ALTO
Internet-Draft
Intended status: Standards Track
Expires: September 5, 2011

J. Medved
D. Ward
Juniper Networks
J. Peterson
Neustar
R. Woundy
Comcast Corporation
D. McDysan
Verizon

March 04, 2011

ALTO Network-Server and Server-Server APIs
draft-medved-alto-svr-apis-00

Abstract

ALTO servers require automated operation, where the topology of the underlying networks is incorporated into network maps automatically. In addition to the Client-to-Server API defined in the ALTO protocol document, two more standardized API are required: an API between the ALTO Server and networking nodes (e.g. routers), through which the ALTO Server can get topology information from the network, and an API between the ALTO Servers, through which they can exchange topology and status information between themselves.

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 September 5, 2011.

Table of Contents

1. Introduction

ALTO Servers are becoming increasingly important technology for finding "the best" or "most preferred" content or server. For example, an ALTO Server can be used to facilitate the selection of the best cache in a CDN, the best set of peers for a P2P node ([RFC5632] or [I-D.lee-alto-chinatelecom-trial]), or the best service instance in a cloud. These use cases will require that network and cost map information accurately reflects the actual network topology and utilization. Static configuration of network and cost maps is not feasible even for moderately sized networks. Therefore, creation of network and cost maps in the ALTO Server should be automated and policy driven.

The ALTO Server can use multiple sources of information to generate the network and cost maps. Network topology data coming directly from routers is required. Additionally, traffic engineering data, geo location data, or network resource utilization data could also be used to further refine the maps, or to generate different maps for different clients. The ALTO Server should use well defined APIs to get the data required to generate maps, since the data will be obtained from different sources provided by a multitude of vendors, and vendor inter-operability will be critical for adoption of ALTO-based solutions. For network topology data, this draft proposes BGP with TE extensions as the ALTO Server-to-Network API.

The ALTO Server will typically only have partial topology data, which will depend on the Server’s location and the sources from which it obtains data to generate the network and cost maps. To obtain a full view of the network topology, the ALTO Server will have to exchange topology data with other ALTO Servers, or redirect Endpoint Cost ranking requests to the best possible ALTO Server. Therefore, a standard Server-to-Server API is also required.

2. Scope

The scope of this draft are the ALTO Server-to-Network APIs and Server-to-Server API that are required for automated operation of the ALTO Service. The Server-to-Network API is used to obtain network topology information from the underlying network. Server-to-Server API is used to exchange topology information between ALTO servers, or to redirect ranking requests from one ALTO Server to another. The ALTO Client-to-Server protocol [I-D.ietf-alto-protocol] itself may be used as the ALTO Server-to-Server protocol; in other words, one ALTO Server may request maps or status from other servers.

3. Terminology

We use the following terms defined in ALTO Problem Statement [RFC5693]: Application, ALTO Service, ALTO Server, ALTO Client, ALTO Query, ALTO Reply, ALTO Transaction.

4. ALTO Server API Reference

In addition to the ALTO protocol, which constitutes the API between the ALTO Server and its clients, the ALTO Server needs several other APIs to get data that are required to generate the network and cost maps. The reference diagram of possible ALTO Server APIs is shown in Figure 1.

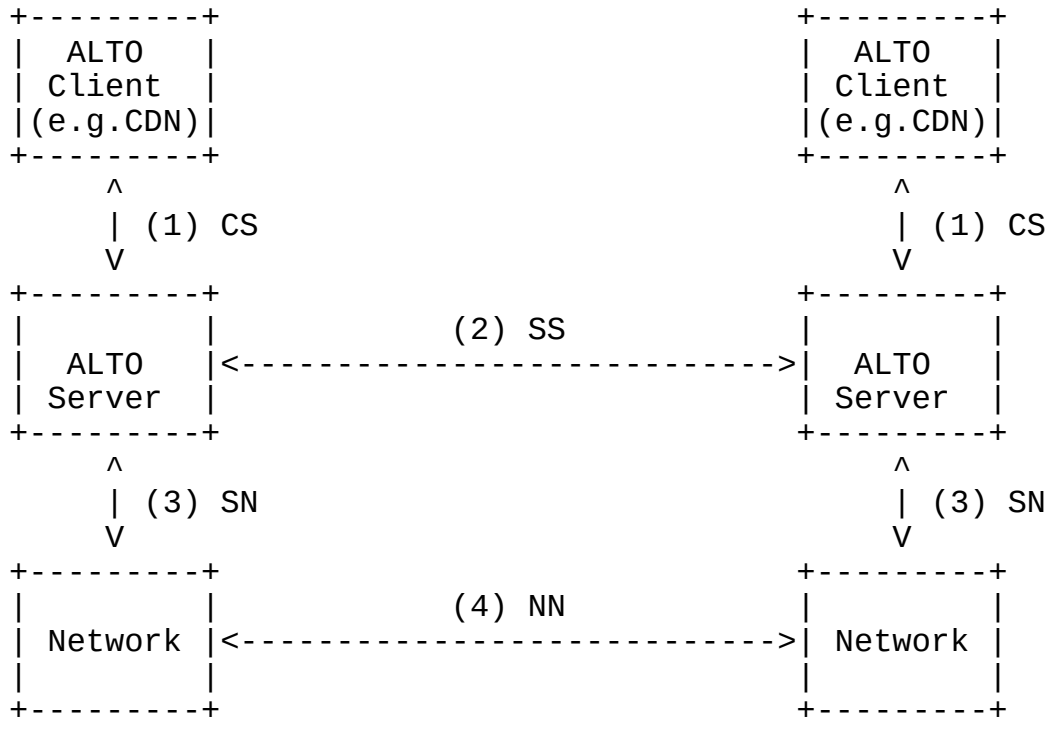

Figure 1: ALTO Server API reference

The ALTO Server interfaces shown in Figure 1 are as follows:

1. CS: The Client-to-Server interface has been the focus of the ALTO WG, and is defined in [I-D.ietf-alto-protocol].

2. SS: The Server-to-Server interface is required to exchange topology data and status between ALTO servers in different networks or administrative domains. For Endpoint Cost queries, the interface can be used to direct the client's request to the

peer ALTO Server that has the best data to respond to the query. The interface may also facilitate other functions, such as ALTO Server discovery.

3. SN: The Server-to-Network interface is used to get the network topology data from the network.

4. NN: The Network-to-Network routing and data interfaces are well-defined in a number of standards (for example, BGP [RFC4271]), and they are not in scope of this draft.

4.1. The ALTO Server-to-Network Interface

4.1.1. Requirements

The Server-to-Network interface should satisfy the following requirements:

- Enable automation of the operation of the ALTO server with minimal human intervention
- Leverage existing sources of network topology data; don’t introduce new (routing) protocols; don’t force un-natural deployment of routing protocols within the ISP network
- Leverage scalable mechanisms for (near real-time) network topology acquisition; don’t use fragile mechanisms to obtain data (e.g. screen-scraping information from looking glass servers)
- Enable centralized and/or distributed deployments of ALTO servers
- Provide network topology information from within the ISP network (intra-AS) as well as from outside the ISP network (inter-AS), as well as from different intra-domain routing areas. (Note that some ISPs use multiple AS’s for different components of the overall network topology.)
- Enable automated ALTO server policy controls above and beyond mere routing metrics
- Provide origin security for network topology information
- Provide the right balance between frequency of updates and accuracy /timeliness of the data. Topology updates from the network should be throttled. For ALTO application, a 15 minute time interval between topology updates from the network should be sufficient.

In addition to having a standardized Server-to-Network interface, the algorithms for generation of ALTO network / cost maps and for endpoint ranking should be normalized as well, to facilitate inter-operability of different ALTO Server implementations.

4.1.2. BGP with TE Extensions

Network topology is best conveyed through routing protocols. BGP carries information about all subnets in the network, and subnet / prefix data from BGP is required to generate ALTO network maps. Intra-AS topology information that is carried in link-state IGPs and inter-AS topology information carried in BGP is required to generate ALTO cost maps. IGP TE data is required if costs in the cost maps have a link utilization component.

This draft proposes to use BGP with TE extensions [I-D.gredler-bgp-te] as the ALTO Server-to-Network API that can carry both the subnet/prefix data for network map generation and the topology data for cost map generation. A BGP Speaker can learn a part or the entire intra-AS topology by participating in the IGP and then distribute the learned topology to other BGP Speakers in the AS. The ALTO Server establishes an iBGP session with a BGP speaker within the AS, typically a Route Reflector, and learns the intra-AS topology from its peer BGP speaker, along with the inter-AS topology and the subnet/prefix data.

Using BGP with TE extensions as the ALTO Server-to-Network API has several advantages:

- Avoid peering with IGP routers, which is more challenging than BGP peering. Moreover, IS-IS, OSPF and EIGRP implementations would be required, although only one IGP peering implementation would typically be used at any given time.

- Unified interface to the network (single protocol), which carries all the network information required to generate the topological component of network and cost maps. The alternative would be for the ALTO Server to interface - in addition to BGP - with IS-IS, OSPF and EIGRP routing protocols.

- Simplified handling of multi-area IGP topologies: if the ALTO Server wants to see the entire multi-area IGP topology, it would need to peer with at least one IGP router in each area. Since the ALTO Server would have to reside in one of the areas, it would have to peer with IGP routers in other areas over GRE tunnels, which is complex and potentially error prone. Alternatively, an ALTO Server would have to be placed in each area, and the ALTO Servers would have to exchange topology information between

themselves via the Server-to-Server API.

- The ALTO Server can peer with a BGP Route Reflector. Route Reflectors are widely deployed, and the Route Reflector control architecture dovetails nicely with the desired ALTO Server control architecture.

- BGP policy and marking capabilities allow the operator to modify or filter / adjust both the prefix and the connectivity information specifically for the ALTO Server's use. This capability is important if the BGP Speaker and the ALTO Server are in different administrative domains.

- BGP has some origin security. This capability is important if the BGP Speaker and the ALTO Server are in different administrative domains.

- BGP carries multicast for future enhancements, where the ALTO Server will be creating multicast network and cost maps.

- Using BGP with TE extensions means that there only needs to be one BGP speaker in each area (or two for redundancy) that gets the area's topology from local IGP routers. The topology information is then distributed throughout the AS and relayed to all interested ALTO Servers. The topology information can be appropriately tagged so that is only stored by those Route Reflectors that talk to ALTO Servers. BGP Input and Output filtering could ensure that only the minimum set of BGP Speakers would need to store the topology information.

- The ALTO Server only needs to peer with a single BGP Speaker to get the entire network topology.

- BGP with TE extensions can be used between eBGP peers to advertise intra-AS topology information between peers in different ASes. Intra-AS topology information from multiple ASes can then be used by an ALTO Server to create more detailed network and cost maps for the combined network.

Due to policy and security considerations, it is assumed that an ALTO Server speaks via the Server-to-Network APIs only to a BGP Speaker in the same Administrative Domain (that may encompass multiple IGP areas and ASes). Any other use cases are for further study.

Note that the network topology received by the ALTO Server must not be summarized beyond what is expressed by the IGP in each area. This is because the network (router) does not understand the application-specific constraints of the ALTO Server for suitable summarization.

Also, where different scaling of metrics and different policies exist inside an Administrative Domain, the Alto Server is instructed via management on how to compare or normalize the data received from the network. The network is not expected to provide translation or normalization.

4.2. The ALTO Server-to-Server Interface

The ALTO Server-to-Server API is required because each ALTO Server will likely have only a partial view of the overall network. The ALTO Server's view of the network depends on which routers are the sources of its topology data. Each router's topology data depends on the administrative domain (Autonomous System) where the router is deployed. In order to generate a combined network/cost map that covers the network beyond its own Autonomous System, an ALTO Server needs to exchange its map information with other ALTO Servers in other network locations and/or administrative domains. To allow generation of combined maps, costs in partial cost maps must be normalized.

The network and cost maps defined in the Client-to-Server ALTO interface provide sufficient semantics to be considered a good candidate for the Server-to-Server information exchange. In other words, the ALTO Client-to-Server interface can be used for communication between ALTO Servers as well.

Note that the idea of sharing information directly between ALTO clients has already been anticipated, as stated in Section 3.1.4 in ALTO Requirements [I-D.ietf-alto-reqs]:

REQ. ARv07-31: The ALTO client protocol SHOULD allow the ALTO server to add information about appropriate modes of re-use to its ALTO responses. Re-use may include redistributing an ALTO response to other parties, as well as using the same ALTO information in a resource directory to improve the responses to different resource consumers, within the specified lifetime of the ALTO response...

Also, although not a formal part of the ALTO protocol, support for redistribution of ALTO data between clients has been anticipated in the ALTO Protocol specification [I-D.ietf-alto-protocol] - see Sections 6.2 and 8. Sharing data between ALTO Servers is similar, but not the same.

Typically, an ALTO Server will handle requests for different services. Moreover, the level of trust between different ALTO Servers can vary. Therefore, topology passed via the Server-to-Server API may be summarized, aggregated, or incomplete as long as they are sufficient to meet the requirements implied by the client's

request.

5. Conclusion

Having well-defined standard APIs will facilitate inter-operation between ALTO Servers and the different sources of information that are required to put together the maps. It will also facilitate inter-operation between the ALTO Servers themselves. Multiple ALTO Servers in different administrative domains may be required to combine partial network maps / cost maps into an overall set of maps that cover a larger multi-provider network or the whole internet. Altogether, having standardized APIs will facilitate inter-operability between ALTO Servers from different vendors.

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

ALTO offers advice to applications on the optimality of various possible Internet destinations for acquiring a resource or service. An attacker who subverts or impersonates an ALTO service might be able to trick many application on the Internet into contacting the same host as a part of a distributed denial of service attack, for example. Interfaces that provision the back-end of ALTO servers are therefore a potentially attractive to attackers, as attackers might attempt to corrupt the ALTO database in order to launch such an attack.

For an ALTO server back-end interface to accept topology data from BGP, the server must trust the source of the information. The ALTO server must peer with a known route reflector, and must authenticate that entity, especially if it is outside the administrative domain of the ALTO server. Any origin security mechanisms will also increase the assurance of the ALTO server. Integrity protection for the channel between the ALTO server and the BGP speaker will also prevent malicious parties from inserting problem information.

Similarly, the ALTO server-to-server mechanism also requires an authentication and data integrity mechanism. If ALTO servers share network maps between one another, for example, assuring the

authenticity and source of data is essential. If ALTO servers share network maps with one another over a public network, a confidentiality mechanism will also be desirable in order to prevent eavesdropping.

8. Acknowledgements

Hannes Gredler from Juniper Networks made significant contributions to concepts presented in this draft. We would like to thank Alia Atlas from Juniper Networks for her input and comments.

9. References

9.1. Normative References

[I-D.gredler-bgp-te]
Gredler, H. and J. Medved, "Advertising Traffic Engineering Information in BGP", draft-gredler-bgp-te-00 (work in progress), March 2011.

[I-D.ietf-alto-protocol]
Alimi, R., Penno, R., and Y. Yang, "ALTO Protocol", draft-ietf-alto-protocol-06 (work in progress), October 2010.

[I-D.ietf-alto-reqs]
Kiesel, S., Previdi, S., Stiemerling, M., Woundy, R., and Y. Yang, "Application-Layer Traffic Optimization (ALTO) Requirements", draft-ietf-alto-reqs-07 (work in progress), January 2011.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC4271] Rekhter, Y., Li, T., and S. Hares, "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, January 2006.

[RFC5693] Seedorf, J. and E. Burger, "Application-Layer Traffic Optimization (ALTO) Problem Statement", RFC 5693, October 2009.

9.2. Informative References

[I-D.lee-alto-chinatelecom-trial]
Li, K. and G. Jian, "ALTO and DECADE service trial within China Telecom", draft-lee-alto-chinatelecom-trial-01 (work

in progress), October 2010.

[RFC5632] Griffiths, C., Livingood, J., Popkin, L., Woundy, R., and Y. Yang, "Comcast's ISP Experiences in a Proactive Network Provider Participation for P2P (P4P) Technical Trial", RFC 5632, September 2009.

Authors' Addresses

Jan Medved
Juniper Networks
1194 N. Mathilda Ave.
Sunnyvale, CA 94089
US

Email: jmedved@juniper.net

David Ward
Juniper Networks
1194 N. Mathilda Ave.
Sunnyvale, CA 94089
US

Email: dward@juniper.net

Jon Peterson
Neustar

Email: jon.peterson@neustar.biz

Richard Woundy
Comcast Corporation
27 Industrial Avenue
Chelmsford, MA 01824
US

Email: Richard_Woundy@cable.comcast.com

David McDysan
Verizon
22001 Loudoun County Pkwy
Ashburn, VA 20147
US

Email: dave.mcdysan@verizon.com