

Workgroup: dmm  
Internet-Draft:  
draft-zzhang-dmm-mup-evolution-03  
Published: 4 February 2023  
Intended Status: Informational  
Expires: 8 August 2023  
Authors: Z. Zhang                      K. Patel      L. Contreras  
          Juniper Networks      Arrcus          Telefonica  
          K. Islam      J. Mutikainen      T. Jiang              L. Jalil  
          Redhat          NTT Docomo          China Mobile      Verizon  
          O. Sejati  
          XL Axiata

## **Mobile User Plane Evolution**

### **Abstract**

[[I-D.zzhang-dmm-5g-distributed-upf](#)] describes evolution of mobile user plane in 5G, including distributed User Plane Functions (UPFs) and alternative user plane implementations that some vendors/operators are promoting without changing 3GPP architecture/signaling. Building on top of that, this document further discusses potentially integrating UPF and Access Node (AN) in a future generation (xG) of mobile network.

This document is not an attempt to do 3GPP work in IETF. Rather, it discusses potential integration of IETF/wireline and 3GPP/wireless technologies - first among parties who are familiar with both areas and friendly with IETF/wireline technologies. If the ideas in this document are deemed reasonable, feasible and desired among these parties, they can then be brought to 3GPP for further discussions.

### **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/>.

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."

This Internet-Draft will expire on 8 August 2023.

## Copyright Notice

Copyright (c) 2023 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

## Table of Contents

- [1. MUP Evolution](#)
  - [1.1. UPF Distribution and RAN Decomposition](#)
  - [1.2. Integrated AN/UP Function in xG](#)
- [2. Some considerations with integrated ANUP](#)
  - [2.1. Separate AN/UP Functions](#)
  - [2.2. Simplified/reduced Signaling and optimized data plane](#)
  - [2.3. Mobility Handover](#)
  - [2.4. Microservice architecture](#)
  - [2.5. Increased burden on previously simple AN](#)
  - [2.6. Use of ULCL I-UPF for MEC Purpose](#)
  - [2.7. VPN PE Function in AN/ANUP](#)
  - [2.8. QoS Handling](#)
  - [2.9. NAT](#)
- [3. Security Considerations](#)
- [4. Acknowledgements](#)
- [5. Informative References](#)
- [Authors' Addresses](#)

### 1. MUP Evolution

[[I-D.zzhang-dmm-5g-distributed-upf](#)] describes evolution of mobile user plane in 5G [[3GPP-23.501](#)], including distributed UPFs and alternative user plane implementations that some vendors/operators are pushing without changing 3GPP architecture/signaling.

This document discusses potential MUP evolution in a future generation (referred to as xG) of mobile networks. It does involve changes in 3GPP architecture and signaling, so the purpose of this section is to share the ideas in IETF/wireline community first. If it gains consensus within IETF/wireline community especially among mobile operators, then the proposal may be brought to 3GPP community for further discussions.

### 1.1. UPF Distribution and RAN Decomposition

In 5G, in the opposite direction of UPF distribution, some RAN components are becoming centralized as a result of the disaggregation and decomposition of baseband processing functions. The AN functionality is now divided into the Radio Unit (RU, comprising the antenna and radiating elements), the Distributed Unit (DU, comprising the functions for the real time processing of the signal), and the Centralized Unit (CU, comprising the remaining signal processing functions). CU is the AN function that handles N3 GTP-U encapsulation for UpLink (UL) traffic and decapsulation for DownLink (DL) traffic.

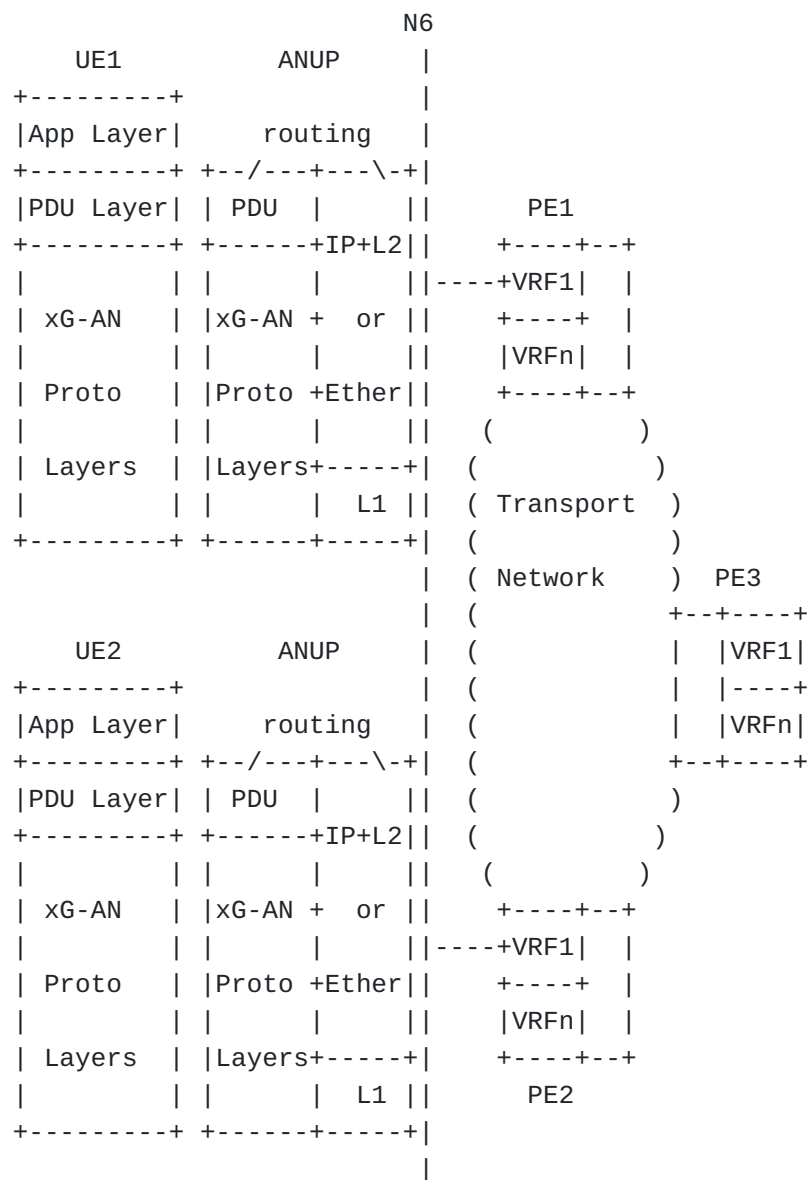
This is also specified in [[ORAN-Arch](#)], with corresponding O-RU, O-DU and O-CU terms.

The placement of the decomposed CU component can converge with the distribution process of the UPF to some optimal and convenient location in the network - they become co-located in an edge or far edge data center (DC) either with direct/short local connections in between or both running as virtual functions on the same compute server.

### 1.2. Integrated AN/UP Function in xG

While the AN (CU) and UPF can be co-located, they are still separate functions connected by N3 tunneling over a short/internal transport connection. Routing happens on the UPF between the DN and UEs over the N3 tunnels, and relay happens on the AN between the N3 tunnels and AN protocol stack.

With AN and UPF functions more and more disaggregated and virtualized even in 5G, it is becoming more and more feasible and attractive to integrate the AN and UPF functions, eliminating the N3 tunneling and the relay on AN entirely. The combined function is referred to as ANUP in this document, which does routing between DN and UEs over the AN protocol stack directly:



With this architecture, 3GPP and IETF technologies are applied where they are best applicable: 3GPP technologies responsible for radio access and IETF technologies for the rest. As IETF technologies continue to evolve, they can be automatically applied in mobile networks without any changes in 3GPP architecture/specification.

One way to view this is that the ANUP is a router/switch with wireless and wired interfaces and it routes/switches traffic among those interfaces. The wireless interface is established by 3GPP technologies (just like an Ethernet interface is established by IEEE technologies) and the routing/switching function follows IETF/IEEE standards.

Some advantages of this new architecture include:

- \*5G-LAN and MEC become transparent applications that wireline networks have been supporting (PDU sessions terminate into the closest ANUP and routed/switched to various DNSs).
- \*MBS becomes very simple - the ANUP gets the multicast traffic in the DN and then use either shared radio bearer or individual bearers to send to interested UEs.
- \*Simplified signaling - instead of seven-steps of separate N2/N4 signaling from separate AMF/SMF to separate AN/UPF and N11 signaling between AMF and SMF to set up the N3 tunneling for a PDU session, a two-step signaling between a new single control plane entity to the single integrated ANUP is enough - see [Section 2.2](#) for details.
- \*Simplified/Optimized data plane - AN-UPF connection and GTP-U encapsulation/decapsulation are not needed anymore. This can significantly improve throughput, especially when compared to AN/UPF functions running on servers.
- \*Natural local break-out in traffic forwarding, by allowing the more efficient routing/switching of traffic according to its destination.
- \*Any kind of tunnels can be used for the DN VPN, whether it is MPLS or SRv6, w/o the overhead of UDP/GTP encapsulation compared to GTP tunneling. Network slicing and QoS functions are still supported (even with current GTP tunneling the transport network need to instantiate slices and implement QoS for N3/N9 tunnels as well).

Because the ANUP already implement the routing/switching functions, even the PE functions (for the DN VPN) could be optionally integrated into it, further streamlining end-to-end communication by reducing NFs and connections between them. While integrating PE function is optional, it is desired and today's AN can be already considered as a PE ([Section 2.7](#)).

## **2. Some considerations with integrated ANUP**

Various considerations/concerns were brought up during the discussions of the ANUP proposal. They are documented in the following sections.

## **2.1. Separate AN/UP Functions**

There are still cases where separate AN/UP functions are desired/required:

- \*An MNO may want to deploy one UPF for a cluster of ANs in proximity in some scenarios/locations

- \*An MNO may support MVNOs who have their own UP functions but make use of the hosting MNO's ANs

- \*Home Routed roaming requires separate HPLMN UPs and VPLMN ANs

Therefore, the integration does not have to be always used. Rather, it is "integration when desired and feasible, separation when necessary".

Note that, the same ANUP can handle both situations - some PDU sessions may be tunneled to a separate UPF while other sessions are terminated and then traffic is routed/switched to either local DN or remote/central DN.

This is also the basis of interworking between 5G and xG:

- \*A 5G AN can have N3 tunneling to an xG UPF

- \*An xG ANUP can have N3 tunneling to a 5G/xG UPF

## **2.2. Simplified/reduced Signaling and optimized data plane**

One may ask why bother with integration when it is still needed to support separate AN and UPF anyway.

When AN and UPF are separate, to set up the N3 tunnel the following seven steps are needed, involving four NFs and three Nx interfaces:

1. SMF sends request to UPF (N4)
2. UPF responds with UPF-TEID (N4)
3. SMF passes <UPF, UPF-TEID> to AMF (N11)
4. AMF sends request to gNB, passing <UPF, UPF-TEID> (N2)
5. gNB responds with AN-TEID (N2)
6. AMF passes <AN, AN-TEID> to SMF (N11)
7. SMF sends <AN, AN-TEID> to UPF (N4)

With integrated ANUP, there is no need for N3 tunnel anymore. A new control plane NF only needs to tell the ANUP which DN that PDU session belongs to.

Additionally, the N3 tunnel is maintained by periodical signaling refreshes - otherwise timeout will happen. This causes significant control plane load on the NFs and interfaces, which no longer exists with ANUP since N3 tunneling is eliminated.

As mentioned before, with ANUP the AN-UPF connection and GTP-U encapsulation/decapsulation are not needed anymore. This can significantly improve performance/throughput, especially when compared to AN/UPF functions running on servers.

### **2.3. Mobility Handover**

Notice that ANUP is for the scenario of distributed UPFs (that are co-located with ANs) and the handover procedures for distributed UPFs (that are not integrated with ANs) applies to ANUP transparently as well. UEs may have persistent IP addresses even when they re-anchor from one ANUP to another, as described in Section 2 of [[I-D.zzhang-dmm-5g-distributed-upf](#)], or they can just get a new address when they re-anchor to a different ANUP, in which case host routes are not needed.

### **2.4. Microservice architecture**

One may argue that the integration of AN and UP functions are against the microservice trend.

The following is a verbatim quote from <https://microservices.io/>:

Microservices - also known as the microservice architecture - is an architectural style that structures an application as a collection of services that are:

- Highly maintainable and testable
  - Loosely coupled
  - Independently deployable
  - Organized around business capabilities
  - Owned by a small team
  - The microservice architecture enables the rapid, frequent and reliable delivery of large, complex applications.
- It also enables an organization to evolve its technology stack.

The counter argument is that microservice is about decomposing complex "applications". ANUP is about integrating co-located and mature data plane entities to streamline and optimize forwarding. It has real and significant benefits of simplified signaling and

optimized data plane - it does not make sense to force microservice here.

## **2.5. Increased burden on previously simple AN**

One may think that the AN only needed to do simple traffic stitching functions while now the ANUP has added UPF burden. However, the main use case of ANUP is where the AN and UPF are co-located even if they are separate functions. Therefore, the ANUP only absorbs the whatever functionalities that the separate UPF at the same site need to do anyway, with reduced signaling and data plane handling - the overall processing at the site actually decreases. While a particular ANUP now has more processing to do, it can offload some sessions to additional ANUPs that are now made possible because of removal of separate UPFs at the same site.

This may also make it easier to allocate resources at the edge DC. Previously, an operator needs to consider how much resources to allocate for the separate UPFs and assign which sessions to which UPFs. Now it simply is to decide which sessions are assigned to which ANUP (just like to decide which sessions are assigned to which AN).

## **2.6. Use of ULCL I-UPF for MEC Purpose**

Notice that the ANUP is to integrate AN and distributed UPF that are co-located in edge DCs, and one use case of distributed UPF (in those edge DCs) is MEC. UpLink Classifier Intermediate UPF (ULCL I-UPF) is an existing way to achieve local breakout routing for MEC purpose, but it is not an optimized/elegant solution compared to ANUP.

The ULCL I-UPF is placed between an AN and a central UPF as a filtering device. While called an UPF it is different from a typical UPF - It inspects *all* GTP-U UL traffic, and based on N4 signaling from SMF certain traffic is intercepted and forwarded to local DN. This places additional control plane burden on SMF in addition to the need of the special traffic-filtering UPF. For example, the SMF will need to know which traffic (to some particular destination address) is to be intercepted.

For comparison, with ANUP there is no need for the additional special UPF and corresponding N4 signaling at all. Everything is standard routing/filtering w/o relying on SMF to determine which traffic is delivered locally:

\*For some PDU sessions, all traffic may be tunneled to a separate UPF.



\*For a particular PDU session, some traffic may be delivered locally while some other delivered to the central/remote DN all based on routing/filtering in the DN.

## 2.7. VPN PE Function in AN/ANUP

As previously mentioned, the ANUP can optionally have the VPN PE function integrated, instead of being a standalone CE device for the VPN for the DN.

While optional, it is a desired optimization. Moreover, even the separate AN itself can be considered as a spoke PE for a hub-and-spoke VPN [[RFC7024](#)] for the DN.

Consider a hub-and-spoke VPN outside the mobile network context:

\*A spoke PE only imports a default route from a hub PE and therefore sends all traffic from its CEs to the hub PE

\*A hub PE imports routes from all PEs and sends traffic to appropriate PEs or its CEs, whether the traffic is from a local CE or another PE

Additionally, consider that a spoke PE advertise different per-prefix (vs. per VRF) VPN labels. When it receives traffic with a per-prefix label, it can send traffic to a local CE purely based on the label without having to do a route lookup in the VRF.

Now consider the AN and the central UPF in a mobile network. Effectively the AN is a spoke PE and the central UPF is a hub PE for the DN:

\*The GTP-U tunnel corresponds to the MPLS label stack.

\*For UL traffic, there is no need for route lookup on the AN because all is to be tunneled to the UPF. The UPF TEID is used by the UPF to determine which DN the traffic belongs to, just like how a VPN label is used to determine VPN the traffic belongs to.

\*For DL traffic, the UPF does a lookup based on the destination address (e.g., that of a UE) and a corresponding GTP-U tunnel is used to send traffic to an AN. When traffic arrives on the AN, the per-UE TEID allows traffic to be relayed to the UE without a route lookup.

In other words, the separate ANs and UPF form a hub-and-spoke VPN for the DN with per-prefix "labels", though no VRF is present on the ANs because there is no need for route lookup at all.

For ANUP with VPN PE function integrated, the only difference is the addition of VRF in the AN. That's so that some sessions will be locally terminated and traffic is locally routed. For DL traffic, the ANUP can either advertise per-VRF label (or SID in case of SR) and do a lookup for DL traffic, or advertises per-prefix/UE label (or SID in case of SR) - just like per-UE TEID - so that it does not to do a lookup before sending traffic to a UE.

## **2.8. QoS Handling**

With separate AN and UPF, the QoS handling happens in the following segments:

- \*Between UE and AN over the air interface

- \*Between AN and UPF over the N3 tunnel, which can be:

- through a transport network, or

- through a local/internal link in co-location case

The QoS over the air interface is the same for both AN and ANUP cases.

For the trivial QoS previously over N3 tunnel through a local/internal link in co-location case, it is now completely eliminated with ANUP.

The QoS over N3 tunnel through a transport network is realized through QoS mechanisms in the transport network. With ANUP, it's likely that similar QoS is needed between the ANUP and a hub router in the DN, which is a VPN over the same transport network. Therefore, it is similar to the QoS over N3 tunnel - only that now it is QoS over VPN tunnel and realized through QoS mechanisms in the transport network.

A central UPF may have rate limiting for N3 tunnels so that each PDU session's DL traffic is limited and the AN won't be overwhelmed by DL traffic. With distributed UPF (whether integrated into AN or not), the routes advertised to the hub DN router may carry QoS information like rate limiting parameters, so that the hub DN router can do rate limiting.

## **2.9. NAT**

Addresses assigned to UEs may be from a private address space and NAT is needed between the private space and public space. In case of central UPFs, the NAT can be done on a central UPF (though NAT is still a logically separate function) or by a separate NAT Gateway (GW) connected to the central UPF.

With distributed UPFs (whether it is a separate UPF or an integrated ANUP), NAT can be done by a central NAT GW connected to the hub router, just like a NAT GW on or next to the previously central UPF.

A large operator may have multiple central UPFs for different regions, and the regions may have overlapping private address spaces. Each UPF will have its own NAT GW, and UE to UE traffic across regions will go through two NAT GWs. With distributed UPFs, each region will have its own hub router with its own NAT GW, and UE to UE traffic across regions will go through two NAT GWs and two hub routers.

### 3. Security Considerations

To be provided.

### 4. Acknowledgements

The authors thank Arda Akam, Constantine Polychronopoulos, Sandeep Patel and Shraman Adhikary for their review, comments and suggestions to make this document and solution more complete.

### 5. Informative References

[I-D.zhang-dmm-5g-distributed-upf] Zhang, Z. J., Patel, K., Jiang, T., and L. M. Contreras, "5G Distributed UPFs", Work in Progress, Internet-Draft, draft-zhang-dmm-5g-distributed-upf-01, 11 July 2022, <<https://www.ietf.org/archive/id/draft-zhang-dmm-5g-distributed-upf-01.txt>>.

[ORAN-Arch] "O-RAN Architecture Description, V06.00", 2022.

[RFC7024] Jeng, H., Uttaro, J., Jalil, L., Decraene, B., Rekhter, Y., and R. Aggarwal, "Virtual Hub-and-Spoke in BGP/MPLS VPNs", RFC 7024, DOI 10.17487/RFC7024, October 2013, <<https://www.rfc-editor.org/info/rfc7024>>.

[\_3GPP-23.501] "System architecture for the 5G System (5GS), V17.3.0", December 2021.

### Authors' Addresses

Zhaohui Zhang  
Juniper Networks

Email: [zzhang@juniper.net](mailto:zzhang@juniper.net)

Keyur Patel  
Arrcus

Email: [keyur@arrcus.com](mailto:keyur@arrcus.com)

Luis M. Contreras  
Telefonica

Email: [luismiguel.contrerasmurillo@telefonica.com](mailto:luismiguel.contrerasmurillo@telefonica.com)

Kashif Islam  
Redhat

Email: [kislam@redhat.com](mailto:kislam@redhat.com)

Jari Mutikainen  
NTT Docomo

Email: [mutikainen@docomolab-euro.com](mailto:mutikainen@docomolab-euro.com)

Tianji Jiang  
China Mobile

Email: [tianjijiang@chinamobile.com](mailto:tianjijiang@chinamobile.com)

Luay Jalil  
Verizon

Email: [luay.jalil@verizon.com](mailto:luay.jalil@verizon.com)

Ori Prio Sejati  
XL Axiata

Email: [ORIP@xl.co.id](mailto:ORIP@xl.co.id)