Improved Fast Centralized Retransmission Scheme for High-Layer Functional Split in 5G Network

Sen XU\textsuperscript{1,2,\ast}, Meng HOU\textsuperscript{1}, Yu FU\textsuperscript{3}, Honglian Bian\textsuperscript{2} and Cheng Gao\textsuperscript{2}

\textsuperscript{1}School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, 100876, China
\textsuperscript{2}China Telecom Corporation Limited Beijing Research Institute, Beijing, 102209, P. R. China
\textsuperscript{3}China Internet Network Information Centre, Beijing, China

\ast xusen.bri@chinatelecom.cn

Abstract. In order to satisfy the varied 5G critical requirements and the virtualization of the RAN hardware, a two-level architecture for 5G RAN has been studied in 3GPP 5G SI stage. The performance of the PDCP-RLC split option and intra-RLC split option, two mainly concerned options for high layer functional split, exist an ongoing debate. This paper firstly gives an overview of CU-DU split study work in 3GPP. By the comparison of implementation complexity, the standardization impact and system performance, our evaluation result shows the PDCP-RLC split Option outperforms the intra-RLC split option. Aiming to how to reduce the retransmission delay during the intra-CU inter-DU handover, the mainly drawback of PDCP-RLC split option, this paper proposes an improved fast centralized retransmission solution with a low implementation complexity. Finally, system level simulations show that the PDCP-RLC split option with the proposed scheme can significantly improve the UE’s experience.

1. Introduction

Three typical families of typical applications including enhanced mobile broadband (eMBB), massive machine type communications (mMTC), and Ultra-Reliable and Low Latency Communications (URLLC) will be supported simultaneously in 5G system [1]. In order to satisfy these varied critical requirements, 5G RAN architecture has the need to be re-designed. 3GPP NR (New Radio) SI phase has been completed on April 2017[2]. Two typical architectures of 5G RAN have been approved to be specified in the ongoing 3GPP 5G WI stage, as illustrated in Figure 1.

One type of gNB is similar to 4G eNB with the full protocol stack supported at one base station node [3]. This architecture is suitable for urban macro scenario and indoor hotspot scenario. To simplify the network and to drive down the operational cost (OPEX) of the network, another type of gNB adopts two-level RAN architecture. The upper layers of the NR radio stacks are centralized in Central Unit (CU) while the lower layers of NR functions are performed in Distributed Unit (DU). In order to exchange the control plane signalling and user plane data between CU and DU, a new interface named F1 interface has also been introduced. One DU can support multiple physical cells and belongs to single CU, which can provide multiple points for CP and UP respectively for the sake of scalability. The drive also involves the virtualization of user data handling, signalling and
applications on commodity hardware as few as possible data centres, where hardware can be maintained inexpensively.

![Diagram of 5G Core Network and Data Centres](image)

**Figure 1** Illustration of two types of 5G RAN architecture

Currently 8 different functional splits between central and distributed units have been captured in [3], as illustrated in Figure 2.

![Diagram of Function Split between central and distributed unit](image)

**Figure 2** Function Split between central and distributed unit

The Option1~4 are defined as high layer split options while the Option5~8 are defined as low layer split options. 3GPP has decided that only one high layer split option will be specified in the 5G WI stage. And the Option2 and Option3 are mainly concerned options in 3GPP SI stage. As the sub-option of Option2, the Option2-1 is similar as 3C architecture in DC [4]. The characterization of Option2-1 is that RRC, PDCP are resided in the CU, while the RLC, MAC, physical layer and RF are resided in the DU. Due to the fundamental architecture for achieving a PDCP-RLC split have already been standardized for LTE Rel-12/13 Dual Connectivity (Option3C), the Option2-1 is regarded as the most straightforward solution to 3GPP standardization. As the sub-option of Option3, the RLC sublayer is split into High RLC and Low RLC sublayers in the Option3-1. For ARQ related functions, such as RLC Acknowledge Mode operation are performed at the High RLC sublayer residing in the central unit, while the segmentation may be performed at the Low RLC sublayer residing in the distributed unit. The benefit of this architecture is to provide the centralization gains of ARQ in CU and use the end-to-end ARQ mechanism to recover the failure over transport network. Since the above options have its cons and pros, respectively. No consensus on the selection of higher-layer option had been reached until the completion of 3GPP NR SI stage.

The Option3-1 with some enhanced solutions is considered for a better performance with Option2-1 for the aspect of missing data retransmission and fastens recovery for DU change in the contributions [5] ~ [7]. An enhanced Option3-1 solution with Transport Link Control (TLC) function is proposed in the contribution [5], which supports RLC-like function for F1 interface. The Option 2-1 with TLC function has a better performance when compared with Option 3-1 in terms of recovering.
the lost packets. Meanwhile, the Option3-1 with an effective TCP mechanism can show better performance for fasten recovery and missing data retransmission in DU overload scenario [6]. And comparing to Option2-1, the Option3-1 can reduce processing and buffer requirements in DU when the DC change occurs [7].

Meanwhile, some contributions have put forward the opposite view with the performance of option2-1 outperforming Option3-1. With respect to Option3-1, the performance of Option2-1 improves significantly while decreasing the CU-DU transport latency [8]. With lower transport capability requirement and less latency sensitive, the Option2-1 is a simpler choice for deployment. Moreover, due to ARQ function lies in DU, the RLC PDU retransmission delay can be avoided [9][10].

Moreover, due to the non-ideal backhaul between CU and DU introducing the additional transmission delay for RLC retransmission, the enhanced solution for the fast-centralized retransmission of lost RLC PDUs is also needed for both split options, especially for Option2-1.

The remainder of this paper is organized as follows. In Section II, we firstly give the analysis for both options from the aspects such as implementation complexity and system performance. And an improved fast centralized retransmission is proposed in Section III. And the simulation results on the comparison of both Option2-1 and Option3-1 are given in Section IV. The conclusions are drawn in Section V.

2. The Evaluation on Two Options

From the operator’s point view, the following evaluation criterions shall be taken into account:

- The better performance: The candidate option can provide a better system performance than the other.
- Easy Maintenance: the candidate option should be easy to configure a reasonable PDCP/RLC/MAC parameter to achieve the better performance.
- Less impact on NR standardization

2.1. The performance comparison

Since the ARQ function is resided in CU for Option3-1, the extra transport delay will be introduced in RLC retransmission. For the polling based RLC PDU retransmission without DU changes, this implies the RLC re-ordering parameters t-Reordering timer shall be increased by 4 times of one-way transport delay. And we take the Figure 3 as the example. For Option3-1, DU detects the instantaneous random error occurred or received the UE’s assist report, it shall provide the error related information to CU for triggering retransmission, due to no ARQ function in DU. Subsequently the CU shall trigger RLC Polling Request to UE. And the RLC Polling Request message shall be transferred to DU entity. For Option2-1, it can omit the above two steps due to ARQ function resided in DU entity. And in step 3 and Step 4, DU transfer or trigger RLC Polling Request to UE and UE feedbacks the RLC Status Report to DU for indicates the lost RLC PDUs. For Option3-1, DU still need to transfer the RLC Status Report to CU and CU retransmits the RLC PDU to UE later. For Option2-1, the Step5 and Step6 can be omitted. Therefore, when one RLC PDU needs to be retransmitted, at least 4 times one-way transport delay will be introduced in ARQ RTT. Moreover, we must mention that the CU can quickly retransmit the lost PDU with 3C-like the flow control mechanism introduced in 3GPP Rel-12 for Option2-1. As the role of SeNB in DC, the DU shall provide the following information to CU via the X2-U interface [11]:

- reporting the highest PDCP PDU sequence number successfully delivered to the UE;
- the desired buffer size;
- the lost X2-U packets.

With the latest transmission information (e.g., the highest PDCP PDU successfully delivered to the UE and the lost X2-U packets), CU can identify which lost PDU needs to be retransmitted. In addition, the delay over air interface (e.g. about 8ms) can also be reduced. Furthermore, the above status information is piggyback with user date PDU, no explicit X2-C signalling is needed. Therefore,
comparing to Option3-1, the delay for retransmission between DUs can be reduced sufficiently for Option2.

The Figure 4 gives simulation results on the PDCP throughput gain of AM bearer with t-Reordering timer of 30ms, 15ms, and 0ms. The three target RLC loss probabilities are considered as: $10^{-4}$, $10^{-3}$, and $10^{-2}$. The t-Reordering timer is used by the receiving side of an AM RLC entity and receiving UM RLC entity in order to detect loss of RLC PDUs at lower layer. And we selected the t-Reordering timer of 40ms as the baseline for each target RLC loss probability. Considering the one-way delay of transport network is about several or ten milliseconds order of magnitude, the assumption on the extra 15ms and 30ms introduced in t-Reordering timer is reasonable. And we can observe that the UE throughput with the t-Reordering timer of 55ms decreases 0.8%, 6.8% and 16.4%, comparing to that with the t-Reordering timer of 0ms on the order of 0.01%, 0.1%, and 1% RLC loss probability.

![Figure 3](image1.png)

**Figure 3** polling based RLC PDU Retransmission without DU changes

![Figure 4](image2.png)

**Figure 4** UE performance with variable RLC ARQ RTT

2.2. Parameters maintenance and optimization

For Option3-1, how to optimize RLC parameters (e.g., t-Reordering in H-RLC) for ARQ to meet various latency requirements and the performance changes of transport network is a big challenge for operators. And we still take the t-Reordering timer for example. This timer can affect the UE’s
experience as the above section described. And a reasonable timer value shall consider the follows factors:

- varying transport network latency
- varying queuing latency at the DU
- the delay over the air interface

If the DU queue and the delay of transport network are stable, a fixed value for $t$-Reordering timer can be configured to CU. However, in the real network the DU queue delay and the transport network delay may vary significantly over time, the operators can hardly configure a reasonable value for $t$-Reordering timer. If this timer is set too short, then the unnecessary retransmissions will be triggered, and network resources will also be wasted. If the timer is too long, UE will wait too much time before triggering retransmissions in case RLC-PDUs have been lost. Therefore, UE’s performance will be decreased with unreasonable parameters for Option3-1. Moreover, in order to satisfy the very short slot length for URLLC, the joint optimization for HARQ and ARQ operation is needed. Through fast adaptation of ARQ, the ARQ and HARQ function located in the same entity (e.g., DU) can effectively improve UE’s performance.

2.3. The impact on NR standardization

From the standardization point of view, the Option3-1 is more complex than Option2-1. Due to the signalling procedure and UP procedure have been specified in the DC topic of 3GPP Rel-12 and onwards, Option2-1 can enjoy this fruit to complete the standardization job as soon as possible. However, since the UP architecture need to be re-designed and specified, the complexity of the standardization work for Option3-1 will be higher than that for Option2-1. And from the implementation point of view, we can also conclude that the complexity of Option2-1 is less than Option3-1. Firstly, the interface and related procedures between PDCP and RLC are clearly defined in 3GPP related specifications. For CU and DU with different vendors, the overall system can be expected to work well. However, for Option3-1 some new kinds of RLC internal information exchange and interface need to be defined. Considering the 3GPP specification doesn’t restrict the implementation way for any protocol stack layers, the different thoughts of implantation way for RLC internal may be exist among vendors. And this may lead to many problems for CU and DU with different vendors. Moreover, another issue is that whether one set or two sets of protocol stack is required to separately support intra-gNB and inter-gNB. As the illustration in Figure 5, it is obvious that two sets of RLC protocol stack are needed for option 3-1, one is working for DU, and another is working for CU/DU separation. While for option 2-1, only one set of protocol stack is enough for both DC and CU/DU separation.

![Figure 5](image_url)

**Figure 5** Two Sets of RLC Protocol Stack for DU in Option3-1

Based on the above analysis, we can conclude that the Option2-1 outperforms the Option3-1 from the aspect of UE’s experience, the complexity of standardization and implantation, and network parameters optimization.
3. The improved fast centralized retransmission scheme

In order to reduce the interrupt delay, how to support the fast retransmission of lost PDCP PDUs during the intra-CU inter-DU handover scenario is regarded as the most difficult challenge for Option2-1. In this section, we firstly give the normal procedure without any improved solution for PDU retransmission in DU changes scenario, then an improved fast retransmission solution is proposed. As a contrast, the delay of the normal procedure for Option3-1 is also analyzed in this section.

3.1. The Basic Solution for Option2-1 and Option3-1

In this solution, after the handover procedure is completed, the UE triggers a PDCP status report to CU via target DU, which has been supported as PDCP data recovery. By decoding the PDCP status report, CU retransmit the lost PDCP PDUs to UE via target DU. As shown in Figure 6, the round trip transport delay will be suffered for the retransmission in this procedure.

![Figure 6 Basic Solution for Option2-1](image)

For Option3-1, the RLC status report is triggered by RLC status report polling. The polling period delay and the transport delay will be suffered before the RLC status report is triggered. As shown in Figure 7, it will lead to additional delay for the lost RLC PDU retransmission.
3.2. The improved solution for PDU retransmission

The key point for this solution is to allow the target DU to buffer a copy of the PDCP PDUs in advance which are transmitting via source DU during the handover procedure. As the illustration in Figure 8, the target DU could buffer a copy of the PDCP PDUs in advance which are transmitting via source DU during the handover procedure. After the handover procedure is completed, UE triggers a PDCP status report to target DU. From the implementation perspective, target DU could be able to derive the SN of the lost PDCP PDUs based on the PDCP status report. Then target DU will retransmit the lost PDUs which have already been buffered in target DU directly and the bi-direction transport delay can be avoided. This solution can further reduce the latency time of lost RLC PDUs recovery.

![Figure 7 Basic Solution for Option3-1](image)

![Figure 8 Improved Solution for lost PDCP PDUs retransmission procedure](image)

The following table summarizes the delay for the lost data PDUs retransmission for the above mentioned solutions for both Option2-1 and Option3-1. As shown in Table 1, the delay performance of the ordinary retransmission procedure for Option 2-1 is better than that in Option 3-1. With the improved solution, the retransmission delay can be further reduced for Option2-1. Besides, the signalling procedure can also be further simplified.
Table 1 Summary of Lost PDU Retransmission Delay for Three Solutions

|                | Fronthaul delay [ms] | Polling period caused delay [ms] |
|----------------|----------------------|----------------------------------|
| Option2-1      | 2 messages transmission | 0                               |
| Option3-1      | 3 messages transmission | Y                               |
| Proposed Solution | 0                     | 0                               |

4. System Simulation for the improved Option2-1

In the section, the System level simulation for CU/DU split is performed to evaluate the performance of different options on in section IV. The traffic model in this simulation adopts FTP model. We consider two types of Download File Sizes in this simulation: 100M bits and 1G bits. The one-way transport delay is set to 15ms. Since the most of appliances are configured with AM mode for RLC in the real network, it is reasonable to set AM mode for RLC in this simulation. Moreover, the HARQ MCS selection is on the target of BLER=10%, the residual RLC BLER should be under 1% after HARQ retransmission processing. We also set RLC residual BLER 0.1% 0.5% 1% 2% in order to simulate both the normal and worse channel situation over air interface.

The Figure 9 shows the TCP throughput efficiency for both Option 2-1 and Option3-1, compared with efficiency simulated in the condition 0% residual RLC BLER TCP throughput. The capability of physical throughput for single user is assumed to 100Mbps.

![Figure 9](image-url)

Figure 9 the TCP throughput efficiency for both Option 2-1 and Option3-1
We can observe that the TCP throughput efficiency with Option2-1 can get the performance gain 3%, 7%, 8%, and 9%, comparing to Option3-1 for 100M file sizes of the order of 0.1%, 0.5%, 1%, and 2% RLC residual BLER. And for 1G bits file sizes, we can also observe that the TCP throughput efficiency with Option2-1 can get the performance gain 1.2%, 6%, 8%, and 9%, comparing to Option3-1 of the order of 0.1%, 0.5%, 1%, and 2% RLC residual BLER. Since the ARQ is located in CU, a round-trip transport delay is introduced into the RLC Status Report and RLC retransmission. With the effective TCP slow start mechanism, the delay of RLC retransmission will lead to some negative impact on the throughput. Therefore, the Option2-1 performance is obviously better than Option3-1 for short time TCP services.

In Figure 10 and 11, FTP download time is selected to comparison the performance of Option2-1 and Option3-1. Two types of user physical throughput for single UE are considered in this simulation: 100Mbps and 1Gbps. The download file size is assumed to 100M bits in Figure 10 and 1G bits in Figure 11.
We can observe that the download time with Option2-1 can reduce 0.5 seconds, 2 seconds, 2.5 seconds, and 4 seconds, compared to Option3-1 for 100M bits File Sizes with 100Mbps physical throughput of single user on the order of 0.1%, 0.5%, 1%, and 2% RLC residual BLER. And for 1G bits file sizes, we can also observe that the download time with Option2-1 can reduce 0.7 seconds, 3.2 seconds, 5 seconds, and 8 seconds, compared to Option3-1 for 100M bits File Sizes with 1Gbps physical throughput of single user on the order of 0.1%, 0.5%, 1%, and 2% RLC residual BLER.

Based on the above simulation results, we can conclude that the Option2-1 has the better performance than Option3-1.

5. Conclusion
In April 2017, the first phase of 3GPP 5G SI has been completed. To simplify the network and to drive down the OPEX of the network, CU/DU functional split has been considered as an important feature for 5G. This paper firstly summarized the standardization work of CU/DU functional split, and then provided our analysis on the comparison of two high layer functional split options from the aspect of UE’s experience, the complexity of standardization and implantation, and network parameters optimization. Finally, the Option2-1 has been verified with the higher TCP throughput efficiency and shorter the download time for the fixed file sizes, comparing with the Option3-1. Then this work has proposed an improved solution to achieve the fast retransmission of lost PDCP PDUs in the intra-CU inter-DU handover scenario. As a result, we propose to adopt Option2-1 with the proposed solutions to improve user’s experience.

References
[1] 3GPP TR 38.913 V14.0.0, Study on Scenarios and Requirements for Next Generation Access Technologies, October 2016
[2] 3GPP Draft MeetingReport_RAN_75_170309_com, 3GPP TSG RAN Meeting #75 Dubrovnik, Croatia, March 2017
[3] 3GPP TR 38.801 V14.0.0, Study on new radio access technology: Radio access architecture and interfaces, March 2017
[4] 3GPP TS 36.300 V14.2.0, Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description, March 2017
[5] IAESI, Thales, Fairspectrum, VTT, Option 2 split with performance and reliable CU-DU connection, R3-170973, 3GPP TSG-RAN WG3 Meeting #95bis, Spokane, USA, April 2017
[6] CATT, Further Clarification of CU-DU Split Option 3-1, R3-170381, 3GPP TSG RAN WG3 Meeting #95 Athens, Greece, February 2017
[7] Nokia, Alcatel-Lucent Shanghai Bell, AT&T, Conclusion on higher layer split option, R3-170419, 3GPP TSG-RAN WG3 Meeting #95 Athens, Greece, February 2017
[8] ZTE, Clarification on CU-DU split options, R3-162852, 3GPP TSG RAN WG3 Meeting #94 Reno, USA, November 2016
[9] Ericsson, Network deployments based on option 2, R3-170683, 3GPP TSG-RAN WG3 Meeting #95 Athens, Greece, February 2017.
[10] Ericsson, Performance evaluation of option 3, R3-170685, 3GPP TSG-RAN WG3 Meeting #95 Athens, Greece, February 2017.
[11] 3GPP TS 36.425 V12.1.0, X2 interface user plane protocol, March 2015