Applying Anchorless Routing to 5G Network

Koji Tsubouchi, Ashiq Khan, Goro Kunito, Shigeru Iwashina
Research Laboratories, NTT DOCOMO, INC.
Yokosuka, Kanagawa, Japan
{kouji.tsubouchi.ua, khan, kunitoug, iwashina}@nttdocomo.com

Abstract—**Ultra Reliable and Low Latency Communication (URLLC) is one of the most important and expected use cases to be realized in 5G mobile networks. In some cases, URLLC requires low latency communication and IP session continuity at the same time. Current 5G system routing defined by 3GPP which adopts conventional tunnel based routing approach, however, does not satisfy those requirements simultaneously due to semi-static anchor point for each end point device. To address this issue, we have proposed to apply a new routing approach in 5G system, which separates communication end point device's identifier from its locator. With this approach, IP session continuity can be guaranteed without the necessity of anchor points found in conventional 3GPP mobile network systems. In this paper, we discuss the possible deployment scenarios of the proposed routing as to where and how to apply the proposed routing to 5G system defined by 3GPP, and provide a qualitative analysis.**

Keywords—5G, Anchorless, Routing, URLLC, IP session continuity, Route optimization, UE-to-UE communication

I. INTRODUCTION

Currently, 5G mobile network are being widely studied in many consortiums, research institutions, and Standards Developing Organizations (SDOs) with a strong emphasis on satisfying various new requirements of future services and markets [1][2]. The 3rd Generation Partnership Project (3GPP), which is an SDO developing standards for the 5G mobile network, states that a key distinguishing feature to be supported by the 5G system (5GS) is flexibility and adaptability so as to simultaneously provide optimized support for different use cases, each with widely different requirements.

Various use cases have been listed in [1], and some of which require Ultra Reliable and Low Latency Communications (URLLC), .e.g., remote control of vehicles and robots, real-time control of flying/driving things, advanced driving of vehicles which enables semi/fully-automated driving. For URLLC, optimized routing is necessary to reduce delay and bandwidth of the network. However, the routing approach defined until 3GPP release 15, which is the initial release to provide functionality for 5GS, still adopts tunnel based routing with an anchor point model. This kind of tunnel based routing using an anchor point causes non-optimal triangular routes, especially when the mobile terminal moves around. This cannot meet the URLLC requirements of Vehicle to Everything (V2X) communication, remote control communication, etc.

IP session continuity is also required in some cases to support seamless UE mobility. For instance, there may be a V2X service which provides a vehicle the list of IP addresses assigned to nearby vehicles so that the vehicle can communicate with them. In such a case, changing the IP addresses assigned to the vehicles will cause interruption in the communication. However, preserving the IP addresses may result in additional delay in the communication if conventional 3GPP routing techniques are used.

In our previous work, we proposed an anchorless routing from control-plane perspective, called ID Routing (IDR) to support low latency as well as IP session continuity during mobility in 5G network [6]. The proposed routing adopts an approach to separate communication end point device's identifier (ID) from its locator. This routing is generally applicable to all types of communication, but especially beneficial to communication services which require communication with mobile edge computing or UE-to-UE communication with ultra-low latency and IP session continuity. 3GPP has defined network slicing feature [3] for 5G network specifications. With network slicing, we can have logically separated 5G networks where different routing mechanisms can be running per network slice or service. Thanks to this feature, this kind of new routing protocols can be easily adopted on a per-slice basis (e.g., in the slice for a mobile broadband/V2X/IoT service). We conducted an extended evaluation between IDR and 3GPP release 15 based routing by using V2X use cases, and showed that IDR can achieve low latency and IP session continuity at the same time during mobility.

In this paper, we detail IDR from user-plane perspective. We further discuss the possible deployment scenarios of IDR as to where and how to apply IDR in 5GS defined by 3GPP. To this end, we evaluate three deployment scenarios, which are 1) IDR is applied to 5GS as a whole, 2) IDR is applied to 5G core only, 3) IDR is applied to outside of 5GS and works with 5GS.

The rest of the paper is organized as follows. Section II introduces existing routing method applicable to 5GS and summarizes its problem. Section III explains our proposed routing. In Section IV, we discuss possible deployment scenarios of our proposed routing in mobile cellular network. Section V presents related work on anchorless routing. Finally, section VI concludes the paper with summary.

II. 5GS ROUTING IN 3GPP RELEASE 15

In 5GS, the end point device connected to Radio Access Network (RAN), i.e., User Equipment (UE) communicates with other end point devices (UEs in 5GS, or devices outside of 5GS) via tunnels created on one or more User Plane Functions (UPFs) [3]. The tunnels are maintained based on the UEs' location in conjunction with the Access Management Function (AMF) and the Session Management Function (SMF). Here, the AMF is a network function which manages UE's location and its mobility, and the SMF is a network function which manages the communication paths or tunnels on the UPFs. For each UE, the SMF selects one of the UPFs which are directly connected to a Data Network (DN), which is outside of 5GS. In a DN, middleboxes (e.g., security functions, network address translation, etc.) and/or application servers are placed to provide services e.g., video, web, etc. to UEs. There are two types of DN; one is in a central site, called central DN (cDN), and the other is in geographically distributed sites, called distributed DN (dDN) in this paper. For example, cDN serves relatively large computing environment and/or provides the Internet connectivity service, while dDN serves Mobile Edge Computing environment and/or provides local services for a certain geographical area. The DN selection is performed based on the UE's location, services provided to the UE, etc. The selected UPF (called anchor-UPF) is the anchor point of the communication for the UE unless it re-connects to the 5GS. When the DN has a packet to be sent to the UE, it firstly forwards the packet to the anchor-UPF, which then forwards it to the UE via the tunnel. Fig. 1 shows that how UEs can communicate with other end point devices via tunnels in 5GS. In 5GS, such anchor points are defined to support seamless mobility when a UE moves from one base station (next generation NodeB; gNB) to another.

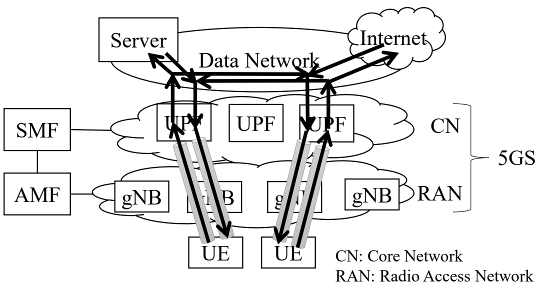


Fig. 1. Tunnel based communication in 5GS.

One of the new feature introduced in 5GS is Uplink Classifier (ULCL). ULCL is an UPF functionality which aims at diverting uplink traffic to a particular DN, e.g., dDN which are not selected as default DN, e.g., cDN. This handling is done by checking the destination address of tunnel inner packets, based on filter rules of ULCL. Fig. 2 shows an example of how ULCL diverts particular traffic to dDN while other traffic goes to cDN. In this example, a unique address space is assigned to dDN. When a UPF forwards a UE's uplink packet, and if the subnet of the destination address is the same as the one assigned to dDN in proximity, then the UPF, with the help of ULCL, diverts the packet to that dDN.

With regards to IP session continuity, 3GPP specifies three Service and Session Continuity (SSC) modes in [3]. In SSC mode 1, the IP address allocated for the UE is preserved even when the UE moves across the different UPF coverage area, which means that IP session continuity for the UE is guaranteed across the UPFs. In SSC mode 2 and 3, the IP session continuity is not guaranteed. Thus, when the UE moves to different UPF coverage area, the UE re-connects to that UPF as the anchor-UPF and a new IP address is allocated for the UE. Since the anchor-UPF is relocated to the one closer to the UE, the communication path for the UE is shortened. Specifically in SSC mode 2, when changing the anchor point, the old path is removed first, and then the new path is established. On the other hand, in SSC mode 3, the new path is established first, and then the old path is removed. SSC mode 3 also involves change in UE's IP address but can achieve the service continuity by using multiple addresses simultaneously during the change of the anchor point.

To realize URLLC, it is necessary to reduce the latency of UE-to-UE as well as that of UE-to-DN, while supporting both mobility and IP session continuity. However, the existing 3GPP network mechanisms in release 15 cannot support all of them simultaneously. For instance, SSC mode 2 can support the low latency use case, while, as for IP session continuity, it can only keep the assigned IP address as long as the UE stays only within the same UPF coverage area. When the UE moves and the anchor-UPF is relocated, the new anchor-UPF allocates a new IP address for the UE. On the other hand, SSC mode 1 can support IP session continuity, but then the UE will experience longer latency when moving out of coverage area of the anchor-UPF. By using SSC mode 3, we could reduce the impact of the change of IP address as well as achieve low latency but would impose complex IP addresses handling in application developments, because developers need to take care of both new and old IP addresses during mobility.

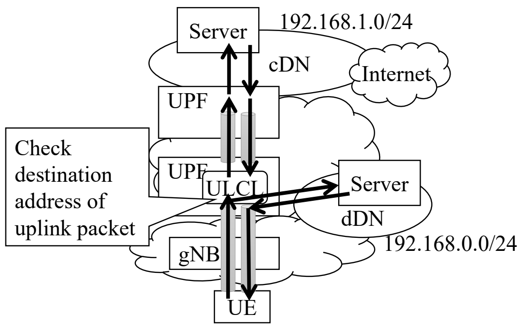


Fig. 2. Traffic diverting by ULCL in UPF.

III. ID ROUTING

In the previous section, we explained that the current tunnel based routing approach with semi-static anchor points cannot meet the requirements of both low latency and IP session continuity during UE's mobility simultaneously. This is due to the fact that the UE's ID (e.g. IP address) is tightly coupled with its anchor-UPF, or UE's locator. That is, UE's ID and its locator is not separated appropriately. To address this issue, we have proposed to apply to 5GS an anchorless routing approach called “ID Routing (IDR)” in which UE’s ID is separated from its locator, and the anchor point of the communication path for each UE is dynamically changed during its mobility [6], and showed that our approach can simultaneously achieve low latency and IP session continuity during mobility.

In terms of control-plane (c-plane), the concept of IDR follows that of Locator/ID Separation Protocol (LISP) [5] specified in Internet Engineering Task Force (IETF). With this concept, IDR can provide the optimal path which achieves low latency for each UE wherever it moves, while ensuring IP session continuity. As for user-plane (u-plane), IDR can adopt tunneling (encapsulation) approach like original LISP u-plane as well as other forwarding mechanisms.

IDR c-plane works based on very simple principles as shown in Fig. 3. Note that IDR u-plane Fig.3 is based on tunneling approach to simplify the explanation, but other approaches can also be applicable. IDR comprises two nodes, one is the Mapping System (MS), and the other is the Forwarder (FWD). In IDR, the mapping information of each UE’s ID (e.g., UE’s IP address) to the ID of location where the UE belongs (e.g., FWD’s IP address) is registered into the MS. When the FWD receives a packet, and if it does not know the location of destination packet, it sends a query message with the destination address to the MS to obtain its location. The MS then replies with the mapping information which matches the destination address. After that, the FWD caches the mapping information locally, encapsulates the received packet, and sends it to the next hop of the FWD indicated by the location ID. Note that the cache of the mapping information created by a UE can be reused for communication between other UEs. According to these principles, IDR can support UE-to-UE communication natively, and also it does not have to manipulate tunnels per UE but per FWD. This may be able to simplify the signaling of c-plane.

In IDR, the availability of the up-to-date mapping information in the cache of FWDs is a key to reduce packet forwarding delay. The learning mechanism of the mapping information at FWDs also affects the efficiency of c-plane usage. Therefore, we consider three modes (reactive/ proactive/hybrid) with regard to the learning mechanism of the map information.

In any of the three modes, if a UE moves to a different gNB or UPF coverage area, since the AMF and the SMF manage the UE’s mobility at gNB level and UPF level respectively, those functions register the latest UE’s location to the MS, which then disseminates the mapping information of the UE to the new FWD as shown in Fig. 4.

The reactive mode, in addition to above, basically follows LISP learning mechanism. That is, each FWD reactively sends a query to the MS, when it receives a packet but does not know where the destination’s location is, i.e., a cache miss of the mapping information happens in the FWD. In case that a query is triggered by the FWD, it causes service interruption due to a certain delay for querying before forwarding the packet. Otherwise, the FWD can forward the incoming packet without any additional delay.

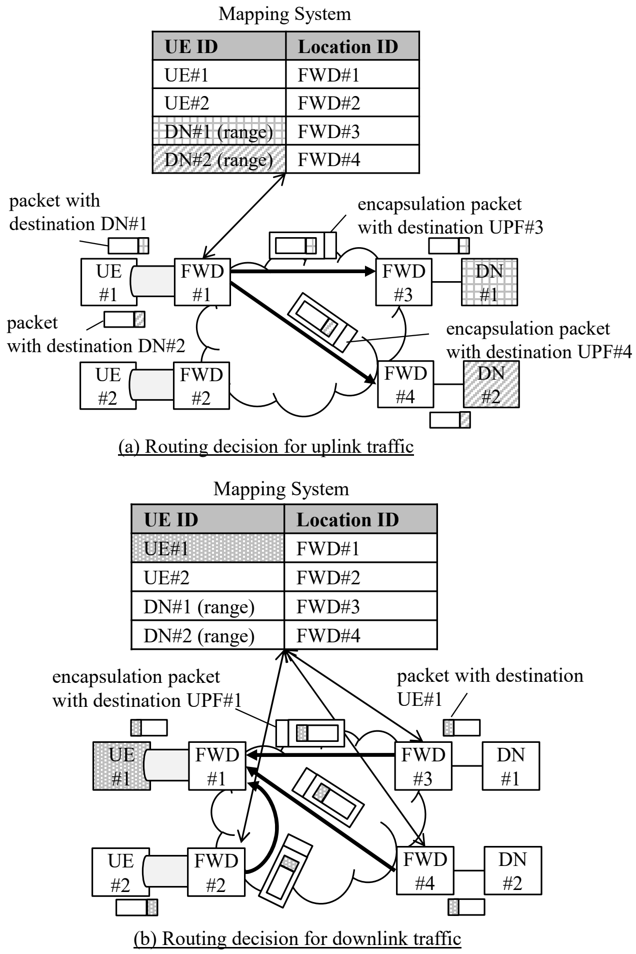


Fig. 3. Basic behavior of IDR

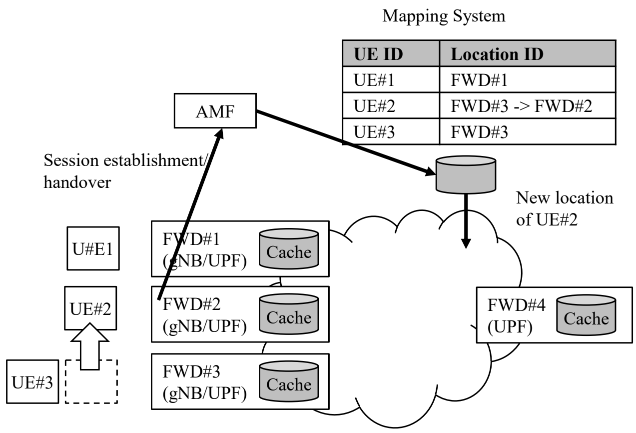


Fig. 4. UE’s location learning

In the proactive mode, when the AMF or the SMF registers the latest UE’s location to the MS, the MS disseminates the mapping information of the UE to every FWD. By doing this, all the FWDs can know the new location of the UE, and can forward the incoming packet without any additional delay for querying. In Fig. 4, when

UE2 moves from FWD 3 to 2, FWD 1, 2, 3 and 4 are notified of the new location of the UE2.

The hybrid mode is a combination of the reactive mode and the proactive mode. In the hybrid mode, UE's location is proactively notified to parts of FWDs, that is, the new UE's location is proactively notified only to the FWDs adjacent to the one the UE has moved to. This is beneficial when UEs only in proximity communicate with each other. In Fig. 4, when UE2 moves from FWD 3 to 2, FWD 1, 2 and 3 are notified of the new location of the UE2. The hybrid mode thus reduces overloading the network with location update messages.

As mentioned earlier, IDR u-plane can have multiple mechanisms. Assuming the u-plane network is composed by IPv4/v6 network, they are mainly categorized into 1) tunneling, 2) transforming, and 3) source routing. Hop by hop policy based routing could also be considered, however, it is taken out from this list in this paper, because this will require mapping information of UE's ID and its locator on every L3 switch/router between FWDs, which may unrealistic for telecom large networks. Fig. 5 shows an overview of these mechanisms.

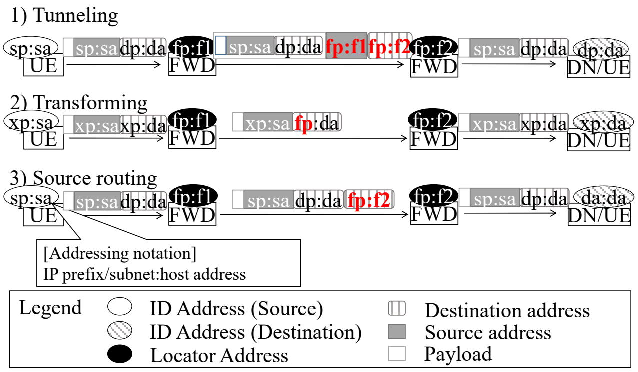


Fig. 5. Overview of different u-plane mechanism variants.

In the tunneling mechanism, the ingress FWD encapsulates an original packet with locator address (f2 assigned to the egress FWD), and forwards it to the egress FWD, which then decapsulates it as previously mentioned. Example available protocols are LISP u-plane, General Packet Radio System (GPRS) Tunneling Protocol U-plane (GTP-U), Generic Routing Encapsulation (GRE), etc.

In the transforming mechanism, each FWD has locator address represented by IP prefix/subnet address, and each UE has an address composed of ID and a prefix which is independent of locator. The ingress FWD transforms the prefix of the destination UE's ID address (xp:da) into the locator's address prefix (fp:da), i.e., the egress FWD, and forwards it to the egress FWD, which then transforms the locator's address into original destination UE's address (xp:da). One of the example protocols is Identifier Locator Addressing (ILA), which is based on IPv6 [12].

In the source routing mechanism, the ingress FWD sends a received packet to the corresponding egress FWD by adding the egress FWD address to the forwarding path list in the packet header. In this mechanism, ingress FWD can optionally specify one or more intermediate underlay addresses of network node which understand the used source routing protocol for forwarding path. Specifying such underlay network node(s) between ingress and egress FWD requires additional functionality apart from MS, which is beyond the scope of this paper. One of the example protocols is Segment Routing [7].

With the tunneling mechanism, network operators do not have to upgrade the whole underlay transport network. However, this mechanism has encapsulation overhead as well as it cannot control underlay network. The transforming mechanism on the other hand does not have such encapsulation overhead. However, if ILA is used, the transport network between ingress/egress FWDs is required to support IPv6. The source routing mechanism introduces additional flexibly on top of the aforementioned mechanisms, for example, this mechanism can specify fine-grained optimal routes between ingress and egress FWD, and can also insert middleboxes for flexible service chaining. Segment Routing can realize this mechanism but it should also be noted that additional header size is required for intermediate nodes(s) as well as the transport network between ingress/egress FWDs is required to support IPv6 or MPLS.

In our previous study based on the use cases of Vehicle to Vehicle communication with network assistance which require both low latency and IP session continuity at the same time [6], proactive mode and hybrid mode outperformed Release 15 5GS routing. As for reactive mode, this cannot meet low latency during the query of mapping information to the MS. On the other hand, reactive mode showed less signaling cost than other modes.

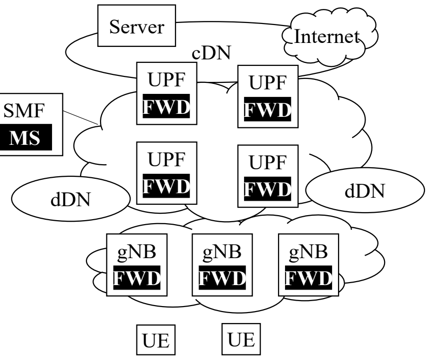
IV. Deployment Scenarios

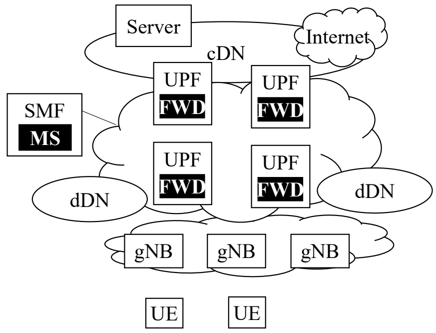
In this section, we explore various scenarios how to apply IDR to 5G network. 3GPP 5GS has network slicing feature, with which we can have logically separated 5G networks where different routing mechanism is running, per network slice or service. Therefore, we assume that the IDR is applied only to certain slices where necessary.

In terms of where in 5G network IDR can be deployed, the following scenarios are considered applicable:

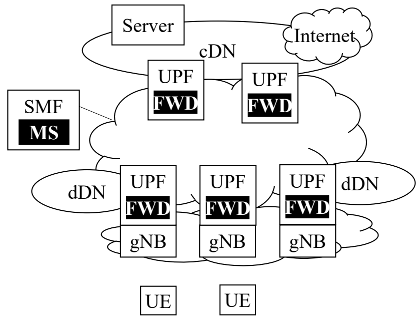
- Scenario 1) 3GPP 5GS (5G RAN + 5G CN part)
- Scenario 2) 5G CN part only
- Scenario 3) DN part (outside of 3GPP 5GS)

Fig.6 shows Scenario 1, where the FWDs are integrated into the UPFs as well as into the gNBs, while the MS is integrated into the SMF. The major benefit of this scenario is route optimization. This scenario is considered to gain the best route optimization among the scenarios especially for UE-to-UE communication because FWDs are placed at the nearest location to UEs among the scenarios. In terms of the amount of nodes required for Software/Hardware (SW/HW) upgrade for MS/FWD, i.e., SW/HW upgrade cost for network operators, it will be considerable as the base stations

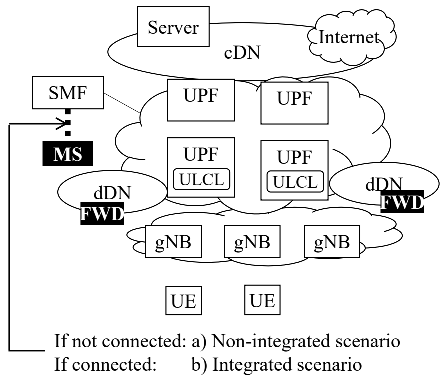




a) Non-co-located scenario



b) Co-located scenario



currently placed per network operator are typically the order of tens of thousands or more. Also, it should be noted that major standardization effort will be needed because FWD function needs to be standardized in the working group which handles RAN nodes as well as the group which handles CN nodes separately.

Fig.7 shows Scenario 2, where the FWDs are only integrated into the UPFs, and not into the gNBs. Thus, conventional GTP u-plane tunnels are used between the gNBs and the UPFs. The MS is integrated into the SMF. In this scenario, basically UPFs are placed in geographically dispersed area but are not co-located in gNB, as shown in scenario 2a. In terms of route optimization, Scenario 2a is the second best among the scenarios. Another valid deployment option, that is, Scenario 2b is that each gNB co-locates with the UPF which has FWD. In this scenario, the route optimization get closer to Scenario 1. In terms of SW/HW upgrade cost for network operators, Scenario 2a will reduce significant cost compared with Scenario 1 as the CN nodes currently placed per network operator are typically the order of tens to hundreds. On the other hand, the cost of Scenario 2b falls back to the same as that of Scenario 1. In terms of standardization effort, both Scenario 2a and 2b will have reduced effort compared with scenario 1 as the MS and the FWDs are involved only in CN nodes, which means that those functions are standardized only in the working group which handles CN nodes. Scenario 2a and Scenario 2b are the relation of the tradeoff in terms of route optimization and SW/HW upgrade cost. It is considered that this trade-off will be eased if Scenario 2a is adopted as a baseline, and Scenario 2b is adopted in selective area where necessary for the severe low latency requirement for UE-to-UE communication.

In Scenario 3 shown in Fig. 8, the FWDs and the MS are deployed in DN. In 3GPP 5GS, although a network slice is used, the network only uses already available 5GS functionalities in Release 15. The details of this scenario is described in [11]. This paper overviews the routing mechanism in this scenario as follows.

This scenario exploits ULCL in geographically dispersed UPFs to divert all the UE-to-UE uplink traffic to nearby dDN, wherein IDR components (MS/FWD) work. That is, ULCL functionality in the geographically dispersed UPFs checks destination address of the tunnel inner packets and if the subnet of the destination address is the one that assigned for UE address, then the UPFs divert those packets to the nearby dDN. In Scenario 3b, the MS is connected to the SMF and the SMF notifies the MS of UE's location information, that is, the mapping information of the UE address and the serving UPF address when the UE changes the serving UPF, while in Scenario 3a, the MS is not connected to the SMF, which means the MS has to learn the mapping information from the FWDs reactively. Thus, IDR in Scenario 3a can support Reactive mode only. In Scenario 3a, once a FWD receives a packet from a nearby UPF, the FWD checks the cache of mapping information of the source UE address. If the FWD does not have the cache for that UE, this means that the UE sent the first packet to the FWD via the new serving UPF, and then the FWD register to the MS the mapping information of the UE and the FWD. Due to this reactive mapping information registration, the UE cannot receives any packet from other UEs without sending out a packet after the UE changes serving UPF as a result of mobility. In Scenario 3b, thanks to the proactive notification of mapping information from the SMF, there are not mode restriction or packet reachability.

In terms of route optimization, Scenario 3 comes third among the scenarios, however, it is improved compared with the pure Release 15 based 5GS routing. In terms of SW/HW upgrade cost for network operators, Scenario 3 will be comparable or less cost compared with Scenario 2 because the amount of the FWDs is less than the amount of total UPFs while Scenario 2 requires all the UPFs to have the FWD. In terms of standardization effort, Scenario 3a does not require any effort for 3GPP. Also, solution/standards of IDR reactive mode, which can only be used in Scenario 3a, is already available, e.g., LISP, ILA. Thus, Scenario 3a has zero effort for standardization. As for Scenario 3b, there may be a possibility for the need of slight effort for mapping information notification in 3GPP if the SMF does not support this capability. As for the standardization of IDR modes, reactive mode is already available as mentioned before. Similar effort will be needed for hybrid/proactive mode in relevant SDOs, e.g., IETF.

V. RELATED WORK

A number of studies for routing optimization (RO) using anchorless approach in mobile networks have been conducted to avoid central and static anchor point e.g., in [8][9][10]. They can be divided into host-based mobility and network-based mobility. As for a study of RO based on Host-based mobility, J. Azevedo et al. enhances Mobile IPv6 (MIPv6) so to achieve routing optimization by direct UE-to-UE communication without traversing the Home Agent, which is the cause of the anchor point [8]. However, it is difficult to introduce this approach in 3GPP 5GS as it requires UE enhancements, which are highly undesired for practical application. In terms of network-based mobility, L. Wang et al. has provided a comprehensive survey for Proxy Mobile IPv6 including RO [9]. M. Portoles-Comeras et al. proposes an enhanced 3GPP 4G system with minor modifications inspired by SDN and LISP to adopt anchorless approach [10]. Different to these studies, our proposal can be based on the enhancement of 3GPP 5GS or it can be applied to dDN with the help of ULCL in UPF.

In terms of ID locator separation approach, there are also numerous research conducted [13][14][15]. B. Feng, et al. provides a comprehensive survey on their principles, mechanisms, and characteristics in this area [13]. It is considered that LISP and ILA are major protocols to realize ID locator separation. D. Farinacci, et al. proposes to apply LISP to 3GPP 5G system in [14], where xTRs (which are equivalent to FWDs) are implemented in gNBs and UPFs, which is categorized to Scenario 1 in the previous analysis. [15] T. Herbert, et al., on the other hand, proposes to apply ILA u-plane to 3GPP 5G CN only in [14], where ILA forwarding nodes and ILA routers (which are equivalent to FWDs) replace UPFs, which is categorized to Scenario 2. S. Homma, et al. proposes to apply ID locator separation approach to DN part, i.e., this can be categorized to Scenario 3 in our analysis.

VI. CONCLUSION

This paper presented the benefits of applying the concept of IDR to 5GS, which can simultaneously achieve low latency and IP session continuity during mobility. It also discussed possible deployment scenarios of IDR for inside and outside of 5GS. As a result, our view of applying IDR to 5G CN only achieves reasonable low latency with affordable cost, while some standardization effort in 3GPP is required. Alternatively, applying IDR to DN reduces the benefit of low latency slightly but reduces standardization effort significantly by using standard technologies outside of the 3GPP.

ACKNOWLEDGEMENT

We thank Shunsuke Homma from NTT for his valuable comments and assistance to this works.

REFERENCES

[1] 3GPP, "Feasibility Study on New Services and Markets Technology Enablers; Stage 1," TR.22.891 v14.2.0, September 2016.

[2] NGMN Alliance, "5G White Paper," v1.0, February 2015.

[3] 3GPP, "System Architecture for the 5G System; Stage 2," TS.23.501 v15.0.0, December 2017.

[4] 3GPP, "Procedures for the 5G System; Stage 2," TS.23.502 v15.0.0, December 2017.

[5] D. Farinacci et al., "The Locator/ID Separation Protocol (LISP)," IETF RFC6830, January 2013.

[6] K. Tsubouchi et al., "Anchorless Routing for URLLC Services in 5G Network," 2018 IEEE 88th Vehicular Technology Conference (VTC2018-Fall) (to be appeared), August 2018.

[7] C. Filsfils et al., "The Segment Routing Architecture," IEEE Global Communications Conference (GLOBECOM '15), December 2015.

[8] J. Azevedo et al., "Distributed IP Mobility in a Real Vehicular Network," IEEE International Conference on Communication Workshop (ICCW), June 2015.

[9] L. Wang et al., "Survey on distributed mobility management schemes for Proxy mobile IPv6," IEEE Consumer Communications and Networking Conference (CCNC), January 2014.

[10] M. Portoles-Comeras et al., "An evolutionary path for the evolved packet system," IEEE Communications Magazine, Vol. 53, Issue 7, July 2015.

[11] S. Homma et al., "Co-existence of 3GPP 5GS and Identifier Locator Separation Architecture," IETF draft-homma-dmm-5gs-id-loc-coexistence-01 (work in progress), May 2018.

[12] T. Herbert, et al., "Identifier-locator addressing for IPv6," IETF draft-herbert-intarea-ila-01 (work in progress), March 2018.

[13] B. Feng, et al., "Locator/identifier split networking: A promising future internet architecture," IEEE Communications Surveys & Tutorials 19.4, pp.2927-2948, July, 2017.

[14] D. Farinacci, et al., "LISP for the Mobile Network," draft-farinacci-lisp-mobile-network-03 (work in progress), March 2018.

[15] T. Herbert, et al., " Identifier Locator Addressing for Mobile User-Plane," IETF draft-herbert-ila-mobile-01 (work in progress), March 2018.