Network Working Group
Internet Draft
Intended status: Standard
Expires: May 8, 2022

L. Dunbar
H. Chen
Futurewei
Aijun Wang
China Telecom

November 8, 2021

IGP Extension for 5G Edge Computing Service
draft-dunbar-lsr-5g-edge-compute-02

Abstract

Routers in 5G Local Data Network (LDN) can use additional site-costs, preference, and other application related metrics on top of the network condition to compute constraint-based SPF within the 5G LDN to enhance performance for selected services. This draft describes those application server related metrics to be used in Flexible Algorithms.

Copyright Notice

Table of Contents

1. Introduction... 3
  1.1. Unbalanced Distribution due to UE Mobility........ 3
  1.2. ANYCAST in 5G EC Environment...................... 4
2. Conventions used in this document..................... 4
3. Solution Overview..................................... 6
4. FAD sub-TLV for Constrained SPF with Site Cost......... 7
  4.1. New Flags added to FAD Flags Sub-TLV.............. 7
5. "Site-Cost" Advertisement in OSPF..................... 8
  5.1. OSPFv3 LSA to Carry the Aggregated Cost........... 8
  5.2. OSPFv2 LSA to Carry the Aggregated Cost........... 8
  5.3. AppMetaData Sub-TLV in OSPF....................... 9
    5.3.1. OSPFv2 Extended Prefix Opaque LSA Extension.. 9
    5.3.2. OSPFv3 Extended LSA to carry the AppMetaData 10
6. "Site-Cost" Advertisement in IS-IS................... 10
  6.1. Aggregated Cost Advertisement in IS-IS........... 10
  6.2. AppMetaData Advertisement in IS-IS............... 11
7. IP Layer App-Metrics (AppMetaData) SubSub-TLVs........ 12
8. AppMetaData Metric Advertisement..................... 14
9. Alternative method for Distributing Aggregated Cost... 15
10. Manageability Considerations........................ 15

11. Security Considerations.............................. 15
12. IANA Considerations.................................. 15
13. References... 16
13.1. Normative References............................. 16
13.2. Informative References........................... 17
14. Appendix:5G Edge Computing Background................ 18
15. 5G EC LDN Characteristics for the Constraint SPF..... 19
15.1. IP Layer Metrics to Gauge EC Server Running Status 19
15.2. App Metrics Constrained Shortest Path First..... 21
15.3. Reason for using IGP Based Solution............. 22
15.4. Flow Affinity to an ANYCAST server.............. 22
16. Acknowledgments...................................... 23

1. Introduction

In 5G Edge Computing (EC) environment, it is common for an application that needs low latency to be instantiated on multiple servers close in proximity to UEs (User Equipment). When those multiple server instances share one IP address (ANYCAST), the transient network and load conditions can be incorporated in selecting an optimal path among server instances and UEs.

Flexible algorithms provide mechanisms for topologies to use different IGP path algorithms. This draft describes using Flexible Algorithms [LSR-FlexAlgo] to indicate the desired constrained SPF behavior for a subset of prefixes, in addition to the encodings for advertising the IP Layer App related metrics that can impact application servers' performance.

1.1. Unbalanced Distribution due to UE Mobility

UEs' frequent moving from one 5G site to another 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 under-utilized. The difference in the routing distance to reach multiple Application Servers might be relatively small. The traffic load at the router where the App Server is attached and the site capacity, when combined, can be more significant from the latency and performance perspective.

Since the condition can be short-lived, it is difficult for the application controller to anticipate the moving and adjusting.

1.2. ANYCAST in 5G EC Environment

ANYCAST makes it possible to load balance across server locations based on network conditions dynamically. With multiple servers having the same IP address, it eliminates the single point of failure and bottleneck at the 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 stale cached IP addresses for an extended period.

But having multiple locations of the same IP address in the 5G Edge Computing environment can be problematic because all those locations can be close in proximity. There might be very small difference in the routing distance to reach an Application Server attached to a different edge router, which can cause packets from one flow to be forwarded to different locations, resulting in service glitches.

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

2. Conventions used in this document

A-ER: Egress Edge 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 hosts 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 an external IP network, in front of a set of application servers. From IP network perspective, this whole group of servers is 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 the 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

LDN: 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

The proposed solution is for the egress edge router (A-ER) with the EC Servers directly attached to advertise the "Site-Cost" [Section 15.1] via IP prefix reachability TLV associated with the (anycast) prefix and use the Flexible algorithms [LSR-FlexAlgo] to advertise the desired constrained SPF behavior, so that constrained IGP path can be computed as desired.

There are two types of "Site-Cost":

a) The IP Layer App related metrics, such as the Load Measurement, the Capacity Index, and the Preference Index that are collected by the egress routers for an attached EC server (i.e., ANYCAST prefix), as described in Section 15.1.

b) The aggregated cost associated with an EC server (i.e., ANYCAST prefix).

The aggregated cost is computed based on the Load Measurement, the Capacity Index, the Preference Index, and other constraints by a consistent algorithm across all A-ERs.

This document, "AppMetaData" refers to the encoding of the IP Layer App related metrics carried by IP Prefix Reachability TLVs.

The solution assumes that the 5G EC controller or management system is aware of the EC ANYCAST addresses that need optimized forwarding. To minimize the processing, only the addresses that match with the ACLs configured by the 5G EC controller will have their Site-Cost collected and advertised.

4. FAD sub-TLV for Constrained SPF with Site Cost

In order for Site-Cost to be considered in IGP path computation, a new Calc-Type, Metric Type and FAD Flag need to be indicated in the FAD sub-TLV [LSR-FlexAlgo].

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Type        | Length        |Flex-Algorithm | Metric-Type   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Calc-Type     | Priority      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               Constrained SPF Flags Sub-TLVs                  |
+                                                               +
|                     ...                                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1: IS-IS Flex Algo Definition Sub-TLV

Flex-Algorithm: SPF.

Metric-Type:
  A new value to be assigned by IANA to indicate the "Site-Cost" included in computing the constrained SPF.

Calc-Type:
  A value chosen by the IGP operator to indicate a specific constrained SPF algorithm that takes the "Site-Cost" into the SPF computation across the routers in the 5G LDN.

4.1. New Flags added to FAD Flags Sub-TLV

New flags are added to indicate the constrained SPF compute methods in the IS-IS FAD Flags Sub-TLV (Section 6.4 of [LSR-FlexAlgo]) or the OSPF FAD Flags Sub-TLV (Section 7.4 of [[LSR-FlexAlgo]), respectively.

Flags:

```
0 1 2 3 4 5 6 7...
+-+-+-+-+-+-+-+-+...
|M|T|A| ...
+-+-+-+-+-+-+-+-+...
```

T-flag: Site-Cost Metrics is only used as tiebreaker.

A-flag: Site-Cost Metrics are the additional metrics for the Calc-Type indicated.

5. "Site-Cost" Advertisement in OSPF

5.1. OSPFv3 LSA to Carry the Aggregated Cost

For EC servers using IPv6 address, the aggregated cost computed by the A-ERs can be encoded in the Metric field [the interface cost] of Intra-Area-Prefix-LSA specified by 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 EC Server             |
  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
  | PrefixLength  | PrefixOptions |             0                 |
  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
  |                           Address Prefix                      |
  |                               ...                             |
  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2: Aggregated Cost to EC Server

5.2. OSPFv2 LSA to Carry the Aggregated Cost

For EC servers in IPv4 address, the aggregated cost can be encoded in the "Metric" field of the Stub Link LSA [Link type =3] specified by Section 12.4 of the [RFC2328].

5.3. AppMetaData Sub-TLV in OSPF

5.3.1. OSPFv2 Extended Prefix Opaque LSA Extension

For IPv4 network, IP layer App related metrics, AppMetaData, can be carried by the OSPFv2 Extended Prefix Opaque LSA with the extended Prefix TLV [RFC7684].

```
 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                                       |
~                                                               ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

5.3.2. OSPFv3 Extended LSA to carry the AppMetaData

For IPv6, OSPFv3 Extended LSA with the Intra-Area-Prefix Address TLV [RFC8362] can be extended to carry the IP Layer App related Metrics as below:

```
    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
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |AppMetaDataType                | Length                        |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |            IPv6 AppServer (ANYCAST) address                   |
   ~                                                               ~
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |            Load measurement SubSub-TLV                        |
   ~                                                               ~
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |            Capability SubSub-TLV                              |
   ~                                                               ~
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |             Preference SubSub-TLV                             |
   ~                                                               ~
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 3: AppMetaData sub TLV

AppMetaData Type (TBD1): AppMetaData-OSPF-IPv6.

6. "Site-Cost" Advertisement in IS-IS

6.1. Aggregated Cost Advertisement in IS-IS

Egress routers can append the Aggregated Cost to the IP Reachability TLVs.

```
   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
  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
  |AggCostSubTLV  | Length        | AggCost to the App Server     |
  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
  | PrefixLength  | PrefixOptions |             0                 |
  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
  |                           Address Prefix                      |
  |                               ...                             |
  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 4: Aggregated cost to EC server

6.2. AppMetaData Advertisement in IS-IS

The IP Layer App related Metrics are encoded in the AppMetaData Advertisement Sub-TLV carried by IP Prefix Reachability TLV 128, 130, or 135.

Here is the AppMetaData Sub-TLV encoding for IS-IS:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|AppMetaDataType| Length        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            IPv6 or IPv4  AppServer (ANYCAST) address          |
~                                                               ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            Load measurement SubSub-TLV                        |
~                                                               ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            Capability SubSub-TLV                              |
~                                                               ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             Preference SubSub-TLV                             |
~                                                               ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 5: AppMetaData sub TLV

AppMetaData Type (TBD1): ISIS-IPv4 or ISIS-IPv6.

7. IP Layer App-Metrics (AppMetaData) SubSub-TLVs

Two types of Load Measurement SubSub-TLVs are specified:

a) The Aggregated Load Index based on a 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 edge 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 6: Aggregated Load Index Sub-TLV

Type=TBD2 (to be assigned by IANA) indicates that the sub-TLV carries the 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 7: 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 (TBD4)         |               Length          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                   Capacity Index                              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 8: 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 (TBD5)         |             Length            |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                   Preference Index                            |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
            Figure 9: Preference Index Sub-TLV
```

Note: "Capacity Index" and "Site preference" can be different for different attached server addresses. For Figure 1, the address S1 can have higher Site Preference when attached to R1 than R2.

8. AppMetaData Metric Advertisement

With Flex-Algorithm, the network administrator can define a function that compute the SPF with consideration of the AppMetaData metrics advertised by the routers to which the EC servers are directly attached.

This document defines a new standard metric type, AppMetaData, for this purpose. The AppMetaData Metric MAY be advertised in the Generic Metric sub-TLV with the metric type set to "AppMetaData Metric". ISIS and OSPF will advertise this new type of metric in their link advertisements. AppMetaData metric is a prefix attribute and for advertisement and processing of this attribute for Flex-algorithm purposes, MUST follow the section 12 of [I-D.ietf-lsr-flex-algo]

Flex-Algorithm uses this metric type by specifying the AppMetaData as the metric type in a FAD TLV. A FAD TLV may also specify an automatic computation of the AppMetaData metric based on a links advertised bandwidth. An explicit advertisement of a link's AppMetaData metric using the Generic Metric sub-TLV overrides this automatic computation. The automatic AppMetaData metric calculation sub-TLVs are advertised in FAD TLV and these parameters are applicable to applications such as Flex-algorithm that make use of the FAD TLV.

9. Alternative method for Distributing Aggregated Cost

Section 7 and Section 8 demonstrate different ways for OSPFv2, OSPFv3, and ISIS to propagate the aggregated cost. It would be better if the aggregated cost could be advertised the same way, regardless of OSPFv2, OSPFv3, or ISIS.

Draft [draft-wang-lsr-stub-link-attributes] introduces the Stub-Link TLV for OSPFv2/v3 and ISIS protocol respectively. Considering the interfaces on an edge router that connects to the EC 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 TLV" that is defined in [draft-wang-lsr-stub-link-attributes]. All the associated prefixes share the same characteristic of the link. Other link related sub-TLVs defined in [RFC8920] can also be attached and applied to the calculation of path to the associated prefixes."

Section 6 for the advertisement of AppMetaData Metric can also utilize the Stub-Link TLV that defined in [draft-wang-lsr-stub-link-attributes]

10. Manageability Considerations

To be added.

11. Security Considerations

To be added.

12. IANA Considerations

The following Sub-TLV types need to be added by IANA to FlexAlgo.

- AppMetaData Type for ISIS or OSPF (TBD1): IPv4 or IPv6

The following SubSub-TLV types need to be added by IANA, to be included in FAD sub-TLV, ISIS Extended-LSA Sub-TLVs, or OSPFv2 Extended Link Opaque LSA TLVs Registry.

- Aggregated Load Index Sub-TLV type (TBD2)
- Raw Load Measurement Sub-TLV type (TBD3)
- Capacity Index Sub-TLV type (TBD4)
- Preference Index Sub-TLV type (TBD5)

13. References

13.1. 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.

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

[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.

[RFC9012] E. Rosen, et al "The BGP Tunnel Encapsulation Attribute", April 2021.

13.2. 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-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.

[BGP-5G-AppMetaData] L. Dunbar, K. Majumdar, H. Wang, "BGP App Metadata for 5G Edge Computing Service", draft-dunbar-idr-5g-edge-compute-app-meta-data-03, work-in-progress, Sept 2020.

[LSR-Flex-Algo] P. Psenak, et al, "IGP Flexible Algorithm", draft-ietf-lsr-flex-algo-17, July 2021.

[LSR-Flex-Algo-BW] S. Hegde, et al, "Flexible Algorithms: Bandwidth, Delay, Metrics and Constraints", draft-ietf-lsr-flex-algo-bw-con-01, July 2021.

[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.

14. Appendix:5G Edge Computing Background

The network connecting the 5G EC servers with the 5G Base stations consists of a small number of dedicated routers that form the 5G Local Data Network (LDN) to enhance the performance of the EC services.

When a User Equipment (UE) initiates application packets using the destination address from a DNS reply or its 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, performs NAT sometimes, before handing the packets from the UEs to the adjacent router, also known as the ingress router to the EC LDN, which 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), the handover procedure is initiated, which includes the 5G SMF (Session Management Function) selecting a new UPF-PSA [3GPP TS 23.501 and TS 23.502]. When the handover process is complete, the IP point of attachment is to the new UPF-PSA. The UE's IP address stays the same unless moving to different operator domain. 5GC may maintain a path from the old UPF to the new UPF for a short 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 10: App Servers in different edge DCs

15. 5G EC LDN Characteristics for the Constraint SPF

15.1. IP Layer Metrics to Gauge EC Server Running Status

Most applications do not expose their internal logic to the network. Their communications are generally encrypted. Most of them do not even respond to PING or ICMP messages initiated by routers.

Here are some IP Layer App related Metrics that can gauge the servers running status and environment:

- Capacity Index:
  a numeric number, configured on all A-ERs in the domain consistently, is used to represent the capacity of an EC server attached to an A-ER. The IP addresses exposed to the A-ER can be the App Layer Load balancers that have many instances attached. At other sites, the IP address exposed is the server itself.
- Site preference index:
  Is used to describe some sites are more preferred than others. For example, a site with less leasing cost has a higher preference value. Note: the preference value is configured on all A-ERs in the domain consistently by the Domain Controller.

- Load Measurement for gauging the load of the attached prefix (i.e., EC Server):
  The Load Measurement for an EC Server is a weighted combination of the number of packets/bytes to the EC server (i.e., its IP address) and the number of packets/bytes from the EC server. The Load Measurement are collected by the A-ER that has the EC Server directly attached.

  An A-ER only collects those measurement for the prefixes instructed by the Domain Controller.

For ease of description, those metrics with more to be added later are called IP Layer App Metrics (or Site-Cost) throughout the document.

15.2. App Metrics Constrained Shortest Path First

The main benefit of using ANYCAST is to leverage the network layer information to balance the traffic among multiple locations of one application server.

For the 5G EC environment, the routers in the LDN need to take consideration of various measurements of the EC servers attached to each A-ER in addition to TE metrics to compute ECMP paths to the servers.

Here is one algorithm that computes the cost to reach the App Servers attached to Site-i relative to another site, say Site-j. When the reference site, Site-j, is plugged in the formula, the cost is 1. So, if the formula returns a value less than 1, the cost to reach Site-i is less than reaching the reference site (Site-j).

$$\text{Cost-i} = \left(w * \left(\frac{\text{CP-j} * \text{Load-i}}{\text{CP-i} * \text{Load-j}}\right) + (1-w) * \left(\frac{\text{Pref-j} * \text{Network-Delay-i}}{\text{Pref-i} * \text{Network-Delay-j}}\right)\right)$$

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 site i, a 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 edge 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 and egress edge nodes. The cost for the egress edge nodes to reach to their attached servers is embedded in the "capacity index".

Pref-i: Preference index for site-i, a higher value means higher preference. Preference can be derived from the total path cost to reach the A-ER [RFC5305], as calculated below: 1/(total-path-cost).

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.

15.3. Reason for using IGP Based Solution

Here are some benefits of using IGP to propagate the IP Layer App-Metrics:

- Intermediate routers can derive the aggregated cost to reach the EC Servers attached to different egress edge nodes, especially:
  - The path to the optimal egress edge 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 packets in transit can be optimally forwarded to another App Server attached to a different egress edge router.
- Doesn't need the ingress nodes to establish tunnels with egress edge nodes.

There are limitations of using IGP too, such as:

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

15.4. Flow Affinity to an ANYCAST server

When multiple servers with the same IP address (ANYCAST) are attached to different A-ERs, Flow Affinity means routers sending the packets of the same flow to the same A-ER even if the cost towards the A-ER is no longer optimal.

Many commercial routers support some forms of flow affinity to ensure packets belonging to one flow be forwarded along the same path.

Editor's note: for IPv6 traffic, Flow Affinity can be achieved by routers forwarding the packets with the same

Flow Label extracted from the IPv6 Header along the same path.

16. Acknowledgments

Acknowledgements to Peter Psenak, Acee Lindem, Shraddha Hegde, Tony Li, Gyan Mishra, 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