Abstract

Layer separation is a powerful concept in system architecture. In the area of mobility, by separating GTP-U that is the overlay tunnel, and the IP transport network that is the underlay, the operation of the mobile network and the transport network can be separated, allowing them to evolve independently.

However, evolving individually at each layer promotes local optimization and may result in non-optimal solutions overall in the long run.

When a drastic architectural transition is required, for example, in the 5G era where various SLAs and completely new data intensive services are assumed, it is necessary to reconsider the architecture holistically, rather than from the viewpoint of individual layer.

One important value propositions of SRv6 mobile user plane is the possible optimization across the existing multiple layers.

This document discusses the architectural implications of applying SRv6 mobile user plane, especially regarding the possible optimization of existing layers. Then it takes 5G requirements and use cases as an example, and describes how these use cases are simply and effectively realized with the inter-layer optimization capability of SRv6 mobile user plane. Thus it show that SRv6 mobile use plane is a right architectural choice for the 5G era.

Status of This Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.
Layer separation is a powerful concept in system architecture. In the area of mobility, by separating GTP-U that is the overlay tunnel, and the IP transport network that is the underlay, the operation of the mobile network and the transport network can be separated, allowing them to evolve independently.
However, evolving individually at each layer promotes local optimization and may result in non-optimal solutions overall in the long run.

The well-known aphorism of David J. Wheeler says:

"All problems in computer science can be solved by adding another level of indirection."

But, as a corollary, it also says:

"...that usually will create another problem."

In other words, excessive layer separation is not good for an overall architecture.

Existing practices have reasonable grounds, so it is usually recommended to follow them. But when a drastic architectural transition is required, for example, in the 5G era where various SLAs and completely new data intensive services are assumed, it is necessary to reconsider the architecture holistically, rather than from the viewpoint of individual layer.

SRv6 mobile user plane has been proposed as an alternative way to complement or replace GTP-U both in IETF [I-D.ietf-dmm-srv6-mobile-uplane] and 3GPP [TR.29892].

SRv6 has also an advantage if it is used as a mobile user plane, because of its flexibility through Service Programming functions and the use of metadata, in addition to the simple and stateless traffic steering capability.

The 3GPP data plane entities such as UPFs and service functions can be implemented either as virtual or physical appliances. The fact that SRv6 has been supported on various platforms including custom ASICs, commercially available NPUs, programmable switches, Smart NICs, Linux Kernel, virtual forwarders on server and container networking, will make the deployment flexible.

Also, the declarative programming nature of SRv6 will provide the necessary distinction to clarify basic reachability vs constraint path vs service path, whereas existing practices depended on the layer separation - service overlay and underlay. In other words, one of the most important value propositions of SRv6 mobile user plane is the possibility to perform cross-layer optimizations.

This document discusses the architectural implications of applying SRv6 mobile user plane, especially regarding the possible
optimization of existing layers. Then it takes 5G requirements and use cases as an example, and describes how these use cases are simply and effectively realized with the inter-layer optimization capability of SRv6 mobile user plane. Thus it show that SRv6 mobile use plane is a right architectural choice for the 5G era.

2. Architecture Consideration and Necessity of Inter-layer Optimization

Historically, Mobile and Transport Network have been designed, standardized and operated separately. GTP-U has been defined as the mobile user plane. This is an overlay tunnel that runs over the Transport Network. Therefore, the underlying network cannot be directly controlled.

5G requires variety of tight-SLA characteristics and flexible traffic steering towards various service functions. While 3GPP has been focused on the mobility overlay, how to map the overlay requirements into the transport network is out of the scope.

IETF has discussed mobile end-to-end slicing in different WGs [I-D.rokui-5g-transport-slice], [I-D.clt-dmm-tn-aware-mobility]. However, all these proposals are based on the current assumption that the underlying network is separated from the overlay and agnostic.

The evolution of architecture requires a review of conventional domain boundaries and practices. This way, inefficiencies caused by traditional practices can be reduced. For example, now that "CUPS" separated the Control Plane and User Plane, UPF, which is dedicated to forwarding, can be considered as an entity on the IP Transport Network.

And, as a matter of fact, layer reduction for efficiency has been done in other domains. Some data centers adopted native IP CLOS, avoiding using VXLAN for simplicity. Also, broadband subscriber managements were simplified by using IPoE instead of PPPoE / L2TP.

In the context of mobile operators, SRv6 provides end-to-end simpler network operations thus decreasing the OPEX. SRv6 has a potential to perform inter-layer optimizations and/or eliminate overlay tunnels, though it does not mandate to do so.

Note that SRv6 can also be applied to the mobility overlay, in which case it also has benefits as the tunnels are removed.
3. Terminology

The terminology used in this document leverages and conforms to [I-D.ietf-dmm-srv6-mobile-uplane].

```
+-----+  
| AMF |  
+-----+  
/    | [N11]
[N2] /  +-----+  
+-----/  | SMF |  
/  +-----+  
/  / \  [N4]
/  /  \  
/  /   \  
[UE]------| gNB |------| UPF1 |--------| UPF2 |--------- \ DN /  
+-+- +-----+ TN +-----+ TN +-----+  
```

Figure 1: Reference Architecture

- UE : User Equipment
- gNB : gNodeB
- UPF : User Plane Function
- SMF : Session Management Function
- AMF : Access and Mobility Management Function
- 3GPP data plane entities : 3GPP entities responsible for data plane forwarding, i.e. gNB and UPF
- TN : Transport Network - IP network where 3GPP data plane entities connected
- DN : Data Network e.g. operator services, Internet access
- CUPS : Control Plane and User Plane Separation

4. SRv6 mobile user plane and the 5G use cases

4.1. Network Slicing

The SRv6 mobile user plane proposal specifies the Traditional mode and the Enhanced mode. The Traditional mode inherits the existing mobile user plane and minimize the impact to the existing the 3GPP architecture. The Enhanced mode can encode any required SID(s) for constraint path steering and service steering purpose, which enable efficient end-to-end network slicing.

How to build network slicing using the Segment Routing based technology is described in [I-D.ali-spring-network-slicing-building-blocks]
In the typical GTP-U over IP/MPLS/SR configuration, 3GPP data plane entity such as UPF is a CE to the transport networks PE. This results in the following facts:

- A certain Extra ID such as VLAN-ID is needed for segregating traffic and mapping it onto a designated slice.
- PE and the PE-CE connection is a single point of failure, so some form of PE redundancy (using routing protocols, MC-LAG, etc.) is required, which makes systems inefficient and complex.

In the past, this was unavoidable. But nowadays 3GPP data plane entities are implemented on servers or dedicated platforms which have virtualized infrastructure, and it is getting common that routing instances are implemented in such servers/platforms.

Furthermore, as stated in the introduction section, SRv6 have been implemented in various forms, it is natural for such servers/platforms to be SRv6 aware.

If the routing administrative domain must be separated between the 3GPP data plane entities and IP network, then BSID (BSID) may be used. With BSID, Topology information is abstracted and not exposed to the 3GPP data plane entities.

The BSID is bound to an SR policy, instantiation of which may involve a list of SIDs. Any packets received with active segment = BSID are steered onto the bound SR Policy, as defined in [RFC8402].

4.2. Edge Computing

Edge computing, where the computing workload is placed closer to users, is recognized as one of the key pillars to meet 5G’s demanding key performance indicators (KPIs), especially with regard to low latency and bandwidth efficiency. The computing workload includes network services, security, analytics, content cache and various applications. UPF can also be viewed as a distributed network service.

SRv6’s flexible traffic steering capabilities and the Network Programming concept of freely describing instructions and meta data are per se suitable for providing Edge Computing.

In addition, since SRv6 can be a common data plane regardless of the domains such as access, WAN, mobility and data center, Service Placement and Service Chain that used to be concentrated in Data Center can be expanded over a wide area.
Furthermore, with SRv6, session and QoS information can be exposed in IP header. It does not affect performance, thanks to the longest match mechanism in the IP routing. Only the services/applications who need the information for granular processing are to lookup. This also allows services/applications to be placed in between N9 paths if needed.

The draft implementation of Firewall using SRv6 meta data is discussed in [I-D.guichard-spring-srv6-simplified-firewall]

4.3. URLLC (Ultra-Reliable Low-Latency Communication) support

3GPP [TR.23725] investigates the key issues for meeting the URLLC requirements on latency, jitter and reliability in the 5G System. The solutions provided in such document are focused at improving the overlay protocol (GTP-U) and limit to provide a few hints into how to map such tight-SLA into the transport network. These hints are based on static configuration or static mapping for steering the overlay packet into the right transport SLA. Such solutions do not scale and hinder network economics.

Some of the issues can be solved more simply without GTP-U tunnel. SRv6 mobile user plane can exposes session and QoS flow information in IP header as discussed in the previous section. This would make routing and forwarding path optimized for URLLC, much simpler than the case with GTP-U tunnel.

Another issue that deserves special mention is the ultra-reliability issue. In 3GPP, in order to support ultra-reliability, redundant user planes paths based on dual connectivity has been proposed. The proposal has two main options.

- Dual Connectivity based end-to-end Redundant User Plane Paths
- Support of redundant transmission on N3/N9 interfaces

In the case of the former, UE and hosts have RHF (Redundancy Handling Function). In sending, RFH is to replicate the traffic onto two GTP-U tunnels, and in receiving, RHF is to merge the traffic.

In the case of the latter, the 3GPP data plane entities are to replicate and merge the packets with the same sequence for specific QoS flow, which requires further enhancements.

SRv6 mobile user plane has some advantages for URLLC traffic. First, it can be used to enforce a low-latency path in the network by means of scalable Traffic Engineering. Additionally, SRv6 provides an automated reliability protection mechanism known as TI-LFA. This is
a sub-50ms FRR mechanism that provides protection regardless of the topology through the optimal backup path.

With the case that dual live-live path is required, the problem is not only the complexity but that the replication point and the merging point would be the single point of failure. The SRv6 mobile user plane also has an advantage in this respect, because any endpoints or 3GPP data plane nodes themselves can be the replication/merging point when they are SRv6 aware.

5. Security Considerations

TBD

6. IANA Considerations

NA

7. Acknowledgements

TBD

8. References

8.1. Normative References

[I-D.hegdeppsenak-isis-sr-flex-algo]

[I-D.ietf-6man-segment-routing-header]

[I-D.ietf-dmm-srv6-mobile-uplane]

[I-D.ietf-spring-srv6-network-programming]
8.2. Informative References

[I-D.ali-spring-network-slicing-building-blocks]

[I-D.clt-dmm-tn-aware-mobility]

[I-D.filsfils-spring-srv6-interop]

[I-D.guichard-spring-srv6-simplified-firewall]

[I-D.ietf-dmm-fpc-cpdp]
[I-D.rokui-5g-transport-slice]


[TS.29244] 3GPP, "Interface between the Control Plane and the User Plane Nodes", 3GPP TS 29.244 15.0.0, December 2017.


Authors’ Addresses

Miya Kohno
Cisco Systems, Inc.
Japan
Email: mkohno@cisco.com

Francois Clad
Cisco Systems, Inc.
France
Email: fclad@cisco.com