Network Working Group
Internet Draft
Intended status: Standard
Expires: June 18, 2021

L. Dunbar
H. Chen
Futurewei
Aijun Wang
China Telecom

December 18, 2020

OSPF extension for 5G Edge Computing Service
draft-dunbar-lsr-5g-edge-compute-ospf-ext-02

Abstract

This draft describes an OSPF extension to distribute the 5G Edge Computing App running status and environment, so that other routers in the 5G Local Data Network can make intelligent decision to optimize forwarding of flows from UEs. The goal is to improve latency and performance for 5G Edge Computing services.

Status of this Memo

This Internet-Draft will expire on April 7, 2021.

Copyright Notice

Table of Contents

1. Introduction

This document describes an OSPF extension to distribute the 5G Edge Computing App running status and environment, so that other routers in the 5G Local Data Network can make intelligent decision to optimize forwarding of flows from UEs. The goal is to improve latency and performance for 5G Edge Computing services.

1.1. 5G Edge Computing Background

As described in [5G-EC-Metrics], one Application can have multiple Application Servers hosted in different Edge Computing data centers that can be close in proximity. Those Edge Computing (mini) data centers are usually very close to, or co-located with, 5G base stations, with the goal to minimize latency and to optimize the user experience.

When a UE (User Equipment) initiates application packets using the destination address from a DNS reply or from its own cache, the packets from the UE are carried in a PDU session through 5G Core [5GC] to the 5G UPF-PSA (User Plan Function - PDU Session Anchor). The UPF-PSA decapsulates the 5G GTP outer header and forwards the packets from the UEs to the Ingress router of the Edge Computing (EC) Local Data Network (LDN). The IP based LDN for 5G EC is responsible for forwarding the packets to the intended destinations.

When the UE moves out of coverage of its current gNB (next generation Node B) (gNB1), handover procedures are initiated and the 5G SMF (Session Management Function) also selects a new UPF-PSA. The standard handover procedures described in 3GPP TS 23.501 and TS 23.502 are followed. When the handover process is complete, the UE has a new IP address and the IP point of attachment is to the new UPF-PSA. 5GC may maintain a path from the old UPF to new the UPF for a short period of time for SSC [Session and Service Continuity] mode 3 to make the handover process more seamless.

```
   +--+
   |UE|---\+---------+                 +------------------+
   +--+    |  5G     |    +---------+  |   S1: aa08::4450 |
   +--+    | Site +--++---+         +-----+               |
   |UE|----|  A   |PSA| Ra|         | R1 | S2: aa08::4460 |
   +--+    |      +---+---+         +-----+               |
  +---+    |         |  |           |  |   S3: aa08::4470 |
  |UE1|---/+---------+  |           |  +------------------+
  +---+                 |IP Network |       L-DN1
                        |(3GPP N6)  |
     |                  |           |  +------------------+
     | UE1              |           |  |  S1: aa08::4450  |
     | moves to         |          +-----+                |
     | Site B           |          | R3 | S2: aa08::4460  |
     v                  |          +-----+                |
                        |           |  |  S3: aa08::4470  |
                        |           |  +------------------+
                        |           |      L-DN3
   +--+                 |           |
   |UE|---\+---------+  |           |  +------------------+
   +--+    |  5G     |  |           |  |  S1: aa08::4450  |
   +--+    | Site +--++-+--+        +-----+               |
   |UE|----|  B   |PSA| Rb |        | R2 | S2: aa08::4460 |
   +--+    |      +--++----+        +-----+               |
   +--+    |         |  +-----------+  |  S3: aa08::4470  |
   |UE|---/+---------+                 +------------------+
   +--+                                     L-DN2
```

Figure 1: App Servers in different edge DCs

1.2. Problem#1: ANYCAST in 5G EC Environment

Increasingly, Anycast is used extensively by various application providers and CDNs because ANYCAST makes it possible to dynamically load balance across server locations based on network conditions.

Application Server location selection using Anycast address leverages the proximity information present in the network (routing) layer and eliminates the single point of failure and bottleneck at the DNS resolvers and application layer load balancers. Another benefit of using ANYCAST address is removing the dependency on how UEs get the IP addresses for their Applications. Some UEs (or clients) might use their cached IP addresses instead of querying DNS for extended period.

But, having multiple locations of the same ANYCAST address in 5G Edge Computing environment can be problematic because all those edge computing Data Centers can be close in proximity. There might be very little difference in the routing cost to reach the Application Servers in different Edge DCs, which can cause packets from one flow to be forwarded to different locations, resulting in service glitches.

1.3. Problem #2: Unbalanced Anycast Distribution due to UE Mobility

Another problem of using ANYCAST address for multiple Application Servers of the same application in 5G environment is that UEs' frequent moving from one 5G site to another, which can make it difficult to plan where the App Servers should be hosted. When one App server is heavily utilized, other App servers of the same address close-by can be very under-utilized. Since the condition can be short lived, it is difficult for the application controller to anticipate the move and adjust.

1.4. Problem 3: Application Server Relocation

When an Application Server is added to, moved, or deleted from a 5G Edge Computing Data Center, the routing protocol must propagate the changes to 5G PSA or the PSA adjacent

routers. After the change, the cost associated with the site [5G-EC-Metrics] might change as well.

Note: for the ease of description, the Edge Computing server, Application server, App server are used interchangeably throughout this document.

2. Conventions used in this document

A-ER: Egress Router to an Application Server, [A-ER] is used to describe the last router that the Application Server is attached. For 5G EC environment, the A-ER can be the gateway router to a (mini) Edge Computing Data Center.

Application Server: An application server is a physical or virtual server that host the software system for the application.

Application Server Location: Represent a cluster of servers at one location serving the same Application. One application may have a Layer 7 Load balancer, whose address(es) are reachable from external IP network, in front of a set of application servers. From IP network perspective, this whole group of servers are considered as the Application server at the location.

Edge Application Server: used interchangeably with Application Server throughout this document.

EC: Edge Computing

Edge Hosting Environment: An environment providing support required for Edge Application Server's execution.

NOTE: The above terminologies are the same as those used in 3GPP TR 23.758

Edge DC: Edge Data Center, which provides the Edge Computing Hosting Environment. It might be co-located with 5G Base Station and not only host 5G core functions, but also host frequently used Edge server instances.

gNB next generation Node B

L-DN: Local Data Network

PSA: PDU Session Anchor (UPF)

SSC: Session and Service Continuity

UE: User Equipment

UPF: User Plane Function

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

3. Solution Overview

From IP Layer, the Application Servers are identified by their IP (ANYCAST) addresses. The 5G Edge Computing controller or management system is aware of the ANYCAST addresses of the Applications that need optimized forwarding in 5G EC environment. The 5G Edge Computing controller or management system can configure the ACLs to filter out packets to or from those applications on routers, especially on the routers adjacent to the 5G PSA and the routers to which the Application Servers are directly attached.

The proposed solution is for the routers, a.k.a. Application Egress Router (A-ER), to which the Application Servers are attached to collect various measurements about the Servers' running status [5G-EC-Metrics] and advertise the metrics to other routers in 5G EC LDN (Local Data Network).

3.1. Flow Affinity to an ANYCAST server

When there are multiple Edge Computing Servers with the same ANYCAST address located in different mini Edge Computing Data Centers, each location is identified by its A-ER unicast address(es). To the routers in an LDN, achieving Flow Affinity is to send the packets of the same flow to the same A-ER's unicast address. A-ER, e.g. R1 in Figure 1, should deliver the packets destined towards the ANYCAST address to its directly attached server even through the same address is also reachable from other routers, unless the directly attached server is no longer reachable due to Server or Link failure.

Many commercial routers today support some forms of flow affinity to ensure packets belonging to one flow be forwarded along the same path. For servers with the same ANYCAST address attached to 3 different egress routers, routers supporting the flow affinity feature should forward the packets of one flow to the same egress router even if the cost towards the egress router changes.

Editor's note: for IPv6 traffic, Flow Affinity can be supported by the routers of the Local Data Network (LDN) forwarding the packets with the same Flow Label in the packets' IPv6 Header along the same path towards the same egress router. For IPv4 traffic, 5 tuples in the IPv4 header can be used to achieve the Flow Affinity

3.2. IP Layer Metrics to Gauge App Server Running Status

Most application don't expose their internal logics to network. Their communications are generally encrypted. Most of them do not even respond to PING or ICMP messages initiated by routers or network gears.

[5G-EC-Metrics] describes the IP Layer Metrics that can gauge the application servers running status and environment:

- IP-Layer Metric for App Server Load Measurement:
  The Load Measurement to an App Server is a weighted combination of the number of packets/bites to the App Server and the number of packets/bytes from the App Server which are collected by the A-ER to which the App Server is directly attached.
  The A-ER is configured with an ACL that can filter out the packets for the Application Server.
- Capacity Index
  Capacity Index is used to differentiate the running environment of the application server. Some data centers can have hundreds, or thousands, of servers behind an Application Server's App Layer Load Balancer that is reachable from external world. Other data centers can have very small number of servers for the application server. "Capacity Index", which is a numeric number, is used to represent the capacity of the application server in a specific location.
- Site preference index:
  [IPv6-StickyService] describes a scenario that some sites are more preferred for handling an application than others for flows from a specific UE.

For ease of description, those metrics, more may be added later, are called IP Layer App-Metrics throughout the document.

3.3. To Equalize traffic among Multiple ANYCAST Locations

The main benefit of using ANYCAST is to leverage the network layer information to balance the traffic among multiple Application Server locations.

For 5G Edge Computing environment, the routers in the LDN need to be notified of various measurement of the App Servers at different EC data centers to make the intelligent decision on where to forward the traffic for the application from UEs.

[5G-EC-Metrics] describes the algorithms that can be used by the routers in LDN to compare the cost to reach the App Servers between the Site-i or Site-j:

```
               Load-i * CP-j              Pref-j * Network-Delay-i
Cost-i=min(w *(----------------) + (1-w) *(------------------------))
              Load-j * CP-i              Pref-i * Network-Delay-j
```

Load-i: Load Index at Site-i, it is the weighted combination of the total packets or/and bytes sent to and received from the Application Server at Site-i during a fixed time period.

CP-i: capacity index at the site I, higher value means higher capacity.

Network Delay-i: Network latency measurement (RTT) to the A-ER that has the Application Server attached at the site-i.
Noted: Ingress nodes can easily measure RTT to all the egress nodes by existing IPPM metrics. But it is not so easy for ingress nodes to measure RTT to all the App Servers. Therefore, "Network-Delay-i", a.k.a. Network latency measurement (RTT), is between the Ingress nodes and egress nodes. The link cost between the egress nodes to their attached servers are embedded in the "capacity index".

Pref-i: Preference index for the site-i, higher value means higher preference.

w: Weight for load and site information, which is a value between 0 and 1. If smaller than 0.5, Network latency and the site Preference have more influence;

otherwise, Server load and its capacity have more influence.

3.4. Reason for using IGP Based Solution

Goal of the proposed OSPF extension is to propagate the IP Layer App-Metrics to other routers in the LDN.
Here are some of the benefits in using IGP to propagate the IP Layer App-Metrics:

- Intermediate routers can derive the aggregated cost to reach the Application Servers attached to different egress nodes, especially:
  - the path to the optimal egress node can be more accurate or shorter
  - convergence is shorter when there is any failure along the way towards the optimal ANYCAST server.
  - When there is any failure at the intended ANYCAST server, all the transient packets can be optimally forwarded to another App Server with the same ANYCAST address attached to a different egress router.
- Doesn't need ingress node to establish tunnels with egress nodes.

Of course, there are limitation of using IGP too, such as:

- The IGP approach might not suit well to 5G EC LDN operated by multiple ISPs networks.
  For LDN operated by multiple IPSs, BGP should be used. AppMetaData NLRI Path Attribute [5G-AppMetaData] describes the BGP UPDATE message to propagate IP Layer App-Metrics crossing multiple ISPs.

4. Aggregated Cost Computed by Egress routers

If all egress routers to which the App Servers are attached can be configured with a consistent algorithm to compute an aggregated cost that take into consideration of Load Measurement, Capacity value and Preference value. Then this aggregated cost can be considered as the Metric of the link to the App Server.

In this scenario, there is no protocol extension needed, but requires all egress routers to agree upon a consistent algorithm to compute the cost to the attached App servers.

4.1. OSPFv3 LSA to carry the Aggregated Cost

If the App Servers use IPv6 ANYCAST address, the aggregated cost computed by the egress routers can be encoded in the Metric field [the interface cost] of Intra-Area-Prefix-LSA specified by the Section 3.7 of the [RFC5340].

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     6 (Intra-Area Prefix)     |         TLV Length            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|          0    | Aggregated Cost to the App Server             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| PrefixLength  | PrefixOptions |             0                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                           Address Prefix                      |
|                               ...                             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2: Aggregated Cost to App Server

4.2. OSPFv2 LSA to carry the Aggregated Cost

For App Servers in IPv4 address, the Aggregated Cost can be encoded in the "Metric" field of the Stub Link LSA [Link type =3] specified by the Section 12.4 of the [RFC2328].

5. IP Layer App-Metrics Advertisements

This section describes the approaches to advertise the IP Layer App-Metrics to other nodes. Those approaches are needed for a scenario when it is not possible for all the egress routers to have a consistent algorithm to compute the aggregated cost, or the ingress routers need all the detailed IP Layer metrics for the App Servers for other purposes.

Under this scenario, the IP Layer Metrics to Gauge App Server running status are encoded in the sub-TLVs, which are specified in Section 5.3.

Since only a subset of routers within an IGP domain need to know those detailed metrics, it makes sense to use the OSPFv2 Extended Prefix Opaque LSA for IPv4 and OSPFv3 Extended LSA with Intra-Area-Prefix TLV to carry the detailed sub-TLVs. For routers that don't care about those metrics, they can ignore them very easily.

It worth noting that not all hosts (prefix) attached to an A-ER are ANYCAST servers that need network optimization among multiple locations. An A-ER only needs to advertise the App-Metrics for the ANYCAST servers that need the network to optimize the forwarding. Therefore, A-ER do not need to include App-Metrics for all attached prefixes.

Draft [draft-wang-lsr-passive-interface-attribute] introduces the Stub-Link TLV for OSPFv2/v3 and ISIS protocol respectively. Considering the interfaces on an edge router that connects to the App servers are normally configured as passive interfaces, these IP-layer App-metrics can also be advertised as the attributes of the passive/stub link. The associated prefixes can then be advertised in the "Stub-Link Prefix Sub-TLV" that defined in [draft-wang-lsr-passive-interface-attribute]. All the associated prefixes share the same characteristic of the link. Other link related sub-TLVs that defined in [RFC8920] can also be attached and applied to the calculation of path to the associated prefixes.

5.1. OSPFv3 Extension to carry the App-Metrics

For App Servers using IPv6, the OSPFv3 Extended LSA with the Intra-Area-Prefix Address TLV specified by the Section 3.7 of RFC8362 can be used to carry the App-Metrics for the attached App Servers.

```
  0                   1                   2                   3
  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |7 (IPv6 Local-Local Address)   |              Length           |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |            IPv6 AppServer (ANYCAST) address                   |
 ~                                                               ~
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |            Load measurement sub-TLV                           |
 ~                                                               ~
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |            Capability sub-TLV                                 |
 ~                                                               ~
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |             Preference sub-TLV                                |
 ~                                                               ~
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 3: IPv6 App Server App-Metrics Encoding

5.2. OSPFv2 Extension to advertise the IP Layer App-Metrics

For App Servers using IPv4 addresses, the OSPFv2 Extended Prefix Opaque LSA with the extended Prefix TLV can be used to carry the App Metrics sub-TLVs, as specified by the Section 2.1 [RFC7684].

Here is the proposed encoding:

```
  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 | Type                          | Length                        |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 | Route Type    | Prefix Length | AF            | Flags         |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 | Address Prefix (variable)                                     |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 | Load Measurement Sub-TLV                                      |
 ~                                                               ~
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 | capacity Index Sub-TLV                                        |
 ~                                                               ~
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 | Site Preference Sub-TLV                                       |
 ~                                                               ~
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 4: App-Metrix Sub-TLVs in OSPFv2 Extended Prefix TLV

5.3. IP Layer App-Metrics Sub-TLVs

Two types of Load Measurement Sub-TLVs are specified:

a) The Aggregated Load Index based on weighted combination of the collected measurements;
b) The raw measurements of packets/bytes to/from the App Server address. The raw measurement is useful when the egress routers cannot be configured with a consistent algorithm to compute the aggregated load index or the raw measurements are needed by a central analytic system.

The Aggregated Load Index Sub-TLV has the following format:

```
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Type (TBD2)           |               Length          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                   Measurement Period                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           Aggregated Load Index to reach the App Server       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 5: Aggregated Load Index Sub-TLV

Type=TBD2 (to be assigned by IANA) indicates that the sub-TLV carries the Aggregated Load Measurement Index derived from the Weighted combination of bytes/packets sent to/received from the App server:

Index=w1*ToPackets+w2*FromPackes+w3*ToBytes+w4*FromBytes

Where wi is a value between 0 and 1; w1+ w2+ w3+ w4 = 1.

The Raw Load Measurement sub-TLV has the following format:

```
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |           Type (TBD3)         |               Length          |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                   Measurement Period                          |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |           total number of packets to the AppServer            |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |           total number of packets from the AppServer          |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |           total number of bytes to the AppServer              |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |           total number of bytes from the AppServer            |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 6: Raw Load Measurement Sub-TLV

Type= TBD3 (to be assigned by IANA) indicates that the sub-TLV carries the Raw measurements of packets/bytes to/from the App Server ANYCAST address.

Measurement Period: A user specified period in seconds, default is 3600 seconds.

The Capacity Index sub-TLV has the following format:

```
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |           Type (TBD3)         |               Length          |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                   Capacity Index                              |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 7: Capacity Index Sub-TLV

The Preference Index sub-TLV has the following format:

```
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           Type (TBD4)         |              Length           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                   Preference Index                            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 8: Preference Index Sub-TLV

Note: "Capacity Index" and "Site preference" can be more stable for each site. If those values are configured to nodes, they might not need to be included in every OSPF LSA.

6. Soft Anchoring of an ANYCAST Flow

This section describes a solution that can anchor an application flow from a UE to a specific ANYCAST Server location even when the UE moves from one 5G Site to another. This is called "Sticky Service" in the 3GPP Edge Computing specification.

Lets' assume one application "App.net" is instantiated on four servers that are attached to four different A-ERs: R1, R2, R3, and R4 respectively. The "App.net" needs to be Sticky means that the packets to the "App.net" from UE-1 needs to stick with one server, say the "App.net" Server attached to R1, even when the UE moves from one 5G site to another. When there is failure at R1 or the Application Server attached to R1, the packets of the flow "App.net" from UE-1 need to be forwarded to the Application Server attached to R2, R3, or R4.

We call this kind of sticky service "Soft Anchoring", meaning that anchoring to the site of R1 is preferred, but other sites can be chosen when the preferred site encounters failure.

Here are the steps to achieving "Soft Anchoring":

- Assign a group of ANYCAST addresses to one application. For example, "App.net" is assigned with

4 ANYCAST addresses, L1, L2, L3, and L4. L1/L2/L3/L4 represents the location preferred ANYCAST addresses.

- For the App.net Server attached to a router, the router has four Stub links to the same Server, L1, L2, L3, and L4 respectively. The cost to L1, L2, L3 and L4 is assigned differently for different routers. For example,
    - o When attached to R1, the L1 has the lowest cost, say 10, when attached to R2, R3, and R4, the L1 can have higher cost, say 30.
    - o ANYCAST L2 has the lowest cost when attached to R2, higher cost when attached to R1, R3, R4 respectively.
    - o ANYCAST L3 has the lowest cost when attached to R3, higher cost when attached to R1, R2, R4 respectively, and
    - o ANYCAST L4 has the lowest cost when attached to R4, higher cost when attached to R1, R2, R3 respectively
- When a UE queries for the "App.net" for the first time, the DNS replies the location preferred ANYCAST address, say L1, based on where the query is initiated.
- When the UE moves from one 5G site-A to Site-B, UE continues sending packets of the "App.net" to ANYCAST address L1. The routers will continue sending packets to R1 because the total cost for the App.net instance for ANYCAST L1 is lowest at R1. If any failure occurs making R1 not reachable, the packets of the "App.net" from UE-1 will be sent to R2, R3, or R4 (depending on the total cost to reach each of them).

If the Application Server supports the HTTP redirect, more optimal forwarding can be achieved.

- When a UE queries for the "App.net" for the first time, the global DNS replies the ANYCAST address G1, which has the same cost regardless where the Application Servers are attached.

- When the UE initiates the communication to G1, the packets from the UE will be sent to the Application Server that has the lowest cost, say the Server attached to R1. The Application server is instructed with HTTPs Redirect to respond back a location specific URL, say App.net-Loc1. The client on the UE will query the DNS for App.net-Loc1 and get the response of ANYCAST L1. The subsequent packets from the UE-1 for App.net are sent to L1.

7. Manageability Considerations

To be added.

8. Security Considerations

To be added.

9. IANA Considerations

The following Sub-TLV types need to be added by IANA to OSPFv4 Extended-LSA Sub-TLVs and OSPFv2 Extended Link Opaque LSA TLVs Registry.

- Aggregated Load Index Sub-TLV type
- Raw Load Measurement Sub-TLV type
- Capacity Index Sub-TLV type
- Preference Index Sub-TLV type

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.

[RFC2328] J. Moy, "OSPF Version 2", RFC 2328, April 1998.

[RFC7684] P. Psenak, et al, "OSPFv2 Prefix/Link Attribute Advertisement", RFC 7684, Nov. 2015.

[RFC8200] S. Deering R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", July 2017.

[RFC8326] A. Lindem, et al, "OSPFv3 Link State advertisement (LSA0 Extensibility", RFC 8362, April 2018.

10.2. Informative References

[3GPP-EdgeComputing] 3GPP TR 23.748, "3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Study on enhancement of support for Edge Computing in 5G Core network (5GC)", Release 17 work in progress, Aug 2020.

[5G-AppMetaData] L. Dunbar, K. Majumdar, H. Wang, "BGP NLRI App Meta Data for 5G Edge Computing Service", draft-dunbar-idr-5g-edge-compute-app-meta-data-01, work-in-progress, Nov 2020.

[5G-EC-Metrics] L. Dunbar, H. Song, J. Kaippallimalil, "IP Layer Metrics for 5G Edge Computing Service", draft-dunbar-ippm-5g-edge-compute-ip-layer-metrics-01, work-in-progress, Nov 2020.

[5G-StickyService] L. Dunbar, J. Kaippallimalil, "IPv6 Solution for 5G Edge Computing Sticky Service", draft-dunbar-6man-5g-ec-sticky-service-00, work-in-progress, Oct 2020.

[RFC5521] P. Mohapatra, E. Rosen, "The BGP Encapsulation Subsequent Address Family Identifier (SAFI) and the BGP Tunnel Encapsulation Attribute", April 2009.

[BGP-SDWAN-Port] L. Dunbar, H. Wang, W. Hao, "BGP Extension for SDWAN Overlay Networks", draft-dunbar-idr-bgp-sdwan-overlay-ext-03, work-in-progress, Nov 2018.

[SDWAN-EDGE-Discovery] L. Dunbar, S. Hares, R. Raszuk, K. Majumdar, "BGP UPDATE for SDWAN Edge Discovery", draft-dunbar-idr-sdwan-edge-discovery-00, work-in-progress, July 2020.

[Tunnel-Encap] E. Rosen, et al "The BGP Tunnel Encapsulation Attribute", draft-ietf-idr-tunnel-encaps-10, Aug 2018.

11. Acknowledgments

Acknowledgements to Acee Lindem, Jeff Tantsura, and Donald Eastlake for their review and suggestions.

This document was prepared using 2-Word-v2.0.template.dot.

Authors' Addresses

Linda Dunbar
Futurewei
Email: ldunbar@futurewei.com

Huaimo Chen
Futurewei
Email: huaimo.chen@futurewei.com

Aijun Wang
China Telecom
Email: wangaj3@chinatelecom.cn