Towards Understanding the Signaling Traffic in 5G Core Networks

Simon Raffeck, Stefan Geißler, Tobias Hoßfeld
University of Würzburg, Institute of Computer Science, Chair of Communication Networks
Würzburg, Germany
prename.surname@uni-wuerzburg.de

I. Introduction

The Fifth Generation (5G) communication technology, its infrastructure and architecture, though already deployed in campus and small scale networks, is still undergoing continuous changes and research. Especially, in the light of future large scale deployments and industrial use cases, a detailed analysis of the performance and utilization with regard to latency and service times constraints is crucial. To this end, a fine granular investigation of the Network Function (NF) based core system and the duration for all the tasks performed by these services is necessary. Understanding the 5G architecture and how all its core NFs interact with each other is crucial in order to optimize and evaluate these novel systems. The focus of this work will be the presentation of a methodology that enables us to capture detailed data regarding the Service Based Interface (SBI) connections between NFs as well as the Next Generation NodeB (gNB). A detailed description of the occurring communication is omitted here, but can be found in the 3rd Generation Partnership Project (3GPP) standards [1] or in tutorials like [2].

The SBI connections manage the connectivity of the various NFs of the 5G core between each other. To this end, the different functions will exchange HTTP2 messages to push and retrieve data from one another. To convey messages to the User Equipment (UE) the core relies on the Next-Generation Application Protocol (NGAP) to communicate with the gNB. Furthermore, the Packet Forwarding Control Protocol (PFCP) is used to establish the connection between control and user plane and therefore between Session Management Function (SMF) and User Plane Function (UPF). Being able to capture the content and the timestamps of these connections helps greatly in verifying, if a core implementation is keeping up to the standards, and in further modeling the core behavior. To this end, we are hosting an Open5GS [3] deployment connected to a single radio cell realized using srsRAN [4] attached to an USRP B210 SDR. The core functions are hosted as separate VMs on top of a single hypervisor host. This leaves us in full control of the connections while minimizing the transport duration of messages and supplying us with a time synchronization between all VMs, since the capture can be performed on the hypervisor machine. In this work, we leverage these benefits to examine the core behavior during UE authentication and to evaluate the service times for different phases of the device registration. The goal of this preliminary investigation is to showcase the methodology of monitoring and parsing the signaling traffic occurring within the 5G core network.

II. Related Work

The performance of 5G networks is the subject of many other works within the research community. However, detailed models of the signaling load within the 5G core architecture is, to the best of our knowledge, still a gap in literature. A fine-granular evaluation of service times for each task a NF has to fulfil as well as a model describing the interactions between NF is yet to be conducted.

The authors of [5] focus on modeling an SDN-based approach to 5G networks, however, their analysis and modeling approach is strictly related to the network as a whole and the impact an SDN-based infrastructure has on it. In [6] various models for optimizing network loads and data-center resources, and NF placement are introduced and evaluated. The focus of these models is more on the network infrastructure and placement of functions. In [7] the authors propose an optimization solution for network slicing in 5G to satisfy end-to-end delay constraints. The authors of [8] model the end-to-end delay for embedded NF chains and propose a queuing model to evaluate the packet delay. All these works, however, only focus on the 5G network as a whole or on very specific NFs. In this work, we instead focus on developing a methodology towards capturing the signaling traffic within the core itself. This allows us to identify service times of every NF, which in return enables us to identify the impact each function has on overall system performance.

III. Methodology and Testbed

This work aims to build towards a methodology to help build a detailed model of the signaling traffic within the 5G core. To this end, we perform a full packet capture of the signaling traffic occurring during device authentication. For this, a SIM8200EA-M2 modem has been configured to authenticate with our 5G Campus network comprised of a srsRAN gNB operated with a B210 SDR, and an Open5GS core. The traffic between the different NFs was recorded without SSL encryption and can therefore be used for further analysis of the observed signaling behavior. The captured

traffic was then further processed to identify different signaling interactions between NFs. Based on the HTTP2 *streamID*, we are able to differentiate and isolate the various traffic streams and our hypervisor based packet capture eliminates the need for time synchronization between NF hosts. In the following, use the captured data to visualize the signaling behavior within the core and provide a preliminary investigation of the service times of different phases during the authentication and attachment procedure of a 5G end device.

IV. Analysis of Signaling Traffic

In the 5G network architecture, the various core functionalities are split up into different NFs. These NFs are furthermore separated into control plane and user plane functions. Thereby, the control plane is responsible for authentication, policing, mobility, billing and related functions while the user plane provides the functionality to establish tunnels towards a data network for user traffic.

In the following, we present the signaling data captured in our testing environment, showcase the available information and provide sequence diagrams of the signaling procedures occurring when a UE attaches to the system.

A. Data Structure and Preprocessing

To gather information about the single interactions between different NFs, we capture the network traffic caused by the 5G core, since all functions run on the same server but separated into different VMs, the message patterns match those of a distributed core deployment. Due to this setup, we can gather all NF traffic in one PCAP file, while having the packet timestamps all stem from the same clock. After sorting the entries by time, we make use of a HTTP2 tuple consisting of source and destination IP addresses and ports and the unique HTTP2 *streamID* to identify the single interactions or dialogs between NFs. After isolating the different packets into dialogs, we reconstructed the different flows between the UE and the different NFs into a sequence diagram of the observed process. Something we further divided into three main parts: authentication, mobility, and PDU session establishment. The phases were separated using messages to or from the UE as reference points for context changes. The *Mobility* phase does not require exchange with the UE and thus, starts and ends between the first and last messages to the gNB of the other two phases. Since we are in full control of the 5G core itself, we were able to extract the service times each NF needed for each of the dialogs. These service times are now fed into a sequence diagram for each of the main three main parts, to deepen our understanding of the impact of signaling traffic within the core.

B. Authentication

As *Authentication* phase, we identify every exchange between the UE, the Access and Mobility Management Function (AMF) and other NFs from the *Initial UE Message* to the *Security Mode Command* and its response. A sequence diagram for this process is depicted in Figure 1. This phase is started with the *Initial UE Message* from the UE to the AMF, it contains the International Mobile Subscriber Identity (IMSI), supported security capabilities, and more information regarding the reached gNB and target AMF. The AMF requests the necessary Authentication Server Function (AUSF) from the Network Repository Function (NRF) afterward, and reports the authentication to it. The AUSF will then chose its target Unified Data Management (UDM) and generate the needed authentication data for the requested UE. The UDM will get the necessary authentication subscription from the target Unified Data Repository (UDR). Afterward, an authentication request is sent to the UE. After the device checks the MAC integrity with its own *Milenage* calculation, an authentication response is sent. The AMF informs the AUSF of the successful authentication, the AUSF notifies the UDM, which forwards this information to the UDR. Lastly, a security mode command is sent to the UE, to which the UE replies to finalize the authentication routine. This entire process has taken 243 ms on average for 10 runs in our system core.

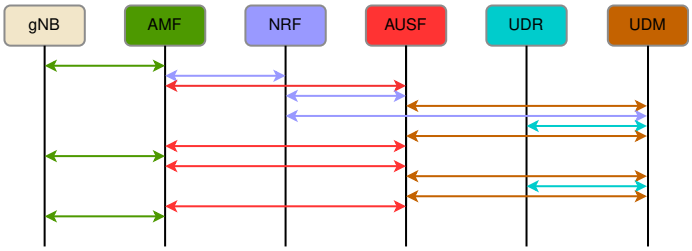

Figure 1: Sequence diagram for the device authentication after the initial UE message to the AMF.

C. Mobility

As *Mobility* phase, we identify every exchange between the AMF, the UDM and other NFs from the initial request from the AMF to the NRF, to the final answer from the Policy Control Function (PCF). The process is visualized in Figure 2. The phase starts out with the AMF inquiring the destination of the target UDM from the NRF. The AMF registers its connection with the newly authenticated device at the UDM, which forwards this information to the UDR. Afterward, the AMF requests access and mobility data for the subscription of the new IMSI from the UDM, which in turn forwards this request to the UDR and passes its received answer to the AMF, which contains information about the maximum uplink and downlink speed for this subscription. Next up, the AMF inquires about the SMF responsible for this UE. The request is again posted to the UDM and in turn to the UDR and answered by providing the AMF with information about the configured Data Network Name (DNN) and network slice. Afterward, the AMF inquires at the UDM if the new UE is already registered with a SMF session and if so, requests the information about this session and changes the Subscriber Data Management (SDM) subscription at the UDM. Lastly, the AMF creates a new UE policy association at the PCF, which in turn gets forwarded to the UDR and the PCF finalizes this exchange by providing all necessary subscription data and policies to the

AMF. This whole process has a mean service time of 8.75 ms for 10 runs in our core.

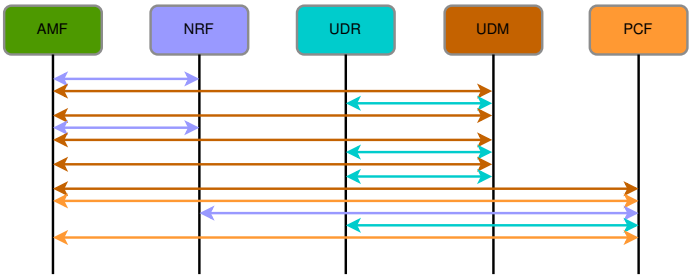

Figure 2: Sequence diagram of the Mobility phase right after device authentication.

D. PDU Session Establishment

As *PDU Session Establishment* phase, we identify every exchange between the UE, the SMF and other NFs from *InitialContextSetupRequest* by the UE to the Session Management (SM) context update from the AMF. This phase is depicted in Figure 3. The process begins with an *InitialContextSetup* between the AMF and the UE. Next, the AMF sets up a SM context with the SMF and sends a Packet Data Unit (PDU) session establishment request. The SMF now requests the target UDM and inquires SM data and information about slicing and DNN. The UDM forwards this to the UDR and supplies vital subscription and SM information to the SMF. Now, the SM policies at the PCF are updated by the SMF. Afterward, the PCF requests subscription and policy data from the UDR and requests the target Binding Support Function (BSF) from the NRF. The PCF now updates the BSF with the subscription information of the UE before updating the session rules at the SMF. The SMF and the UPF now establish a PFCP session for the user plane connection within the 5G network between the UPF and gNB via tunneling. Lastly, the SMF relays all this new information to the AMF before sending out a PDU session establishment accept message. The AMF then sends a PDU session resource setup request to the UE and the and after the reply from the UE, it updates the SM context at the SMF. In response to this, the SMF modifies the PFCP session with the UPF before concluding this procedure. This process has taken a mean of 279 ms for 10 runs in our Open5GS deployment.

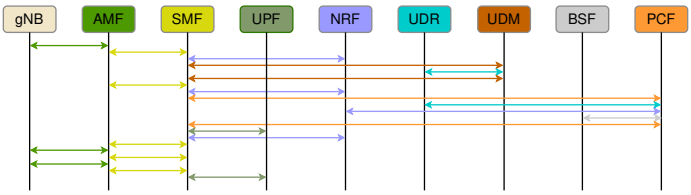

Figure 3: Sequence diagram for the PDU session establishment procedure.

E. Evaluation of Service Times

In the following, we evaluate the service times for each of the three phases. The statistical characteristics for each of the phases is further detailed in Table I. Furthermore, Figure 4 provides a stacked bar plot of the measured durations for each phase. The x-axis shows the individual runs. On the y-axis, the time for each phase is displayed in ms and each phase is color-coded as displayed in the legend. The data shows that the *Mobility* phase exhibits significantly faster service times compared to the other two phases. This can be traced back to the fact, that this phase is solely taking place in the core system itself and does not need to communicate with the UE via the air-interface. Transmission and modulation seems to significantly impact the time it takes to complete the tasks required for *Authentication* and *PDU Session Establishment*, thus the gNB and UE communication has the biggest impact on the system. This can also be seen in the variance for the service times of the *Authentication* and *PDU Session Establishment*.

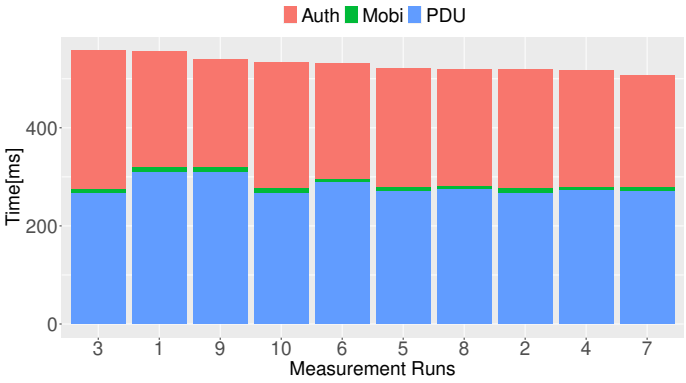

Figure 4: Service times of the three phases for 10 runs within our core.

Table I: Statistical characteristics of service times[ms].

Phase	Mean	Std.	CV	Min	Max
Authentication	242.98	16.72	0.069	220.11	281.81
Mobility	8.75	1.74	0.20	6.61	11.57
PDU Session Establ.	278.97	16.82	0.06	266.51	308.81

V. Conclusion and Discussion

In this work, we present first steps towards analyzing the signaling behavior within 5G core networks. We outline our methodology to capture and process the signaling traffic occurring between NFs within the core network as well as the gNB and UE.

To identify individual NF tasks, the HTTP2 messages used by the SBI connections are split up and ordered using their unique *streamID* and aggregated into signaling interactions between NFs. The sequence diagrams are merged and presented in his work, organized into three distinct phases, the *Authentication*, *Mobility* and *PDU Session Establishment*. Furthermore, we show preliminary results regarding the service times for each of these phases for 10 runs within a running Open5GS core network and looked closer into what factors might impact these times. Especially, the involvement of the air-interface between gNB and UE appears to impact the service times, while introducing jitter into the distribution.

For future work, we aim to automate the process of traffic capture and preprocessing while retaining as many details as possible. We further plan to perform detailed measurements in order to obtain the information required to develop detailed models of the signaling behavior within the 5G core network. To this end, we not only need to obtain detailed processing time distributions of each of the signaling procedures within the core, but also develop models for the behavior of UEs, to be able to generate realistic load scenarios for the system.

REFERENCES

[1] Technical realization of service based architecture;stage 3(3gpp ts 29.500 version 16.5.0 release 16. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/129500_129599/129500/16.05.00_60/ts_129500v160500p.pdf

[2] M. Shafi, A. F. Molisch, P. J. Smith, T. Haustein, P. Zhu, P. De Silva, F. Tufvesson, A. Benjebbour, and G. Wunder, "5g: A tutorial overview of standards, trials, challenges, deployment, and practice," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 6, pp. 1201–1221, 2017.

[3] S. Lee. Introduction to open5gs. [Online]. Available: https://open5gs.org/open5gs/docs/guide/01-quickstart/

[4] srsran project. [Online]. Available: https://www.srsran.com//5g

[5] A. Abdulghaffar, A. Mahmoud, M. Abu-Amara, and T. Sheltami, "Modeling and evaluation of software defined networking based 5g core network architecture," *IEEE Access*, vol. 9, pp. 10 179–10 198, 2021.

[6] A. Basta, A. Blenk, K. Hoffmann, H. J. Morper, M. Hoffmann, and W. Kellerer, "Towards a cost optimal design for a 5g mobile core network based on sdn and nfv," *IEEE Transactions on Network and Service Management*, vol. 14, no. 4, pp. 1061–1075, 2017.

[7] D. Sattar and A. Matrawy, "Optimal slice allocation in 5g core networks," *IEEE Networking Letters*, vol. 1, no. 2, pp. 48–51, 2019.

[8] Q. Ye, W. Zhuang, X. Li, and J. Rao, "End-to-end delay modeling for embedded vnf chains in 5g core networks," *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 692–704, 2019.