INTERNET-DRAFT
Intended Status: Informational draft
Expires: April 5, 2015

Arunkumar Arumuga Nainar
Tata Communications Ltd

October 2, 2014

Layer 4 Path preference negotiation for DPS
draft-arumuganainar-rtgwg-dps-l4-ppn-01

Abstract

This document is a supporting draft to draft-arumuganainar-rtgwg-DPS-Requirements-01. The DPS draft talks about high level architecture to implement dynamic path selection based on application. The DPS draft suggests the implementation to be done in three steps:

1) DPS Signaling: Here participating routers communicate with each other to exchange application related information

2) Profile Based Filter: This section describes how packets can be classified and filtered

3) DPS Routing Frame Work: This ensures that separated traffic remains separated through out the network

While overall architecture is still valid, this draft suggests an enhancement to the DPS Signaling component. The currently implemented technique uses BGP for the signaling requirements. Whilst this is good for certain cases, applications that can be off loaded to the secondary link are pre-decided. It restricts behavior from responding to dynamic network conditions. For example, a network administrator would want to off load some of the non-critical applications over the secondary link, however when there is acute congestion within the network, they might want the router to behave aggressively by off loading more applications to the secondary circuit. Yet while doing so there should not be any asymmetric routing on the network.

Since BGP is essentially a control plane protocol, it is not aware of what is happening on the network in the forwarding plane, hence there is a need to do the signaling in the forwarding plane. This drafts suggest one such mechanism. Here the idea is to exchange the Path Preference information at layer 4 level. Such signaling could happen during TCP connection establishment phase. When done this way, a decision can be taken for each of the session and hence making it more dynamic than the one that can achieved through BGP.

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the

Copyright and License Notice

Table of Contents

1 Introduction

BGP is used for signaling in the current implementation of DPS . BGP provides a consistent mechanism to exchange the Path Preference information between the two routers/sites. However, since BGP is a control plane protocol, we will not be able to exchange the dynamically changing network conditions. Here there is need to do the signaling in the forwarding plane.

This draft proposes to do the signaling during the TCP connection establishment phase. When done this way the signaling is done for each session. This enables the router to be aware of network conditions dynamically and take the most appropriate path.

1.1 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 RFC 2119 [RFC2119].

2. DPS Layer 4 Signaling Overview.

The TCP connection establishment follows the three way process. This can be summarised in 3 steps:

1) Client M/C first sends a SYN packet

2) Sever M/C accepts the SYN and responds back with a SYN/ACK

3) Client Acknowledges the SYN/ACK with an ACK

During the whole process, many parameters will be exchanged and agreed upon. Some router implementations intercept these session initiation packets and influence the exchanged information. For example, routers could intercept the SYN packet and force the end hosts to negotiate a reduced MSS size.

Similar legal intercept technique could be used to exchange Path Preference information.

Client -----Router1 ====(Network)======Router2------Server

Here the interception works in the below fashion:

1) Client sends a packet to server

2) Router 1 intercepts the SYN packet and adds TCP options. This option could be loaded with Path Preference information. For example, let us assume there is no congestion on the client site's tail circuit. Hence, router 1 wants to send traffic that belongs to this session over the primary circuit.

3) When router 2 receives the SYN packet with that known TCP option. It comes to know that the SYN is coming from a DPS capable site. It then caches the TCP state, strips the option and sends it to the server.

4) When the server responds back with a SYN/ACK, router 2 intercepts the SYN/ACK and adds its Path Preference information in to the SYN/ACK packet using the same TCP option. Let's assume there is a congestion on the network at router 2, it replies back saying my side is congested so let us off load this session to the secondary link.

5) Router 1, when it receives back the SYN/ACK it knows it is coming from a DPS capable site and the remote site wants the application to be off loaded to the secondary link.

6) When the client sends back ACK, router 1 intercepts it and adds an option confirming that it agrees to the remote site choice to offload that session

Subsequently when the data packets are sent over a path that is negotiated and agreed. Here the preference of a particular path can be taken based on dynamic local network conditions. Since the path is negotiated and agreed across both ends, the network will respond to both local and remote network conditions.

3.Application Classification :-

The key thing here is how does a router decide which application needs to be off loaded. This draft makes the following recommendations.

To begin with, the administrator of the network should prepare 3 lists: a) white list b) grey list c) black list

White List:- Application listed in the white list always goes over the primary link. The behavior does not change even if there is any congestion on the link.

Black List:- Application listed in the black list always goes over secondary link. Even when there is no congestion on the primary link.

Grey List:- Application listed in the grey list behaves like a white list application when the primary link is not congested and behaves like a black list application when there is congestion.

It should be noted that the grey list behavior can also be a result of remote side congestion. For example, when given Site prefers the primary link for the grey list application but the remote site prefers the back up link, then both sites will agree on back path for the session.

Also the definition of application could also include a layer 7 signature. For example:

1) HTTP (TCP Port: 80) made to URL www.youtube.com. Could be put on a black list

2) HTTP (TCP Port: 80) made to URL www.gmail.com. Could be put on a grey list

Hence it is mandatory that router implementations should support deep packet inspection for the purpose of classification (based on layer 7 signature)

4. Application List And Grey Scales

While the 3 list hierarchy does seems to work. Under practical circumstances this is not sufficient. Network administrators will want to off load applications less aggressively when the utilization levels are very low. However, they may want to progressively off load more applications or TCP sessions. To illustrate this, the following example is given:

Let us say following applications are listed as grey applications:

Grey Shade 1:- Email, Intranet

Grey Shade 2:- Video streaming application such as You Tube

Grey shade 3:- All peer to peer applications.

1) When link utilization is <=60 all grey applications will go over the primary link

2) When link utilisation is >60 but <=70, shade 3 applications(peer-to-peer) will go over the secondary link

3) When link link utilisation is >70 but <=90, shade 3 and shade 2 applications (peer-to-peer & video streaming) will go over the secondary

4) When link utilisation is >90, all grey applications will be off loaded

5. Layer 7 Classification Issue:-

In order to make the off loading very effective, we should be able to do two things.

1) Classification and grouping of protocol to be done based on L7 signatures

2) Path Preferences should be negotiated per session. This actually provides the ability to react to local or remote network conditions

Whilst negotiating the Path Preferences during the connection establishment, there will be problem reading the layer 7 Signatures during that phase. Layer 7 signature will be only know after the connection is established. Hence if the grey list is based on layer 7 signatures, then direct negotiation of Path Preferences will not be possible. Hence an alternate approach needs to be worked out.

To overcome the above problem, the draft suggests, rather than negotiating Path Preferences directly, we could get pre-authorization for the use of specific path. This will work based on the following rules.

1) When the SYN packet hits router, it intercepts that packet and adds a TCP option. This option will communicate what is the maximum grey shade that router can send via the primary link. For example, if you have 6 shades of grey defined and if the TCP option indicates option 4, then it indicates router is seeking pre-authorization for offloading applications that belongs to shade 5 or 6.

2) However it could happen that in the remote site primary link is heavily congested and hence it would want to offload all applications that belong to shades 3, 4, 5 & 6. Hence it's router responds back with TCP option set to 2.

3) Since 2 is less than 4, Originating site-router accepts Remote site router's choice and confirms it agreement by sending back the acceptance in the ACK message.

When the connection is established routers in both site would have formally agreed what path to choose when they determine the layer 7 signatures. Because of this pre-authorization, there is no chance of asymmetric routing. In this scheme of things following rules apply:

1) Administrator should clearly indicate what is LAN facing interface and what is WAN facing interface on the router.

2) Any SYN packet with no Path Preference information set would indicate the packet is coming from a site that does not have DPS enabled. In such a case normal routing will be followed to forward the packet.

3) If the two routers specify different Path Preference information, the lowest number always wins.

Additionally it is based on the assumption that classification of application and grey scale definition will have to be done globally. This is because only grey scale levels are negotiated. The premise of the negotiation is based on the fact that once the grey scale is negotiated, the path of application could determined with certainty. This is only possible when the definitions are made globally. Hence the draft recommends any implementations should support central definition of an application list. And any individual routers/devices participating in DPS should be able to pull the definition from central location so as avoid inconsistencies in path selection.

UDP and other non-TCP based application may not be able to participate in the negotiation. Hence they could be either placed in a white list or a black list.

6. Customization Of Path Selection

Though the definition of the white list, grey list and black list is universal across the network, customization is still possible. For example a smaller site with a smaller primary circuit can be configured to support only two shades grey. While Larger site that have large bandwidths can be configured to support 5 shades of grey. Also grey shade selection could also individually configured. For illustration following example is provided.

A given network 3 type of site.

Site type A: Primary circuit : 100 MBPS , Secondary : 100 MBPS

Site Type B: Primary circuit : 10 MBPS , Secondary : 10 MBPS

Site Type C: Primary circuit : 2 MBPS , Secondary : 2 MBPS

Site type A and B can support all the 5 shades of grey. And site type C can support only 2 shades of grey. Further site, type A starts off loading traffic when the utilization level goes above 80% while site type B does it when it crosses 60%.

Even if we tweak this per site, the chance of asymmetric routing does not arise because two sites that are communicating will always negotiate the same grey levels to off load.

8. Implementation Details.

Currently many non-routing devices also perform deep packet inspection. For example, WAN optimization appliances. The signaling portion can also be off loaded to devices behind routers. In such a case DPS will be implemented on the router based on the traditional model. However the router will be configured to trust the marking done on the sites that hosts these WAN optimization devices. i.e Coloring will be done on an external device that is capable of doing deep packet inspection.

7. Summary

Currently in the DPS implementation, signaling is done in the signaling plane via BGP. Because the signaling happens in the control plane, we are limited to define applications statically. It does not provide room for off loading applications based on dynamically changing network condition.

By moving the signaling to forwarding plane, the draw back of the previous method of signaling can be done away with.

This draft suggests doing the signaling when the TCP connection is established. This draft also acknowledges difficulty to integrate layer 7 signatures based classification. To overcome this, the implementation should support global definition application lists. The two sites, whilst negotiating the Path Preferences, only negotiates the highest shade of grey that can sent over the primary circuit.

In summary, it is possible to do the signaling of path preference information at layer 4. This would enable DPS implementation to alter the DPS behavior based on network level changes that are local or remote to the given site.

```
   Definitions and code {
     line 1
     line 2
   }
```

Special characters examples:

The characters , , ,
However, the characters \0, \&, \%, \" are displayed.

.ti 0 is displayed in text instead of used as a directive.
.\" is displayed in document instead of being treated as a comment

C:\dir\subdir\file.ext Shows inclusion of backslash "\".

8 Security Considerations

TBD

9 IANA Considerations

TBD

10 References

10.1 Normative References

10.2 Informative References

11 Acknowledgements

The authors would like to thank Hesham Moussa for his review and comments.

Authors' Addresses

Arunkumar Arumuga Nainar
Tata Communications (UK)
1st Floor
20 Old Bailey
London EC4M 7AN
United Kingdom

EMail: arun.arumuganainar@tatacommunications.com