P4-NetFPGA-based network slicing solution for 5G MEC architectures

Ricart-Sanchez, Rubén; Malagon, Pedro; Alcaraz-Calero, Jose M.; Wang, Qi

Published in:
Proceedings of the 2019 ACM/IEEE Symposium on Architectures for Networking and Communications Systems (ANCS)

DOI:
10.1109/ANCS.2019.8901889

Published: 18/11/2019

Document Version
Peer reviewed version

P4-NetFPGA-based network slicing solution for 5G MEC architectures

1st Ruben Ricart-Sanchez
School of Computing and Engineering
University of the West of Scotland
Paisley, United Kingdom
Ruben.Ricart-Sanchez@uws.ac.uk

2nd Pedro Malagon
Departamento de Ingenieria Electronica
Universiad Politecnica de Madrid
Madrid, Spain
pedro.malagon.marzo@upm.es

3rd Jose M. Alcaraz-Calero
School of Computing and Engineering
University of the West of Scotland
Paisley, United Kingdom
Jose.Alcaraz-Calero@uws.ac.uk

4th Qi Wang
School of Computing and Engineering
University of the West of Scotland
Paisley, United Kingdom
Qi.Wang@uws.ac.uk

***Abstract*—Network Slicing is one of the fundamental capabilities of the new Fifth-generation (5G) networks. It is defined as several logical networks that are created to fulfil specific Quality of Service (QoS) and Quality of Experience (QoE) requirements and are available over the same physical infrastructure. This paper proposes a novel extension to P4-NetFPGA framework to achieve network slicing between different 5G users in the edge-to-core network segment. This solution provides hardware-isolation of the performance in terms of bandwidth, latency and packet loss of 5G network traffic. The work proposed has been validated in a real 5G infrastructure.**

***Index Terms*—5G mobile networks, P4, NetFPGA, FPGA, Network Slicing**

I. Introduction

This new generation of mobile network communications requires to shift the current "best-effort" approach to an approach that allows different vertical businesses to address ambitious use cases in terms of ultra-reliable low-latency communications (uRLLC), enhanced mobile broadband (eMBB), massive internet of things (mIoT) and massive machine-type communications (mMTC). In this context, network slicing is required to allow the creation of different logical networks that guarantee the Service Level Agreements (SLAs) of the ambitious KPIs required by the vertical businesses.

This paper demonstrates a novel P4-based framework which allows the definition and control of the slicing in a 5G Mobile Edge Computing (MEC) architecture, guaranteeing the complete isolation of 5G network traffic while fulfilling the ambitious 5G end-to-end latency KPI, defined as <1ms for critical traffic and <10ms for normal traffic [1]. The definition of network slicing has been based on an extended usage of the P4 language. The proposed system is based on a NetFPGA [2] network card and the NetFPGA-SUME project [3] and provides an extension to support network slicing. A functional prototype has been deployed in a real 5G infrastructure, carrying out an exhaustive empirical evaluation to demonstrate the scalability and reliability of the system.

Identify applicable funding agency here. If none, delete this.

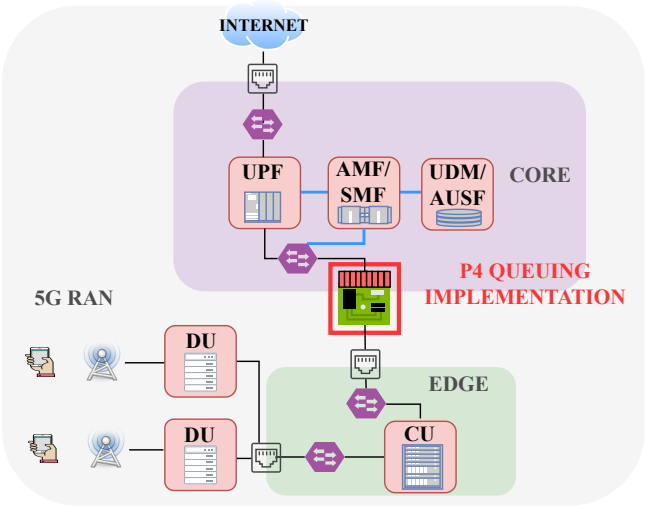

Fig. 1. 5G Edge to Core Architecture based on a P4 queuing implementation.

II. QoS-aware 5G architecture overview

The architecture shown in Figure 1 is clearly divided in three parts; i) 5G RAN with antennas and 5G users connected to such antennas; ii) Edge with CUs to allow users mobility; and, iii) Core with the data centre hosting the core of the network. Our solution allows us to define QoS network slicing over the 5G network traffic exchanged between Edge and Core. In this edge-to-core network segment, there is a General Packet Radio Service (GPRS) Tunneling Protocol (GTP) encapsulation to establish a tunnel between User Equipments (UEs) and Core Network. This tunnel is required to achieve user mobility across the different Distributed Units (DUs) of the architecture. Traditional network cards do not provide support for GTP and this research is an step forward to overcome such limitation,

978-1-7281-4387-3/19/$31.00 © 2019 IEEE

while also providing the possibility to define network slices over such traffic.

III. UWS NetFPGA-based core

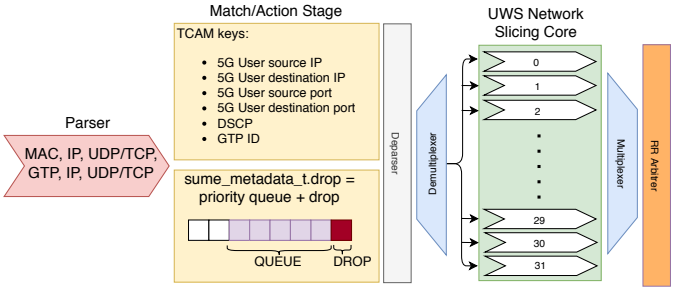

Fig. 2. NetFPGA pipeline extension with UWS Network Slicing Core implemented.

A 5G flow is composed by a traditional IP flow which is encapsulated over another IP flow using a GTP header (see Parser in Fig 2).

The parser extracts the 5G flow data that will be used as part of the definition of a network slice. A network slice is defined by the following 6-tuple: 5G user source and destination IPs, 5G user source and destination ports, Differentiated Services Code Point (DSCP) and GTP Tunnel ID. The foundations of this parser stage are based on our previous publication, just in case the reader is interested [4].

The match/action stage contains a TCAM table where the previous 6-tuple have been used to define a network slice and determine the actions to be taken over such slice. This stage has been extended to allow P4 to control the new UWS Network Slicing Core. The connection between our new core and the match/action stage has been achieved by modifying the *sume metadata* structure provided by the NetFPGA-SUME project. Fig. 2 shows the field used in *sume metadata* structure to drop a network packet. This is also used to choose the priority queue where the different 5G flows are sent. This mechanism allows us to control the latency of a given network slice.

At deparser stage, packets are rebuild and relayed to the UWS network slicing core created. This core guarantees the QoS and allows the isolation of the 5G traffic.

The UWS network slicing core provides a queuing discipline composed by 32 different queues which allows the definition of 32 types of QoS. It guarantees the complete isolation of the traffic in each queue. These queues have differentiated priorities, with queue 31 receiving the highest priority traffic and queue 0 the lowest. The queuing discipline is based on a priority-based algorithm where a queue with less priority will not be processed until the queues with higher priority are empty, enforcing an strict priority-based discipline.

The novel P4 pipeline is fully controlled by a software API which allows the insertion and deletion of rules in the TCAM table. The match aspect of the rule corresponds to the 6-tuple previously defined. The action aspect of the rule determines the queue ID. That ID is received in the pipeline when a flow matches with a rule of the table. Thus, the API provides the administrator of the network with full control of the queuing discipline implemented and also, the possibility to define up to 32 types of QoS for the 5G traffic that is being processed.

IV. Experimental validation

The solution proposed has been validated in a 5G edge-to-core infrastructure fully deployed in our data centre. This validation has been performed by sending traffic from 32 separate user equipments (UE) simultaneously and defining 32 different types of QoS by using 32 rules. The traffic sent congested the network. The framework under such circumstances achieved full isolation between the traffic transmitted by the different UEs in terms of latency, packet losses and bandwidth. The results obtained in terms of end-to-end latency have been under 0.5ms for the highest priority queues and under 3ms for the lowest priority queues in an end-to-end communication. These results achieved the ambitious 5G KPIs in terms of end-to-end latency, defined as $<$1ms and $<$10ms for critical and non-critical traffic [1].

V. Conclusions

This paper has proposed a novel QoS-aware network slicing solution based on P4-NetFPGA to meet the SLA requirements for network slicing. This solution guarantees total isolation between the different 5G users of the network based on the SLA established by the network operator and it fulfils the KPIs required to make our framework suitable for 5G networks.

Future work will analyse the different queuing algorithms in an attempt to find the most efficient and fair system. The work can be further extended to allow multi-tenancy scenarios which are a key feature to reduce capital and operational costs for telecommunication industries.

Acknowledgment

This work was funded by the European Commission Horizon 2020 5G-PPP Program under Grant Agreement Number H2020-ICT-2016-2/761913 (SliceNet: End-to-End Cognitive Network Slicing and Slice Management Frame-790 work in Virtualised Multi-Domain, Multi-Tenant 5G Networks), and by the Spanish Ministry of Economy and Competitiveness under contract RTC-2016-5434-8.

References

[1] N. Alliance, "5g white paper," *Next generation mobile networks, white paper*, pp. 1–125, 2015.

[2] J. W. Lockwood, N. McKeown, G. Watson, G. Gibb, P. Hartke, J. Naous, R. Raghuraman, and J. Luo, "Netfpga-an open platform for gigabit-rate network switching and routing," in *2007 IEEE International Conference on Microelectronic Systems Education (MSE'07)*. IEEE, 2007, pp. 160–161.

[3] N. Zilberman, Y. Audzevich, G. A. Covington, and A. W. Moore, "Netfpga sume: Toward 100 gbps as research commodity," *IEEE micro*, vol. 34, no. 5, pp. 32–41, 2014.

[4] R. Ricart-Sanchez, P. Malagon, J. M. Alcaraz-Calero, and Q. Wang, "Hardware-accelerated firewall for 5g mobile networks," in *2018 IEEE 26th International Conference on Network Protocols (ICNP)*. IEEE, 2018, pp. 446–447.