ALTO Path Vector: Open Issues

draft-ietf-alto-path-vector-19
Kai Gao

IETF 112
Nov 10, 2021
Outline

- GENART & ARTART LC reviews
- OPSDIR LC reviews
ARTART & GENART LC Reviews

ARTART LC review:

- Open issue 1: Use of Content-ID
- Open issue 2: IPv6 example

GENART LC review:

- Open issue 1: IPv6 example

1) Minor: Use of Content-ID

This document makes use of Content-ID with constrained syntax and values. This usage "ought" to be consistent with http and mime definitions of Content-ID syntax.

I'm actually having some difficulty in identifying the most recent authoritative document for this syntax. The IANA registry of http header field names refers to RFC4229, and it refers to "The HTTP Distribution and Replication Protocol", W3C NOTE NOTE-drp-19970825, August 1997. That reference seems oddly specialized to be the authoritative definition. In any case it defines it as:

```
content-identifier = URI { "\", URI }
```

A traditional one is `<left|right>` with some constraints on the content of left and right.

I'm not going to be the judge of what is authoritative here, but I suggest you find out and then conform to it.

2) NIT: Use of IPv6

As noted by IdNits, you should consider including some IPv6 addresses in examples.

Nits/editorial comments:

Please add some IPv6 examples in Section 8.
Use of Content-ID

- Define the format to be compatible with RFC 2387 & 5322
- Protocol specification and examples are revised accordingly

---

**6.6. Part Resource ID and Part Content ID**

A Part Resource ID is encoded as a JSON string with the same format as that of the type ResourceID (Section 10.2 of [RFC7285]).

Even though the client-id assigned to a Path Vector request and the Part Resource ID MAY contain up to 64 characters by their own definition, their concatenation (see Section 5.3.2) MUST also conform to the same length constraint. The same requirement applies to the resource id of the Path Vector resource, too. Thus, it is RECOMMENDED to limit the length of resource ID and client ID related to a Path Vector resource to 31 characters.

A Part Content ID conforms to the format of msg-id as specified in [RFC2387] and [RFC5322]. Specifically, it has the following format:

```
"<" PART-RESOURCE-ID "@" DOMAIN-NAME ">"
```

PART-RESOURCE-ID: PART-RESOURCE-ID has the same format as the Part Resource ID. It is used to identify whether a part message is a Path Vector or a Property Map.

The body MUST consist of two parts:

- The Path Vector part MUST include "Content-ID" and "Content-Type" in its header. The value of "Content-ID" MUST have the same format as the Part Content ID as specified in Section 6.6.

```
--example-1
Content-ID: propmap
Content-Type: application/alto-propmap+json

{ "meta": { "dependent-vtags": [ { "resource-id": "ecs-pv.costmap", "tag": "d827f484cb66ce6df6b5077cb8562b0a" } ] }
```

---

```
--example-1
Content-ID: <propmap@alto.example.com>
Content-Type: application/alto-propmap+json

{ "meta": { "dependent-vtags": [ { "resource-id": "ecs-pv.ecs", "tag": "d827f484cb66ce6df6b5077cb8562b0a" } ] }
```
OPSDIR Review

- Open issue 1: Real use cases & examples
- Open issue 2: Clarification and examples for ANE
- Open issue 3: Clarification of terminologies (request, flow, ANE domain)
- Open issue 4: Examples/Pointers for traffic orchestration using PV
Use Cases & Examples

The use cases defined are quite varied - large scale analytics, mobile and CDNs. SENSE and LHC are not specifically data analytics use cases in the usual sense of the word, rather SENSE is a model for orchestrating network links (and capacity) between sites, and the LHC provides large scale data sets for four major experiments that are distributed and computed upon via the WLCG (worldwide large hadron collider computing grid).

For LHC, QoE is not so much about time to complete; the important point is not to have data backlogging if performance drops.

For the WLCG, two networks have evolved over many years to carry the traffic from the four main experiments; LHCOPN, the optical network, and LHCONET, the overlay network, both of which are 'manually' configured, and with enough capacity for the traffic thanks to regular network forward look exercises. While a little complex to administer, other emerging disciplines have expressed interest in using LHCONET to move data, and some have established agreements (e.g. SKA, I believe). While a means to provision capacity on demand would be attractive, the R&E networks typically have capacity, LHCOPN/LHCONET carry the LHC traffic, and bottlenecks are in the end sites (hence the evolution of the Science DMZ principles).

Some specific examples of ANEs would be very helpful. While the document does contain examples, they are not grounded around a use case I can readily relate to, such as the orchestration of a large data flow between two sites in different R&E networks. Can the doc show some real examples?

In 4.2.1 it talks of ALTO client identifying bottlenecks; a little more discussion and examples of that would be useful, for practical use cases such as an international R&E data transfer.

Open issue 1: Real use cases & examples

- Use case section does not reflect real traffic steering objectives of existing systems (e.g., LHC)
- Some specific examples of ANEs are needed to show real use cases
- Discussion and examples for identifying bottlenecks are needed
Use case section does not reflect real traffic steering objectives of existing systems (e.g., LHC)

In ’19, multiple examples are given covering different traffic steering objectives, different control capabilities and different information:

- Case 1: data analytics system optimizing query completion time over geo-distributed data centers
- Case 2: data analytics system collecting bottleneck information for transfer orchestration (with pointers to potential clients such as G2)
- Case 3: cloud-edge service discovery exposing available computation/storage resource information
Some specific examples of ANEs are needed to show real use cases.

Pointers to reported use cases of the described scenarios are given.

Specific ANE examples and potential properties for these use cases are given.

---

**Real Use Cases & Examples**

- Some specific examples of ANEs are needed to show real use cases.
- Pointers to reported use cases of the described scenarios are given.
- Specific ANE examples and potential properties for these use cases are given.

---

**Figure 7: Example Service Edge Query Results**

```
a: { b: [anel1], c: { d: [anel1, ane11, ane111], e: [f: [anel1v1]] }
anel1: bw = 10 Gbps (link: A->B)
anel2: bw = 10 Gbps (link: B->GW)
anel3: bw = 50 Gbps (link: GW->Core)

and site 2 returns

c: { d: [anel1, ane11, ane111] }
e: { f: [anel1v1] }
anel1: bw = 5 Gbps (link Y->X)
anel11: bw = 10 Gbps (link GW->Y)
anel111: bw = 20 Gbps (link Core->GW)
anel1v1: bw = 10 Gbps (link Y->GW)
```

---

**References**

Real Use Cases & Examples

- Discussion and examples for identifying bottlenecks are needed
- Pointers to related studies are given
- Use case 2 gives a concrete example on how PV can help expose bottleneck information
Section 3 talks of definitions of ANEs being “similar to” Network Elements in RFC2216, but this is vague. The topology in Figure 5 is quite simple, as an example; something more realistic would be interesting. Ultimately, if ALTO clients have the full network topology even then they may not know about the routing that occurs by default, so implicitly there’s an assumption of a capability to steer traffic to meet a request. What is the “request” referred to in 5.1.2, for example?

It seems that the document argues that ‘bottlenecks’ are typically capacity based; do ANEs include specific links, rather than routers, firewalls, etc? A stateful firewall can be a significant bottleneck on throughput, for example.

- Open issue 2: Clarification and examples for ANE
- Examples are given in use cases (Sec 4) and where ANE is specified
OPSDIR Review

- Open issue 3: Clarification of terminologies (request, flow, ANE domain)

Section 3 talks of definitions of ANEs being "similar to" Network Elements in RFC2216, but this is vague. The topology in Figure 5 is quite simple, as an example: something more realistic would be interesting. Ultimately, if ALTO clients have the full network topology even then they may not know about the routing that occurs by default, so implicitly there's an assumption of a capability to steer traffic to meet a request. What is the "request" referred to in 5.1.2, for example?

By design, ANEs are ephemeral and not to be used in further requests to other ALTO resources. More precisely, the corresponding ANE names are ephemeral and not to be used in further requests to other ALTO resources.

The discussion on p.9 about multiple flows is a little odd: in practice in R&E networks large transfers use tools like GridFTP which uses multiple parallel TCP flows, such that loss on individual flows does not severely impact throughput. Of course, BBR also reduces this concern.

It must be noted that the term "domain" here does not refer to a network domain. Rather, it is inherited from the "entity domain" defined in Sec 3.2 in [I-D.ietf-alto-unified-props-new] that represents the set of valid entities defined by an ALTO information resource (called the defining information resource).

- The terms are revised to avoid ambiguity

The discussion on p.9 about multiple flows is a little odd: in practice in R&E networks large transfers use tools like GridFTP which uses multiple parallel TCP flows, such that loss on individual flows does not severely impact throughput. Of course, BBR also reduces this concern.

Is the use of ALTO designed for single domain, or can it span multiple domains? It seems the latter, given the definition of ANE domains, but for the latter there is no specific model for the common definition of ANEs.
Open issue 4: Examples/Pointers for traffic orchestration using PV

Given the definition of ANEs and PVs, how is traffic then orchestrated or optimised? Some pointers here would be useful. SENSE may be one example. From my own discussion with people involved with SENSE (and AutoGOLÉ which uses it) there is as yet no use of ALTO (rather SENSE uses its own methods to orchestrate based on intent-based descriptors), but it is something that may be considered in the future.

What of non-ALTO traffic on the same links; is the approach to reserve x% capacity of a link for ALTO orchestrated traffic (the SENSE approach, I believe)?

Examples/pointers are provided along with use cases

It is emphasized that ALTO is used for information exposure and traffic steering is done by the application.
Summary

- Issues from LC reviews are addressed
- Waiting for reviewers’ feedback
- How to proceed?
Thanks!