

ALTO WG  
Internet-Draft  
Intended status: Standards Track  
Expires: April 28, 2011

R. Alimi, Ed.  
Google  
R. Penno, Ed.  
Juniper Networks  
Y. Yang, Ed.  
Yale University  
October 25, 2010

ALTO Protocol  
draft-ietf-alto-protocol-06.txt

Abstract

Networking applications today already have access to a great amount of Inter-Provider network topology information. For example, views of the Internet routing table are easily available at looking glass servers and entirely practical to be downloaded by clients. What is missing is knowledge of the underlying network topology from the ISP or Content Provider (henceforth referred as Provider) point of view. In other words, what a Provider prefers in terms of traffic optimization -- and a way to distribute it.

The ALTO Service provides information such as preferences of network resources with the goal of modifying network resource consumption patterns while maintaining or improving application performance. This document describes a protocol implementing the ALTO Service. While such service would primarily be provided by the network (i.e., the ISP), content providers and third parties could also operate this service. Applications that could use this service are those that have a choice in connection endpoints. Examples of such applications are peer-to-peer (P2P) and content delivery networks.

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 [1].

Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/lid-abstracts.txt>.

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on April 28, 2011.

#### Copyright Notice

Copyright (c) 2010 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 BSD License.

## Table of Contents

1. Introduction . . . . .	6
1.1. Background and Problem Statement . . . . .	6
1.2. Design History and Merged Proposals . . . . .	6
1.3. Solution Benefits . . . . .	6
1.3.1. Service Providers . . . . .	6
1.3.2. Applications . . . . .	7
2. Architecture . . . . .	7
2.1. Terminology . . . . .	7
2.1.1. Endpoint Address . . . . .	7
2.1.2. ASN . . . . .	8
2.1.3. Network Location . . . . .	8
2.1.4. ALTO Information . . . . .	8
2.1.5. ALTO Information Base . . . . .	8
2.2. ALTO Service and Protocol Scope . . . . .	8
3. Protocol Structure . . . . .	9
3.1. Server Information Service . . . . .	10
3.2. ALTO Information Services . . . . .	11
3.2.1. Map Service . . . . .	11
3.2.2. Map Filtering Service . . . . .	11
3.2.3. Endpoint Property Service . . . . .	11
3.2.4. Endpoint Cost Service . . . . .	11
4. Network Map . . . . .	11
4.1. PID . . . . .	12
4.2. Endpoint Addresses . . . . .	13
4.2.1. IP Addresses . . . . .	13
4.3. Example Network Map . . . . .	13
5. Cost Map . . . . .	14
5.1. Cost Attributes . . . . .	14
5.1.1. Cost Type . . . . .	14
5.1.2. Cost Mode . . . . .	15
5.2. Cost Map Structure . . . . .	15
5.3. Network Map and Cost Map Dependency . . . . .	16
6. Protocol Design Overview . . . . .	16
6.1. Existing Infrastructure . . . . .	17
6.2. ALTO Information Reuse and Redistribution . . . . .	17
7. Protocol Messaging . . . . .	18
7.1. Notation . . . . .	18
7.2. Message Format . . . . .	18
7.2.1. Protocol Versioning . . . . .	18
7.2.2. Content Type . . . . .	19
7.2.3. Request Message . . . . .	19
7.2.4. Response Message . . . . .	20
7.3. General Processing . . . . .	22
7.4. ALTO Status Codes . . . . .	22
7.5. Client Behavior . . . . .	23
7.5.1. Successful Response . . . . .	23

7.5.2.	Error Conditions . . . . .	24
7.6.	HTTP Usage . . . . .	24
7.6.1.	Authentication and Encryption . . . . .	24
7.6.2.	Cookies . . . . .	24
7.6.3.	Caching Parameters . . . . .	24
7.7.	ALTO Types . . . . .	24
7.7.1.	PID Name . . . . .	24
7.7.2.	Cost Mode . . . . .	25
7.7.3.	Cost Type . . . . .	25
7.8.	ALTO Messages . . . . .	25
7.8.1.	Server Information Service . . . . .	26
7.8.2.	Map Service . . . . .	30
7.8.3.	Map Filtering Service . . . . .	34
7.8.4.	Endpoint Property Service . . . . .	38
7.8.5.	Endpoint Cost Service . . . . .	41
8.	Redistributable Responses . . . . .	43
8.1.	Concepts . . . . .	43
8.1.1.	Service ID . . . . .	43
8.1.2.	Expiration Time . . . . .	44
8.1.3.	Signature . . . . .	44
8.2.	Protocol . . . . .	46
8.2.1.	Response Redistribution Descriptor Fields . . . . .	47
8.2.2.	Signature . . . . .	48
9.	Use Cases . . . . .	48
9.1.	ALTO Client Embedded in P2P Tracker . . . . .	48
9.2.	ALTO Client Embedded in P2P Client: Numerical Costs . . . . .	50
9.3.	ALTO Client Embedded in P2P Client: Ranking . . . . .	51
10.	Discussions . . . . .	51
10.1.	Discovery . . . . .	52
10.2.	Hosts with Multiple Endpoint Addresses . . . . .	52
10.3.	Network Address Translation Considerations . . . . .	52
10.4.	Mapping IPs to ASNs . . . . .	53
10.5.	Endpoint and Path Properties . . . . .	53
11.	IANA Considerations . . . . .	53
11.1.	application/alto Media Type . . . . .	54
11.2.	ALTO Cost Type Registry . . . . .	55
12.	Security Considerations . . . . .	56
12.1.	Privacy Considerations for ISPs . . . . .	56
12.2.	ALTO Clients . . . . .	56
12.3.	Authentication, Integrity Protection, and Encryption . . . . .	57
12.4.	ALTO Information Redistribution . . . . .	57
12.5.	Denial of Service . . . . .	58
12.6.	ALTO Server Access Control . . . . .	58
13.	References . . . . .	59
13.1.	Normative References . . . . .	59
13.2.	Informative References . . . . .	59
Appendix A.	TO BE MOVED . . . . .	61
A.1.	Discovery . . . . .	61

A.2. P2P Peer Selection . . . . .	61
A.2.1. Client-based Peer Selection . . . . .	62
A.2.2. Server-based Peer Selection . . . . .	62
A.2.3. Location-Only Peer Selection . . . . .	62
Appendix B. Acknowledgments . . . . .	63
Appendix C. Authors . . . . .	64
Authors' Addresses . . . . .	64

## 1. Introduction

### 1.1. Background and Problem Statement

Today, network information available to applications is mostly from the view of endhosts. There is no clear mechanism to convey information about the network's preferences to applications. By leveraging better network-provided information, applications have the potential to become more network-efficient (e.g., reduce network resource consumption) and achieve better application performance (e.g., accelerated download rate). The ALTO Service intends to provide a simple way to convey network information to applications.

The goal of this document is to specify a simple and unified protocol that meets the ALTO requirements [11] while providing a migration path for Internet Service Providers (ISP), Content Providers, and clients that have deployed protocols with similar intentions (see below). This document is a work in progress and will be updated with further developments.

### 1.2. Design History and Merged Proposals

The protocol specified here consists of contributions from

- o P4P [12], [13];
- o ALTO Info-Export [14];
- o Query/Response [15], [16];
- o ATTP [ATTP];
- o Proxidor [17].

See Appendix B for a list of people that have contributed significantly to this effort and the projects and proposals listed above.

### 1.3. Solution Benefits

The ALTO Service offers many benefits to both end-users (consumers of the service) and Internet Service Providers (providers of the service).

#### 1.3.1. Service Providers

The ALTO Service enables ISPs to influence the peer selection process in distributed applications in order to increase locality of traffic,

improve user-experience, amongst others. It also helps ISPs to efficiently engineer traffic that traverses more expensive links such as transit and backup links, thus allowing a better provisioning of the networking infrastructure.

### 1.3.2. Applications

Applications that use the ALTO Service can benefit in multiple ways. For example, they may no longer need to infer topology information, and some applications can reduce reliance on measuring path performance metrics themselves. They can take advantage of the ISP's knowledge to avoid bottlenecks and boost performance.

An example type of application is a Peer-to-Peer overlay where peer selection can be improved by including ALTO information in the selection process.

## 2. Architecture

Two key design objectives of the ALTO Protocol are simplicity and extensibility. At the same time, it introduces additional techniques to address potential scalability and privacy issues. After an introduction to the terminology, the ALTO architecture and the ALTO Protocol's place in the overall architecture are defined.

### 2.1. Terminology

We use the following terms defined in [18]: Application, Overlay Network, Peer, Resource, Resource Identifier, Resource Provider, Resource Consumer, Resource Directory, Transport Address, Host Location Attribute, ALTO Service, ALTO Server, ALTO Client, ALTO Query, ALTO Reply, ALTO Transaction, Local Traffic, Peering Traffic, Transit Traffic.

We also use the following additional terms: Endpoint Address, ASN, and Network Location.

#### 2.1.1. Endpoint Address

An endpoint address represents the communication address of an endpoint. An endpoint address can be network-attachment based (IP address) or network-attachment agnostic. Common forms of endpoint addresses include IP address, MAC address, overlay ID, and phone number.

#### 2.1.2. ASN

An Autonomous System Number.

#### 2.1.3. Network Location

Network Location is a generic term denoting a single endpoint or group of endpoints.

#### 2.1.4. ALTO Information

ALTO Information is a generic term referring to the network information sent by an ALTO Server.

#### 2.1.5. ALTO Information Base

Internal representation of the ALTO Information maintained by the ALTO Server. Note that the structure of this internal representation is not defined by this document.

### 2.2. ALTO Service and Protocol Scope

An ALTO Server conveys the network information from the perspective of a network region; the ALTO Server presents its "my-Internet View" [19] of the network region. A network region in this context can be an Autonomous System, an ISP, or perhaps a smaller region or set of ISPs; the details depend on the deployment scenario and discovery mechanism.

To better understand the ALTO Service and the role of the ALTO Protocol, we show in Figure 1 the overall system architecture. In this architecture, an ALTO Server prepares ALTO Information; an ALTO Client uses ALTO Service Discovery to identify an appropriate ALTO Server; and the ALTO Client requests available ALTO Information from the ALTO Server using the ALTO Protocol.

The ALTO Information provided by the ALTO Server can be updated dynamically based on network conditions, or can be seen as a policy which is updated at a larger time-scale.

More specifically, the ALTO Information provided by an ALTO Server may be influenced (at the operator's discretion) by other systems. Examples include (but are not limited to) static network configuration databases, dynamic network information, routing protocols, provisioning policies, and interfaces to outside parties. These components are shown in the figure for completeness but outside the scope of this specification.



Note that it may also be possible for ALTO Servers to exchange network information with other ALTO Servers (either within the same administrative domain or another administrative domain with the consent of both parties) in order to adjust exported ALTO information. Such a protocol is also outside the scope of this specification.

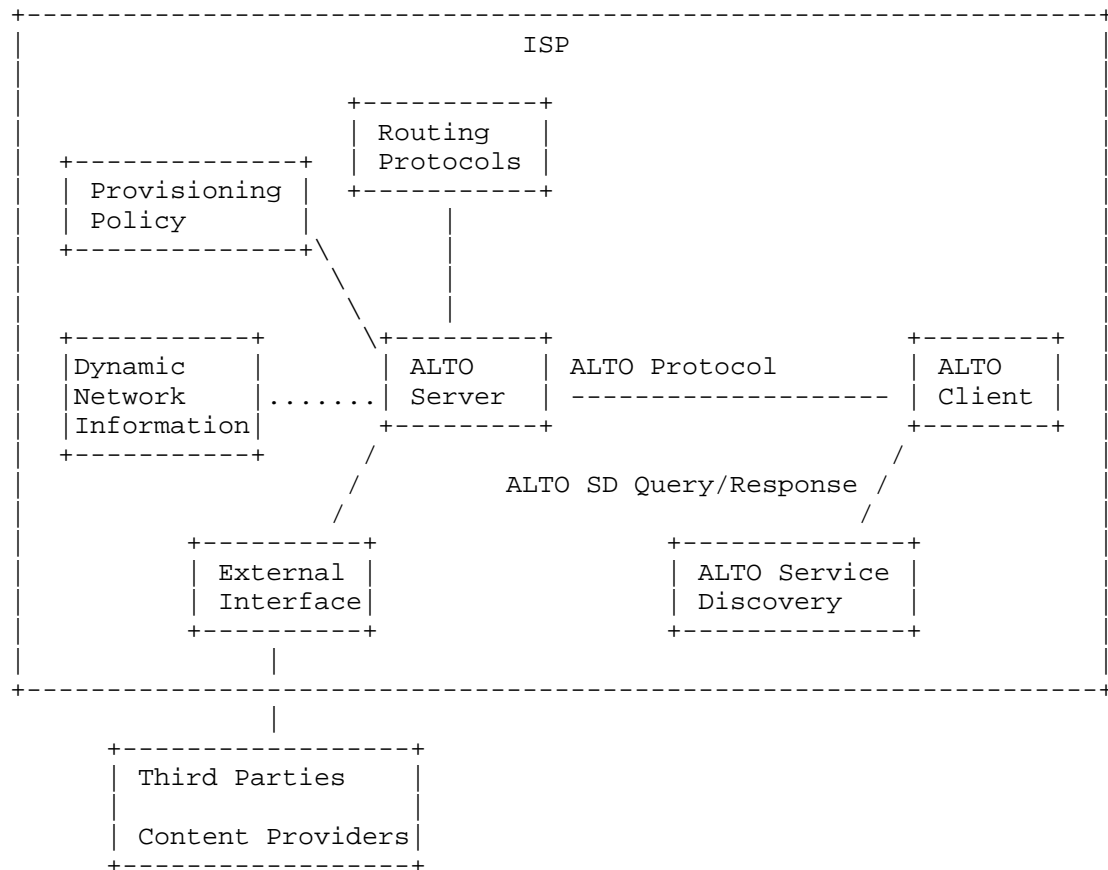


Figure 1: Basic ALTO Architecture

### 3. Protocol Structure

The ALTO Protocol uses a simple extensible framework to convey network information. In the general framework, the ALTO protocol will convey properties on both Network Locations and the paths between Network Locations.

In this document, we focus on a particular Endpoint property to denote the location of an endpoint, and provider-defined costs for paths between pairs of Network Locations.

The ALTO Protocol is built on a common transport protocol, messaging structure and encoding, and transaction model. The protocol is subdivided into services of related functionality. ALTO-Core provides the Server Information Service and the Map Service to provide ALTO Information. Other ALTO Information services can provide additional functionality. There are three such services defined in this document: the Map Filtering Service, Endpoint Property Service, and Endpoint Cost Service. Additional services may be defined in in companion documents. Note that functionality offered in different services are not totally non-overlapping (e.g., the Map Service and Map Filtering Service).

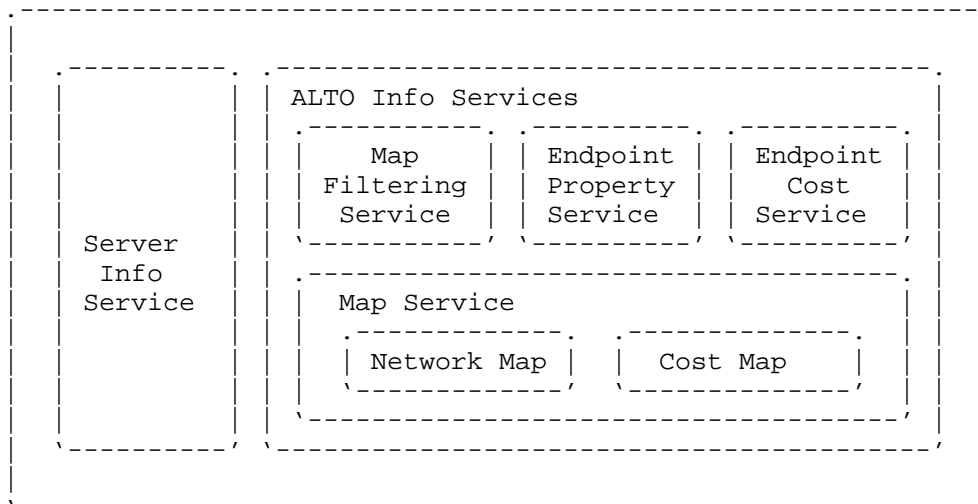


Figure 2: ALTO Protocol Structure

### 3.1. Server Information Service

The Server Capability Service lists the details on the information that can be provided by an ALTO Server and perhaps other ALTO Servers maintained by the network provider. The configuration includes, for example, details about the operations and cost metrics supported by the ALTO Server and other related ALTO Servers that may be usable by an ALTO Client. The capability document can be downloaded by ALTO Clients. The capability information could also be provisioned to devices, but care must be taken to update it appropriately.

### 3.2. ALTO Information Services

Multiple, distinct services are defined to allow ALTO Clients to query ALTO Information from an ALTO Server. The ALTO Server internally maintains an ALTO Information Base that encodes the network provider's preferences. The ALTO Information Base encodes the Network Locations defined by the ALTO Server (and their corresponding properties), as well as the provider-defined costs between pairs of Network Locations.

#### 3.2.1. Map Service

The Map Service provides batch information to ALTO Clients in the form of a Network Map and Cost Map. The Network Map (See Section 4) provides the full set of Network Location groupings defined by the ALTO Server and the endpoints contained with each grouping. The Cost Map (see Section 5) provides costs between the defined groupings.

These two maps can be thought of (and implemented as) as simple files with appropriate encoding provided by the ALTO Server.

#### 3.2.2. Map Filtering Service

Resource constrained ALTO Clients may benefit from query results being filtered at the ALTO Server. This avoids an ALTO Client spending network bandwidth or CPU collecting results and performing client-side filtering. The Map Filtering Service allows ALTO Clients to query for the ALTO Server Network Map and Cost Map based on additional parameters.

#### 3.2.3. Endpoint Property Service

This service allows ALTO Clients to look up properties for individual endpoints. An example endpoint property is its Network Location (its grouping defined by the ALTO Server) or connectivity type (e.g., ADSL, Cable, or FioS).

#### 3.2.4. Endpoint Cost Service

Some ALTO Clients may also benefit from querying for costs and rankings based on endpoints. The Endpoint Cost Service allows an ALTO Server to return either numerical costs or ordinal costs (rankings) directly amongst Endpoints.

## 4. Network Map

In reality, many endpoints are very close to one another in terms of

network connectivity, for example, endpoints on the same site of an enterprise. By treating a group of endpoints together as a single entity in ALTO, we can achieve much greater scalability without losing critical information.

The Network Location endpoint property allows an ALTO Server to group endpoints together to indicate their proximity. The resulting set of groupings is called the ALTO Network Map.

The definition of proximity varies depending on the granularity of the ALTO information configured by the provider. In one deployment, endpoints on the same subnet may be considered close; while in another deployment, endpoints connected to the same PoP may be considered close.

As used in this document, the Network Map refers to the syntax and semantics of the information distributed by the ALTO Server. This document does not discuss the internal representation of this data structure within the ALTO Server.

#### 4.1. PID

Each group of Endpoints is identified by a provider-defined Network Location identifier called a PID. There can be many different ways of grouping the endpoints and assigning PIDs.

A PID is an identifier that provides an indirect and network-agnostic way to specify a network aggregation. For example, a PID may be defined by the ALTO service provider to denote a subnet, a set of subnets, a metropolitan area, a PoP, an autonomous system, or a set of autonomous systems. Aggregation of endpoints into PIDs can indicate proximity and can improve scalability. In particular, network preferences (costs) may be specified between PIDs, allowing cost information to be more compact and updated at a smaller time scale than the network aggregations themselves.

Using PIDs, the Network Map may also be used to communicate simple preferences with only minimal information from the Cost Map. For example, an ISP may prefer that endpoints associated with the same PoP (Point-of-Presence) in a P2P application communicate locally instead of communicating with endpoints in other PoPs. The ISP may aggregate endhosts within a PoP into a single PID in the Network Map. The Cost Map may be encoded to indicate that peering within the same PID is preferred; for example,  $\text{cost}(\text{PID}_i, \text{PID}_i) == c^*$  and  $\text{cost}(\text{PID}_i, \text{PID}_j) > c^*$  for  $i \neq j$ . Section 5 provides further details about Cost Map structure.

## 4.2. Endpoint Addresses

Communicating endpoints may have many types of addresses, such as IP addresses, MAC addresses, or overlay IDs. The current specification only considers IP addresses.

### 4.2.1. IP Addresses

The endpoints aggregated into a PID are denoted by a list of IP prefixes. When either an ALTO Client or ALTO Server needs to determine which PID in a Network Map contains a particular IP address, longest-prefix matching **MUST** be used.

A Network Map **MUST** define a PID for each possible address in the IP address space. A **RECOMMENDED** way to satisfy this property is to define a PID containing the 0.0.0.0/0 prefix for IPv4 or ::/0 (for IPv6).

## 4.3. Example Network Map

Figure 3 illustrates an example Network Map. PIDs are used to identify network-agnostic aggregations.

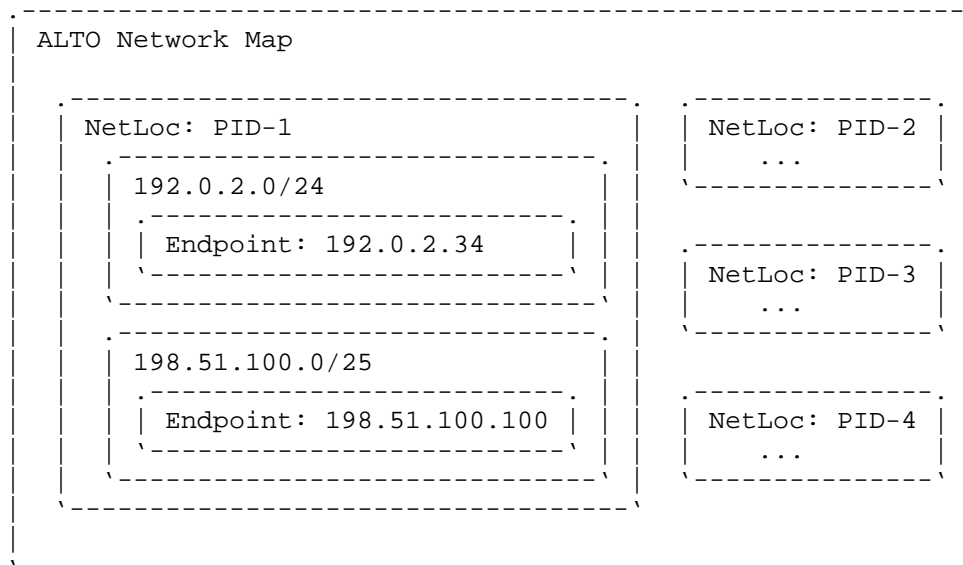


Figure 3: Example Network Map

## 5. Cost Map

An ALTO Server indicates preferences amongst network locations in the form of Path Costs. Path Costs are generic costs and can be internally computed by a network provider according to its own needs.

An ALTO Cost Map defines Path Costs pairwise amongst sets of source and destination Network Locations.

One advantage of separating ALTO information into a Network Map and a Cost Map is that the two components can be updated at different time scales. For example, Network Maps may be stable for a longer time while Cost Maps may be updated to reflect dynamic network conditions.

As used in this document, the Cost Map refers to the syntax and semantics of the information distributed by the ALTO Server. This document does not discuss the internal representation of this data structure within the ALTO Server.

### 5.1. Cost Attributes

Path Costs have attributes:

- o Type: identifies what the costs represent;
- o Mode: identifies how the costs should be interpreted.

Certain queries for Cost Maps allow the ALTO Client to indicate the desired Type and Mode.

#### 5.1.1. Cost Type

The Type attribute indicates what the cost represents. For example, an ALTO Server could define costs representing air-miles, hop-counts, or generic routing costs.

Cost types are indicated in protocol messages as strings.

##### 5.1.1.1. Cost Type: routingcost

An ALTO Server **MUST** define the 'routingcost' Cost Type.

This Cost Type conveys a generic measure for the cost of routing traffic from a source to a destination. Lower values indicate a higher preference for traffic to be sent from a source to a destination.

Note that an ISP may internally compute routing cost using any method

it chooses (e.g., air-miles or hop-count) as long as it conforms to these semantics.

#### 5.1.1.2. Cost Mode

The Mode attribute indicates how costs should be interpreted. An ALTO Server return costs that are interpreted as either numerical values or ordinal rankings.

It is important to communicate such information to ALTO Clients, as certain operations may not be valid on certain costs returned by an ALTO Server. For example, it is possible for an ALTO Server to return a set of IP addresses with costs indicating a ranking of the IP addresses. Arithmetic operations, such as summation, that would make sense for numerical values, do not make sense for ordinal rankings. ALTO Clients may handle such costs differently.

Cost Modes are indicated in protocol messages as strings.

An ALTO Server MUST support at least one of 'numerical' and 'ordinal' costs. ALTO Clients SHOULD be cognizant of operation when a desired cost mode is not supported. For example, an ALTO Client desiring numerical costs may adjust behavior if only the ordinal Cost Mode is available. Alternatively, an ALTO Client desiring ordinal costs may construct ordinal costs given numerical values if only the numerical Cost Mode is available.

##### 5.1.2.1. Cost Mode: numerical

This Cost Mode is indicated by the string 'numerical'. This mode indicates that it is safe to perform numerical operations (e.g. summation) on the returned costs.

##### 5.1.2.2. Cost Mode: ordinal

This Cost Mode is indicated by the string 'ordinal'. This mode indicates that the costs values to a set of Destination Network Locations from a particular Source Network Location are a ranking, with lower values indicating a higher preference.

It is important to note that the values in the Cost Map provided with the ordinal Cost Mode are not necessarily the actual cost known to the ALTO Server.

#### 5.2. Cost Map Structure

A query for a Cost Map either explicitly or implicitly includes a list of Source Network Locations and a list of Destination Network

Locations. (Recall that a Network Location can be an endpoint address or a PID.)

Specifically, assume that a query has a list of multiple Source Network Locations, say [Src\_1, Src\_2, ..., Src\_m], and a list of multiple Destination Network Locations, say [Dst\_1, Dst\_2, ..., Dst\_n].

The ALTO Server will return the Path Cost for each communicating pair (i.e., Src\_1 -> Dst\_1, ..., Src\_1 -> Dst\_n, ..., Src\_m -> Dst\_1, ..., Src\_m -> Dst\_n). We refer to this structure as a Cost Map.

If the Cost Mode is 'ordinal', the Path Cost of each communicating pair is relative to the m\*n entries.

### 5.3. Network Map and Cost Map Dependency

If a Cost Map contains PIDs in the list of Source Network Locations or the list of Destination Network Locations, the Path Costs are generated based on a particular Network Map (which defines the PIDs). Version Tags are introduced to ensure that ALTO Clients are able to use consistent information even though the information is provided in two maps.

A Version Tag is an opaque string associated with a Network Map maintained by the ALTO Server. When the Network Map changes, the Version Tag SHOULD also be changed. (Thus, the Version Tag is defined similarly to HTTP's ETag.) Possibilities for generating a Version Tag included the last-modified timestamp for the Network Map, or a hash of its contents.

A Network Map distributed by the ALTO Server includes its Version Tag. A Cost Map referring to PIDs also includes the Version Tag of the Network Map on which it is based.

## 6. Protocol Design Overview

The ALTO Protocol design uses a REST-like interface with the goal of leveraging current HTTP [2] [3] implementations and infrastructure, as well as familiarity with existing REST-like services in popular use. ALTO messages use JSON [4] to encode message bodies.

This document currently specifies both services and the message encoding in a descriptive fashion. Care is taken to make descriptions precise and unambiguous, but it still lacks benefits of automatic tooling that exists for certain encoding formats.



Standards such as WSDL 2.0 and WADL are capable of describing available interfaces. JSON Schema [20] allows message encodings to be specified precisely and messages may be verified against the schema. It is not yet clear whether such an approach should be taken in this document.

Other benefits enabled by these design choices include easier understanding and debugging, flexible ALTO Server implementation strategies, and more importantly, simple caching and redistribution of ALTO information to increase scalability.

#### 6.1. Existing Infrastructure

HTTP is a natural choice for integration with existing applications and infrastructure. In particular, the ALTO Protocol design leverages:

- o the huge installed base of infrastructure, including HTTP caches,
- o mature software implementations,
- o the fact that many P2P clients already have an embedded HTTP client, and
- o authentication and encryption mechanisms in HTTP and SSL/TLS.

#### 6.2. ALTO Information Reuse and Redistribution

ALTO information may be useful to a large number of applications and users. For example, an identical Network Map may be used by all ALTO Clients querying a particular ALTO Server. At the same time, distributing ALTO information must be efficient and not become a bottleneck.

Beyond integration with existing HTTP caching infrastructure, ALTO information may also be cached or redistributed using application-dependent mechanisms, such as P2P DHTs or P2P file-sharing. This document does not define particular mechanisms for such redistribution, but it does define the primitives (e.g., digital signatures) needed to support such a mechanism. See [21] for further discussion.

Note that if caching or redistribution is used, the Response message may be returned from another (possibly third-party) entity. Reuse and Redistribution is further discussed in Section 12.4. Protocol support for redistribution is specified in Section 8.

## 7. Protocol Messaging

This section specifies client and server processing, as well as messages in the ALTO Protocol. Details common to ALTO Server processing of all messages is first discussed, followed by details of the individual messages.

### 7.1. Notation

This document uses an adaptation of the C-style struct notation to define the required and optional members of JSON objects. Unless explicitly noted, each member of a struct is REQUIRED.

The types 'JSONString', 'JSONNumber', 'JSONBool' indicate the JSON string, number, and boolean types respectively.

This document only includes object members used by this specification. It is possible that protocol extensions include additional members to JSON objects defined in this document; such additional members will be silently ignored by ALTO Servers and Clients only implementing the base protocol defined in this document.

### 7.2. Message Format

Request and Response follow the standard format for HTTP Request and Response messages [2] [3].

The following subsections provide an overview of how ALTO Requests and Responses are encoded in HTTP, and discusses rationale for certain design decisions.

#### 7.2.1. Protocol Versioning

The ALTO Protocol uses a simple versioning approach that permits evolution between versions even if ALTO information is being served as static, pre-generated files.

It is assumed that a single host responding to ALTO Requests implements a single protocol version. Virtual hosting may be used if multiple protocol versions need to be supported by a single physical server.

A common query (Server List, detailed in Section 7.8.1.1) to be present in all ALTO protocol versions allows an ALTO Client to discover additional ALTO Servers and the ALTO Protocol version number of each.

This approach keeps the ALTO Server implementation free from parsing

and directing each request based on version number. Although ALTO Requests are free from protocol version numbers, the protocol version number is echoed in each ALTO Response to keep responses self-contained to, for example, ease reading persisted or redistributed ALTO responses.

Using virtual hosting with TLS may require the Server Name Indication extension for TLS [5] [22].

This document specifies ALTO Protocol version 1.

#### 7.2.2. Content Type

All ALTO Request and Response messages MUST set the Content-Type HTTP header to "application/alto".

#### 7.2.3. Request Message

An ALTO Request is a standard HTTP Request generated by an ALTO Client, with certain components defined by the ALTO Protocol.

The basic syntax of an ALTO Request is:

```
<Method> /<Resource> HTTP/1.1
Host: <Host>
```

For example:

```
GET /info/capability HTTP/1.1
Host: alto.example.com:6671
```

##### 7.2.3.1. Standard HTTP Headers

The Host header MUST follow the standard rules for the HTTP 1.1 Host Header.

The Content-Length header MUST follow the standard rules defined in HTTP 1.1.

The Content-Type HTTP Header MUST have value "application/alto" if the Body is non-empty.

##### 7.2.3.2. Method and Resource

Next, both the HTTP Method and URI-Path (denoted as Resource) indicate the operation requested by the ALTO Client. In this example, the ALTO Client is requesting basic capability information from the ALTO Server.

#### 7.2.3.3. Input Parameters

Certain operations defined by the ALTO Protocol (e.g., in the Map Filtering Service) allow the ALTO Client to supply additional input parameters. Such input parameters are encoded in a URI-Query-String where possible and appropriate. However, due to practical limitations (e.g. underlying HTTP implementations may have limitations on the total length of a URI and the Query-String is better-suited for simple unstructured parameters and lists), some operations in the ALTO Protocol use input parameters encoded in the HTTP Request Body.

#### 7.2.4. Response Message

A Response message is a standard HTTP Response generated by an ALTO Server with certain components defined by the ALTO Protocol.

The basic syntax of an ALTO Response is:

```
HTTP/1.1 <StatusCode> <StatusMsg>
Content-Length: <ContentLength>
Content-Type: <ContentType>

<ALTOResponse>
```

where the HTTP Response Body is an ALTOResponse JSON Object (defined in Section 7.2.4.3). For example:

```
HTTP/1.1 200 OK
Content-Length: 1000
Content-Type: application/alto

{
  "meta" : {
    "version": 1,
    "status" : {
      "code" : 1,
      "reason" : "Success"
    },
    ...
  },
  "type" : "capability",
  "data" : {
    ...
  }
}
```

#### 7.2.4.1. Standard HTTP Headers

The Content-Length header MUST follow the standard rules defined in HTTP 1.1.

The Content-Type HTTP Header MUST have value "application/alto" if the Body is non-empty.

#### 7.2.4.2. Status Code and Message

Two sets of status codes are used in the ALTO Protocol. First, an ALTO Status Code provides detailed information about the success or failure of a particular operation. Second, an HTTP Status Code indicates to HTTP processing elements (e.g., intermediaries and clients) how the response should be treated.

#### 7.2.4.3. HTTP Body

The Response body MUST encode a single top-level JSON object of type ALTOResponse:

```
object {  
    RspMetaData    meta;  
    JSONString     type;  
    [RspDataType] data;  
} ALTOResponse;
```

The ALTOResponse object has distinct sections for:

- o meta information encoded in an extensible way,
- o the type of ALTO Information to follow, and
- o the requested ALTO Information.

##### 7.2.4.3.1. Meta Information

Meta information is encoded as a JSON object with type RspMetaData:

```
object {  
    JSONString     code;  
    JSONString     reason;           [OPTIONAL]  
} RspStatus;  
  
object {  
    JSONNumber     version;  
    RspStatus      status;  
    RspRedistDesc  redistribution;   [OPTIONAL]  
}
```

```
    } RspMetaData;
```

with members:

- o version: the ALTO Protocol version
- o status: an ALTO Status Code from Section 7.4 and corresponding reason (free-form string) providing a human-readable explanation of the particular status code.
- o redistribution: see Section 8.

#### 7.2.4.3.2. ALTO Information

If the Response is successful (see Section 7.4), then the "type" and "data" members of the ALTOResponse object are REQUIRED. "type" encodes a Response-specific string which indicates to the ALTO Client the type of data encoded in the message. The "data" member encodes the actual Response-specific data; the structure of this member is detailed later in this section for each particular ALTO Response.

#### 7.2.4.4. Signature

An ALTO Server MAY additionally supply a signature asserting that it generated a particular response. See Section 8.2.2.

### 7.3. General Processing

The protocol is structured in such a way that, independent of the query type, there are a set of general processing steps. The ALTO Client selects a specific ALTO Server with which to communicate, establishes a TCP connection, and constructs and sends ALTO Request messages which MUST conform to Section 7.8. In response to Request messages, an ALTO Server constructs and sends ALTO Response messages which also MUST conform to Section 7.8.

### 7.4. ALTO Status Codes

This document defines ALTO Status Codes to support the operations defined in this document. Additional status codes may be defined in companion or extension documents.

An ALTO Server MUST return the SUCCESS status code if and only if the Request message is successfully processed and the requested ALTO information is returned by the ALTO Server.

The HTTP Status Codes corresponding to each ALTO Status Code are defined to provide correct behavior with HTTP intermediaries and

clients. When an ALTO Server returns a particular ALTO Status Code, it MUST indicate one of the corresponding HTTP Status Codes in Table 1.

If multiple errors are present in a single ALTO Request (e.g., a request uses a JSONString when a JSONInteger is expected and a required field is missing), then the ALTO Server MUST return exactly one of the detected errors. However, the reported error is implementation defined, since specifying a particular order for message processing encroaches needlessly on implementation technique.

ALTO Status Code	HTTP Status Code(s)	Description
SUCCESS	2xx	Success
E_JSON_SYNTAX	400	JSON parsing error in request
E_JSON_FIELD_MISSING	400	Required field missing
E_JSON_VALUE_TYPE	400	JSON Value of unexpected type
E_INVALID_OPERATION	501	Invalid operation requested
E_INVALID_COST_TYPE	501	Invalid cost type

Table 1: Defined ALTO Status Codes

Status codes described in Table 1 are a work in progress. This document will be modified to update the available status codes as implementation experience is gained. Feedback is welcomed.

In addition, feedback from implementers of ALTO Clients is welcomed to identify if there is a need to communicate multiple status codes in a single response.

## 7.5. Client Behavior

### 7.5.1. Successful Response

This specification does not indicate any required actions taken by ALTO Clients upon receiving a successful response from an ALTO Server. Although ALTO Clients are suggested to interpret the received ALTO Information and adapt application behavior, ALTO Clients are not required to do so.

### 7.5.2. Error Conditions

If an ALTO Client does not receive a successful response from the ALTO Server, it can either choose another server or fall back to a default behavior (e.g., perform peer selection without the use of ALTO information). An ALTO Client may also retry the request at a later time.

## 7.6. HTTP Usage

### 7.6.1. Authentication and Encryption

An ALTO Server MAY support SSL/TLS to implement server and/or client authentication, as well as encryption. See [6] for considerations regarding verification of server identity.

An ALTO Server MAY support HTTP Digest authentication.

### 7.6.2. Cookies

Cookies MUST NOT be used.

### 7.6.3. Caching Parameters

If the Response generated by the ALTO Server is cachable, the ALTO Server MAY include 'Cache-Control' and 'Expires' HTTP headers.

If a Response generated by the ALTO Server is not cachable, the ALTO Server MUST specify the "Cache-Control: no-cache" HTTP Header.

## 7.7. ALTO Types

This section details the encoding for particular data values used in the ALTO Protocol.

### 7.7.1. PID Name

A PID Name is encoded as a US-ASCII string. The string MUST be no more than 32 characters, and MUST NOT contain characters other than alphanumeric characters or the '.' separator. The '.' separator is reserved for future use and MUST NOT unless specifically indicated by a companion or extension document.

The type 'PIDName' is used in this document to indicate a string of this format.



#### 7.7.2. Cost Mode

A Cost Mode is encoded as a US-ASCII string. The string MUST either have the value 'numerical' or 'ordinal'.

The type 'CostMode' is used in this document to indicate a string of this format.

#### 7.7.3. Cost Type

A Cost Type is encoded as a US-ASCII string. The string MUST be no more than 32 characters, and MUST NOT contain characters other than alphanumeric characters or the ':' separator.

Identifiers prefixed with 'priv:' are reserved for Private Use [7]. Identifiers prefixed with 'exp:' are reserved for Experimental use. All other identifiers appearing in an ALTO Request or Response MUST be registered in the ALTO Cost Types registry Section 11.

The type 'CostType' is used in this document to indicate a string of this format.

#### 7.8. ALTO Messages

This section documents the individual operations supported in the ALTO Protocol. See Section 7.2.3 and Section 7.2.4 for specifications of HTTP Request/Response components common to all operations in the ALTO Protocol.

Table 2 provides an summary of the HTTP Method and URI-Paths used for ALTO Requests:

Service	Operation	HTTP Method and URI-Path
Server Info Server Info	List Servers	GET /info/servers
	Capability	GET /info/capability
Map Map	Network Map	GET /map/core/pid/net
	Cost Map	GET /map/core/pid/cost
Map Filtering Map Filtering	Network Map	POST /map/filter/pid/net
	Cost Map	POST /map/filter/pid/cost
Endpoint Prop. Endpoint Prop.	Lookup	GET /endpoint/prop/<name>
		POST /endpoint/prop/lookup
Endpoint Cost	Lookup	POST /endpoint/cost/lookup

Table 2: Overview of ALTO Requests

#### 7.8.1. Server Information Service

The Server Information Service provides information about available ALTO Servers and their capabilities (e.g., supported services).

An ALTO Server **MUST** support the Server Information Service and **MUST** implement all operations defined in this section.

##### 7.8.1.1. Server List

The Server List request allows an ALTO Client to discover other ALTO Servers provided by the ALTO Service Provider. Upon discovering an additional ALTO Server, the ALTO Client may then query the server capabilities (see Section 7.8.1.2) to test if it supports desired functionality.

The Server List request is intended to help an ALTO Client find an ALTO Server supporting the desired ALTO Protocol version and capabilities. It is not intended to serve as a substitute for the ALTO Server Discovery which helps an ALTO Client locate an initial ALTO Server.

This operation **MUST** be supported by the ALTO Server.

##### 7.8.1.1.1. Request Syntax

```
GET /info/servers HTTP/1.1
Host: <Host>
```

## 7.8.1.1.2. Response Syntax

```
HTTP/1.1 200 <StatusMsg>
Content-Length: <BodyLength>
Content-Type: application/alto
```

```
<ALTOResponse>
```

where the ALTOResponse object has "type" member equal to the string "server-list" and "data" member of type RspServerList:

```
object {
    JSONString    uri;
    JSONNumber    version;
} ServerItem;

object {
    ServerItem    servers<0..*>;
} RspServerList;
```

RspServerList has members:

- o servers: Array of available ALTO Servers, detailing the URI of the ALTO Server and the ALTO Protocol version that it implements. The array must at least contain an entry corresponding to the ALTO Server at the URI from which it is retrieving the server list.

## 7.8.1.1.3. Example

```
GET /info/servers HTTP/1.1
Host: alto.example.com:6671
```

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

{
  "meta" : {
    "version" : 1,
    "status" : {
      "code" : 1
    }
  },
  "type" : "server-list",
  "data" : {
    "servers" : [
      {
        "uri": "http://alto.example.com:6671",
        "version" : 1
      }
    ]
  }
}
```

#### 7.8.1.2. Server Capability

The Server Capability request allows an ALTO Client to determine the functionality supported by the queried ALTO Server.

This operation MUST be supported by the ALTO Server.

##### 7.8.1.2.1. Request Syntax

```
GET /info/capability HTTP/1.1
Host: <Host>
```

##### 7.8.1.2.2. Response Syntax

```
HTTP/1.1 200 <StatusMsg>
Content-Length: <BodyLength>
Content-Type: application/alto

<ALTOResponse>
```

where the ALTOResponse object has "type" member equal to the string "capability" and "data" member of type RspCapability:

```
enum {
    map,
    map-filtering,
    endpoint-property,
    endpoint-cost
} ServiceType;          [Note: encoded as JSONString's]

object {
    ServiceType  services<0..*>;
    CostMode     cost-modes<0..*>;          [OPTIONAL]
    CostType     cost-types<0..*>;          [OPTIONAL]
    JSONBool     cost-constraints;          [OPTIONAL]
    JSONString   service-id;                [OPTIONAL]
    JSONString   certificates<0..*>;        [OPTIONAL]
} RspCapability;
```

RspCapability has members:

- o services: Lists the services supported by the ALTO Server. The service names defined in this document are "map", "map-filtering", "endpoint-property", and "endpoint-cost".
- o cost-modes: Array of supported ALTO Cost Modes.
- o cost-types: Array of supported ALTO Cost Types.
- o cost-constraints: Indicates if the ALTO Server supports cost constraints. The value 'false' is implied if this member is not present.
- o service-id: UUID [8] indicating an one or more ALTO Servers serving equivalent ALTO Information.
- o certificates: List of PEM-encoded X.509 certificates used by the ALTO Server in the signing of responses.

If an ALTO Server denotes a response as redistributable, the 'service-id' and 'certificates' fields are REQUIRED instead of OPTIONAL. See Section 8 for detailed specification.

#### 7.8.1.2.3. Example

```
GET /info/capability HTTP/1.1
Host: alto.example.com:6671
```

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

```
{
  "meta" : {
    "version" : 1,
    "status" : {
      "code" : 1
    }
  },
  "type" : "capability",
  "data" : {
    "services" : [ "map", "map-filtering" ],
    "cost-modes": [
      "numerical",
      "ordinal"
    ],
    "cost-types": [
      "routingcost",
      "hopcount"
    ],
    "cost-constraints": false
  }
}
```

#### 7.8.2. Map Service

The Map Service provides batch information to ALTO Clients in the form of two maps: a Network Map and Cost Map.

An ALTO Server **MUST** support the Map Service and **MUST** implement all operations defined in this section.

##### 7.8.2.1. Network Map

The full Network Map lists for each PID, the network locations (endpoints) within the PID.

###### 7.8.2.1.1. Request Syntax

```
GET /map/core/pid/net HTTP/1.1
Host: <Host>
```

## 7.8.2.1.2. Response Syntax

```
HTTP/1.1 200 <StatusMsg>
Content-Length: <BodyLength>
Content-Type: application/alto
```

```
<ALTOResponse>
```

where the ALTOResponse object has "type" member equal to the string "network-map" and "data" member of type RspNetworkMap:

```
object {
  CIDRString [pidname]<0..*>;
  ...
} NetworkMapData;

object {
  JSONString      map-vtag;
  NetworkMapData map;
} RspNetworkMap;
```

RspNetworkMap has members:

- o map-vtag: The Version Tag of the Network Map (Section 5.3)
- o map: The network map data itself.

NetworkMapData is a JSON object with each member representing a single PID and its associated set of IP Prefixes (encoded as a string in CIDR notation). A member's name is a PIDName string denoting the PID's name.

## 7.8.2.1.3. Example

```
GET /map/core/pid/net HTTP/1.1
Host: alto.example.com:6671
```

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

```
{
  "meta" : {
    "version" : 1,
    "status" : {
      "code" : 1
    }
  },
  "type" : "network-map",
  "data" : {
    "map-vtag" : "1266506139",
    "map" : {
      "PID1" : [
        "192.0.2.0/24",
        "198.51.100.0/25"
      ],
      "PID2" : [
        "198.51.100.128/25"
      ],
      "PID3" : [
        "0.0.0.0/0"
      ]
    }
  }
}
```

#### 7.8.2.2. Cost Map

The Map Service Cost Map query is a batch operation in which the ALTO Server returns the Path Cost for each pair of source/destination PID defined by the ALTO Server.

The ALTO Server provides costs using the default Cost Type ('routingcost') and default Cost Mode ('numerical').

##### 7.8.2.2.1. Request Syntax

```
GET /map/core/pid/cost HTTP/1.1
Host: <Host>
```



## 7.8.2.2.2. Response Syntax

```
HTTP/1.1 200 <StatusMsg>
Content-Length: <BodyLength>
Content-Type: application/alto
```

```
<ALTOResponse>
```

where the ALTOResponse object has "type" member equal to the string "cost-map" and "data" member of type RspCostMap:

```
object DstCosts {
  JSONNumber [dstname];
  ...
};

object {
  DstCosts [srcname]<0..*>;
  ...
} CostMapData;

object {
  JSONString  map-vtag;
  CostType    cost-type;
  CostMode    cost-mode;
  CostMapData map;
} RspCostMap;
```

RspCostMap has members:

- o map-vtag: The Version Tag of the Network Map used to generate the Cost Map (Section 5.3).
- o cost-type: Cost Type used in the map (Section 5.1.1)
- o cost-mode: Cost Mode used in the map (Section 5.1.2)
- o map: The cost map data itself.

CostMapData is a JSON object with each member representing a single Source PID; the name for a member is the PIDName string identifying the corresponding Source PID. For each Source PID, a DstCosts object denotes the associated cost to a set of destination PIDs (Section 5.2); the name for each member in the object is the PIDName string identifying the corresponding Destination PID. DstCosts has a single member for each destination PID in the map.

## 7.8.2.2.3. Example

```
GET /map/core/pid/cost HTTP/1.1
Host: alto.example.com:6671
```

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

```
{
  "meta" : {
    "version" : 1,
    "status" : {
      "code" : 1
    }
  },
  "type" : "cost-map",
  "data" : {
    "map-vtag" : "1266506139",
    "cost-type" : "routingcost",
    "cost-mode" : "numerical",
    "map" : {
      "PID1": { "PID1": 1, "PID2": 5, "PID3": 10 },
      "PID2": { "PID1": 5, "PID2": 1, "PID3": 15 },
      "PID3": { "PID1": 20, "PID2": 15, "PID3": 1 }
    }
  }
}
```

## 7.8.3. Map Filtering Service

The Map Filtering Service allows ALTO Clients to specify filtering criteria to return a subset of the full maps available in the Map Service.

An ALTO Server MAY support the Map Filtering Service. If an ALTO Server supports the Map Filtering Service, all operations defined in this section MUST be implemented.

## 7.8.3.1. Network Map

ALTO Clients can query for a subset of the full network map (see Section 7.8.2.1).

## 7.8.3.1.1. Request Syntax

```
POST /map/filter/pid/net HTTP/1.1
Host: <Host>
Content-Length: <BodyLength>

<ReqNetworkMap>
```

where:

```
object {
    PIDName pids<0..*>;
} ReqNetworkMap;
```

The Body of the request encodes an array of PIDs to be included in the resulting Network Map. If the list of PIDs is empty, the ALTO Server MUST interpret the list as if it contained a list of all currently-defined PIDs.

## 7.8.3.1.2. Response Syntax

The Response syntax is identical to that of the Map Service's Network Map Response (Section 7.8.2.1.2).

The ALTO Server MUST only include PIDs in the Response that were specified (implicitly or explicitly) in the Request. If the Request contains a PID name that is not currently defined by the ALTO Server, the ALTO Server MUST behave as if the PID did not appear in the request.

## 7.8.3.1.3. Example

```
POST /map/filter/pid/net HTTP/1.1
Host: alto.example.com:6671
Content-Length: <BodyLength>
```

```
{
  pids: [ "PID1", "PID2" ]
}
```

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

```
{
  "meta" : {
    "version" : 1,
    "status" : {
      "code" : 1
    }
  },
  "type" : "network-map",
  "data" : {
    "map-vtag" : "1266506139",
    "map" : {
      "PID1" : [
        "192.0.2.0/24",
        "198.51.100.0/24",
      ],
      "PID2" : [
        "198.51.100.128/24",
      ]
    }
  }
}
```

## 7.8.3.2. Cost Map

ALTO Clients can query for the Cost Map (see Section 7.8.2.2) based on additional parameters.

## 7.8.3.2.1. Request Syntax

```
POST /map/filter/pid/cost?<URI-Query-String> HTTP/1.1
Host: <Host>

<ReqCostMap>
```

where:

```
object {  
    PIDName srcs<0..*>;  
    PIDName dsts<0..*>;  
} ReqCostMap;
```

The Query String may contain the following parameters:

- o type: The requested Cost Type (Section 5.1.1). If not specified, the default value is "routingcost". This parameter MUST NOT be specified multiple times.
- o mode: The requested Cost mode (Section 5.1.2). If not specified, the default value is "numerical". This parameter MUST NOT be specified multiple times.
- o constraint: Defines a constraint on which elements of the Cost Map are returned. This parameter MUST NOT be used if the Server Capability Response (Section 7.8.1.2) indicates that constraint support is not available. A constraint contains two entities separated by whitespace (before URL encoding): (1) an operator either 'gt' for greater than, 'lt' for less than or 'eq' for equal to with 10 percent on either side, (2) a target numerical cost. The numerical cost is a number that MUST be defined in the units specified in the Server Capability Response. If multiple 'constraint' parameters are specified, the ALTO Server assumes they are related to each other with a logical AND. If no 'constraint' parameters are specified, then the ALTO Server returns the full Cost Map.

The Request body MAY specify a list of Source PIDs, and a list of Destination PIDs. If a list is empty, it is interpreted by the ALTO Server as the full set of currently-defined PIDs. The ALTO Server returns costs between each pair of source/destination PID. If the Request body is empty, both lists are interpreted to be empty.

#### 7.8.3.2.2. Response Syntax

The Response syntax is identical to that of the Map Service's Cost Map Response (Section 7.8.2.2.2).

The Response MUST NOT contain any source/destination pair that was not indicated (implicitly or explicitly) in the Request. If the Request contains a PID name that is not currently defined by the ALTO Server, the ALTO Server MUST behave as if the PID did not appear in the request.

## 7.8.3.2.3. Example

```
POST /map/filter/pid/cost?type=hopcount HTTP/1.1
```

```
Host: alto.example.com:6671
```

```
{
  "srcs" : [ "PID1" ],
  "dsts" : [ "PID1", "PID2", "PID3" ]
}
```

```
HTTP/1.1 200 OK
```

```
Content-Length: [TODO]
```

```
Content-Type: application/alto
```

```
{
  "meta" : {
    "version" : 1,
    "status" : {
      "code" : 1
    }
  },
  "type" : "cost-map",
  "data" : {
    "map-vtag" : "1266506139",
    "cost-type" : "hopcount",
    "cost-mode" : "numerical",
    "map" : {
      "PID1": { "PID1": 0, "PID2": 1, "PID3": 2 }
    }
  }
}
```

## 7.8.4. Endpoint Property Service

The Endpoint Property Lookup query allows an ALTO Client to lookup properties of Endpoints known to the ALTO Server. If the ALTO Server provides the Endpoint Property Service, the ALTO Server MUST define at least the 'pid' property for Endpoints.

An ALTO Server MAY support the Endpoint Property Service. If an ALTO Server supports the Endpoint Property Service, all operations defined in this section MUST be implemented.

## 7.8.4.1. Endpoint Property Lookup

## 7.8.4.1.1. Request Syntax

```
POST /endpoint/prop/lookup?<URI-Query-String> HTTP/1.1
Host: <Host>
Content-Length: <BodyLength>

<ReqEndpointProp>
```

where:

```
object {
  JSONString endpoints<0..*>;
} ReqEndpointProp;
```

The Query String may contain the following parameters:

- o prop: The requested property type. This parameter MUST be specified at least once, and MAY be specified multiple times (e.g., to query for multiple different properties at once).

The body encodes a list of endpoints (IP addresses) as strings.

An alternate syntax is supported for the case when properties are requested for a single endpoint:

```
GET /endpoint/prop/<Endpoint>?<URI-Query-String> HTTP/1.1
Host: <Host>
```

where the Query String is the same as in the first form.

## 7.8.4.1.2. Response Syntax

```
HTTP/1.1 200 <StatusMsg>
Content-Length: <BodyLength>
Content-Type: application/alto

<ALTOResponse>
```

where the ALTOResponse object has "type" member equal to the string "endpoint-property" and "data" member of type RspEndpointProperty:

```
object {
  JSONString [propertyname];
  ...
} EndpointProps;

object {
  EndpointProps [endpointname]<0..*>;
  ...
} RspEndpointProperty;
```

RspEndpointProperty has one member for each endpoint indicated in the Request. The requested properties for each endpoint are encoded in a corresponding EndpointProps object, which encodes one name/value pair for each requested property. Note that property values are JSON Strings. If the ALTO Server does not define a requested property for a particular endpoint, then it **MUST** omit it from the Response for only that endpoint.

#### 7.8.4.1.3. Example

```
POST /endpoint/prop/lookup?prop=pid HTTP/1.1
Host: alto.example.com:6671
Content-Length: [TODO]
```

```
{
  "endpoints" : [ "192.0.2.34", "203.0.113.129" ]
}
```

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

```
{
  "meta" : {
    "version" : 1,
    "status" : {
      "code" : 1
    }
  },
  "type" : "endpoint-property",
  "data": {
    "192.0.2.34" : { "pid": "PID1" },
    "203.0.113.129" : { "pid": "PID3" }
  }
}
```



#### 7.8.5. Endpoint Cost Service

The Endpoint Cost Service allows ALTO Clients to directly supply endpoints to an ALTO Server. The ALTO Server replies with costs (numerical or ordinal) amongst the endpoints.

In particular, this service allows lists of Endpoint addresses to be ranked (ordered) by an ALTO Server.

An ALTO Server MAY support the Endpoint Cost Service. If an ALTO Server supports the Endpoint Cost Service, all operations defined in this section MUST be implemented.

##### 7.8.5.1. Endpoint Cost Lookup

###### 7.8.5.1.1. Request Syntax

```
POST /endpoint/cost/lookup?<URI-Query-String> HTTP/1.1
Host: <Host>
Content-Length: <BodyLength>

<ReqCostMap>
```

The request body includes a list of source and destination endpoints that should be assigned a cost by the ALTO Server. The allowed Query String parameters are defined identically to Section 7.8.3.2.

The request body MUST specify a list of source Endpoints, and a list of destination Endpoints, using an structure identical to Section 7.8.3.2 with the exception that identifiers are endpoints instead of PIDs. If the list of source Endpoints is empty (or it is not included), the ALTO Server MUST treat it as if it contained the Endpoint address of the requesting client. The list of destination Endpoints MUST NOT be empty. The ALTO Server returns costs between each pair of source/destination Endpoint.

###### 7.8.5.1.2. Response Syntax

```
HTTP/1.1 200 <StatusMsg>
Content-Length: <BodyLength>
Content-Type: application/alto

<ALTOResponse>
```

where ALTOResponse is encoded identically to Section 7.8.2.2.2 with the following exceptions:

- o ALTO Response's "type" member must be equal to "endpoint-cost-map",
- o The "map-vtag" member of RspCostMap MUST be omitted, and
- o Identifiers refer to endpoints instead of PIDs.

#### 7.8.5.1.3. Example

```
POST /endpoint/cost/lookup?mode=ordinal HTTP/1.1
Host: alto.example.com:6671
Content-Length: [TODO]
```

```
{
  "src": [ "192.0.2.2" ],
  "dst": [ "192.0.2.89", "198.51.100.34", "203.0.113.45" ]
}
```

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

```
{
  "meta" : {
    "version" : 1,
    "status" : {
      "code" : 1
    }
  },
  "type" : "endpoint-cost-map",
  "data" : {
    "cost-type" : "routingcost",
    "cost-mode" : "ordinal",
    "map" : {
      "192.0.2.2": {
        "192.0.2.89" : 1,
        "198.51.100.34" : 2,
        "203.0.113.45" : 3
      }
    }
  }
}
```

## 8. Redistributable Responses

This section defines how an ALTO Server enables certain responses to be redistributed by ALTO Clients. Concepts are first introduced, followed by the protocol specification.

### 8.1. Concepts

#### 8.1.1. Service ID

The Service ID is a UUID that identifies a set of ALTO Servers that would provide identical ALTO Information for any ALTO Request for any ALTO Client. Each ALTO Server within such a set is configured with an identical Service ID.

If a pair of ALTO Servers would provide different ALTO Information in response to a particular ALTO Client request, then the pair of ALTO Servers MUST have a different Service ID.

##### 8.1.1.1. Rationale

For scalability and fault tolerance, multiple ALTO Servers may be deployed to serve equivalent ALTO Information. In such a scenario, ALTO Responses from any such redundant server should be seen as equivalent for the purposes of redistribution. For example, if two ALTO Servers A and B are deployed by the service provider to distribute equivalent ALTO Information, then clients contacting Server A should be able to redistribute ALTO Responses to clients contacting Server B.

To accomplish this behavior, ALTO Clients must be able to determine that Server A and Server B serve identical ALTO Information. One technique would be to rely on the ALTO Server's DNS name. However, such an approach would mandate that all ALTO Servers resolved by a particular DNS name would need to provide equivalent ALTO information, which may be unnecessarily restrictive. Another technique would be to rely on the server's IP address. However, this suffers similar problems as the DNS name in deployment scenarios using IP Anycast.

To avoid such restrictions, the ALTO Protocol allows an ALTO Service Provider to explicitly denote ALTO Servers that provide equivalent ALTO Information by giving them identical Service IDs. Service IDs decouple the identification of equivalent ALTO Servers from the discovery process.

#### 8.1.1.2. Server Capability Response

If an ALTO Server generates redistributable responses, the Server Capability response's 'service-id' field MUST be set to the ALTO Server's Service ID.

#### 8.1.1.3. Configuration

To help prevent ALTO Servers from mistakenly claiming to distribute equivalent ALTO Information, ALTO Server implementations SHOULD by default generate a new UUID at installation time or startup if one has not explicitly been configured.

#### 8.1.2. Expiration Time

ALTO Responses marked as redistributable should indicate a time after which the information is considered stale and should be refreshed from the ALTO Server (or possibly another ALTO Client).

If an expiration time is present, the ALTO Server SHOULD ensure that it is reasonably consistent with the expiration time that would be computed by HTTP header fields. This specification makes no recommendation on which expiration time takes precedence, but implementers should be cognizant that HTTP intermediaries will obey only the HTTP header fields.

#### 8.1.3. Signature

ALTO Responses marked as redistributable include a signature used to assert that the ALTO Server Provider generated the ALTO Information.

##### 8.1.3.1. Rationale

Verification of the signature requires the ALTO Client to retrieve the ALTO Server's public key. There are multiple possibilities through which the ALTO Protocol could be designed to retrieve it:

- o SSL/TLS connection with the ALTO Server: The public key algorithm and public key may be retrieved from the ALTO Server's X.509 Certificate used on an HTTPS connection between the ALTO Server and ALTO Client.
- o Included in ALTO Server's Server Capability Response: An X.509 certificate (including the public key and public key algorithm) can be included in the Server Capability Response. This could be achieved even if the ALTO Server and ALTO Client do not have a SSL/TLS channel.

To reduce requirements on the underlying transport (i.e., requiring SSL/TLS), the ALTO Protocol uses the latter option.

#### 8.1.3.2. Certificates

##### 8.1.3.2.1. Local Certificate

The ALTO Server's public key is encoded within an X.509 certificate. The corresponding private key **MUST** be used to sign redistributable responses. This certificate is termed the Local Certificate for an ALTO Server.

##### 8.1.3.2.2. Certificate Chain

To ease key provisioning, the ALTO Protocol is designed such that each ALTO Server with an identical Service ID may have a unique private key (and hence certificate).

The ALTO Service Provider may configure a certificate chain at each such ALTO Server. The Local Certificate for a single ALTO Server is the bottom-most certificate in the chain. The Certificate Chains of each ALTO Server with an identical Service ID **MUST** share a common Root Certificate.

Note that there are two simple deployment scenarios:

- o One-Level Certificate Chain (Local Certificate Only): In this deployment scenario, each ALTO Server with an identical Service ID may be provisioned with an identical Local Certificate.
- o Two-Level Certificate Chain: In this deployment scenario, a Root Certificate is maintained for a set of ALTO Servers with the same Service ID. A unique Local Certificate signed by this CA is provisioned to each ALTO Server.

There are advantages to using a Certificate Chain instead of deploying the same Local Certificate to each ALTO Server. Specifically, it avoids storage of the CA's private key at ALTO Servers. It is possible to revoke and re-issue a key to a single ALTO Server.

##### 8.1.3.2.3. Server Capability Response

If an ALTO Server generates redistributable responses, the Server Capability response's 'certificates' field **MUST** be populated with the ALTO Server's full certificate chain. The first element **MUST** be the ALTO Server's Local Certificate, followed by the remaining Certificate Chain in ascending order to the Root Certificate.

#### 8.1.3.3. Signature Verification

ALTO Clients SHOULD verify the signature on any ALTO information received via redistribution before adjusting application behavior based on it.

An ALTO Client SHOULD cache its ALTO Server's Service ID and corresponding Certificate Chain included in the Server Capability response. Recall that the last certificate in this chain is the Root Certificate. The retrieval of the Service ID and certificates SHOULD be secured using HTTPS with proper validation of the server endpoint of the SSL/TLS connection [6].

An ALTO Response received via redistribution from Service ID S is declared valid if an ALTO Client can construct a transitive certificate chain from the certificate (public key) used to sign the ALTO Response to the Root Certificate corresponding to Service ID S obtained by the ALTO Client in a Server Capability response.

To properly construct the chain and complete this validation, an ALTO Client may need to request additional certificates from other ALTO Clients. A simple mechanism is to request the certificate chain from the ALTO Client that received the ALTO Response. Note that these additional received certificates may be cached locally by an ALTO Client.

ALTO Clients SHOULD verify ALTO Responses received via redistribution.

#### 8.1.3.4. Redistribution by ALTO Clients

ALTO Clients SHOULD pass the ALTO Server Certificate, Signature, and Signature Algorithm along with the body of the ALTO Response. The mechanism for redistributing such information is not specified by the ALTO Protocol, but one possibility is to add additional messages or fields to the application's native protocol.

#### 8.2. Protocol

An ALTO Server MAY indicate that a response is suitable for redistribution by including the "redistribution" member in the RspMetaData JSON object of an ALTO Response message. This additional member, called the Response Redistribution Descriptor, has type RspRedistDesc:

```
object {  
    JSONString service-id;  
    JSONString request-uri;  
    JSONValue  request-body;  
    JSONString expires;  
} RspRedistDesc;
```

The fields encoded in the Response Redistribution Descriptor allows an ALTO Client receiving redistributed ALTO Information to understand the context of the query (the ALTO Service generating the response and any input parameters) and to interpret the results.

Information about ALTO Client performing the Request and any HTTP Headers passed in the request are not included in the Response Redistribution Descriptor. If any such information or headers influence the response generated by the ALTO Server, the response SHOULD NOT be indicated as redistributable.

#### 8.2.1. Response Redistribution Descriptor Fields

This section defines the fields of the Response Redistribution Descriptor.

##### 8.2.1.1. Service ID

The 'service-id' member is REQUIRED and MUST have a value equal to the ALTO Server's Service ID.

##### 8.2.1.2. Request URI

The 'request-uri' member is REQUIRED and MUST specify the HTTP Request-URI that was passed in the HTTP Request.

##### 8.2.1.3. Request Body

If the HTTP Request body was non-empty, the 'request-body' member MUST specify full JSON value passed in the HTTP Request (note that whitespace may differ, as long as the JSON Value is identical). If the HTTP Request was empty, then the 'request-body' MUST NOT be included.

##### 8.2.1.4. Expiration Time

The 'expires' element is RECOMMENDED and, if present, MUST specify a time in UTC formatted according to [9].

### 8.2.2. Signature

The Hash Algorithm, Signature Algorithm, and Signature are included as either HTTP Headers or Trailers. Headers may be useful if Responses are pre-generated, while Trailers may be useful if Responses are dynamically generated (e.g., to avoid buffering large responses in memory while the hash value is computed).

The following HTTP Headers (the ALTO Server MAY specify them as HTTP Trailers instead) MUST be used to encode the Signature parameters for redistributable ALTO Responses:

```
ALTO-HashAlgorithm: <HashAlgorithm>
ALTO-SignatureAlgorithm: <SignatureAlgorithm>
ALTO-SignatureDigest: <Signature>
```

where <HashAlgorithm> and <SignatureAlgorithm> are an integer values from the IANA TLS HashAlgorithm and SignatureAlgorithm registries, and <Signature> is the corresponding PEM-encoded signature.

## 9. Use Cases

The sections below depict typical use cases.

### 9.1. ALTO Client Embedded in P2P Tracker

Many P2P currently-deployed P2P systems use a Tracker to manage swarms and perform peer selection. P2P trackers may currently use a variety of information to perform peer selection to meet application-specific goals. By acting as an ALTO Client, an P2P tracker can use ALTO information as an additional information source to enable more network-efficient traffic patterns and improve application performance.

A particular requirement of many P2P trackers is that they must handle a large number of P2P clients. A P2P tracker can obtain and locally store ALTO information (the Network Map and Cost Map) from the ISPs containing the P2P clients, and benefit from the same aggregation of network locations done by ALTO Servers.



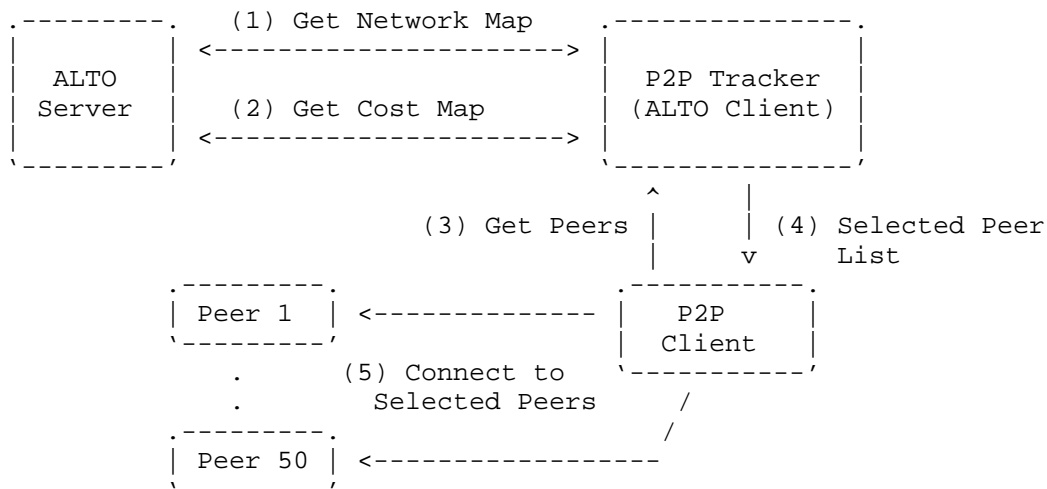


Figure 4: ALTO Client Embedded in P2P Tracker

Figure 4 shows an example use case where a P2P tracker is an ALTO Client and applies ALTO information when selecting peers for its P2P clients. The example proceeds as follows:

1. The P2P Tracker requests the Network Map covering all PIDs from the ALTO Server using the Network Map query. The Network Map includes the IP prefixes contained in each PID, allowing the P2P tracker to locally map P2P clients into a PIDs.
2. The P2P Tracker requests the Cost Map amongst all PIDs from the ALTO Server.
3. A P2P Client joins the swarm, and requests a peer list from the P2P Tracker.
4. The P2P Tracker returns a peer list to the P2P client. The returned peer list is computed based on the Network Map and Cost Map returned by the ALTO Server, and possibly other information sources. Note that it is possible that a tracker may use only the Network Map to implement hierarchical peer selection by preferring peers within the same PID and ISP.
5. The P2P Client connects to the selected peers.

Note that the P2P tracker may provide peer lists to P2P clients distributed across multiple ISPs. In such a case, the P2P tracker may communicate with multiple ALTO Servers.

## 9.2. ALTO Client Embedded in P2P Client: Numerical Costs

P2P clients may also utilize ALTO information themselves when selecting from available peers. It is important to note that not all P2P systems use a P2P tracker for peer discovery and selection. Furthermore, even when a P2P tracker is used, the P2P clients may rely on other sources, such as peer exchange and DHTs, to discover peers.

When an P2P Client uses ALTO information, it typically queries only the ALTO Server servicing its own ISP. The my-Internet view provided by its ISP's ALTO Server can include preferences to all potential peers.

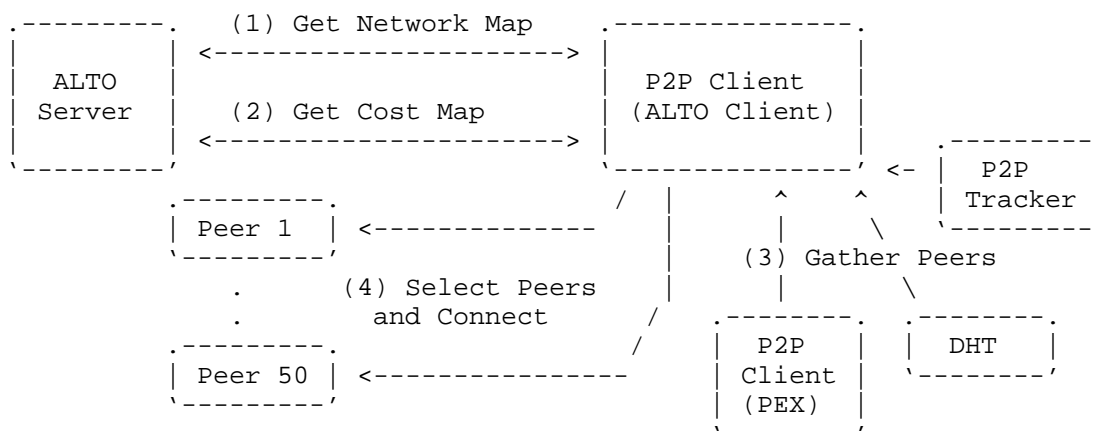


Figure 5: ALTO Client Embedded in P2P Client

Figure 5 shows an example use case where a P2P Client locally applies ALTO information to select peers. The use case proceeds as follows:

1. The P2P Client requests the Network Map covering all PIDs from the ALTO Server servicing its own ISP.
2. The P2P Client requests the Cost Map amongst all PIDs from the ALTO Server. The Cost Map by default specifies numerical costs.
3. The P2P Client discovers peers from sources such as Peer Exchange (PEX) from other P2P Clients, Distributed Hash Tables (DHT), and P2P Trackers.
4. The P2P Client uses ALTO information as part of the algorithm for selecting new peers, and connects to the selected peers.

### 9.3. ALTO Client Embedded in P2P Client: Ranking

It is also possible for a P2P Client to offload the selection and ranking process to an ALTO Server. In this use case, the ALTO Client gathers a list of known peers in the swarm, and asks the ALTO Server to rank them.

As in the use case using numerical costs, the P2P Client typically only queries the ALTO Server servicing its own ISP.

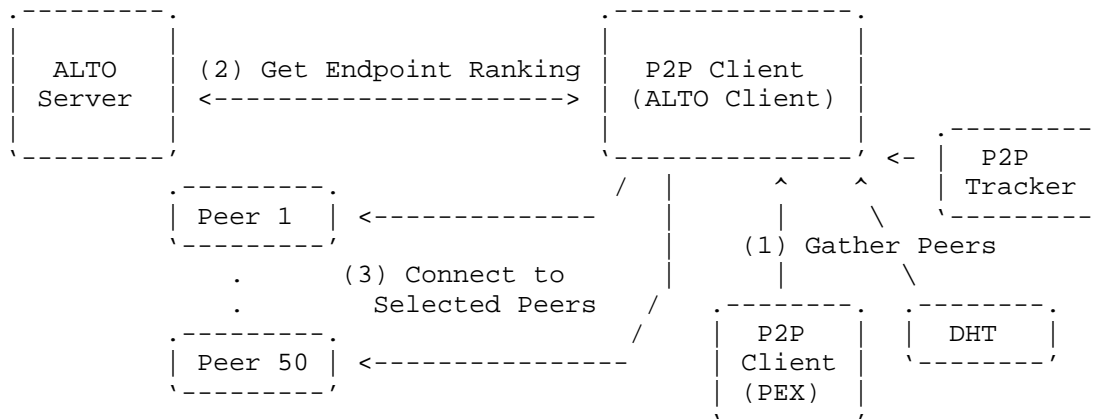


Figure 6: ALTO Client Embedded in P2P Client: Ranking

Figure 6 shows an example of this scenario. The use case proceeds as follows:

1. The P2P Client discovers peers from sources such as Peer Exchange (PEX) from other P2P Clients, Distributed Hash Tables (DHT), and P2P Trackers.
2. The P2P Client queries the ALTO Server's Ranking Service, including discovered peers as the set of Destination Endpoints, and indicates the 'ordinal' Cost Mode. The response indicates the ranking of the candidate peers.
3. The P2P Client connects to the peers in the order specified in the ranking.

## 10. Discussions

### 10.1. Discovery

The discovery mechanism by which an ALTO Client locates an appropriate ALTO Server is out of scope for this document. This document assumes that an ALTO Client can discover an appropriate ALTO Server. Once it has done so, the ALTO Client may use the Server List query Section 7.8.1.1 to locate an ALTO Server with capabilities necessary for its application.

### 10.2. Hosts with Multiple Endpoint Addresses

In practical deployments, especially during the transition from IPv4 to IPv6, a particular host may be reachable using multiple addresses. Furthermore, the particular network path followed when sending packets to the host may differ based on the address that is used. Network providers may prefer one path over another (e.g., one path may have a NAT64 middlebox). An additional consideration may be how to handle private address spaces (e.g., behind carrier-grade NATs).

Note that to support such behavior, Endpoints must be associated with a particular address type (e.g., IPv4 or IPv6). One simple possibility may be to prefix each endpoint address with its type (e.g., "ipv4:198.51.100.128/25"). However, we may want to discuss if a more efficient/compact encoding is possible in some cases (e.g., all addresses in the same PID are IPv6).

There are limitations as to what information ALTO can provide in this regard. In particular, a particular ALTO Service provider may not be able to determine if connectivity with a particular endhost will succeed over IPv4 or IPv6, as this may depend upon information unknown to the ISP such as particular application implementations.

Exploration of these issues is being considered in a separate Internet Draft [23]. Once a suitable solution emerges, it will be included in this document.

### 10.3. Network Address Translation Considerations

At this day and age of NAT v4<->v4, v4<->v6 [24], and possibly v6<->v6[25], a protocol should strive to be NAT friendly and minimize carrying IP addresses in the payload, or provide a mode of operation where the source IP address provide the information necessary to the server.

The protocol specified in this document provides a mode of operation where the source network location is computed by the ALTO Server (via the Endpoint Property Lookup interface) from the source IP address found in the ALTO Client query packets. This is similar to how some

P2P Trackers (e.g., BitTorrent Trackers - see "Tracker HTTP/HTTPS Protocol" in [26]) operate.

The ALTO client SHOULD use the Session Traversal Utilities for NAT (STUN) [10] to determine a public IP address to use as a source Endpoint address. If using this method, the host MUST use the "Binding Request" message and the resulting "XOR-MAPPED-ADDRESS" parameter that is returned in the response. Using STUN requires cooperation from a publicly accessible STUN server. Thus, the ALTO client also requires configuration information that identifies the STUN server, or a domain name that can be used for STUN server discovery. To be selected for this purpose, the STUN server needs to provide the public reflexive transport address of the host.

#### 10.4. Mapping IPs to ASNs

It may be desired for the ALTO Protocol to provide ALTO information including ASNs. Thus, ALTO Clients may need to identify the ASN for a Resource Provider to determine the cost to that Resource Provider.

Applications can already map IPs to ASNs using information from a BGP Looking Glass. To do so, they must download a file of about 1.5MB when compressed (as of October 2008, with all information not needed for IP to ASN mapping removed) and periodically (perhaps monthly) refresh it.

Alternatively, the Network Map query in the Map Filtering Service defined in this document could be extended to map ASNs into a set of IP prefixes. The mappings provided by the ISP would be both smaller and more authoritative.

For simplicity of implementation, it's highly desirable that clients only have to implement exactly one mechanism of mapping IPs to ASNs.

#### 10.5. Endpoint and Path Properties

An ALTO Server could make available many properties about Endpoints beyond their network location or grouping. For example, connection type, geographical location, and others may be useful to applications. The current draft focuses on network location and grouping, but the protocol may be extended to handle other Endpoint properties.

### 11. IANA Considerations

### 11.1. application/alto Media Type

This document requests the registration of a new media type:  
"application/alto":

Type name: application

Subtype name: alto

Required parameters: n/a

Optional parameters: n/a

Encoding considerations: Encoding considerations are identical to those specified for the 'application/json' media type. See [4].

Security considerations: Security considerations relating to the generation and consumption of ALTO protocol messages are discussed in Section 12.

Interoperability considerations: This document specifies format of conforming messages and the interpretation thereof.

Published specification: This document.

Applications that use this media type: ALTO Servers and ALTO Clients either standalone or embedded within other applications.

Additional information:

Magic number(s): n/a

File extension(s): This document uses the mime type to refer to protocol messages and thus does not require a file extension.

Macintosh file type code(s): n/a

Person & email address to contact for further information: See "Authors' Addresses" section.

Intended usage: COMMON

Restrictions on usage: n/a

Author: See "Authors' Addresses" section.

Change controller: See "Authors' Addresses" section.

## 11.2. ALTO Cost Type Registry

This document requests the creation of an ALTO Cost Type registry to be maintained by IANA.

This registry serves two purposes. First, it ensures uniqueness of identifiers referring to ALTO Cost Types. Second, it provides references to particular semantics of allocated Cost Types to be applied by both ALTO Servers and applications utilizing ALTO Clients.

New ALTO Cost Types are assigned after Expert Review [7]. The Expert Reviewer will generally consult the ALTO Working Group or its successor. Expert Review is used to ensure that proper documentation regarding ALTO Cost Type semantics and security considerations has been provided. The provided documentation should be detailed enough to provide guidance to both ALTO Service Providers and applications utilizing ALTO Clients as to how values of the registered ALTO Cost Type should be interpreted. Updates and deletions of ALTO Cost Types follow the same procedure.

Registered ALTO Cost Type identifiers MUST conform to the syntactical requirements specified in Section 7.7.3. Identifiers are to be recorded and displayed as ASCII strings.

Identifiers prefixed with 'priv:' are reserved for Private Use. Identifiers prefixed with 'exp:' are reserved for Experimental use.

Requests to add a new value to the registry MUST include the following information:

- o Identifier: The name of the desired ALTO Cost Type.
- o Intended Semantics: ALTO Costs carry with them semantics to guide their usage by ALTO Clients. For example, if a value refers to a measurement, the measurement units must be documented. For proper implementation of the ordinal Cost Mode (e.g., by a third-party service), it should be documented whether higher or lower values of the cost are more preferred.
- o Security Considerations: ALTO Costs expose information to ALTO Clients. As such, proper usage of a particular Cost Type may require certain information to be exposed by an ALTO Service Provider. Since network information is frequently regarded as proprietary or confidential, ALTO Service Providers should be made aware of the security ramifications related to usage of a Cost Type.

This specification requests registration of the identifier 'routingcost'. Semantics for the this Cost Type are documented in Section 5.1.1.1, and security considerations are documented in Section 12.1.

## 12. Security Considerations

### 12.1. Privacy Considerations for ISPs

ISPs must be cognizant of the network topology and provisioning information provided through ALTO Interfaces. ISPs should evaluate how much information is revealed and the associated risks. On the one hand, providing overly fine-grained information may make it easier for attackers to infer network topology. In particular, attackers may try to infer details regarding ISPs' operational policies or inter-ISP business relationships by intentionally posting a multitude of selective queries to an ALTO server and analyzing the responses. Such sophisticated attacks may reveal more information than an ISP hosting an ALTO server intends to disclose. On the other hand, revealing overly coarse-grained information may not provide benefits to network efficiency or performance improvements to ALTO Clients.

### 12.2. ALTO Clients

Applications using the information must be cognizant of the possibility that the information is malformed or incorrect. Even if an ALTO Server has been properly authenticated by the ALTO Client, the information provided may be malicious because the ALTO Server and its credentials have been compromised (e.g., through malware). Other considerations (e.g., relating to application performance) can be found in Section 6 of [18].

ALTO Clients should also be cognizant of revealing Network Location Identifiers (IP addresses or fine-grained PIDs) to the ALTO Server, as doing so may allow the ALTO Server to infer communication patterns. One possibility is for the ALTO Client to only rely on Network Map for PIDs and Cost Map amongst PIDs to avoid passing IP addresses of their peers to the ALTO Server.

In addition, ALTO clients should be cautious not to unintentionally or indirectly disclose the resource identifier (of which they try to improve the retrieval through ALTO-guidance), e.g., the name/identifier of a certain video stream in P2P live streaming, to the ALTO server. Note that the ALTO Protocol specified in this document does not explicitly reveal any resource identifier to the ALTO Server. However, for instance, depending on the popularity or other



specifics (such as language) of the resource, an ALTO server could potentially deduce information about the desired resource from information such as the Network Locations the client sends as part of its request to the server.

### 12.3. Authentication, Integrity Protection, and Encryption

SSL/TLS can provide encryption of transmitted messages as well as authentication of the ALTO Client and Server. HTTP Basic or Digest authentication can provide authentication of the client (combined with SSL/TLS, it can additionally provide encryption and authentication of the server).

An ALTO Server may optionally use authentication (and potentially encryption) to protect ALTO information it provides. This can be achieved by digitally signing a hash of the ALTO information itself and attaching the signature to the ALTO information. There may be special use cases where encryption of ALTO information is desirable. In many cases, however, information sent out by an ALTO Server may be regarded as non-confidential information.

ISPs should be cognizant that encryption only protects ALTO information until it is decrypted by the intended ALTO Client. Digital Rights Management (DRM) techniques and legal agreements protecting ALTO information are outside of the scope of this document.

### 12.4. ALTO Information Redistribution

It is possible for applications to redistribute ALTO information to improve scalability. Even with such a distribution scheme, ALTO Clients obtaining ALTO information must be able to validate the received ALTO information to ensure that it was generated by an appropriate ALTO Server. Further, to prevent the ALTO Server from being a target of attack, the verification scheme must not require ALTO Clients to contact the ALTO Server to validate every set of information. Contacting an ALTO server for information validation would also undermine the intended effect of redistribution and is therefore not desirable.

Note that the redistribution scheme must additionally handle details such as ensuring ALTO Clients retrieve ALTO information from the correct ALTO Server. See [21] for further discussion. Details of a particular redistribution scheme are outside the scope of this document.

To fulfill these requirements, ALTO Information meant to be redistributable contains a digital signature which includes a hash of

the ALTO information signed by the ALTO Server with its private key. The corresponding public key is included in the Server Capability response Section 7.8.1.2, along with the certificate chain to a Root Certificate generated by the ALTO Service Provider. To prevent man-in-the-middle attacks, an ALTO Client SHOULD perform the Server Capability Query over SSL/TLS and verify the server identity according to [6].

The signature verification algorithm is detailed in Section 8.1.3.3.

#### 12.5. Denial of Service

ISPs should be cognizant of the workload at the ALTO Server generated by certain ALTO Queries, such as certain queries to the Map Filtering Service and Ranking Service. In particular, queries which can be generated with low effort but result in expensive workloads at the ALTO Server could be exploited for Denial-of-Service attacks. For instance, a simple ALTO query with  $n$  Source Network Locations and  $m$  Destination Network Locations can be generated fairly easily but results in the computation of  $n*m$  Path Costs between pairs by the ALTO Server (see Section 5.2). One way to limit Denial-of-Service attacks is to employ access control to the ALTO server. Another possible mechanism for an ALTO Server to protect itself against a multitude of computationally expensive bogus requests is to demand that each ALTO Client to solve a computational puzzle first before allocating resources for answering a request (see, e.g., [27]). The current specification does not use such computational puzzles, and discussion regarding tradeoffs of such an approach would be needed before including such a technique in the ALTO Protocol.

ISPs should also leverage the fact that the the Map Service allows ALTO Servers to pre-generate maps that can be useful to many ALTO Clients.

#### 12.6. ALTO Server Access Control

In order to limit access to an ALTO server (e.g., for an ISP to only allow its users to access its ALTO server, or to prevent Denial-of-Service attacks by arbitrary hosts from the Internet), an ALTO server may employ access control policies. Depending on the use-case and scenario, an ALTO server may restrict access to its services more strictly or rather openly (see [28] for a more detailed discussion on this issue).

### 13. References

## 13.1. Normative References

- [1] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [2] Berners-Lee, T., Fielding, R., and H. Nielsen, "Hypertext Transfer Protocol -- HTTP/1.0", RFC 1945, May 1996.
- [3] 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.
- [4] Crockford, D., "The application/json Media Type for JavaScript Object Notation (JSON)", RFC 4627, July 2006.
- [5] Blake-Wilson, S., Nystrom, M., Hopwood, D., Mikkelsen, J., and T. Wright, "Transport Layer Security (TLS) Extensions", RFC 4366, April 2006.
- [6] Saint-Andre, P. and J. Hodges, "Representation and Verification of Domain-Based Application Service Identity within Internet Public Key Infrastructure Using X.509 (PKIX) Certificates in the Context of Transport Layer Security (TLS)", draft-saintandre-tls-server-id-check-10 (work in progress), October 2010.
- [7] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 5226, May 2008.
- [8] Leach, P., Mealling, M., and R. Salz, "A Universally Unique IDentifier (UUID) URN Namespace", RFC 4122, July 2005.
- [9] Klyne, G., Ed. and C. Newman, "Date and Time on the Internet: Timestamps", RFC 3339, July 2002.
- [10] Rosenberg, J., Mahy, R., Matthews, P., and D. Wing, "Session Traversal Utilities for (NAT) (STUN)", draft-ietf-behave-rfc3489bis-18 (work in progress), July 2008.

## 13.2. Informative References

- [11] Kiesel, S., Popkin, L., Previdi, S., Woundy, R., and Y. Yang, "Application-Layer Traffic Optimization (ALTO) Requirements", draft-kiesel-alto-reqs-01 (work in progress), November 2008.
- [12] Alimi, R., Pasko, D., Popkin, L., Wang, Y., and Y. Yang, "P4P: Provider Portal for P2P Applications", draft-p4p-framework-00 (work in progress), November 2008.

- [13] Wang, Y., Alimi, R., Pasko, D., Popkin, L., and Y. Yang, "P4P Protocol Specification", draft-wang-alto-p4p-specification-00 (work in progress), March 2009.
- [14] Shalunov, S., Penno, R., and R. Woundy, "ALTO Information Export Service", draft-shalunov-alto-infoexport-00 (work in progress), October 2008.
- [15] Das, S. and V. Narayanan, "A Client to Service Query Response Protocol for ALTO", draft-saumitra-alto-queryresponse-00 (work in progress), March 2009.
- [16] Das, S., Narayanan, V., and L. Dondeti, "ALTO: A Multi Dimensional Peer Selection Problem", draft-saumitra-alto-multi-ps-00 (work in progress), October 2008.
- [17] Akonjang, O., Feldmann, A., Previdi, S., Davie, B., and D. Saucez, "The PROXIDOR Service", draft-akonjang-alto-proxidior-00 (work in progress), March 2009.
- [18] Seedorf, J. and E. Burger, "Application-Layer Traffic Optimization (ALTO) Problem Statement", RFC 5693, October 2009.
- [19] Yang, Y., Popkin, L., Penno, R., and S. Shalunov, "An Architecture of ALTO for P2P Applications", draft-yang-alto-architecture-00 (work in progress), March 2009.
- [20] Zyp, K., "A JSON Media Type for Describing the Structure and Meaning of JSON Documents", draft-zyp-json-schema-02 (work in progress), March 2010.
- [21] Yingjie, G., Alimi, R., and R. Even, "ALTO Information Redistribution", draft-gu-alto-redistribution-03 (work in progress), July 2010.
- [22] 3rd, D., "Transport Layer Security (TLS) Extensions: Extension Definitions", draft-ietf-tls-rfc4366-bis-12 (work in progress), September 2010.
- [23] Penno, R. and J. Medved, "ALTO and IPv4/IPv6 Co-existence and Transition", draft-penno-alto-ipv4v6-00 (work in progress), June 2010.
- [24] Baker, F., Li, X., and C. Bao, "Framework for IPv4/IPv6 Translation", draft-baker-behave-v4v6-framework-02 (work in progress), February 2009.

- [25] Wasserman, M. and F. Baker, "IPv6-to-IPv6 Network Address Translation (NAT66)", draft-mrw-behave-nat66-02 (work in progress), March 2009.
- [26] "Bittorrent Protocol Specification v1.0", <http://wiki.theory.org/BitTorrentSpecification>, 2009.
- [27] Jennings, C., "Computational Puzzles for SPAM Reduction in SIP", draft-jennings-sip-hashcash-06 (work in progress), July 2007.
- [28] Stiernerling, M. and S. Kiesel, "ALTO Deployment Considerations", draft-stiernerling-alto-deployments-05 (work in progress), October 2010.
- [29] H. Xie, YR. Yang, A. Krishnamurthy, Y. Liu, and A. Silberschatz., "P4P: Provider Portal for (P2P) Applications", In SIGCOMM 2008.

#### Appendix A. TO BE MOVED

The text in this section is intended to be moved to a more appropriate document.

##### A.1. Discovery

Some ISPs have proposed the possibility of delegation, in which an ISP provides information for customer networks which do not wish to run ALTO Servers themselves. A consideration for delegation is that customer networks may wish to explicitly configure such delegation.

##### A.2. P2P Peer Selection

This section discusses possible approaches to peer selection using ALTO information (Network Location Identifiers and associated Costs) from an ALTO Server. Specifically, the application must select which peers to use based on this and other sources of information. With this in mind, the usage of ALTO Costs is intentionally flexible, because:

Different applications may use the information differently. For example, an application that connects to just one address may have a different algorithm for selecting it than an application that connects to many.

Though initial experiments have been conducted [29], more investigation is needed to identify other methods.

In addition, the application might account for robustness, perhaps using randomized exploration to determine if it performs better without ALTO information.

#### A.2.1. Client-based Peer Selection

One possibility is for peer selection using ALTO costs to be done entirely by a P2P client. The following are some techniques have been proposed and/or used:

- o Prefer network locations with lower ordinal rankings (i.e., higher priority) [17] [14].
- o Optimistically unchoking low-cost peers with higher probability [14].

#### A.2.2. Server-based Peer Selection

Another possibility is for ALTO costs to be used by an Application Tracker (e.g., BitTorrent Tracker) when returning peer lists. The following are techniques that have been proposed and/or used:

- o Using bandwidth matching (e.g., at an Application Tracker) and choosing solution (within bound of optimal) with minimal network cost [29].

#### A.2.3. Location-Only Peer Selection

This section discusses a promising peer selection algorithm that was recently used in experiments with a P2P live streaming application.

Experiments in the context of live streaming have shown significant benefits of a simple "location-only" algorithm that primarily makes use of the Network Map. A benefit of this algorithm is that it can provide a simple integration path for applications wishing to utilize ALTO.

In particular, the algorithm proceeds as follows to select an ordered list of peers for a particular incoming (or existing peer):

1. Insert into the result list a number (up to a threshold) of peers from the same PID as the incoming peer.
2. Insert into the result list a number (up to a threshold) of peers from the same ISP as the incoming peer.
3. Insert into the result list a number (up to a threshold) of peers from different ISPs than the incoming peer.

In the experiments, this algorithm was implemented at a tracker and executed for peer selection when peers initially join and when requesting new peers.

This algorithm makes two assumptions about the preferences communicated by the Network Map:

- o The ISP prefers peers within the same PID to peer with each other (see Section 4); and
- o The ALTO Client can distinguish between peers within the same ISP and peers outside of the ISP. In implementation at the ALTO Client, it may estimate a threshold based on costs read from the Cost Map.

#### Appendix B. Acknowledgments

Thank you to Jan Seedorf for contributions to the Security Considerations section. We would like to thank Yingjie Gu and Roni Even for helpful input and design concerning ALTO Information redistribution.

We would like to thank the following people whose input and involvement was indispensable in achieving this merged proposal:

Obi Akonjang (DT Labs/TU Berlin),  
Saumitra M. Das (Qualcomm Inc.),  
Syon Ding (China Telecom),  
Doug Pasko (Verizon),  
Laird Popkin (Pando Networks),  
Satish Raghunath (Juniper Networks),  
Albert Tian (Ericsson/Redback),  
Yu-Shun Wang (Microsoft),  
David Zhang (PPLive),  
Yunfei Zhang (China Mobile).

We would also like to thank the following additional people who were involved in the projects that contributed to this merged document:

Alex Gerber (AT&T), Chris Griffiths (Comcast), Ramit Hora (Pando Networks), Arvind Krishnamurthy (University of Washington), Marty Lafferty (DCIA), Erran Li (Bell Labs), Jin Li (Microsoft), Y. Grace Liu (IBM Watson), Jason Livingood (Comcast), Michael Merritt (AT&T), Ingmar Poesse (DT Labs/TU Berlin), James Royalty (Pando Networks), Damien Saucez (UCL) Thomas Scholl (AT&T), Emilio Sepulveda (Telefonica), Avi Silberschatz (Yale University), Hassan Sipra (Bell Canada), Georgios Smaragdakis (DT Labs/TU Berlin), Haibin Song (Huawei), Oliver Spatscheck (AT&T), See-Mong Tang (Microsoft), Jia Wang (AT&T), Hao Wang (Yale University), Ye Wang (Yale University), Haiyong Xie (Yale University).

#### Appendix C. Authors

[[Comment.1: RFC Editor: Please move information in this section to the Authors' Addresses section at publication time.]]

Stefano Previdi  
Cisco

Email: sprevidi@cisco.com

Stanislav Shalunov  
BitTorrent

Email: shalunov@bittorrent.com

Richard Woundy  
Comcast

Richard\_Woundy@cable.comcast.com

#### Authors' Addresses

Richard Alimi (editor)  
Google  
1600 Amphitheatre Parkway  
Mountain View CA  
USA

Email: ralimi@google.com



Reinaldo Penno (editor)  
Juniper Networks  
1194 N Mathilda Avenue  
Sunnyvale CA  
USA

Email: rpenno@juniper.net

Y. Richard Yang (editor)  
Yale University  
51 Prospect St  
New Haven CT  
USA

Email: yry@cs.yale.edu



Network Working Group  
Internet-Draft  
Intended status: Informational  
Expires: April 28, 2011

S. Kiesel, Ed.  
University of Stuttgart  
S. Previdi  
Cisco Systems, Inc.  
M. Stiemerling  
NEC Europe Ltd.  
R. Woundy  
Comcast Corporation  
Y R. Yang  
Yale University  
October 25, 2010

Application-Layer Traffic Optimization (ALTO) Requirements  
draft-ietf-alto-reqs-06.txt

Abstract

Many Internet applications are used to access resources, such as pieces of information or server processes, which 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, which have to select one or several hosts from a set of candidates, that are able to provide 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 resource consumption in the underlying network infrastructure.

This document enumerates requirements for ALTO, which should be considered when specifying, assessing, or comparing protocols and implementations, and it solicits feedback and discussion.

Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/lid-abstracts.txt>.

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on April 28, 2011.

#### Copyright Notice

Copyright (c) 2010 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 BSD License.

## Table of Contents

1. Introduction . . . . .	4
2. Terminology and Architectural Framework . . . . .	5
2.1. Requirements Notation . . . . .	5
2.2. ALTO Terminology . . . . .	5
2.3. Architectural Framework for ALTO . . . . .	6
2.4. Sample Use Cases . . . . .	6
3. ALTO Requirements . . . . .	9
3.1. ALTO Client Protocol . . . . .	9
3.1.1. General Requirements . . . . .	9
3.1.2. Host Group Descriptor Support . . . . .	9
3.1.3. Rating Criteria Support . . . . .	10
3.1.4. Placement of Entities and Timing of Transactions . . . . .	11
3.1.5. Protocol Extensibility . . . . .	13
3.1.6. Error Handling and Overload Protection . . . . .	13
3.2. ALTO Server Discovery . . . . .	14
3.3. Security and Privacy . . . . .	15
4. Host Group Descriptors . . . . .	16
5. Rating Criteria . . . . .	17
5.1. Distance-related Rating Criteria . . . . .	17
5.2. Charging-related Rating Criteria . . . . .	17
5.3. Performance-related Rating Criteria . . . . .	18
5.4. Inappropriate Rating Criteria . . . . .	19
6. IANA Considerations . . . . .	20
7. Security Considerations . . . . .	21
7.1. High-level security considerations . . . . .	21
7.2. Classification of Information Disclosure Scenarios . . . . .	21
7.3. Security Requirements . . . . .	23
8. References . . . . .	24
8.1. Normative References . . . . .	24
8.2. Informative References . . . . .	24
Appendix A. Contributors . . . . .	25
Appendix B. Acknowledgments . . . . .	26
Authors' Addresses . . . . .	27

## 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 which can help peer-to-peer (P2P) applications to 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 resource consumption in 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 logical 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. They 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 operated by the network provider, 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, Overlay Network, Application protocol, Peer, P2P, Resource, Resource Identifier, Resource Provider, Resource Consumer, Resource Directory, Transport Address, ALTO Service, ALTO Server, ALTO Client, ALTO Client Protocol, ALTO Query, ALTO Reply, ALTO Transaction, Provisioning protocol, Inter ALTO-Server Protocol, Local Traffic, Peering Traffic, Transit Traffic.

Furthermore, the following additional terms will be used:

- o Host Group Descriptor: Information used to describe the resource consumer which seeks ALTO guidance, or one or several candidate resource providers. This can be, for example, a single IP address, an address prefix or address range that contains the host(s), or an autonomous system (AS) number. Different options 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. For a discussion, see Section 4.
- o Host Characteristics Attribute: Properties of a host (other than the host group descriptor), in particular related to its attachment to the network. This information may be stored in the ALTO server and transmitted in the ALTO protocol. It may be evaluated according to the rating criteria.
- o 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, i. e., "low topological distance from a given resource consumer", which can be described by means of a host group descriptor.

### 2.3. Architectural Framework for ALTO

There are various architectural options how ALTO could be implemented, and specifying or mandating one specific architecture is out of the scope of this document.

The ALTO Working Group Charter [ALTO-charter] itemizes several key components, which shall be elaborated and specified by the ALTO Working Group. The ALTO problem statement [RFC5693] defines a terminology (see Section 2.2) and presents a figure that gives a high-level overview of protocol interaction between ALTO elements.

This document itemizes requirements for the following components of the abovementioned architecture:

- o The ALTO client protocol, which is used for sending ALTO queries and ALTO replies between ALTO client and ALTO server.
- o The discovery mechanism, which will be used by ALTO clients in order to find out where to send ALTO requests.
- o The overall architecture, especially with respect to security and privacy issues.

Furthermore, this document describes the following data structures, which might be used in the ALTO client protocol:

- o Host group descriptors, which are used to describe the location of a host in the network topology.
- o Rating criteria, i. e., conditions that shall be evaluated in order to generate the ALTO guidance.

Requirements regarding other components are not considered in the current version of this document, but may be added later.

### 2.4. Sample Use Cases

The ALTO problem statement [RFC5693] presents a figure that gives a high-level overview of protocol interaction between ALTO elements. The following figures are somewhat more elaborated and extended versions of it, in order to give some non-normative examples of ALTO usage. It can also be seen that, in some use cases, some of the requirements presented in later sections are more relevant than in others.

Figure 1 shows an ALTO use case with a DHT-based P2P application. Using this distributed lookup mechanism, a peer can figure out which



other peers are candidate resource providers for a desired resource. Every peer software includes an ALTO client, in order to request and receive guidance on peer selection from the ALTO servers.

From an ALTO perspective this means that the ALTO servers will receive ALTO queries from a rather large number of different ALTO clients. The performance of many clients and their Internet connectivity may be rather limited and therefore, this puts certain restrictions on the amount of guiding data that can be sent to them. Furthermore, the privacy-sensitive IP addresses of the peers are visible to the (operators of the) ALTO servers, as these are also the source addresses of the ALTO query messages.

Figure 2 shows an ALTO use case with a P2P application that makes use of a centralized resource directory (in some specific P2P implementations called a "tracker"). In this scenario the ALTO servers receive queries only from few entities, i.e., the resource directories. As these resource directories must be powerful machines anyway, it may be reasonable to send large amounts of ALTO guidance data to them, which will be cached there. Furthermore, in this scenario it may be possible to hide the exact addresses of the peers from the ALTO server.

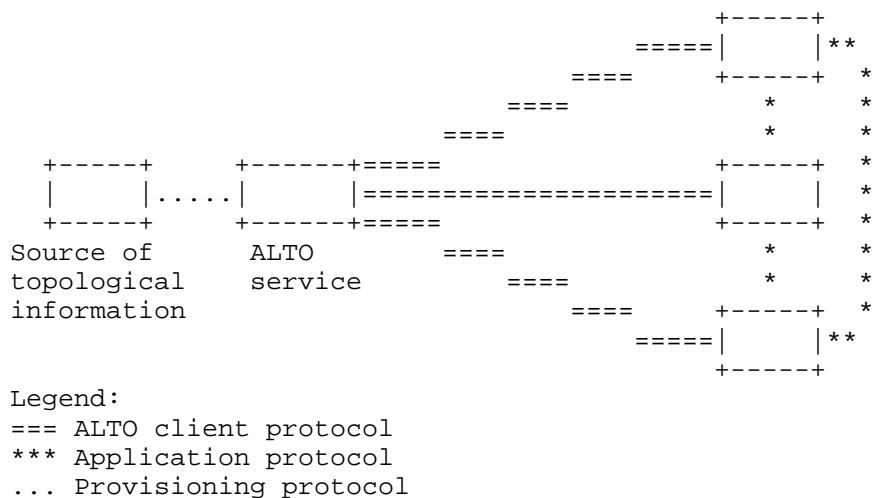
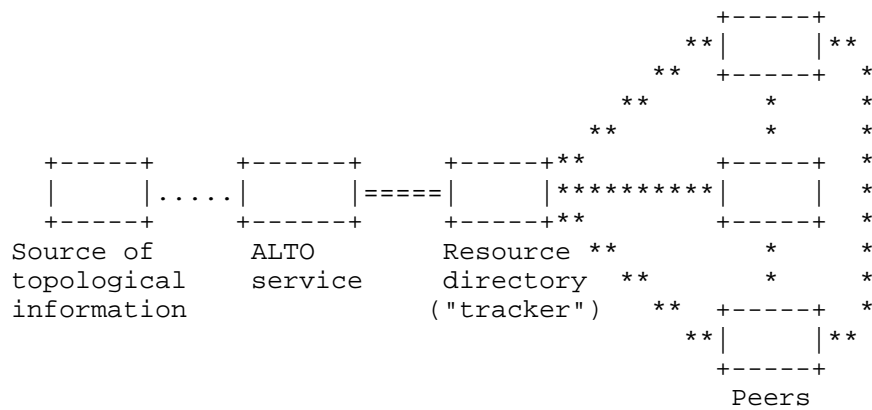


Figure 1: Overview of protocol interaction between ALTO elements, scenario without resource directory



Legend:

=== ALTO client protocol

\*\*\* Application protocol

... Provisioning protocol

Figure 2: Overview of protocol interaction between ALTO elements, scenario with resource directory

### 3. ALTO Requirements

#### 3.1. ALTO Client Protocol

##### 3.1.1. General Requirements

REQ. ARv06-1: The ALTO service is provided by one or more ALTO servers. ALTO servers MUST implement the ALTO client protocol, for receiving ALTO queries from ALTO clients and for sending the corresponding ALTO replies.

REQ. ARv06-2: ALTO clients MUST implement the ALTO client protocol, for sending ALTO queries to ALTO servers and for receiving the corresponding ALTO replies.

REQ. ARv06-3: The format of the ALTO query message MUST allow the ALTO client to solicit guidance for selecting appropriate resource providers.

REQ. ARv06-4: The format of the ALTO reply message MUST allow the ALTO server to express its guidance for selecting appropriate resource providers.

REQ. ARv06-5: The detailed specification of a protocol is out of the scope of this document. However, any protocol specification that claims to implement the ALTO client protocol MUST be compliant to the requirements itemized in this document.

##### 3.1.2. Host Group Descriptor Support

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

REQ. ARv06-6: The ALTO client protocol MUST support the usage of several different host group descriptor types.

REQ. ARv06-7: The ALTO client protocol specification MUST define a basic set of host group descriptor types, which MUST be supported by all implementations of the ALTO client protocol.

REQ. ARv06-8: The 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 CIDR notation), which contains all hosts in question. It is also possible to specify a broader address range (i.e., a shorter prefix length) than the intended group of hosts

actually uses, in order to conceal their exact identity.

REQ. ARv06-9: The ALTO client protocol specification MUST define an appropriate procedure for adding new host group descriptor types, e.g., by establishing an IANA registry.

See Section 4 for a discussion of possible other host group descriptor types.

REQ. ARv06-10: ALTO clients and ALTO servers MUST clearly identify the type of each host group descriptor sent in ALTO queries or replies.

REQ. ARv06-11: 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, which can be used to translate host group descriptors of that type to IPv4/IPv6 address prefixes, e.g., by means of a mapping table or an algorithm.

REQ. ARv06-12: Protocol functions for mapping other host group descriptor types to IPv4/IPv6 address prefixes SHOULD be designed and specified as part of the 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 the 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 downloaded as file via HTTP.

REQ. ARv06-13: The ALTO client protocol specification MUST define mechanisms, which can be used by the ALTO client and the ALTO server to indicate that a host group descriptor used by the other party is of an unsupported type, or that the indicated mapping mechanism could not be used.

### 3.1.1.3. Rating Criteria Support

REQ. ARv06-14: The ALTO client protocol MUST support the usage of several different rating criteria types.

REQ. ARv06-15: The ALTO client protocol specification MUST define a basic set of rating criteria types, which MUST be supported by all implementations of the ALTO client protocol.

REQ. ARv06-16: The ALTO client protocol specification MUST support the rating criteria type "relative operator's preference." This is a relative measure, i.e., it is not associated with any unit of measurement. A higher rating according to this criterion indicates

that the application should prefer the respective candidate resource provider over others with lower ratings (if no other reasons speak against it, 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. ARv06-17: The ALTO client protocol specification MUST define an appropriate procedure for adding new rating criteria types, e.g., by establishing an IANA registry.

See Section 5 for a discussion of possible other rating criteria.

REQ. ARv06-18: The ALTO query message SHOULD allow the ALTO client to express which rating criteria should be considered, as well as their relative relevance for the specific application that will eventually make use of the guidance.

REQ. ARv06-19: The ALTO reply message SHOULD allow the ALTO server to express which rating criteria have been considered when generating the reply.

REQ. ARv06-20: The ALTO client protocol specification MUST define mechanisms, which 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:

- o One mode of ALTO operation is that ALTO clients may be embedded directly in the resource consumer (e.g., peer of a DHT-based P2P application), which wants to access a resource.
- o Another mode of operation is to perform ALTO queries indirectly, via resource directories (e.g., tracker of a P2P application), which may issue ALTO queries to solicit preference on potential resource providers, considering the respective resource consumer.

REQ. ARv06-21: The ALTO client protocol MUST support the mode of operation, in which the ALTO client is directly embedded in the resource consumer.

REQ. ARv06-22: The ALTO client protocol MUST support the mode of operation, in which the ALTO client is embedded in the resource

directory.

REQ. ARv06-23: The ALTO client protocol MUST be designed in a way that the ALTO service can be provided by an entity which is not the operator of the IP access network.

REQ. ARv06-24: The ALTO client protocol MUST be designed in a way that different instances of the ALTO service operated by different providers can coexist.

With respect to the timing of ALTO queries, several modes of operation exist:

- o In target-aware query mode, an ALTO client performs the ALTO query when the desired resource and a set of candidate resource providers are already known, i. e., after DHT lookups, queries to the resource directory, etc.
- o In target-independent query mode, ALTO queries are performed in advance or periodically, in order to receive comprehensive, "target-independent" guidance, which will be cached locally and evaluated later, when a resource is to be accessed.

REQ. ARv06-25: The ALTO client protocol MUST support at least one of these two modes, either the target-aware or the target-independent query mode.

REQ. ARv06-26: The ALTO client protocol SHOULD support both the target-aware and the target-independent query mode.

REQ. ARv06-27: The ALTO client protocol SHOULD support lifetime attributes, to enable caching of recommendations at ALTO clients.

REQ. ARv06-28: The ALTO client protocol SHOULD specify an aging mechanism, which allows to give newer recommendations precedence over older ones.

REQ. ARv06-30: The ALTO client protocol SHOULD allow the ALTO server to add information about appropriate modes of re-use to its ALTO replies. Re-use may include redistributing an ALTO reply to other parties, as well as using the same ALTO information in a resource directory to improve the replies to different resource consumers, within the specified lifetime of the ALTO reply. The ALTO server SHOULD be able to express that

- o no re-use should occur

- o re-use 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 reply, which is only a subset of the known actual "target audience", e.g., if required by operator policies
- o re-use is appropriate for any resource consumer that would send (or cause a third party sending 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
- o re-use is appropriate for any resource consumer that would send (or cause a third party sending 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

REQ. ARv06-31: The ALTO client protocol MUST support scenarios with the ALTO client 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. ARv06-32: The ALTO client protocol MUST include support for adding protocol extensions in a non-disruptive, backward-compatible way.

REQ. ARv06-33: The 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. ARv06-34: Any application designed to use ALTO MUST also work if no ALTO servers can be found or if no responses to ALTO queries are received, e.g., due to connectivity problems or overload situation.

REQ. ARv06-35: The ALTO client protocol MUST use TCP based transport.

REQ. ARv06-36: An ALTO server, which is operating close to its capacity limit, MUST be able to inform clients about its impending overload situation, and require them to throttle their query rate.

REQ. ARv06-37: An ALTO server, which is operating close to its capacity limit, MUST be able to inform clients about its impending overload situation, and redirect them to another ALTO server.

REQ. ARv06-38: An ALTO server, which is operating close to its capacity limit, MUST be able to inform clients about its impending overload situation, and terminate the conversation with the ALTO client.

REQ. ARv06-39: An ALTO server, which is operating close to its capacity limit, MUST be able to inform clients about its impending overload situation, and reject new conversation attempts.

### 3.2. ALTO Server Discovery

The ALTO client protocol is supported by one or several ALTO server discovery mechanisms, which will be used by ALTO clients in order to find out where to send ALTO requests.

REQ. ARv06-40: ALTO clients which are embedded in the resource consumer MUST be able to use the ALTO server discovery mechanism, in order to find one or several ALTO servers that can provide ALTO guidance suitable for the resource consumer. This mode of operation is called "resource consumer initiated ALTO server discovery".

REQ. ARv06-41: ALTO clients which are embedded in a resource directory and perform third-party ALTO queries on behalf of a remote resource consumer MUST be able to use the ALTO server discovery mechanism, in order to find one or several ALTO servers that can provide ALTO guidance suitable for the respective resource consumer. This mode of operation is called "third-party ALTO server discovery".

REQ. ARv06-42: ALTO clients MUST be able to perform resource consumer initiated ALTO server discovery, even if they are located behind a network address translator (NAT).

REQ. ARv06-43: ALTO clients MUST be able to perform third-party ALTO server discovery, even if they are located behind a network address translator (NAT).

REQ. ARv06-44: 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 network address translator (NAT).

REQ. ARv06-45: The ALTO server discovery mechanism may be specified and provided using an existing protocol or mechanism, such as DNS, DHCP, or PPP based automatic configuration, etc. These candidate "base protocols" differ with respect to their availability in various access network architectures and their suitability for third-party queries. When evaluating different options this should be taken into account, in order to limit the total number of ALTO server discovery



mechanisms that have to be specified for supporting a reasonably wide range of deployment scenarios.

REQ. ARv06-46: The ALTO server discovery mechanism SHOULD be able to return the respective contact information for several ALTO servers.

REQ. ARv06-47: The ALTO server discovery mechanism SHOULD be able to indicate preferences for each returned ALTO server contact information.

### 3.3. Security and Privacy

REQ. ARv06-48: The ALTO client protocol MUST support mechanisms for the authentication of ALTO servers.

REQ. ARv06-49: The ALTO client protocol MUST support mechanisms for the authentication of ALTO clients.

REQ. ARv06-50: The ALTO client protocol MUST support different levels of detail in queries and responses, in order for the operator of an ALTO service to be able to control how much information (e.g., about the network topology) is disclosed.

REQ. ARv06-51: The operator of an ALTO server MUST NOT assume that an ALTO client will implement mechanisms or comply with rules that limit the ALTO client's ability to redistribute information retrieved from the ALTO server to third parties.

REQ. ARv06-52: The 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. ARv06-53: The ALTO client protocol SHOULD be defined in a way, that the operator of one ALTO server cannot easily deduce the resource identifier (e.g., file name in P2P file sharing) which the resource consumer seeking ALTO guidance wants to access.

REQ. ARv06-54: The ALTO client protocol MUST include appropriate mechanisms to protect the ALTO service against DoS attacks.

#### 4. Host Group Descriptors

Host group descriptors are used in the ALTO client protocol to describe the location of a host in the network topology. The ALTO client protocol specification defines a basic set of host group descriptor types, which have to be supported by all implementations, and an extension procedure for adding new descriptor types (see Section 3.1.2). The following list gives an overview on further host group descriptor types that have been proposed in the past, or which are in use by ALTO-related prototype implementations. This list is not intended as normative text. Instead, the only purpose of the following list is to document the descriptor types that have been proposed so far, and to solicit further feedback and discussion:

- o Autonomous System (AS) number
- o Protocol-specific group identifiers, which expand to a set of IP address ranges (CIDR) and/or AS numbers. In one specific solution proposal, these are called Partition ID (PID).

## 5. Rating Criteria

Rating criteria are used in the ALTO client protocol to express topology- or connectivity-related properties, which are evaluated in order to generate the ALTO guidance. The ALTO client protocol specification defines a basic set of rating criteria, which have to be supported by all implementations, and an extension procedure for adding new criteria (see Section 3.1.3). The following list gives an overview on further rating criteria that have been proposed in the past, or which are in use by ALTO-related prototype implementations. This list is not intended as normative text. Instead, the only purpose of the following list is to document the rating criteria that have been proposed so far, and to solicit further feedback and discussion:

### 5.1. Distance-related Rating Criteria

- o Relative topological distance: relative means that a larger numerical value means greater distance, but it is up to the ALTO service how to compute the values, and the ALTO client will not be informed about the nature of the information. One way of generating this kind of information MAY be counting AS hops, but when querying this parameter, the ALTO client MUST NOT assume that the numbers actually are AS hops.
- o Absolute topological distance, expressed in the number of traversed autonomous systems (AS).
- o Absolute topological distance, expressed in the number of router hops (i.e., how much the TTL value of an IP packet will be decreased during transit).
- o Absolute physical distance, based on knowledge of the approximate geolocation (continent, country) of an IP address.

### 5.2. Charging-related Rating Criteria

- o Traffic volume caps, in case the Internet access of the resource consumer is not charged by "flat rate". For each candidate resource provider, the ALTO service could indicate the amount of data that may be transferred from/to this resource provider until a given point in time, and how much of this amount has already been consumed. Furthermore, it would have to be indicated how excess traffic would be handled (e.g., blocked, throttled, or charged separately at an indicated price). The interaction of several applications running on a host, out of which some use this criterion while others don't, as well as the evaluation of this criterion in resource directories, which issue ALTO queries on

behalf of other peers, are for further study.

### 5.3. Performance-related Rating Criteria

The following rating criteria are subject to the remarks below.

- o The minimum achievable throughput between the resource consumer and the candidate resource provider, which is considered useful by the application (only in ALTO queries), or
- o An arbitrary upper bound for the throughput from/to the candidate resource provider (only in ALTO replies). This may be, but is not necessarily the provisioned access bandwidth of the candidate resource provider.
- o The maximum round-trip time (RTT) between resource consumer and the candidate resource provider, which is acceptable for the application for useful communication with the candidate resource provider (only in ALTO queries), or
- o An arbitrary lower bound for the RTT between resource consumer and the candidate resource provider (only in ALTO replies). This may be, for example, based on measurements of the propagation delay in a completely unloaded network.

The ALTO client MUST be aware, that with high probability, the actual performance values differ significantly from these upper and lower bounds. In particular, an ALTO client MUST NOT consider the "upper bound for throughput" parameter as a permission to send data at the indicated rate without using congestion control mechanisms.

The discrepancies are due to various reasons, including, but not limited to the facts that

- o the ALTO service is not an admission control system
- o the ALTO service may not know the instantaneous congestion status of the network
- o the ALTO service may not know all link bandwidths, i.e., where the bottleneck really is, and there may be shared bottlenecks
- o the ALTO service may not know whether the candidate peer itself is overloaded
- o the ALTO service may not know whether the candidate peer throttles the bandwidth it devotes for the considered application

- o the ALTO service may not know whether the candidate peer will throttle the data it sends to us (e.g., because of some fairness algorithm, such as tit-for-tat)

Because of these inaccuracies and the lack of complete, instantaneous state information, which are inherent to the ALTO service, the application must use other mechanisms (such as passive measurements on actual data transmissions) to assess the currently achievable throughput, and it **MUST** use appropriate congestion control mechanisms in order to avoid a congestion collapse. Nevertheless, these rating criteria may provide a useful shortcut for quickly excluding candidate resource providers from such probing, if it is known in advance that connectivity is in any case worse than what is considered the minimum useful value by the respective application.

#### 5.4. Inappropriate Rating Criteria

Rating criteria that **SHOULD NOT** be defined for and used by the ALTO service include:

- o Performance metrics that are closely related to the instantaneous congestion status. The definition of alternate approaches for congestion control is explicitly out of the scope of ALTO. Instead, other appropriate means, such as using TCP based transport, have to be used to avoid congestion.

## 6. IANA Considerations

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

## 7. Security Considerations

### 7.1. 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].

### 7.2. Classification of Information Disclosure Scenarios

The unwanted disclosure of information is one key concern related to ALTO. The following list gives a classification of information disclosure scenarios, which may be considered more or less critical by different parties:

- o (1) Excess disclosure of 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 somehow obfuscated information to an ALTO client.
- o (2) Disclosure of the application behavior to the 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.
- o (3) Disclosure of ALTO server operator's data (e.g., network topology information) to an unauthorized third party. There are a couple of sub-cases here:
  - \* (3a) An ALTO server sends the information directly to an unauthorized ALTO client.
  - \* (3b) An unauthorized party snoops on the data transmission from the ALTO server to an authorized ALTO client.
  - \* (3c) An authorized ALTO client knowingly forwards the information it had received from the ALTO server to an unauthorized party.

- o (4) Disclosure of the application behavior to an unauthorized third party.
- o (5) Excess retrieval of ALTO server operator's data by collaborating ALTO clients. Several authorized ALTO clients could ask an ALTO server for guidance, and redistribute the replies among each other (see also case 3c). By correlating the ALTO replies they could find out more information than intended to be disclosed by the ALTO server operator.

(1) may be addressed by the ALTO server operator choosing the level of detail of the information to be populated into the ALTO server. Furthermore, access control mechanisms for filtering ALTO replies 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 (3c) and (5), see below.

(2) is addressed 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, (2) can be addressed by ALTO clients by obfuscating the identity of candidate resource consumers, e.g., by zeroing-out or randomizing the last few bits of the IP addresses. However, there is the potential side effect of yielding inaccurate results.

(3a), (3b), and (4) 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 (3c) and (5), see below.

Straightforward authentication and encryption schemes won't help solving (3c) and (5), 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, and therefore they are not considered as a viable solution. That is, ALTO server operators must be aware that (3c) and (5) cannot be prevented from happening, and therefore they should feed only such data into an ALTO server, which they do not consider sensitive with respect to (3c) and (5).

These insights are reflected by the requirements presented in this document.



### 7.3. Security Requirements

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

## 8. References

### 8.1. Normative References

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

### 8.2. Informative References

- [ALTO-charter] Marocco, E. and V. Gurbani, "Application-Layer Traffic Optimization (ALTO) Working Group Charter", February 2009.
- [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.
- [RFC5693] Seedorf, J. and E. Burger, "Application-Layer Traffic Optimization (ALTO) Problem Statement", RFC 5693, October 2009.

## Appendix A. Contributors

The authors were supported by the following people, who have contributed to this document:

- o Richard Alimi <richard.alimi@yale.edu>
- o Zoran Despotovic <despotovic@docomolab-euro.com>
- o Jason Livingood <Jason\_Livingood@cable.comcast.com>
- o Saverio Niccolini <saverio.niccolini@nw.neclab.eu>
- o Jan Seedorf <jan.seedorf@nw.neclab.eu>

The authors would like to thank the members of the P2PI and ALTO mailing lists for their feedback.

## Appendix B. Acknowledgments

The initial version of this document was co-authored by Laird Popkin.

The authors would like to thank

- o Vijay K. Gurbani <vkg@alcatel-lucent.com>
- o Enrico Marocco <enrico.marocco@telecomitalia.it>

for fostering discussions that lead to the creation of this document, and for giving valuable comments on it.

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.

Saverio Niccolini, Jan Seedorf, and Martin Stiernerling are partially supported by the NAPA-WINE project (Network-Aware P2P-TV Application over Wise Networks, <http://www.napa-wine.org>), a research project supported by the European Commission under its 7th Framework Program (contract no. 214412). 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 NAPA-WINE project or the European Commission.

Authors' Addresses

Sebastian Kiesel (editor)  
University of Stuttgart Computing Center  
Allmandring 30  
Stuttgart 70550  
Germany

Email: [ietf-alto@skiesel.de](mailto:ietf-alto@skiesel.de)  
URI: <http://www.rus.uni-stuttgart.de/nks/>

Stefano Previdi  
Cisco Systems, Inc.

Email: [sprevidi@cisco.com](mailto:sprevidi@cisco.com)

Martin Stiernerling  
NEC Laboratories Europe/University of Goettingen

Email: [martin.stiernerling@neclab.eu](mailto:martin.stiernerling@neclab.eu)  
URI: <http://ietf.stiernerling.org>

Richard Woundy  
Comcast Corporation

Email: [Richard\\_Woundy@cable.comcast.com](mailto:Richard_Woundy@cable.comcast.com)

Yang Richard Yang  
Yale University

Email: [yry@cs.yale.edu](mailto:yry@cs.yale.edu)



ALTO  
Internet-Draft  
Intended status: Standards Track  
Expires: April 28, 2011

S. Kiesel  
University of Stuttgart  
M. Tomsu  
Alcatel-Lucent  
N. Schwan  
M. Scharf  
Alcatel-Lucent Bell Labs  
M. Stiemerling  
NEC Europe Ltd.  
October 25, 2010

ALTO Server Discovery Protocol  
draft-kiesel-alto-3pdisc-04

Abstract

The goal of Application-Layer Traffic Optimization (ALTO) is to provide guidance to applications, which have to select one or several hosts from a set of candidates that are able to provide a desired resource.

Entities seeking guidance need to discover and possibly select an ALTO server to ask. This is called ALTO server discovery. This memo describes an ALTO server discovery mechanism based on several alternative mechanisms that are applicable in a diverse set of ALTO deployments.

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 April 28, 2011.

Copyright Notice

Copyright (c) 2010 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
1.1. Requirements . . . . .	3
1.2. Pre-Conditions . . . . .	4
2. Protocol Overview . . . . .	5
3. Retrieving the URI by DHCP . . . . .	7
3.1. ALTO Server Domain Name Encoding . . . . .	7
3.2. ALTO Server DHCPv4 Option . . . . .	7
3.3. ALTO Server DHCPv6 Option . . . . .	8
4. Retrieving the URI by U-NAPTR . . . . .	10
4.1. U-NAPTR Resolution . . . . .	10
4.2. Retrieving the Domain Name . . . . .	10
4.2.1. Option 1: User input . . . . .	11
4.2.2. Option 2: DHCP . . . . .	12
4.2.3. Option 3: Reverse DNS Lookup . . . . .	12
5. Applicability . . . . .	13
5.1. Applicability for Resource Consumer Server Discovery . . . . .	13
5.2. Applicability for Third Party Server Discovery . . . . .	13
6. IANA Considerations . . . . .	15
7. Security Considerations . . . . .	16
7.1. General . . . . .	16
7.2. For U-NAPTR . . . . .	16
8. Open Issues . . . . .	18
9. Conclusion . . . . .	19
10. References . . . . .	20
10.1. Normative References . . . . .	20
10.2. Informative References . . . . .	20
Appendix A. Acknowledgments . . . . .	22
Authors' Addresses . . . . .	23



## 1. Introduction

The goal of Application-Layer Traffic Optimization (ALTO) is to provide guidance to applications, which have to select one or several hosts from a set of candidates, that are able to provide a desired resource [RFC5693]. The requirements for ALTO are itemized in [I-D.ietf-alto-reqs]. ALTO is realized by a client-server protocol. ALTO clients send queries to ALTO servers, in order to solicit guidance.

ALTO clients have to discover suitable ALTO servers. Therefore the output of the herein defined ALTO discovery procedure tells the ALTO client which ALTO servers to send the queries to. The ALTO discovery procedure, as part of the the ALTO client, can be embedded in the resource consumer, which will eventually access the desired resource. As an alternative, they can be embedded in a resource directory, which assists resource consumers in finding appropriate resource providers. In some specific peer-to-peer application protocols these resource directories are called "trackers". Finally the ALTO server discovery procedure can be embedded in the resource consumer, whereas the ALTO client is embedded in the resource directory. ALTO queries, which are issued by a resource directory on behalf of a resource consumer, are referred to as third-party ALTO queries. The various possibilities to place ALTO servers and the placement of ALTO clients is discussed in [I-D.stiemerling-alto-deployments]. [I-D.song-alto-server-discovery] compares different protocol options and identifies DHCP and DNS as two approaches for the ALTO server discovery without detailing on the exact solution.

No matter where ALTO server and client are located, clients have to first find out if there is an ALTO server deployed that is in charge for them, and second they have to get the contact information of that server, i.e., the IP address, port number, and probably transport protocol (which defaults to TCP for [I-D.ietf-alto-protocol]).

The goal of this memo is to propose a uniform mechanism for all types of ALTO client deployments that is implementable and deployable at a fast pace, i.e., without creating other deployment dependencies for ALTO. We propose to use a combination of DHCP and DNS to retrieve the URL of the responsible ALTO server.

Comments and discussions about this memo should be directed to the ALTO working group: [alto@ietf.org](mailto:alto@ietf.org).

### 1.1. Requirements

There is other related works on server discovery, for instance GEOPRIV has rather strong security requirements (for good reasons),

which are documented in [I-D.ietf-geopriv-lis-discovery]. However, these requirements do not apply for the ALTO server discovery, as ALTO as such has very different requirements (see [I-D.ietf-alto-reqs]).

The result of the guidance provided to the application via the ALTO protocol is input to improve the initial peer selection process for peer-to-peer applications, or any other application applicable. A missing ALTO server, i.e., no result returned as part of the ALTO server discovery procedure, does not prevent the application to operate. A wrong or forged guidance from the ALTO server may only impact the overall operational result of the peer-to-peer system for a limited time, as these systems fine-tune their behavior depending on the experience network behavior.

This means that a wrong, missing, or forged ALTO guidance will not cause damage to the application or peer-to-peer system. This is in sharp contrast to the GEOPRIV use case, where a failure may have severe impact, including loss of human life. This is not the case for ALTO, as it is intended to be used today and as it is explored right now from the networking community.

## 1.2. Pre-Conditions

The whole document assumes certain pre-conditions, such as:

- o The ALTO server discovery procedure is executed on a per IP address base. Multiple IP addresses per interface or multiple IP addresses assigned to different IP interfaces require to repeat the procedure for every IP address. It may be fine to group IP addresses according their domain suffixes and to perform the procedure for such a group. However, this is out of scope of this document.
- o The ALTO server discovery procedure is executed on a per IP family base, i.e., separate for IPv4 and IPv6. It is up to the ALTO client to decide which of the possible multiple results of different IP address families to use. The choice of whether to use IPv4 or IPv6 is out of scope of this document.
- o A change of the IP address at an interface invalidates the result of the ALTO server discovery procedure. For instance, if the IP address assigned to a mobile host changes due to host mobility, it is required to run the ALTO server discovery procedure for the new IP address without relying on earlier gained information.

## 2. Protocol Overview

We define multiple alternatives to discover the IP address of the ALTO server, as there are a number of ways possible how such information can be provided to the ALTO client. The choice of method is up to the local network deployment. For instance, there can be deployments where the ALTO server in charge for ALTO client is provisioned by the network operator and communicated to the ALTO client's host via a DHCP option, while in other deployments no such means may exist.

The following figure illustrates the different protocols that are used to find the URI of a suitable ALTO server.

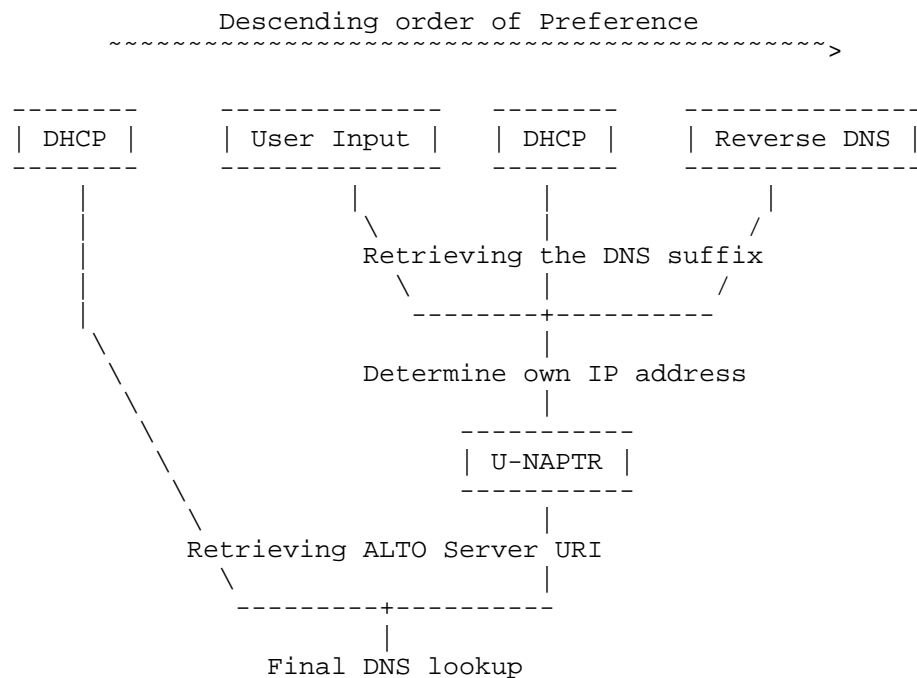


Figure 1: Protocol Overview

One option to retrieve the URI directly from the access network provider is DHCP. However for DHCP there are problems with residential gateways or broadband routers with NAT. If the network operator gives information about ALTO serves to the residential gateway via DHCP, the residential gateway would have to forward this information to the hosts with the (P2P) applications within the local network. This is not supported by already deployed residential gateways. Also DHCP poorly supports third-party ALTO server

discovery, i.e., in scenarios where the ALTO client is co-located with a resource directory ("tracker"), which is located in a different administrative domain than the client which will eventually access the resource.

Thus in deployment scenarios where DHCP is not possible, we specify a U-NAPTR based resolution process as a second option to retrieve the URL. As a precondition for resolution the U-NAPTR process needs the right domain name as input. This domain name is determined by the IP address of the client and the DNS suffix of the access network where the client is registered in. In order to retrieve the DNS suffix we specify three options:

User input: a user may manually specify the DNS suffix on its own, either to access a 3rd party ALTO service provider or as it does know such information.

DHCP: a network provider provides the DNS suffix through a DHCP option.

Reverse DNS: the DNS system can be used to retrieve the DNS suffix through reverse lookup of an FQDN associated with an IP address. This is the last resort if all other options failed.

### 3. Retrieving the URI by DHCP

One way of directly configuring the ALTO server URI for an access network provider is the DHCP protocol. The ALTO server URI consists of a domain name and the protocol the client should use to contact the server. While the domain name can vary and is configured by DHCP, the protocol is always HTTP.

For example a client may retrieve the domain name `altoserver.example.com` by the DHCP option as described in the remaining section. The client uses this domain name to contact the ALTO server under

```
http://altoserver.example.com/
```

#### 3.1. ALTO Server Domain Name Encoding

This section describes the encoding of the domain name used in the DHCPv4 option shown in Section 3.2 and also used in the DHCPv6 option shown in Section 3.3.

The domain name is encoded according to Section 3.1 of [RFC1035] whereby each label is represented as a one-octet length field followed by that number of octets. Since every domain name ends with the null label of the root, a domain name is terminated by a length byte of zero. The high-order two bits of every length octet MUST be zero, and the remaining six bits of the length field limit the label to 63 octets or less. To simplify implementations, the total length of a domain name (i.e., label octets and label length octets) is restricted to 255 octets or less.

#### 3.2. ALTO Server DHCPv4 Option

The ALTO server DHCPv4 option carries a DNS ([RFC1035]) fully-qualified domain name (FQDN) to be used by the ALTO client to locate a ALTO server.

The DHCP option for this encoding has the following format:

Code	Len	ALTO Server Domain Name				
tba	n	s1	s2	s3	s4	s5   ...

Figure 2: ALTO FQDN DHCPv4 Option

The values `s1`, `s2`, `s3`, etc. represent the domain name labels in the domain name encoding. Note that the length field in the DHCPv4

option represents the length of the entire domain name encoding, whereas the length fields in the domain name encoding (see Section 3.1) is the length of a single domain name label.

Code: to be assigned by IANA

Len: Length of the 'ALTO Server Domain Name' field in octets; variable.

ALTO Server Domain Name: The domain name of the ALTO server for the client to use.

A DHCPv4 client MAY request a ALTO server domain name in a Parameter Request List option, as described in [RFC2131].

The encoding of the domain name is described in Section 3.1.

This option contains a single domain name and, as such, MUST contain precisely one root label.

### 3.3. ALTO Server DHCPv6 Option

This section specifies the DHCP option for IPv6 (DHCPv6) to carry the domain name of the ALTO server. It is similar formatted to the DHCPv4 option

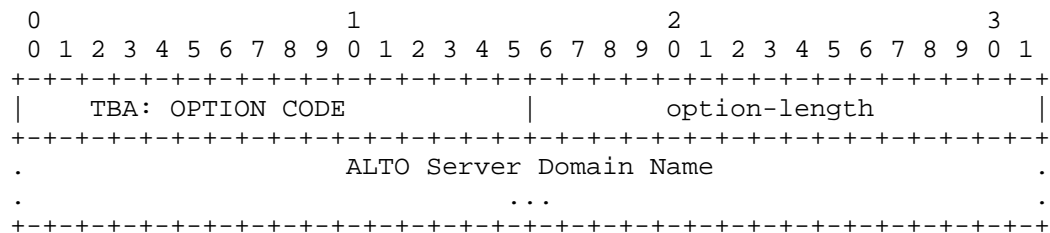


Figure 3: ALTO Server Domain Name DHCPv4 Option

option-code: to be assigned by IANA

option-length: The length of the 'ALTO Server Domain Name' field in octets; variable.

ALTO Server Domain Name: The domain name of the ALTO server for the client to use.

A DHCPv6 client MAY request a ALTO server domain name in an Options Request Option (ORO), as described in [RFC3315].

The encoding of the domain name is described in Section 3.1.

This option contains a single domain name and, as such, **MUST** contain precisely one root label.

#### 4. Retrieving the URI by U-NAPTR

As already described a direct DHCP configuration may not always be possible, for example due to deployment restrictions of the access network. Alternatively the ALTO server URI can be discovered by a U-NAPTR resolution process, as specified in this section.

The section is divided in two parts: Section 4.1 describes the U-NAPTR resolution process itself. As a precondition this process requires the domain name of the access network where the resource consumer is registered in. How the client identifies this DNS suffix is described in Section 4.2.

##### 4.1. U-NAPTR Resolution

ALTO servers are identified by U-NAPTR/DDDS (URI-Enabled NAPTR/Dynamic Delegation Discovery Service) [RFC4848] application unique strings, in the form of a DNS name. An example is 'altoserver.example.com'.

Clients need to use the U-NAPTR [RFC4848] specification described below to obtain a URI (indicating host and protocol) for the applicable ALTO service. In this document, only the HTTP and HTTPS URL schemes are defined. Note that the HTTP URL can be any valid HTTP URL, including those containing path elements.

The following two DNS entries show the U-NAPTR resolution for "example.com" to the HTTPS URL <https://altoserver.example.com/secure> or the HTTP URL <http://altoserver.example.com>, with the former being preferred.

```
example.com.
```

```
IN NAPTR 100 10 "u" "ALTO:https"  
"!.*!https://altoserver.example.com/secure!" ""
```

```
IN NAPTR 200 10 "u" "ALTO:http"  
"!.*!http://altoserver.example.com!" ""
```

##### 4.2. Retrieving the Domain Name

The U-NAPTR resolution process requires a domain name as input. The algorithm that is applied to determine this domain name is described in this section. We specify three different options. In option 1 the user manually configures a specific ALTO service instance that he wants to use. Option 2 defines a DHCP option to allow the network service provider a remote configuration of the client. In option 3



the client tries to get the domain name by performing a reverse DNS lookup on its IP address.

The resource consumer may have private IP addresses and public IP addresses and depending on the deployment it might be necessary to determine for all IP addresses the ALTO server in charge of. To determine its public IP address the resource consumer may need to use STUN[RFC5389] or BEP24[bep24]. For the following examples we assume that the IP address of the resource consumer is a.b.c.d.

#### 4.2.1. Option 1: User input

A user may want to use a third party ALTO service instance. Therefore we allow the user to specify a DNS suffix on its own, for example in a config file option. The DNS suffix given by the user is combined with the IP address of the resource consumer to allow the third party ALTO service to direct the client to a suitable ALTO server based on the location of the client. A possible DNS suffix entered by the user may be:

myaltoprovider.org

This DNS suffix is prepended with the IP address of the resource consumer in reverse order to compose the domain name used for the final U-NAPTR lookup Section 4.1. In case there are multiple ALTO servers deployed, the third party ALTO service instance can direct the ALTO client to the ALTO server closest to the client based on the IP address.

Multiple lookups with different domain names might be necessary to complete the U-NAPTR resolution process. If there is no response for a lookup the domain name is shortened by one part for the succeeding lookup, until a lookup is successful, as for example

d.c.b.a.myaltoprovider.org.

c.b.a.myaltoprovider.org.

b.a.myaltoprovider.org.

a.myaltoprovider.org.

myaltoprovider.org.

#### 4.2.2. Option 2: DHCP

As a second option network operators can configure the domain name to be used for service discovery within an access network. RFC 5986[RFC5986] defines DHCP IPv4 and IPv6 access network domain name options that identify a domain name that is suitable for service discovery within the access network. The ALTO server discovery procedure uses these DHCP options to retrieve the domain name as an input for the U-NAPTR resolution. One example could be:

example.com

#### 4.2.3. Option 3: Reverse DNS Lookup

The last option to get the domain name is to use a DNS PTR query for the IP address of the resource consumer. The local DNS server resolves the IP address to the FQDN that also contains the DNS suffix for the respective IP address. A possible answer for a PTR lookup for d.c.b.a.in-addr.apra might be, for example:

d-c-b-a.dsl.westcoast.myisp.net

This domain name can be used for the final U-NAPTR lookup Section 4.1. Again, if there is no response to the lookup the domain name is shortened by one part for the succeeding lookup. The domain names used for the example as described above are:

d-c-b-a.dsl.westcoast.myisp.net.

dsl.westcoast.myisp.net.

westcoast.myisp.net.

myisp.net.

## 5. Applicability

This section discusses the applicability of the proposed solution with respect to the resource consumer server discovery and the third party deployment scenarios. Each section discusses the proposed steps that are needed to determine the ALTO Server URI.

### 5.1. Applicability for Resource Consumer Server Discovery

In this scenario the ALTO server discovery procedure is performed by the resource consumer, for example a peer in a P2P system. After the discovery the peer does the ALTO query on its own, or it might share the ALTO server contact information with a third party, for example a tracker, which then does the ALTO query on behalf of the peer.

The access network provider has two options based on DHCP to remotely configure the ALTO client to use its ALTO server. The first option is to provide the ALTO server URI directly by a DHCP option as described in Section 3, the second option is to provide the access network domain name as described in Section 4.2.2. It is up to the access network provider to choose one of both options.

To complete the ALTO server discovery process the resource consumer first SHOULD try to retrieve the ALTO server URI by the DHCP option as described in Section 3. In case this is successful the discovery process is finished, in case it fails, either as the access network provider has not configured the specified option or through deployment restrictions, the resource consumer SHOULD subsequently check whether the user has provided the domain name through manual configuration. If this is also not the case the next step SHOULD be to check for the access network domain name DHCP option (Section 4.2.2). Finally the client SHOULD try to retrieve the domain name by the last option, the DNS reverse lookup on its IP address as described in Section 4.2.3.

In case the ALTO discovery client has determined the domain name through one of the described options it proceeds with the U-NAPTR lookup as described in Section 4.1.

If the ALTO server URI could not be retrieved either through direct configuration by the access network provider through DHCP nor through the U-NAPTR lookup the discovery process fails.

### 5.2. Applicability for Third Party Server Discovery

In case of the third party server discovery deployment scenario the entity performing the ALTO server discovery process is different from the resource consumer. Typically the resource consumer is a peer

whereas the ALTO client is a resource directory which seeks for ALTO guidance on behalf of the peer. Another use case for the third party discovery is an application that looks for ALTO guidance transparently for the resource consumer, for example a CDN.

Here the ALTO server discovery process can also retrieve guidance through one of the DHCP options or manual user configuration, but only if the provided discovery information is forwarded by the resource consumer to the third party entity. In this case, additional mechanisms for the forwarding of this discovery information need to be specified. However these mechanisms are out of scope of this document.

If the third party entity cannot obtain this discovery information, the ALTO server discovery process relies on retrieving the domain name used as input to the U-NAPTR lookup through reverse DNS lookup of the IP address of the resource consumer as described in Section 4.2.3. Usually the third party entity already knows the IP address of the resource consumer which was used to establish the initial connection. In general this IP address is a public address, either of the resource consumer or of the last NAT on the path to the ALTO client. This makes the IP address a good candidate for the DNS PTR query. Thus, we expect that the DNS query will be successfully resolved to the FQDN of the domain where the resource consumer is registered in.

In case the resource consumer needs guidance for a different IP address, for example one from a private network, we recommend that the resource consumer discovers the server itself and forwards the ALTO server contact information directly to the third party entity, which in turn can then do the third party ALTO query. Again, forwarding the contact information from the resource consumer to the third party entity is out of scope of this document.

## 6. IANA Considerations

This document registers the following U-NAPTR application service tag:

Application Service Tag: ALTO

Defining Publication: The specification contained within this document.

This document registers the following U-NAPTR application protocol tags:

- o Application Protocol Tag: http

Defining Publication: RFC 2616 [RFC2616]

- o Application Protocol Tag: https

Defining Publication: RFC 2818 [RFC2818]

## 7. Security Considerations

### 7.1. General

This is still to be done in later revision of this draft, as the draft evolves heavily right now.

### 7.2. For U-NAPTR

The address of an ALTO server is usually well-known within an access network; therefore, interception of messages does not introduce any specific concerns.

The primary attack against the methods described in this document is one that would lead to impersonation of a ALTO server since a device does not necessarily have a prior relationship with a ALTO server.

An attacker could attempt to compromise ALTO discovery at any of three stages:

1. providing a falsified domain name to be used as input to U-NAPTR
2. altering the DNS records used in U-NAPTR resolution
3. impersonation of the ALTO

This document focuses on the U-NAPTR resolution process and hence this section discusses the security considerations related to the DNS handling. The security aspects of obtaining the domain name that is used for input to the U-NAPTR process is described in respective documents, such as [I-D.ietf-geopriv-lis-discovery].

The domain name that is used to authenticated the ALTO server is the domain name in the URI that is the result of the U-NAPTR resolution. Therefore, if an attacker were able to modify or spoof any of the DNS records used in the DDDS resolution, this URI could be replaced by an invalid URI. The application of DNS security (DNSSEC) [RFC4033] provides a means to limit attacks that rely on modification of the DNS records used in U-NAPTR resolution. Security considerations specific to U-NAPTR are described in more detail in [RFC4848].

An "https:" URI is authenticated using the method described in Section 3.1 of [RFC2818]. The domain name used for this authentication is the domain name in the URI resulting from U-NAPTR resolution, not the input domain name as in [RFC3958]. Using the domain name in the URI is more compatible with existing HTTP client software, which authenticate servers based on the domain name in the URI.

An ALTO server that is identified by an "http:" URI cannot be authenticated. If an "http:" URI is the product of the ALTO discovery, this leaves devices vulnerable to several attacks. Lower layer protections, such as layer 2 traffic separation might be used to provide some guarantees.

## 8. Open Issues

Here are a few open issues to be clarified:

Handling of reverse DNS lookups for IPv6: Refer to [RFC4472] for a discussion about the issues.

Missing reverse DNS entries for an IP address: There may be cases where the reverse DNS lookup does not yield any result. However, this will leave the ALTO client with no choice, other than giving up. This needs better documentation.

How to handled multiple results: For instance, a host behind a NAT that yields an ALTO server in the private IP address domain and one in the public IP address domain. Whom to ask?

Suffix Issues Document issues with suffix information provided by DHCP or by other means. For instance, a host behind a NAT may have a configured DNS suffix ".local". This suffix is not usable for the server discovery procedure.



## 9. Conclusion

This document describes a general ALTO server discovery process and discusses how the process can be applied in different deployment scenarios, including the resource consumer discovery as well as the third party discovery.

## 10. References

### 10.1. Normative 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.
- [RFC2818] Rescorla, E., "HTTP Over TLS", RFC 2818, May 2000.
- [RFC3958] Daigle, L. and A. Newton, "Domain-Based Application Service Location Using SRV RRs and the Dynamic Delegation Discovery Service (DDDS)", RFC 3958, January 2005.
- [RFC4033] Arends, R., Austein, R., Larson, M., Massey, D., and S. Rose, "DNS Security Introduction and Requirements", RFC 4033, March 2005.
- [RFC5389] Rosenberg, J., Mahy, R., Matthews, P., and D. Wing, "Session Traversal Utilities for NAT (STUN)", RFC 5389, October 2008.

### 10.2. Informative References

- [I-D.ietf-alto-protocol]  
Alimi, R., Penno, R., and Y. Yang, "ALTO Protocol", draft-ietf-alto-protocol-05 (work in progress), July 2010.
- [I-D.ietf-alto-reqs]  
Kiesel, S., Previdi, S., Stiernerling, M., Woundy, R., and Y. Yang, "Application-Layer Traffic Optimization (ALTO) Requirements", draft-ietf-alto-reqs-06 (work in progress), October 2010.
- [I-D.ietf-geopriv-lis-discovery]  
Thomson, M. and J. Winterbottom, "Discovering the Local Location Information Server (LIS)", draft-ietf-geopriv-lis-discovery-15 (work in progress), March 2010.
- [I-D.song-alto-server-discovery]  
Yongchao, S., Tomsu, M., Garcia, G., Wang, Y., and V. Avila, "ALTO Service Discovery", draft-song-alto-server-discovery-03 (work in progress), July 2010.
- [I-D.stiernerling-alto-deployments]  
Stiernerling, M. and S. Kiesel, "ALTO Deployment

Considerations", draft-stiemerling-alto-deployments-05 (work in progress), October 2010.

- [RFC1035] Mockapetris, P., "Domain names - implementation and specification", STD 13, RFC 1035, November 1987.
- [RFC2131] Droms, R., "Dynamic Host Configuration Protocol", RFC 2131, March 1997.
- [RFC3315] Droms, R., Bound, J., Volz, B., Lemon, T., Perkins, C., and M. Carney, "Dynamic Host Configuration Protocol for IPv6 (DHCPv6)", RFC 3315, July 2003.
- [RFC4472] Durand, A., Ihren, J., and P. Savola, "Operational Considerations and Issues with IPv6 DNS", RFC 4472, April 2006.
- [RFC4848] Daigle, L., "Domain-Based Application Service Location Using URIs and the Dynamic Delegation Discovery Service (DDDS)", RFC 4848, April 2007.
- [RFC5693] Seedorf, J. and E. Burger, "Application-Layer Traffic Optimization (ALTO) Problem Statement", RFC 5693, October 2009.
- [RFC5986] Thomson, M. and J. Winterbottom, "Discovering the Local Location Information Server (LIS)", RFC 5986, September 2010.
- [bep24] Harrison, D., "Tracker Returns External IP", BEP [http://bittorrent.org/beps/bep\\_0024.html](http://bittorrent.org/beps/bep_0024.html).

## Appendix A. Acknowledgments

The authors would like to thank Haibin Song, Richard Alimi, and Roni Even for fruitful discussions during the 75th IETF meeting.

Hannes Tschofenig provided the initial input to the U-NAPTR solution part. Hannes and Martin Thomson provided excellent feedback and input to the server discovery.

Marco Tomsu and Nico Schwan are partially supported by the ENVISION project (<http://www.envision-project.org>), a research project supported by the European Commission under its 7th Framework Program (contract no. 248565). 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 ENVISION project or the European Commission.

Michael Scharf is supported by the German-Lab project (<http://www.german-lab.de>) funded by the German Federal Ministry of Education and Research (BMBF).

Martin Stiernerling 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  
University of Stuttgart Computing Center  
Allmandring 30  
Stuttgart 70550  
Germany

Email: [ietf-alto@skiesel.de](mailto:ietf-alto@skiesel.de)  
URI: <http://www.rus.uni-stuttgart.de/nks/>

Marco Tomsu  
Alcatel-Lucent  
Lorenzstrasse 10  
Stuttgart 70435  
Germany

Email: [marco.tomsu@alcatel-lucent.com](mailto:marco.tomsu@alcatel-lucent.com)  
URI: [www.alcatel-lucent.com/bell-labs](http://www.alcatel-lucent.com/bell-labs)

Nico Schwan  
Alcatel-Lucent Bell Labs  
Lorenzstrasse 10  
Stuttgart 70435  
Germany

Email: [nico.schwan@alcatel-lucent.com](mailto:nico.schwan@alcatel-lucent.com)  
URI: [www.alcatel-lucent.com/bell-labs](http://www.alcatel-lucent.com/bell-labs)

Michael Scharf  
Alcatel-Lucent Bell Labs  
Lorenzstrasse 10  
Stuttgart 70435  
Germany

Email: [michael.scharf@alcatel-lucent.com](mailto:michael.scharf@alcatel-lucent.com)  
URI: [www.alcatel-lucent.com/bell-labs](http://www.alcatel-lucent.com/bell-labs)

Martin Stiemerling  
NEC Laboratories Europe/University of Goettingen  
Kurfuerstenanlage 36  
Heidelberg 69115  
Germany

Phone: +49 6221 4342 113  
Email: martin.stiemerling@neclab.eu  
URI: <http://ietf.stiemerling.org>



ALTO  
Internet-Draft  
Intended status: Informational  
Expires: April 22, 2011

Kai.Lee  
China Telecom  
GuangYao.Jian  
Xunlei network  
October 22, 2010

ALTO and DECADE service trial within China Telecom  
draft-lee-alto-chinatelecom-trial-01.txt

#### Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at  
<http://www.ietf.org/ietf/lid-abstracts.txt>.

The list of Internet-Draft Shadow Directories can be accessed at  
<http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on April 22, 2011.



#### Copyright Notice

Copyright (c) 2010 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 BSD License.

#### Abstract

This document reports the experience of China Telecom in a recent experiment with the ALTO service and P2P caches deployment. It is found that the deployment of the ALTO service significantly improves the capability of a Service Provider to affect the distribution of P2P traffic. It is also found that a traffic localized ALTO policy may decrease the download speed of a P2P user. However, the deployment of some P2P caches can compensate such influence.

## Table of Contents

1. Introduction .....	3
2. High level description of the trial .....	4
2.1. Difference between standard ALTO protocol .....	4
2.2. Difference with Comcast's trial .....	5
3. Trial results .....	6
3.1. ALTO server policy test.....	7
3.2. P2P cache test .....	8
4. Methods of data collection.....	9
5. Configurations and algorithms in trial .....	10
5.1. Configuration of PID MAP.....	10
5.2. Algorithms of Xunlei using ALTO information .....	10
5.3. Configuration of cache system.....	12
6. Next steps .....	13
7. Security Considerations.....	14
8. IANA Considerations .....	14
9. References .....	14
Author's Addresses .....	14

## 1. Introduction

Although another trial on P4P, the predecessor of the ALTO, is available by Comcast, the impact of ALTO on a large scale real network has never publicly reported. Such real network should post no limitation on either the number of contents or the number of users. This draft reports the experience of China Telecom in a recent experiment with the deployment of the ALTO service and P2P caches.

With over 60 million fixed-line broadband subscribers, China Telecom is the largest broadband service provider in China. It has one IP backbone network that cover all of the 31 provinces and about 200 MAN networks managed by the provinces respectively. This trial was taken place in one province with 7 million broadband subscribers and about 11 MAN networks.

Xunlei, the cooperater of this trial, is a leading P2P service provider in China. Xunlei supports both file downloads and real time media streaming. In 2009, when was this trail occurring, it serves over 20 million users each day.

This trial is a joint effort of China Telecom and Xunlei. During this trial, China Telecom provided the following devices: an alto server for distribute ALTO information, some P2P caches to test its influence on traffic localization and user experience. China Telecom

also monitored the traffic load within its backbone. Xunlei provided the P2P client and users. To support this trial, Xunlei modified its platform to support ALTO, and recorded operational information on its platform according to the requirement of China Telecom. Note that the client of Xunlei was not changed.

## 2. High level description of the trial

### 2.1. Difference between standard ALTO protocol

Note that ALTO protocol is still on progressing, in this trail, some modifications were made to the ALTO.

First, a notification mechanism for the ALTO server is introduced. With this mechanism, the ALTO server notifies its clients the changes of network maps and cost maps. Thus, ALTO clients can respond fast to the change of traffic optimizing policy.

One problem that this trail met is to find the effect of ALTO&Cache deployment. The traffic within the IP backbone is highly periodical. For example, the traffic on each weekend is higher than the workday.

As such, data should be collected in the same workday in different week. This can facilitate the comparison of the effects on p2p traffic under different ALTO configuration and different policy, and to evaluate the effect of ALTO service

In this trail, ALTO clients were just embedded in the trackers of Xunlei, not in the Xunlei clients. The reason for this is mainly for deployment consideration. There are hundreds of millions of Xunlei clients in use, To update these clients as the ALTO client in a short time is not feasible. However, according to the analysis of Xunlei, although both tracker based and tracker-less technology are adopted, the traffic does not controlled by the trackers is less than 15% of its total traffic. Based on this analysis, in this trial, Xunlei clients are not involved in the ALTO service which has negligible influence on the final evaluation of this trial. Such design can also reduce the load on the ALTO server.

Secondly, only map service is provided in this trial. Other services of ALTO service were not deployed, as they are not essential for this trial.

## 2.2. Difference with Comcast's trial

Comcast has a trial with limited swarms, with the cooperation of Pando. According to (ref), there are five swarms, and overall 57,000 peers are involved in that trial.

There are several differences between our trial and Comcast's trial:

1. The scope of the trail: This trial covers the whole province with over 700 million broadband users. It lasted for over 4 months. There are countless swarms with all kinds of contents. Thus, this trial is more realistic than the previous trial from Comcast.
2. The usage of P2P cache: This trail differs from the previous trail by the utilization of P2P cache. In this trail, the average download speed of a Xunlei client decreases with the increase of the level of traffic localization. Thus the usage of P2P cache was introduced to compensate the decrease of download speed.
3. The evaluation method: In contrast to that all test data was collected by Pando client in Comcast's trial, we collect test data from two ways. Besides the data from Xunlei P2P client, we simultaneously collect the data from network operator's NMS system.(such as data from SNMP reports and DPI(deep package inspection) device deployed on backbone). We can do this because Xunlei's p2p traffic occupy 20% of backbone traffic flow. This traffic flow will all be affected by our alto policy and it is big enough to be observed by network operator's NMS system.
4. The implementation of ALTO: In this trial, only the P2P trackers are ALTO clients, but not those Xunlei clients. There are some reasons to do this:
  - a) To avoid the update all Xunlei clients and simplify the deployment of trial.
  - b) To lessen the alto server load.
  - c) Above 85% of Xunlei traffic flow is controlled by Xunlei tracker, the traffic flow from DHT mechanism is less than 15%. An alto server dedicated for Xunlei tracker can control majority of Xunlei traffic flow.

### 3. Trial results

This trial used all Xunlei p2p client in the province and all contents that are requested or served by Xunlei P2P client in the province. The trial environment is more realistic than comcast's. A primary objective of this trial is to measure the effects of traffic localization and change of users download speed in comparison to normal p2p activity.

The test process is divided into two parts: first part is just applied the ALTO server to measure the effects of traffic localization and change of P2P user experience. The second part is to introduce the P2P cache to the trial, to measure the improvement of user download speed, the bandwidth consumption and their relationship with the scale of cache and.

Our trial starts at 2009.6.12 and ends at 2009.10.18, lasting nearly four months. We do this trial by applying different ALTO policy to Xunlei tracker. There are two kinds of ALTO policy: One is optimized policy and the other is normal policy. The optimized policy will try to localize the traffic as much as possible by utilizing the information from ALTO server. The normal policy will just use the original Xunlei peer selection and traffic control rules and no alto policy are involved. We usually change the alto policy in midnight of a day and send a notification to Xunlei tracker with notification mechanism. (<http://tools.ietf.org/id/draft-sun-alto-notification-02.txt>)

Before we do the trial , we collect the information about Xunlei's peer and traffic distribution

No	Data Item	Description	The way of collection
1	Peer distribution	24.6% is within the province, 75.4% is out of the province	Random sampling by Xunlei tracker 24 times one day
2	Traffic distribution	76.9% is intra-province traffic 23.1% is inter-province traffic	Random selecting peers to report their traffic flow

### 3.1. ALTO server policy test

After we applied the alto optimized policy about 60% inter-province traffic has became The intra-province traffic. Below is the result that we observed on china telecom's network NMS system:

No	Data Item	Description	The way of collection
1	Outbound bandwidth	Decreased 42.77Gbps, about 50.61% of total Xunlei outbound traffic	Collecting max average outbound traffic of a day from the DPI system
2	Inbound/outbound bandwidth	outbound bandwidth decreased 31.58Gbps inbound bandwidth decreased 10.46Gbps	Collecting max average inbound/outbound traffic of a day from the snmp system

User's average download speed will decreased if traffic localization policy is applied

### 3.2. P2P cache test

In this trial we deployed 16 cache devices, each with 1.8TB SAS hard disks. The P2P cache system has 15Gbps links connected to the internet. We cached the content according to its popularity.

No	Data Item	Description	The way of collection
1	Outbound bandwidth	Decreased 40Gbps, about 54.47% of total Xunlei outbound traffic	Collecting max average outbound traffic of a day from the DPI system
2	Inbound/outbound bandwidth	outbound bandwidth decreased 39.18Gbps inbound bandwidth decreased 28.3 Gbps	Collecting max average inbound/outbound traffic of a day from the snmp system
3	Average download speed	From 279KBps up to 294.5KBps	Collection from Xunlei OAM system

The P2P cache system occupancy ratio is about 80%. Bandwidth consumed is about 4-5Gbps.

After deployed the P2P cache system, the traffic flow in the the province has decreased a lot. Meanwhile the average download speed of Xunlei client has been increased.

#### 4. Methods of data collection

In this trial we have two ways for information collection; one is to collect from p2p service provider such as Pando and Xunlei just like comcast's trial. The other is to collect from ISP's network OAM system. Because the Xunlei's inter-province traffic flow is about 80Gbps that is large enough to be observed by ISP's network OAM system

1. Information from ISP's network OAM system and DPI system include



- a) Inbound/outbound traffic flow statistic
- b) Xunlei traffic flow detected by DPI system. The DPI system just monitored the uplink of the province to China telecom's backbone.

## 2. Information from Xunlei

- a) Inter-province/intra-province traffic flow.
- b) User average download speed.

## 5. Configurations and algorithms in trial

### 5.1. Configuration of PID MAP

- a) PID Map: We define 11 PIDs
  - PID1-PID11 represent the 11 MANs of the trial network
  - PID12 represents rest of the Internet
- b) Cost Map:
  - Bidirectional cost between any PIDs from PID1 to PID11 has the same value 1
  - Bidirectional cost between PID12 and PIDi ( $1 \leq i \leq 11$ ) has the same value 2

### 5.2. Algorithms of Xunlei using ALTO information

Xunlei is a hybrid application utilizing both trackers and DHT, About 85% of Xunlei traffic controlled by Xunlei trackers. In this trail ALTO clients just include the xunlei trackers not include the xunlei client. Just the traffic controlled by xunlei tracker has been affected.

Before the trial Xunlei tracker peer selection algorithm is:

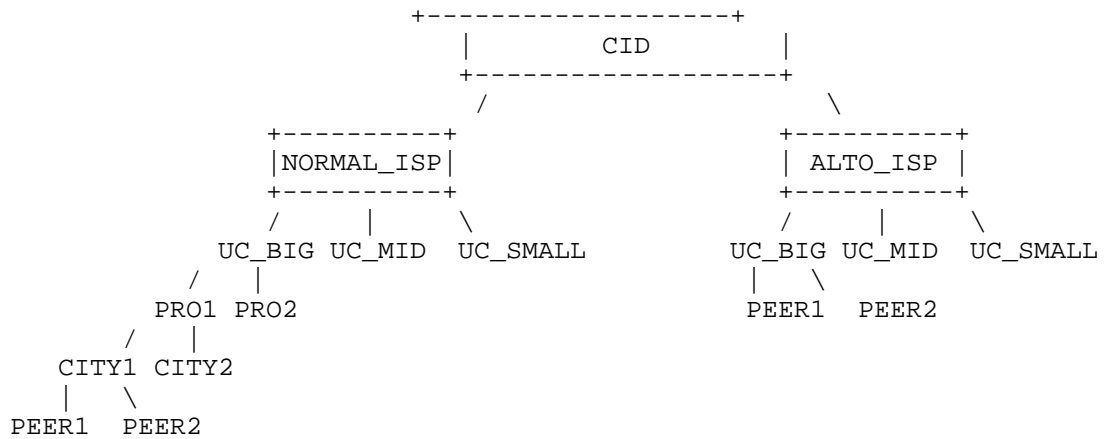
Xunlei Peer selection algorithm depends on two properties: ISP ID and UC (upload capability), the peer selection priority is :

Same ISP ID > different ISP ID

Higher UC > lower UC

The peers with same ISP ID with the requesting peer have higher priority than those with different ISP ID. If peers have same ISP ID then the peers with higher UC have higher priority than those with lower UC.

After applying the ALTO information into the xunlei peer selection algorithm. Xunlei changed his Peers select mechanism. All xunlei peers are organized in a tree structure which is indexed by CID(content ID), in the second level ALTO\_ISP and normal\_ISP represent the network of ISP with and without alto information. In this trial 11 MANs in trial province became 11 ALTO\_ISPs. The third level is defined by different upload capability(UC) of peers. The fourth level of normal\_ISP branch is the different provinces(PRO1,PRO2) of ISP, the fifth level of the normal\_ISP is different city of ISP.



The algorithms of cost between origination peer(peer\_o) and destination peer(peer\_d) is :

If (peer\_o and peer\_d both from ALTO\_ISP)

If (peer\_o and peer\_d in the same ALTO\_ISP) then cost = 0;

```
Else cost = 100000;

Else if (peer_o from ALTO_ISP and peer_d from normal_ISP) cost =
100000;

Else if (peer_o from normal_ISP and peer_d from ALTO_ISP) cost =
1000;

Else if (peer_o and peer_d both from normal_ISP){

    If (peer_o and peer_d from different normal_ISP) cost =1000;

    Else if (peer_o and peer_d from different province) cost = 100;

    Else if (peer_o and peer_d from different city) cost = 10;

    Else cost =0;

}
```

The peer select mechanism is lower cost peers will have higher priority

The updated peer selection mechanism is not the best mechanism. For example a peer in MAN2 is supposed to be better choice than the peers which not located in china telecom's network when a peer in MAN1 send a content request to tracker. But this mechanism will select the peer out of china telecom's network first then select the peer in the MAN2. Before we defined the network map with 12 PIDs. We first defined a network map with just 2 PIDs. PID1 represent the trial province and PID2 represent the other network to test the backbone traffic saving effect of ALTO service. The test result show that the network map with 12PIDs has almost same backbone traffic saving effect compared to the network map with 2 PIDs. So in the trial we deployed this mechanism.

The other change is the number of returned peers from xunlei tracker . If a listing request is from the trial province, the maximum # of returned peers from xunlei tracker is set to 120, not the normal case of 500.

### 5.3. Configuration of cache system

Before we deploy the cache system we have made some statistics about relationship of content popularity and network traffic caused by content with different popularity in trial province.

content	total	total	proportion
popularity	size(GB)	traffic	of total
		(Gbps)	traffic(%)
top 10	18.9	1.34	9.3
top 20	29.3	1.68	11.7
top 50	51.8	2.28	15.9
top 100	93.6	2.89	20.1
top 500	418.7	4.74	33
top 1000	812.4	5.88	40.9
top 2000	1518.6	7.16	49.8
top 5000	3551	8.89	61.9

Our cache system has limited storage and access bandwidth so we need to know which content is most "valuable" to be cached. According the statistics from xunlei if a downloading task is fed over 100 peers , this task always can get the maximum download speed(this speed depends on the peer's access network, in the trial the average access speed of user is about 2Mbps). The top 2000 popular content almost all have over 100 seeds in trial province. That means the top 2000 popular contents don't need be cached. Our cache policy is just cache the content which's popularity rank behind 2000.

## 6. Next steps

The alto mechanism is very effective to optimize the traffic flow. But when the traffic is localized, the user average download speed

is slowed down simultaneously. If alto can cooperate with p2p cache or other service performance enhancement mechanism, it will be more practical.

The ALTO service's effect depends on the SP such as Xunlei, pando how to use it. The mechanism such as peer selection mechanism and content cache mechanism need to be studied.

## 7. Security Considerations

High-level security considerations can be found in the [draft-ietf-alto-problem-statement].

## 8. IANA Considerations

This document requests the registration of a new media type:  
"application/alto"

## 9. References

[RFC 5693]

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

[I-D.ietf-alto-reqs]

Kiesel, S., Popkin, L., Previdi, S., Woundy, R., and Y. Yang, "Application-Layer Traffic Optimization (ALTO) Requirements", draft-ietf-alto-reqs-01 (work in progress), July 2009.

[I-D.penno-alto-protocol]

Penno, R. and Y. Yang, "ALTO Protocol", draft-ietf-alto-protocol-01 (work in progress), July 2009.

## Author's Addresses

Kai Lee  
China Telecom Beijing Research Institute  
Email: leekai@ctbri.com.cn

GuangYao Jian  
Xunlei Network  
Email: jianguangyao@xunlei.com



Network Working Group  
Internet-Draft  
Intended status: Experimental  
Expires: April 28, 2011

R. Penno  
S. Raghunath  
J. Medved  
Juniper Networks  
R. Alimi  
Google  
R. Yang  
Yale University  
S. Previdi  
Cisco Systems  
October 25, 2010

ALTO and Content Delivery Networks  
draft-penno-alto-cdn-02

Abstract

Networking applications can request through the ALTO protocol information about the underlying network topology from the ISP or Content Provider (henceforth referred as Provider) point of view. In other words, information about what a Provider prefers in terms of traffic optimization -- and a way to distribute it. The ALTO Service provides information such as preferences of network resources with the goal of modifying network resource consumption patterns while maintaining or improving application performance.

One of the main use cases of the ALTO Service is its integration with Content Delivery Networks (CDN). The purpose of this draft is twofold: first, to describe how ALTO can be used in existing and new CDNs, both within an ISP and in separate organizational entities from the ISP; second, to collect requirements for ALTO usage in CDNs and to provide recommendations into the development of the ALTO protocol for better support of CDNs.

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 to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that

other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/lid-abstracts.txt>.

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on April 28, 2011.

#### Copyright Notice

Copyright (c) 2010 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 BSD License.



## Table of Contents

1. Introduction . . . . .	4
2. Scope . . . . .	4
3. Terminology . . . . .	4
4. Request Routing as an Integration Point of ALTO into CDN . . . . .	5
4.1. HTTP Redirect . . . . .	5
4.2. DNS Request Routing . . . . .	6
5. Basic Scheme of CDN/ALTO Integration . . . . .	6
5.1. Basic Integration Scheme . . . . .	6
5.1.1. ALTO for HTTP Redirect . . . . .	7
5.1.2. ALTO for DNS Resolution . . . . .	8
5.2. Multi-hop Redirection . . . . .	8
5.3. CDN Node Discovery and Status Notification . . . . .	8
5.3.1. CDN Node Status Updates received by Request Router . . . . .	9
5.3.2. CDN Node Status Updates received by ALTO . . . . .	10
6. Request Routing using ALTO Services . . . . .	10
6.1. Request Routing using the Map Service . . . . .	10
6.2. Request Routing using the Endpoint Cost Service . . . . .	11
7. Multiple Administrative Domains . . . . .	12
7.1. CDN nodes/Request Router in a separate administrative domain from that of ISP . . . . .	12
7.2. Managed DNS Domain with Three Administrative Domains . . . . .	15
7.2.1. Managed DNS Redirect to Local CDN . . . . .	15
7.2.2. Managed DNS with CDN-Provided Request Routing . . . . .	16
8. Protocol Recommendations . . . . .	17
8.1. Necessary Additions . . . . .	17
8.1.1. NA1: PID Attributes . . . . .	17
8.1.2. NA2: PID Attributes and Query . . . . .	17
8.2. Helpful Additions . . . . .	18
8.2.1. HA1: Push Mechanism . . . . .	18
8.2.2. HA2: Incremental Map Updates . . . . .	18
8.2.3. HA3: ALTO Border Router PID attribute . . . . .	18
8.2.4. HA4: CDN ALTO Server Discovery . . . . .	18
8.2.5. HA5: Extensible ALTO Cost Maps . . . . .	18
8.2.6. NA4: Federated Deployment of ALTO Servers . . . . .	19
9. IANA Considerations . . . . .	19
10. Security Considerations . . . . .	19
11. Acknowledgements . . . . .	19
12. References . . . . .	19
12.1. Normative References . . . . .	19
12.2. Informative References . . . . .	19
Authors' Addresses . . . . .	20

## 1. Introduction

Content Delivery Networks are becoming increasingly important in the Internet [ARBOR] and many CDNs today already use some form of proximity through geolocation. But in many cases the content provider/distributor and the Internet Service Provider are disjoint and even if content servers are co-located into the ISP's networks, there is no standardized way to share server location and/or network topology information. Therefore a natural step forward would be to use ALTO to share this information.

Another key aspect of ALTO in the context of CDNs deployments is that it is desirable that no changes to the hosts are needed (or that changes to hosts would be transparent to the user). In other words, a traditional web browser is all there is needed to take advantage of ALTO information. This is a significant difference from the P2P applications where a special client is typically needed and ALTO is normally used as a way to reduce operational expense.

## 2. Scope

This document discusses how Content Delivery Networks can benefit from ALTO through integration of the ALTO Service with the main request routing techniques. There are two objectives:

- o Present basic integration schemes of ALTO into CDNs.
- o Provide protocol recommendations to ALTO: Whenever a new requirement on protocol functionality is identified to achieve integration with CDNs, it will be enumerated with 'REQ-<N>'. Each requirement is documented in a section of its own in order to foster parallel discussions and possible adoption.

## 3. Terminology

**Content-aware Proximity Request Router:** The Request Router knows about locations and presence of content & media objects in the network. Therefore the redirection to a CDN node is made based on both the availability of content or content-type in that CDN node and the proximity of the CDN node to the requesting user.

**Service-aware Proximity Request Router:** The Request Router knows about locations of CDN nodes in the network and redirects user to the closest CDN node. A redirection is made irrespective of content presence in the CDN node; if content is not present, the node will be populated with the content while the content is

served to the user.

**HTTP Request Router:** a Content-aware or Service-aware Proximity Request Router for HTTP. It embeds an HTTP Server that performs HTTP Redirects, an ALTO client that retrieves network mapping from the ALTO Server, and a Location Database which stores network mappings received from the ALTO Client. The HTTP Server consults the Location Database when making redirection decisions.

#### 4. Request Routing as an Integration Point of ALTO into CDN

Content Distribution is a rich and evolving field. New architectures and approaches (e.g., a hybrid architecture using both servers and P2P) continue to be developed in the research community and industry and some are being deployed in production networks. While we would like to provide a survey of each possible CDN architecture and show how it may be integrated with ALTO, it would be a daunting task to track such a rapidly-changing field.

One scheme that is out of the scope of this document is P2P-only CDNs, where the application tracker takes the role of the ALTO Client, fetching the Network and Cost Maps from the ALTO Server and integrating them with its peer database. The result is a peer database that takes into account both the current peer metrics, such as peer availability or content availability, and network metrics, such as topological localization. This architecture in context of file sharing was extensively studied and trialed by ISPs such as Comcast [RFC5632] and China Telecom [I-D.lee-alto-chinatelecom-trial] under the ALTO/P4P [P4P] protocol. Thus, P2P-only CDNs are not discussed in this document.

Today, multiple request routing approaches can be used even in CDNs with purely server-based infrastructure. Thus, we take the approach of developing a basic request routing scheme covering all major CDN types. Specifically, the Request Routing Component of a CDN directs a request to a serving CDN node, and thus is the major integration point to utilize information available through ALTO. There are multiple request routing mechanisms, including HTTP Redirect, DNS name resolution, and anycast. We focus on HTTP Redirect and DNS name resolution. We briefly review the two mechanisms.

##### 4.1. HTTP Redirect

In this mechanism, an HTTP GET request from a host is received by an HTTP Request Router which sends back an HTTP responses with Status-Code 302 (Redirect) informing the host of the most optimal location to fetch the content. The HTTP Redirection method is already

commonly used in production CDNs as described in RFC3568 [RFC3568]. ALTO integration provides localization services where the device that performs the redirection becomes an ALTO client.

#### 4.2. DNS Request Routing

In this mechanism, the DNS server handling host requests provides the Request Routing Component. When the host performs a DNS query/lookup, the IP address contained in the response is already optimal for that query.

DNS queries can be either iterative or recursive. Iterative queries can be used with ALTO if the host itself queries the DNS Servers, or if the DNS Proxy used by the host is topologically close to the host. If the Host queries the DNS Servers, the authoritative DNS Server can see directly the host's IP address. If the DNS Proxy's is topologically close to the Host, its IP address is a good approximation for the host's location. In recursive queries, the authoritative DNS Server sees the IP address of the previous DNS Server in the resolution chain, and the IP address of the host is unknown. DNS-based request routing does not work with recursive DNS queries.

In an iterative DNS lookup with DNS Proxy, the host queries the Proxy, which in turn first queries one of the root servers to find the server authoritative for the top-level domain (com in our example). The Proxy then queries the obtained top-level-domain DNS server for the address of the DNS server authoritative for the CDN domain. Finally, the Proxy queries the DNS server that is authoritative for the cdn.com domain. The authoritative DNS Server for the cdn.com will perform the request routing to the most appropriate CDN node, based on the source IP address of the requestor. The host will then request the content directly from the CDN Node.

### 5. Basic Scheme of CDN/ALTO Integration

Although HTTP Redirect and DNS are quite different mechanisms to direct a request to a serving CDN node, as we will see, the basic structure of integrating ALTO with them can be quite similar. Thus, we first present common structures. We refer to the HTTP Redirect component or the DNS component of a CDN as a CDN Request Router.

#### 5.1. Basic Integration Scheme

Figure 1 shows a general structure to embed an ALTO Client into a CDN Request Router.

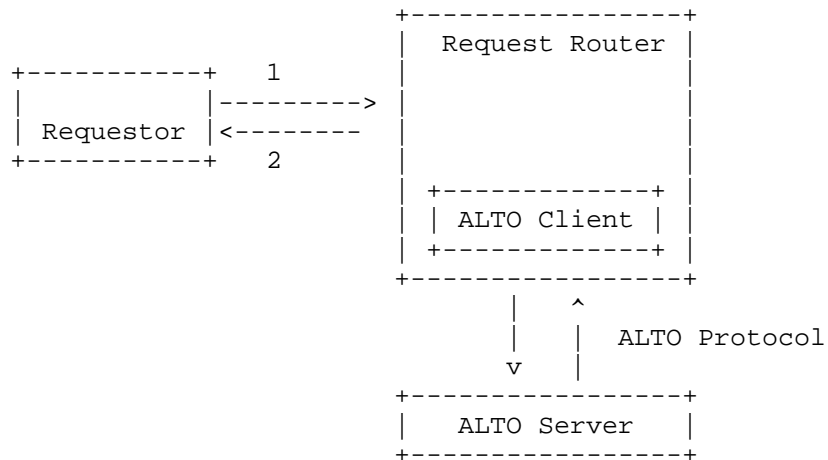


Figure 1: Request Router with ALTO

#### 5.1.1. ALTO for HTTP Redirect

To make the basic scheme more concrete, Figure 2 shows the case that the Request Router uses HTTP Redirect.

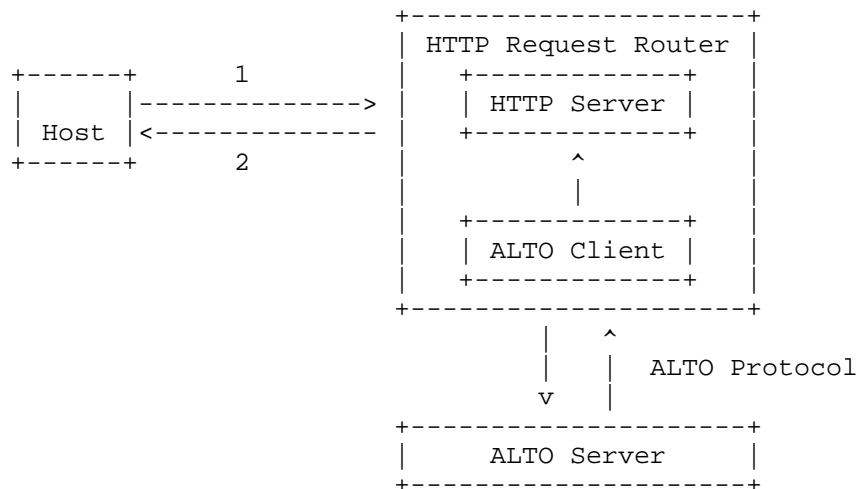


Figure 2: ALTO for HTTP Request Router

### 5.1.2. ALTO for DNS Resolution

Figure 3 shows the case that the Request Router uses DNS Resolution.

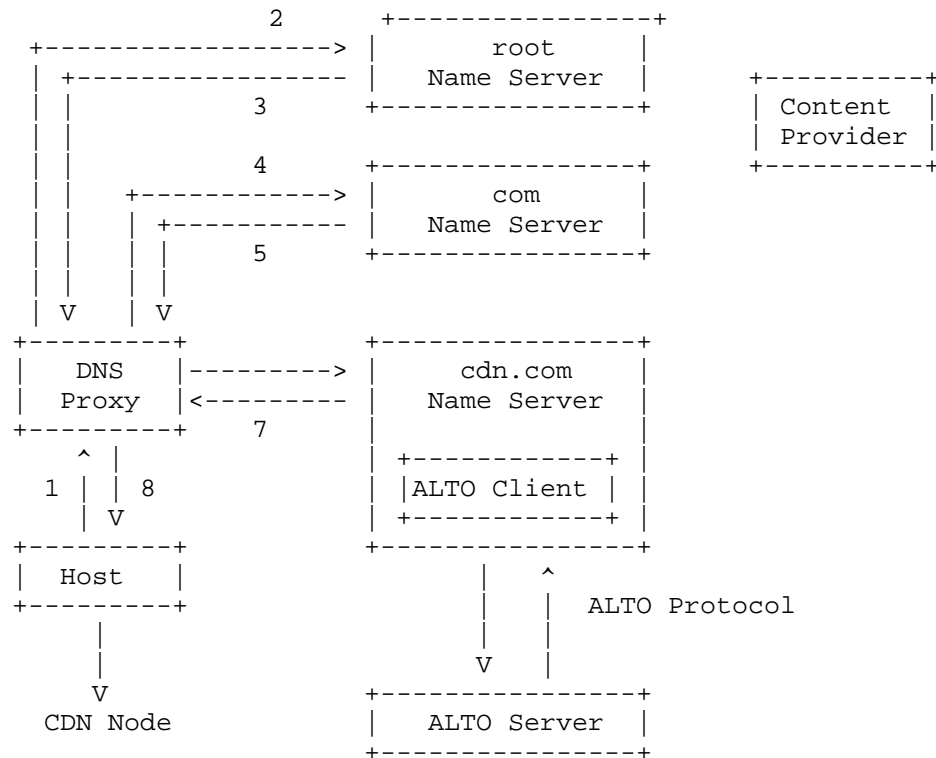


Figure 3: ALTO for DNS Resolution.

### 5.2. Multi-hop Redirection

The preceding examples show the logical flow for redirection. It is important to state that there maybe multiple redirection hops.

For HTTP Redirect, the requestor may be redirected again by the first CDN node. For DNS, the first DNS server may direct, using aggregated ALTO information (e.g., from multiple ALTO Servers of multiple ISPs), the DNS resolution to a second level DNS server, which then may use more specific ALTO information as well as CDN node status.

### 5.3. CDN Node Discovery and Status Notification

Since ALTO for HTTP Redirect and that for DNS have many common issues, we use the basic general scheme unless stated otherwise.

One common issue is how Request Router discovers the available CDN nodes and their locations. The exact mechanism is outside the scope of this document.

It is desirable that not only CDN node locations, but also real-time CDN node status (like health, load, cache utilization, CPU, etc.) is communicated to the CDN.

Specifically, CDN node status can be retrieved from the existing Load Balancer infrastructure. Most Load Balancers today have mechanisms to poll caches/servers via ping, HTTP Get, traceroute, etc. Most LBs have SNMP trap capabilities to let other devices know about these thresholds.

[yry: move]In addition to the CDN node status, network status can also be retrieved from TE/RP databases.

We see two ways that CDN node status can be communicated into the request routing decision process.

#### 5.3.1. CDN Node Status Updates received by Request Router

In this use case the Request Router receives CDN Status updates directly.

Specifically, the Request Router can implement an SNMP agent and get to know whatever is needed.

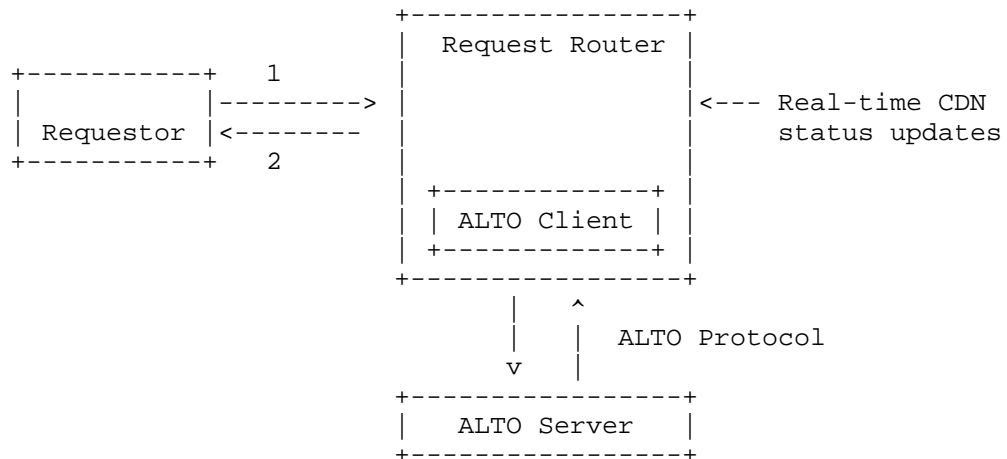


Figure 4: CDN Node Status to Request Router

### 5.3.2. CDN Node Status Updates received by ALTO

This model generally simplifies the Request Router. It allows an easier distribution of the Request Router, and to keep real time CDN status data updates in a logically centralized ALTO Server or in an ALTO Server Cluster. It allows for the Request Router and the ALTO Server to be in different administrative domains. For example, the Request Router can be in a Content Provider's domain, the ALTO Server and CDN Nodes in a Network Service Provider's domain.

Specifically, ALTO Server could provide an API (for example, a Web Service or XMPP-based API) that could be used by CDN nodes to communicate their status to the ALTO server directly.

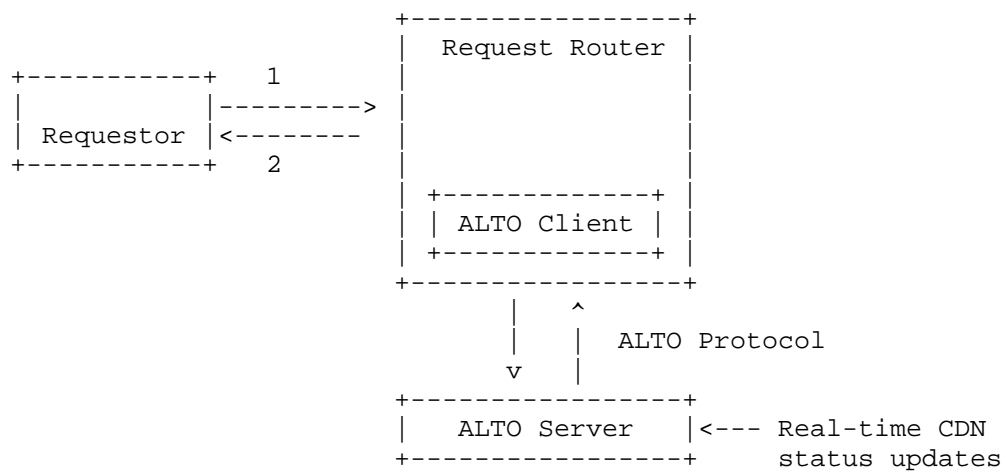


Figure 5: CDN Node Status to ALTO

## 6. Request Routing using ALTO Services

Either the Map Service or the Endpoint Cost Service of ALTO can be used by the Request Router.

### 6.1. Request Routing using the Map Service

The ALTO client embedded in the Request Router fetches the Network and Cost Maps from the ALTO Server and provides that information to the Request Router.

As an illustrative example, we consider the case of HTTP Redirect. A simple Request Router may be given (from an external source) the list of available CDN nodes. The Request Router precomputes a redirection



table indexed by source PID with values being the closest CDN nodes. This redirection table can be built based on information from Network and Cost Maps. Then when the Request Router receives an HTTP GET request, it looks up the PID of the source IP address on the request, indexes the redirection table using the request PID to select a CDN node, and finally returns a response that is an HTTP redirect with the URL of the selected CDN node. The URL in 302 Redirect may contain the IP address of the selected CDN node or a domain name instead of IP address due to virtual hosting. Therefore the IP addresses contained in the cost maps may need to be correlated to domain names a priori. In practice, the redirection table may be indexed by both source and content to provide better redirection.

The illustrative example can also be extended to DNS.

The Network Maps generated by the ALTO Server will contain both Host PIDs and CDN Node PIDs, i.e., Host PIDs contain host subnets; CDN PIDs contain IP addresses of available CDN nodes. Cost Maps may contain only cost from each host PID to each CDN PID and not the full matrix across all PIDs. The reason is that the Request Router may redirect a host only to a CDN node, not to another host as in the P2P case. Moreover, there is no generic way to disambiguate PIDs containing only hosts from PIDs containing CDN nodes.

It is possible that a Request Router may be designated as being responsible only for a fixed set of Host PIDs. This information can be made available to the Request Router before it receives requests from hosts. If the set of Host PIDs is not known ahead of time, the latency for serving requests will be impacted by the capabilities of the ALTO server.

With such information ahead of time, a Request Router that uses the Network Maps Service may pre-download the Network Map for the interesting Host PIDs and the CDN PIDs. It can also start periodically pulling Cost Map for relevant PID 2-tuples.

The Request Router can rely on the ALTO Server generated Cache-Control headers to decide how often to fetch CDN PID network map and Host PID network maps.

For Alto protocol requirements related to request routing with the Map Service see Section 8.1.1 and Section 8.1.2.

## 6.2. Request Routing using the Endpoint Cost Service

Alternatively, the Request Router may request the Endpoint service from the ALTO client.

Specifically, the Request Router requests the Endpoint Cost Service in order to rank/rate the content locations (i.e., IP addresses of CDN nodes) based on their distance/cost (by default the Endpoint Cost Service operates based on Routing Distance) from/to the user address.

Once the Request Router obtained from the ALTO Server the ranked list of locations (for the specific user) it can incorporate this information into its selection mechanisms in order to point the user to the most appropriate location.

A Request Router that uses the Endpoint Cost Service may query the ALTO Server for rankings of CDN Node IP addresses for each interesting Host and cache the results for later usage.

## 7. Multiple Administrative Domains

The preceding discussion works well in a single administrative domain setting: the CDN nodes are in the administrative domain of the ISP. However, the CDN nodes, the ISP, and the Request Router can be in different administrative domains. In this section, we consider a few such deployment cases. We use DNS as an example.

### 7.1. CDN nodes/Request Router in a separate administrative domain from that of ISP

In many situations, the CDN nodes and the Request Router are in a separate network managed by an entity that is distinct from the ISP. Consequently, the CDN nodes belong to a network with its own ALTO server that is distinct from the ALTO server of the ISP where the subscriber belongs.

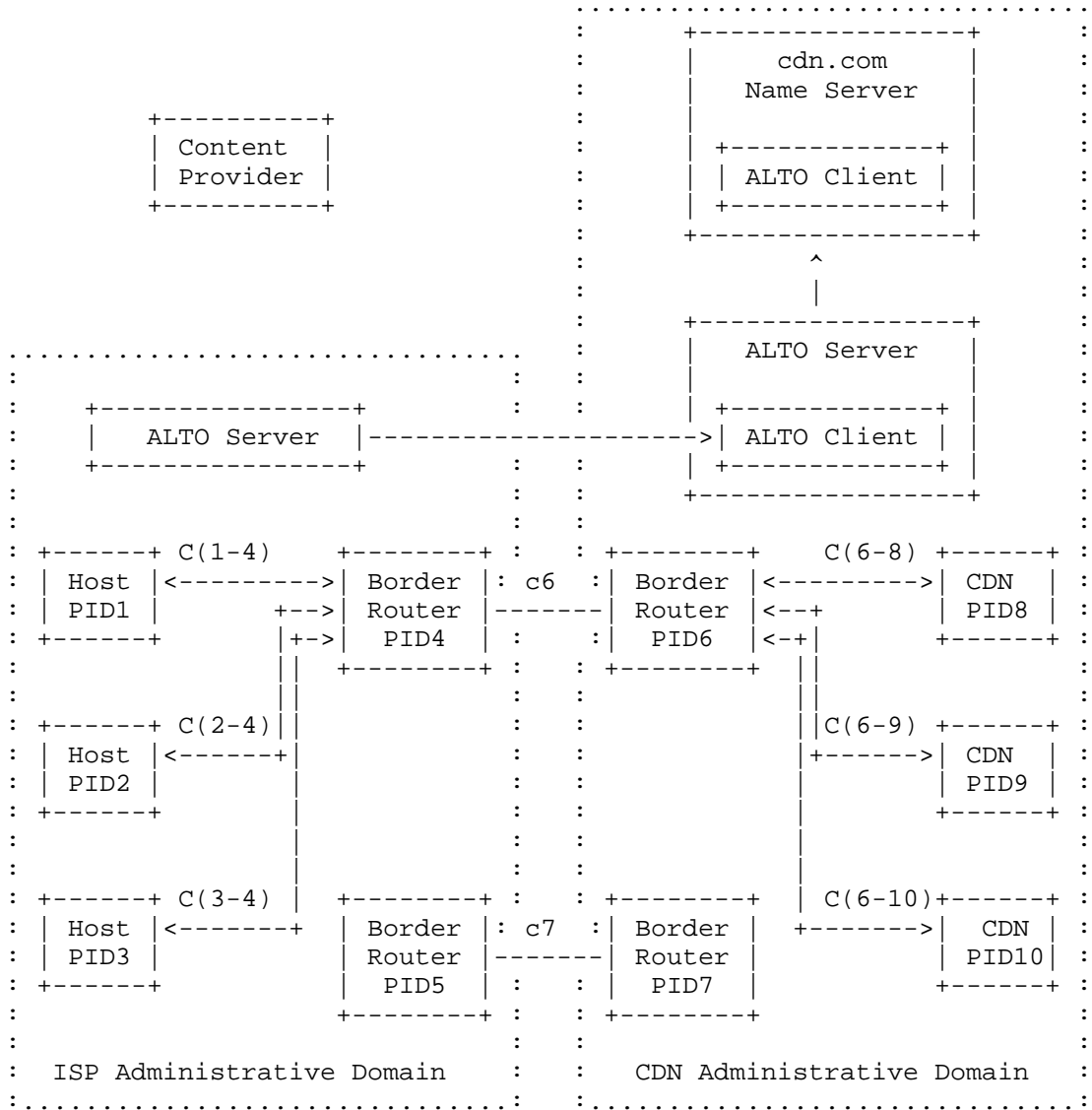


Figure 6: Map advertising between ISP and CDN domains

The ALTO server in the CDN provider network is assumed to be initialized with information about the ISP networks it serves. For every such ISP network, it consults the routing plane to find the set of Border routers. The CDN network ALTO server computes the cost of reaching each Border router from every CDN node (say,  $C_{cdn}$ ).

Next, the CDN ALTO server contacts the ISP network's ALTO server and downloads the network map. In order to help the CDN ALTO server compute the cost from a CDN node to a subscriber's PID, we break it down into two parts - the cost from the CDN node to the Border Router ( $C_{cdn}$ ) and the cost from the Border Router to the subscriber's PID (say,  $C_{isp}$ ). Note that for any chosen exit point,  $C_{cdn}$  may be computed locally by the CDN ALTO Server. However, the fundamental issue is that  $C_{isp}$  depends on the exit point (Border router) chosen by the CDN. There are multiple ways for the CDN ALTO Server to compute  $C_{isp}$  given the Network Map and Cost Map from the ISP's ALTO Server.

One possibility is for the ISP ALTO Server to define a special Border Router PID (denoted by a PID attribute) which also indicates the corresponding Border Router PID in the CDN. The attributes and values may be agreed-upon by the ISP and CDN when the ALTO Services are configured. For example, in the example shown in Figure 5, the ISP ALTO Server indicates that its PID4 and PID5 are Border PIDs, with corresponding PIDs in the CDN as PID6, and PID7, respectively. Then, CDN ALTO Server can locally compute  $C_{isp} = \text{cost}(\text{ISP Border Router PID}, \text{Subscriber PID})$ .

A second possibility for computing  $C_{isp}$  is to make use of Border Router IP addresses. The CDN's Border Router can locally determine the IP address of the connected border router in the ISP. In this approach, neither the CDN ALTO Server nor the ISP ALTO Server define PID attributes. The ISP ALTO Server is not required to define special PIDs for Border Routers - it only needs to ensure that Border Router IP addresses are aggregated appropriately in its Network Map.

Specifically, we identify two scenarios for the CDN ALTO Server to compute  $C_{isp}$  and  $C_{cdn}$ .

In the first scenario, the CDN does not conduct CDN-level multi-path routing from the CDN nodes to the subscriber hosts. Thus, the routing path from a CDN IP address to a subscriber host IP address is typically uniquely (if no ECMP) determined by the network routing system. In this scenario, for a given CDN node IP address to a subscriber host IP address, the CDN ALTO Server uses the routing system to compute the Border Egress router inside the CDN, and the corresponding Border Ingress router inside the ISP. Then the CDN ALTO Server has  $C_{cdn}$ (CDN node IP, Border Egress router IP inside the CDN), and  $C_{isp}$ (Border Ingress router IP inside the ISP, Subscriber IP). The computation of  $C_{cdn}$  and  $C_{isp}$  can be done using ALTO in the traditional way through either the Network Map and Cost Map or the Endpoint Cost Service.

In the second scenario, the CDN may support CDN-level multi-path

routing from the CDN nodes to the subscriber hosts. In particular, from each CDN node, the CDN has a capability (e.g., through tunneling) to send to a subscriber host IP through multiple Border Egress routers (e.g., through any Egress router that receives an announcement from the ISP of the subscriber host IP). In this case, the cost of reaching a host PID from a given CDN node is then determined as the minimum cost among all possible intermediate Border Routers.

If the network is homogeneous, then a good approximation of the cost between each host PID and a given CDN node can be given as:  $C_{cdn}(\text{CDN Node, Border router}) + C_{isp}(\text{Border router, Subscriber PID})$ . In this computation, the Border Router is the one that is on the best path from the CDN node to the Subscriber PID.

The CDN ALTO server now has a cost map that provides the cost from each CDN node to all known Subscriber PIDs. The ALTO client in the CDN DNS server downloads this cost map in preparation for subscriber DNS requests.

When a subscriber DNS request arrives at the CDN provider's DNS server, it looks up the network map and maps the source IP address to a Subscriber PID. It then uses the cost map to pick the best CDN node for this Subscriber PID.

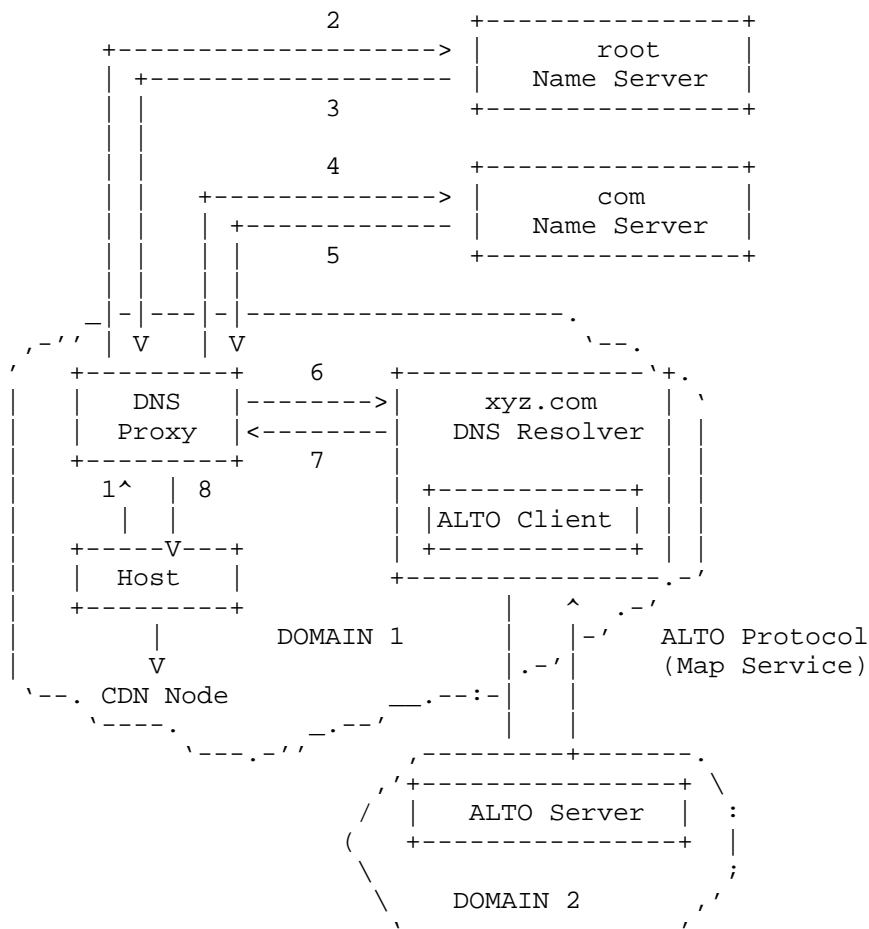
## 7.2. Managed DNS Domain with Three Administrative Domains

Many organizations / content providers outsource DNS management to the external vendors for various reasons like reliability, performance improvement, DNS security etc. Managed DNS service could be used either with caches owned by the organization itself (section 6.3.1) OR with external CDNs (section 6.3.2)

### 7.2.1. Managed DNS Redirect to Local CDN

One of the common functions offered by managed DNS service vendor is DNS traffic management where DNS resolver can load balance traffic dynamically across CDN servers.

Typically managed DNS service provider has DNS resolvers spread across geographical locations to improve performance. This also makes easier for DNS resolver to redirect host to the nearest cache. Such a DNS resolver would be an ideal candidate to implement ALTO client where it can fetch network map and cost map from ALTO servers located in the same geographical area only. Load balancing implemented with the knowledge of network and cost map would be more efficient than other mechanisms like round robin.



In the figure above, there exists 2 possibilities:

Case 1: Domain 1 and Domain 2 are connected to the same service provider network. This case is similar to section 6.1

Case 2: Domain 1 and Domain 2 are connected to different service provider network. This case is similar to section 6.2

#### 7.2.2. Managed DNS with CDN-Provided Request Routing

It is also possible to utilize a Managed DNS service and still rely on a CDN's request routing. For example, this could be done if a network provider wishes to utilize a Managed DNS provider, but also wishes to integrate its own CDN using ALTO with DNS-based request routing.

To support this, the network provider may submit any necessary configuration files (e.g., indicating necessary CNAME records) to redirect CDN requests to the CDN's DNS request routing mechanism. Requests for the CDN (e.g., 'cdn.isp.com') will then be directed by DNS request routing, while requests for other hosts are handled by the Managed DNS solution.

## 8. Protocol Recommendations

In the previous sections, this document has taken the approach of providing information on existing CDN approaches and possible benefits of utilizing ALTO. However, in developing the taxonomy, use cases, and deployment scenarios, we have identified cases where the ALTO Protocol [I-D.ietf-alto-protocol] and Server Discovery [I-D.kiesel-alto-3pdisc] [I-D.song-alto-server-discovery] [I-D.stiemerling-alto-dns-discovery] may be lacking capabilities that may be helpful and/or necessary for usage with CDNs. We now focus on detailing these gaps with the goal of providing feedback and recommendations. Note that some protocol changes may be necessary in the core protocol, while others may be implemented as extensions.

This section will be updated to track changes in the ALTO Protocol, ALTO Server Discovery, and accompanying protocols.

### 8.1. Necessary Additions

This section details changes to the ALTO protocols that would be necessary to make use of ALTO within CDN infrastructures. We classify a change as "necessary" if there is a core feature of a CDN/ALTO integration that is not possible to implement with the existing protocols.

#### 8.1.1. NA1: PID Attributes

In order to disambiguate between PIDs that contain endpoints of a specific class, a PID property is needed. A PID can be classified as containing "CDN nodes", "Mobile Hosts", "Wireline Hosts", etc. This mechanism can be used to provide an ALTO Client a list of nodes of a particular type, along with the ALTO Costs to each node.

#### 8.1.2. NA2: PID Attributes and Query

PID attributes can be used by the ALTO Client to select a appropriate host and also passed as a constraint in the map filtering service.

## 8.2. Helpful Additions

This section details changes to the ALTO Protocol that would be helpful to make use of ALTO within CDN infrastructures. We classify a change as "helpful" if there is a compelling extension to existing CDNs that would be possible with additional functionality within ALTO, or if there is a component of CDN/ALTO integration that could be made more efficient or otherwise improved with additional ALTO functionality.

### 8.2.1. HA1: Push Mechanism

It is important for the ALTO Service through the ALTO protocol or a companion protocol to provide a push mechanism from server to client. The push mechanism can be a notification that new data is available or the data itself.

### 8.2.2. HA2: Incremental Map Updates

A natural evolution to the protocol if maps are large and change often is to allow for incremental map updates. In this sense the map contained in the reply would be considered the delta from the previous version.

### 8.2.3. HA3: ALTO Border Router PID attribute

In order for administrative domains to collate costs across domain boundaries, the border routers may be placed in their own PIDs. Such PIDs may be identified by a Border Router attribute.

### 8.2.4. HA4: CDN ALTO Server Discovery

In certain deployment scenarios, it may be beneficial for an ALTO client to directly query a CDN's ALTO Server (instead of the CDN's ALTO Server only being consulted as a backend process). For example, this can provide more accurate guidance than DNS request routing since the client's IP address may be directly used by the CDN in order to select a cache node. This would require an ALTO Client (e.g., an ISP subscriber) to be able to discover an ALTO Server owned and/or managed by a CDN. This could be done by an extension to the discovery protocol, or it could be done by allowing an ISP's ALTO Server to redirect certain queries to a CDN ALTO Server.

### 8.2.5. HA5: Extensible ALTO Cost Maps

Certain deployment scenarios may benefit from additional information being carried within ALTO information. For example, a trusted neighboring ISP B may be able to help ISP A optimize multihoming



costs. To provide an extensible way to communicate additional data, the ALTO Protocol could be extended to include opaque data strings (in addition to numeric and ordinal values) in an ALTO Cost Map.

#### 8.2.6. NA4: Federated Deployment of ALTO Servers

There is a need to define how ALTO servers may communicate with each other in a federated model.

### 9. IANA Considerations

This document makes no request of IANA.

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

### 10. Security Considerations

When the ALTO Server and Client are operated by different entities the issue of trust and security comes forward. The exchange of information could be done using the encryption methods already present in HTTP but preventing unauthorized redistribution comes into play. A further issue is if the ALTO information information is transitive, which modifications are allowed.

### 11. Acknowledgements

We would like to thank Mayuresh Bakshi for valuable input and contributions to this draft. We would also like to thank Nabil Bitar, Manish Bhardwaj, Michael Korolyov, Steven Luong and Ferry Sutanto for their comments.

### 12. References

#### 12.1. Normative References

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

#### 12.2. Informative References

[ARBOR] Labovitz, "Internet Traffic and Content Consolidation", 2009, <<http://www.ietf.org/proceedings/10mar/slides/plenaryt-4.pdf>>.

- [I-D.ietf-alto-protocol]  
Alimi, R., Penno, R., and Y. Yang, "ALTO Protocol",  
draft-ietf-alto-protocol-05 (work in progress), July 2010.
- [I-D.kiesel-alto-3pdisc]  
Kiesel, S., Tomsu, M., Schwan, N., Scharf, M., and M.  
Stiemerling, "Third-party ALTO server discovery",  
draft-kiesel-alto-3pdisc-03 (work in progress), July 2010.
- [I-D.lee-alto-chinatelecom-trial]  
Li, K., Wang, A., and K. Zhou, "ALTO and DECADE service  
trial within China Telecom",  
draft-lee-alto-chinatelecom-trial-00 (work in progress),  
July 2010.
- [I-D.song-alto-server-discovery]  
Yongchao, S., Tomsu, M., Garcia, G., Wang, Y., and V.  
Avila, "ALTO Service Discovery",  
draft-song-alto-server-discovery-03 (work in progress),  
July 2010.
- [I-D.stiemerling-alto-dns-discovery]  
Stiemerling, M. and H. Tschofenig, "A DNS-based ALTO  
Server Discovery Procedure",  
draft-stiemerling-alto-dns-discovery-00 (work in  
progress), July 2010.
- [P4P]  
Xie, H., Yang, YR., Krishnamurthy, A., Liu, Y., and A.  
Silberschatz, "P4P: Provider Portal for (P2P)  
Applications", March 2009.
- [RFC3568]  
Barbir, A., Cain, B., Nair, R., and O. Spatscheck, "Known  
Content Network (CN) Request-Routing Mechanisms",  
RFC 3568, July 2003.
- [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

Reinaldo Penno  
Juniper Networks

Email: rpenno@juniper.net

Satish Raghunath  
Juniper Networks

Email: [satishr@juniper.net](mailto:satishr@juniper.net)

Jan Medved  
Juniper Networks

Email: [jmedved@juniper.net](mailto:jmedved@juniper.net)

Richard Alimi  
Google

Email: [ralimi@google.com](mailto:ralimi@google.com)

Richard Yang  
Yale University

Email: [yry@yale.edu](mailto:yry@yale.edu)

Stefano Previdi  
Cisco Systems

Email: [sprevidi@cisco.com](mailto:sprevidi@cisco.com)



Network Working Group  
Internet-Draft  
Intended status: Experimental  
Expires: April 18, 2011

S. Randriamasy, Ed.  
Alcatel-Lucent Bell Labs  
October 15, 2010

Multi-Cost ALTO  
draft-randriamasy-alto-multi-cost-00

Abstract

IETF is designing a new service called ALTO (Application Layer traffic Optimization) that includes a "Network Map Service", an "Endpoint Cost Service" and an "Endpoint (EP) Ranking Service" and thus incentives for application clients to connect to ISP preferred Endpoints. These services provide a view of the Network Provider (NP) topology to overlay clients.

The present draft proposes a light way to extend the information provided by the current ALTO protocol. The purpose is to broaden the possibilities of the Application Clients in two ways: firstly by providing a better mapping of the Selected Endpoints to needs of the growing diversity of Content Networking Applications and to the network conditions, secondly by producing a more robust choice of multiple Endpoints, helping thus out for efficient Multi-Path transfer.

There are 2 parts in this draft: the first part proposes protocol extensions to support requests on multiple CostTypes in 1 transaction; the second part proposes additional CostTypes and Cost attributes such as validity period, timeframe and reliability.

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 April 18, 2011.

#### Copyright Notice

Copyright (c) 2010 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 . . . . .	4
2. Scope . . . . .	5
3. Terminology . . . . .	5
4. Proposed ALTO services updates . . . . .	6
4.1. Endpoint Cost Service with multiple Cost Types . . . . .	6
4.2. All Costs Types in one response with vector cost values . . . . .	6
4.3. Proposed additional Cost Types . . . . .	7
4.4. Statistical costs with a timeframe . . . . .	7
5. Proposed ALTO protocol updates . . . . .	7
5.1. Proposed updates for Multi-Cost ALTO . . . . .	8
5.1.1. Multi-Cost Attributes . . . . .	8
5.2. Proposed additional Properties and Costs . . . . .	9
5.2.1. Proposed additional Endpoints properties . . . . .	9
5.2.2. Scoping ALTO information . . . . .	10
5.2.3. Proposed additional Cost Types . . . . .	10
5.3. ALTO Status Codes for Multi-Cost ALTO . . . . .	11
5.4. Examples of Multi-Cost ALTO messages . . . . .	11
6. Use case . . . . .	11
6.1. Scenario . . . . .	11
6.2. Illustrative ALTO use case . . . . .	12
7. IANA Considerations . . . . .	15
8. Acknowledgements . . . . .	15
9. References . . . . .	15
9.1. Normative References . . . . .	15
9.2. Informative References . . . . .	15
Author's Address . . . . .	15

## 1. Introduction

IETF is designing a new service called ALTO that provides guidance to P2P applications, which have to select one or several hosts from a set of candidates that are able to provide 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 Quality of Experience (QoE) in the application while reducing resource consumption in the underlying network infrastructure. The ALTO protocol conveys the Internet View from the perspective of a Provider Network region that spans from a region to one or more Autonomous System (AS). Together with this Network Map, it provides the Provider determined Cost Map between locations of the Network Map. Last, it provides the Ranking of Endpoints w.r.t. their routing cost.

The term Network Provider in this document includes both ISPs, who provide means to transport the data and Content Delivery Network (CDN) operators who care for the dissemination, persistent storage and possibly identification of the best/closest content copy.

The last ALTO protocol draft see [ID-alto-protocol5], gives the possibility to query multiple Endpoint properties at once (see S.7.7.4.1). However section 7.7.3.2 on Cost Map states about both parameters Cost Type and Cost Mode that: "This parameter MUST NOT be specified multiple times". The ALTO requirements draft, see [ID-ALTO-Requirements] also states in REQ. ARv05-14: "The ALTO client protocol MUST support the usage of several different rating criteria types". In the current protocol draft, there is no specified way to get values for several Cost Types altogether. Currently, the costs are provided in a scalar form, one by one. So that an ALTO Client wanting information for several Cost Types must place a request and receive a response as many times as desired Cost Types. However, vector costs provide a robust and natural input to multi-path connections and getting all costs in one single query/response transaction saves time and ALTO traffic, thus resources, thus energy.

The ALTO Problem Statement, see [RFC5693] and the ALTO requirements draft, see [ID-ALTO-Requirements] stress that: "information that can change very rapidly, such as transport-layer congestion, is out of scope for an ALTO service. Such information is better suited to be transferred through an in-band technique at the transport layer instead", as "ALTO is not an admission control system "and does not necessarily know about the instant load of endpoints and links. However, longer term statistics or empirical ratings on performance oriented information may still be useful for a reliable choice of candidate endpoints. In addition, given the QoE requirements of



nowadays and future Internet applications, more and more NPs compute and store such information to optimize their traffic. Last, specific ALTO servers can be specified for mobile core networks, which have a smaller scale and can afford and take advantage of using smaller time-scale network information.

Adding QoE-enabling metrics to the Network Provider established routing cost could meet the interests of both the end users and the Providers. Besides, keeping the shortest or cheapest possible path, in addition, saves resources, time and energy.

## 2. Scope

This draft generalizes the case of a P2P client to include the case of a CDN client, a GRID application client and any Client having the choice in several connection points for data or resource exchange. To do so, it uses the term "Application Client" (AC).

This draft focuses on the use case where the ALTO client is embedded in the Application Client. For P2P applications, the use case where the ALTO Client is embedded in the P2P tracker is also applicable.

It is assumed that Applications likely to use the ALTO service have a choice in connection endpoints as it is the case for most of them. The ALTO service is managed by the Network Provider and reflects its preferences for the choice of endpoints. The NP defines in particular the network map, the routing cost among Network Locations, and which ALTO services are available at a given ALTO server.

The solution proposed in this draft is applicable to fixed networks. It is also meant for smaller networks such as mobile networks.

## 3. Terminology

Endpoint (EP): can be a Peers, a CDN storage location, a Party in a resource sharing swarm such as Grid or online gaming.

Endpoint Discovery (EP Discovery) : this term embraces the different types of processes used to discover different types of endpoints.

Network provider: includes both ISPs, who provide means to transport the data and Content Delivery Network (CDN) who care for the dissemination, persistent storage and possibly identification of the best/closest content copy.

Application Client (AC): this term generalizes the case of a P2P

client to include the case of a CDN client and of any Client having the choice in several connection points for data or resource exchange.

Traffic Engineered End Point Optimization Tool (TEEPOT): this is a functional entity introduced in this draft, that is linked to an ALTO Client and to an Application Client. Its role is to assist the selection of Endpoints upon Allocation needs and the ALTO responses. It can be a specific group of functions or an already existing function.

#### 4. Proposed ALTO services updates

The currently available ALTO services supporting Endpoint evaluation are: Endpoint Cost Service, Cost Map and Filtered Cost Map. The ALTO client may want to simultaneously use a number  $N > 1$  of cost metrics referred to as Cost Types in ALTO. The only possibility in the current ALTO protocol is to sequentially place as many requests as desired cost types. This draft proposes to add the following features:

##### 4.1. Endpoint Cost Service with multiple Cost Types

Some application clients may want to consider several metrics to select the endpoints appropriately w.r.t. the application needs. Clients may also want to use multiple paths for the transfer of particular data bulks, possibly selected with several metrics. Therefore the Endpoint Cost Lookup and the Cost Map Services should have the possibility to handle several metrics.

##### 4.2. All Costs Types in one response with vector cost values

Providing all the numerical costs simultaneously with only one request and response exchange saves time, resources and energy. To avoid overloading the network with ALTO traffic with multiple requests for Cost Types, we propose that the Cost values provided by the ALTO server be arranged in a vector. This requires:

- o firstly to add an ALTO Cost Attribute called for instance "Cost Length" that provides the number  $N$  of desired Cost Types,
- o secondly to put the requested cost values in a vector having a number  $N$  of components, where  $N$  is equal to Cost Length.

As specified in the ALTO Requirements [ID-ALTO-Requirements] "REQ. ARv05-19: The ALTO reply message SHOULD allow the ALTO server to express which rating criteria have been considered when generating

the reply." That is, the ALTO response indicates the mapping between vector components and Cost Types.

Note that in this case, the ALTO client MUST require the Cost Mode "numerical" that is the Mode MUST NOT be "ordinal".

#### 4.3. Proposed additional Cost Types

The current ALTO protocol draft provides examples of metrics in section 5.1.1, that are: air miles, hop-counts or generic routing costs. Statistics or longer term ratings on path bandwidth and latency may also be considered. Additional Endpoint properties may be useful, such as the memory capacity or statistical scores on the load and possibilities of an Endpoint.

#### 4.4. Statistical costs with a timeframe

The ALTO Requirements Draft [ID-ALTO-Requirements] advises against instant performance-related cost metrics as they may be easily captured by online mechanisms and in addition, the ALTO service does not know how a Peer manages its sending rate. Application clients however may have good reasons and wise ways to use performance related information in the mid to long term ,on Endpoints that they don't know in advance and on which they therefore cannot plan measurements. Other applications may wisely use static performance indicators such as nominal memory capacity.

Dynamic performance indicators can be represented by scores, reflecting some overall performance, in a static way or with values periodically updated according to a timeframe. A timeframe SHOULD be sent along with the statistical Cost Types if the latter are available. By default this timeframe corresponds to permanent validity.

### 5. Proposed ALTO protocol updates

This section proposes updates or additions to the ALTO protocol to support Multi Cost ALTO Services or provide additional ALTO information. The applicable ALTO services are:

- o Cost Map Service,
- o Cost Map Filtering Service,
- o Endpoint Property Lookup Service,

- o Endpoint Cost Lookup Service.

#### 5.1. Proposed updates for Multi-Cost ALTO

If an ALTO client desires several Cost Types, instead of placing as many requests as costs, it may request and receive all the desired cost types in one transaction. The correspondence between the components and the cost type MUST be indicated in the ALTO request.

The ALTO server then, provided it supports the desired cost, and provided it supports the vector cost values, sends one single response where for each {source, destination} pair, the cost values are arranged in a vector, whose component each corresponds to a specified Cost Type. The correspondence between the components and the cost types MUST be indicated in the ALTO response.

The following ALTO protocol services and features need to be updated to enable Multi Cost ALTO transactions.

- o Endpoint (EP) Cost (see [ID-alto-protocol5], S. 3.2.4 and S. 7.7.5).
- o Cost attributes (see [ID-alto-protocol5], S. 5.1).
- o Cost Map (see [ID-alto-protocol5] S. 5 and 7.7.2.2):
  - \* between Network Locations (that are groups of 1 or several endpoints).
- o Cost Map filtering: need the same updates as for the Cost Map.

##### 5.1.1. Multi-Cost Attributes

To enable Multi-Cost ALTO Cost Services, we propose the following updates to the Cost Attributes, described in [ID-alto-protocol5] S. 5.1.

- o addition of attribute "Cost Length", a numerical value equal to the number of requested EP Cost Types.
- o extension of the attribute Cost Type from a single value to a vector of  $N \geq 1$  values. If  $N > 1$ , then the values WILL be interpreted as numerical values.
- o addition of definitions that list and identify the Cost Types supported by the acting ALTO server. These definitions can be formulated with alphanumeric strings,

- o definition of the correspondence between an index "i\_typecost" in [1,N] in a cost vector and the ID of the defined alphanumeric cost types.
- o optional addition of a reliability vector having the same dimension as the cost vector and that reflects, for each component of the vector, the reliability of the provided cost value, for instance in statistical terms or as a percentage. Values lying in [0,1] can also be a good option.
  - \* by default, the reliability is considered as total,
  - \* the unit of validity values MUST be specified.
- o optional association of a validity timeframe to the reliability vector, indicating how long the information can be considered as up to date.
  - \* by default the validity timeframe WILL be considered infinite.

To the attribute Cost Mode in S.5.1: addition of a rule stipulating that when multiple cost types are requested, then the requested Cost Mode MUST be numerical. If the attribute Cost Length is > 1 and the Cost Mode is set to "ordinal", then one option is that the ALTO Server returns the 'Sucess' code "E\_INVALID\_COST\_TYPE".

## 5.2. Proposed additional Properties and Costs

### 5.2.1. Proposed additional Endpoints properties

The Endpoint Properties given as example in [ID-alto-protocol5] S.3.2.3 mostly apply to fixed end nodes. We propose to add other properties, that are static, contribute to reflect the potential physical abilities of end nodes and therefore may guide their selection. In addition, these properties apply to end nodes connected by any access technology. Example additional properties include:

- o EP capacity in memory,
- o EP nominal bandwidth,
- o EP access technology.

Note that if this service is not supported, it is possible although less convenient to get the information at the overlay level, thus without the ALTO server.

### 5.2.2. Scoping ALTO information

One way to moderate the ALTO traffic load while maintaining some reliability is to associate the following attributes to the applicable ALTO information:

- o a Time Frame attribute: this is the period during which an information is considered applicable, for example 5 minutes, 2 hours, one month. When a time framed Property Service is supported by the ALTO server, the Time Frame parameter can be by default set to "permanent".
- o a Time To Expire counter associated to some lifetime attribute and the Time Frame as proposed in REQ ARv05-27 of [ID-ALTO-Requirements] . By default, this parameter can be set to infinity.
- o RELIABILITY LEVEL: reflects the degree of likelihood of the property, either a statistical value or a percentage.

The Time Frame and Time To Expire values can be used by the aging mechanism as proposed in REQ ARv05-28 of [ID-ALTO-Requirements] for a better synchronization of Cost Information collected at various times and places.

### 5.2.3. Proposed additional Cost Types

Additional Cost Types may be used in either the Cost Map or the Endpoint Cost Lookup Services and include:

- o Endpoint availability: indicating how often an Endpoint is reachable, preferably as a percentage. To be further specified. Possibly with associated Time frame and Time To Expire.
- o Endpoint reliability: indicating how easily an Endpoint is reachable, and / or the degree of continuity of its reachability, preferably as a percentage. To be further specified. Possibly with associated Time frame and Time To Expire.
- o Endpoint Load: indicating the average load, preferably as a percentage, or a quantitative coarse grain index indicating whether this Endpoint is in a rush period or calm period. To be further specified. Possibly with associated Time frame and Time To Expire.
- o Path robustness: one or more timeframed indicators related to statistical evaluations of the path performance on bandwidth, delay, packet loss, or other such metrics. This Cost can also be

represented by a quantitative coarse grain index indicating whether this Endpoint is in a rush period or calm period. To be further specified. Possibly with associated Time frame and Time To Expire.

### 5.3. ALTO Status Codes for Multi-Cost ALTO

If the vector cost structure is not supported, then the ALTO server sends an ALTO status code 7 corresponding to HTTP status code 501 indicating "Invalid cost structure". The ALTO client may then needs to place as many requests as needed Cost Types, and the ALTO server sends as many cost maps or EP cost as needed.

To the attribute Cost Mode in S.5.1 should be associated a rule stipulating that when multiple cost types are requested, then the requested Cost Mode MUST be numerical. If the attribute Cost Length is > 1 and the Cost Mode is set to "ordinal", an option is that the ALTO Server returns the 'Sucess' code "E\_INVALID\_COST\_TYPE".

### 5.4. Examples of Multi-Cost ALTO messages

Request and Response syntax. To be further specified.

## 6. Use case

### 6.1. Scenario

A Multi-Cost ALTO transaction is illustrated in a simple scenario, where an application client in a terminal wants to use several paths for a data transfer. This scenario applies to a terminal having access to the network via one or several interfaces.

The application client for example wants 3 paths per transfer:

- o 1 path optimising the Cost Type "routingcost",
- o 2 paths optimizing 2 metrics: the Cost Type "routingcost" and an Endpoint property named "EP memory".
  - \* The application client in addition wants these 2 paths to optimize the first criterion with a weight W\_PATH\_LENGTH equal for example to 0.4 and the second criterion with a weight W\_EP\_MEMORY equal to 0.6.
  - \* If the EP Property Service provides the information on Endpoint Load, then the application client wants this information in the available time frame closest to 1 hour.

A TEEPOT connected with the ALTO Client and the Application Client takes in the list of candidate Endpoints from the Application Client and prepares for the ALTO Client the request to the ALTO Server, in particular the following values: EP Cost Length, vector EP Cost Type [EP Cost Length], vector TimeFrame[EP Cost Length], with components equal to either a value or an indication of "not applicable".

## 6.2. Illustrative ALTO use case

Figure 1 shows the example scenario in the last IETF ALTO protocol draft, where the ALTO client is embedded in the P2P Client and requires an ALTO server servicing its own ISP to provide the Endpoint Cost for a list of gathered peers.

As written in [ID-alto-protocol5], the use case proceeds as follows:

1. The P2P Client discovers peers from sources such as Peer Exchange (PEX) from other P2P Clients, Distributed Hash Tables (DHT), and P2P Trackers.
2. The P2P Client queries the ALTO Server's Ranking Service, including discovered peers as the set of Destination Endpoints, and indicates the 'ordinal' Cost Mode. The response indicates the ranking of the candidate peers.
3. The P2P Client connects to the peers in the order specified in the ranking.

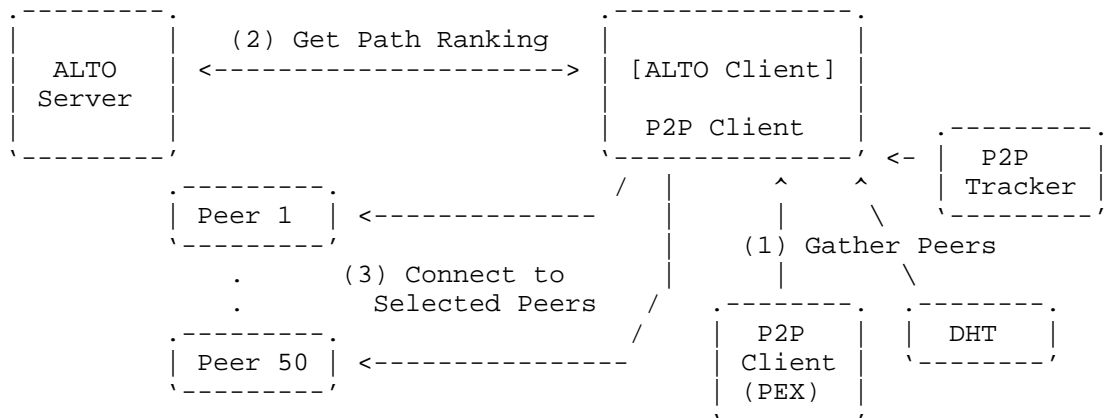


Figure 1: example scenario in the last IETF ALTO protocol draft, where the ALTO client is embedded in the P2P Client



Figure 2 depicts the features and mechanisms added to the current ALTO scenario for Multi-Cost ALTO services, for the use case of Figure 1. The EPs have already been discovered. In this figure, the term Peer is replaced by the term Endpoint (EP), the term P2P Client by Application Client and an Endpoint Tracker for resource Sharing Applications is added to the tools involved in Step (1) Gather Endpoints.

We focus on the ALTO use case where the ALTO client is co-located with an Application client in a terminal node, as not all P2P systems use a P2P tracker for peer discovery and selection as written in section 8.2 of [ID-alto-protocol5]. In Figure 2, the entity called P2P Client mentioned in the current protocol draft is zoomed to an entity called in this draft "Client Block" and that links: the Application Client (AC), its ALTO Client and the Traffic Engineered EP Optimization Tool (TEEPOT).

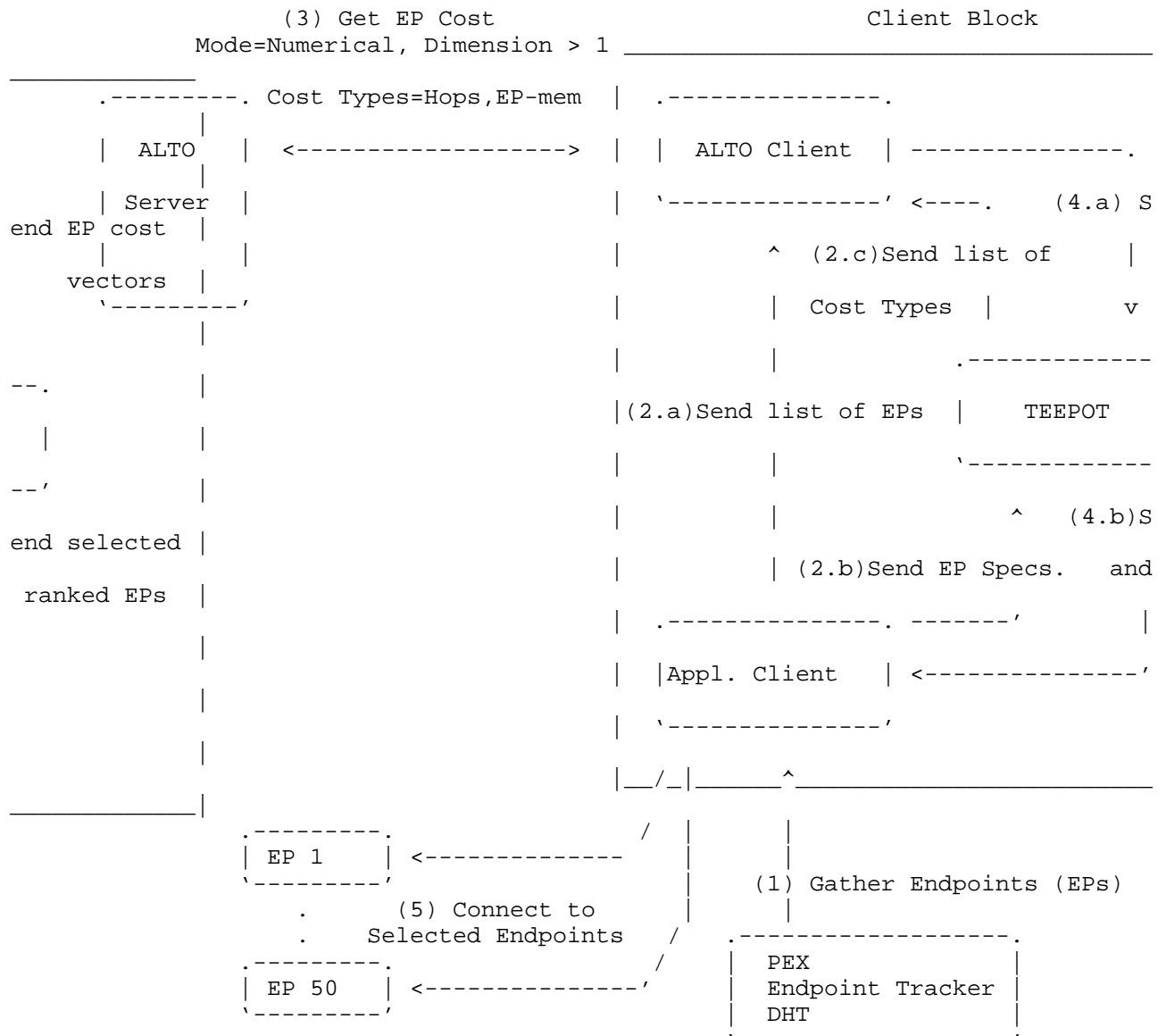


Figure 2: features and mechanisms added to the current ALTO scenario for Multi-Cost ALTO services

Randriamasy

Expires April 18, 2011

[Page 13]

The use case in Figure 2 proceeds as follows:

1. The Application Client discovers Endpoints (EPs) from sources such as Peer Exchange (PEX) from other P2P Clients, Distributed Hash Tables (DHT), P2P Trackers or other types of EP trackers.
2. In the "Client Block" gathering the Application Client (AC), its ALTO Client and the Traffic Engineered EP Optimization Tool (TEEPOT):
  - A. the Application Client (AC) sends to the ALTO Client the list of the discovered peers as the set of Destination Endpoints.
  - B. the Application Client (AC) sends to the TEEPOT the specifications on the EPs to select, according to the needs of the application. For example, AC needs 3 EPs, with 1 EP optimizing the Path Length Metric and 2 EPs optimizing the Path Length and the EP Memory Capacity Score, with respective weights of 0.4 and 0.6.
  - C. the TEEPOT indicates to the ALTO Client that the Service to request is EP Cost, with the Cost Mode set to "Numerical", and the Cost Dimension equal to the number of requested metrics and with the index of the requested Cost Types.
3. The ALTO Client queries the ALTO Server's EP Cost Service, sends the list of the discovered peers as the set of Destination Endpoints and indicates the 'numerical' Cost Mode, with a Cost Dimension equal to 2 and the index of requested metrics, corresponding in this example to: "Path Length" and "EP Memory Capacity Score". The response is the set of metric values associated to each EP.
4. In the Client block:
  - A. The ALTO Client hands to the TEEPOT the list of EPs and their associated value set.
  - B. The TEEPOT ranks the EPs with some smart algorithm, given the metric weights and then sends the ranked list to the Application Client.
5. The Application Client connects to the selected EPs.

## 7. IANA Considerations

This document makes no request of IANA.

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

## 8. Acknowledgements

## 9. References

### 9.1. Normative References

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

[RFC5693] "Application Layer Traffic Optimization (ALTO) Problem Statement", October 2009.

### 9.2. Informative References

[ID-ALTO-Requirements]  
"draft-ietf-alto-reqs-05.txt", June 2010.

[ID-alto-protocol5]  
"ALTO Protocol" draft-ietf-alto-protocol-05.txt",  
July 2010.

## Author's Address

Sabine Randriamasy (editor)  
Alcatel-Lucent Bell Labs  
Route de Villejust  
NOZAY 91460  
FRANCE

Email: Sabine.Randriamasy@alcatel-lucent.com



ALTO  
Internet-Draft  
Intended status: Standards Track  
Expires: January 13, 2011

H. Song  
Huawei  
M. Tomsu  
Alcatel-lucent Bell Labs  
G. Garcia  
Telefonica I+D  
Y. Wang  
Microsoft Corp.  
V. Pascual  
Consultant  
July 12, 2010

ALTO Service Discovery  
draft-song-alto-server-discovery-03

Abstract

Application-Layer Traffic Optimization (ALTO) service aims to provide distributed applications with information to perform better-than-random initial peer selection when multiple peers in the network are available to provide a resource or service. In order to discover an Application-Layer Traffic Optimization (ALTO) Server, a set of mechanisms are required. These mechanisms enable applications to find an information source which provides them with information regarding the underlying network. This document discusses various scenarios of ALTO discovery and specifies the use of several available options such as DHCP or DNS.

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 13, 2011.

Copyright Notice

Copyright (c) 2010 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 . . . . .	4
1.1. History . . . . .	4
1.2. Overview . . . . .	4
2. Terminology . . . . .	4
3. ALTO Service Deployment . . . . .	5
3.1. ISP-Centric ALTO Service Deployments . . . . .	6
3.2. Cross-domain vs. Localized ALTO Server Discovery . . . . .	7
4. ALTO Service Discovery Scenarios . . . . .	7
4.1. Discovery Metrics . . . . .	7
4.1.1. Discovery Clients . . . . .	7
4.1.2. Service Location . . . . .	8
4.1.3. Layering Perspective . . . . .	8
4.2. Discovery Scenarios . . . . .	9
4.2.1. Local ALTO service discovery by end terminals . . . . .	9
4.2.2. Local ALTO service discovery by application trackers . . . . .	10
5. ALTO Service Discovery Mechanisms . . . . .	11
5.1. ALTO service discovery using Domain Name System (DNS) . . . . .	11
5.1.1. DNS-based ALTO discovery . . . . .	12
5.1.2. Determine Service Name of Local ALTO servers . . . . .	12
5.1.2.1. Using DHCP option for access domain name . . . . .	13
5.1.2.2. Use IANA Database . . . . .	13
5.1.2.3. Reverse DNS lookup . . . . .	14
5.2. DHCP . . . . .	14
5.3. XRD . . . . .	14
5.4. Provisioning . . . . .	15
5.5. Manual Configuration . . . . .	15
5.6. Multicast and broadcast . . . . .	15
5.7. Caching . . . . .	16
6. Security Considerations . . . . .	16
7. IANA Considerations . . . . .	16
8. Acknowledgements . . . . .	16
9. References . . . . .	17
9.1. Normative References . . . . .	17
9.2. Informative References . . . . .	18
Authors' Addresses . . . . .	19



## 1. Introduction

### 1.1. History

This document represents a merge of features from two previous drafts:

(1). draft-wang-alto-discovery-00

(2). draft-song-alto-server-discovery-00

The ALTO service architecture and protocol are currently under discussion and development within the IETF ALTO working group.

Although it is identified in the charter that a discovery mechanism is needed, the preference is to adopt one or more existing mechanisms for ALTO discovery rather than designing a new one. This document is consistent with the ALTO framework[I-D.ietf-alto-protocol], and presents different scenarios and available options based on prior and related discovery mechanisms. This document will be updated to track the progress of the ALTO requirements and solution.

### 1.2. Overview

The ALTO problem statement [RFC5693] describes that in P2P applications or client/server applications, resources or services are often available through multiple replicas and each of those are sometimes provided by different providers. ALTO service gives guidance to a consumer or directory about which resource provider(s) to select, in order to optimize the client's performance or quality of experience while optimizing resource consumption in the underlying network infrastructure.

In order to query the ALTO server, clients must first know one or more ALTO servers that might provide useful information. The purpose of the ALTO discovery mechanism is to find those servers in different network and application scenarios.

Section 3 and Section 4 discuss various scenarios of ALTO service deployment and discovery. Section 5 provides a description of available discovery mechanisms and its application to the ALTO service discovery use case addressing potential issues and consideration for each.

## 2. Terminology

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].

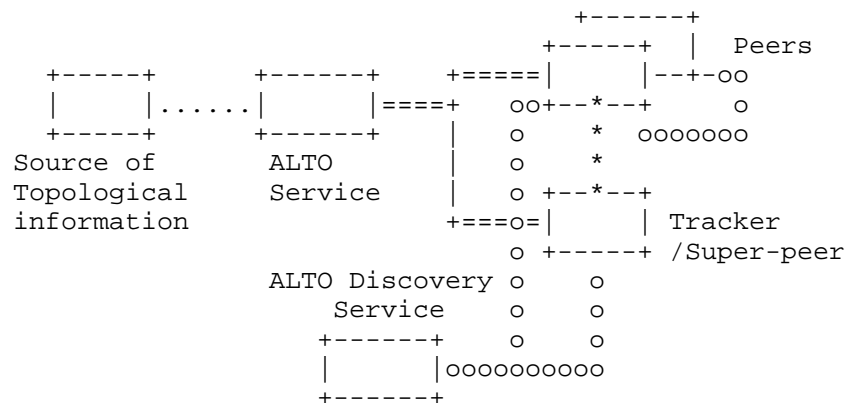
The document uses terms defined in [RFC5693].

In addition to the generic ALTO descriptions, the following terms are used to describe the discovery mechanisms in this document:

- o ALTO Discovery Client: The logical entity discovering the ALTO Service. Depending on the scenario, this could be a Peer or a Super-peer/Tracker.
- o ALTO Discovery Server: The logical entity providing information to locate the ALTO Service. Depending on the discovery mechanism, this could be another Peer or a dedicated entity in the network.
- o ALTO Discovery Domain: The scope of the network handled by a particular ALTO Discovery Server.

### 3. ALTO Service Deployment

This section explores the various dimensions of the ALTO service deployment and access scenarios, and briefly discusses their implications to the discovery mechanisms. Figure 1 below shows a generic ALTO framework diagram with discovery. .



Legend:

=== ALTO query protocol  
 ooo ALTO service discovery protocol  
 \*\*\* Application protocol (out of scope)  
 ... Provisioning or initialization (out of scope)

Figure 1 ALTO Discovery Diagram

### 3.1. ISP-Centric ALTO Service Deployments

(Haibin: we delete the application-centric ALTO service deployment scenario as to keep consistent with the ALTO framework in the working group)

An ALTO Server is the logical entity that provides query interfaces for ALTO Clients. ALTO servers are deployed in an ISP-centric deployment.

A network operator which wants to optimize its traffic, e.g. to reduce its transit traffic volume across the network boundaries; a third party on behalf of one or even several ISPs.

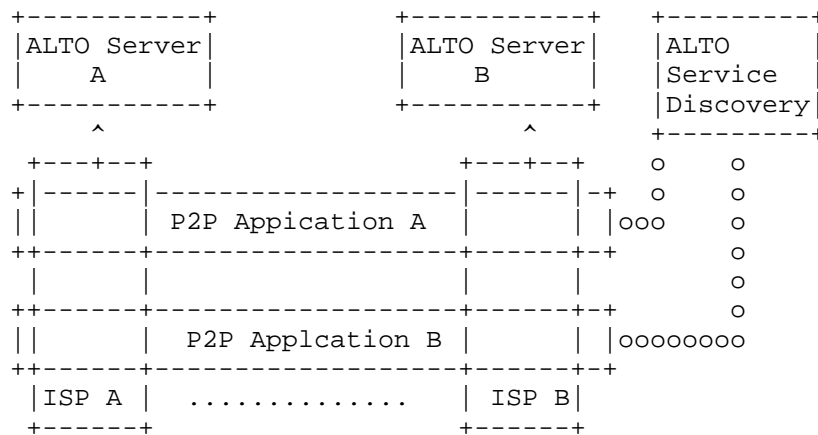


Figure 2: ISP-centric ALTO service deployment example

### 3.2. Cross-domain vs. Localized ALTO Server Discovery

For cross domain scenarios, the ALTO client is embeded in the application tracker or Resource Directory.

There may be several ALTO servers distributed in different operator's networks. Each operator may provide the ALTO service using their own ALTO servers. Each network operator may have its own traffic optimization policy based on his network topology, however it may not know other network operator's policies, nor be clear of other network operator's topologies (e.g. topology hiding). Each of the ALTO servers may have a FQDN.

The ALTO client (e.g. the Tracker) must be able to discover and choose the ALTO server that has the information that is specific to those clients located within that network.

In localized discovery deployments one or several ALTO servers provide the service only to clients in their own network or autonomous system.

#### 4. ALTO Service Discovery Scenarios

### 4.1. Discovery Metrics

#### 4.1.1. Discovery Clients

The ALTO Client can be the Peer in the end-user host or an external entity like a Super-peer or Resource Directory (aka Tracker) on

behalf of the Peer [RFC5693]. If a Super-Peer or Tracker acts as an ALTO Client it needs to know and select the suitable ALTO Service for the Peer being served. Possible mechanisms for third party ALTO discovery have been proposed in[I-D.kiesel-alto-3pdisc]

In a hybrid model the address info of the ALTO Server could be communicated from the Peer to the Super-Peer using the application protocol. It could also be discovered by the Super-Peer from other Peer information received implicitly (like the Peer public IP address) or received explicitly.

There could be scenarios where only the Peer (and not the Super-Peer/Tracker) is able to access the ALTO Service, for example if the ALTO Server is located in a private network. Also the ALTO server might not allow requests from the IP addresses that are out of its administrative domain.

#### 4.1.2. Service Location

The ALTO service is provided by a centralized entity (the ALTO Server) for a given scope. A centralized ALTO Server is implicitly or explicitly assigned to a specific network scope, an out-of-band discovery mechanism is often required.

The ALTO Server for a Peer could be in the same Local Area Network (LAN), within the same ISP Network but not on the same LAN, or in the Global Internet outside the ISP Network. Different network scopes place different constraints on the discovery mechanisms. Multicast discovery generally works within a single LAN only, whereas DNS-based or DHCP-aabased discovery can span multiple subnets within a single ISP or a single network administrative domain. Internet scope discovery usually requires cross-domain indexing or directory services. Note that peers participating in a single P2P application may reside on the same or different ISP networks. Scenarios like this may require hybrid discovery solutions that can adapt to multiple network scopes at the same time. The discovery mechanisms listed in this document should take into account possible limitations of the ALTO service deployment in those network scopes.

#### 4.1.3. Layering Perspective

The discovery process takes place before the first access to the ALTO server. This discovery process could be done at host network initialization time, at application initialization time or just before the first ALTO query is sent.

## 4.2. Discovery Scenarios

The ALTO service discovery scenarios are classified into two types: one is the ALTO server discovery by end terminals, and the other is the ALTO server discovery by application trackers.

### 4.2.1. Local ALTO service discovery by end terminals

In p2p applications without a tracker like DHTs and other conventional client/server applications, an end device needs to discover the local ALTO server by itself.

P2P application which has tracker(s) may also embed the ALTO client within the peers . And the peers can do the remote peer selection after retrieving peer list from the application tracker. Or the peer can send its ALTO server address information to the application tracker, and the application tracker will contact the specific ALTO server and do the peer selection for peers.

After the discovery of an ALTO server, the p2p client can get guidance from the ALTO server directly or through its application tracker.

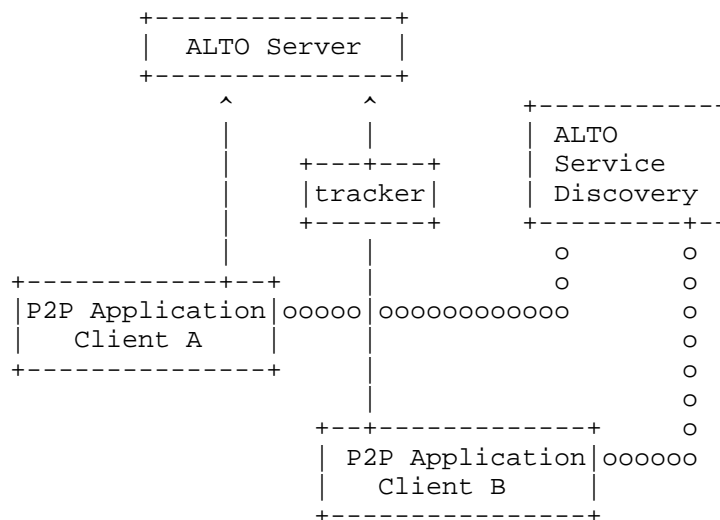


Figure 3: Local ALTO service discovery by end terminals (Example)

#### 4.2.2. Local ALTO service discovery by application trackers

Some p2p applications have trackers, and these applications might not need to have their clients looking for the ALTO server guidance. Trackers query the ALTO servers for guidance, and then return the final ranked result to the application clients. However, application clients are distributed among different network operators and autonomous systems. Trackers must find different ALTO servers for the clients located in different network operators or autonomous systems.

Figure 4 shows an example for a tracker's ALTO server discovery. For client 1, the tracker has not cached yet the mapping between client 1's network operator and its ALTO server address, so it queries the DNS server for the ALTO server address in that operator's domain. And then the tracker interacts with the ALTO server on behalf of client 1 (to get the network map and cost map), finally, the ranked list is sent back to client 1. For client 2, the tracker has cached the mapping between client 2's network operator and its ALTO server address, so it does not need to query the DNS for the address of ALTO server 2. If the Application tracker already has the network map and cost map from ALTO Server2, then it does not to query the ALTO Server for network map and cost map frequently.

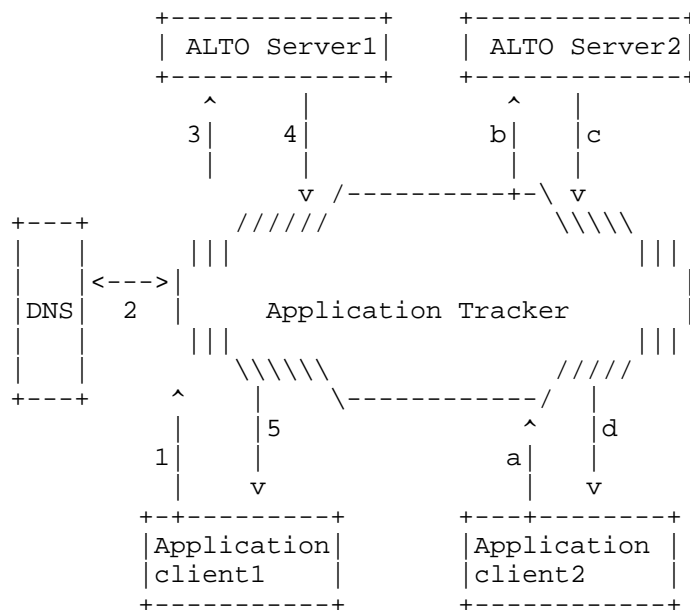


Figure 4: Local ALTO service discovery by application trackers (Example)

## 5. ALTO Service Discovery Mechanisms

One ALTO client should use one or several of the introduced discovery mechanisms according to its application scenario until it finally finds an appropriate ALTO server.

The following issues should be considered when designing the ALTO service discovery mechanism.

**Load Balance:** When more than one ALTO server provide identical service for the same area, we must find a mechanism to balance the processing load between the ALTO servers;

**Well known port:** If ALTO server provides service through a well known port, then the discovery mechanism only needs to discover the IP address of an ALTO server that can provide service for a client, otherwise, the discovery mechanism must discover both IP address and port number through which the ALTO server provides the service.

**Note:** It will depend on the ALTO protocol whether a well known port is used for the ALTO server. If there is no well known port for the ALTO server, we need to discover the port information with the discovery process.

**IP address change:** The IP address of the ALTO server may change in some circumstances. The ALTO service discovery mechanism must be well adaptable to this case when necessary.

**Mobile Scenarios:** When the end terminals are mobile equipments, the data traffic may route via a roaming client's home agent's router to the client, or route to the client directly. Which ALTO server to choose should depend on the routing optimization mode adopted for mobility. If the data traffic routes via the client's home agent, it should choose the ALTO server that serves its home area network, otherwise, it should choose the ALTO server that serves its current network.

### 5.1. ALTO service discovery using Domain Name System (DNS)

DNS is widely used on the Internet to discover the server address for applications. ALTO service is a conventional client/server mode service, which can use DNS lookup for its service discovery.

NAPTR [RFC2915] and SRV [RFC2782] DNS resource records are appropriate to provide service discovery mechanisms. The concrete application of these resource records depends on the final ALTO requests/response protocol. The use of NAPTR or SRV records is a



trade-off between flexibility and simplicity. S-NAPTR [RFC3958] and U-NAPTR [RFC4848] mechanisms provide a Dynamic Delegation Discovery System (DDDS) Application to map domain name, service name and protocol name to a target host and port or to a target URI. SRV records provide a mechanism to map domain name, service name and transport protocol name to a target host and port. The use of a NAPTR or SRV solution is open to discussion and depends on the requirements of the ALTO protocol. Next section will assume the use of SRV records.

#### 5.1.1.1. DNS-based ALTO discovery

Figure 5. shows a general DNS ALTO server discovery mechanism. A server must register its SRV resource record with a well known service name (e.g. `_ALTO._TCP.example.com`) in the DNS system. The service name in this document refers to the name used for DNS SRV query, which includes the service label, protocol name (TCP or UDP) and domain name. Any ALTO client that wants to get the IP address and port of the ALTO server sends a DNS SRV query to the DNS server in that domain .

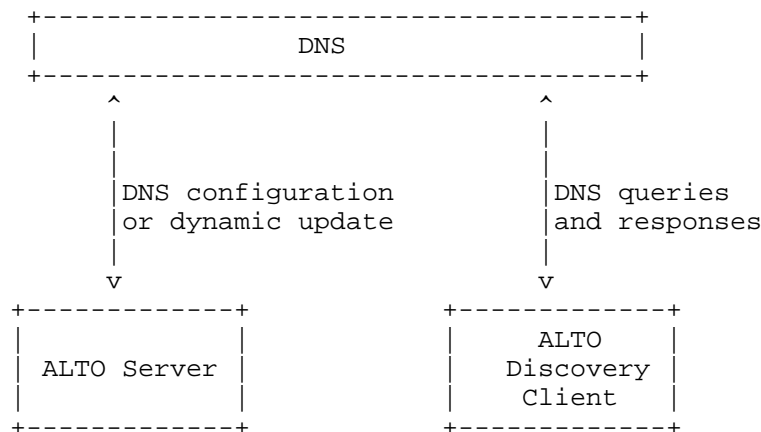


Figure 5: DNS query for well known ALTO servers

#### 5.1.1.2. Determine Service Name of Local ALTO servers

An ALTO discovery client must know its ALTO service name for it before sending a query to the DNS system. Some ALTO servers may provide service to the overall network, they may have well-known service name. But in most cases, one ALTO server will only provide service to its own local access network or autonomous system. There will be multiple ALTO servers in the overall network. An ALTO discovery client needs to find the service name of its local ALTO

server.

#### 5.1.2.1. Using DHCP option for access domain name

There are DHCP options (OPTION\_V4\_ACCESS\_DOMAIN and OPTION\_V6\_ACCESS\_DOMAIN) proposed in [I-D.ietf-geopriv-lis-discovery] to discover the local access domain names. The retrieved access domain name can be used to form a SRV name by prefixing the ALTO service label to the access domain name. If it failed with the SRV lookup with this service name, then it will remove one tag from the left hand of the access domain name and prefix the ALTO service label to form a new SRV name. It will iterate the process until it succeeds in getting an ALTO server information or failed.

It should be noticed that there are many residential gateways (RG) which can act as DHCP servers themselves. RG becomes a hindrance between the end terminals and the ALTO service provider's DHCP server if we use DHCP. It should not depend on the update of all these RGs to support this new DHCP Option for ALTO server discovery. A DHCP Container Option [I-D.ietf-dhc-container-opt] for server configuration should be used here. With the Container Option, the DHCP server for the consumer domain (e.g. RGs) can just pass the server configuration to the end terminals without explicit knowledge of the DHCP options contained in the Container. The DHCP Option for the access domain name could be contained in the Container Option.

#### 5.1.2.2. Use IANA Database

The service name of a client's local ALTO server could be formed by adding the service and protocol label before its domain information. IANA and its subsidiary organizations (e.g. APNIC) database can be used to lookup the physical domain of a client through its public IP address, i.e. which network operator and/or autonomous system the client belongs to. The WHOIS service [WWW.WHOIS] on the Internet is also available for this purpose. This mechanism requires ISPs assign the domain names to their ALTO servers according to the AS and ISP information (e.g. they have a rule to format the domain name, AS.ISP.COM), then you can rebuild the domain name with the information retrieved from WHOIS. Otherwise, you can't.

However, the mapping information may be changed due to the business deals and network adjustment. For example, an ISP could sell some part of its network (include all equipments, IP addresses, AS number, and so on) to another ISP, and the ISP does not have the responsibility to notify the IANA, and then the information in the IANA database is wrong.

#### 5.1.2.3. Reverse DNS lookup

BEP 22 [BEP-22] framework uses reverse DNS lookup to determine the domain name of a client through its public address. And then use service label and the domain name to lookup the local server in DNS. The following limitations should be considered when use this mechanism.

(1) This method assumes that the access network provider also provides the reverse DNS record and they control the domain that is indicated in the "PTR" record. (In most cases it is true, but not always)

(2) Furthermore, this method might not apply where a host is given a domain name that is different from the domain name of the access network.

(3) In case of NAT and a public ALTO server, it requires the ALTO client to know its public IP address.

The advantage is that it doesn't require any update/configuration/change in the DHCP servers of any residential gateway.

#### 5.2. DHCP

There are other ways using DHCP to locate an ALTO server. One suggestion is to use DHCP to obtain the ALTO server IP address and port information directly. New DHCP options are needed for this purpose. The residential gateways consideration for DHCP option must be considered as described in . (Section 5.1.2.1)

With this mechanism, the DHCP server needs to support load balance if there are more than one ALTO servers for this access domain. The maintenance is costly when the address of ALTO server changes.

#### 5.3. XRD

XRD is described in [XEP-1.0]. In order to begin the XRD discovery you need the URL (or XRI) of the resource you want to discover links/services related to. In other XRD use cases like OpenId or OpenSocial, it is clear that you know that URL (the OpenId url of the user, or the url of the OS container). But in case of ALTO Server Discovery, the obtainment of this initial URL also needs to use some discovery framework.

#### 5.4. Provisioning

A network operator can simply provide a configuration file that contains the ALTO server address for its clients, provided that there are only one or a few ALTO servers which provide identical service for its network. An application can also provide such a configuration file containing the ALTO server address if an existing ALTO server provides identical service to the overall network.

#### 5.5. Manual Configuration

Manual configuration of the ALTO service location(s) could work in a single ISP network scope, but is not scalable when multiple ISPs or cross-domain ALTO services are required. P2P applications often connect peers from ISPs that they may not have contacted before, and manual configuration will not work without any prior knowledge of the ALTO servers.

#### 5.6. Multicast and broadcast

Multicast or broadcast MAY be used in some scenarios for ALTO discovery.

IP-multicast-based discovery generally works in two ways:

1. Clients send out multicast discovery requests and listen for responses (usually unicast) from available servers or service providers.
2. Servers or service providers send out multicast announcements when they become available or periodically, and clients wait for the next available multicast announcement to identify the servers or service providers.

The on-demand requests and periodic announcements are not mutually exclusive. An implementation can choose to utilize both simultaneously. The configuration effort of multicast discovery is fairly straightforward, only the multicast address and port are needed. Service types and additional information are often encoded in the requests or announcements messages, enabling the same multicast channel to support discovery of different resources or services. There are two main constraints of multicast-based discovery - scopes and flooding messages. Routers disable multicast forwarding by default, making it practically a single-subnet solution. Some forms of discovery proxies are needed to extend the scope of multicast discovery to multiple subnets. The second issue is the flooding of multicast messages to all hosts on the same subnet. The total bandwidth consumed by multicast depends on the

arrival rate the client application requests, and/or the frequency of the service announcements. Older generations of 802.11-based wireless access points often slow down the transmission of multicast messages or generally have a higher packet loss rate for those, causing some multicast discovery implementation to automatically re-send multicast requests or announcements by default. This mitigation further increases the amount of flooding messages on the LAN. Examples of multicast-based discovery include [I-D.cheshire-dnsext-multicastdns], [I-D.cai-ssdp-v1], [WSD], SLP [RFC2165], and LLMNR [RFC4795].

### 5.7. Caching

Once a client has located an ALTO server for the first time, it can cache it for use as future ALTO server. There are implications in case of mobility of devices.

## 6. Security Considerations

As this document mainly proposes to use DNS and DHCP for ALTO service discovery, it will depend on the DHCP security and DNS security for this service discovery.

## 7. IANA Considerations

The service label for the ALTO service will depend on the final protocol name for application-layer traffic optimization(TBD).

## 8. Acknowledgements

The authors would like to give special thanks to Roni Even for his continuous contribution to this document.

We would also like to thank the following experts for their contribution.

Sebastian Kiesel

Yunfei Zhang

Y. Richard Yang

Xingfeng Jiang

Jay Gu

Ning Zong

David Bryan

Enrico Marocco

## 9. References

### 9.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC2782] Gulbrandsen, A., Vixie, P., and L. Esibov, "A DNS RR for specifying the location of services (DNS SRV)", RFC 2782, February 2000.
- [RFC3958] Daigle, L. and A. Newton, "Domain-Based Application Service Location Using SRV RRs and the Dynamic Delegation Discovery Service (DDDS)", RFC 3958, January 2005.
- [RFC4848] Daigle, L., "Domain-Based Application Service Location Using URIs and the Dynamic Delegation Discovery Service (DDDS)", RFC 4848, April 2007.
- [RFC5693] Seedorf, J. and E. Burger, "Application-Layer Traffic Optimization (ALTO) Problem Statement", RFC 5693, October 2009.
- [RFC2915] Mealling, M. and R. Daniel, "The Naming Authority Pointer (NAPTR) DNS Resource Record", RFC 2915, September 2000.
- [I-D.ietf-alto-protocol]  
Alimi, R., Penno, R., and Y. Yang, "ALTO Protocol",  
draft-ietf-alto-protocol-04 (work in progress), May 2010.
- [I-D.ietf-geopriv-lis-discovery]  
Thomson, M. and J. Winterbottom, "Discovering the Local Location Information Server (LIS)",  
draft-ietf-geopriv-lis-discovery-15 (work in progress),  
March 2010.
- [I-D.ietf-dhc-container-opt]  
Droms, R., "Container Option for Server Configuration",  
draft-ietf-dhc-container-opt-05 (work in progress),

March 2009.

## 9.2. Informative References

- [RFC2629] Rose, M., "Writing I-Ds and RFCs using XML", RFC 2629, June 1999.
- [RFC3552] Rescorla, E. and B. Korver, "Guidelines for Writing RFC Text on Security Considerations", BCP 72, RFC 3552, July 2003.
- [RFC2165] Veizades, J., Guttman, E., Perkins, C., and S. Kaplan, "Service Location Protocol", RFC 2165, June 1997.
- [RFC4795] Aboba, B., Thaler, D., and L. Esibov, "Link-local Multicast Name Resolution (LLMNR)", RFC 4795, January 2007.
- [I-D.kiesel-alto-3pdisc]  
Kiesel, S., Tomsu, M., Schwan, N., Scharf, M., and M. Stiemerling, "Third-party ALTO server discovery", draft-kiesel-alto-3pdisc-03 (work in progress), July 2010.
- [I-D.cheshire-dnsext-multicastdns]  
Cheshire, S. and M. Krochmal, "Multicast DNS", draft-cheshire-dnsext-multicastdns-11 (work in progress), March 2010.
- [I-D.wang-alto-p4p-specification]  
Wang, Y., Alimi, R., Pasko, D., Popkin, L., and Y. Yang, "P4P Protocol Specification", draft-wang-alto-p4p-specification-00 (work in progress), March 2009.
- [I-D.narten-iana-considerations-rfc2434bis]  
Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", draft-narten-iana-considerations-rfc2434bis-09 (work in progress), March 2008.
- [I-D.cai-ssdp-v1]  
Goland, Y., Cai, T., Leach, P., Gu, Y., and S. Albright, "Simple Service Discovery Protocol/1.0 Operating without an Arbiter", October 1999, <draft-cai-ssdp-v1-03>.
- [WWW.WHOIS]  
"http://www.whois.net".

- [BEP-22] Harrison, D., Shalunov, S., and G. Hazel, "BitTorrent Local Tracker Discovery Protocol", October 2008, <[http://bittorrent.org/beps/bep\\_0022.html](http://bittorrent.org/beps/bep_0022.html)>.
- [XEP-1.0] Hammer-Lahav, E., "Extensible Resource Descriptor (XRD) Version 1.0", May 2009, <<http://www.oasis-open.org/committees/download.php/32686/xrd-1.0-wd01.html>>.
- [WSD] Beatty, J., "Web Services Dynamic Discovery (WS-Discovery)", April 2005, <<http://specs.xmlsoap.org/ws/2005/04/discovery/ws-discovery.pdf>>.

## Authors' Addresses

Haibin Song  
Huawei

Email: [melodysong@huawei.com](mailto:melodysong@huawei.com)

Marco Tomsu  
Alcatel-lucent Bell Labs  
Lorenzstrasse 10  
70435 Stuttgart  
Germany

Email: [marco.tomsu@alcatel-lucent.com](mailto:marco.tomsu@alcatel-lucent.com)  
URI: [www.alcatel-lucent.com/bell-labs](http://www.alcatel-lucent.com/bell-labs)

Gustavo Garcia  
Telefonica I+D  
Emilio Vargas  
Madrid, Madrid  
Spain

Phone: +34 913129826  
Email: [ggb@tid.es](mailto:ggb@tid.es)



Yu-Shun Wang  
Microsoft Corp.  
One Microsoft Way  
Redmond, WA 98052  
USA

Email: [yu-shun.wang@microsoft.com](mailto:yu-shun.wang@microsoft.com)

Victor Pascual  
Consultant

Email: [victor.pascual.avila@gmail.com](mailto:victor.pascual.avila@gmail.com)



ALTO  
Internet-Draft  
Intended status: Informational  
Expires: April 28, 2011

M. Stiemerling  
NEC Europe Ltd.  
S. Kiesel  
University of Stuttgart  
October 25, 2010

ALTO Deployment Considerations  
draft-stiemerling-alto-deployments-05

Abstract

Many Internet applications are used to access resources, such as pieces of information or server processes, which 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 these applications, which have to select one or several hosts from a set of candidates, that are able to provide a desired resource. The protocol is under specification in the ALTO working group. This memo discusses deployment related issues of ALTO for peer-to-peer and CDNs, some preliminary security considerations, and also initial guidance for application designers using ALTO.

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 April 28, 2011.

Copyright Notice

Copyright (c) 2010 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. Overview . . . . .	4
2.1. General Placement of ALTO . . . . .	4
2.2. Provided Guidance . . . . .	6
2.2.1. Keeping Traffic Local in Network . . . . .	6
2.2.2. Off-Loading Traffic from Network . . . . .	7
2.2.3. Intra-Network Localization/Bottleneck Off-Loading . . . . .	8
3. Using ALTO for Peer-to-Peer . . . . .	11
3.1. Using ALTO for Tracker-based Peer-to-Peer Applications . . . . .	13
3.2. Expectations of ALTO . . . . .	15
4. Using ALTO for CDNs . . . . .	16
5. Cascading ALTO Servers . . . . .	17
6. Known Limitations of ALTO . . . . .	19
6.1. Limitations of Map-based Approaches . . . . .	19
6.2. Limitations of Non-Map-based Approaches . . . . .	20
6.3. General Challenges . . . . .	20
7. API between ALTO Client and Application . . . . .	22
8. Security Considerations . . . . .	23
8.1. Information Leakage from the ALTO Server . . . . .	23
8.2. ALTO Server Access . . . . .	23
8.3. Faking ALTO Guidance . . . . .	24
9. Conclusion . . . . .	25
10. References . . . . .	26
10.1. Normative References . . . . .	26
10.2. Informative References . . . . .	26
Appendix A. Acknowledgments . . . . .	28
Authors' Addresses . . . . .	29

## 1. Introduction

Many Internet applications are used to access resources, such as pieces of information or server processes, which are available in several equivalent replicas on different hosts. This includes, but is not limited to, peer-to-peer file sharing applications and Content Delivery Networks (CDNs). The goal of Application-Layer Traffic Optimization (ALTO) is to provide guidance to applications, which have to select one or several hosts from a set of candidates, that are able to provide a desired resource. The basic ideas of ALTO are described in the problem space of ALTO is described in [RFC5693] and the set of requirements is discussed in [I-D.ietf-alto-reqs].

However, there are no considerations about what operational issues are to be expected once ALTO will be deployed. This includes, but is not limited to, location of the ALTO server, imposed load to the ALTO server, or from whom the queries are performed.

Comments and discussions about this memo should be directed to the ALTO working group: [alto@ietf.org](mailto:alto@ietf.org).

## 2. Overview

The ALTO protocol is a client/server protocol, operating between a number of ALTO clients and an ALTO server, as sketched in Figure 1. The ALTO working groups defines the ALTO protocol [I-D.ietf-alto-protocol].

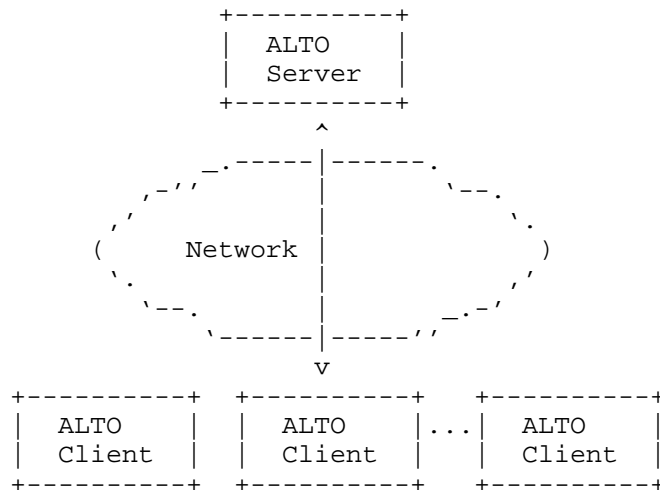


Figure 1: Network Overview of ALTO Protocol

### 2.1. General Placement of ALTO

The ALTO server and ALTO clients can be situated at various entities in a network deployment. The first differentiation is whether the ALTO client is located on the actual host that runs the application, as shown in Figure 2, (e.g., peer-to-peer filesharing application) or if the ALTO client is located on resource directory, as shown in Figure 3 (e.g., a tracker in peer-to-peer filesharing).

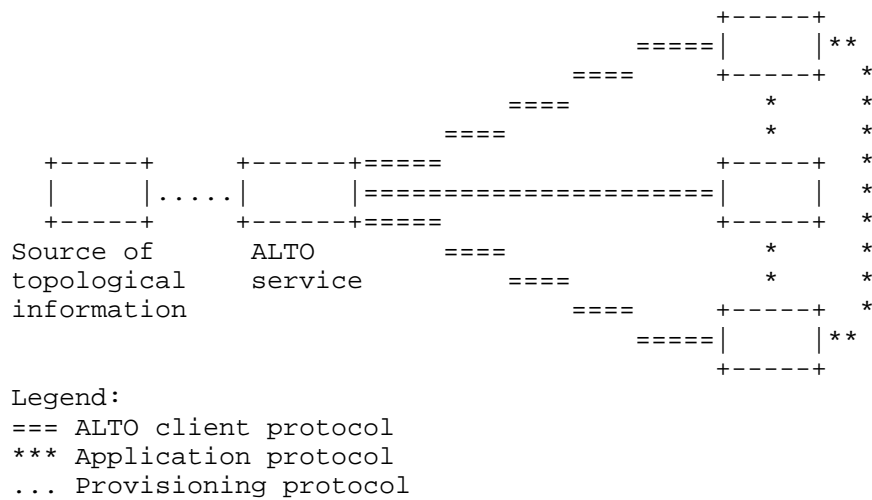


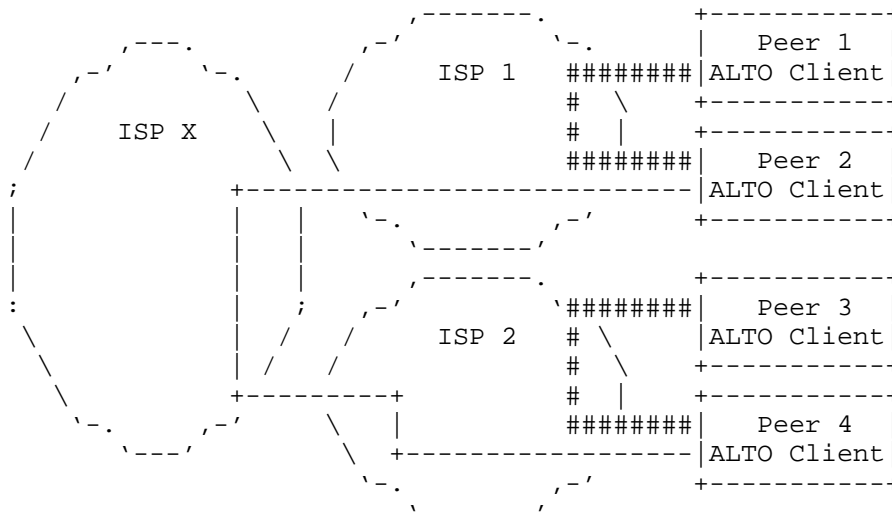
Figure 2: Overview of protocol interaction between ALTO elements, scenario without tracker

Figure 2 shows the operational model for applications that do not use a tracker, such as, edonky, or in if the tracker should be the querying party. This use case also holds true for CDNs. The ALTO server can also be queried by CDNs to get a guidance about where the a particular client accessing data in the CDN is exactly located in the ISP’s network.





the same network (e.g., Peer 1 and Peer 2 in ISP1 and Peer 3 and Peer 4 in ISP2).



Legend:

### preferred "connections"

--- non-preferred "connections"

Figure 4: ALTO Traffic Network Localization

TBD: Describes limits of this approach (e.g., traffic localization guidance is of less use if the peers cannot upload); describe how maps would look like.

#### 2.2.2. Off-Loading Traffic from Network

Another scenario where the use of ALTO can be beneficial is in mobile broadband networks, e.g., CDMA200 or UMTS, but where the network operator may have the desire to guide peers in its own network to use peers in remote networks. One reason can be that the wireless network is not made for the load cause by, e.g., peer-to-peer applications, and the operator has the need that peers fetch their data from remote peers in other parts of the Internet.

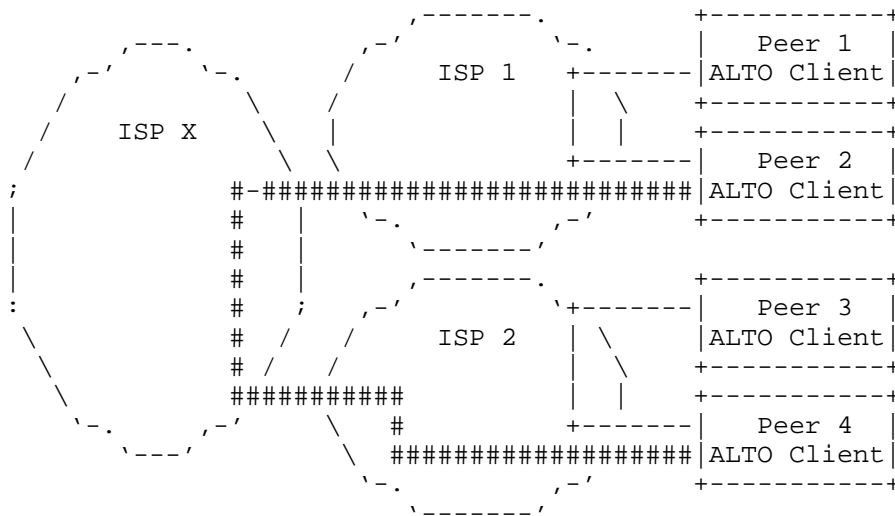


Figure 5: ALTO Traffic Network De-Localization

Figure 5 shows the result of such a guidance process where Peer 2 prefers a connection with Peer4 instead of Peer 1, as shown in Figure 4.

TBD: Limits of this approach in general and with respect to p2p. describe how maps would look like.

### 2.2.3. Intra-Network Localization/Bottleneck Off-Loading

The above sections described the results of the ALTO guidance on an inter-network level. However, ALTO can also be used to guide peers on which internal peers are to be preferred. For instance, to guide Peers on a remote network side to prefer to connect to each other, instead of crossing a bottleneck link, a backhaul link to connect the side to the network core. Figure 6 shows such a scenario where Peer 1 and Peer 2 are located in Net 2 of ISP1 and connect via a low capacity link to the core (Net 1) of the same ISP1. Peer1 and Peer 2 would both exchange their data with remote peers, probably clogging the bottleneck link.

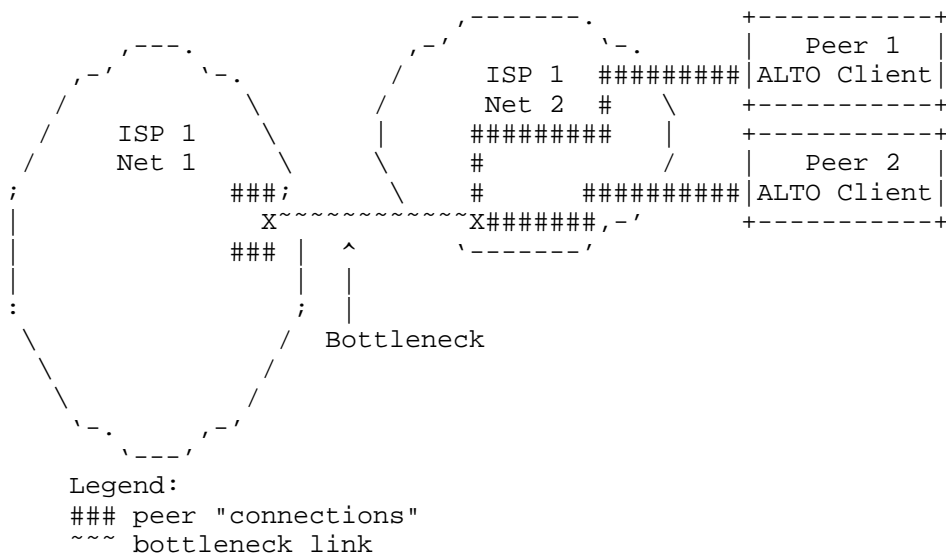


Figure 6: Without Intra-Network ALTO Traffic Localization

The operator can guide the peers in such a situation to try first local peers in the same network islands, avoiding or at least lowering the effect on the bottleneck link, as shown in Figure 7.

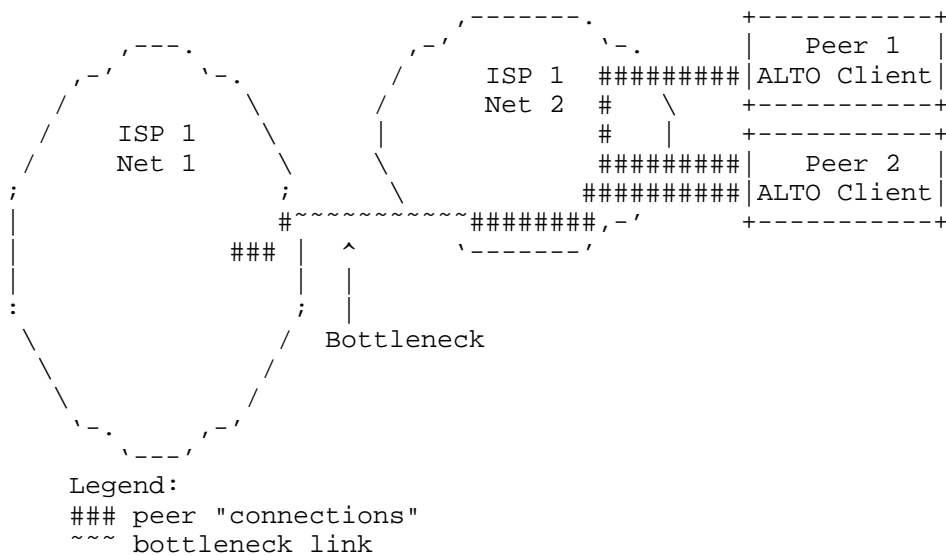


Figure 7: With Intra-Network ALTO Traffic Localization

TBD: describe how maps would look like.

## 3. Using ALTO for Peer-to-Peer

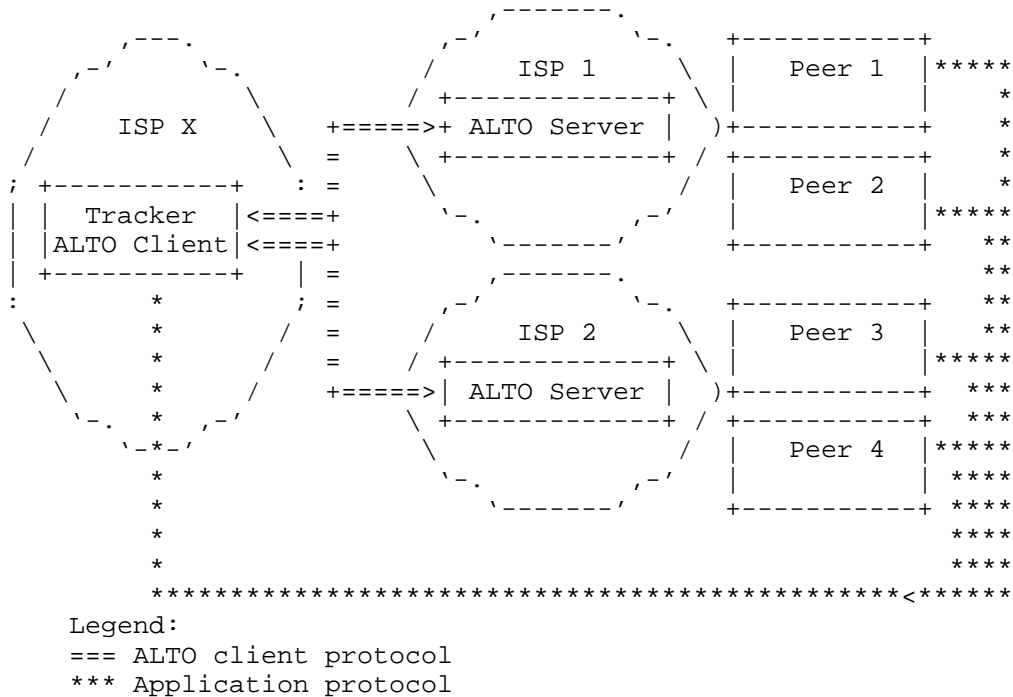


Figure 8: Global tracker accessing ALTO server at various ISPs

Figure 8 depicts a tracker-based system, where the tracker embeds the ALTO client. The tracker itself is hosted and operated by an entity different than the ISP hosting and operating the ALTO server. Initially, the tracker has to look-up the ALTO server in charge for each peer where it receives a ALTO query for. Therefore, the ALTO server has to discover the handling ALTO server, as described in [I-D.kiesel-alto-3pdisc]. However, the peers do not have any way to query the server themselves. This setting allows to give the peers a better selection of candidate peers for their operation at an initial time, but does not consider peers learned through direct peer-to-peer knowledge exchange, AKA peer exchange in various peer-to-peer protocols.

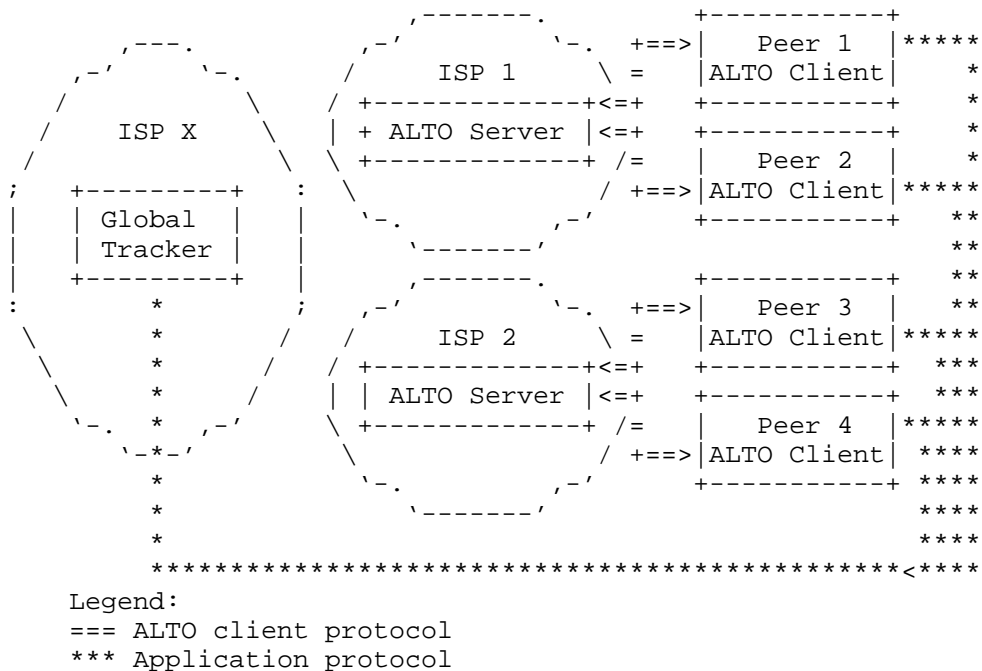


Figure 9: Global Tracker - Local ALTO Servers

The scenario in Figure 9 lets the peers directly communicate with their ISP’s ALTO server (i.e., ALTO client embedded in the peers), giving thus the peers the most control on which information they query for, as they can integrate information received from trackers and through direct peer-to-peer knowledge exchange.

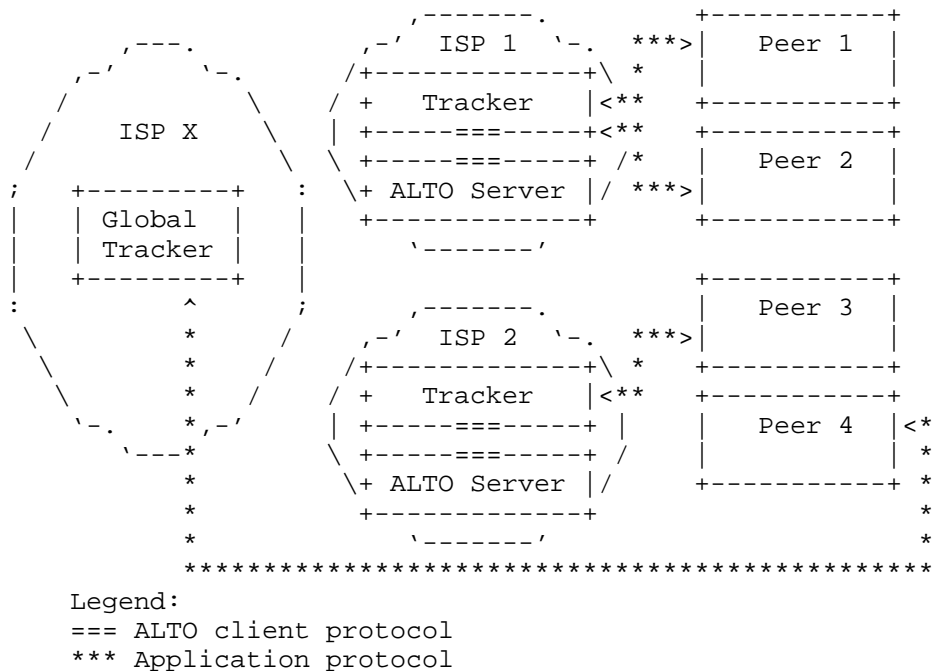


Figure 10: P4P approach with local tracker and local ALTO server

There are some attempts to let ISP's to deploy their own trackers, as shown in Figure 10. In this case, the client has no chance to get guidance from the ALTO server, other than talking to the ISP's tracker. However, the peers would have still chance the contact other trackers, deployed by entities other than the peer's ISP.

Figure 10 and Figure 8 ostensibly take peers the possibility to directly query the ALTO server, if the communication with the ALTO server is not permitted for any reason. However, considering the plethora of different applications of ALTO, e.g., multiple tracker and non-tracker based P2P systems and or applications searching for relays, it seems to be beneficial for all participants to let the peers directly query the ALTO server. The peers are also the single point having all operational knowledge to decide whether to use the ALTO guidance and how to use the ALTO guidance. This is a preference for the scenario depicted in Figure Figure 9.

### 3.1. Using ALTO for Tracker-based Peer-to-Peer Applications

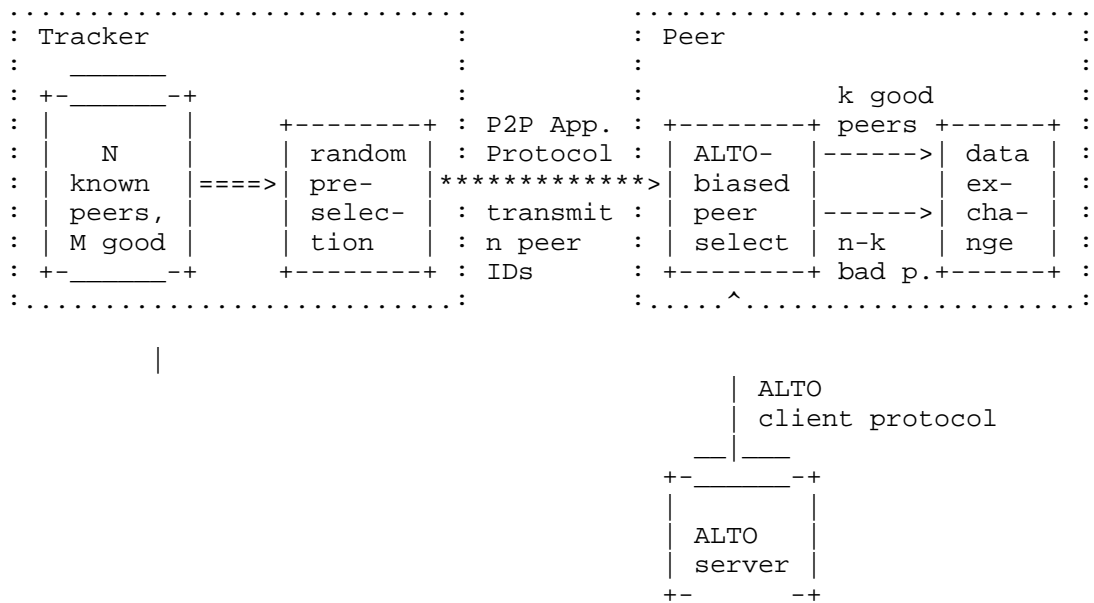


Figure 11: Tracker-based P2P Application with random peer preselection

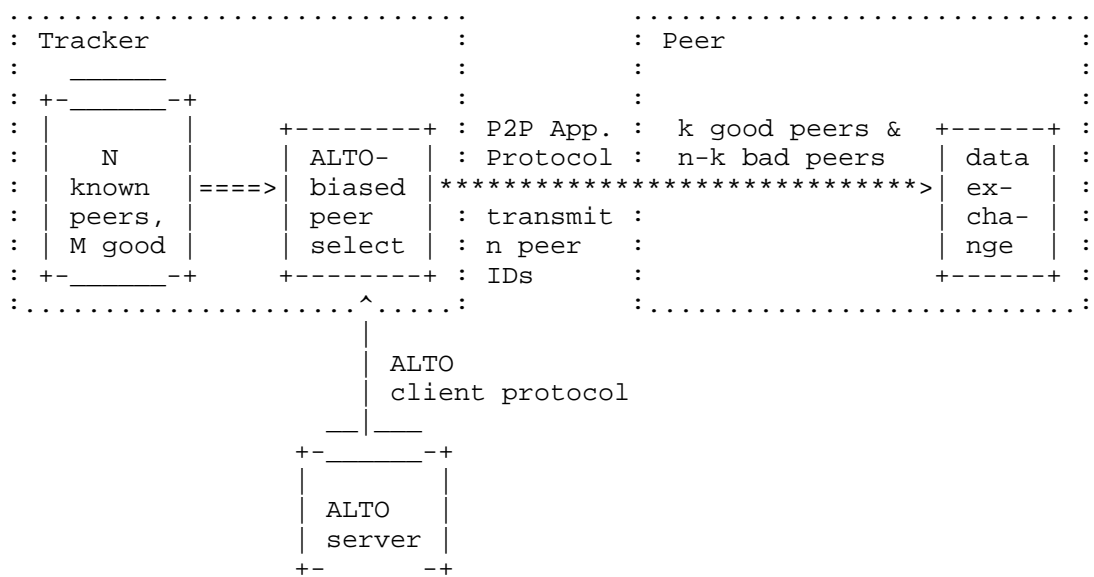


Figure 12: Tracker-based P2P Application with ALTO client in tracker



TBD: explain why Figure 12 usually will yield better results wrt. peer selection than Figure 11.

### 3.2. Expectations of ALTO

This section hints to some recent experiments conducted with ALTO-like deployments in Internet Service Provider (ISP) network's. NTT performed tests with their HINT server implementation and dummy nodes to gain insight on how an ALTO-like service influence a peer-to-peer systems [I-D.kamei-p2p-experiments-japan]. The results of an early experiment conducted in the Comcast network are documented here[RFC5632]

#### 4. Using ALTO for CDNs

Section 3 discussed the placement and usage of ALTO for P2P systems, but not beyond. This section discusses the usage of ALTO for Content Delivery Networks (CDNs). CDNs are used to bring a service (e.g., a web page, videos, etc) closer to the location of the user - where close refers to shorten the distance between the client and the server in the IP topology. CDNs use several techniques to decide which server is closest to a client requesting a service. One common way to do so, is relying on the DNS system, but there are many other ways, see [RFC3568].

The general issue for CDNs, independent of DNS or HTTP Redirect based approaches (see, for instance, [I-D.penno-alto-cdn]), is that the CDN logic has to match the client's IP address with the closest CDN cache. This matching is not trivial, for instance, in DNS based approaches, where the IP address of the DNS original requester is unknown (see [I-D.vandergaast-edns-client-ip] for a discussion of this and a solution approach).

## 5. Cascading ALTO Servers

The main assumptions of ALTO seems to be each ISP operates its own ALTO server independently, irrespectively of the ISP's situation. This may true for most envisioned deployments of ALTO but there are certain deployments that may have different settings. Figure 13 shows such setting, were for example, a university network is connected to two upstream providers. ISP2 if the national research network and ISP1 is a commercial upstream provider to this university network. The university, as well as ISP1, are operating their own ALTO server. The ALTO clients, located on the peers will contact the ALTO server located at the university.

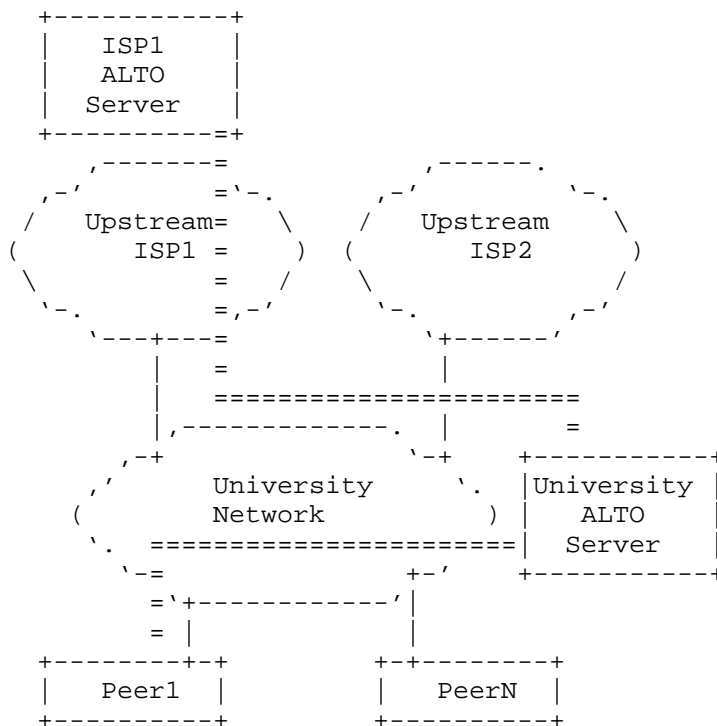


Figure 13: Cascaded ALTO Server

In this setting all "destinations" useful for the peers within ISP2 are free-of-charge for the peers located in the university network (i.e., they are preferred in the rating of the ALTO server). However, all traffic that is not towards ISP2 will be handled by the ISP1 upstream provider. Therefore, the ALTO server at the university has also to include the guidance given by the ISP1 ALTO server in its

replies to the ALTO clients. This can be called cascaded ALTO servers.

## 6. Known Limitations of ALTO

This section describes some known limitations of ALTO in general or specific mechanisms in ALTO.

### 6.1. Limitations of Map-based Approaches

The specification of the ALTO protocol [I-D.ietf-alto-protocol] uses, amongst others mechanism, so-called network maps. The network map approach uses Host Group Descriptors that group one or multiple subnetworks (i.e., IP prefixes) to a single Host Group Descriptor. A set of IP prefixes is called partition and the associated Host Group Descriptor is called partition ID. The "costs" between the various partition IDs is stored in a second map, the cost map. Map-based approaches are chosen as they lower the signaling load on the server, as the maps have only to be retrieved if they are changed.

The main assumption for map-based approaches is that the information provided in these maps is static for a longer period of time, where this period of time refers to days, but not hours or even minutes. This assumption is fine, as long as the network operator does not change any parameter, e.g., routing within the network and to the upstream peers, IP address assignment stays stable (and thus the mapping to the partitions). However, there are several cases where this assumption is not valid, as:

1. ISPs reallocate IPv4 subnets from time to time;
2. ISPs reallocate IPv4 subnets on short notice;
3. IP prefix blocks may be assigned to a single DSLAM which serves a variety of access networks.

For 1): ISPs reallocate IPv4 subnets within their infrastructure from time to time, partly to ensure the efficient usage of IPv4 addresses (a scarce resource), and partly to enable efficient route tables within their network routers. The frequency of these "renumbering events" depend on the growth in number of subscribers and the availability of address space within the ISP. As a result, a subscriber's household device could retain an IPv4 address for as short as a few minutes, or for months at a time or even longer.

Some folks have suggested that ISPs providing ALTO services could sub-divide their subscribers' devices into different IPv4 subnets (or certain IPv4 address ranges) based on the purchased service tier, as well as based on the location in the network topology. The problem is that this sub-allocation of IPv4 subnets tends to decrease the efficiency of IPv4 address allocation. A growing ISP

that needs to maintain high efficiency of IPv4 address utilization may be reluctant to jeopardize their future acquisition of IPv4 address space.

However, this is not an issue for map-based approaches if changes are applied in the order of days.

For 2): ISPs can use techniques, such as ODAP (XXX) that allow the reallocation of IP prefixes on very short notice, i.e., within minutes. An IP prefix that has no IP address assignment to a host anymore can be reallocate to areas where there is currently a high demand for IP addresses.

For 3): In DSL-based access networks, IP prefixes are assigned to DSLAMs which are the first IP-hop in the access-network between the CPE and the Internet. The access-network between CPE and DSLAM (called aggregation network) can have varying characteristics (and thus associated costs), but still using the same IP prefix. For instance one IP addresses IP11 out of a IP prefix IP1 can be assigned to a VDSL (e.g., 2 MBit/s uplink) access-line while the subsequent IP address IP12 is assigned to a slow ADSL line (e.g., 128 kbit/s uplink). These IP addresses are assigned on a first come first served basis, i.e., the a single IP address out of the same IP prefix can change its associated costs quite fast. This may not be an issue with respect to the used upstream provider (thus the cross ISP traffic) but depending on the capacity of the aggregation-network this may raise to an issue.

## 6.2. Limitations of Non-Map-based Approaches

The specification of the ALTO protocol [I-D.ietf-alto-protocol] uses, amongst others mechanism, a mechanism called Endpoint Cost Service. ALTO clients can ask guidance for specific IP addresses to the ALTO server. However, asking for IP addresses, asking with long lists of IP addresses, and asking quite frequent may overload the ALTO server. The server has to rank each received IP address which causes load at the server. This may be amplified by the fact that not only a single ALTO client is asking for guidance, but a larger number of them.

Caching of IP addresses at the ALTO client or the usage of the H12 approach [I-D.kiesel-alto-h12] in conjunction with caching may lower the query load on the ALTO server.

## 6.3. General Challenges

An ALTO server stores information about preferences (e.g., a list of preferred autonomous systems, IP ranges, etc) and ALTO clients can retrieve these preferences. However, there are basically two

different approaches on where the preferences are actually processed:

1. The ALTO server has a list of preferences and clients can retrieve this list via the ALTO protocol. This preference list can be partially updated by the server. The actual processing of the data is done on the client and thus there is no data of the client's operation revealed to the ALTO server .
2. The ALTO server has a list of preferences or preferences calculated during runtime and the ALTO client is sending information of its operation (e.g., a list of IP addresses) to the server. The server is using this operational information to determine its preferences and returns these preferences (e.g., a sorted list of the IP addresses) back to the ALTO client.

Approach 1 (we call it H1) has the advantage (seen from the client) that all operational information stays within the client and is not revealed to the provider of the server. On the other hand, does approach 1 require that the provider of the ALTO server, i.e., the network operator, reveals information about its network structure (e.g., AS numbers, IP ranges, topology information in general) to the ALTO client.

Approach 2 (we call it H2) has the advantage (seen from the operator) that all operational information stays with the ALTO server and is not revealed to the ALTO client. On the other hand, does approach 2 require that the clients send their operational information to the server.

Both approaches have their pros and cons and are extensively discussed on the ALTO mailing list. But there is basically a dilemma: Approach 1 is seen as the only working solution by peer-to-peer software vendors and approach 2 is seen as the only working by the network operators. But neither the software vendors nor the operators seem to willing to change their position. However, there is the need to get both sides on board, to come to a solution.

## 7. API between ALTO Client and Application

This sections gives some informational guidance on how the interface between the actual application using the ALTO guidance and the ALTO client can look like.

This is still TBD.



## 8. Security Considerations

The ALTO protocol itself, as well as, the ALTO client and server raise new security issues beyond the one mentioned in [I-D.ietf-alto-protocol] and issues related to message transport over the Internet. For instance, Denial of Service (DoS) is of interest for the ALTO server and also for the ALTO client. A server can get overloaded if too many TCP requests hit the server, or if the query load of the server surpasses the maximum computing capacity. An ALTO client can get overloaded if the responses from the sever are, either intentionally or due to an implementation mistake, too large to be handled by that particular client.

### 8.1. Information Leakage from the ALTO Server

The ALTO server will be provisioned with information about the owning ISP's network and very likely also with information about neighboring ISPs. This information (e.g., network topology, business relations, etc) is consider to be confidential to the ISP and must not be revealed.

The ALTO server will naturally reveal parts of that information in small doses to peers, as the guidance given will depend on the above mentioned information. This is seen beneficial for both parties, i.e., the ISP's and the peer's. However, there is the chance that one or multiple peers are querying an ALTO server with the goal to gather information about network topology or any other data considered confidential or at least sensitive. It is unclear whether this is a real technical security risk or whether this is more a perceived security risk.

### 8.2. ALTO Server Access

Depending on the use case of ALTO, several access restrictions to an ALTO server may or may not apply. For an ALTO server that is solely accessible by peers from the ISP network (as shown in Figure 9), for instance, the source IP address can be used to grant only access from that ISP network to the server. This will "limit" the number of peers able to attack the server to the user's of the ISP (however, including botnet computers).

On the other hand, if the ALTO server has to be accessible by parties not located in the ISP's network (see Figure Figure 8), e.g., by a third-party tracker or by a CDN system outside the ISP's network, the access restrictions have to be more loose. In the extreme case, i.e., no access restrictions, each and every host in the Internet can access the ALTO server. This might no the intention of the ISP, as the server is not only subject to more possible attacks, but also on

the load imposed to the server, i.e., possibly more ALTO clients to serve and thus more work load.

### 8.3. Faking ALTO Guidance

It has not yet been investigated how a faked or wrong ALTO guidance by an ALTO server can impact the operation of the network and also the peers.

Here is a list of examples how the ALTO guidance could be faked and what possible consequences may arise:

**Sorting** An attacker could change to sorting order of the ALTO guidance (given that the order is of importance, otherwise the ranking mechanism is of interest), i.e., declaring peers located outside the ISP as peers to be preferred. This will not pose a big risk to the network or peers, as it would mimic the "regular" peer operation without traffic localization, apart from the communication/processing overhead for ALTO. However, it could mean that ALTO is reaching the opposite goal of shuffling more data across ISP boundaries, incurring more costs for the ISP.

**Preference of a single peer** A single IP address (thus a peer) could be marked as to be preferred all over other peers. This peer can be located within the local ISP or also in other parts of the Internet (e.g., a web server). This could lead to the case that quite a number of peers try to contact this IP address, possibly causing a Denial of Service (DoS) attack.

This section is solely giving a first shot on security issues related to ALTO deployments.

## 9. Conclusion

This is the first version of the deployment considerations and for sure the considerations are yet incomplete and imprecise.

## 10. References

### 10.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC3568] Barbir, A., Cain, B., Nair, R., and O. Spatscheck, "Known Content Network (CN) Request-Routing Mechanisms", RFC 3568, July 2003.

### 10.2. Informative References

- [I-D.ietf-alto-protocol]  
Alimi, R., Penno, R., and Y. Yang, "ALTO Protocol", draft-ietf-alto-protocol-05 (work in progress), July 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-06 (work in progress), October 2010.
- [I-D.kamei-p2p-experiments-japan]  
Kamei, S., Momose, T., Inoue, T., and T. Nishitani, "ALTO-Like Activities and Experiments in P2P Network Experiment Council", draft-kamei-p2p-experiments-japan-03 (work in progress), May 2010.
- [I-D.kiesel-alto-3pdisc]  
Kiesel, S., Tomsu, M., Schwan, N., Scharf, M., and M. Stiemerling, "Third-party ALTO server discovery", draft-kiesel-alto-3pdisc-03 (work in progress), July 2010.
- [I-D.kiesel-alto-h12]  
Kiesel, S. and M. Stiemerling, "ALTO H12", draft-kiesel-alto-h12-02 (work in progress), March 2010.
- [I-D.penno-alto-cdn]  
Penno, R., Raghunath, S., Medved, J., Bakshi, M., Alimi, R., and S. Previdi, "ALTO and Content Delivery Networks", draft-penno-alto-cdn-01 (work in progress), July 2010.
- [I-D.vandergaast-edns-client-ip]  
Contavalli, C., Gaast, W., Leach, S., and D. Rodden, "Client IP information in DNS requests", draft-vandergaast-edns-client-ip-01 (work in progress), May 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.
- [RFC5693] Seedorf, J. and E. Burger, "Application-Layer Traffic Optimization (ALTO) Problem Statement", RFC 5693, October 2009.

## Appendix A. Acknowledgments

Martin Stiernerling is partially supported by the NAPA-WINE project (Network-Aware P2P-TV Application over Wise Networks, <http://www.napa-wine.org>), a research project supported by the European Commission under its 7th Framework Program (contract no. 214412). 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 NAPA-WINE project or the European Commission.

Authors' Addresses

Martin Stiernerling  
NEC Laboratories Europe/University of Goettingen  
Kurfuerstenanlage 36  
Heidelberg 69115  
Germany

Phone: +49 6221 4342 113  
Fax: +49 6221 4342 155  
Email: martin.stiernerling@neclab.eu  
URI: <http://ietf.stiernerling.org>

Sebastian Kiesel  
University of Stuttgart, Computing Center  
Allmandring 30  
Stuttgart 70550  
Germany

Email: [ietf-alto@skiesel.de](mailto:ietf-alto@skiesel.de)





ALTO  
Internet-Draft  
Intended status: Informational  
Expires: April 28, 2011

X. Sun  
China Telecom  
Y. Yang  
Yale University  
October 25, 2010

ALTO Deployment Considerations: Configuration and Monitoring by ISPs  
draft-sun-deployment-01.txt

## Abstract

As ALTO specification continues in the ALTO Working Group and some applications start to conduct integration with ALTO, more ISPs start to evaluate key issues in the deployment of ALTO in their networks. In this document, we discuss key issues that an ISP needs to consider when deploying ALTO. In particular, we discuss issues on how to configure ALTO information as well as how to monitor the effectiveness of ALTO.

## 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 [1].

## Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at  
<http://www.ietf.org/ietf/lid-abstracts.txt>.

The list of Internet-Draft Shadow Directories can be accessed at  
<http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on April 28, 2011.

## Copyright Notice

Copyright (c) 2010 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 BSD License.

## Table of Contents

1. Introduction	3
2. ALTO Server Placement and Configuration	3
2.1. Server Placement	4
2.1.1. Optimization Area	4
2.1.2. Server Load Balancing and Fault Tolerance	4
2.2. Network and Cost Map Configuration	4
2.2.1. Network Map and PID	4
2.2.2. Cost Map	5
3. ALTO Deployment Monitoring	5
3.1. Monitoring Metrics	5
3.1.1. Network Metrics	5
3.1.2. Application Metrics	6
3.2. Monitoring Data Sources	6
3.2.1. Application Log Server	6
3.2.2. P2P Clients	6
3.2.3. OAM	6
3.2.4. DPI	6
3.3. Application/ISP Monitoring Integration	7
3.3.1. Structure	7
3.3.2. ALTO Monitoring Report Protocol	7
4. IANA Considerations	8
5. Security Considerations	8
6. References	8
6.1. Normative References	8
6.2. Informative References	8
Appendix A. Acknowledgments	8
Authors' Addresses	9

## 1. Introduction

A basic service of ALTO is to provide information from network service providers to applications, in order to improve network efficiency and application performance. Some applications start to or have shown interests to conduct integration with ALTO. Some major ISPs (e.g., China Telecom) are in the process of deploying production ALTO services in some of their production networks. As a result, more ISPs start to evaluate key issues in the deployment of ALTO in their networks. Thus, a document highlighting some key issues that an ISP should consider in the deployment process can be a highly valuable reference.

The objective of this document is to provide such a reference. The document will try to draw on many valuable discussions in the ALTO mailing list as well as the predecessor p2pi mailing list. In addition, it will try to draw on the trial experiences of multiple ISPs (e.g., [CTTrial,ComcastTrial]).

The deployment of ALTO involves both ISPs and network applications. We can identify four major issues in ALTO deployment:

1. How does an ISP deploy and configure its ALTO servers?  
Specifically, an ALTO Server provides the Network Map and the Cost Map. How does an ISP configure these maps? Where does an ISP deploy ALTO servers?
2. Which application entities fetch ALTO information?
3. How does an application integrate ALTO information into its decision process?
4. How does an ISP (potentially with collaboration from applications) monitor the deployment of ALTO, so that the ISP can better understand the status as well as the policy impacts of its ALTO deployment?

This document focuses more on the ISP perspective. Therefore, it focuses more on the first and the fourth issues. There are additional deployment documents in the ALTO working group that focus more on the second issue and the third issue. Our document is complementary to these other documents.

## 2. ALTO Server Placement and Configuration

## 2.1. Server Placement

### 2.1.1. Optimization Area

An ISP deploys ALTO service to optimize traffic for a given network area. We define a network area for which traffic need be optimized using the ALTO service as an optimization area. A typical optimization objective of an ISP is to reduce the inbound and outbound traffic across the optimization area, due to the higher cost of such traffic.

An optimization area can be an access network, a MAN, or a larger network consisting of both access works and MANs. An ISP with a relatively small network can define a single optimization area and deploy an ALTO server for the area.

An ISP with a larger network may partition its network into multiple optimization areas. Each optimization area may include one or more MANs. Alternatively, the ISP may choose to use a large optimization area and distribute a group of ALTO servers.

### 2.1.2. Server Load Balancing and Fault Tolerance

## 2.2. Network and Cost Map Configuration

Key components for an ISP to configure when it deploys its ALTO service are the Network Map and Cost Map. They have impacts on both the load and the effectiveness of the service.

### 2.2.1. Network Map and PID

Different ISPs use different technologies to build their infrastructures. Some ISPs have only a relatively small network, focusing mainly on access. On the other hand, some large ISPs have access networks, MANs, and a Core network.

There are tradeoffs when a large ISP defines its Network Map. If the partition of the network in the Network Map is too fine-grained, it may lead to higher complexity and overhead. On the other hand, a too coarse-grained Network Map may lead to suboptimal optimization.

Specifically, first consider an access network, say an ADSL or Ethernet based access network. A BAS server may be deployed to provide access service for its subscribers. Because all subscribers' traffic must be transmitted through the BAS server, one technique is to identify each such access network by one PID. It is generally unnecessary to further divide such access networks. On the other hand, it can be beneficial to combine several such access networks

into a single PID.

A MAN usually consists of several access networks. The MANs are connected to a core network, whose network bandwidth resource may be costly for some networks. Thus, the ISP can define one or several MANS as one PID. It is also possible that the ISP deploys ALTO independently in some MANs.

#### 2.2.2. Cost Map

### 3. ALTO Deployment Monitoring

In addition to providing configuration, an ISP providing ALTO may want to deploy a monitoring infrastructure to assess the benefits of ALTO and adjust its ALTO configuration.

To construct an effective monitoring infrastructure, the ISP should (1) define the performance metrics to be monitored; (2) and identify and deploy devices to collect data to compute the performance metrics. We discuss both below.

#### 3.1. Monitoring Metrics

The monitoring of some performance metrics can be dependent on specific applications, and ALTO can be applied to multiple applications such as P2P and CDN. We focus on P2P applications.

##### 3.1.1. Network Metrics

An ISP may monitor the impacts of ALTO on its network through a set of performance metrics. We enumerate some key metrics. We define the term domain as one or many groups of Endpoints. That is, one domain includes one PID or some PIDs. Endpoints and PID are defined in "draft-ietf-alto-protocol-05".

A specific set of metrics measuring the impacts of ALTO on networks can include the following:

- o Inter-domain ALTO-Integrated Application Traffic (Network metric): This metric includes total cross domain traffic generated by applications that utilize ALTO guidance. This metric evaluates the impacts of ALTO on the inbound and outbound traffic of a domain.
- o Total Inter-domain Traffic (Network metric): This is similar to the preceding but focuses on all of the traffic, ALTO aware or not. One possibility is that some of the reduction of interdomain

traffic by ALTO aware applications may This metric is always used with the preceding and the following metrics.

- o Intra-domain ALTO-Integrated Application Traffic (Network metric).
- o Network hop count (Network metric): This metric provides the average number of hops that traffic traverses inside a domain. ALTO may reduce not only traffic volume but also the hops. The metric can also indirectly reflect some application performance (e.g., latency).

### 3.1.2. Application Metrics

Each specific application can have its specific set of performance metrics. We give one example for file sharing.

- o Application download rate (Application metric): This metric measures application performance directly. Download means inbound traffic to one user. Global average means the average value of all users' download rates in one or more domains.

## 3.2. Monitoring Data Sources

The preceding metrics are derived from data sources. We identify four data sources.

### 3.2.1. Application Log Server

Many P2P applications deploy Log Servers to collect data.

### 3.2.2. P2P Clients

Some P2P applications may not have Log Servers. When available, P2P client logs can provide data.

### 3.2.3. OAM

Many ISPs deploy OAM systems to monitor IP layer traffic. An OAM provides traffic monitoring of every network device in its management area. It provides data such as link physical bandwidth and traffic volumes.

### 3.2.4. DPI

A DPI system can be deployed in an ISPs' network to understand the traffic of specific classes of applications. Different from OAM, A DPI system can provide application specific information.

### 3.3. Application/ISP Monitoring Integration

#### 3.3.1. Structure

As discussed in the preceding section, some data sources are from ISP while some others are from application. When there is a collaboration agreement between the ISP and an application, there can be an integrated monitoring system as shown in the figure below. In particular, an application developer may deploy Monitor Clients to communicate with Monitor Server of the ISP to transmit raw data from the Log Server or P2P clients of the application to the ISP.

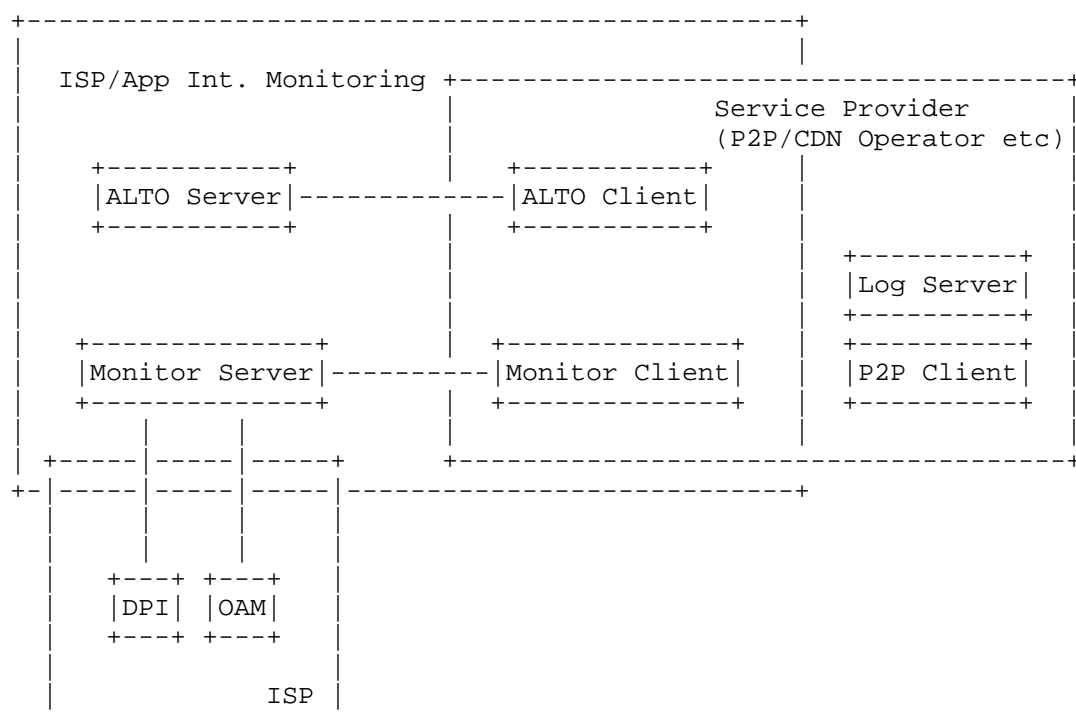


Figure 1

#### 3.3.2. ALTO Monitoring Report Protocol

A potential report message format from the Monitor Client to the Monitor Server can be:

```
HTTP/1.1 200 OK
Content-Length: [TODO]
Content-Type: application/alto
{
  "meta" : {
    "version" : 1,
    "status" : {
      "code" : 1
    }
  },
  "metric1 name" : "value",
  "metric2 name" : "value",
}
```

Figure 2

#### 4. IANA Considerations

This document makes no request of IANA.

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

#### 5. Security Considerations

Multiple documents in the ALTO WG discuss security perspectives. These documents complement this document.

#### 6. References

##### 6.1. Normative References

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

##### 6.2. Informative References

- [2] H. Xie, Y.R. Yang, A. Krishnamurthy, Y. Liu, and A. Silberschatz., "P4P:", In SIGCOMM 2008.

#### Appendix A. Acknowledgments

We thank the discussions with Kai Li.



Authors' Addresses

Xianghui Sun  
China Telecom

Email: alto.deployment@gmail.com

Yang Richard Yang  
Yale University

Email: yry@cs.yale.edu



ALTO WG  
Internet-Draft  
Intended status: Experimental  
Expires: April 20, 2011

Y. Yang  
Yale University  
R. Alimi  
Google  
Y. Wang  
Yale University  
D. Zhang  
PPLive  
K. Lee  
China Telecom  
October 17, 2010

Tracker-Based Peer Selection using ALTO Map Information  
draft-yang-tracker-peer-selection-00.txt

Abstract

As ALTO core information starts to become available from some ISPs, how to effectively utilize such information by P2P applications has become a major issue. In this document, we discuss some techniques that a P2P application tracker can incorporate ALTO information in initial peer selection.

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 [1].

Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/lid-abstracts.txt>.

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on April 20, 2011.

#### Copyright Notice

Copyright (c) 2010 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 BSD License.

## Table of Contents

1. Introduction . . . . .	4
2. Challenges . . . . .	4
3. Peer Classification Data Structures . . . . .	5
3.1. One-Level Key Partitioning . . . . .	6
3.2. Hierarchical Partitioning . . . . .	6
3.3. Comments . . . . .	7
4. Peer Selection Using Peer Classification . . . . .	7
4.1. Overview of Scheme . . . . .	7
4.2. Extensions and Issues . . . . .	8
5. Peering Matrix . . . . .	9
5.1. Overview . . . . .	9
5.2. Partition Tree . . . . .	9
5.3. Computing the Peering Matrix: Bandwidth Matching . . . . .	10
5.4. Computing the Peering Matrix: Generic . . . . .	11
5.5. Live Streaming Results Using Planetlab . . . . .	11
6. IANA Considerations . . . . .	11
7. Security Considerations . . . . .	11
8. References . . . . .	11
8.1. Normative References . . . . .	11
8.2. Informative References . . . . .	11
Appendix A. Acknowledgments . . . . .	12
Authors' Addresses . . . . .	12

## 1. Introduction

ALTO provides information to network applications to improve network efficiency [2]. There are many ways that a network application can utilize ALTO information. For example, an application may choose to utilize only the Network Map, another may use both the Network Map and the Cost Map, and yet another may use the Endpoint ranking. It can be either a more centralized entity such as a tracker in a P2P application or the P2P clients that utilize ALTO information. One P2P application may choose to use the information at the P2P clients during piece selection or rate scheduling, while another P2P application may use it during peer selection. How to effectively utilize ALTO information is a challenge that P2P applications considering integrating with ALTO need to address.

In this document, we present example techniques of how to integrate ALTO information into the peer selection process at a P2P tracker. We first present some key challenges. We then present some techniques used in real trial examples that address the challenges.

## 2. Challenges

A P2P tracker selects a set of peers upon receiving a LISTING request of a peer. Since a tracker may receive a large number of such LISTING requests, it is important that the tracker can handle each request with high efficiency while achieving effectiveness in utilizing available information. The design of the data structures and algorithms at the tracker for peer selection is challenging and can have a major impact on the efficiency and effectiveness of peer selection.

Specifically, there are two challenges in integrating ALTO information into the peer selection of a P2P tracker.

- o Scalability: A P2P developer may have a small number of trackers to handle a large number of channels (files) each with multiple peers. The peers might be distributed across multiple ISPs that provide ALTO information. Thus, the storage and processing overhead caused by using ALTO information must be considered in order to scale to the increasingly larger P2P applications. In addition, it may be necessary to scale the tracker of a particularly popular channel from a single machine to multiple machines. In practice, many P2P applications may use multiple physical P2P trackers for a single channel for fault tolerance (e.g., when one tracker crashes) and/or connectivity reasons (e.g., poor connectivity between networks).

- o Application-Network Information Fusion: When selecting peers, a tracker should consider not only ALTO information, but also peer properties known only to the application (e.g., instantaneous peer upload capacity) as well as application requirements. In particular, a key concern of a P2P application is that solely considering ALTO information may lead to degraded application performance (e.g., slower download rate in a P2P file sharing application.)

One of the simplest ways to implement peer selection is random peer selection using a single array storing all current peers. Upon receiving a LISTING request, the tracker picks a random position in the array, and returns a set of peers starting from the chosen position. A slightly different random peer selection algorithm is to repeatedly pick random numbers in the range of the size of the array to pick multiple random peers.

An advantage of the preceding algorithm is scalability. But it is lacking in network-application information fusion. It does not consider peer properties during peer selection. On the other hand, many P2P trackers already select peers considering peer properties. For example, one type of peer property often considered is peer upload capabilities. Another type of peer property is the playpoint of a peer, in particular, in an VoD setting. Also, in addition to using ALTO information, some existing P2P trackers already consider network location properties such as the ASN, the IP prefix, the geo location (e.g., city, country or latitude/longitude), or the set of nearest landmarks of a peer.

### 3. Peer Classification Data Structures

When peers are annotated with properties, we might envision that the peers are stored at a tracker in a table similar in format to Figure 1:

peer_id	IP	upld_cap	play_point	ASN	country	cty	..
...	...	...	...	...	...	...	..

Figure 1: Using a Table to Store Peers

A problem of a flat table is that it does not support peer classification to find peers with given properties. Just as many databases build indices, many P2P trackers build inverted data structures such as map/hash in order to index to the pool of peers

with a given property.

In an abstract formulation, for each peer A requesting LISTING, the peer selection algorithm at the tracker determines a probability that any peer B will be returned to A, where the probability depends on the relative "match" between the properties of A and B. However, too much fine-grained tuning of the probabilities (there are  $O(N^2)$  such values, where N is number of peers) may not be necessary or feasible. Peer classification is a technique to aggregate peers into equivalent classes before peer selection to improve scalability.

### 3.1. One-Level Key Partitioning

Multiple examples exist in this category. In one example, the key is a category of the uploading capacity of a peer. For example, the tracker may classify peers as having high upload capacity, medium upload capacity, or low upload capacity. Then according to the property of the peer issuing the LISTING request, the tracker selects fractions of peers from each category.

In another example, the key can be the ASN or ISP name. When the tracker receives a LISTING request from a peer, the tracker looks up the ASN/ISP of the peer, and indexes to the ASN/ISP to select peers. To avoid partition of the P2P topology or when there are not sufficient numbers of peers in the ASN/ISP, the tracker may select peers from other ASNs/ISPs.

### 3.2. Hierarchical Partitioning

In addition to a single level flat map, some trackers classify peers using multiple attributes and/or build multiple levels of indexing, utilizing a hierarchical partitioning of peers according to peer properties.

One example is to use the hierarchical geo partitioning of peers first into country, then state/province, and then city.

Utilizing the Network Map of ALTO, the tracker can classify each peer into the PID of each ISP providing ALTO Network Map.

Extending the preceding examples of using one type of peer property, the partition keys at different levels can be from different categories. For example, at the first level, the tracker might use peer upload capacity, the next level uses ISP as the key, and the third level uses PID.



### 3.3. Comments

There are several comments. First, a tracker may partition peers from multiple perspectives. For example, each ISP providing ALTO Network Map provides a classification of peers into its set of PIDs. Thus, a single IP address may belong to different PIDs from different Network Maps of different ISPs. The tracker may build a classification tree for each ISP. It is possible that these multiple trees can be merged into a single tree with a dummy ROOT. We use a single classification tree as an example.

Second, the nodes of different non-overlapping branches may overlap regarding the sets of peers contained in them.

Second, it may not be straightforward or necessary to partition peers using certain properties. For example, landmarks may not lead to easy partitioning of peers.

## 4. Peer Selection Using Peer Classification

### 4.1. Overview of Scheme

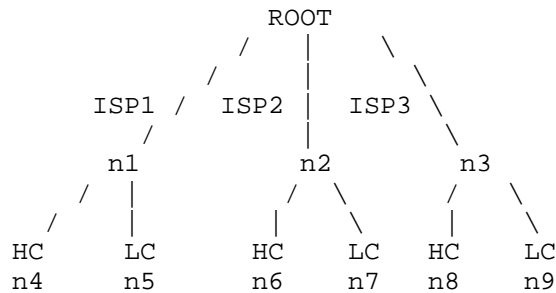
We now look at a class of tracker peer selection techniques that utilize peer classification. Each peer is located at a leaf node of a classification tree. We consider the set of algorithms where the peer selection depends on the properties of the peer issuing the request.

We introduce a concept called the "home" node of a peer issuing the LIST request. The home node is identified first before peer selection.

The peer selection is specified in the following way. Associated with each "home" leaf node of the classification tree is an ordered list, where each element in the list contains two fields: the first is a pointer to a node in a peer classification tree, and the second indicates how and how much to select peers from the node. It is important to notice that the list is ordered as the tracker picks peers in order. It is straightforward to extend that the nodes may come from multiple classification trees.

Figure 2 is an example. We refer to the data structure containing the selected peers as the bucket. The example specifies that when a peer A with "home" leaf node at n4 (high capacity peers from ISP1) issues a LISTING request, first fill 50% of the bucket containing peers to be returned to A from n4 (high capacity peers from same ISP) or no more peers available from n4, then continue to fill the bucket

by choosing peers from n5 (low capacity peers from same ISP) until the bucket is 80% full or no peers available from n5, then continue to fill the bucket by picking peers from n2 (peers from ISP2) so that the bucket can be 95% full, and finally fill the remaining of the bucket from n3. Note that the scheme intends that for n4, to pick more from the same ISP (80%), then ISP2 (15%), and then ISP3. It also tries to pair high capacity peers more with high capacity peers.



```

leaf n4: [n4, 50%] [n5, 80%] [n2, 95%] [n3, 100%]
leaf n5: [n4, 20%] [n5, 60%] [n2, 95%] [n3, 100%]
...
leaf n9: ...
  
```

Figure 2: Example: Peer Selection using Classification Tree.

#### 4.2. Extensions and Issues

The preceding peer selection scheme is simple and flexible. It can be efficiently implemented. There can be multiple ways to extend the scheme.

There are two remaining issues:

- o First, how to design the classification tree?
- o Second, how to create the traversal list of each leaf node?

In addition to addressing the two preceding issues, this scheme may not be a general representation of some existing peer selection schemes. For example, when the network location of a peer is represented by its set of close-by landmarks, a straightforward partition tree may not exist. Instead, some other data structures and algorithms may be needed to pick the peers that are the closest measured by a special metric space measured by "closeness" of landmark sets.

## 5. Peering Matrix

### 5.1. Overview

The Peering Matrix approach is an instance of the scheme of Peer Selection Using Partition Tree. It has been used in several trials, including the Pando/Comcast trial [3]. The scheme has also been evaluated in the context of P2P Live Streaming. Below, we present more details on Peering Matrix. We also briefly summarize a set of results applying Peering Matrix to P2P Live streaming on the Planetlab.

### 5.2. Partition Tree

Recall that we name the peer issuing the LISTING request as A. There is a unique path  $\text{Path}(A)$  going up from the leaf node containing A to ROOT in the partition tree. We consider the class of tracker peer selection algorithms that specify a (upper bound) target fraction of peers to be selected at each node  $n$  along  $\text{Path}(A)$ . The peer selection algorithm goes up along  $\text{Path}(A)$ . To be consistent, the fraction value at a node  $n$  should be no larger than that of the parent of  $n$  named  $\text{Parent}(n)$ .

Also, at each node  $n$  along  $\text{Path}(A)$ , for each child  $c$  of  $n$ , the peer selection algorithm specifies how peers are distributed among the siblings of  $c$ .

Consider an example in Figure 3:

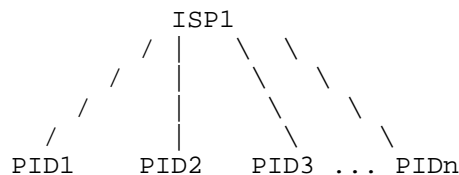


Figure 3: Example: Using ALTO Network Map for building a Classification Tree.

Specifically, Figure 3 is a two level classification tree for an ISP, and each second level node represents a PID of the ISP. Each PID is labeled with Fraction = 75%. The ROOT has a fraction of 100%. The sibling distribution of node PID1 node is 50%, 30%, 20% to PID2, PID3, and PID4 respectively. This means that when a peer A from PID1 asks for a list of peers, the tracker selects up to 75% peers at PID1, and fills the remaining (25%) at PID2 (up to 25% \* 50%), PID3 (25% \* 30%), and PID4 (25% \* 20%).

During P4P trials, we used a three level partition tree for each ISP.

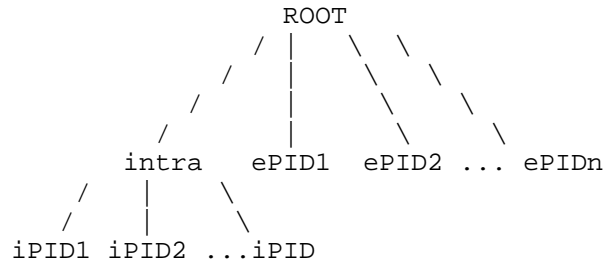


Figure 4: Three-Level Peer Classification.

One nice property of using per ISP classification tree is to implement a distributed tracker, where a tracker is responsible for the peers within a set of ISPs. A peer may request LISTING from multiple trackers (e.g., located at different ISPs) that together are responsible for the channel. The tracker hosting the "home" leaf of the peer uses peering matrix, while the other trackers return a small number of random peers for robustness.

### 5.3. Computing the Peering Matrix: Bandwidth Matching

To compute the sibling distribution at node intra and ROOT, the tracker estimates the aggregated upload capacity (a seed can use full upload capacity and a leecher achieves 70%) and demand of each PID and then conducts bandwidth matching as specified in [4].

Specifically, The following diagram shows how the information flow as well as how to transform ALTO Network Maps and Cost Maps into peering matrix, considering application states.

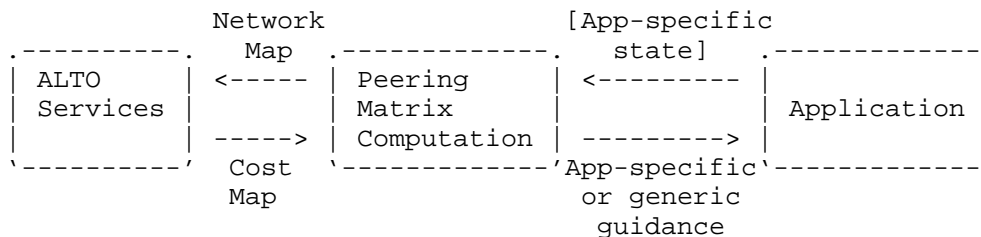


Figure 5: Information Flow to Compute Peering Matrix.

The interface to the Peering Matrix Computation Component, for a BitTorrent like file sharing application can be:

GetPeeringWeights: The request optionally includes swarm state information as a list of PIDs, and for each PID, the number of seeds and leechers and the aggregated download and upload capacity of clients within the PID. The response is a matrix of peering weights amongst the PIDs included in the request, as computed from the set of Costs currently pulled from the ALTO Server. If the request included swarm information, the returned weight matrix is tailored for the current state of the swarm.

#### 5.4. Computing the Peering Matrix: Generic

Similar to the preceding, but instead of using estimated capacity and demand, it assumes that each PID has one peer.

#### 5.5. Live Streaming Results Using Planetlab

### 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

This document does not evaluate security considerations. Multiple other documents in the ALTO working group considers the security perspective of using ALTO information.

### 8. References

#### 8.1. Normative References

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

#### 8.2. Informative References

- [2] Alimi, R., Penno, R., and Y. Yang, "ALTO Protocol", draft-ietf-alto-protocol-03 (work in progress), March 2010.
- [3] 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.

- [4] H. Xie, Y.R. Yang, A. Krishnamurthy, Y. Liu, and A. Silberschatz., "P4P:", In SIGCOMM 2008.

#### Appendix A. Acknowledgments

This document benefits substantially from trials and designs involving P2P applications Pando (Laird Pasko), PPLive (David Zhang), Digimeld (Gene Qin), and Xunlei. The data structure designs and algorithms include the contributions of Hao Wang, Ye Wang, and Harry Liu. We appreciate their contributions. The design is quite similar to that of Xunlei, whose more details will be included in the updated China Telecom trial document.

#### Authors' Addresses

Y. Richard Yang  
Yale University

Email: yry@cs.yale.edu

Richard Alimi  
Google

Email: ralimi@google.com

Ye Wang  
Yale University

Email: ye.wang@yale.edu

David Zhang  
PPLive

Email: davidzhang@pplive.com

Kai Lee  
China Telecom

Email: leek4u@gmail.com

