2022 18th International Conference on Distributed Computing in Sensor Systems (DCOSS) | 978-1-6654-9512-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/DCOSS54816.2022.00075

Prototyping a Fine-Grained QoS Framework for 5G and NextG Networks using POWDER

Udhaya Kumar Dayalan, Rostand A. K. Fezeu, Timothy J. Salo, Zhi-Li Zhang
Department of Computer Science & Engineering, University of Minnesota – Twin Cities, U.S.A.
{dayal007, fezeu001, salox049, zhang089},@umn.edu

***Abstract*—Unlike previous generation cellular technologies, 5G networks support diverse radio bands from low-band, mid-band to (mmWave) high-band, and offer a wide variety of new and enhanced features. In particular, 3GPP 5G standards adopt a flow-based 5G Quality-of-Service (QoS) framework that allows more flexibility in mapping QoS "flows" to data radio bearers. Nonetheless, the 5G QoS classes are pre-defined and QoS treatment is limited to the "flow" level. As we will argue in an earlier paper, the 5G QoS framework cannot fully and intelligently utilize the diversity of 5G radio bands and other capabilities to cope with fast varying channel conditions, and is therefore inadequate in meeting the quality-of-experience (QoE) requirements of many emerging applications such as augmented/virtual realities (AR/VR) and connected and autonomous vehicles (CAV). This has led us to advance a novel *software-defined*, *fine-grained* QoS framework for 5G/NextG networks.**

In this "work in progress" paper, we share our initial experience in prototyping the proposed fine-grained QoS framework. Our framework extends both the 5G core network and 5G radio access network (RAN) functionality to enable intelligent control of radio resources in a fashion that exploits application semantics to improve user QoE. We discuss in detail about the changes in different systems and its individual components, share the current state of implementation progress (work completed and in-progress) and finally our evaluation plan to validate the framework when the implementation is complete.

***Index Terms*—QoS framework, 5G and NextG, application semantics, software-defined, fine-grained**

I. Introduction

Emergent 5G networks introduce a wide spectrum of radio bands, from 5G *low-band* (sub-1GHz spectrum bands) to 5G *mid-band* (1 GHz – 7.125 GHz frequencies) and 5G *high-band* (24GHz – 60 GHz) including mmWave radio bands. These spectrum radio bands coupled with the so-called flexible numerology[1] and frame structure, dynamic single/mini-slot scheduling (with dedicated Downlink (DL) and Uplink (UL) symbols) and semi-persistent scheduling (SPS), slot configuration and aggregation with preemptive scheduling mechanisms and more is envisage to pillar new use cases, from Ultra-Reliable Low-Latency Communication (URLLC), and massive Machine Type Communication (mMTTC) to enhanced Mobile Broadband (eMBB) that will give rise to new application services like vehicle-to-pedestrians (V2P), vehicle-to-networks (V2N), Telemedicine, ultra-high resolution (UHR) (8K & volumetric) live video streaming applications, and augmented/virtual reality (AR/VR).

These 5G promises and new capabilities trigger an important question: Is 5G's unique and complex architectural design and proposed flexible features "*Enough*" to support envisioned 5G applications and services, some of which have high reliability, confidentiality and security demands, others demand ultra-high bandwidth and ultra-low latency with massive connectivity and supreme user quality of experience, or any combination of the above? Specifically, can 5G's high sensitivity to obstruction, especially mmWave 5G (e.g., moving cars/people, building, trees etc.) support existing video streaming services?

In this paper, partly motivated by our recent measurement study [2], [3], [4], we explore this question. We find that existing adaptive bit rate (ABR) algorithms fail to keep-up with the high, and varying mmWave 5G throughput, as a results degrades video streaming quality and overall quality-of-experience (QoE) [4], [5]. In other words, we argue that *the current 5G QoS architecture fall-short to effectively exploit 5G NR new capabilities when there is a need to make fast decisions and dynamically adapt to changing channel conditions to support ultra-high-resolution (UHR) video streaming over 5G.* 5G has introduced a new *service data adaption protocol* (SDAP) sublayer in the air interface protocol to manage the QoS via a QoS *flow-based* mechanism. Each QoS flow is *predefined* with a QoS profile and fixed QoS metrics which allows the user-plane functions (UPFs) to tag each IP flow with its appropriate QoS flow identifiers (QFIs) based on its profile.

To effectively take advantage of 5G new diverse features and frequent high 5G throughput fluctuations, we advanced *scalable layered coding* (SVC). SVC will provide the ultra-high bandwidth needed for UHR video streaming. For instance, using SVC video over mmWave 5G, the best, Line-of-Sight (LoS) beam could be used to deliver the layer one (base) chunk of the video and the secondary beam, non-LoS can be used to deliver subsequent enhanced layers based on the available radio resources. However, the existing *coarse* and *inflexible* 5G QoS framework cannot differentiate the base layer chunk from subsequent enhancement layer chunks within the same SVC video IP flow. Moreover, 5G QoS profiles are carrier specific (*carrier-centric*) and cannot utilize application semantics (even if available) for intelligent decision making, thus maximizing user QoE.

In an earlier position paper [6], we argue that due to the fixed QoS classes defined by 3GPP, the current 5G QoS framework (see §II-A for a quick overview) – while improving upon the so-called "EPS bearer"-based QoS framework of 4G LTE – is still too rigid; the "flow"-level QoS treatment

2325-2944/22/$31.00 ©2022 IEEE
DOI 10.1109/DCOSS54816.2022.00075

is also inadequate for diverse future applications such as AR/VR and connected and autonomous vehicles (CAV) that require both low-latency and high bandwidth. In particular, the 5G QoS framework cannot fully utilize the diversity of 5G radio bands and other capabilities to cope with fast varying channel conditions, and thus cannot intelligently meet the quality-of-experience (QoE) requirements of many emerging applications. This has led us to advance a novel *software-defined*, *fine-grained* QoS framework for 5G/NextG networks.

The basic idea behind our proposed *software-defined, fine-grained* QoS framework is a mechanism for cross-layer signalling of QoS requirements specified by the applications and dynamic adaptation of radio resources to meet the QoS requirements. More specifically: 1) We want to enable the application (or application service providers) to specify their application semantic tags (for data sub-streams and objects) and their associated QoS profiles and metrics. The applications will then negotiate and signal these QoS profiles to the 5G carriers and mark the data with the predefined QoS semantic tags. 2) Using the application specific semantic QoS tags, the 5G carrier can then install appropriate UPFs in the 5G core network to classify/filter, steer and preempt data packets accordingly based on the tags and set QFIs. By doing so, the 5G radio access network will be able to intelligently utilize the varying radio resources, beams and radio band characteristics to transport the right type of data for the best QoE, thereby maximizing the radio resources utilization. Our proposed framework design and key components are outlined in §II-B.

In this *work-in-progress* paper, we share our initial experience in prototyping the proposed fine-grained QoS framework. We use Open Air Interface (OAI) [7] 5G Core, OAI RAN (gNodeB/gNB) and OAI User Equipment (UE) to simulate an end-to-end 5G test bed and SVC for volumetric video streaming as a case study with trace driven emulation to demonstrate the potential of our proposed framework. Although our design and evaluation are in the infant stage, our goal in this paper is to demonstrate the need for a *software-defined*, *fine-grained* QoS framework for cross-layer signaling for application and network integration to support emergent new 5G applications and services.

II. 5G Fine-Grained QoS Framework

We propose to further extend the Semantically Aware, Mission-Oriented (SAMO) network framework [8] to enhance the ability of 5G networks to make decisions that offer applications an improved QoE. The SAMO framework, originally formulated for IPv4 and IPv6 networks, includes two components. First, applications can signal their QoS needs and other characteristics to the network by attaching semantic information as metadata on network packet headers with application data. Second, the network may consider the SAMO metadata when processing a packet. In its perhaps simplest application, the SAMO framework could permit an application to signal the priority of application data contained in the packet, and the network could endeavor to, when necessary, discard the packets that *the application* has signaled are the least important.

A. 5G QoS Framework

The 5G QoS architecture supports multiple QoS flows per PDU session, in contrast to 4G, which supports only one. A ”PDU session” is a data connection between a user equipment (UE), such as a cellular handset, and a host attached to an external data network (DN). The UPF tunnels PDUs (such as IPv4 or IPv6 packets) through the 5G core network between the DN and the RAN; these tunnels are effected by the GPRS Tunneling Protocol for the User Plane (GPT-U) protocol. PDUs tunneled through the UPF are encapsulated within the Service Data Adaptation Protocol (SDAP). The SDAP header carries the QoS Flow ID (QFI), which identifies the QoS Flow. PDUs within a QoS flow (i.e., that have the same QFI) receive similar QoS treatment. The QFI is set by the PDU Session Anchor, UPF function that receives PDUs from the DN and encapsulates them within a GPT-U tunnel. The PDU Session Anchor is responsible for setting the QFI in the SDAP header. That is, the PDU Session Anchor examines packets from the DN and classifies it into a particular QoS category. The PDU Session Anchor might, for example, translate the Differentiated Services Code Point (DSCP) value in the IP packet header into a QFI value. The 3GPP specifies a set of standard QFI values and their interpretations. The gNodeB, upon receiving a PDU from the UPF, examines the QFI value associated with the PDU and Data Bearer (e.g., a particular RF channel).

B. Proposed 5G Fine-Grained QoS Framework

Previously, we described an application of the SAMO framework to 5G networks, ”Fine-Grained QoS Framework for 5G and Beyond Networks”. This application of the fine-grained 5G QoS framework enables 5G networks to allocate 5G radio access network (RAN) resources, which guidance from the application, so as to maximize the application’s QoE. This use of the 5G QoS framework is particularly beneficial in 5G environments where the bandwidth available through the RAN may vary rapidly and dramatically. In this use case, an application that streams a layer coded video stream tags each packet with metadata that specifies the layer (or priority) of the application data in that packet. On ingress to the 5G network, the UPF translates the metadata to a (perhaps subscriber-specific) QoS Flow Identifier (QFI) value. This QFI value is transported through the UPF to the RAN: the QFI is carried in the Service Data Adaptation Protocol (SDAP) header. (The SDAP protocol is new in 5G, and carries QoS information as IP packets are tunneled through the 5G network by the UPF.) The RAN uses this QFI (which was derived from the application-provided metadata) to when assigning packets to specific RF resources (e.g., RF spectrum or beam).

In this section, we elaborate on the architecture of the fine-grained framework. Figure 2 includes the key components of the fine-grained framework. Fine-grain framework doesn’t add any new components to the 5G core or RAN or UE, instead

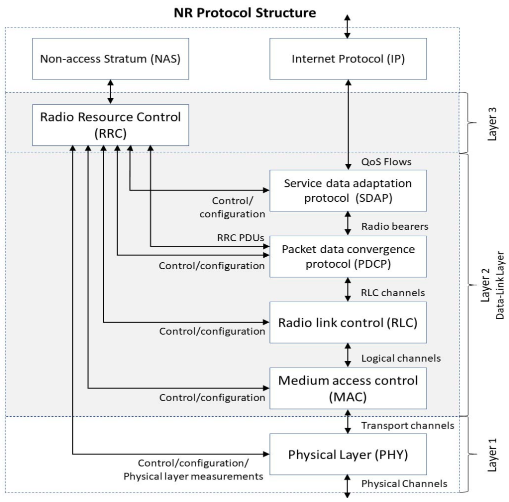

Fig. 1: 5G Radio Network Protocol Stack.

it enhances the existing components of the 5G core or RAN or UE.

III. Implementation

In this section, we elaborate on the implementation of the fine-grained framework. Figure 2 shows the system architecture of various systems involved in this paper. In this section, we will discuss in detail about the changes being made to individual system and its components. Fine-grain framework doesn't add any new components to the 5G core or RAN or UE, instead it enhances the existing components of the 5G core or RAN or UE. We used OAI 5G Core, OAI RAN and OAI UE which are open source.

We are working on a new application service provider (ASP) which resides in the mobile edge cloud (MEC) within or close to the 5G carrier network for reduced latency. The servers are running inside the MEC with direct connectivity to the 5G core network. The ASP provides the 5G control plane (via an ASP controller) with (service-specific) application semantics manifests and QoS profiles e.g., in the form of XML or JSON files (similar to the manifest files used in video streaming applications). The semantic tags are specified in these files which will be used by its applications to define (fine-grained) data objects or data streams such as layered video frames or chunks. The application marks the data with appropriate semantic tags for the (desired) QoS metrics and treatments over 5G. These semantic tags are implemented using IPv6 flow labels or extension headers [8]. The semantics manifests and QoS profiles can be dynamically updated and pushed to the 5G network by altering the number of data streams such as SVC layers based on bandwidth prediction or feedback from the 5G network. Next, we developed an Application Server Endpoint, which acts as an video-streaming application server. We developed an application data refactoring mechanisms, using scalable video coding (SVC) [6] for volumetric video streaming as a case study. Application constructs stream with layered coding (e.g., SVC). Each packet contains data for only one layer. In this application, the metadata identifies the SVC layer. We will modify the OAI RAN to use these metadata to assign the packet to the most appropriate channel. The application and the 5G core network share the semantics-QFI mapping. We extend the 5G/NextG UPF and SDAP to tag and encode application semantic metadata.

In the 5G/NextG control plane, the QFI values were defined for the desired service objectives and QoS metrics were defined specific to this ASP and its service based on the business agreement with the ASP. In order to set up and authenticate Packet Data Network (PDN) sessions for application flows and track the mobility of mobile client end points, and PCFs (policy control functions) to install services-specific QoS tables at the relevant UPFs in the user plane, we institute appropriate control functions, such as SMFs and AMFs. The 5G core control plane will instruct and configure the 5G RAN for intelligent radio resource control (RRC) functions. The QoS flows extends them to include (service-specific) semantics tags to map IP flows into finer-grained QoS "subflows" or QoS data streams similar to the flow tables used by SDN switches. Here is the form ⟨ flow header; semantic tags — QFI ⟩(including perhaps also ARP bits) of each entry of in the QoS table. Upon receiving the data packets from the application, the UPFs in the 5G user plane, assisted by service specific AFs, will process them based on the (service-specific) QoS tables for the desired QoS treatments – in particular, convert and encapsulate them to 5G packet data units (PDU) with appropriate QFI values.

On the 5G/NextG RAN side, O-RAN RIC [9] and RRC will intelligently allocate radio resources (channels, beams, Transmission Time Intervals (TTIs) based on radio channel quality and other characteristics/considerations. RAN uses QFI to select most appropriate radio channel for data. And finally, we are working on a ASP client application which runs on the end device such as an autonomous vehicle, an IoT device or a smart phone and functions as a client endpoint. Tthe end device is also equipped with an 5G radio interface that executes the radio protocol stack shown in Figure 1, it also signals the data semantics in-band to the 5G RAN to effect changes in data transmission. The UE receives the data through the best throughput radio channel based on the current signal conditions of the UE.

IV. Planned Evaluation

In this section, we elaborate on our plan to evaluate the fine-grained framework. We will conduct comparative and in-depth experiments to study the fine-grained framework for the support of emerging and future 5G/NextG services, and develop new solutions for tackling these challenges. The experiments will be conducted and include detailed experimental design and specifications, evaluation metrics, datasets, software artifacts that summarize our experimental results.

We will use volumetric video streaming where video frames are represented as 3D point cloud with *layered video coding* (SVC) over 5G/NextG as a use case to demonstrate and evaluate the potential benefits of our proposed software-defined,

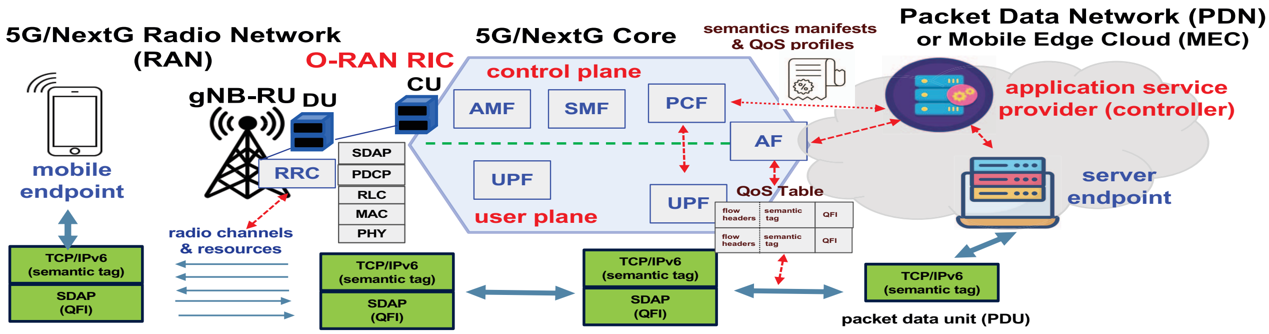

Fig. 2: Software-Defined, Fine-Grained QoS Framework for 5G and Beyond 5G (NextG) Networks. As stated in the text, semantic tag can be implemented using IPv6 flow labels or extension headers, or other mechanisms.

fine-grained QoS framework. We re-purpose IPv6 flow labels as "semantic tags" for the application service endpoints. By exposing application semantics (via "fine-grained" data sub-streams marked by semantic tags, in this case, *layered video chunks*), our framework will enable the 5G/NextG network to intelligently match and map the layered video chunks of differing utility to diverse radio channels of varying qualities, and dynamically allocate radio resources for their transmission so as to deliver the best user QoE. The video frames are encoded using SVC with a total of 5 layers, a base layer with a resolution corresponding using roughly 1/5 of the total points per frame, and 4 enhancement layers which progressively enhance the frame resolution. We will allocate up to 5 radio channels with differing channel qualities. In the case of the current **flow-based QoS** framework, data packets from the same frame will be *striped* across the allocated channels – i.e., *independent of the layers the data packets belong to*, as they all carry the same QFI value. In the case of our **fine-grained QoS** framework, we assign and transmit the base layer data packets using the radio channel with the best quality, the next (enhancement) layer data packet using the radio channel with the next best quality, and so forth. We will apply this framework to the evaluation of our proposed experimental activities on the NSF PAWR platforms.

Using our prototype implementation running on the POWDER platform, we will evaluate and compare our *fine-grained QoS* framework with the current flow-based 5G QoS framework. For this, we will leverage various empirical datasets (including our commercial 5G service measurement datasets), and perform trace-driven evaluation. We will also utilize the POWDER's "in the lab" RAN/RF frontend testbed [10] as well as mobile/fixed user end points for "in-the-wild" empirical evaluation and field trials. We will measure the success of our experiments along several key dimensions – radio resource *utilization*; *agility* in matching (fast varying) radio resources with application data streams with different semantics); *(end-to-end) performance* such as throughput, latency as well as *user QoE*.

V. Conclusions

In this paper, we proposed a novel software-defined, fine-grained QoS framework for cross-layer network and application integration that will improve future 5G and NextG applications. In particular, our framework exposes application semantics to the network for intelligent decision making, especially matching the diverse, constantly changing 5G radio conditions to application QoS requirements.

VI. Acknowledgement

This research was in part supported by NSF under grants CNS-1814322, CNS-1831140, CNS-1836772, CNS-1901103, CNS-2106771 and CNS-212848.

References

[1] 3GPP, "5G NR; Physical channels and modulation," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 38.211, 04 2018, version 14.2.2. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/138200_138299/138211/15.03.00_60/ts_138211v150300p.pdf

[2] A. Narayanan, E. Ramadan, J. Carpenter, Q. Liu, Y. Liu, F. Qian, and Z.-L. Zhang, "A first look at commercial 5g performance on smartphones," in *Proceedings of The Web Conference 2020*, ser. WWW '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 894–905. [Online]. Available: https://doi.org/10.1145/3366423.3380169

[3] A. Narayanan, E. Ramadan, R. Mehta, X. Hu, Q. Liu, R. A. K. Fezeu, U. K. Dayalan, S. Verma, P. Ji, T. Li, F. Qian, and Z. Zhang, "Lumos5G: Mapping and predicting commercial mmwave 5G throughput," in *IMC '20: ACM Internet Measurement Conference, Virtual Event, USA, October 27-29, 2020*. ACM, 2020, pp. 176–193. [Online]. Available: https://doi.org/10.1145/3419394.3423629

[4] A. Narayanan, X. Zhang, R. Zhu, A. Hassan, S. Jin, X. Zhu, X. Zhang, D. Rybkin, Z. Yang, Z. M. Mao, F. Qian, and Z. Zhang, "A variegated look at 5g in the wild: Performance, power, and qoe implications," *ACM SIGCOMM'21*, 2021.

[5] E. Ramadan, A. Narayanan, U. K. Dayalan, R. A. K. Fezeu, F. Qian, and Z. Zhang, "Case for 5g-aware video streaming applications," in *Proceedings of the ACM SIGCOMM Workshop on 5G Measurements, Modeling, and Use Cases*, ser. 5G-MeMU'21, 2021.

[6] Z. Zhang, U. K. Dayalan, E. Ramadan, and T. J. Salo, "Towards a software-defined, fine-grained qos framework for 5g and beyond networks," in *Proceedings of the ACM SIGCOMM Workshop on Network Meets AI & ML*, ser. NetAI'21, 2021.

[7] (2022) Open air interface. [Online]. Available: https://openairinterface.org/

[8] T. J. Salo and Z. Zhang, "Semantically aware, mission-oriented (SAMO) networks: A framework for application/network integration," in *Proceedings of the 2020 Workshop on Network Application Integration/CoDesign, NAI@SIGCOMM 2020, Virtual Event, USA, August 14, 2020*. ACM, 2020, pp. 41–42. [Online]. Available: https://doi.org/10.1145/3405672.3409490

[9] (2022) O-ran alliance. [Online]. Available: https://www.o-ran.org/

[10] J. Breen, A. Buffmire, J. Duerig, K. Dutt, E. Eide, A. Ghosh, M. Hibler, D. Johnson, S. K. Kasera, E. Lewis *et al.*, "Powder: Platform for open wireless data-driven experimental research," *Computer Networks*, vol. 197, p. 108281, 2021.