TEAS Working Group                                          A.Wang
Internet Draft                                          China Telecom

Intended status: Standard Track                          June 30,2016
Expires: December 30, 2016

PCE in Native IP Network
draft-wang-teas-pce-native-ip-00.txt

Status of this Memo

Abstract

This document defines the PCE use case and solution that can be deployed within the native IP network, using Multi-BGP session strategy and PCE-based central control to assure the end2end traffic performance, and proposes the corresponding extension to PCEP protocol to transfer the key parameters between PCE and the underlying network device (PCC).

Table of Contents



## 1. Introduction

Currently, PCE based traffic assurance requires the underlying network devices support MPLS and the network must deploy multiple LSPs to assure the end-to-end traffic performance. LDP/RSVP-TE or Segment Routing should be enabled within the network to establish various MPLS paths. Such solution will certainly work but the main drawback of it is that all the LSP paths are divided logically, that is to say, all the LSP paths that go through one physical link will share and compete the





same resource and MPLS technology has no better solution to meet the requirements for determined QoS effect.
On the other hand, there are some legacy networks that does not deploy the MPLS control and forward plane technology, but also need to assure the QoS of application traffic. Deploy some dedicated links statically to meet such requirements is one option but it is not feasible in the service provider network, because the volume and path of application traffic will be vary from time to time.
In summary, there are scenarios that the current PCE-based MPLS solution can't be deployed within the network, because the following user requirements:

1) End to End traffic assurance.
2) Determined Qos Effect.
3) No complex MPLS signaling procedure, support Native IP environment.
4) Flexible deployment
5) Central control.

This document defines the PCE use case and solution that can be deployed within the native IP network, using PCE-based central control and Multi BGP sessions strategy to assure the end2end traffic performance, meet the above requirements in dynamical and central control mode, proposes the corresponding extension to PCEP protocol to transfer the key parameters between PCE and the underlying network device(PCC).

## 2. Conventions used in this document

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].

## 3. Dual-BGP solution for simple topology.

This section introduces first the dual-BGP solution for simple topology that illustrated in Fig.1, which is comprised by SW1, SW2, R1, R2. There are multiple physical links between R1 and R2. Traffic between IP11 and IP21 are normal traffic, traffic between IP12 and IP22 are priority traffic that should be treated differently.
There is only Native IP protocol being deployed between R1 and R2. The traffic between each address pair will be changed timely and the

corresponding source/destination addresses of the traffic may also be changed dynamically.
The key idea of the Dual-BGP solution for this simple topology is the





following:

1) Build two BGP sessions between R1 and R2, via the different loopback address lo0,lo1 on these routers.
2) Send different prefixes via the two BGP sessions.(For example, IP11/IP21 via the BGP pair 1 and IP12/IP22 via the BGP pair 2).
3) Set the static route on R1 and R2 respectively for BGP next hop of lo0,lo1 to different physical link address between R1 and R2.

So, the traffic between the IP11 and IP12, and the traffic between IP21 and IP22 will go through different physical links between R1 and R2, each type of traffic occupied the different dedicated physical links and will not influence with each other.

If there is more traffic between IP12 and IP13 need to be assured, one can reassign more physical links on R1 and R2 to reach the loopback address lo1(also the next hop for BGP Peer pair2), the prefixed that advertised by two BGP peer need not be changed.

If, for example, there are traffic from another address pair need to be assured (for example IP13/IP23), but the total volume of assured traffic does not exceed the capacity of the previous appointed physical links, then one need only to advertise the newly added source/destination prefixes via the BGP peer pair2, then the traffic between IP13/IP23 will go through the assigned dedicated physical links as the traffic between IP12/IP22.

Such decouple philosophy gives the network operator more flexible control ability on the network traffic, get the determined QoS assurance effect to meet the application's requirement. No complex MPLS signal procedures is introduced, the router need only support native IP protocol.

```
                         |                  |
                         |  BGP Peer Pair2  |
                         +------------------+
                         |lo1           lo1 |
                         |                  |
                         |  BGP Peer Pair1  |
                         +------------------+
            IP12       |lo0           lo0 |       IP22
            IP11       |                  |       IP21
```

```
                     SW1-------R1-----------------R2-------SW2
                                  Links Group
```





Fig.1 Design Philosophy for Dual-BGP Solution

## 4. Dual-BGP in large Scale Topology

When the assured traffic spans across one large scale network, as that illustrated in Fig.2, the dual BGP sessions cannot be established neighbor by neighbor especially for the iBGP within one AS. For such scenario, we should consider to use the Route Reflector (RR) to achieve the similar Dual-BGP effect, that is to say, select one router which performs the role of RR (for example R3 in Fig.2 - Dual-BGP Solution using Route Reflector for large scale network), every other router will establish two BGP sessions with the RR, using their different loopback addresses respectively. The other two steps for traffic differentiation are same as one described in the Dual-BGP simple topology usage case.

For the example shown in Fig.2, if we select the R1-R2-R4-R7 as the dedicated path, then we should set the static routes on theses router respectively, point the BGP next hop (loopback addresses of R1 and R7, which are used to send the prefix of the assured traffic) to the actual address of the physical link

```
                     +------------R3--------------+
                     |                            |
          SW1-------R1-------R5---------R6-------R7--------SW2
                     |        |          |        |
                     +--------R2---------R4--------+
```

Fig.2 Dual-BGP solution using route reflector for large scale network

## 5. Multi-BGP for Extended Traffic Differentiation

Discussed in the document so far, is the requirement for traffic differentiation to classify traffic into two classes: Assured traffic or best effort (normal) traffic. Dual-BGP solution(simple topology or large scale topology) can meet above requirements. In general situations, several additional traffic differentiation criteria exist, including:

? Traffic requires low latency links and not sensitive to packet loss
? Traffic requires low packet loss but can endure higher latency

?  Traffic requires lowest jitter path





?  Traffic requires high bandwidth links

These varying traffic requirements may be summarized in the following table:

| Flow No. | Latency | Packet Loss | Jitter |
|---|---|---|---|
| 1 | Low | Normal | Don't care |
| 2 | Normal | Low | Dont't care |
| 3 | Normal | Normal | Low |

Table 1. Traffic Requirement Criteria

For Flow No.1, we can select the shortest distance path to carry the traffic; for Flow No.2, we can select the idle links to form its end to end path; for Flow No.3, we can let all the traffic pass one single path, no ECMP distribution on the parallel links is required.

It is difficult and almost impossible to provide an end-to-end (E2E) path with latency, latency variation, packet loss, and bandwidth utilization constraints to meet the above composition requirements in large scale network via the traditional distributed routing protocol, but these requirements can be solved using the PCE-based architecture since the PCE has the overall network view, can collect real network topology and network performance information about the underlying network, select the appropriate path to meet the various network performance requirements of different traffic type.

## 6. PCE based solution for Multi-BGP strategy deployment.

With the advent of SDN concepts within IP network, it is possible to deploy the PCE related technology into the underlying native IP network, to accomplish the central and dynamic control of network traffic according to the application's various requirements.

The procedure to implement the dynamic deployment of Multi-BGP strategy is the following:

1) PCE gets underlying topology information via the BGP-LS protocol via one router, such as the route reflector R3 in Fig.3
2) It collects also the link utilization information via the SNMP

protocol.





3)       Upon the application's requirement, for example, the bi-direction traffic assurance between SW1/SW2, the PCE will calculate the appropriate link path, which can be assigned to such traffic in dedicated mode, other normal traffic will not pass through such physical links.

4)       PCE will then send the key parameters to R1 and R7 respectively, to let R1 and R7 build another i/eBGP neighbor with R3, advertise the prefixes that owned by SW1/SW2.

5)       If the calculated dedicated path is via some physical links that belong to R1-R2-R4-R7, then PCE need also build the connection with these on-path routers, send some key parameters to them to build the path to the BGP next-hop via the address of physical links between R1/R2,R2/R4,R4/R7.

6)       If the assured traffic prefixes are changed and the total volume of assured traffic is not exceed the physical capacity of the previous end-to-end path, then the PCE need only change the related information on R1 and R7.

7)       If the volume of the assured traffic exceeds the capacity of previous calculated path, then PCE must recalculate the appropriate path to accommodate the exceeding traffic in some new end-to-end physical link. It then need to send some relevant key parameters to the on-path routers to build such path hop by hop.

```
                               +-----+
                     **********+PCE +*************
                     *         +--*-+            *
                     *           / * \            *
                     *             *              *
                 PCEP*             *BGP-LS/SNMP   *PCEP
                     *             *              *
                     *             *          \   * /
                   \ * /           *           \ */
                    \*/-----------R3--------------*
                     |                            |
                     |                            |
          SW1-------R1-------R5---------R6-------R7--------SW2
                     |        |          |        |
                     |        |          |        |
                     +-------R2---------R4--------+
```

Fig.3 PCE based solution for Multi-BGP deployment

## 7. PCEP extension for key parameter transformation.

In order to pass the key parameters to the underlying routers and keep the overall implementation as simple as possible, it is appropriate to extend the PCEP protocol to transfer the key parameters.
Based on the design philosophy of afore mentioned Multi-BGP deployment scenario, the key parameters should include the following information:

1) BGP peer address and assured prefixes that will be advertised via this BGP session
2) Static route information/Destination(BGP next hop) and Next Physical Link Address.

Once the router receive such information, it should establish the BGP session with the peer appointed in the PCEP message, advertises the prefixes that contained in the corresponding PCEP message, and build the end to end dedicated path hop by hop.

The detail format and the processing procedure of the above two extensions will be provided in another draft.

## 8. Security Considerations

TBD

## 9. IANA Considerations

TBD

## 10. Conclusions

TBD

## 11. References

### 11.1. Normative References

[RFC4655] Farrel, A., Vasseur, J.-P., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006,<http://www.rfc-editor.org/info/rfc4655>.

[RFC5440]Vasseur, JP., Ed., and JL. Le Roux, Ed., "Path





Computation Element (PCE) Communication Protocol
(PCEP)", RFC 5440, March 2009,
<http://www.rfc-editor.org/info/rfc5440>.

## 11.2. Informative References

TBD

## 12. Acknowledgments

TBD

## Authors' Addresses

Aijun Wang
China Telecom
Beiqijia Town, Changping District
Beijing,China

Email: wangaj@ctbri.com.cn