Internet Engineering Task Force (IETF)
Request for Comments: 6708
Category: Informational
ISSN: 2070-1721

S. Kiesel, Ed.
University of Stuttgart
S. Previdi
Cisco Systems, Inc.
M. Stiemerling
NEC Europe Ltd.
R. Woundy
Comcast Corporation
Y. Yang
Yale University
September 2012

Application-Layer Traffic Optimization (ALTO) Requirements

Abstract

Many Internet applications are used to access resources, such as pieces of information or server processes that are available in several equivalent replicas on different hosts. This includes, but is not limited to, peer-to-peer file sharing applications. The goal of Application-Layer Traffic Optimization (ALTO) is to provide guidance to applications that have to select one or several hosts from a set of candidates capable of providing a desired resource. This guidance shall be based on parameters that affect performance and efficiency of the data transmission between the hosts, e.g., the topological distance. The ultimate goal is to improve performance or Quality of Experience in the application while reducing the utilization of the underlying network infrastructure.

This document enumerates requirements for specifying, assessing, or comparing protocols and implementations.

Status of This Memo

This document is not an Internet Standards Track specification; it is published for informational purposes.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Not all documents approved by the IESG are a candidate for any level of Internet Standard; see Section 2 of RFC 5741.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at http://www.rfc-editor.org/info/rfc6708.

Table of Contents

1. Introduction

The motivation for Application-Layer Traffic Optimization (ALTO) is described in the ALTO problem statement [RFC5693].

The goal of ALTO is to provide information that can help peer-to-peer (P2P) applications make better decisions with respect to peer selection. However, ALTO may be useful for non-P2P applications as well. For example, clients of client-server applications may use information provided by ALTO to select one of several servers or information replicas. As another example, ALTO information could be used to select a media relay needed for NAT traversal. The goal of these informed decisions is to improve performance or Quality of Experience in the application while reducing the utilization of the underlying network infrastructure.

Usually, it would be difficult or even impossible for application entities to acquire this information by other mechanisms, e.g., using measurements between the peers of a P2P overlay, because of complexity or because it is based on network topology information, network operational costs, or network policies, which the respective network provider does not want to disclose in detail.

The functional entities that provide the ALTO service do not take part in the actual user-data transport, i.e., they do not implement functions for relaying user data. These functional entities may be placed on various kinds of physical nodes, e.g., on dedicated servers, as auxiliary processes in routers, on "trackers" or "super peers" of a P2P application, etc.

2. Terminology and Architectural Framework

2.1. Requirements Notation

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 [RFC2119].

2.2. ALTO Terminology

This document uses the following ALTO-related terms, which are defined in [RFC5693]:

Application, Peer, P2P, Resource, Resource Identifier, Resource Provider, Resource Consumer, Transport Address, Overlay Network, Resource Directory, ALTO Service, ALTO Server, ALTO Client, ALTO

Query, ALTO Response, ALTO Transaction, Local Traffic, Peering Traffic, Transit Traffic, Application Protocol, ALTO Client Protocol, and Provisioning Protocol.

Furthermore, the following additional terms will be used:

- Host-Group Descriptor: Information used to describe one or more Internet hosts (such as the resource consumer that seeks ALTO guidance, or one or more candidate resource providers) and their location within the network topology. There can be several different types of host-group descriptors, for example, a single IP address, an address prefix or address range that contains the host(s), or an Autonomous System (AS) number. Different host-group descriptor types may provide different levels of detail. Depending on the system architecture, this may have implications on the quality of the guidance ALTO is able to provide, on whether recommendations can be aggregated, and on how much privacy-sensitive information about users might be disclosed to additional parties.

- Rating Criterion: The condition or relation that defines the "better" in "better-than-random peer selection", which is the ultimate goal of ALTO. Examples may include "host's Internet access is not subject to volume-based charging (flat rate)" or "low topological distance". Some rating criteria, such as "low topological distance", need to include a reference point, e.g., "low topological distance from a given resource consumer". This reference point can be described by means of a host-group descriptor.

- Host-Characteristics Attribute: Properties of a host, other than the host-group descriptor. It may be evaluated according to one or more rating criteria. This information may be stored in an ALTO server and transmitted via an ALTO protocol. One example for a host-characteristics attribute would be a data field indicating whether a host's Internet access is subject to volume-based charging or not (flat rate).

- Target-Aware Query Mode: In this mode of operation, an ALTO client performs the ALTO query when the desired resource and a set of candidate resource providers are already known, i.e., after Distributed Hash Table (DHT) lookups, queries to the resource directory, etc. To this end, the ALTO client transmits a list of host-group descriptors and optionally one or more rating criteria to the ALTO server. The ALTO server evaluates the host-group descriptors according to the indicated criteria or a default

criterion. It returns a list of these host-group descriptors to the ALTO client, which is sorted according to the rating criteria and/or enriched with host-characteristics attributes.

- Target-Independent Query Mode: In this mode of operation, ALTO queries are performed in advance or periodically, in order to receive comprehensive guidance. The ALTO client indicates the desired host-characteristics attributes in the ALTO query. The ALTO server answers with a list that indicates for all known host-group descriptors (possibly subject to the server's policies) the desired host-characteristics attributes. These lists will be cached locally and evaluated later, when a resource is to be accessed.

2.3. Architectural Framework for ALTO

There are various architectural options for ALTO implementation. Specifying or mandating one specific architecture is out of the scope of this document.

In addition to the terminology (see Section 2 of [RFC5693] and Section 2.2 of this document), [RFC5693] presents a figure that gives a high-level overview of protocol interaction between these components.

This document itemizes requirements for the following components: ALTO client protocols, ALTO server discovery mechanisms, host-group descriptors, rating criteria, and host-characteristics attributes. Furthermore, requirements regarding the overall architecture, especially with respect to security and privacy issues, are presented.

Note that the detailed specification of such protocols and mechanisms is out of the scope of this document. In fact, this document does not even assume that there will be only one single specification for each of these components, respectively. However, this document enumerates requirements for ALTO to be considered when specifying, assessing, or comparing protocols and implementations.

3. ALTO Requirements

3.1. ALTO Client Protocol

3.1.1. General Requirements

Req. AR-1: The ALTO service is provided by one or more ALTO servers. It may be queried by ALTO clients seeking guidance for selecting appropriate resource providers. ALTO clients and ALTO servers MUST

implement an ALTO client protocol. An ALTO client protocol MUST be able to transmit ALTO queries from an ALTO client to an ALTO server, and it MUST be able to transmit the corresponding ALTO replies from the ALTO server to the ALTO client.

The detailed specification of an ALTO client protocol is out of the scope of this document. In fact, this document does not even assume that there will be only one single protocol specification. However, this document enumerates requirements for ALTO, to be considered when specifying, assessing, or comparing protocols and implementations.

Req. AR-2: An ALTO client protocol MUST provide adequate mechanisms for operations and management support, as outlined in RFC 5706 [RFC5706].

3.1.2. Host-Group Descriptor Support

The ALTO guidance is based on the evaluation of several resource providers or groups of resource providers, considering one or more rating criteria. The resource providers or groups of resource providers are characterized by means of host-group descriptors.

Req. AR-3: An ALTO client protocol MUST support the usage of multiple host-group descriptor types.

Req. AR-4: ALTO clients and ALTO servers MUST clearly identify the type of each host-group descriptor sent in ALTO queries or responses. An ALTO protocol specification MUST provide appropriate protocol elements.

Req. AR-5: An ALTO client protocol MUST support the host group descriptor types "IPv4 address prefix" and "IPv6 address prefix". They can be used to specify the IP address of one host, or an IP address range (in Classless Inter-Domain Routing (CIDR) notation) containing all hosts in question.

Req. AR-6: An ALTO client protocol MUST be extensible to enable future support of other host-group descriptor types. An ALTO client protocol specification MUST define an appropriate procedure for adding new host-group descriptor types, e.g., by establishing an IANA registry.

Req. AR-7: For host-group descriptor types other than "IPv4 address prefix" and "IPv6 address prefix", the host-group descriptor type identification MUST be supplemented by a reference to a facility that can be used to translate host-group descriptors of this type to IPv4/IPv6 address prefixes, e.g., by means of a mapping table or an algorithm.

Req. AR-8: Protocol functions for mapping other host-group descriptor types to IPv4/IPv6 address prefixes SHOULD be designed and specified as part of an ALTO client protocol, and the corresponding address mapping information SHOULD be made available by the same entity that wants to use these host-group descriptors within an ALTO client protocol. However, an ALTO server or an ALTO client MAY also send a reference to an external mapping facility, e.g., a translation table to be obtained via an alternative mechanism.

Rationale for the previous two requirements: The preferred type of host-group descriptors are IPv4 and IPv6 prefixes. However, in some situations, one party may prefer to use another type, e.g., AS numbers. Usually, applications seeking ALTO guidance work with IP addresses, e.g., when establishing connections. Understanding guiding information that is based on other host-group descriptor types, i.e., mapping from these other types to IP prefixes and back, may be a non-trivial task. Therefore, before a party may use other host-group descriptor types, they must provide a mapping mechanism to IP prefixes.

Req. AR-9: An ALTO client protocol specification MUST define mechanisms that can be used by the ALTO server to indicate that a host-group descriptor used by the ALTO client is of an unsupported type, or that the indicated mapping mechanism could not be used.

Req. AR-10: An ALTO client protocol specification MUST define mechanisms that can be used by the ALTO client to indicate that a host-group descriptor used by the ALTO server is of an unsupported type, or that the indicated mapping mechanism could not be used.

3.1.3. Rating Criteria Support

Req. AR-11: An ALTO client protocol specification MUST define a rating criterion that can be used to express and evaluate the "relative operator’s preference". This is a relative measure, i.e., it is not associated with any unit of measurement. A preferred rating, according to this criterion, indicates that the application should prefer the respective candidate resource provider over others with less preferred ratings (unless information from non-ALTO sources suggests a different choice, such as transmission attempts suggesting that the path is currently congested). The operator of the ALTO server does not have to disclose how and based on which data the ratings are actually computed. Examples could be: cost for peering or transit traffic, traffic engineering inside the network, and other policies.

Req. AR-12: An ALTO client protocol MUST be extensible to enable future support of other rating criteria types. An ALTO client protocol specification MUST define an appropriate procedure for adding new rating criteria types, e.g., by establishing an IANA registry.

Req. AR-13: ALTO client protocol specifications MUST NOT define rating criteria closely related to the instantaneous network congestion state, i.e., rating criteria that have the primary aim to serve as an alternative to established congestion control strategies, such as using TCP-based transport.

Req. AR-14: Applications using ALTO guidance MUST NOT rely solely on the ALTO guidance to avoid causing network congestion. Instead, they MUST use other appropriate means, such as TCP-based transport, to avoid causing excessive congestion.

> Rationale for the previous requirement: One design assumption for ALTO is that it is acceptable for the host-characteristics attributes, which are stored and processed in the ALTO servers for giving guidance, to be updated rather infrequently. Typical update intervals may be several orders of magnitude longer than the typical network-layer packet round-trip time (RTT). Therefore, ALTO cannot be a replacement for TCP-like congestion control mechanisms.

Req. AR-15: In the target-independent query mode, the ALTO query message SHOULD allow the ALTO client to express which host-characteristics attributes should be returned.

Req. AR-16: In the target-aware query mode, the ALTO query message SHOULD allow the ALTO client to express which rating criteria should be considered by the server, as well as their relative relevance for the specific application that will eventually make use of the guidance. The corresponding ALTO response message SHOULD allow the ALTO server to express which rating criteria have been considered when generating the response.

Req. AR-17: An ALTO client protocol specification MUST define mechanisms that can be used by the ALTO client and the ALTO server to indicate that a rating criteria used by the other party is of an unsupported type.

3.1.4. Placement of Entities and Timing of Transactions

With respect to the placement of ALTO clients, several modes of operation exist:

- One mode of ALTO operation is that an ALTO client may be embedded directly in the resource consumer, i.e., the application protocol entity that will eventually initiate data transmission to/from the selected resource provider(s) in order to access the desired resource. For example, an ALTO client could be integrated into the peer of a P2P application that uses a distributed algorithm such as "query flooding" for resource discovery.

- Another mode of operation is to integrate the ALTO client into a third party, such as a resource directory. This third party may issue ALTO queries to solicit preference on potential resource providers, considering the respective resource consumer. For example, an ALTO client could be integrated into the tracker of a tracker-based P2P application, in order to request ALTO guidance on behalf of the peers contacting the tracker.

Req. AR-18: An ALTO client protocol MUST support the mode of operation in which the ALTO client is directly embedded in the resource consumer.

Req. AR-19: An ALTO client protocol MUST support the mode of operation in which the ALTO client is embedded in a third party. This third party performs queries on behalf of resource consumers.

Req. AR-20: An ALTO client protocol MUST be designed in a way that the ALTO service can be provided by an entity that is not the operator of the underlying IP network.

Req. AR-21: An ALTO client protocol MUST be designed in a way that different instances of the ALTO service operated by different providers can coexist.

Req. AR-22: An ALTO client protocol specification MUST specify at least one query mode, either the target-aware or the target-independent query mode.

Note that this requirements document does not assume that there will be only one single protocol specification.

Req. AR-23: An ALTO client protocol specification SHOULD specify both the target-aware and the target-independent query mode. If an ALTO client protocol specification specifies more than one query mode, it MUST define at least one of these modes as REQUIRED to implement by

ALTO clients and ALTO servers. Furthermore, it MUST specify an appropriate protocol mechanism for negotiating between the ALTO client and ALTO server, which query mode to use.

Req. AR-24: An ALTO client protocol SHOULD support version numbering, TTL (time-to-live) attributes, and/or similar mechanisms in ALTO transactions, in order to enable time validity checking for caching, and to enable comparisons of multiple recommendations obtained through redistribution.

Req. AR-25: An ALTO client protocol SHOULD allow the ALTO server to add information about appropriate modes of reuse to its ALTO responses. Reuse 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. The ALTO server SHOULD be able to express that

- no reuse should occur.
- reuse is appropriate for a specific "target audience", i.e., a set of resource consumers explicitly defined by a list of host-group descriptors. The ALTO server MAY specify a "target audience" in the ALTO response that is only a subset of the known actual "target audience", e.g., if required by operator policies.
- reuse is appropriate for any resource consumer that would send (or cause a third party to send on behalf of it) the same ALTO query (i.e., with the same query parameters, except for the resource consumer ID, if applicable) to this ALTO server.
- reuse is appropriate for any resource consumer that would send (or cause a third party to send on behalf of it) the same ALTO query (i.e., with the same query parameters, except for the resource consumer ID, if applicable) to any other ALTO server that was discovered (using an ALTO discovery mechanism) together with this ALTO server.
- reuse is appropriate for any resource consumer that would send (or cause a third party to send on behalf of it) the same ALTO query (i.e., with the same query parameters, except for the resource consumer ID, if applicable) to any ALTO server in the whole network.

Req. AR-26: An ALTO client protocol MUST support the transport of ALTO transactions, even if the ALTO client is located in the private address realm behind a network address translator (NAT). There are different types of NAT, see [RFC4787] and [RFC5382].

3.1.5. Protocol Extensibility

Req. AR-27: An ALTO client protocol MUST include support for adding protocol extensions in a non-disruptive, backward-compatible way.

Req. AR-28: An ALTO client protocol MUST include protocol versioning support, in order to clearly distinguish between incompatible versions of the protocol.

3.1.6. Error Handling and Overload Protection

Req. AR-29: An ALTO client protocol MUST use congestion-aware transport, e.g., by using TCP.

Req. AR-30: An ALTO client protocol specification MUST specify mechanisms for an ALTO server to inform clients about an impending or occurring overload situation, or how to leverage appropriate mechanisms provided by underlying protocol layers. The mechanisms MUST provide all of the following options to the server:

- terminate the conversation with the client,
- redirect the client to another ALTO server, and
- request that the client throttle its query rate.

  In particular, a simple form of throttling is to let an ALTO server answer a query with an error message advising the client to retry the query later (e.g., using a protocol function such as HTTP's Retry-After header ([RFC2616], Section 14.37)). Another simple option is to actually answer the query with the desired information, but adding an indication that the ALTO client should not send further queries to this ALTO server before an indicated period of time has elapsed.

Req. AR-31: An ALTO client protocol specification MUST specify mechanisms for an ALTO server to inform clients about its inability to answer queries due to technical problems or system maintenance, or how to leverage appropriate mechanisms provided by underlying protocol layers. The mechanisms MUST provide all of the following options to the server:

- terminate the conversation with the client,
- redirect the client to another ALTO server, and
- request that the client retry the query later.

Note: The existence of the above-mentioned protocol mechanisms does not imply that an ALTO server must use them when facing an overload, technical problem, or maintenance situation, respectively. Some servers may be unable to use them in that situation, or they may prefer to simply refuse the connection or not to send any answer at all.

3.2. ALTO Server Discovery

An ALTO client protocol is supported by one or more ALTO server discovery mechanisms, which may be used by ALTO clients to determine one or more ALTO servers, to which ALTO requests can be sent. This section enumerates requirements for an ALTO client, as well as general requirements to be fulfilled by the ALTO server discovery mechanisms.

Req. AR-32: An ALTO server discovery mechanism MUST support features allowing ALTO clients that are embedded in the resource consumer to find one or several ALTO servers that can provide ALTO guidance suitable for the resource consumer, using an ALTO protocol version compatible with the ALTO client. This mode of operation is called "resource consumer initiated ALTO server discovery".

Req. AR-33: An ALTO server discovery mechanism MUST support features allowing ALTO clients that are embedded in a resource directory and perform third-party ALTO queries on behalf of a remote resource consumer to find one or several ALTO servers that can provide ALTO guidance suitable for the respective resource consumer, using an ALTO protocol version compatible with the ALTO client. This mode of operation is called "third-party ALTO server discovery".

Req. AR-34: ALTO clients MUST be able to perform resource consumer initiated ALTO server discovery, even if they are located behind a NAT.

Req. AR-35: ALTO clients MUST be able to perform third-party ALTO server discovery, even if they are located behind a NAT.

Req. AR-36: ALTO clients MUST be able to perform third-party ALTO server discovery, even if the resource consumer, on behalf of which the ALTO query will be sent, is located behind a NAT.

Req. AR-37: ALTO server discovery mechanisms SHOULD leverage an existing protocol or mechanism, such as DNS-, DHCP-, or PPP-based automatic configuration, etc. A single mechanism with a broad spectrum of applicability SHOULD be preferred over several different mechanisms with narrower scopes.

Req. AR-38: Every ALTO server discovery mechanism SHOULD be able to return the respective contact information for multiple ALTO servers.

Req. AR-39: Every ALTO server discovery mechanism SHOULD be able to indicate preferences for each returned ALTO server contact information.

3.3. Security and Privacy

Note: The following requirements mandate the inclusion of certain security mechanisms at a protocol specification level. Whether it makes sense to enable these mechanisms in a given deployment scenario depends on a threat analysis for this specific scenario. For a classification of potential information disclosure risks, refer to Section 5.2.

Req. AR-40: An ALTO client protocol specification MUST specify mechanisms for the authentication of ALTO servers or specify how to leverage appropriate mechanisms provided by underlying protocol layers.

Req. AR-41: An ALTO client protocol specification MUST specify mechanisms for the authentication of ALTO clients or specify how to leverage appropriate mechanisms provided by underlying protocol layers.

Req. AR-42: An ALTO client protocol specification MUST specify mechanisms for the encryption of messages or specify how to leverage appropriate mechanisms provided by underlying protocol layers.

Req. AR-43: An ALTO client is not required to implement mechanisms or to comply with rules that limit its ability to redistribute information retrieved from the ALTO server to third parties.

Req. AR-44: An ALTO client protocol MUST support different levels of detail in queries and responses in order to protect the privacy of users, to ensure that the operators of ALTO servers and other users of the same application cannot derive sensitive information.

Req. AR-45: An ALTO client protocol MAY include mechanisms that can be used by the ALTO client when requesting guidance to specify the resource (e.g., content identifiers) it wants to access. An ALTO server MUST provide adequate guidance, even if the ALTO client prefers not to specify the desired resource (e.g., keeps the data field empty). The mechanism MUST be designed in a way that the operator of the ALTO server cannot easily deduce the resource identifier (e.g., file name in P2P file sharing) if the ALTO client prefers not to specify it.

Req. AR-46: An ALTO client protocol specification MUST specify appropriate mechanisms for protecting the ALTO service against Denial-of-Service (DoS) attacks or specify how to leverage appropriate mechanisms provided by underlying protocol layers.

4. IANA Considerations

This requirements document does not mandate any immediate IANA actions. However, such IANA considerations may arise from future ALTO specification documents that try to meet the requirements given here.

5. Security Considerations

5.1. High-Level Security Considerations

High-level security considerations for the ALTO service can be found in the "Security Considerations" section of the ALTO problem statement document [RFC5693].

5.2. Information Disclosure Scenarios

The unwanted disclosure of information is one key concern related to ALTO. Neither the ALTO server nor a third party using or misusing the ALTO service should be able to infer the application behavior or correlate data in such a way that would violate user privacy, e.g., who is exchanging which files with whom using a P2P file-sharing application. Furthermore, many network operators are concerned about the amount of information related to their network infrastructure (e.g., topology information, number of "premium customers", or utilization statistics) that might be released through ALTO. This section presents a classification and discussion of information disclosure scenarios and potential countermeasures.

5.2.1. Classification of Information Disclosure Scenarios

The following issues may be considered a risk for the operator of an ALTO server, depending on the specific deployment scenario:

(1) Excess disclosure of the ALTO server operator’s data to an authorized ALTO client. The operator of an ALTO server has to feed information, such as tables mapping host-group descriptors to host-characteristics attributes, into the server, thereby enabling it to give guidance to ALTO clients. Some operators might consider the full set of this information confidential (e.g., a detailed map of the operator’s network topology) and might want to disclose only a subset of it or disclose somehow obfuscated information to an ALTO client.

(2) Disclosure of the ALTO server operator's data (e.g., network topology information) to an unauthorized third party. There are three subcases here:

(2a) An ALTO server receives and answers queries originating from an unauthorized ALTO client.

(2b) An unauthorized party snoops on the data transmission from the ALTO server to an authorized ALTO client.

(2c) An authorized ALTO client knowingly forwards the information it has received from the ALTO server to an unauthorized party.

(3) Excess retrieval of the ALTO server operator's data by collaborating ALTO clients. Several authorized ALTO clients could ask one or more ALTO servers for guidance, possibly several times during an extended period of time, and redistribute the responses among each other (see also case 2c). By aggregating and correlating the ALTO responses, they could find out more information than intended to be disclosed by the ALTO server operator(s).

The following issues may be considered a risk for the user of an ALTO client, depending on the specific deployment scenario:

(4) Disclosure of the application behavior or other user private data to the (authorized) ALTO server. The operator of an ALTO server could infer the application behavior (e.g., content identifiers in P2P file sharing applications, or lists of resource providers that are considered for establishing a connection) from the ALTO queries sent by an ALTO client.

(5) Disclosure of the application behavior or other user private data to an unauthorized third party. There are three subcases here:

(5a) An ALTO client willingly sends queries directly to an untrusted or malicious ALTO server, possibly due to a forged response of the ALTO server discovery mechanism.

(5b) An unauthorized party snoops on the data transmission from the ALTO client to an authorized ALTO server.

(5c) An authorized ALTO server knowingly forwards the information it has received from the ALTO client to an unauthorized party.

(6) One or several collaborating (see case 5c) ALTO servers could try to infer the application behavior or other user private data by aggregating and correlating queries from one or more ALTO clients, possibly over an extended period of time.

5.2.2. Discussion of Information Disclosure Scenarios

An ALTO server operator should consider:

- Issue (1) may be addressed by the ALTO server operator choosing the level of detail of the information to be populated into the ALTO server and returned in the responses. For example, by specifying a broader address range (i.e., a shorter prefix length) than a group of hosts in question actually uses, an ALTO server operator may control to some extent how much information about the network topology is disclosed. Furthermore, access control mechanisms for filtering ALTO responses according to the authenticated ALTO client identity might be installed in the ALTO server, although this might not be effective given the lack of efficient mechanisms for addressing (2c) and (3), see below.

- (2a) and (2b) may be addressed by authentication, access control, and encryption schemes for the ALTO client protocol. However, deployment of encryption schemes might not be effective given the lack of efficient mechanisms for addressing (2c) and (3), see below.

- Straightforward authentication and encryption schemes will not help solving (2c) and (3), and there is no other simple and efficient mechanism known. The cost of complex approaches, e.g., based on Digital Rights Management (DRM), might easily outweigh the benefits of the whole ALTO solution; therefore, they are not considered as a viable solution. That is, ALTO server operators must be aware that (2c) and (3) cannot be prevented from happening; therefore, they should feed only such data into an ALTO server that they do not consider sensitive with respect to (2c) and (3).

A user of an ALTO client should consider:

- Issue (4) can and needs to be addressed in several ways: If the ALTO client is embedded in the resource consumer, the resource consumer’s IP address (or the "public" IP address of the outermost NAT in front of the resource consumer) is disclosed to the ALTO server as a matter of principle, because it is in the source address fields of the IP headers. By using a proxy, the disclosure of source addresses to the ALTO server can be avoided at the cost of disclosing them to said proxy. If, in contrast,

the ALTO client is embedded in a third party (e.g., a resource directory), which issues ALTO requests on behalf of resource consumers, it is possible to hide the exact addresses of the resource consumers from the ALTO server, e.g., by zeroing out or randomizing the last few bits of IP addresses. However, there is the potential side effect of yielding inaccurate results.

The disclosure of candidate resource providers' addresses to the ALTO server can be avoided by allowing ALTO clients to use the target-independent query mode. In this mode of operation, guiding information (e.g., "maps") is retrieved from the ALTO server and used entirely locally by the ALTO client, i.e., without sending host-location attributes of candidate resource providers to the ALTO server. In the target-aware query mode, this issue can be addressed by ALTO clients through obfuscating the identity of candidate resource consumers, e.g., by specifying a broader address range (i.e., a shorter prefix length) than a group of hosts in question actually uses, or by zeroing out or randomizing the last few bits of IP addresses. However, there is the potential side effect of yielding inaccurate results.

- (5a) may be addressed by mandating that the ALTO server discovery procedure, as a whole, must be secure against spoofing.

  Note: Given that this document does not mandate a specific system architecture, it is difficult to specify more details than that the discovery procedure, as a whole, should be secure against spoofing. There are many different architectural options, e.g., have an insecure discovery mechanism and use server certificates to later verify its response (cf. the DNS + HTTPS security model widely used in the World Wide Web). Therefore, at this requirements stage, it is not mandatory for the discovery mechanism itself to be secure against spoofing attacks.

- (5b) may be addressed by encryption schemes for the ALTO client protocol. However, the effort vs. benefit should be evaluated for any specific deployment scenario, while also considering the risks and solution approaches for issues (4), (5c), and (6).

- Straightforward authentication and encryption schemes will not help solving (5c) and (6). However, potential risks can be mitigated using the same approaches as used for issue (4), see above.

These insights are reflected in the requirements in this document.

5.3. ALTO Server Discovery

See discussion of (5a) above.

5.4. Security Requirements

For a set of specific security requirements, please refer to Section 3.3 of this document.

6. References

6.1. Normative References

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

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

6.2. Informative References

[RFC2616] Fielding, R., Gettys, J., Mogul, J., Frystyk, H., Masinter, L., Leach, P., and T. Berners-Lee, "Hypertext Transfer Protocol -- HTTP/1.1", RFC 2616, June 1999.

[RFC4787] Audet, F. and C. Jennings, "Network Address Translation (NAT) Behavioral Requirements for Unicast UDP", BCP 127, RFC 4787, January 2007.

[RFC5382] Guha, S., Biswas, K., Ford, B., Sivakumar, S., and P. Srisuresh, "NAT Behavioral Requirements for TCP", BCP 142, RFC 5382, October 2008.

[RFC5706] Harrington, D., "Guidelines for Considering Operations and Management of New Protocols and Protocol Extensions", RFC 5706, November 2009.

Appendix A. Contributors

Early draft versions of this document were co-authored by Laird Popkin.

Appendix B. Acknowledgments

The authors would like to thank Vijay K. Gurbani and Enrico Marocco for fostering discussions that lead to the creation of this document, and for giving valuable comments on it.

The authors would like to thank the members of the P2PI and ALTO mailing lists for contributions and feedback, in particular: Richard Alimi, Jason Livingood, Michael Scharf, Nico Schwan, and Jan Seedorf.

Laird Popkin and Y. Richard Yang are grateful to the many contributions made by the members of the P4P working group and Yale Laboratory of Networked Systems. The P4P working group is hosted by DCIA.

Martin Stiemerling is partially supported by the COAST project (COntent Aware Searching, retrieval and sTreaming, http://www.coast-fp7.eu), a research project supported by the European Commission under its 7th Framework Program (contract no. 248036). The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the COAST project or the European Commission.

Authors' Addresses

Sebastian Kiesel (editor)
University of Stuttgart Computing Center
Networks and Communication Systems Department
Allmandring 30
70550 Stuttgart
Germany

EMail: ietf-alto@skiesel.de
URI: http://www.rus.uni-stuttgart.de/nks/

Stefano Previdi
Cisco Systems, Inc.

EMail: sprevidi@cisco.com

Martin Stiemerling
NEC Laboratories Europe

EMail: martin.stiemerling@neclab.eu
URI: http://ietf.stiemerling.org

Richard Woundy
Comcast Corporation

EMail: Richard_Woundy@cable.comcast.com

Yang Richard Yang
Yale University

EMail: yry@cs.yale.edu