11
4.4.2. Example: FCM Response for contextual ARFcost	12
4.5. Use case 2: example ALTO transactions for the ECS	13
4.5.1. Use case 2: example with logical context parameter combinations	13
5. Deployment case: local ALTO Server cascaded with global ALTO Server	16
5.1. Cascaded ALTO Servers with one network map each	16
6. IANA Considerations	17
7. Security Considerations	17
8. Acknowledgements	17
9. References	17
9.1. Normative References	17
9.2. Informative References	17
Appendix A. An Appendix	18
Author's Address	18

1. Introduction

The IETF ALTO protocol specified in [RFC7285] provides guidance to over the top applications which have to select one or several hosts or endpoints from a set of candidates that are able to provide a desired data resource, or which need some provider-centric insight on the cost of application paths to these endhosts. The ALTO Service has defined network and cost maps to provide basic network information, where the cost maps allow only one JSON value for a requested metric.

This draft brings a use case where providing different values for the same cost metric can help in optimizing the application path selection. Typically, when an end host can connect to the network via multiple technologies or access points, the path performance for a metric may be accordingly impacted.

The present draft proposes to extend the cost information specified in [RFC7285] by providing, for the same cost metric, several possible cost values. Which value to provide depends on qualitative criteria as opposed to quantitative criteria such as time. The purpose is to allow a finer grained decision on which application endpoint or sub network to access.

Previous ALTO WG discussions have suggested introducing "the ability to "name" cost maps so that a single Information Resource Directory can link multiple cost maps with the same cost type to a single network map." The goal was to provide, for a given cost metric, multiple cost values depending on qualitative conditions named "circumstance", where a circumstance reflects a given policy.

For applications such as video download or streaming, a user equipment (UE) may use [RFC7285] to choose the best possible application resource location.

Currently, the insight of ALTO information on the path between a UE and a connection node (or say Endpoint) does not provide details below IP hops. However the major QoE challenges of wireless network users arise in the access network, that is, in the first hop between the UE and its one or more serving packet data network gateway (PGW). The path of a UE to its serving PGW(s) impacts the path to the content and thus the related QoE. Therefore, it is necessary to inform the UE, which could take the appropriate decisions w.r.t. the utilized access path. The access technology in current ALTO documents is accounted at the content location (last hop) side by distinguishing whether the requested content is located in a fixed or a wireless access network, as described in [draft-ietf-alto-deployment]

This document introduces several protocol extensions to allow ALTO clients to support use cases such as context-based connection selection in cellular networks and calendaring for unattended data. This document refers to other extension proposals posted in the ALTO WG that can support the present use cases as well. Likewise, some of the proposed extensions may serve other ALTO use cases.

2. Use cases

This section presents motivating use cases for contextual ALTO Costs with a focus on conditional RF costs in cellular networks. In these 2 use cases, a terminal UE is located in a LTE network and associated to a "local" ALTO Server(LAOS) that covers this access network, say up to the Packet Data Network (PDN) Gateway PGW and can itself connect to another ALTO Server having a more global view covering up to the whole ISP network. Such a deployment is proposed in section Section 5 of this draft.

2.1. Use Case 1: conditional RF costs in cellular networks

Let's assume a terminal UE located in a cellular network. An ALTO Client (LAOC) associated to the UE queries the local ALTO Server in order to know via which cell it should connect to the network. So in a first place, LAOC will query the connection cost associated to cells C1,... CK.

The present example assumes that the connection cost conveyed by the ALTO Server is a unitless value abstracting a non-real-time metric reflecting traffic conditions, aggregated over time and space. This metric aims at providing guidance to applications. It may integrate abstractions by the network provider, of actual costs impacted by other values such as congestion or available bandwidth are assumed to be not easily available to UEs or applications otherwise. The ALTO Server does not aim at reporting accurate radio conditions that indeed vary in time and space.

Let us call this metric: "ARF cost", standing for Abstracted RF cost.

Our example however includes 2 additional considerations:

- the ARF cost to a cell may be impacted by its load,
- a UE usually transmits a fair amount of "unattended data" (UD).

UD is considered in one of the key features for LTE enhancements in Release 13 and defined in 3GPP TS22.101 as follows: "Unattended Data Traffic : Data traffic of which the user is unaware he/she initiated, e.g. based on the screen/keypad lock being activated, length of time

since the UE last received any input from the user, known type of app (e.g. an application monitoring a user's health "mHealth" may need its data never treated as Unattended Data Traffic.)". UD traffic is often delay tolerant and it would be beneficial for the network if the UE can schedule its transmission. To this end, the UE can use an instant UD Indicator (UDI) sent by the LTE network. The UDI, accepted for LTE Release 13 is a single bit sent to the UE indicating whether UD in a cell is allowed (UDA) or not (UDNA). The status change of a UDI from UDA to UDNA is presumably triggered when the cell load exceeds a given threshold $T(udna)$. The value of $T(udna)$ may change across cells and in time but is not provided to UEs. If the UE had an ALTO calendar for either $T(udna)$ or for the abstracted cell load values, it could appropriately schedule the transmission of its UD, that will have to occur anyway. The UE could combine this calendar with the UDI it receives from the cellular network. The UDI state may change within sub-seconds and impact the data exchange. What is missing in the provided LTE information is:

- knowing whether the UDI threshold relates to downlink or uplink congestion.
- knowing the level of congestion that triggers a change in UDI and how it may evolve in time.

The UE thus can advantageously combine the non-real time ALTO information with the real-time UDI provided by the LTE network. The present draft illustrates how ALTO can fill these gaps with the support of:

- ALTO Cost Calendars,
- the proposed protocol extension providing context-dependent ALTO Cost values.

In this use case: ALTO calendars need to be requested via for the ALTO Filtered Cost Map (FCM) Service, the context parameters impacting the cost values are: "uda" (Unattended Data Allowed), "udna" (Unattended Data Not Allowed), "uplink", "downlink".

2.2. Use case 2: access-aware endpoint selection

In a second use case, an end-system called UEP is located in a LTE network and may connect via several access technologies, e.g. Cellular or WiFi. UEP may also benefit from a given Service Level Agreement SLA-m. Other parameters may characterize the UEP generated traffic.

Currently the insight of ALTO information in the path between a UE and a connection node (or say Endpoint) does not provide details below IP hops. However the major QoE challenges of wireless network users arise in the access network, that is, in the first hop between the UE and its one or more associated packet data network gateway (PGW). The path of a UE to its associated PGW(s) impacts the path to the content and thus the related QoE. Therefore, it is necessary to inform the UE, which could take the appropriate decisions w.r.t. the utilized access path. The access technology in current ALTO proposals is accounted at the content location (last hop) side by distinguishing whether the requested content is located in a fixed or a wireless access network, as described in [draft-ietf-alto-deployments] that states: "For ISPs with mobile network and fixed network, the traffic optimizing problems they focus will be optimizing the mobile traffic, except problems on last hop section."

For Mobile Network Operators (MNO) and their users, being connected via e.g. cellular or trusted Wifi can hugely impact the QoE and routing cost. Sometimes a 4G connection is preferable for users than a poor WiFi connection although potentially more expensive. Sometimes, MNOs have spare data resources or offer them for given SLAs. For both parties, access-aware Endpoint selection for Users is thus beneficial. One way to achieve this is that ALTO provides cost values depending on qualitative contextual parameters such as access technology and the access technology and SLA.

3. Required ALTO extensions

The aforementioned use cases can be supported with a few simple extensions to the ALTO protocol. A number of them have already been discussed in other WG drafts and use cases. The proposed extensions include:

- Cost value context parameters: a capability to allow exposing several possible context-dependent values for one metric, as proposed in the present document,
- Entities with associated domain and properties for cellular and wireless networks, that could be added to [draft-roome-alto-unified-props],
- Cost metrics for cellular and wireless networks: these features would extend current proposals in the WG, that could be added to [draft-ietf-alto-performance-metrics],
- Extended input for the Filtered Cost Map Service: to allow the input to comprise several (source-array, destination-array) pairs, as it has been proposed in [draft-yang-alto-path-vector].

4. Design options and examples

Similarly to Multi-Cost and Cost Calendar ([draft-ietf-alto-cost-calendar]), this proposal does not introduce new cost modes or new media-types. It ensures backwards compatibility with legacy ALTO Clients, that is: "A legacy ALTO Client must be able to send legacy requests to a Cost Context aware ALTO Server and get legacy responses as specified in RFC7285".

"A Cost Context aware ALTO Server must be able to receive and process requests sent by legacy ALTO Clients, as specified in RFC 7285".

Besides, the proposed extension is designed to be compatible with Multi-Cost ALTO and ALTO Cost Calendars ([draft-ietf-alto-cost-calendar]).

In the present draft version, the IRD indicates the supported context attributes as values encoded in JSON strings. This design simplifies the transactions, as it allows a limited number of context attributes or their combinations, say 1 to 5. Context attributes taking numerous or unpredictable values should be handled as values properties or metrics expressed in constraints.

- A cost context aware ALTO Server MUST provide metric values, as specified in RFC 7285, without any context consideration for all the Cost Types indicated in its "meta".

4.1. Overview of context features

Cost context attributes are strings with values such as "wifi", "cellular", "uda".

Cost context attributes are indicated in the IRD as capabilities of an information resource. They are associated to cost type names.

- A cost context aware ALTO may indicate in its IRD capabilities, whether and how context attributes may be combined in ALTO requests.

- A cost context aware ALTO Server MUST return metric values, without any context consideration, as specified in RFC 7285, if the value for a context attribute or a combination of attributes requested by the client is not available.

- A cost context aware ALTO may indicate a maximum number of context attributes or their combinations authorised in context-aware Client requests.

4.1.1. Applicable ALTO services

Draft [draft-bertz-alto-mobilitynets] proposes to identify network points of attachment (PoA) such as cells to PIDs, as PoAs are endpoint types not currently supported in ALTO. The current proposal is to represent cellular PIDs in an ALTO Network Map with no routes. PID properties as specified in [draft-roome-alto-unified-props] could be used to indicate the type of the PoA, together with other properties. ALTO properties are well suited for almost static attributes such as access type.

To abstract and convey connection properties with frequently changing values such as ARF Cost, load or congestion, the ALTO Filtered Cost Map service can be used. Connection properties may also be conveyed with the Endpoint property service or its extensions defined in [draft-roome-alto-unified-props].

Costs and properties with the extensions proposed in this document may be conveyed with different values depending on the context parameter. The present version of this draft focuses on context parameters associated to costs.

4.2. Example IRD

The purpose of ALTO is to guide the behavior of the end systems or applications without the need for networks to explicitly expose their performance values. In this example, the IRD does not expose the real load percentage of a cell to UE. Instead, it abstracts the cell congestion by a metric called 'ARFcost' represented by a number between 0 and 100, where the optimal value is 0. The values of 'ARFcost' are provided as an ALTO Calendar as specified in [draft-ietf-alto-cost-calendar-00] in shorter time intervals. In addition they differ, depending on the context values "uda" and "udna".

Besides, the IRD provides metric 'routingcost' as a MUST specified in [RFC7285], that may represent a more administrative or monetary access cost.

The IRD could publish the capability of a resource to provide context dependent 'routingcost' values as expressed for resource "filtered-cost-calendar-map".

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-ARFcost": {
        "cost-mode" : "numerical",
        "cost-metric": "ARFcost",
      }
    }
    ... other meta ...
  },
  "resources" : {
    "filtered-cost-calendar-map" : {
      "uri" : "http://alto.local.example.com/costmap/filtered/calendar/cont
ext",
      "media-types" : [ "application/alto-endpointcost+json" ],
      "accepts" : [ "application/alto-endpointcostparams+json" ],
      "capabilities" : {
        "cost-constraints" : true,
        "cost-type-names" : [ "num-routingcost",
                              "num-ARFcost"], // ++NEW
        "calendar-attributes" : [
          { "cost-type-names" : "num-routingcost",
            "time-interval-size" : "1 hour",
            "number-of-intervals" : 24}, // MAY ALSO BE SINGLE VALUE
          { "cost-type-names" : "num-ARFcost", // ++NEW
            "time-interval-size" : "5 minute",
            "number-of-intervals" : 12}
        ],
        "cost-context" : [ // ++NEW
          { "cost-type-names" : "num-ARFcost",
            "context-params" : [ ["uda", "udna"],
                                ["uplink", "downlink"]]
          }
        ], // ++NEW
        "max-context-attributes" : 10,
        "uses": [ "my-default-network-map" ]
      } // end FCM capab
    }
    ... other resources ...
  } // end resources
} // end IRD
```

4.3. Use case 1: Example scenario for the FCM Service

We assume an example scenario where a UE has the choice to connect to 2 cells C1 and C2.

As suggested in [draft-bertz-alto-mobilitynets], we may represent the cellular topology with an ALTO Network Map comprising PIDs representing the cells and named "Cell1", "Cell2", ... "Celln". A format for a cell identifier has been proposed in [draft-rauschenbach-alto-wireless-access] and is not being discussed here.

As a Network Map may cover a large number of cells, the Filtered Cost Map Service can be used to reduce data exchange and get information on a restricted number of cells.

We assume that the ALTO Client in the UE wants to get calendared values for ALTO metric "ARFcost" in order to appropriately schedule its unattended data transmission. The ALTO information resource 'ALTO Calendar' provides an array of time-dependent cost values and is being specified in [draft-ietf-alto-cost-calendar]. In addition, the ALTO Client wants these values for both the "uda" and "udna" context. Last, we assume that the UE needs the Cost values for both the uplink (UE to Cell-k) and downlink (Cell-k to UE) directions. We assume that the UE is located in the PID called "Cell1".

In this scenario, Cell1 is limited by its uplink capacity and Cell2 is limited by its downlink capacity. ALTO can be used to convey the following information:

At time interval T1 of the next Calendar:

- if Cell1 indicates "unattended data allowed" the downlink ARF cost is 20, and the uplink ARF cost is 70
- if Cell1 indicates "unattended data NOT allowed", the downlink ARF cost is 20, and the uplink ARF cost is 90
- if Cell2 indicates "unattended data allowed" the downlink ARF cost is 70, and the uplink ARF cost is 20
- if Cell2 indicates "unattended data NOT allowed", the downlink ARF cost is 90, in the uplink ARF cost is 20.

The ALTO Calendar provides values for the other 11 time intervals Ti.

4.4. Design option: Network Map with cells as PIDs

In this design, the cellular topology is represented with an ALTO Network Map comprising PIDs named "Cell1", "Cell2", ... "Celln". The UE is located in one of these PIDs. A Cost Map is associated to this Network Map and conveys metrics indicated in the IRD. The Cost Map can be to convey connection costs between firstly the UE to its serving cell (that is the PID to itself) and secondly the UE and neighboring cells (that is the PID to another one) and last, for both uplink and downlink directions.

The ALTO Server can regularly update the Cost Map and send filtered information to the ALTO Client. The proposed IRD design announces additional context attributes "uplink", "downlink". In this case and other potential cases, the context parameters need to be arranged w.r.t. their possible combinations (to be further specified in the IRD). For example, the IRD may announce that costs are provided for contexts "uda" and "udna" and this in both contexts "uplink" and "downlink". Or that costs are provided for contexts "uplink" and "downlink" and this in both contexts "udna" and "uda". In such a case, the IRD capability member may list the possible combinations in the capabilities as follows:

```
"cost-context" : [ // ++NEW
  { "cost-type-names" : "num-ARFcost",
    "context-params" : [{"uda", "uplink"},
                        {"uda", "downlink"},
                        {"udna", "uplink"},
                        {"udna", "downlink"}] // ++NEW
  }
]
```

This arrangement indicates that for the metric named "num-ARFcost", the ALTO Server can provide 4 different values: v1 for ["uda" AND "uplink"], ... v4 for ["udna" AND "downlink"].

Further versions of this draft will specify more elaborated logical combinations of context attributes, to moderate the length of the ALTO request and support use cases as described in section 4.5.1.

4.4.1. Example: FCM Request for contextual 'ARFcost'

The ALTO Client can specify the desired cost value context parameters in the request input. In particular, it can select one or more combinations indicated in the IRD. Its input parameter "context-params" is an array of all the desired combinations. In this

example, the ALTO Client wants to know the ALTO connection costs within Cell1 and Cell2. For each cell, the Client wants the 4 values, corresponding to all the combinations indicated below.

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

{
  "cost-type" : { "cost-mode": "numerical", "cost-metric": "ARFcost"},
  "calendared" : true,
  "context-params" : [{"uda", "uplink"}, // ++NEW
                      ["uda", "downlink"],
                      ["udna", "uplink"],
                      ["udna", "downlink"]],
  "pids" : [
    { "srcs" : [ "Cell1"], "dsts" : [ "Cell1"]},
    { "srcs" : [ "Cell2"], "dsts" : [ "Cell2"]}
  ]
}
```

4.4.2. Example: FCM Response for contextual ARFcost

The ALTO response provides, for each requested ("src", "dest") pair, a calendar of 12 JSON values, where each is an array of cost values arranged as specified in the "meta" of the ALTO response.

```
HTTP/1.1 200 OK
Content-Type: application/alto-costmap+json
Content-Length: ###

{
  "meta" : {
    "dependent-vtags" : [
      { "resource-id": "my-default-network-map",
        "tag": "3ee2cb7e8d63d9fab71b9b34cbf764436315542e"
      }
    ],
    "cost-type" : { "cost-mode": "numerical", "cost-metric": "ARFcost" },
    "calendar-response-attributes" :
      { "calendar-start-time" : Tue, 1 Sept 2016 13:00:00 GMT,
        "time-interval-size" : "5 minute",
        "numb-intervals" : 12 },
    "context-params" : [ [ "uda", "uplink" ], // ++NEW
                        [ "uda", "downlink" ],
                        [ "udna", "uplink" ],
                        [ "udna", "downlink" ] ]
  } // end meta
  "cost-map" : {
    "Cell1": { "Cell1": [[70, 20, 90, 20], ... , [50, 20, 70, 20]],
    "Cell2": { "Cell2": [[20, 70, 20, 90], ... , [20, 50, 20, 70]]
  }
}
```

4.5. Use case 2: example ALTO transactions for the ECS

In this use case, the UE requests the ECS to a local ALTO server for the routingcost to the PGW and wants the metric values varying w.r.t. the "access-type" and "SLA-id". Note that the "context" related design feature can be easily transposed for the Cost Map Service.

4.5.1. Use case 2: example with logical context parameter combinations

This section proposes a design, allowing a Client to arrange input context parameters in logical combinations. The purpose is to show how such combinations of context parameters avoids specifying as many metrics and moderates the amount of exchanged data.

In this example the ALTO Server indicates in its IRD that it can provide endpoint cost maps for the example metrics "routingcost" and "bandwidthscore". Values for metric "routingcost" are provided w.r.t. 2 types of context parameters. The ALTO Client may query values for metric "routingcost" for either of these types of parameters or both or none.

For each type, the parameters are listed in an array. We have 2 arrays:

- ["cell", "wifi", "lan"]
- ["SLA-1", "SLA-2", "SLA-3"]

This indicates that in each array, the client can pick one or more parameters and combine them with one or more parameters in the second array. The ALTO Server will provide costs w.r.t. the AND combination across and within arrays.

In the present example, if the Client requests cost values for the combination:

```
[["cell", "wifi"], ["SLA-3"]]
```

The server will provide 2 values: one for ("cell" AND "SLA-3") and the second one for ("wifi" and "SLA-3").

4.5.1.1. Example IRD with logical context parameter combinations

The IRD below specifies the possibility to combine parameters from the two arrays of the example above.

```
"resources" : {
  "filtered-cost-calendar-map" : {
    "uri" : "http://alto.local.example.com/endpointcostmap/lookup/context
",
    "media-types" : [ "application/alto-endpointcost+json" ],
    "accepts" : [ "application/alto-endpointcostparams+json" ],
    "capabilities" : {
      "cost-constraints" : true,
      "cost-type-names" : [ "num-routingcost",
                           "num-bandwidthscore" ],
      "cost-context" : [ // ++NEW
        { "cost-type-names" : "num-routingcost",
          "context-params" : [ ["cell", "wifi", "lan"],
                              ["SLA-1", "SLA-2", "SLA-3"] ]
        }
      ]
    } // end ECM capab
    ... other resources ...
  } // end resources
```

4.5.1.2. Use case 2: example ECS request with logical context parameter combinations

The ALTO Client queries the ECS between 2 endpoints for the following combinations: ("cell" AND "SLA-3") and ("wifi" and "SLA-3") and thus arranges its input context parameters as follows:

```
POST /endpointcost/lookup/context HTTP/1.1
Host: alto.local.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"},
  "context-params" : [["cell", "wifi"], ["SLA-3"]],
  "endpoints" : {
    "srcs": [ "ipv4:192.0.2.2" ],
    "dsts": [
      "ipv4:192.0.2.89",
      "ipv6:2000::1:2345:6789:abcd"
    ]
  }
}
```

4.5.1.3. Use case 2: example ECS response with logical context parameter combinations

Following the ALTO Client request of the above example, The ALTO Server provides a response as follows:

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

{
  "meta" : {
    "cost-type" : {"cost-mode" : "numerical", "cost-metric" : "routingcost"},
    "context-params" : [{"cell", "wifi"}, ["SLA-3"]]
  } // end meta

  "endpoint-cost-map" : {
    "ipv4:192.0.2.2": {
      "ipv4:192.0.2.89" : [10, 4],
      "ipv6:2000::1:2345:6789:abcd" : [4, 6]
    }
  }
}
```

5. Deployment case: local ALTO Server cascaded with global ALTO Server

To maintain scalability, the ALTO coverage network zone can be decomposed in one "local" ALTO Server part covering a restricted local network zone, for instance within the first IP hop range and another "global" part covering the rest of the ISP network, similarly to what is proposed in [draft-ietf-alto-deployments]. The local ALTO server may include the guidance given by the ISP ALTO server and compose it with the "global" guidance in its replies to its ALTO clients. Recent ALTO WG discussions open the possibility for one IRD to indicate multiple network maps having different levels of detail.

5.1. Cascaded ALTO Servers with one network map each

In the "cascaded" use case, the ALTO Service is preferably distributed among two ALTO Servers as follows:

The ALTO Client serving the UE is referred to as the LAOC and can be located either in the UE or in the network.

1. A Local ALTO Server (LAOS)

- * Hosts the information on the local EPS network, covering the paths between e.g. the UEs and the cells or the PGWs,
- * Hosts an ALTO Client that sends an ALTO request to a "global" ALTO Server, covering the zone beyond the PGW. It can possibly get the global Server updates using the extensions specified in [draft-ietf-alto-incr-update-sse].

- * receives the ALTO request issued by the ALTO Client associated to the UE.

2. a "core" ALTO Server covers the whole ISP network view, as it would if the "local ALTO Service" is not available or deactivated.

6. IANA Considerations

This document makes no request of IANA.

Note to RFC Editor: this section may be removed on publication as an RFC.

7. Security Considerations

8. Acknowledgements

Many thanks to Dawn Chan, Li Geng, Xin Wan, Yichen Qian for their feedback on this draft.

9. References

9.1. 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>>.
- [RFC7285] Alimi, R., Penno, R., Yang, Y., Kiesel, S., Previdi, S., Roome, W., Shalunov, S., and R. Woundy, "Application-Layer Traffic Optimization (ALTO) Protocol", RFC 7285, September 2014.

9.2. Informative References

- [draft-bertz-alto-mobilitynets] Bertz, L., "Mobility Network Models in ALTO", October 2015.
- [draft-ietf-alto-cost-calendar] Randriamasy, S., Yang, Y., Wu, Q., Deng, L., and N. Schwan, "ALTO Cost Calendar", February 2017.
- [draft-ietf-alto-deployment] Stiernerling, M., Kiesel, S., Scharf, M., Seidel, H., and S. Previdi, "draft-ietf-alto-deployments-16", July 2016.

[draft-ietf-alto-incr-update-sse]

Roome, W. and Y. Yang, "ALTO Incremental Updates Using Server-Sent Events (SSE)", Septembre 2016.

[draft-ietf-alto-performance-metrics]

Wu, Q., Yang, Y., Lee, Y., Dhody, D., and S. Randriamasy, "ALTO Performance Cost Metrics", March 2017.

[draft-rauschenbach-alto-wireless-access]

Rauschenbach, U., "ALTO in wireless access networks", October 2014.

[draft-roome-alto-unified-props]

Roome, W., "Extensible Property Maps for the ALTO Protocol", July 2016.

[draft-yang-alto-path-vector]

Bernstein, G., Gao, K., Lee, Y., Roome, W., Scharf, M., and Y. Yang, "ALTO Extension: Path Vector Cost Mode", July 2016.

Appendix A. An Appendix

Author's Address

Sabine Randriamasy
Nokia Bell Labs
Route de Villejust
Nozay 91460
FRANCE

Email: sabine.randriamasy@nokia-bell-labs.com