ACCEPTED FROM OPEN CALL

RRC State Handling for 5G

Sofonias Hailu, Mikko Säily, and Olav Tirkkonen

Standardization work for the 5th Generation of mobile communications is under way in the 3rd Generation Partnership Project (3GPP). The design of the UE state machine is one of the central questions related to the overall control plane design. In this article, we present a state machine for 5G that consists of a novel RRC Connected Inactive state in addition to the conventional RRC Idle and RRC Connected states.

Standardization work for the 5th Generation of mobile communications is under way in the 3GPP. The design of the UE state machine is one of the central questions related to the overall control plane design. In this article, we present a state machine for 5G that consists of a novel RRC Connected Inactive state in addition to the conventional RRC Idle and RRC Connected states.

Introduction

The design of the control plane for the 5th Generation of mobile communications (5G) is ongoing in standardization. One of the important control plane topics is the mobility framework, which consists of the state machine design for the Radio Access Network (RAN). State machine design for 5G RAN is challenging due to the high number of 5G use cases, which have divergent and sometimes contradictory requirements. For example, there are more than 50 use cases identified for 5G in the 3rd Generation Partnership Project (3GPP) [1].

In the Universal Mobile Telecommunications System (UMTS), the state machine consists of one idle mode state and four connected mode states [2]. The idle mode is optimized for low consumption of power and network resources. No user equipment (UE) RAN context is stored in the UE nor in the network. The connected mode states are optimized for high UE activity and fast connection re-establishment. Accordingly, both the UE and the network store the UE's RAN context. However, the high number of radio resource control (RRC) states, some with significantly overlapping characteristics, is complex to implement and increases the standardization effort.

The initial release of Long Term Evolution (LTE) scaled down the number of RRC states to two, RRC Idle and RRC Connected [3]. The state machine is optimized for the Mobile Broadband (MBB) use case, as was the overall LTE design. RRC Idle is optimized for low consumption of power and network resources. The UE's access stratum (AS) context is discarded from the network and the UE upon transition to RRC Idle. RRC Connected is optimized for high UE activity. However, the state machine is inefficient for infrequent transmissions of small packets due to the heavy signaling procedure required for RRC Idle to RRC Connected transition [4]. Several studies have been conducted in 3GPP to tackle the problem, as an independent study item [5] and as part of Narrow Band (NB) Cellular Internet of Things (C-IoT) study [6].

RRC Suspended is introduced in LTE release 13 for NB C-IoT [7]. The objective of this state is to minimize the state transition signaling overhead for NB C-IoT devices. To achieve this, part of the UE AS context, for example, AS security context, is stored in the UE and the RAN during RRC Suspended. The data radio bearer (DRB)/signaling radio bearer (SRB) configurations are released, which requires re-establishment during state transition. For this purpose, an RRC Resume procedure is introduced where DRBs/SRBs are re-established and security is reactivated. The RRC Suspended to RRC Connected state transition procedure is lighter than the RRC Idle to RRC Connected transition, which reduces the signaling overhead over the air interface. However, the RAN/core network (CN) connection is torn down during transition to RRC Suspended. Thus, RRC Suspended does little to reduce the signaling overhead over the RAN/CN interface during state transition. The re-establishment of the RAN/CN interface also constitutes a significant contribution to the overall control plane latency.

To further lighten the state transition procedure, especially from the RAN/CN interface perspective, a work item for a Light Connected state is ongoing for LTE [8]. Based on high-level agreements, the CN/RAN connection is kept in addition to storing part of the UE's AS context in Light Connected, to allow faster and lighter state transition compared to RRC Suspended to RRC Connected transition.

According to high-level agreements during the 3GPP study item phase for New Radio (NR) [9], the state machine for 5G consists of three states: RRC Idle, RRC Connected and RRC Inactive. The high-level characteristics of RRC Inactive are similar to LTE's Light Connected. However, RRC Inactive is expected to have more features such as support for small data transmission with-

Digital Object Identifier:
10.1109/MCOM.2018.1700957

Sofonias Hailu and Olav Tirkkonen are with Aalto University; Mikko Säily is with Nokia, Finland.

out state transition to RRC Connected and inactive mode inter-NR/LTE mobility, which will be studied for Release 16 along optimizations for fast system access, paging, location tracking, etc, as summarized in Table 1. Stage-3 details of RRC Inactive with basic features and Light Connected are expected to be completed for Release 15.

In this article, we propose a state machine for 5G RAN that consists of RRC Idle and RRC Connected, and a novel RRC Connected Inactive state.[1] Among the characteristics of RRC Connected Inactive are:

- The UE AS context is kept in the UE and the network.
- The CN/RAN connection is kept.
- UE mobility is based on cell reselections.
- The behavior of a UE in this state is configurable based on the service requirements of the UE [10, 11].

The configurability of RRC Connected Inactive is a key feature to serve UEs with divergent service requirements with a single flexible low activity state rather than having multiple low activity states optimized for each use case. We also present an RRC Connected Inactive to RRC Connected state transition procedure with novel features that enable fast first packet transmission and inter-5G/LTE mobility in inactive mode. Performance analysis shows that RRC Connected Inactive outperforms RRC Idle in terms of latency, signaling overhead and UE power consumption.

This article is organized as follows. In the following section, we discuss the lessons that can be learned from state handling in existing systems. Then we describe the overview of a state handling mechanism for 5G. The motivations on the need for configurable state is then presented. Following that we explain the RRC Connected Inactive to RRC Connected state transition procedure with novel features for fast first packet transmission and enabling tight LTE/5G interworking. We then discuss the performance of RRC Connected Inactive. In the final section we conclude the discussion.

Lessons from State Handling in Existing Cellular Systems

Several lessons can be learned from the state handling mechanism in existing systems. One lesson from UMTS state handling is to avoid RRC states with significantly overlapping characteristics. As shown in Table 2, the connected mode states CELL_FACH, CELL_PCH and URA_PCH have similar characteristics with minor differences, targeting different levels of UE power consumption [2]. For example, CELL_PCH and URA_PCH have similar characteristics except that the UE location is tracked on a cell level in CELL_PCH, while on the level of a group of cells in URA_PCH. Thus, URA_PCH leads to lower UE power consumption and location update signaling overhead. Considering the high number of 5G use cases, aiming at having multiple states, each optimized for different use cases, requires complex implementation and excessive standardization effort.

Another lesson that can be learned from state handling in UMTS and LTE is that storing UE AS context in the UE and the network during low activity reduces the state transition latency and signaling overhead. Thus, it is crucial to have at least one such low activity state. It is also important to limit the use of a low activity state where no UE AS context is stored in the UE and the network to be a bootstrap or fault recovery state.

Another lesson that can be learned from LTE state handling is that the re-establishment of the RAN/CN interface contributes a significant component to the overall control plane latency. For example, re-establishing a RAN/CN connection contributes up to 68 percent of the overall

Topics	RRC Inactive Release 15	RRC Inactive Release 16	RRC Connected Inactive
Basic characteristics description	✓	✓	✓
Basic connection control	✓	✓	✓
Small data transmission optimization	×	✓	✓
Fast system access optimization	×	✓	✓
Autonomous inter-RAT cell reselection	×	✓	✓
Multi-connectivity setup during state transition	×	✓	✓

Table 1. Comparison of topics to be addressed in NR RRC Inactive Release 15 and Release 16 with the topics addressed in the proposed RRC Connected Inactive. Note that other potential optimizations, e.g. paging signaling optimization, location update signaling optimization, configurability to address diverse use cases, will also be addressed in Release 16.

UMTS state	Mobility procedure	Monitoring dedicated physical channels	DL channel monitoring	Location update	Uplink activity allowed	Storage of RAN context
CELL_DCH	Network controlled handover	Yes	Continuous (DCH)	Active set update	Yes	Yes
CELL_FACH	Cell reselection	No	Continuous (FACH)	Cell update	Yes	Yes
CELL_PCH	Cell reselection	No	Discontinuous with DRX (PCH)	Cell update	No	Yes
URA_PCH	Cell reselection	No	Discontinuous with DRX (PCH)	URA update	No	Yes
IDLE	Cell reselection	No	N/A	N/A	N/A	No

Table 2. Comparison of RRC states in UMTS. N/A means not applicable.

[1] Part of the paper was presented at the IEEE International Conference on Communications Workshops (ICCW) 2016 [10], where we introduced RRC Connected Inactive, and the Global Wireless Submit (GWS) 2016 [11].

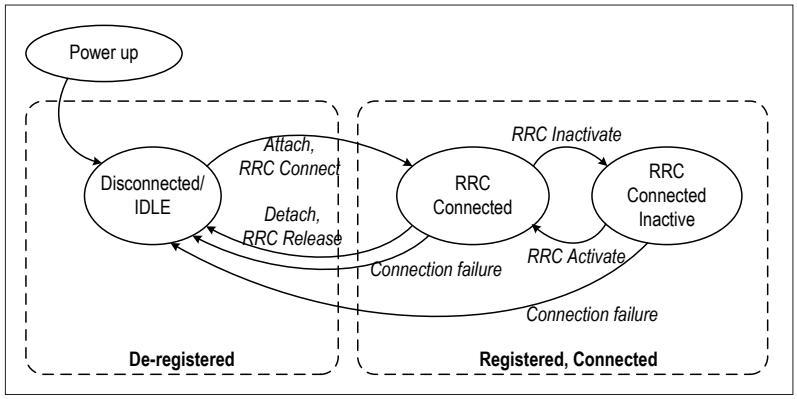

Figure 1. Potential state model for the 5G radio access network

control plane latency in LTE [12]. Therefore, it is important to maintain the RAN/CN connection during low activity state.

RRC State Machine for 5G

We propose a state machine for 5G RAN that consists of RRC Idle, RRC Connected and a novel RRC Connected Inactive (see Fig. 1). RRC Connected is optimized for high UE activity and has similar characteristics to LTE's RRC Connected, for example, UE AS context is stored in the network and the UE, and mobility is network controlled. In RRC Idle, the UE AS context is stored in neither the network nor the UE. The use of this state should be limited, it should only be used as a bootstrap state during power on and as a fault recovery state. RRC Connected Inactive is proposed as the main low activity state.

Some of the characteristics of RRC Connected Inactive are:

- The RAN/CN connection is kept, which reduces state transition latency and signaling overhead over the RAN/CN interface.
- UE AS context, for example, AS security context and DRB/SRB configurations, is stored in the UE and in the RAN. This is crucial to reduce state transition latency and signaling overhead over the air interface.
- UE mobility is based on autonomous cell reselection, where UE may move within its tracking area without updating its location. This leads to lower UE power and network resource consumption, compared to network controlled mobility. Paging is used to reach the UE.
- UE behavior in the RRC Connected Inactive state is configured based on the requirements of the applications running in the UE. This can be achieved by providing a service tailored configuration to the UE, for example, using a dedicated RRC message that transitions the UE to RRC Connected Inactive.

In general, the memory requirement to store the UE AS context and RAN/CN connection configuration is not critical in a centralized deployment as memory capacity is scalable. It may not be critical in a distributed deployment either. For example, with an expected size of up to 65 KB of UE AS context, based on the size of the HandoverPreparationInformation message [7], storing the UE AS context of around 15,000 UEs requires only 1 GB of memory.

On the Need for Configurability of RRC Connected Inactive

The need for configurability of RRC Connected Inactive is motivated by several factors that require flexibility and programmability such as diverse requirements of 5G use cases, future proofness and quick time to market requirements for new services.

Diverse Requirements of 5G Use Cases

5G use cases have highly diverse and sometimes contradictory requirements in terms of, for example, mobility, security and privacy, reliability, bandwidth, latency, battery life, and so on [1]. For example, the E2E latency requirement (which may need to be complemented by a state transition latency requirement) varies from <1ms for use cases requiring ultra-low latency such as automated driving to latencies in the order of seconds for Massive low-cost/long-range/low-power Machine Type Communication (MTC)' use cases. Battery life requirements are irrelevant for, for example, automated driving where the device can get unlimited power from the car; whereas for battery operated devices it ranges from three days for smartphones to 15 years for low-cost MTC device. Similar requirement diversity can also be observed in terms of, for example, mobility, reliability and bandwidth.

RRC Connected Inactive needs to be configurable so that UE behavior during low activity is adjusted according to its service requirement. In [11], several configurable procedures are discussed that tune the behavior of a UE in RRC Connected Inactive, for example, configurations for Discontinuous Reception (DRX), measurement, paging and location tracking, small data transmission, state transition, and so on. For example, a static massive MTC (mMTC) device with 15 years of battery life requirement can be configured with very long DRX cycles, for example, in hours or days, not to perform idle mode measurement, and to use contention based small packet transmission. On the other hand, an MBB UE may be configured with short DRX cycles, idle mode measurements for cell reselection, and to transition to RRC Connected to transmit data. Thus, there is no need for two independent inactivity states, one optimized for mMTC devices and another for MBB UE. RRC Connected Inactive can handle the low activity state of a wide range of 5G devices and UEs without increasing the number of RRC states.

Future Proofness and Quick Time to Market

It is not possible to predict all the 5G use cases that may arise from IoT, for example. The 5G use cases listed in [1] are not meant to be exhaustive. They rather serve as a tool to ensure that the level of flexibility required in 5G is well captured. Thus, future proofness should be a key criterion for 5G design.

Adapting the state handling of a new service to a static and highly standardized state handling mechanism might compromise the desired performance of the use case. Yet, modifications to a standardized functionality have proven in many cases to be complex and slow. If standardization of a new state handling and transition mechanism

is required for a new service, time-to-market is prolonged. This may require the 5G state handling model to have room for proprietary solutions so that new services can enter the market quickly.

The configurability of RRC Connected Inactive ensures that a level of future proofness can be achieved and new services can enter the market quickly. The behavior of a UE in RRC Connected Inactive can be tuned using standardized mechanisms such as DRX, and non-standardized solutions, for example, for effectively collecting information from sensors. Thus, the network can configure devices with a new service using a combination of solutions and parameters that fulfill its requirements. If there is room for tuning, UEs with new services whose requirements cannot be met with standardized solutions can enter the market quickly. However, such flexibility comes at the cost of implementational complexity. The upside is that the complexity resides on the network side.

State Transition Procedures

The state transition between RRC Idle and RRC Connected is assumed to follow the same procedure as in LTE. No special signaling procedure might be required for RRC Connected Inactive to RRC Idle transition. RRC Connected to RRC Connected Inactive transition can be achieved using a procedure similar to LTE's RRC Release procedure or RRC Reconfiguration procedure, where the signaling that orders the UE to move to RRC Connected Inactive includes a service tailored configuration [11].

RRC Connected Inactive to RRC Connected transition is initiated when UE has data to transmit, received a paging message, and so on. The state transition procedure can be implemented using a three-way message exchange, for example, like the RRC Resume procedure [6]. It may also be optimized to use only two messages. Due to the UE autonomous cell reselection procedure, the state transition may occur in the anchor gNB or a new gNB that does not have the UE context. If the state transition occurs in the anchor gNB, it is faster and has lower signaling overhead as the UE context is already available in the anchor gNB.

However, when the transition occurs in a new gNB, it introduces additional control plane latency and signaling overhead if UE context is not proactively prepared in the new gNB. If the RRC connection needs to be reestablished before data transmission can be started, like in LTE, the latency due to context fetching contributes significantly to the overall control plane latency. The signaling overhead due to context fetching and RAN/CN connection relocation is also significant especially if the UE transmits only a small packet. One way to tackle the problem is to proactively prepare UE context throughout UE's RAN notification area, a list of cells where UE may move without updating its location. However, this may not be feasible due to high memory consumption and signaling overhead for large RAN notification areas. Therefore, other optimization approaches are required for state transition in a new gNB to minimize the control plane latency and signaling overhead impact.

We propose an RRC Connected Inactive to RRC Connected state transition procedure in a new gNB, as shown in Fig. 2. The state transition is defined over the logical link between the UE and the anchor gNB, unlike the conventional approach where it is defined over the radio interface between the UE and the new gNB. That is, the RRC Activation Request message is forwarded to the anchor gNB, and it is up to the anchor gNB's decision how to respond to the request, for example, whether to relocate the serving gNB role to the new gNB. The state transition procedure also decouples the user plane (UP) data transmission from the procedure to re-establish the RRC connection. These approaches have several advantages as discussed below.

Small Packet Relaying

The network may configure devices that intermittently transmit small packets for contention based small data transmission. If such a device initiates a state transition with a small packet included in RA message 3, the anchor gNB may decide not to re-establish the RRC connection, and thus not initiate context transfer to the new gNB nor RAN/CN connection relocation, as shown in Fig. 2a. Instead, the anchor gNB sends an indication to the device to remain in RRC Connected Inactive in RRC Activate message. This avoids signaling overhead and additional latency due to anchor gNB relocation. Such small data transmission without state transition and anchor gNB relocation is beneficial, especially when a UE is moving as the UE is likely to move out of the new gNB anyway.

Establishing Multi-Connectivity During State Transition

When a UE initiates a state transition in a new gNB, it may also be located under the coverage of the anchor gNB, especially in a heterogeneous deployment scenario. In such a situation, the anchor gNB may decide to reactivate the RRC connection between the UE and the anchor gNB as a Master RRC (M-RRC) connection and add a Slave RRC (S-RRC) connection between the UE and the new gNB, as shown in Fig. 2b. The anchor gNB transfers all the necessary S-RRC context information to the new gNB if the context is not already stored in the new gNB. Note that the new gNB becomes a secondary gNB (S-gNB). Such an approach allows a quick multi-connectivity setup during state transitions so that the UE can immediately start utilizing radio resources from multiple cells after transitioning to RRC Connected.

Fast System Access

When a UE receives an IP packet from the application layer, it already has security keys and parameters required to encrypt the packet for transmission over the radio interface. Therefore, in principle, the UE can start UP transmission before an RRC connection is established if the new gNB grants radio resources to the UE. One example to achieve this is shown in Fig. 2c. The UE includes its buffer status report (BSR) in RA message 3, which is a MAC PDU, multiplexed with RRC control PDU. The new gNB grants radio resources to the UE according to the received BSR, while re-establishing the RRC connection.

However, as the new gNB does not have the UE context and the anchor gNB is yet to verify the UE, there must be a mechanism to mitigate grant-

The behavior of a UE in RRC Connected Inactive can be tuned using standardized mechanisms such as DRX, and non-standardized solutions, for example, for effectively collecting information from sensors. Thus, the network can configure devices with a new service using a combination of solutions and parameters that fulfill its requirements.

Although most state transitions occur in the anchor gNB, they may also occur in a new gNB. This may lead to additional signaling due to context fetching and RAN/CN connection relocation. As a result, signaling reduction is smaller, 43 percent and 27 percent when compared to RRC Idle and RRC Suspended, if context fetching and RAN/CN connection relocation are required.

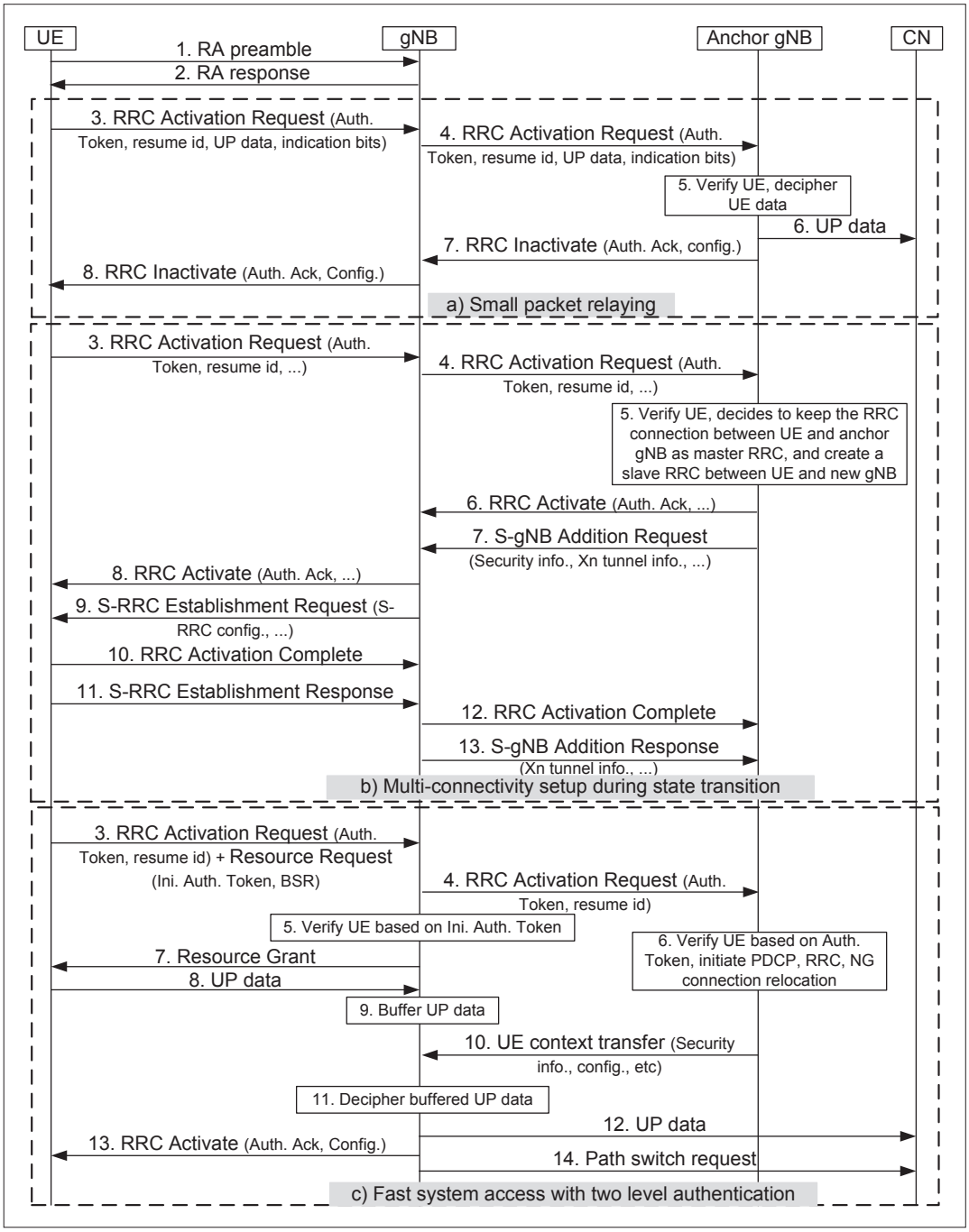

Figure 2. RRC Connected Inactive to RRC Connected transition from a new gNB. Note that RRC Activation Request consists of an authentication token prepared using UE specific key. It may optionally include an authentication token prepared using non-UE specific key that works in a group of gNBs.

ing resources to malicious UEs. One approach is to use a two-level authentication process. An authentication token prepared using a security key valid over a group of gNBs, referred to as an initial authentication token, can be used. The new gNB verifies the UE using the initial authentication token before granting resources for UP transmission. When the anchor gNB receives the RRC Activation request, it verifies the UE using the authentication token prepared using a UE specific key. The new gNB buffers the received UP data during state transition. Once it receives the UE context from the anchor gNB, it deciphers and forwards the data to the CN.

Inter-LTE/5G Inactive Mode Mobility Support

Supporting autonomous inter-Radio Access Technology (RAT) cell reselection, more specifically between LTE connected to 5G CN and NR, minimizes UE power consumption and increases the usability of RRC Connected Inactive. This requires extending the notification area coverage to include cells from LTE and NR. However, this will be challenging to implement if LTE does not support RRC Connected Inactive; a UE would be required to autonomously change its RRC state to an equivalent state, for example, Light Connected. With the proposed state transition approach, LTE does not necessarily need to support RRC Con-

nected Inactive; rather, it needs to know how to forward an RRC Activation request to the anchor gNB. This can be done by including a RAT ID as part of resumeID, which currently includes the anchor gNB ID and UE context ID, assuming the resume ID encoding is shared between the RATs. Upon receiving the RRC Activation Request, the anchor gNB decides whether to initiate RAN/CN connection relocation in the same manner as inter-gNB, intra-RAN state transition.

Performance Analysis

Here, the performance of RRC Connected Inactive is assessed in terms of signaling overhead, control plane latency and UE power consumption. The performance of LTE's RRC Idle and RRC Suspended are taken as the baseline.

Signalling Overhead

State transition signaling overhead is an important Key Performance Indicator (KPI) considering the high number of applications that transmit small packets. The state transition from RRC Idle to RRC Connected involves 9, 3, 2 messages over air, the S1 and S11 interfaces, respectively [13]. On the other hand, the transition from RRC Suspended requires 5, 2, 2 messages, correspondingly [6]. However, only four messages over the air interface are required for state transition from RRC Connected Inactive to RRC Connected as the RAN/CN connection is kept. This leads to an overall state transition signaling overhead reduction of 71 percent and 55 percent when compared to RRC Idle and RRC Suspended, respectively.

Although most state transitions occur in the anchor gNB, they may also occur in a new gNB. This may lead to additional signaling due to context fetching and RAN/CN connection relocation. As a result, signaling reduction is smaller, 43 percent and 27 percent when compared to RRC Idle and RRC Suspended, if context fetching and RAN/CN connection relocation are required.

Control Plane Latency

Control plane latency is the time interval from the first message in a power efficient state to the time a UE starts to transmit UP data [13]. In LTE, the control plane latency equals the state transition latency, that is, the time interval for a UE to transition from a low activity state to RRC connected; an LTE UE can only transmit data in RRC Connected. However, control plane latency in RRC Connected Inactive is not necessarily the same as the state transition latency due to the decoupling of data transmission and RRC connection reestablishment, as discussed earlier. For example, a UE configured for small data transmission can transmit small data during state transition piggybacked to an RRC Activation request message. Similarly, a UE with a large first packet that initiates a state transition in a new gNB can start data transmission in RA message 5 before the UE context is fetched to the new gNB and the RRC connection is reestablished. The 5G RRC state machine is required to have at least one power efficient state with a control plane latency of 10 ms [14].

In Fig. 3a the average control plane latency for RRC Connected Inactive is compared with that of RRC Suspended and RRC Idle. Performance can be analyzed combining average state transition, processing and RACH latencies with analysis of the state transition diagrams for RRC Idle [13], RRC Suspended [6], and RRC Connected Inactive (Fig. 2). For fair comparison, the average state transition, processing and RACH latencies are obtained based on the values for LTE, with the assumption of a 1 ms transmission time interval (TTI) [13], as summarized in Table 3. For example, the control plane latency of RRC Connected Inactive with a small packet, referring to Fig. 2 and Table 2, becomes 0.5 + 1 + 3 + 5 = 9.5ms. The control plane latency in RRC Connected Inactive is significantly lower than the figures for RRC Suspended and RRC Idle. When the state transition occurs in the anchor gNB, it is 88 percent and 84 percent lower with a small first packet (57 percent and 41 percent lower with a large first packet) than in RRC Idle and RRC suspended, respectively. If the state transition occurs in a new gNB, the figures are 88 percent and 87 percent lower with a small first packet (57 percent and 54 percent lower with a large first packet). Notice that the CP latency of RRC Connected Inactive remains the same whether the state transition occurs in a new gNB or the anchor gNB due to the decoupling of data transmission and RRC connection re-establishment. Therefore, the gain, compared to RRC

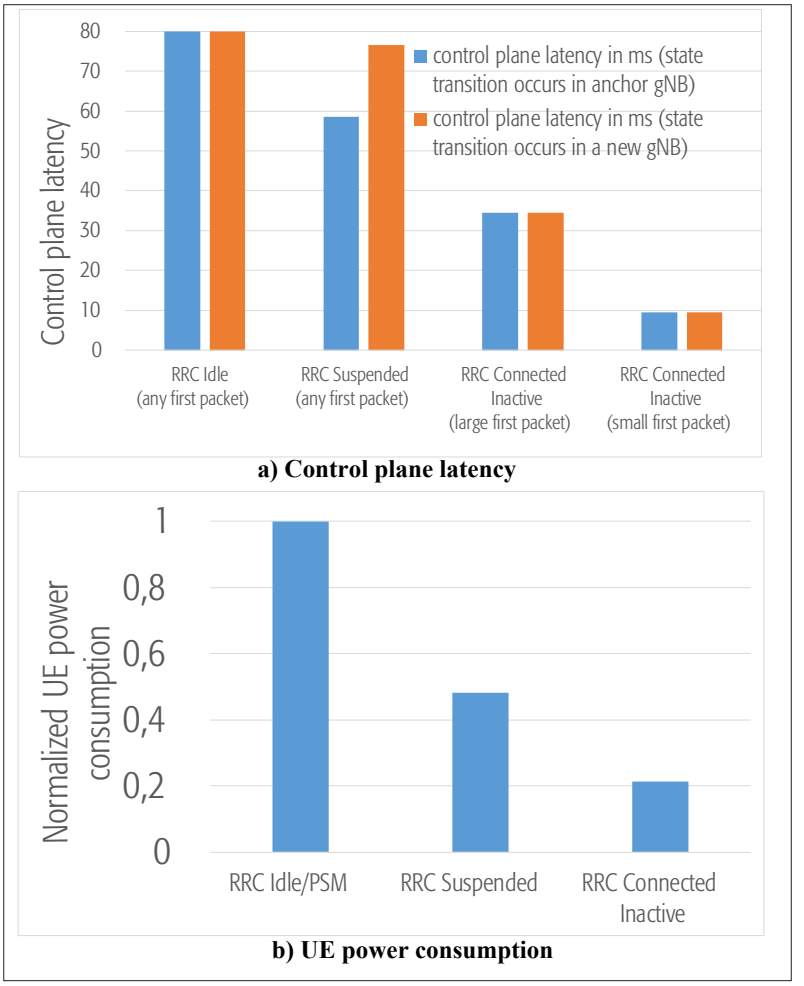

Figure 3. Comparison of control plane latency and UE power consumption to transmit small packet for RRC Connected Inactive, RRC Idle PSM and RRC Suspended.

Description	[ms]
Average delay due to RACH scheduling period	0,50
RACH Preamble	1,00
Preamble detection and transmission of RA response (time between the end RACH transmission and UE's reception of scheduling grant and timing adjustment)	3,00
UE processing delay (decoding of scheduling grant, timing alignment and C-RNTI assignment + L1 encoding of RRC connection request)	5,00
Transmission of RRC message between UE and eNB/gNB over air interface	1,00
Processing delay in eNB (L2 and RRC)	4,00
Processing delay in the UE (L2 and RRC)	15,00
Processing delay in eNB (RAN/CN interface message)	4,00
RAN/CN interface transfer delay	T_S1
MME processing delay (including UE context retrieval of 10ms)	15,00
Transmission of RRC security mode command and connection reconfiguration (+TTI alignment)	1,50
Processing delay in UE (L2 and RRC that includes RRCConnectionReconfiguration)	20,00

Table 3. Summary of processing and transmission latencies based on values for LTE [13]. T_S1 indicates the RAN/CN interface transfer delay, which may vary from 2 ms to 15 ms depending on the transmission media used.

Suspended is higher when the state transition occurs in a new gNB since RRC Suspended CP latency becomes higher due to context fetching. It is also worthwhile to note that the processing delay in 5G will be much lower than in LTE due to improved UE and base station processing capacity, shorter TTI length, and lighter signaling procedure.

UE Power Consumption

The UE power consumption of RRC Connected Inactive is analyzed for a use case with devices that transmit small packets and have long battery life requirement, for example, sensors. Such devices can be configured for contention based small data transmission with long DRX cycle, for example, equal to the periodic notification area update timer. If the device transmits data periodically, for example, temperature reports, the data transmission time can be further aligned with the periodic RAN area update timer. Thus, overall power consumption becomes highly correlated to the power consumed to transmit a small packet.

Figure 3b shows the power consumption for small packet transmission from RRC Connected Inactive, LTE's RRC Idle Power Saving Mode (PSM) and RRC Suspended. The UE power consumption model is based on [14], with parameters DL receive power -40 dBm, UL transmit power 15 dBm, 50 UL physical resource blocks, DL Modulation and Coding Scheme (MCS) index 12 and UL MSC index 12. The UE power is calculated as the total power consumed from the time a UE wakes up from DRX mode to transmit a small packet until it returns to DRX mode. RRC Connected Inactive consumes significantly less power than RRC Idle and RRC Suspended as UE power consumption is correlated to the time a UE is kept active. The power required to transmit a small packet from RRC Connected Inactive is 79 percent and 55 percent less than the power consumed from RRC Idle PSM and RRC Suspended, respectively.

Notice that the simulations do not consider error cases, for example, HARQ retransmission due to RACH failure, for simplicity. Since average values are used, the absolute gains remain the same even if error cases are considered. However, the relative gains would be slightly lower.

Conclusion

In this article, we discussed an RRC state machine for 5G. The state machine includes a novel intermediate state called RRC Connected Inactive, in addition to the conventional RRC Idle and RRC Connected states. The characteristics of RRC Connected Inactive are presented and shown to overcome the shortcomings of the state machines in existing systems such as LTE and HSPA. It is shown that RRC Connected Inactive has the required level of flexibility and configurability to fulfil the requirements of diverse 5G use cases. This minimizes the overall number of RRC States, which avoids implementation complexity and standardization effort arising from multi-state machines. Performance analysis for small packet transmission shows that RRC Connected Inactive is superior in terms of signaling overhead, latency and power consumption, with improvements of up to 71 percent, 88 percent and 79 percent over LTE's RRC Idle, respectively.

Acknowledgment

Part of this work has been performed in the framework of the H2020 project METIS-II co-funded by the EU. This information reflects the consortium's view, but the consortium is not liable for any use that may be made of any of the information contained therein. The authors would like to acknowledge the contributions of their colleagues.

References

[1] 3GPP, "Feasibility Study on New Services and Markets Technology Enablers stage 1," TR 22.891, Nov. 2015.
[2] 3GPP, "Interlayer Procedures in Connected Mode (Release 13)," TS 25.303, Dec. 2015.
[3] S. Sesia, M. Baker, and I. Toufik, *LTE — The UMTS Long Term Evolution: From Theory to Practice*, John Wiley & Sons, Jan. 2011.
[4] 3GPP, "LTE Radio Access Network (RAN) enhancements for diverse data applications (Release 11)," TR 36.822, Sept. 2012.
[5] 3GPP, "Study on Machine-Type Communications (MTC) and other mobile data applications communications enhancements," TR 23.887, Dec. 2013.
[6] 3GPP, "Study on architecture enhancements for Cellular Internet of Things (Release 13)," TR 23.720, Mar. 2016.
[7] 3GPP, "Evolved universal terrestrial radio access (E-UTRA); radio resource control (RRC); protocol specification," TS 36.331, Oct. 2016.
[8] Huawei *et al.*, "Introduction of Light Connection in LTE Stage-2 Specification," R2-1702016, Feb. 2017.
[9] 3GPP, "Study on New Radio Access Technology; Radio Interface Protocol Aspects (Release 14)," TR 38.804, Mar. 2017.
[10] I. L. Da Silva *et al.*, "A Novel State Model for 5G Radio Access Networks," *Proc. IEEE Int'l. Conf. Commun. (ICC) Workshops*, May 2016, pp. 632–37.
[11] S. Hailu, M. Säily, and O. Tirkkonen, "Towards a Configurable State Model for 5G Radio Access Networks," *Proc. Global Wireless Submit 2016 (GWS-2016)*, River Publishers, Dec. 2016, pp. 123–27.
[12] 3GPP, "Feasibility study for evolved Universal Terrestrial Radio Access (UTRA) and Universal Terrestrial Radio Access Network (UTRAN) (Release 13)," TR 25.912, Dec. 2015.

[13] 3GPP, "Feasibility study for Further Advancements for E-UTRA (LTE-Advanced) (Release 14)," TR 36.912, Mar. 2017.
[14] 3GPP, "Study on Scenarios and Requirements for Next Generation Access Technologies; (Release 14)," TR 38.913, Mar. 2016.
[15] A. R. Jensen *et al.*, "LTE UE Power Consumption Model: For System Level Energy and Performance Optimization," *Proc. IEEE Vehicular Technology Conf. (VTC Fall)*, Sept. 2012, pp. 1–5.

Biographies

Sofonias Hailu (sofonias.hailu@aalto.fi) received his B.Sc. in electrical engineering from Mekelle University, Ethiopia in 2007 and his M.Sc. degree in communications engineering from Aalto University, Finland in 2014. Currently, he is pursuing a D.Sc. (Tech.) degree in communication engineering from Aalto University in the field of 5G mobility, spectrum and radio resource management.

Mikko Säily (mikko.saily@nokia-bell-labs.com) received his B.Sc. in embedded systems and computer engineering from Raahe School of Engineering and Business. He joined Nokia in 1994 and worked as a senior specialist in the areas of wireless communication algorithms and radio performance. He has published 30+ papers and book chapters and holds 100+ granted patents or patent applications in wireless communications. He is currently leading research in 5G connectivity and mobility at Nokia bell labs.

Olav Tirkkonen (olav.tirkkonen@aalto.fi) received his M.Sc. and Ph.D. degrees in theoretical physics from Helsinki University of Technology, Finland. Currently he is an associate professor in communication theory at the Department of Communications and Networking in Aalto University, Finland. His current research interests are in coding theory, multi-antenna techniques and cognitive management of 5G cellular systems.