A Novel Architecture for Multihop Relaying in 3GPP LTE and 5G Networks

Shashi Ranjan*, Pranav Jha*, Prasanna Chaporkar* and Abhay Karandikar*†
*Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai, India
Email: {shashi.svk, pranavjha, chaporakar, karandi}@ee.iitb.ac.in
†Director, Indian Institute of Technology, Kanpur, India
Email: karandi@iitk.ac.in

***Abstract*—Multihop relaying with mobility support is a long due necessity in a cellular network. Researchers have made proposals to enable multihop relaying in the LTE network, but it garnered a little interest from the industry as these proposals require significant modifications in the standard network elements and signaling procedures. With the advent of 5G, the multihop relaying has become essential. This work proposes a novel architecture that supports multihop relay communications in 3GPP LTE and 5G networks. We introduce two new nodes into the network, namely Relay Node and Proxy-eNB (Proxy-gNB, resp.), that constitute radio and core stack of an eNB (gNB, resp.). These nodes are connected via LTE (5G, resp.) IP network and in conjunction provide eNB (gNB, resp.) functionality to the UE connected to a relay node. The architecture does not need modifications to any standard network element, and control and user data plane procedures. Further, the new nodes reuse the existing standard interfaces to communicate either on the radio or to the core network. We also provide handover procedures due to UE and/or relay mobility.**

***Index Terms*—Multihop relaying, mobile relays, LTE networks, 5G networks, 3GPP standards**

I. Introduction

Traditionally, cellular networks are designed for a single hop communication; a cell is covered by a Base Station (BS), and a Mobile Station (MS) inside the cell communicates with the BS through a direct radio link. However, novel cost-effective ways to increase the capacity of cellular networks are required to support increasing traffic demand. Heterogeneous network (HetNet) that uses small cells by employing femto-BS and pico-BS is a step in this direction [1]. A key challenge in deploying HetNet is that the small cell BSs need to be connected to the Core Network (CN), mostly using high capacity optical fiber. In this paper, we propose an alternative that uses Relay Nodes (RNs). An RN is a low power node that can be deployed as an intermediate node between an MS and a BS. MS communicates with RN on a radio access link, and the RN in turn communicates with a BS on a wireless backhaul link. Unlike BS, RNs do not need wired backhaul, air-conditioned rooms or cooling units, high antenna and large power. Further, RNs can be installed on street lamps or walls due to their small sizes. RNs can be used to meet several objectives: on the cell edge to enhance coverage, in dead-spot areas to repair coverage holes or in urban hotspots to cope with the high user density. Thus, RNs provides all the advantages of the small cell BSs without needing to be connected to the CN directly. RNs getting associated with BSs instead of CN provide additional flexibility as they can be mobile. Mobility is critical in applications like networks during emergency situations and high-speed access in public transport. For mobile systems like public transport, only RN is required to perform handovers whereas MSs retain their connections to the RN. This significantly reduces control overhead in the network and prolongs battery life of MS.

In our architecture, RNs can also be deployed in a multihop manner; an MS communicates with an RN, and its signals hop through one or more RNs before reaching the BS. Multihop relaying may provide coverage to MSs that are located in rural or hard-to-reach places, or dead spot areas of the cellular network. In particular, multihop relaying is a must-have feature for a cellular network to be designed for military tactical communications and disaster recovery and rescue operations. Such networks demand more stringent requirements than the commercial cellular network in terms of higher reliability and resiliency. Thus, multihop relaying is a cost-effective solution that extends cell coverage, enhances cell capacity, improves battery life, and is easy to install and deploy, robust to channel conditions and emergency impairments, reliable and resilient. Due to these potential benefits, there has been an interest in deploying multihop relaying in the cellular networks.

Relaying can be out-band or in-band. In out-band relaying, the access and backhaul links operate in different carrier frequencies; in in-band relaying, they operate in the same carrier frequency. Due to isolation of the access and backhaul links in frequency, an out-band RN can work in full-duplex mode and experiences lower inter-cellular interference compared to an in-band RN. Nevertheless, from an operator point-of-view, the allocation of a separate carrier frequency for access link might not always be feasible due to lack of available spectrum, high cost involved in spectrum licensing and inefficiency in dedicating this frequency band only for MSs connected to the RN. As in-band RNs use the same radio hardware for access and backhaul, no extra equipment is required, and deploying them is usually cheaper than out-band RNs. Therefore most modern studies have focused on in-band relaying as a solution.

It is evident that RN performance is restricted by the capacity of the backhaul link. The spectrum shortage in sub-6 GHz band and high capacity demands in urban regions are the reasons operators are not ready to leave any part of cellular bandwidth for wireless backhaul. With the emergence of millimeter Wave (mmWave) communication, however, the capacity constraint is not a concern as high bandwidth in the mmWave spectrum (at least $10\times$ the cellular bandwidth in sub-6 GHz) will be available. As the upcoming Fifth Generation (5G) cellular networks are expected to operate in the mmWave band, a cell will have small coverage area and hence a large number of cells will be required to cover a given geographical area. In such an ultra-dense deployment, backhauling enormous access traffic from every BS to CN via wired fiber connection is not a scalable and economical solution. As a result, Integrated Access and Backhaul (IAB) has gained substantial interest recently. IAB is an ambitious solution where BS will use the same spectrum to serve MSs in the access link and communicate with CN or other BSs in the backhaul link. The 3rd Generation Partnership Project (3GPP) has also started working on a new study item to investigate the performance of IAB-enabled 5G network [2]. Since IAB functions similar to in-band relaying, an architecture for supporting multihop relays in the 5G network may also be compliant with IAB.

A. *Things expected from a multihop relay cellular network*

After carefully considering the benefits and usability of relaying, we list following as the key architectural requirements for a multihop relay cellular system:

- RN must be layer 3 relay. A layer 3 relay demodulates and decodes the incoming signal and re-modulates and re-encodes the signal before its transmission. Also, layer 3 RN posses a unique physical cell ID to control its cell. Hence, the RN appears to the MSs to be a normal BS and thus can manage radio resources among them.
- RN should have cost-effective hardware and software design. Adding RN in the network must be possible with minimum changes in network elements and interfaces. RN should be quickly deployed when needed, like a plug-and-play device.
- The relaying function should be transparent to the operations of MSs and core network elements. This is required to support legacy or standard MSs.
- The system should be able to support in-band backhauling.

In this article, we present an architecture to support multihop relays in a cellular system that fulfills all these requirements categorically.

II. Related Work

All the advantages discussed before have motivated significant research work both in industry and academia to enable relaying in cellular network specifically in the Long Term Evolution (LTE) network. In the LTE network, BS is referred to as evolved NodeB (eNB), MS as User Equipment (UE) and CN as Evolved Packet Core (EPC). The EPC is primarily composed of three nodes for different purposes – Mobility Management Entity (MME), Serving Gateway (SGW) and Packet Data Network Gateway (PGW). The eNB communicates to UE via Uu interface, to EPC via S1 interface and to other eNBs via X2 interface. The 3GPP identified four alternative architectures, A.1 to A.4, for supporting layer 3 relays in LTE [3]. Out of these, the A.2 architecture became a part of the LTE standard in Release 10 [4]. In A.2 architecture, a new node called Donor eNB (DeNB) is introduced to serve RNs. The DeNB is a modified eNB that embeds and provides S1 and X2 proxy functionalities between an RN and other network nodes viz. MME, SGW and DeNBs. Thus, the DeNB appears as an MME for S1, as an eNB for X2, and also as an SGW to the RN. Communication between the RN and DeNB happens via a new radio interface called Un interface, a modification of the Uu interface. As RN does not know which cells support RN operation, it obtains information on the attachable DeNB cell list. Then, the RN makes a radio connection with the cell having the best radio quality in the list.

The standardized architecture, however, has several shortcomings. First, the DeNB is a bulky modification of eNB – to act as a proxy for S1 functionality, protocol stacks of CN is also integrated into the DeNB. Second, at least one DeNB must be deployed beforehand for proper RN operation. Thus the deployment of DeNBs and RNs needs careful network planning which makes their deployment economically inefficient. Third, RNs require a separate MME/SGW than that for UEs, which is an extra burden on the deployment side. Finally, the Un protocol should support the transfer of S1 and X2 messages making Un interface challenging to implement. Furthermore, all four architectures, A.1 to A.4, only support relays that are operator-deployed, two-hop and stationary. As a result, the 3GPP investigated several possible architectures for mobile relays in [5]. Each of these architectures requires modification of multiple protocols and thus a considerable standardization effort [6]. Due to these reasons, despite being the part of the standards and having several advantages, relays have seen limited deployment in LTE cellular networks.

Most of the other research works are mainly an extension of the 3GPP architectures, and thereby inherit their shortcomings. To support mobile relay, authors of [7] developed a hybrid architecture of the alternatives A.1 and A.2. In the architecture, mobility region of a mobile RN is split into several mobility areas, and user plane is tunneled between mobility areas. Authors of [8] designed an architecture compliant with the alternative A.1 to support mobile relays on trains. The architecture needs two LTE/EPC networks; the first network is the general network that manages UEs within the train. The second network is deployed by the transport operator to manage mobile relays. In [9], the authors presented a method for multihop relaying by implementing X2 proxy – next-hop node serves as an X2 proxy of a previous-hop node. This architecture requires changes in the X2 establishment message and has significantly higher control overhead.

Few research works provide an entirely new architecture [10], [11]. Both these architectures require a significant change in the existing elements, their functionalities and the interfaces. Thus, all the above architectures require significant modifications to the standard 3GPP LTE or 5G system. Not only protocol stacks at the eNB are made complex but also Attach and handover procedures are modified substantially. Moreover, network planning and deployment of nodes need to be in place before a relay communication starts functioning. In this article, we present a novel architecture to support multihop relays in 3GPP LTE and 5G networks. We incorporate the requirements expected of a multihop cellular system as discussed in section I-A. Our main contributions are listed as follows:

- We propose an architecture that supports mobile and multihop RNs with minimal changes in the existing network elements, functionalities and interfaces. The architecture does not define any new interface and reuses the existing interfaces when needed. A key merit of our architecture is that a 3GPP LTE or 5G network can be converted to a multihop relay cellular network on demand basis.
- We introduce an additional node called Proxy-eNB (P-eNB) in the CN. P-eNB is not a bulky modified eNB like DeNB, but it is a simple node that can be a fast computer. Additionally, a single P-eNB is capable of controlling a large number of RNs (upto 256) within the cellular network.
- In the proposed architecture, RNs are plug-and-play devices, which means a multihop cellular network is formed as soon as RNs become operational. The architecture also supports handovers due to the mobility of MSs and RNs. The handovers are similar to the existing UE handovers.
- We also propose a possible enhancement of the architecture to enable multi-Radio Access Technology (RAT) multihop relay communication in a next-generation cellular network.

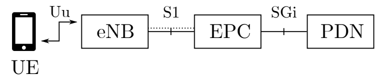
The rest of the paper is organized as follows: Section III describes the proposed architecture for 3GPP LTE network. Here, we specifically discuss two-hop relay architecture and signaling plane procedures. Subsequently, we extend these ideas to multihop relaying in Section IV. The handover procedures due to UE mobility and RN mobility for LTE multihop network are described in Section V and VI, respectively. Section VII describes a similar architecture to support multihop relaying in the 5G system. Possible extensions of our architecture are discussed in Section VIII. Section IX concludes the paper.

III. Description of the Proposed Architecture

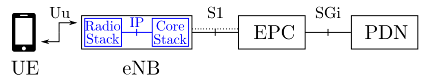
We propose a recursive architecture to enable multihop relays in 3GPP LTE and 5G networks. Here, we present our two-hop relay architecture in detail and elaborate on control and user plane procedures in an LTE system. In subsequent sections, we show how this can be extended to the multihop scenario in LTE and 5G networks.

The main idea of the proposed architecture is to visualize an eNB as a composition of two separate logical units, eNB radio unit and eNB core network unit. The eNB communicates to a UE via eNB radio unit and to CN via core network unit. Even if these units are not co-located but connected via an IP based interface (as shown in Fig 1b), together they can still provide eNB functionality to a UE. In the proposed architecture, we place the radio unit in the RN and the core network unit in a newly defined node called P-eNB. The main novelty of the architecture is that the IP connectivity between the radio unit in the RN and the core network unit in the P-eNB may itself be enabled by the LTE network as shown in Fig 1c.

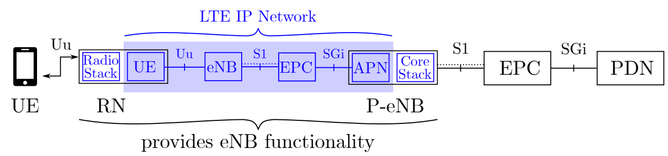
An RN has two interfaces: eNB radio and UE radio interfaces. The eNB radio interface is used by RN to communicate with its UEs (to serve them as an eNB), and the UE radio interface is used by RN to communicate with an eNB in its neighborhood. The P-eNB is located close to the CN and simultaneously connected to it through two different interfaces



(a) Abstract view of the standard 3GPP LTE architecture.



(b) Splitting of an eNB as radio unit and core network unit. These units are connected via an IP network, and together they provide eNB functionality.



(c) Abstract view of the proposed architecture. The two protocol stacks of an eNB are distributed between RN and P-eNB and connected via LTE IP network. The eNB radio stack of the RN and the CN stack of the P-eNB in conjunction provide eNB functionality to the UE connected to the RN.

Fig. 1: Evolution of the proposed multihop relay architecture from the standard 3GPP LTE architecture.

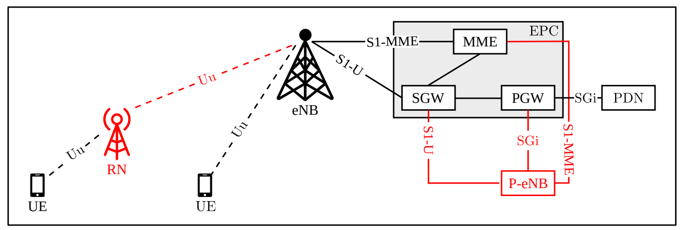


Fig. 2: An illustration of proposed architecture with network elements and their interfaces. The newly defined nodes (RN and P-eNB) and their interfaces are in red color. Note that the architecture reuses the existing standard interfaces: RN uses only Uu interface while P-eNB uses S1-U, S1-MME and SGi interfaces.

– SGi interface with PGW, and S1 interface with MME and SGW. Thus, on one side P-eNB acts as an Application Server connected to the PGW over SGi interface, and on the other side, it is attached as an eNB with the MME and SGW over S1 interface. Each network element in the proposed architecture with its interfaces is shown in Fig 2.

To provide multihop relaying in the LTE network, RNs are pre-configured to designate the P-eNB as its Access Point Name (APN) irrespective of their hop level. Thus, the RN (UE stack) communicates to the P-eNB (APN of RN) via the IP connectivity provided by the LTE network. An abstract view of the proposed architecture is illustrated in Fig 1c. The comprehensive working of two-hop relay communication in the proposed architecture is described below:

- When an RN is instantiated, it acts as a UE and establishes a radio connection to a neighboring eNB over UE radio interface.
- The RN makes an Attach procedure to the CN similar to a standard UE Attach procedure. As a result, the LTE network allocates an IP address to the RN and creates a data bearer between the RN and the PGW. It enables IP connectivity between the RN and the P-eNB.
- Once the IP connectivity between the RN and P-eNB is established, together, they can act as an eNB in the same LTE network. The RN brings up its eNB radio stack to provide LTE radio connectivity to its (prospective) UEs. Simultaneously,

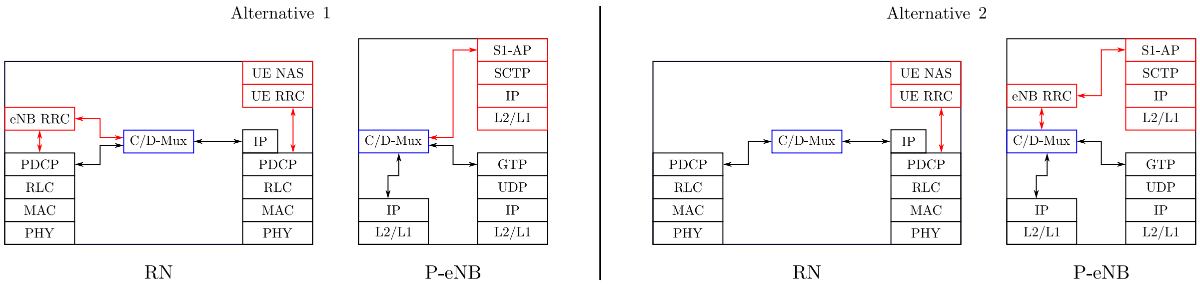


Fig. 3: The protocol stacks at RN and P-eNB for the both Alternatives. Red color is used for control related stacks. Other network elements are omitted as their protocol stacks are unchanged in the proposed architecture.

the P-eNB establishes a connection to the CN over the S1 interface and starts acting as an eNB on behalf of the RN. Thus, from the CN perspective, the relay cells appear as if they belong to the P-eNB. Hence, RNs in an area may be treated as distributed radio units of a single eNB (i.e., P-eNB) and the relay cells are administered/controlled by the P-eNB.

- The UEs now may connect to the RN over the LTE radio link, and the RN needs to forward/receive their control (e.g., Attach procedure) and data packets to/from the CN. These control and data packets are exchanged between the RN and the P-eNB over the IP connection established earlier. The P-eNB then communicates these packets with the CN over its S1 interface.

From an architectural perspective, the eNB Radio Resource Control (RRC) layer can be placed either at RN or at the P-eNB. These two alternatives result in different protocol stacks at both the nodes as shown in Fig 3. In Alternative 1, as eNB RRC is situated at the RN, UE makes RRC Connection directly to the RN. In Alternative 2, eNB RRC is situated at the P-eNB, and UE must communicate to the P-eNB to establish an RRC Connection. The RN thus becomes a simple data plane node and thereby can be manufactured at a lower cost. Transmission delay for exchanging RRC messages would be more than that of the Alternative 1. However, since a single RRC is managing all RNs, a high-performance P-eNB can process RRC messages much faster and can potentially compensate for the increased transmission delay. We explain control plane procedures for both alternatives next. The user plane procedures for both the alternatives are the same.

A. Control Plane Procedure for Alternative 1

As described before, the RN, using its UE stack, makes an Attach procedure to the CN (shown as green line in Fig 4). In response, the LTE network creates a data bearer between the RN and the PGW via the eNB and SGW. A bearer is a virtual connection between two end points which identifies an IP packet flow with specific QoS attribute. As the P-eNB is the APN for the RN, the PGW enables further IP connectivity to the P-eNB through the SGi interface (RN data path shown as magenta line). After the RN is connected to the P-eNB, the P-eNB behaves as an eNB on its other interface. The RN brings up an eNB radio stack and allows a UE to establish RRC Connection with it. The UE then sends an Attach or Non-access stratum (NAS) message to the RN which needs to be communicated to the CN. In the standard case, UE NAS message is sent to the RRC of eNB, and the eNB puts it inside S1-AP message and sends it to

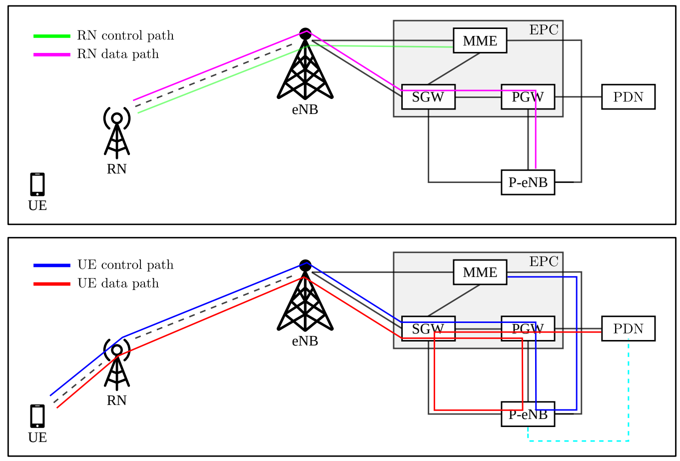


Fig. 4: An illustration of control and data paths for both UE and RN. As RN behaves as an UE, the control and data paths for RN follows the corresponding standard 3GPP UE paths with P-eNB as APN. The UE uses data path of the RN for both its control and data paths. The control path for the UE follows UE→RN→(RN data path)→P-eNB→MME, and the data path for the UE follows UE→RN→(RN data path)→P-eNB→SGW→PGW→PDN.

the MME. However, in our architecture, the RN takes the NAS message from RRC, puts it inside IP and sends it to the P-eNB along its data path that terminates at the P-eNB. At this point, it should be noted that the RN forwards both control and data packets of the UE through its data path. At P-eNB, it becomes necessary to identify the received UE packet from RN as either data or control so that the P-eNB forwards the packet to an appropriate node (MME or SGW). We define a simple transport layer protocol called Control-Data Multiplexing Protocol (C/D-Mux) to work over IP to distinguish a UE packet as data or control at the RN and the P-eNB. The C/D-Mux protocol functions similar to the relay module in an eNB specified in the 3GPP standard.

Therefore, an IP packet at the RN contains two layers of headers - C/D-Mux header and IP header of the RN. The NAS message goes through the data bearer of the RN to reach the P-eNB. The P-eNB then extracts the IP packet of the RN and determines that the message is a control packet based on its C/D-Mux header. The P-eNB thus puts the NAS message inside the S1-AP message and finally sends it to the MME. It means that from the MME perspective, the UE is connected to the P-eNB and not to the RN. The UE Attach procedure is shown as a blue path in Fig 4. The MME then creates a bearer over S1-U and S5, that is, between P-eNB and PGW via SGW for the UE and the PGW assigns an IP address to the UE. It should be noted that this bearer is separate from the bearer between the RN and the P-eNB.

Instead of defining a new transport layer protocol as mentioned above, the C/D-Mux can also be implemented at the application layer by utilizing two different types of transport to exchange the control and data with peers. The control messages can be exchanged over a TCP/SCTP and the data messages can be sent over UDP.

B. Control Plane Procedure for Alternative 2

As eNB RRC is situated at the P-eNB, UE needs to make RRC Connection directly with the P-eNB. Thus, when UE sends RRC setup message to the RN, RN being a data plane node, takes the message from PDCP layer, puts it inside IP and forwards it through RN data bearer to reach the P-eNB. After the RRC Connection is established between UE and the P-eNB, UE is now ready to attach to the LTE network. UE encapsulates the NAS message as part of an RRC message and sent to the RN. The RN encapsulates the RRC message inside an IP packet using the C/D-Mux protocol and forwards the packet to the P-eNB. The P-eNB takes the packet, determines it to control packet, retrieves the NAS message (as eNB RRC is located at the P-eNB) and sends it to the MME.

C. User Plane Procedure

When UE wants to transmit data, it sends it to the RN over radio bearer that has been established between them. The RN encapsulates the UE data inside an IP packet using its own IP address and forwards to the P-eNB via RN data path. The P-eNB determines that the received IP packet is a data packet based on its C/D-Mux header, and extracts the UE data. The P-eNB then puts the UE data into the GTP tunnel, already established for the UE from the S1-U side. The tunnel takes the packet to SGW, then PGW and finally external IP network. The UE data path is shown as the red path in Fig 4.

Observe that control and data paths of UE have additional steps, SGW→PGW→P-eNB, as compared to the 3GPP standard. The delay due to these additional steps would be insignificant as mostly SGW and PGW are incorporated as one entity, and P-eNB, being a high processing unit can be combined with this entity. Optionally, it is also possible for the P-eNB to forward the UE data directly to the PDN using an IP interface (cyan dotted line in Fig 4) after receiving it from R1 and not sending it through SGW and PGW. This is similar to Local IP Access (LIPA) from the 3GPP standard and would ensure reduced delay on the user plane.

D. QoS Provisioning

Multiple data bearers with different QoS requirements may be created between an RN and LTE network. Since an RN acts as a UE from the network perspective, we can establish up to eight data bearers between RN and P-eNB of different QoS attributes. When a UE sends or receives a data packet, the RN selects a appropiate relay bearer based on the QoS requirement of the flow. Many-to-one mapping is used between radio bearers of UE in access link and RN in backhaul link.

IV. Support for Multihop communication

Our architecture is designed to support multihop relay communications. To illustrate this, consider a system comprising of two RNs (R1 and R2), and UE seeks to connect to the CN via theses RNs as shown in Figure 5. As R2 acts as a UE for R1 (serving as an eNB), the control and user plane procedures of R2 is similar to that of UE in a two-hop architecture described in the previous section. The only difference is that R2 data path terminates at the P-eNB (designated as APN of R2).

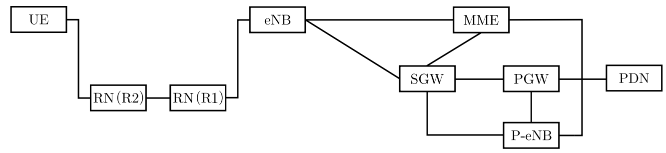


Fig. 5: An illustration of multihop relay communications in LTE network. R2 uses R1 data bearer, and UE uses R2 data bearer to communicate to CN.

In Alternative 1, UE makes RRC Connection to R2 directly. The UE can then send control or data messages to the CN via R2 data bearer (or path R2→R1→eNB→SGW→PGW→P-eNB). To achieve this, the UE message is encapsulated inside IP of R2 and sent to R1 over R2 radio bearer. At R1, the outer IP header (the IP header of R2) is removed to retrieve the UE message, and R1 encapsulates it inside IP and forwards to the P-eNB. Similarly, in Alternative 2, the UE makes RRC Connection to the P-eNB via R2 data bearer. The exchange of control or data messages between UE and CN are similar to that of Alternative 1.

The removal of outer IP header at R1 eliminates extra loop (SGW→PGW→P-eNB) that would have incurred in UE's data path due to the presence of the second level RN, R2. This strategy ensures that UE's data path has exactly one loop irrespective of the number of relay hops between UE and eNB. Another method to achieve the same is as follows. When R1 receives a packet from R2, R1 does not remove the outer IP header and simply forwards to the P-eNB. The P-eNB peels of each IP header one-by-one (first of R1 and then of R2) to obtain the UE message. Note that both these methods are employed as a part of RN and P-eNB, and thus no standard network element needs to be modified.

V. Handovers due to UE mobility

A UE handover should not depend on the number of relay hops between UE and eNB and be transparent to other UEs and standard network elements. In the proposed architecture, a handover of UE from RN to another RN or eNB and vice-versa employs standard X2 or S1 handover. We explain the handover procedure for Alternative 1; handover procedure for Alternative 2 follows along similar lines. Before providing details, we briefly recall UE handover procedures in 3GPP LTE networks on S1 interface in the following two scenarios: 1) inter-cell intra-eNB handover and 2) inter-eNB handover. Handovers on X2 interface are similar.

Each cell has a unique cell ID that is 28 bits long with the first 20 bits representing the ID of its eNB and the last 8 bits to uniquely identify a cell. In a standard S1 handover, a UE sends measurement reports with cell ID of the target cell to its eNB. The source eNB derives the target eNB ID from the cell ID. If the derived ID is the same as the eNB's ID (inter-cell intra-eNB handover), then UE is instructed to change the cell; otherwise (inter-eNB handover) the eNB sends handover request to the MME which is forwarded to the target eNB. The target eNB then responds to the MME regarding what bearers are being accepted and rejected along with RRC Configuration. The MME sends a handover command to the source eNB with this information. The source eNB then forwards the information to the UE and command it to perform handover to the target eNB. Now, we describe UE handover procedures in our architectures.

• *Handover Between RNs:* The RNs attached to the same P-eNB are treated as different cells of the same P-eNB. Hence, UE handovers between RNs is the same as the standard inter-cell intra-eNB handover. For example, consider a scenario where source and destination relays (say R1 and R2, resp.) are associated with eNBs (say E1 and E2, resp.), and let P-eNB has decided to handover UE from R1 to R2. Then P-eNB configures radio resources on R2 for the UE, and ask the UE to do a handover to R2 through an RRC Reconfiguration message. The RRC Reconfiguration message to the UE is sent via R1 (source cell). As soon as the P-eNB decides to handover, it stops sending the UE specific downlink data towards the R1 and starts buffering them instead. Upon receiving the RRC Reconfiguration message, the UE synchronizes to R2 (target cell) and establishes a radio connection with R2. With this, the UE handover completes, and R1 intimates the P-eNB about it. The P-eNB then starts sending the buffered data of the UE via R2. As the handover does not change UE specific data tunnel created between the P-eNB and SGW, the CN is not informed about the handover. Here, E1 can be same as E2.

• *Handover Between eNB and RN:* Note that a handover between eNB and RN is actually a handover between eNB and P-eNB as RN is just a cell in P-eNB. Thus, the handover procedure is identical to inter-eNB handovers in LTE networks. A handover between eNB and P-eNB can happen via X2 or S1 interfaces. The same handover procedure is applicable irrespective of the RN's association with the eNB.

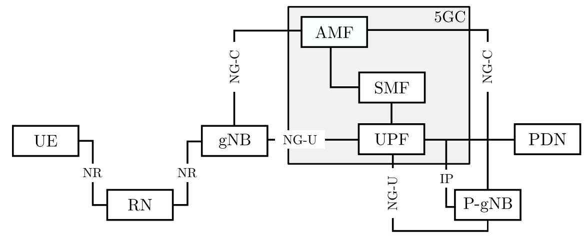
VI. Handovers Due to RN mobility

Handling of RN mobility is one of the prime requirements expected from a multihop relay network. Our architecture supports RN mobility with ease. As RNs are connected to their superordinate RN or eNB as a UE, a handover due to RN mobility is similar to a standard UE handover. Like a UE, RNs send their measurement reports to their superordinate RN or eNB using their UE stack. So standard X2 or S1 handover is performed between the source and target cells. We identified three types of handovers, each of which is described next.

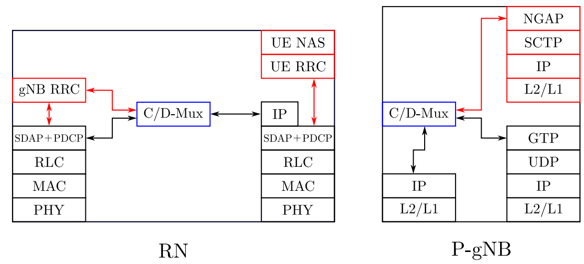
• *Handover Between eNBs:* A handover of RN between eNBs is the same as a standard UE handover. Note that the handover between two eNBs does not change the IP connectivity of the RN with the P-eNB as the PGW through which they are connected does not change during the handover.

• *Handover Between RNs:* A handover of RN from one superordinate RN to another is an inter-cell intra-eNB handover. Thus, the handover procedure is similar to that of UE handover between RNs as discussed in Section V.

• *Handover Between RN and eNB:* There are two cases possible, a handover of RN from its superordinate RN to another eNB, and from an eNB to an RN. As RNs are cells within the P-eNB, the handovers effectively happen between P-eNB and eNB. Thus, a standard X2 or S1 handover is performed between the P-eNB to eNB.



(a) An illustration of proposed architecture with interfaces.



(b) The protocol stacks at RN and P-gNB for Alternative 1. Protocol stacks for Alternative 2 follows in a similar fashion.

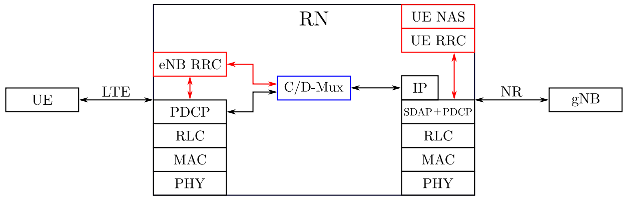
Fig. 6: Proposed 5G architecture to support multihop relays.

VII. 5G Architecture to Support Multihop Relays

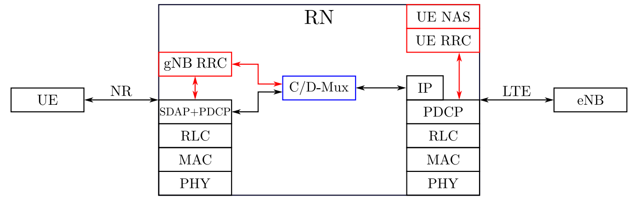
The 5G system along with its network elements to support multihop relay communications is shown in Fig 6a. The overall architecture of the 5G system is similar to the LTE system. On the CN side, 5G uses 5G Core (5GC) similar to LTE uses EPC; Access and Mobility management Function (AMF) replaces the MME, and User Plane Function (UPF) replaces the SGW/PGW. On the radio side, Next generation NodeB (gNB), a replacement of eNB, communicates to UE over New Radio (NR) protocols and the 5GC via Next Generation (NG) interface. The interfaces Uu, S1-MME and S1-U in LTE are replaced by NR, NG-C, and NG-U in 5G. Interested readers can refer to [12] for details on the 3GPP 5G standards.

To enable multihop relay communications, RNs and Proxy gNB (P-gNB) are introduced in the 5G system. If we extend the idea of splitting the protocol stacks to the 5G system, then protocol stacks at RN and P-gNB are as shown in Fig 6b. Similar to the LTE multihop relay architecture, the gNB RRC layer can be placed either at the RN or P-gNB resulting in two alternatives. In fact, RN may contain three radio protocol layers – PHY, MAC and RLC – similar to the gNB distributed units in the 5G standard. The higher layers, PDCP and SDAP, may also be on P-gNB, and P-gNB may have the same functionality as the standard gNB centralized unit.

The RRC Connection, Registration (Attach procedure), UE bearer setup procedures and handover procedures in the 3GPP 5G system are similar to that of the LTE system. Thus, each message from UE is first communicated to the RN, and the RN encapsulates it within its IP packet to send to the P-gNB though the 5G IP network. The P-gNB extracts the UE message and forwards it to the 5GC via CN stack. The flexibility of our architecture to work with both LTE and 5G networks makes it a future-proof solution.



(a) LTE access over NR backhaul.



(b) NR access over LTE backhaul.

Fig. 7: Protocol stacks at RN, for Alternative 1, to support different combinations of access and backhaul for LTE and 5G.

VIII. Possible Extensions of the Architecture

The architecture can support all the different combinations of access and backhaul for LTE and 5G, that is, LTE access over LTE backhaul, NR access over NR backhaul, LTE access over NR backhaul and NR access over LTE backhaul. Here, LTE access over NR backhaul means that UE is connected to RN over LTE access, but the RN is connected to its superordinate RN or gNB over NR backhaul. The first scenario is the base scenario upon which our architecture is built. The second scenario is for a 5G system which has been discussed already in Section VII. To achieve LTE access over NR backhaul, RN must have an LTE eNB radio stack towards UE and a UE NR radio stack towards gNB as depicted in Figure 7a. The RN then connects to the P-eNB to provide EPC connectivity to the UE. As a consequence, a connection is established between LTE UE and EPC over 5G IP network. Similarly, a connection can be established between 5G UE and 5GC over LTE IP network if the RN has protocol stacks as shown in Fig 7b.

A general extension to the architecture is to have the two core networks, EPC and 5GC, and the two proxy nodes, P-eNB and P-gNB, simultaneously in a cellular system. Therefore, the architecture can be enhanced to become a multi-RAT multihop relay architecture that enables the flexibility of utilizing LTE or 5G network to a UE based on its service requirement and availability of a particular network. Such a type of hybrid architecture is useful, especially in emergency situations where a particular network (LTE or 5G) has been partially destroyed by a natural disaster or other emergency events.

IX. Conclusion

In this article, we proposed a novel multihop relay architecture for LTE and 5G networks that also supports UE and relay mobility. The architecture does not need modifications to any standard network element and control and user data plane procedures. Our architecture is transparent to UEs, eNB/gNB and CN. The main novelty of the architecture is that the functionality of eNB (gNB) is distributed among two nodes that are connected by LTE (5G) IP network. These two nodes, namely RN and P-eNB (P-gNB) host radio and core stack of an eNB (gNB), respectively. These new nodes reuse the existing standard interfaces to communicate either on the radio or to CN. Such a feature made multihop relaying uncomplicated. We presented two alternative architectures based on whether the eNB (gNB) RRC is located at RN or P-eNB (P-gNB). We also provided handover mechanisms for both UE and RN which are actually standard X2 or S1 handover. In fact, as RN acts as a UE for the LTE network, all the 3GPP standard signaling and handover procedures that apply to a UE also apply to an RN.

Finally, we provided a multi-RAT multihop relay architecture, an enhanced architecture for a hybrid network consisting of LTE and 5G. Such a network poses a tremendous potential in future cellular communications, due to its reliable, resilience and scalable nature. Moreover, as relays need only radio interface, they can be deployed anywhere like a plug and play device. It is then possible to rapidly deploy a dynamic and on-demand network, particularly for emergency and disaster relief situations.

We believe that the proposed architecture provides a cost-effective, scalable and easy to implement solution for enabling multihop relays in LTE and 5G networks. Since, none of the existing network elements, interfaces and procedures require changes, the proposed solution can be rapidly standardized and deployed in the existing network. The architecture can be considered as a contender for IAB, ongoing work in the 3GPP.

References

[1] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, "A survey on 3GPP heterogeneous networks," *IEEE Wireless communications*, vol. 18, no. 3, pp. 10–21, 2011.

[2] 3GPP TR 38.874 v16.0.0, "NR; Study on integrated access and backhaul," January 2019.

[3] 3GPP TR 36.806 v9.0.0, "Evolved universal terrestrial radio access (E-UTRA); Relay architectures for E-UTRA (LTE-Advanced)," April 2010.

[4] 3GPP TS 36.300 V15.4.0, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description," January 2019.

[5] 3GPP TR 36.836 v12.0.0, "Evolved Universal Terrestrial Radio Access (E-UTRA); Study on mobile relay," June 2014.

[6] L. Chen, Y. Huang, F. Xie, Y. Gao, L. Chu, H. He, Y. Li, F. Liang, and Y. Yuan, "Mobile relay in LTE-advanced systems," *IEEE Communications Magazine*, vol. 51, no. 11, pp. 144–151, 2013.

[7] A. Papadogiannis, M. Farber, A. Saadani, M. D. Nisar, P. Weitkemper, T. M. de Moraes, J. Gora, N. Cassiau, D. Ktenas, J. Vihriala *et al.*, "Pass it on: advanced relaying concepts and challenges for networks beyond 4G," *IEEE Vehicular Technology Magazine*, vol. 9, no. 2, pp. 29–37, 2014.

[8] T. Kerdoncuff, T. Galezowski, and X. Lagrange, "Mobile relay for LTE: Proof of concept and performance measurements," *IEEE Vehicular Technology Conference (VTC Spring)*, pp. 1–5, 2018.

[9] L. Chu, Y. Gao, and Y. Wu, "Method and system for implementing X2 proxy," Dec. 26 2017, US Patent 9,853,902.

[10] R. Favraud and N. Nikaein, "Wireless mesh backhauling for LTE/LTE-A networks," *IEEE Military Communications Conference (MILCOM)*, pp. 695–700, 2015.

[11] T. Ochiai, K. Matsueda, T. Kondo, H. Takano, R. Kimura, R. Sawai, and F. Teraoka, "MocLis: A Non-Tunneling Moving Cell Support Protocol Based on Locator/ID Split for 5G System," *IEEE International Conference on Communications (ICC)*, pp. 1–7, 2018.

[12] 3GPP TS 38.300 v15.4.0, "NR; NR and NG-RAN Overall Description," January 2019.