Network Working Group
Internet Draft
Intended status: Standard
Expires: September 7, 2022

L. Dunbar
J. Kaippallimalil
Futurewei

March 7, 2022

IPv6 Solution for 5G Edge Computing Sticky Service
draft-dunbar-6man-5g-edge-compute-sticky-service-06

Abstract

This draft describes the IPv6-based solutions that can stick an application flow originated from a mobile device to the same ANYCAST server location when the mobile device moves from one 5G cell site to another.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79. This document may not be modified, and derivative works of it may not be created, except to publish it as an RFC and to translate it into languages other than English.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html

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

Table of Contents

1. Introduction

1.1. 5G Edge Computing Background

As described in [5G-EC-Metrics], one application in 5G Edge Computing environment can have multiple application servers hosted in different Edge Computing data centers close in proximity. Those Edge Computing (mini) data centers are usually very close to, or co-located with, 5G base stations, to minimize latency and optimize the performances.

When a mobile device sends packets using the destination address from a DNS reply or its own cache, the packets are carried by a GTP tunnel from the 5G eNB 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 mobile devices to the Ingress router of the Edge Computing (EC) Local Data Network (LDN). The LDN for 5G EC, the IP Networks, is responsible for forwarding the packets to the intended destinations.

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

1.2. 5G Edge Computing Network Properties

In this document, 5G Edge Computing Network refers to multiple Local IP Data Networks (LDN) in one region that interconnect the Edge Computing mini-data centers. Those IP LDN networks are the N6 interfaces from 3GPP 5G perspective.

The ingress routers to the 5G Edge Computing Network are directly connected to 5G UPFs. The egress routers to the 5G Edge Computing Network are the routers that have a direct link to the Edge Computing servers. The servers and the egress routers are co-located. Some of those mini Edge Computing Data centers may have Virtual switches or Top of Rack switches between the egress routers and the servers. But transmission delay between the egress routers and the Edge Computing servers is very small, which is considered negligible in this document.

When multiple Edge Computing Servers attached to one App Layer Load Balancer, only the App Layer Load Balancer address is visible to the 5G Edge Computing Network. How the App Layer Load balancer manages the individual servers is out of the scope of the document.

The Edge Computer Services are registered services that need to utilize the network topology and balance among multiple mini Edge Computing Data Centers with the same ANYCAST address. Majority services are not registered 5G Edge Computing Services.

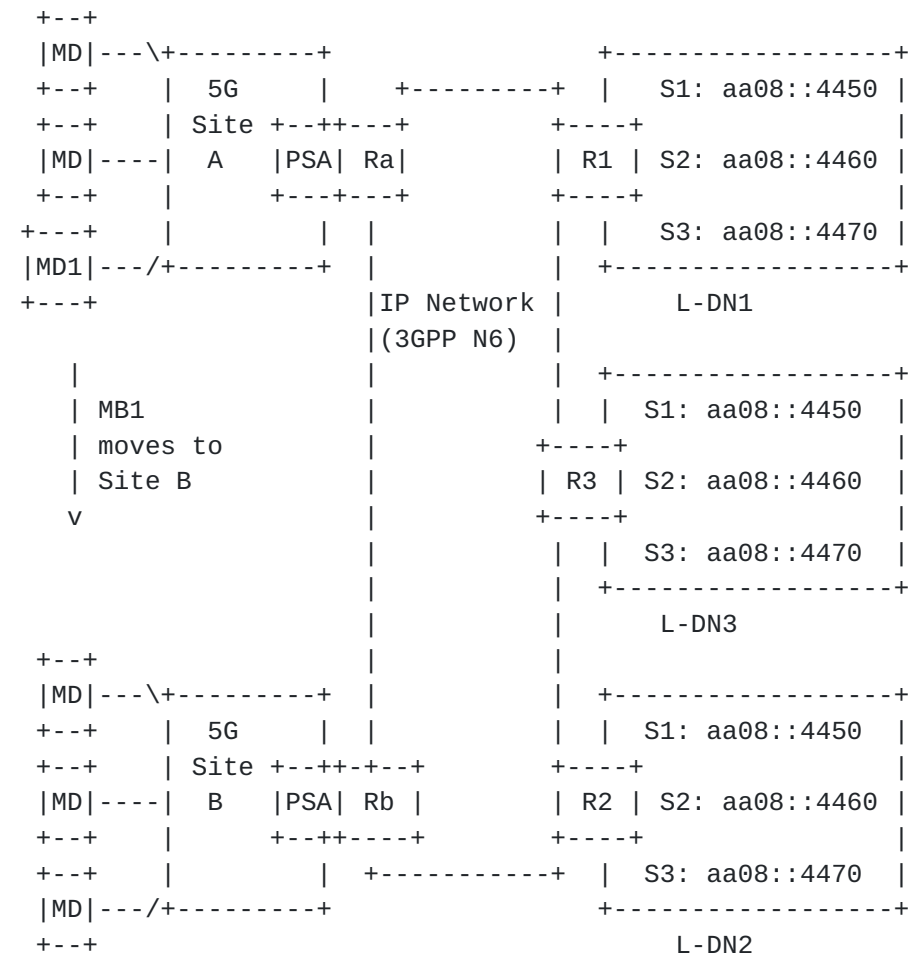


Figure 1: App Servers in different edge DCs

1.3. Problem #1: Discovery of Edge Application Server

Key Issue #1 identified by 3GPP Edge Computing Study [TR 23.748] is that one application service might be served by multiple Edge Application Servers typically deployed in different sites. These multiple Edge Application Server instances that host same content or service may use a single IP address (anycast address) or different IP addresses.

Key Issue #2 identified by 3GPP Edge Computing Study [TR 23.748] is Edge server relocation.

Application Server discovery and relocation can be achieved by running IGP/BGP routing protocols among the routers in LDN.

Increasingly, ANYCAST is used extensively by various application providers because it is possible to dynamically load balance across multiple locations of the same address based on network conditions. When multiple servers in different locations have the same IP address (ANYCAST), the routers see multiple paths to the IP address. The IGP/BGP routing protocols can inform all the nodes where the servers are and when servers move to new locations.

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 mobile devices that use their cached IP addresses instead of querying DNS when they move to a new location.

However, having multiple locations for the same ANYCAST address in the 5G Edge Computing environment can be problematic because all those edge computing Data Centers can be close in proximity. There might not be any difference in the routing cost to reach the Application Servers in different Edge DCs. The same routing cost to multiple locations can cause packets from one flow to be forwarded to different locations, which can cause service glitches.

1.4. Problem #2: sticking to original App Server

When a mobile device moves to a new location but continues the same application flow, the router connected to the new UPF might choose the App Server closer to the new location. As shown in the figure below, when the MD1 in 5G-site-A moves to the 5G-Site-B, the router directly connected to 5G PSA2 might forward the packets destined towards the S1: aa08::4450 to the server located in L-DN2 because L-DN2 has the lowest cost based on routing. This is not the desired behavior for some services, which are called Sticky Services in this document.

Even for some advanced applications with built-in mechanisms to re-sync the communications at the application layer after switching to a new location, service glitches are often experienced.

It worth noting that not all services need to be sticky. We assume only a subset of services are, and the Network is informed of the services that need to be sticky, usually by requests from application developers or controllers.

This document describes an IPv6-based network layer solution to stick the packets belonging to the same flow of a mobile device to its original App Server location after the mobile device is anchored to a new nearby UPF-PSA.

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

2. Conventions used in this document

APN6 Application aware network using IPv6. The term "Application" has very broad meanings. In this document the term "Application" refers to any applications that use ANYCAST servers in the 5G Edge Computing Environment.

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 or very close to a 5G Base Station.

gNB next generation Node B

L-DN: Local Data Network

MD: Mobile Device, which is the same as the UE (User Equipment) used in 3GPP. The term "mobile device" is used instead of UE to emphasize on sticking services originated from the devices that are mobile to same server.

PSA: PDU Session Anchor (UPF)

SSC: Session and Service Continuity

UE: User Equipment. UE is same as a mobile device in this document.

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. Stick a Flow to an ANYCAST Server

When servers attached to different egress routers are assigned with the same IP address, the routers in the LDN see multiple paths to the IP address. The Egress nodes' unicast addresses are the Next Hops (i.e., R1, R2, and R3) to reach the Edge Computing server ANYCAST address.

The routers choose the lowest cost path. [5G-EC-OSPF-EXT] and [5G-EC-BGP-EXT] describe the OSPF and BGP extension to propagate additional costs about the site where the servers are located so that the site costs can be incorporated into the path computation.

Flow sticking to one server is not the same as flow nailing down to the same server. When the network cost is significantly increased, such as the mobile device moving to a very far away location or the extreme case of link failure to the original server, another server with the same IP address is selected.

The Flow Affinity feature, which most commercial routers support today, can ensure packets belonging to one flow be forwarded along the same path to the same egress router, which then delivers the packets to the attached server.

Editor's note: for IPv6 traffic, Flow Affinity can be supported by the Local Data Network (LDN) routers 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.

When a UE moves to a different cell site, the packets from the UE might enter the 5G LDN from a different UPF. Suppose the handover to the new cell site is in the middle of a flow from the UE. In that case, the new ingress router directly connected to the new UPF needs to have the original egress router information to stick the flow from the UE to the original egress router. The original egress router is called Sticky Egress throughout this document.

4. Sticky flow for QUIC based Applications

For applications using QUIC transport protocol, ANYCAST stickiness are supported natively. During the initial handshake, QUIC servers can provide a "preferred address" (IP or IPv6 and port number), and the client can immediately migrate the connection to use that address. This was

specifically designed to support servers listening on anycast addresses, so the connection can be pinned to a unicast address specific to the server.

5. Other Solutions within a Limited Domain

This section describes some sticky flow solutions within a limited domain [RFC8799] for applications not based on QUIK.

Within a limited domain [RFC8799], mobile devices, edge servers, and network functions are under one administrative domain. Therefore, it is feasible for mobile devices to perform specific actions.

5.1. Use Case of 5G Edge Computing in a limited domain.

Some 5G Connected devices, such as drones for fighting natural disasters or robots in Industry 4.0 environments, need ultra-low latency responses from their analytic servers. To reach ultra-low latency, those analytic functions can be hosted on servers very close to radio towers.

All the functions (including networking and analytics) and devices are administrated by one operator. Network devices within the 5G LDN limited domain might be provided by different vendors, therefore needing interoperable solutions.

5.2. End Node Based Sticky Service Solution

The End-Node-based Sticky Service solution needs IPv6 mobile devices to insert the Destination Option header extracted from the packet received from the network side to the IPv6 Header of the next packet if the next packet belongs to the same flow. This action dramatically simplifies the processing at the LDN's Ingress routers.

Here are some assumptions for the End-Node based Sticky Service solution:

- The mobile devices are under the same administrative control as the Edge computing servers.
- If an Edge Computing service needs to be sticky in the 5G Edge Computing environment, the corresponding service ID is registered with the 5G Edge Computing controller. The Sticky Service ID can be the IP address (unicast or ANYCAST) of the server.

Here is the overview of the End-Node based Sticky Service solution:

- Each ANYCAST Edge Computing server either learns or is informed of the unicast Sticky Egress address (Section 3). The goal is to deliver packets belonging to one flow to the same Sticky Egress address for the ANYCAST address.
- When an Edge Computing server sends data packets back to a client (or the mobile device), it inserts the Sticky-Dst-SubTLV (described in Section 4.4) into the packets' Destination Option Header.
- The client (or the mobile device) needs to copy the Destination Option Header from the received packet to the next packet's Destination Header if the next packet belongs to the same flow as the previous packet.
- If the following conditions are true, the ingress router encapsulates the packet from the client in a tunnel whose outer destination address is set to the Sticky Egress Address extracted from the packet's Sticky-Dst-SubTLV:
  - The destination of the packet from the client-side matches with one of the Sticky Service ACLs configured on the ingress router of the LDN,
  - the packet header has the Destination Option present with Sticky-Dst-SubTLV.
- Else (i.e., one of the conditions above is not true), the ingress node uses its algorithm, such as the least cost as described in [5G-EC-Metrics], to select the optimal Sticky Egress address for forwarding the packet.

5.2.1. Edge Controller Based Solution.

To be added.
[Editor's note: can consider adding something along the line of the following, which is suggested by the email: say 5G/MEC control plane can tell the UE what address to use, it does NOT mean a UE will query whenever it is anchored to a new UPF. The initial query when it needs a service will return the unicast address of a server based

on all kinds of information/constraints, including the server load information talked about in draft-dunbar-idr-5g-edge-compute-app-meta-data. After that, the server won't change until new server is indeed needed (this is what "sticky service" is about, right). When a server change is indeed needed, the 5G/MEC control plane will tell the UE the new unicast address to use and tell the servers to move the corresponding application data when necessary.
]

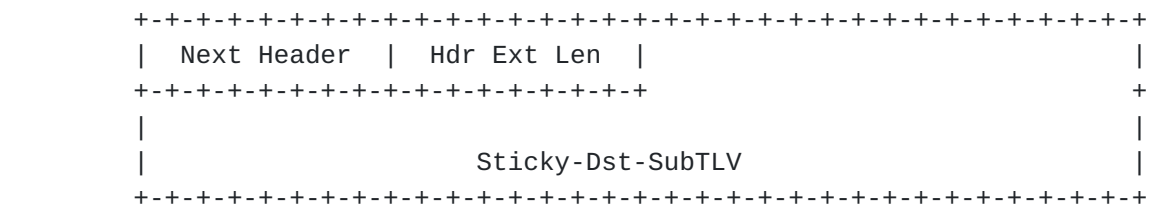
5.3. Sticky Egress Address Discovery

To an App server with ANYCAST address, the Sticky Egress address is the same as its default Gateway address.

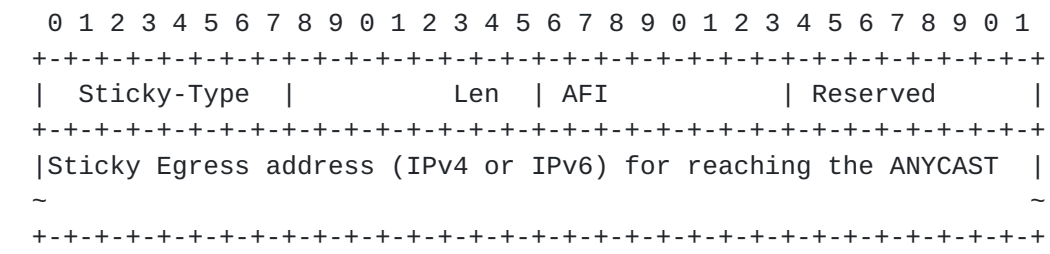
To prevent malicious entities sending DDOS attacks to routers within 5G EC LDN, e.g., the Sticky Egress address that is encoded in the Destination option header in the packets sent back to the clients, a proxy Sticky Egress address can be encoded in the Destination option header. The proxy Sticky Egress address is only recognizable by the 5G EC LDN ingress nodes, i.e., the Ra and Rb in Figure 1, but not routable in other networks. The LDN ingress routers can translate the proxy Sticky Egress to a routable address for the Sticky Egress node after the source addresses of the packets are authenticated.

5.4. Sticky-Dst-SubTLV in Destination Extension Header

A new Sticky-Dst-SubTLV is specified as below, which can be inserted into the IPv6 Destination Options header. The IPv6 Destination Option Header is specified by [RFC8200] as having a Next Header value of 60:



Sticky-Dst-SubTLV is specified as:



Sticky-Type = 1: indicate the Sticky Egress unicast address at encoded in the Sticky-Dst-SubTLV.

5.5. Processing at the Ingress router

- An Ingress router is configured with an ACL for filtering out the applications that need sticky service.

  Note, not all applications need sticky service. Using ACL can significantly reduce the processing on the routers.

- When an Ingress router receives a packet from the 5G side that matches the ACL, the Ingress router extracts the Sticky-Dst-SubTLV from the packet IPv6 header if the field exists in the packet header.
- Encapsulate the packet with the tunnel type that are supported by the original Sticky Egress node, using the extracted Sticky Egress address in the destination field of the outer Header, and forward the packet.

  Note: if the proxy Sticky Egress address is encoded in the Sticky-Dst-SubTLV, the ingress router needs to translate the proxy Sticky Egress address to a routable address.

If none of the above conditions are met, the ingress router uses its algorithm to select the optimal Sticky Egress node to forward the packet.

6. Tunnel based Sticky Service Solution

For environments that mobile devices cannot change their processing behavior as described in Section 4, a Tunnel based

Sticky Service solution can be used. This solution does not depend on mobile device's behavior. However, this solution does require ingress routers to filter out the registered sticky services and might need some level of assistance from the LDN network controller.

6.1. Desired functions by the Network Controller

6.2. Ingress and Egress Routers Processing Behavior

The solution assumes that both ingress routers and egress routers support at least one type of tunnel and are configured with ACLs to filter out packets whose destination or source addresses match with the Sticky Service Identifier. The solution also assumes there are only limited number of Sticky Services to be supported.
An ingress router needs to build a Sticky-Service-Table, with the following minimum attributes. The Sticky-Service-Table is initialized to be empty.

- Sticky Service ID
- Flow Label
- Sticky Egress address
- Timer

Editor's Note:
When a mobile device moves from one 5G Site to another, the same mobile device will have a new IP address. "Flow Label + Sticky Service ID" stays the same when a mobile device is anchored to a new PSA. Therefore, this solution uses "Flow Label + Sticky Service ID" to identify a sticky flow. Since the chance of different mobile devices sending packets to the same ANYCAST address using the same Flow Label is very low, it is with high probability that "Flow Label + Sticky Service ID" can uniquely identify a flow. When multiple mobile devices using the same Flow Label sending packets to the same ANYCAST address, the solution described in this section will stick the flows to the same ANYCAST server attached to the Sticky Egress router. This behavior doesn't cause any harm.

Each entry in the Sticky-Service-Table has a Timer because a sticky service is no longer sticky if there are no packets of the same flow destined towards the service ID for a period of time. The Timer should be larger than a typical TCP session Timeout value. An entry is automatically removed from the Sticky-Service-Table when its timer expires.
Note: since there are only small number of Sticky services, the Sticky-Service-Table is not very large.
When an ingress router receives a packet from a mobile device matching with one of the Sticky Service ACLs and there is no entry in the Sticky-Service-Table matching the Flow Label and the Sticky Service ID, the ingress router considers the packet to be the first packet of the flow. There is no need to sticking the packet to any location. The ingress router uses its own algorithm to select the optimal egress node as the Sticky Egress address for the ANYCAST address, encapsulates the packet with a tunnel that is supported by the egress node. The tunnel's destination address is set to the egress node address.

When an egress router receives a packet from an attached host with the packet's source address matching with one of the Sticky Service IDs, the egress router encapsulates the packet with a tunnel that is supported by the ingress router and the tunnel's destination address is set to the ingress router address. An Egress router learns the ingress router address for a mobile device IP address via BGP UPDATE messages.

When an ingress router receives a packet in a tunnel from any egress router and the packet's source address matches with a Sticky Service ID, the egress router address is set as the Sticky Egress address for the Sticky Service ID. The ingress router adds the entry of "Sticky-Service-ID + Flow Label + the associated Sticky Egress address + Timer" to the Sticky-Service-Table if the entry doesn't exist yet in the table. If the entry exists, the ingress router refreshes the Timer of the entry in the table.

When the ingress router receives the subsequent packets of a flow from the 5G side matching with an Sticky Service ID and the Sticky-Service ID exists in the Sticky-Service-Table, the ingress router uses the Sticky Egress address found in the Sticky-Service-Table to encapsulate the packet and refresh the Timer of the entry. If the Sticky-Service ID doesn't exist in the table, the ingress router considers the packet as the first packet of a flow.

The subsequent sections describe how ingress nodes prorogate their Sticky-Service-Table to their neighboring ingress nodes. The propagation is for neighboring ingress nodes to be informed of the Sticky Egress address to a sticky service if a mobile device moves to a new neighboring 5G site resulting in anchoring to a new ingress node.

6.3. A Solution without the Communication with 5G system.

When a mobile device moves to a very far away 5G site, say a different geographic region, the benefit of sticking to the original ANYCAST server is out weighted by network delay. Then, there is no point sending packets to the Sticky Egress node if the ingress router very far away. Therefore, it is necessary for each ingress router to have a group of neighboring ingress routers that are not too far away from the potential Sticky Egress nodes selected by the ingress router. This group of ingress routers is called the Neighboring Ingress Group. Each ingress router can either automatically discover its Neighboring Ingress Group by routing protocols or is configured by its controller. It is out of the scope of this document on how ingress nodes discover its Neighboring Ingress Group.

Each ingress node needs to periodically advertise its Sticky-Service-Table to the routers within its Neighboring Ingress Group.

Upon receiving the Sticky-Service-Table from routers in its Neighboring Ingress Group, each ingress router merges the entries from the received Sticky-Service-Table to its own.

The ingress and the egress nodes perform the same actions as described in Section 5.1.

6.4. A Solution that depends on the communication with 5G system

In this scenario, there is communication with 5G System and network get notified by a mobile device is anchored to a new PSA.

When a mobile device is re-anchoring from PSA1 to PSA2, 5GC EC management system sends a notification to the router that is directly connected to PSA1. The notification includes the address of the new PSA that the mobile device is to be anchored, i.e. the PSA2, and the mobile device's new IP address.

In this scenario, the Sticky Service can be uniquely identified by "Sticky Service ID" + "mobile device address". the Sticky-Service-Table should include the following attributes:

- Sticky Service ID
- mobile device address
- Sticky Egress address
- Timer

Upon receiving the notification from the 5G EC management system, the ingress router (i.e. the one directly connected to the old PSA) sends the specific entry of the Sticky-Service Table, i.e. "Sticky Service ID" + mobile device address + Sticky Egress + Timer to the router directly connected to the new PSA.

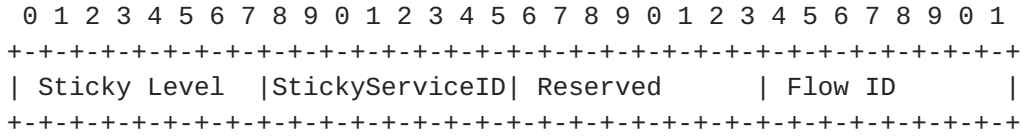
Upon receiving the entry, the ingress router merges the entry into its own Sticky-Service-Table.

The ingress and egress router processing are the same as described in Section 5.1 except a flow is now uniquely identified by the "Sticky Service ID" + "mobile device address" instead of "Sticky Service ID" + "Flow Label".

7. Expanding APN6 for Sticky Service information

The Application-aware ID and Service-Para Option described [APN6] can be expanded to include the sticky service information.

7.1. Sticky Service ID encoded in the Application-aware ID



Sticky Level: represent how important for an application to stick to its ANYCAST servers. Some applications may prefer one flow sticking to the original ANYCAST server, but not required. Some applications may require the stickiness.

StickyServiceID: the ANYCAST address of the application servers.

The Reserved field can be used for future to identifier the 5G access domain for the flow.

Flow ID: the identifier for the flow that needs to stick to a specific ANYCAST server.

7.2. Sticky Service Sub-TLV encoded in APN6 Service-para option

The Sticky-Dst-SubTLV described in the Section 4.2 of this document can be included in the Service-Para Sub-TLVs field.

8. Manageability Considerations

To be added.

9. Security Considerations

To be added.

10. IANA Considerations

To be added.

11. References

11.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC4364] E. rosen, Y. Rekhter, "BGP/MPLS IP Virtual Private networks (VPNs)", Feb 2006.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

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

11.2. 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-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-00, work-in-progress, Oct 2020.

[5G-EC-OSPF-EXT] L. Dunbar, H.Chen, A. Wang, "OSPF extension for 5G Edge Computing Service", draft-dunbar-lsr-5g-edge-compute-ospf-ext-05, work-in-progress, March 2021.

[5G-EC-BGP-EXT] 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-02, work-in-progress, March 2021.

[APN6] Z. Li, et al, "Application-aware IPv6 Networking (APN6) Encapsulation", draft-li-6man-app-aware-ipv6-network-03, work-in-progress, Feb 2021.

[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.

12. Acknowledgments

Acknowledgements to Gyan Mishra, Jeffrey Zhang, Joel Halpern, Ron Bonica, Donald Eastlake, and Eduard Vasilenko for their review and contributions.

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

Authors' Addresses

Linda Dunbar
Futurewei
Email: ldunbar@futurewei.com

John Kaippallimalil
Futurewei
Email: john.kaippallimalil@futurewei.com