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 disucss 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

Table of Contents

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:

- 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.

- 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.

- Intra-domain ALTO-Integrated Application Traffic (Network metric).

- 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.

- 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.

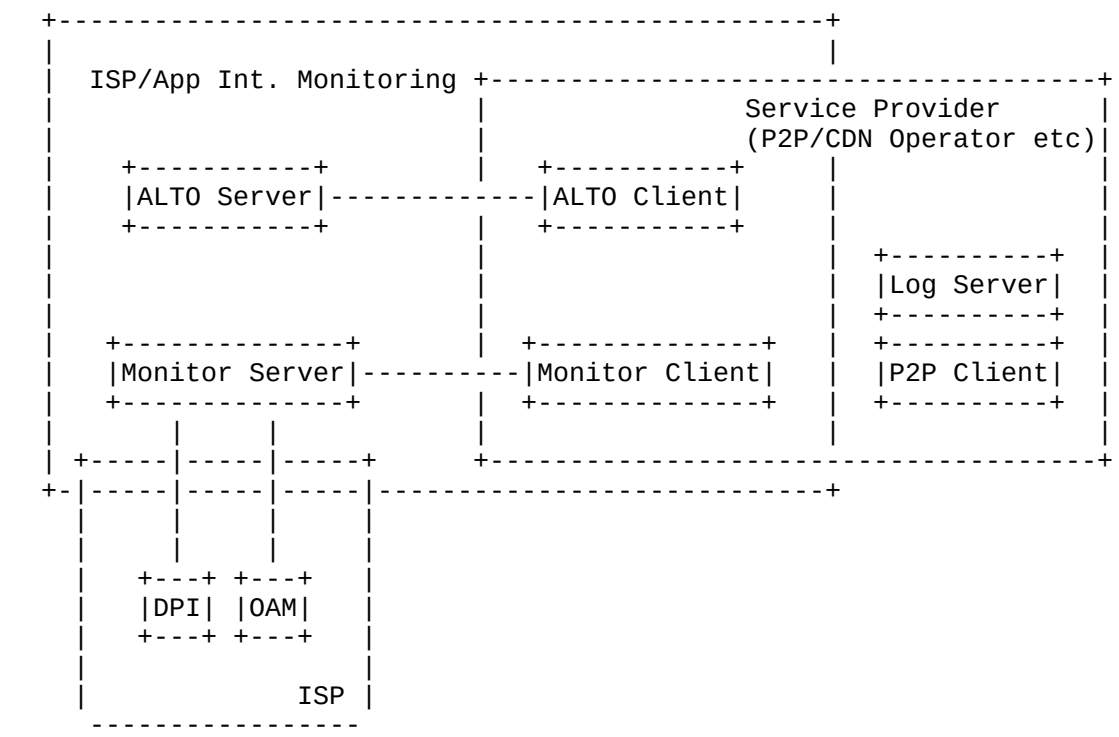

Figure 1

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