        Application-Layer Traffic Optimization (ALTO) Server Discovery

Abstract

   The goal of Application-Layer Traffic Optimization (ALTO) is to
   provide guidance to applications that have to select one or several
   hosts from a set of candidates capable of providing a desired
   resource. ALTO is realized by a client-server protocol. Before an
   ALTO client can ask for guidance, it needs to discover one or more
   ALTO servers.

   This document specifies a procedure for resource-consumer-initiated
   ALTO server discovery, which can be used if the ALTO client is
   embedded in the resource consumer.

Status of This Memo

   This is an Internet Standards Track document.

   This document is a product of the Internet Engineering Task Force
   (IETF). It represents the consensus of the IETF community. It has
   received public review and has been approved for publication by the
   Internet Engineering Steering Group (IESG). Further information on
   Internet Standards is available in Section 2 of RFC 5741.

   Information about the current status of this document, any errata,
   and how to provide feedback on it may be obtained at
   http://www.rfc-editor.org/info/rfc7286.

Copyright Notice

Table of Contents

1. Introduction

   The goal of Application-Layer Traffic Optimization (ALTO) is to
   provide guidance to applications that have to select one or several
   hosts from a set of candidates capable of providing a desired
   resource [RFC5693]. ALTO is realized by a client-server protocol;
   see requirement AR-1 in [RFC6708]. Before an ALTO client can ask for
   guidance it needs to discover one or more ALTO servers that can
   provide guidance to this specific client.

   This document specifies a procedure for resource-consumer-initiated
   ALTO server discovery, which can be used if the ALTO client is
   embedded in the resource consumer. In other words, this document
   meets requirement AR-32 in [RFC6708] while AR-33 is out of scope. A
   different approach, which tries to meet requirement AR-33, i.e.,
   third-party ALTO server discovery, is addressed in [3PDISC].

   A more detailed discussion of various options on where to place the
   functional entities comprising the overall ALTO architecture can be
   found in [ALTO-DEPLOY].

   The ALTO protocol specification [RFC7285] is based on HTTP and
   expects the discovery procedure to yield the HTTP(S) URI of an ALTO
   server's Information Resource Directory (IRD). Therefore, this
   procedure is based on a combination of the Dynamic Host Configuration
   Protocol (DHCP) or local configuration and URI-enabled Naming
   Authority Pointer (U-NAPTR) resource records in the Domain Name
   System (DNS), in order to deliver such URIs.

1.1. Terminology and Requirements Language

   This document makes use of the ALTO terminology defined in [RFC5693].

   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].

2. ALTO Server Discovery Procedure Overview

   The ALTO protocol specification [RFC7285] expects that the ALTO
   discovery procedure yields the HTTP(S) URI [RFC7230] of the ALTO
   server's Information Resource Directory (IRD), which gives further
   information about the capabilities and services provided by that ALTO
   server.

   On hosts with more than one interface or address family (IPv4/v6),
   the ALTO server discovery procedure has to be run for every interface
   and address family. For more details see Section 4.2.

The ALTO server discovery procedure is performed in two steps:

1. One DNS domain name is retrieved for each combination of
    interface and address family, either by local configuration
    (e.g., manual input into a menu or configuration file) or by
    means of DHCP.

2. These DNS domain names are used for U-NAPTR lookups yielding one
    or more URIs. Further DNS lookups may be necessary to determine
    the ALTO server's IP address(es).

The primary means for retrieving the DNS domain name is DHCP.
However, there may be situations where DHCP is not available or does
not return a suitable value. Furthermore, there might be situations
in which the user wishes to override the value that could be
retrieved from DHCP. In these situations, local configuration may be
used. Consequently, the algorithm first checks for a locally
configured override, before it tries to retrieve a value from DHCP.

Typically, but not necessarily, the DNS domain name is the domain
name in which the client is located, i.e., a PTR lookup on the
client's IP address (according to [RFC1035], Section 3.5 for IPv4 or
[RFC3596], Section 2.5 for IPv6) would yield a similar name.
However, due to the widespread use of Network Address Translation
(NAT), trying to determine the DNS domain name through a PTR lookup
on an interface's IP address is not recommended for resource consumer
initiated ALTO server discovery (see also [RFC3424]).

3. ALTO Server Discovery Procedure Specification

    As already outlined in Section 2, the ALTO server discovery procedure
    is performed for every address family on every interface the
    application considers for communicating with resource providers.

    First, the algorithm checks for a locally configured domain name, as
    specified in Section 3.1.1. If no such name was configured, it tries
    to retrieve one from DHCP, as specified in Section 3.1.2. If still
    no domain name could be found, the procedure has failed and
    terminates with an appropriate error code.

    If one or more domain names were found, they will be used as U-NAPTR/
    DDDS (URI-Enabled NAPTR/Dynamic Delegation Discovery Service)
    [RFC4848] application-unique strings for a DNS lookup, as specified
    in Section 3.2.

3.1. Step 1: Retrieving the Domain Name

3.1.1. Step 1, Option 1: Local Configuration

   The preferred way to acquire a domain name related to an interface's
   point of network attachment is the use of DHCP (see Section 3.1.2).
   However, in some network deployment scenarios, there is no DHCP
   server available. Furthermore, a user may want to use an ALTO
   service instance provided by an entity that is not the operator of
   the underlying IP network. Therefore, we allow the user to specify a
   DNS domain name, for example, in a configuration file option. An
   example domain name is:

      my-alternative-alto-provider.example.org

   Implementations MAY give the user the opportunity (e.g., by means of
   configuration file options or menu items) to specify an individual
   domain name for every address family on every interface.
   Implementations SHOULD allow the user to specify a default name that
   is used if no more specific name has been configured.

3.1.2. Step 1, Option 2: DHCP

   Network operators may provide the domain name to be used for service
   discovery within an access network using DHCP.

   RFC 5986 [RFC5986] defines DHCP IPv4 and IPv6 access network domain
   name options to identify a domain name that is suitable for service
   discovery within the access network. RFC 2132 [RFC2132] defines the
   DHCP IPv4 domain name option. While this option is less suitable, it
   still may be useful if the RFC 5986 option is not available.

   For IPv6, the ALTO server discovery procedure MUST try to retrieve
   DHCP option 57 (OPTION_V6_ACCESS_DOMAIN). If no such option can be
   retrieved the procedure fails for this interface. For IPv4, the ALTO
   server discovery procedure MUST try to retrieve DHCP option 213
   (OPTION_V4_ACCESS_DOMAIN). If no such option can be retrieved, the
   procedure SHOULD try to retrieve option 15 (Domain Name). If neither
   option can be retrieved, the procedure fails for this interface. If
   a result can be retrieved, it will be used as an input for the next
   step (U-NAPTR resolution). One example result could be:

      example.net

3.2. Step 2: U-NAPTR Resolution

The first step of the ALTO server discovery procedure (see
Section 3.1) retrieved one or -- in case of multiple interfaces and/
or IPv4/v6 dual-stack operation -- several domain names, which will
be used as U-NAPTR/DDDS (URI-Enabled NAPTR/Dynamic Delegation
Discovery Service) [RFC4848] application unique strings. An example
is:

    example.net

In the second step, the ALTO server discovery procedure uses a
U-NAPTR [RFC4848] lookup with the "ALTO" Application Service Tag and
either the "http" or the "https" Application Protocol Tag to obtain
one or more URIs (indicating protocol, host, and possibly path
elements) for the ALTO server's Information Resource Directory. In
this document, only the HTTP and HTTPS URI schemes are defined, as
the ALTO protocol specification defines the access over both
protocols, but no other [RFC7285]. Note that the result can be any
valid HTTP(S) URI.

The following two U-NAPTR resource records can be used for mapping
"example.net" to the HTTPS URIs "https://alto1.example.net/ird" and
"https://alto2.example.net/ird", with the former being preferred.

    example.net.

    IN NAPTR 100 10   "u"   "ALTO:https"
        "!.*!https://alto1.example.net/ird!" ""

    IN NAPTR 100 20   "u"   "ALTO:https"
        "!.*!https://alto2.example.net/ird!" ""

If no ALTO-specific U-NAPTR records can be retrieved, the discovery
procedure fails for this domain name (and the corresponding interface
and IP protocol version). If further domain names retrieved by Step
1 are known, the discovery procedure may perform the corresponding
U-NAPTR lookups immediately. However, before retrying a lookup that
has failed, a client MUST wait a time period that is appropriate for
the encountered error (NXDOMAIN, timeout, etc.).

4. Deployment Considerations

4.1. Issues with Home Gateways

   Section 3.1.2 describes the usage of a DHCP option that provides a
   means for the network operator of the network in which the ALTO
   client is located to provide a DNS domain name. However, this
   assumes that this particular DHCP option is correctly passed from the
   DHCP server to the actual host with the ALTO client, and that the
   particular host understands this DHCP option. This memo assumes the
   client to be able to understand the proposed DHCP option; otherwise,
   there is no further use of the DHCP option, but the client has to use
   the other proposed mechanisms.

   There are well-known issues with the handling of DHCP options in home
   gateways. One issue is that unknown DHCP options are not passed
   through some home gateways, effectively eliminating the DHCP option.

   Another well-known issue is the use of home-gateway-specific DNS
   domain names that "override" the DNS domain name provided by the
   network operator. For instance, a host behind a home gateway may
   receive a DNS domain name ".local" instead of "example.net". In
   general, this domain name is not usable for the server discovery
   procedure, unless a DNS server in the home gateway resolves the
   corresponding NAPTR lookup correctly, e.g., by means of a DNS split
   horizon approach.

4.2. Issues with Multihoming, Mobility, and Changing IP Addresses

   If the user decides to enter only one (default) DNS domain name in
   the local configuration facility (see Section 3.1.1), only one set of
   ALTO servers will be discovered, irrespective of multihoming and
   mobility. Particularly in mobile scenarios, this can lead to
   undesirable results.

   The DHCP-based discovery method can discover different sets of ALTO
   servers for each interface and address family (i.e., IPv4/v6). In
   general, if a client wishes to communicate using one of its
   interfaces and using a specific IP address family, it SHOULD query
   the ALTO server or servers that have been discovered for this
   specific interface and address family. How to select an interface
   and IP address family as well as how to compare results returned from
   different ALTO servers are out of the scope of this document.

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 re-run the ALTO server discovery procedure without
relying on earlier gained information.

There are several challenges with DNS on hosts with multiple
interfaces [RFC6418], which can affect the ALTO server discovery. If
the DNS resolution is performed on the wrong interface, it can return
an ALTO server that could provide suboptimal or wrong guidance.
Finding the best ALTO server for multi-interfaced hosts is outside
the scope of this document.

When using Virtual Private Network (VPN) connections, there is
usually no DHCP. The user has to enter the DNS domain name in the
local configuration facility. For good optimization results, a DNS
domain name corresponding to the VPN concentrator, not corresponding
to the user's current location, has to be entered. Similar
considerations apply for Mobile IP.

5. IANA Considerations

IANA has registered the following U-NAPTR [RFC4848] application
service tag for ALTO:

Application Service Tag: ALTO

Intended usage: see [RFC5693] or: "The goal of Application-Layer
    Traffic Optimization (ALTO) is to provide guidance to applications
    that have to select one or several hosts from a set of candidates
    capable of providing a desired resource."

Defining Publication: The specification contained within this
    document

Contact information: The authors of this document

Author/Change controller: The IESG

Interoperability considerations: No interoperability issues are
    known or expected. This tag is to be registered specifically for
    ALTO, which is a new application without any legacy deployments.

Security considerations: see Section 6 of this document.

   Related publications: This document specifies a procedure for
      discovering an HTTP or HTTPS URI of an ALTO server. HTTP and
      HTTPS are specified in [RFC7230]. The HTTP(S)-based ALTO protocol
      is specified in [RFC7285].

   Application Protocol Tag: This document specifies how to use the
      application service tag "ALTO" with the application protocol tags
      "http" and "https", which have already been registered in the
      relevant IANA registry. Therefore, IANA is not requested by this
      document to register any new application protocol tag.

6. Security Considerations

   A high-level discussion of security issues related to ALTO is part of
   the ALTO problem statement [RFC5693]. A classification of unwanted
   information disclosure risks, as well as specific security-related
   requirements can be found in the ALTO requirements document
   [RFC6708].

   The remainder of this section focuses on security threats and
   protection mechanisms for the ALTO server discovery procedure as
   such. Once the ALTO server's URI has been discovered and the
   communication between the ALTO client and the ALTO server starts, the
   security threats and protection mechanisms discussed in the ALTO
   protocol specification [RFC7285] apply.

6.1. Integrity of the ALTO Server's URI

   Scenario Description
      An attacker could compromise the ALTO server discovery procedure
      or infrastructure in a way that ALTO clients would discover a
      "wrong" ALTO server URI.

   Threat Discussion
      This is probably the most serious security concern related to ALTO
      server discovery. The discovered "wrong" ALTO server might not be
      able to give guidance to a given ALTO client at all, or it might
      give suboptimal or forged information. In the latter case, an
      attacker could try to use ALTO to affect the traffic distribution
      in the network or the performance of applications (see also
      Section 15.1. of [RFC7285]). Furthermore, a hostile ALTO server
      could threaten user privacy (see also Section 5.2.1, case (5a) in
      [RFC6708]).

      However, it should also be noted that, if an attacker was able to
      compromise DHCP and/or DNS servers used for ALTO server discovery
      (see below), (s)he could also launch significantly more serious
      other attacks (e.g., redirecting various application protocols).

  Protection Strategies and Mechanisms
     The ALTO server discovery procedure consists of three building
     blocks (local configuration, DHCP, and DNS) and each of them is a
     possible attack vector.

     The problem of users possibly following "bad advice" that tricks
     them into manually configuring unsuitable ALTO servers cannot be
     solved by technical means and is out of the scope of this
     document.

     Due to the nature of the protocol, DHCP is rather prone to
     attacks. As already mentioned, an attacker that is able to inject
     forged DHCP replies into the network may do significantly more
     harm than only configuring a wrong ALTO server. Best current
     practices for safely operating DHCP should be followed.

     A further threat is the possible alteration of the DNS records
     used in U-NAPTR resolution. If an attacker was able to modify or
     spoof any of the DNS records used in the DDDS resolution, this URI
     could be replaced by a forged 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].

     A related risk is the impersonation of the ALTO server (i.e.,
     attacks after the correct URI has been discovered). This threat
     and protection strategies are discussed in Section 15.1 of
     [RFC7285]. Note that if Transport Layer Security (TLS) is used to
     protect ALTO, the server certificate will contain the host name
     (CN). Consequently, only the host part of the HTTPS URI will be
     authenticated, i.e., the result of the ALTO server discovery
     procedure. The U-NAPTR based mapping within the ALTO server
     discovery procedure needs to be secured as described above, e.g.,
     by using DNSSEC.

     In addition to active protection mechanisms, users and network
     operators can monitor application performance and network traffic
     patterns for poor performance or abnormalities. If it turns out
     that relying on the guidance of a specific ALTO server does not
     result in better-than-random outcomes, the use of the ALTO server
     may be discontinued (see also Section 15.2 of [RFC7285]).

6.2. Availability of the ALTO Server Discovery Procedure

Scenario Description
   An attacker could compromise the ALTO server discovery procedure
   or infrastructure in a way that ALTO clients would not be able to
   discover any ALTO server.

Threat Discussion
   If no ALTO server can be discovered (although a suitable one
   exists), applications have to make their decisions without ALTO
   guidance. As ALTO could be temporarily unavailable for many
   reasons, applications must be prepared to do so. However, the
   resulting application performance and traffic distribution will
   correspond to a deployment scenario without ALTO.

Protection Strategies and Mechanisms
   Operators should follow best current practices to secure their
   DHCP, DNS, and ALTO (see Section 15.5 of [RFC7285]) servers
   against Denial-of-Service (DoS) attacks.

6.3. Confidentiality of the ALTO Server's URI

Scenario Description
   An unauthorized party could invoke the ALTO server discovery
   procedure, or intercept discovery messages between an authorized
   ALTO client and the DHCP and DNS servers, in order to acquire
   knowledge of the ALTO server's URI.

Threat Discussion

   In the ALTO use cases that have been described in the ALTO problem
   statement [RFC5693] and/or discussed in the ALTO working group,
   the ALTO server's URI as such has always been considered as public
   information that does not need protection of confidentiality.

Protection Strategies and Mechanisms
   No protection mechanisms for this scenario have been provided, as
   it has not been identified as a relevant threat. However, if a
   new use case is identified that requires this kind of protection,
   the suitability of this ALTO server discovery procedure as well as
   possible security extensions have to be re-evaluated thoroughly.

6.4. Privacy for ALTO Clients

Scenario Description
     An unauthorized party could intercept discovery messages between
     an ALTO client and the DHCP and DNS servers, and thereby find out
     the fact that said ALTO client uses (or at least tries to use) the
     ALTO service.

Threat Discussion
     In the ALTO use cases that have been described in the ALTO problem
     statement [RFC5693] and/or discussed in the ALTO working group,
     this scenario has not been identified as a relevant threat.

Protection Strategies and Mechanisms
     No protection mechanisms for this scenario have been provided, as
     it has not been identified as a relevant threat. However, if a
     new use case is identified that requires this kind of protection,
     the suitability of this ALTO server discovery procedure as well as
     possible security extensions have to be re-evaluated thoroughly.

7. References

7.1. Normative References

[RFC2119]   Bradner, S., "Key words for use in RFCs to Indicate
            Requirement Levels", BCP 14, RFC 2119, March 1997,
            <http://www.rfc-editor.org/info/rfc2119>.

[RFC2132]   Alexander, S. and R. Droms, "DHCP Options and BOOTP Vendor
            Extensions", RFC 2132, March 1997,
            <http://www.rfc-editor.org/info/rfc2132>.

[RFC4848]   Daigle, L., "Domain-Based Application Service Location
            Using URIs and the Dynamic Delegation Discovery Service
            (DDDS)", RFC 4848, April 2007,
            <http://www.rfc-editor.org/info/rfc4848>.

[RFC5986]   Thomson, M. and J. Winterbottom, "Discovering the Local
            Location Information Server (LIS)", RFC 5986, September
            2010, <http://www.rfc-editor.org/info/rfc5986>.

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

7.2. Informative References

   [3PDISC]   Kiesel, S., Krause, K., and M. Stiemerling, "Third-Party
              ALTO Server Discovery (3pdisc)", Work in Progress,
              draft-kist-alto-3pdisc-05, January 2014.

   [ALTO-DEPLOY]
              Stiemerling, M., Kiesel, S., Previdi, S., and M. Scharf,
              "ALTO Deployment Considerations", Work in Progress,
              draft-ietf-alto-deployments-10, July 2014.

   [RFC1035] Mockapetris, P., "Domain names - implementation and
              specification", STD 13, RFC 1035, November 1987,
              <http://www.rfc-editor.org/info/rfc1035>.

   [RFC3424] Daigle, L. and IAB, "IAB Considerations for UNilateral
              Self-Address Fixing (UNSAF) Across Network Address
              Translation", RFC 3424, November 2002,
              <http://www.rfc-editor.org/info/rfc3424>.

   [RFC3596] Thomson, S., Huitema, C., Ksinant, V., and M. Souissi,
              "DNS Extensions to Support IP Version 6", RFC 3596,
              October 2003, <http://www.rfc-editor.org/info/rfc3596>.

   [RFC4033] Arends, R., Austein, R., Larson, M., Massey, D., and S.
              Rose, "DNS Security Introduction and Requirements", RFC
              4033, March 2005,
              <http://www.rfc-editor.org/info/rfc4033>.

   [RFC5693] Seedorf, J. and E. Burger, "Application-Layer Traffic
              Optimization (ALTO) Problem Statement", RFC 5693, October
              2009, <http://www.rfc-editor.org/info/rfc5693>.

   [RFC6418] Blanchet, M. and P. Seite, "Multiple Interfaces and
              Provisioning Domains Problem Statement", RFC 6418,
              November 2011, <http://www.rfc-editor.org/info/rfc6418>.

   [RFC6708] Kiesel, S., Previdi, S., Stiemerling, M., Woundy, R., and
              Y. Yang, "Application-Layer Traffic Optimization (ALTO)
              Requirements", RFC 6708, September 2012,
              <http://www.rfc-editor.org/info/rfc6708>.

   [RFC7230] Fielding, R. and J. Reschke, "Hypertext Transfer Protocol
              (HTTP/1.1): Message Syntax and Routing", RFC 7230, June
              2014, <http://www.rfc-editor.org/info/rfc7230>.

Acknowledgments

Contributors

   The initial version of this document was coauthored by Marco Tomsu.

   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.

   The authors would also like to thank the following persons for their
   contribution to this document or its predecessors: Richard Alimi,
   David Bryan, Roni Even, Gustavo Garcia, Jay Gu, Xingfeng Jiang,
   Enrico Marocco, Victor Pascual, Y. Richard Yang, Yu-Shun Wang, Yunfei
   Zhang, Ning Zong.

Authors' Addresses

    Sebastian Kiesel
    University of Stuttgart Information Center
    Networks and Communication Systems Department
    Allmandring 30
    Stuttgart 70550
    Germany

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

    Martin Stiemerling
    NEC Laboratories Europe
    Kurfuerstenanlage 36
    Heidelberg 69115
    Germany

    Phone: +49 6221 4342 113
    EMail: mls.ietf@gmail.com
    URI: http://ietf.stiemerling.org

    Nico Schwan
    Thales Deutschland
    Thalesplatz 1
    Ditzingen 71254
    Germany

    EMail: ietf@nico-schwan.de

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

    EMail: michael.scharf@alcatel-lucent.com

    Haibin Song
    Huawei

    EMail: haibin.song@huawei.com