ALTO WG
Internet-Draft
Intended status: Informational
Expires: January 12, 2023

LM. Contreras
Telefonica

July 11, 2022

Considering ALTO as IETF Network Exposure Function
draft-contreras-alto-ietf-nef-01

Abstract

This document proposes ALTO as the means for exposure of underlay network capabilities for multiple overlays on top of the network.

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 https://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 12, 2023.

Table of Contents

1. Introduction 2
2. Exposing network capabilities for enhancing service delivery 2
3. ALTO versus network controller conceptualization 3
4. Modes of usage 4
4.1. Existing use cases 4
4.1.1. Network topology and reachability 4
4.1.2. Network performance metrics per path 4
4.1.3. Segmented paths and associated characteristics . . . 5
4.1.4. In-time view of dynamic IP addressing allocation . . 5
4.2. Prospective use cases 5
4.2.1. Determination of optimal compute facility taking into account network information 5
4.2.2. Information related to Service Functions and Service Function chains 6
4.2.3. Visibility of underlying network information in overlay networks 6
4.2.3.1. Cellular case 6
4.2.3.2. Media distribution case 6
5. ALTO as IETF Network Exposure Function 7
6. TODO for next versions of this document 8
7. Security Considerations 9
8. IANA Considerations 9
9. References 9
Acknowledgments 11
Author's Address 11

1. Introduction

Networks are turning on consumable objects by external applications and services. In order to facilitate that, it is necessary to expose the capabilities offered by the networks in such a way that the applications and services can produce informed decisions that assist in the improvement of the service delivery.

Thus it is convenient to define mechanisms for capabilities exposure that could provide required information for IETF networks. ALTO [RFC7285] can play such a role. This memo describes existing and foreseen capabilities that can be exposed by leveraging on ALTO.

2. Exposing network capabilities for enhancing service delivery

More and more, services and applications rely on information retrieved from the network in order to make decisions positively affecting the service delivery, by adapting the applications to the reality observed through the retrieved information. This information is typically offered by specific components in the network with the

mission of aggregating, processing and securely exposing such information.

Several initiatives are being developed in order to facilitate such exposure of capabilities and information at different network levels. For example, 3GPP defines the Network Exposure function (NEF) [TS29.522] as a secure, scalable and simplified exposing tool for capabilities (as well as events) supported by the 5G Core (5GC) network. Main capabilities of NEF are the following:

- Securely expose 3GPP Network Functions (NFs) capabilities to Application Functions (AF).
- Secure provision of information to 5GC, including authentication and authorization to AF.
- Rate limit AF access to 5GC functions and information, including charging functions.
- Translation of internal-external information, e.g. identities.

This is done though a number of APIs defined in [TS29.522]. A specific NEF instance may support only a subset of the APIs specified for capability exposure.

Further examples are present in other network concerns. Thus, in ETSI Multi-Access Edge Computing (MEC) group a number of APIs allow the retrieval of specific network information at the edge (e.g., location API [MEC-013]), or the O-RAN Alliance which is working on exposing information to applications running on top of the non-real time Radio Informacion Controller (RIC) [O-RAN].

The purpose of this document is to consider ALTO as the means for exposure of underlay network capabilities to multiple overlays on top of the network. In other words, serve as "ground truth" from the network provider perspective to the applications consuming network capabilities in the scope of IETF.

3. ALTO versus network controller conceptualization

A relevant question that could arise is about the difference on purpose between ALTO and a network controller in the network.

Primarily, the final purpose of these components is quite different. In this respect, a network controller (i.e., SDN controller [RFC7149]) can be seen as the element devoted to orchestration, control and management of the network assets, that is, the component in charge of administering network objects. Typically, a network

controller leverages on another IETF functional component used for network control, such as the Path Computation Element (PCE) [RFC4655], which is used to compute paths for forwarding purposes based on network constraints. In contrast to these two elements, ALTO acts as a "one-stop-shop" for retrieving (and correlating) network related information, potentially leveraging on the capabilities of the othe rtow (i.e., SDN controller and/or PCE).

Moreover, ALTO has been included as part of some architectural frameworks, such ABNO [RFC7491], with the mission of allowing joint network and application-layer optimization precisely by exposing to applications an abstract network topology containing only information relevant to such application. In this manner the application can take an informed decision and request specific control actions in the network.

4. Modes of usage

This section presents different modes of usage of ALTO network exposure capabilities to improve network operations. Some of these usages can be implemented nowadays based on existing specifications, while a set of other use cases is considered as prospective since more specification work is yet needed in IETF.

4.1. Existing use cases

This subsection presents a number of use cases already described that can leverage on ALTO as IETF Network Exposure Function.

4.1.1. Network topology and reachability

The basic ALTO capabilities [RFC7285] provide network maps associated with costs in a manner that for any pair of source and destination can be retrieved information about topology and reachability. This can be considered as the fundamental or baseline information on top of which the other modes of usage are built on.

4.1.2. Network performance metrics per path

Extensions defined in [I-D.ietf-alto-performance-metrics] permit the reporting of standard-based performance metrics associated to the paths generated in the network map. With that view, applications consuming ALTO (i.e., ALTO clients) can determine the performance expectation for the possible paths between an origin and a destination. Thus, not only puere cost but also performance can be considered as an element for decision.

4.1.3. Segmented paths and associated characteristics

Original ALTO concentrates on end-to-end paths. However it may result of interest to get knowledge of specific parts of the end-to-end paths that could produce problems such as e.g. congestion. Then, having means of segmenting the end-to-end paths becomes useful. [I-D.ietf-alto-path-vector] allows for that defining a new abstraction called Abstract Network Element (ANE) to represent components constituting an entire end-to-end network path as a vector of ANEs.

4.1.4. In-time view of dynamic IP addressing allocation

Some architectures allow for dynamic allocation of IP address subnets across the network. An example of that is the Control and User Plane Separation (CUPS) architecture for Broadband Network Gateways (BNGs) [I-D.wadhwa-rtgwg-bng-cups], [TR-459]. In that architecture, the control place of the BNG has the possibility of dynamically assigning IP address subnets to different elements distributed in the network, acting as user plane functions of the BNG. This dynamic allocation implies that certain IP prefixes could be allocated in different parts of the network along the time. By means of ALTO and its network map is it possible to obtain an up-to-date view of the topological location of each subnet in runtime, facilitating the optimization of some services (e.g. media distribution) in an automated manner.

4.2. Prospective use cases

This subsection presents a number of use cases that could be enabled by ALTO as IETF Network Exposure Function.

4.2.1. Determination of optimal compute facility taking into account network information

ALTO can be used as a component to provide insights on the reachability of suitable compute facilities. An initial case has been documented in [I-D.contreras-alto-service-edge]. The rationale for this case is that ALTO receives information of connected compute capabilities in terms of e.g. CPU, memory and storage. This information can be put together with the network map, in a way that the cost of reaching those capabilities can be easily determined.

Note that if further information apart of cost is included in the map (e.g., performance metrics) then the resulting information provided to applications becomes enriched.

4.2.2. Information related to Service Functions and Service Function chains

ALTO can provide information relative to the paths characteristics associated with a single Service Function or with a number of chained Service Functions. This can be useful at the definition phase of a network service, either considering specific instances of the constituent Service Functions, or as a mean of identifying the more appropriate Service Functions to compose a service.

[I-D.lcsr-alto-service-functions] proposes different situations of interest and explores augmentations in ALTO to support the retrieval of information associated to Service Functions. Internal IETF solutions as the ones for Service Function Chaining or SRv6 programmability can benefit of this insight, but also other solutions like ETSI NFV, 3GPP, O-RAN or any other requiring efficient decisions in relation with chains of Service Functions can be benefitted for their own automation, management and control processes.

4.2.3. Visibility of underlying network information in overlay networks

Different overlay networks run today leveraging the connectivity provided by the basic underlying transport network. Since specific situations on the transport network can result in relevance for the service being provided by the overlays, it is crucial to facilitate the observation of such situations from the underlay to the overlay.

4.2.3.1. Cellular case

Mobile networks leverage transport networks to connect mobile access nodes with core management and control entities (e.g., for mobility management, policing, etc), running in an overlay mode through tunneling (i.e., the GTP protocol). [I-D.li-alto-cellular-use-cases] presents the benefits of exposing network information for applications running on access devices of a cellular network.

4.2.3.2. Media distribution case

Media delivery systems, as traditional CDNs, deliver content to end-users in an over-the-top fashion. The key aspect for an efficient and optimal delivery of the content is to select the proper delivery point for whatever end-user requesting it is to have a clear view of the network topology (including the associated costs or any other information that could enrich the decision, such as performance metrics). In this respect, the information exposed by ALTO in reference to the requesting end-user can be consumed by CDN control elements for improving the decision on what delivery point to select [RFC7971].

Further than that, additional scenarios can benefit from ALTO network information exposure capabilities. For instance, in scenarios of interconnection of CDNs, such as the one described in [I-D.ryan-cdni-capacity-insights-extensions] for advertising capacity associated with the CDN internal to an operator, could leverage on ALTO capabilities for that purpose (with the necessary augmentations).

5. ALTO as IETF Network Exposure Function

From its inception, ALTO was defined as a way of informing applications about network-related aspects for improving the overall service.

The applications under scope can be either internal or external to the operator of the network. The implications can differ in the level of aggregating, abstracting and securely exposing the information, but the purpose keeps being the same.

Figure 1 illustrates the role of ALTO as IETF Network Exposure Function.

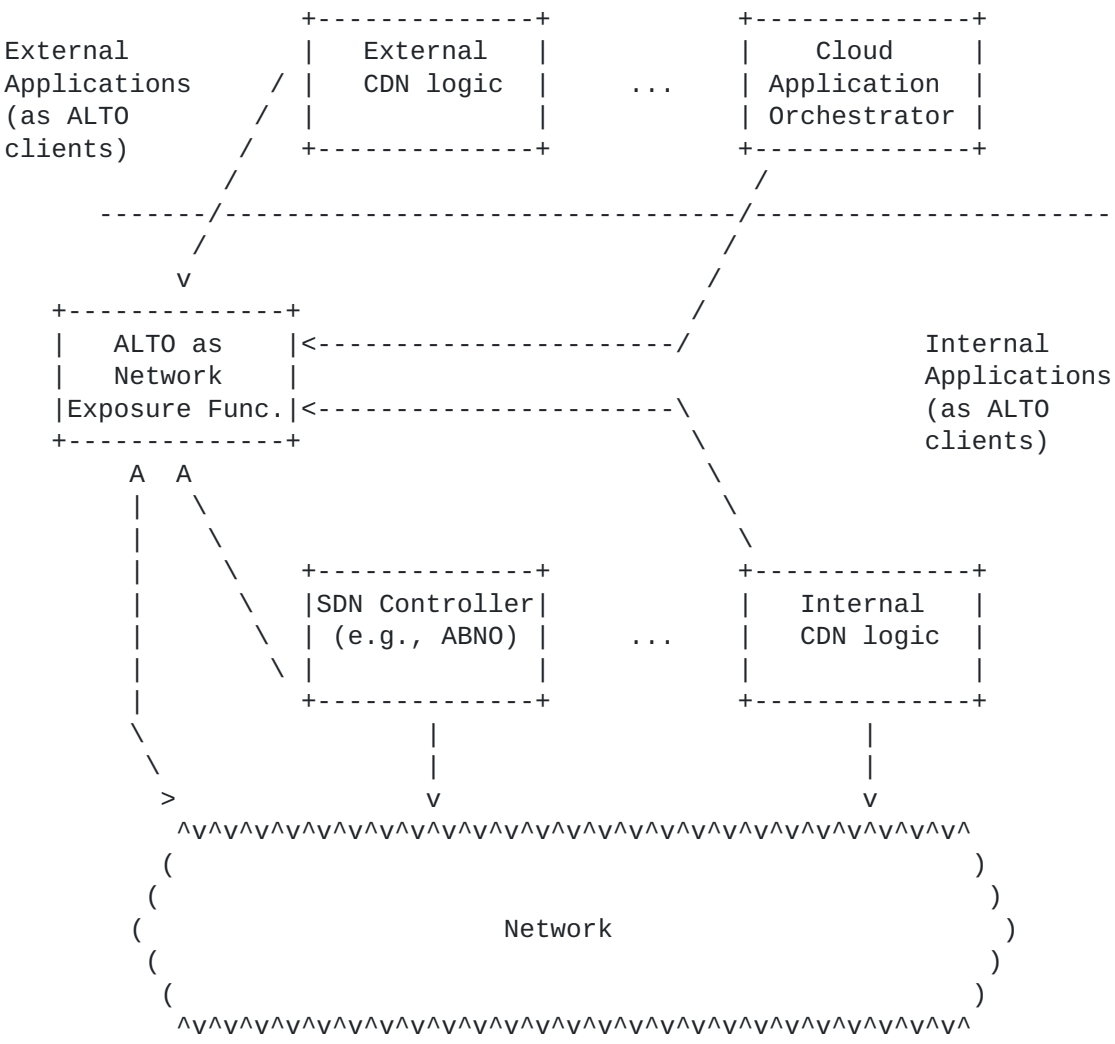

Figure 1: ALTO as IETF Network Exposure Function

Each particular ALTO instance in a certain network could support only a subset of the capabilities discussed in the use cases described before. In this respect, ways of advertising supported capabilities should be defined.

6. TODO for next versions of this document

This version is a work-in-progress. Next versions of the document will address somo further aspects such as:

- Identification of further network capabilities of interest for being exposed by ALTO.

- Complete security aspects when exposing information to internal and external applications.

7. Security Considerations

ALTO security considerations as reflected in [RFC7285] apply to this document.

Apart from that, the following aspects should be taken into consideration:

- Authentication between ALTO and any external entitiy consuming ALTO, to prevent malicious behaviors.

- Privacy of the information shared between components, especially when those components pertain to different administrative domain (e.g., an external CDN retieving network information from a network of a different administrative domain).

- Secure transport of the information in the communication with ALTO Server (e.g., TLS, etc).

8. IANA Considerations

This draft does not include any IANA considerations

9. References

[I-D.contreras-alto-service-edge]
Contreras, L. M., Lachos, D. A., Rothenberg, C. E., and S. Randriamasy, "Use of ALTO for Determining Service Edge", draft-contreras-alto-service-edge-05 (work in progress), July 2022.

[I-D.ietf-alto-path-vector]
Gao, K., Lee, Y., Randriamasy, S., Yang, Y. R., and J. J. Zhang, "An ALTO Extension: Path Vector", draft-ietf-alto-path-vector-25 (work in progress), March 2022.

[I-D.ietf-alto-performance-metrics]
Wu, Q., Yang, Y. R., Lee, Y., Dhody, D., Randriamasy, S., and L. M. C. Murillo, "ALTO Performance Cost Metrics", draft-ietf-alto-performance-metrics-28 (work in progress), March 2022.

[I-D.lcsr-alto-service-functions]
Contreras, L. M. and S. Randriamasy, "ALTO extensions for handling Service Functions", draft-lcsr-alto-service-functions-00 (work in progress), July 2022.

[I-D.li-alto-cellular-use-cases]
Gang, L., Randriamasy, S., and C. Xiong, "ALTO Uses Cases for Cellular Networks", draft-li-alto-cellular-use-cases-00 (work in progress), July 2021.

[I-D.ryan-cdni-capacity-insights-extensions]
Ryan, A., Rosenblum, B., and N. B. Sopher, "CDNI Capacity Capability Advertisment Extensions", draft-ryan-cdni-capacity-insights-extensions-02 (work in progress), March 2022.

[I-D.wadhwa-rtgwg-bng-cups]
Wadhwa, S., Shinde, R., Newton, J., Hoffman, R., Muley, P., and S. Pani, "Architecture for Control and User Plane Separation on BNG", draft-wadhwa-rtgwg-bng-cups-03 (work in progress), March 2019.

[MEC-013] "GS MEC 013 Location API V2.1.1", ETSI GS MEC 013 V2.1.1 , September 2019.

[O-RAN] "Non-RT RIC Architecture", O-RAN.WG2.Non-RT-RIC-ARCH-TS-v01.00.02 , July 2021.

[RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, DOI 10.17487/RFC4655, August 2006, <https://www.rfc-editor.org/info/rfc4655>.

[RFC6241] Enns, R., Ed., Bjorklund, M., Ed., Schoenwaelder, J., Ed., and A. Bierman, Ed., "Network Configuration Protocol (NETCONF)", RFC 6241, DOI 10.17487/RFC6241, June 2011, <https://www.rfc-editor.org/info/rfc6241>.

[RFC7149] Boucadair, M. and C. Jacquenet, "Software-Defined Networking: A Perspective from within a Service Provider Environment", RFC 7149, DOI 10.17487/RFC7149, March 2014, <https://www.rfc-editor.org/info/rfc7149>.

[RFC7285] Alimi, R., Ed., Penno, R., Ed., Yang, Y., Ed., Kiesel, S., Previdi, S., Roome, W., Shalunov, S., and R. Woundy, "Application-Layer Traffic Optimization (ALTO) Protocol", RFC 7285, DOI 10.17487/RFC7285, September 2014, <https://www.rfc-editor.org/info/rfc7285>.

[RFC7491] King, D. and A. Farrel, "A PCE-Based Architecture for Application-Based Network Operations", RFC 7491, DOI 10.17487/RFC7491, March 2015, <https://www.rfc-editor.org/info/rfc7491>.

[RFC7971] Stiemerling, M., Kiesel, S., Scharf, M., Seidel, H., and S. Previdi, "Application-Layer Traffic Optimization (ALTO) Deployment Considerations", RFC 7971, DOI 10.17487/RFC7971, October 2016, <https://www.rfc-editor.org/info/rfc7971>.

[TR-459] "Control and User Plane Separation for a disaggrgated BNG", Broadband Forum TR-459 , June 2020.

[TS29.522] "TS 29.522 Network Exposure Function Northbound APIs V16.9.0.", 3GPP TS 29.522 V16.9.0 , September 2021.

Acknowledgments

...

Author's Address

Luis M. Contreras
Telefonica
Ronda de la Comunicacion, s/n
Sur-3 building, 3rd floor
Madrid 28050
Spain

Email: luismiguel.contrerasmurillo@telefonica.com
URI: http://lmcontreras.com/