NetFPGA-based Firewall Solution for 5G Multi-Tenant architectures

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

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

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

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

***Abstract*—Future fifth-generation (5G) mobile networks entails architectural and network changes, mainly motivated by the idea of sharing resources between different network operators, which implies a reduction of the costs, thanks to the deployment of virtualised scenarios in shared infrastructures, and an improvement of the network usability. These architectural changes should guarantee that security and 5G Key Performance Indicators (KPIs) are achieved in 5G multi-tenant scenarios. The deployment of advanced architectures and network scenarios for the emerging 5G networks involves a renovation of the elements that compose them. Nowadays, there is no hardware solution which ensures the protection in 5G edge to core multi-tenant scenarios, therefore this paper proposes a fully functional 5G firewall based on a Field Programmable Gate Array (FPGA) that allows effective detention of cyber-attacks in 5G multi-tenant scenarios with user mobility support. The prototyped 5G firewall has been empirically evaluated to validate new capabilities in a 5G edge-to-core scenario. Moreover, an extensive performance and scalability test of the prototyped system has been carried out in a realistic testbed.**

***Index Terms*—P4-NetFPGA, FPGA, 5G, network protection, mobility networks, Firewall**

I. Introduction

In the following years, mobile network infrastructures will be changing with the emergence of the Fifth-Generation (5G) networks, and thus the architectural elements that compose the new telecommunication infrastructures will evolve accordingly. The exponential growth of the number of connected devices is one of the main security concerns in 5G infrastructures, due to the fact that IoT and mobile devices can be easily manipulated to generate attacks using the high bandwidth provided by 5G infrastructures. These attacks can be used to collapse the core of the network operator in a short period of time, as much as 100 times faster than the Fourth-Generation (4G) scenarios as promised by 5G network data rates.

To achieve this data rates, networks infrastructures should be enough mature to be able to manage a huge volume of network traffic. The adaptation of the 4G scenarios to the emerging 5G infrastructures involve the improvement of the current network elements. Following this approach, this paper presents a novel 5G firewall that helps to prevent the overload of the 5G network architectures dropping network traffic during a cyber-attack. 5G networks consists of a Radio Access Network (RAN) segment, an Edge segment and a Core segment, but this paper is mainly focused in the segment between Edge and Core. The Core of the network is govern by a cloud computing stack which allow multi-tenancy isolation and softwarization, whereas the Edge segment also provide multi-tenancy and softwarization support thanks to a mobile edge stack.

The main aim of this contribution is to provide a prototype based on NetFPGA Network Interface Card (NIC) that makes use of the P4 language in order to develop a new hardware-accelerated 5G firewall for the novel 5G traffic. As well as a full prototype has been implemented and extensive empirical validation of the prototype has been carried out using a 5G Edge to Core infrastructure currently deployed in our data center.

The rest of the paper is organized as follows. Section 2 describes related works with other firewall solutions. Section 3 provides an overview of the 5G MEC architecture describing the main network elements of the architecture with the main 5G firewall architecture proposed. Section 4 shows the testbed setup used to validate and evaluate the solution prototyped. In this section, empirical results are also presented. Finally, section 5 concludes this paper and outlines future research work.

II. Related Work

Traffic filtering has been a hot topic in security for a number of years. There are well-known hardware-based and software-based filtering tools already available. From the software side, Linux Netfilter [1] is an open source packet filtering framework, which allows manipulating packets. This framework can be defined as a set of hooks that work inside the Linux kernel, and using kernel modules, they register to callback functions in the Linux networking system stack. These callbacks are called

when some packets go through their hooks. This framework is employed in Iptables [2], a user program that allows configuring the firewall implemented by Netfilter, thanks to a command-line tool provided by Iptables. There exits a Netfilter extension, developed by Don Cohen, called u32 [3] that allows the inspection of packet beyond the traditional IP packet. This extension allows jumping between packet headers and extracting 4 bytes, which are in any location. The main problem of this extension is that it only provides support for 100 characters per predicate. It is a strong limitation when the packets being analyzed have a lot of headers, for example with encapsulated traffic imposed by the Virtual Extensible Local Area Network (VXLAN) and GPRS Tunneling Protocol (GTP) required in 5G infrastructures to support user mobility and multi-tenancy. Salva-Garcia et al. [4] and Serrano et al. [5] propose the only two software tools which are able to work in a multi-tenant 5G network achieving traffic control over a Edge to Core scenario. However, these contributions do not provide a good scalability mainly due to the use software-based architectures proposed with no hardware-acceleration.

From the hardware side, a design of an Ethernet Firewall Based on Field Programmable Gate Arrays (FPGA) is proposed by [6]. The solution allows the filtering and inspection of packets under a traditional IP network architecture, reducing the latency with respect to software-based solutions and allowing achieve up to 950 Mbps of throughput. In [7], a hybrid hardware/software-based firewall approach is introduced using NetFPGA SUME [8] and NetFilter [1]. This solution allows packet header analysis and classification in hardware and makes use of Netfilter to apply complex decisions over traffic flows. In [9], it is proposed to use Simple Classification and Regression Trees (Simple CART) matching learning algorithm for traffic classification developed on FPGA. In this solution, the authors use the dual Read/Write ports of the BRAM memory to process two search in every clock. Ricart-Sanchez et al. [10] proposes a hardware-based data-path for 5G multi-tenant scenarios. Authors provide an exhaustive analysis in terms of performance, scalability and reliability of the data-path developed, however they do not present any security solution for 5G multi-tenant scenarios which has been the main motivation of this research work. In [11] authors provide a firewall for 5G Edge to Core scenarios, but they do not provide multi-tenancy support and do not present experimental data about the solution proposed.

Based on the the review of related work and to the best of our knowledge, there is a gap of firewall solutions to fit in the novel 5G infrastructure where user mobility, multi-tenancy and the VXLAN and GTP protocols need to be supported in order to allow 5G user protection against cyber-attacks, which has been the main motivation and contribution of this work.

III. Proposed 5G multi-tenant Firewall Architecture

A. Architectural design

Figure 1 provides an overview of a 5G multi-tenant infrastructure deployed employing a Mobile Edge Computing

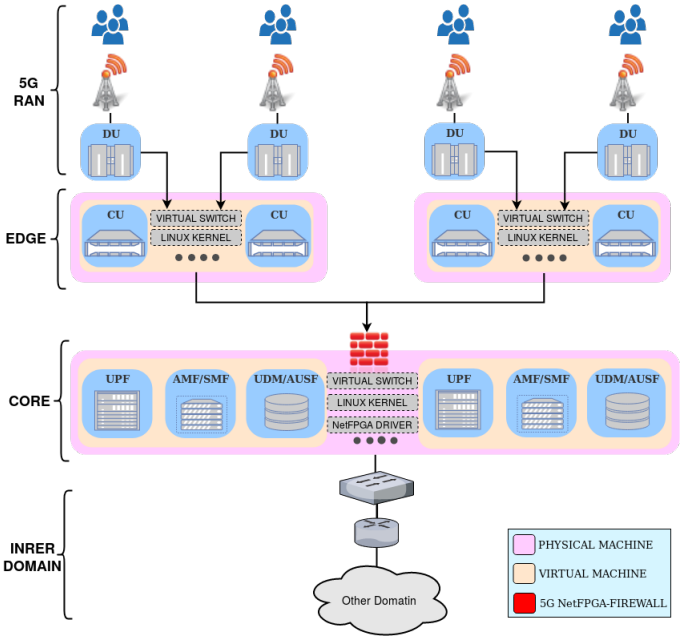

Fig. 1. 5G architecture overview

architecture and 5G components. A 5G Cloud-RAN (C-RAN) is deployed allowing communication between User Equipment (UE) and the Distributed Units (DU). Each DU is connected with the UE with a wireless system based on 5G Radio Interface [12]. A group of DUs can be interconnected with a pool of Centralised Units (CUs) [13]; those connections are traditionally deployed using optical fibers. This pool of CUs process the radio samples and provide the IP traffic into the network in the Edge-to-Core network segment. These CUs are deployed as Virtual Network Functions (VNFs) deployed in virtual machines. The 5G firewall developed in this paper can be allocated just after each CU, allowing the detection of traffic going to or coming from the Core and also can be allocated just before the User Plane Function (UPF) allowing the detection of traffic going to or coming from Edges, as can be seen in Figure 1. It is noticed that the Core concentrates the communications from multiple edges. It is exactly on this concentration where it will allow stopping traffic from users when they change their location to other CUs, i.e. in a handover process. Therefore, that is the best location for the solution to be deployed in order to protect all the internal infrastructure, which contains the different tenants, while providing mobility support for UEs of the network operator infrastructure.

The Edge is connected with optical fibers to the Core of the network. In the control plane of the 5G architecture, the Access and Mobility Management Function/Session Management Function (AMF/SMF) control the registration, connection and mobility management of users. The Unified Data Management/Authentication Server Function (UDM/AUSF) comprise databases of the system where all user information and subscription is stored. In the data plane, the UPF allows

the interconnection with the infrastructure and the maintenance of the session for the handover processes. Each tenant is allocated in a isolated virtual machine which contains an AMF/SMF, an UDM/AUSF and an UPF. The Core consist of different virtual machines that represent the multi-tenant support of the infrastructure. These elements are protected thanks to the 5G firewall developed.

B. Proposed P4-based 5G firewall

The main aim is to protect the internal 5G infrastructure of different communication operators, and thus the traffic investigated in this research needs to fulfill such requirements of the 5G infrastructure, mainly by providing support for the VXLAN and GTP encapsulation protocol require to handle user mobility across all the elements of the 5G infrastructure. The corresponding 5G packets have the structure shown in the Figure 2. The first part of the figure shows the headers used to establish a communication between physical machines, the second part is conform by a VXLAN encapsulation, which is use to identify the tenant where the packet is sent, and the third part is used to represent the traffic of a concrete user connected to a tenant of a mobile operator infrastructure.

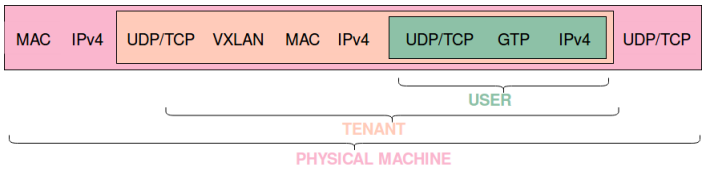

Fig. 2. Packet structure used in the Edge-Core 5G infrastructure.

Following subsections describe how the stages of the P4-NetFPGA pipeline have been implemented along the design of the proposed 5G firewall.

1) Parsing stage: Previous the parsing stage, following headers are defined: MAC, IP, UDP/TCP, VXLAN and GTP. Once the headers are created, it is needed to develop a data path able to parse packets with the structure defined in the Figure 2. To do so, firstly the MAC header of the packets is checked. In this study, Ethernet is adopted as the MAC layer and thus it is checked whether the Ethernet header has a hexadecimal value 0x0800 in the EtherType field, which indicates that the next header is an IP header used by the machines of the infrastructure to communicate. In the parsing of the IP header, the field protocol is checked to see what is the transport layer protocol. If the value of field protocol is 17 it is UDP; if it is 6, TCP is employed. After this, it should be checked if there is any further header following them. For this purpose, the destination port field is checked. In the solution proposed, it is important to see if the next header is VXLAN, which usually utilises port 4789 as is indicated by IANA. If the destination port matches the VXLAN port, the parser continues and then the second MAC header of the 5G tenant is parsed; in the same manner that happened with the first MAC header, the EtherType and the protocol field are checked to reveal the network and transport layer protocol, respectively. In the same way that happened with the VXLAN encapsulation, the destination port of the transport header is checked to see if the port match with 2152, which is the port assassinated by the IANA for the GTP header. After the GTP encapsulation the packet parsed have two headers more, IP and UDP/TCP which follow the same patron previously described. The encapsulated UDP or TCP protocol is the last header parsed, and it will have the payload. Traffic that matches a predefined traffic rule can be blocked before entering the infrastructure of the mobile operator as explained below.

2) Match/Action pipeline: Match/Action pipeline is situated after the parser stage. Whether a packet has been parsed correctly by the parser stage, it goes directly to the Match/Action pipeline, where the packet is analysed and the data extracted of the packet is used to see if the packet should be dropped by the NetFPGA.

The internal design of the TCAM table implemented in P4 is composed of several keys: 5G User source and destination IP, L4 Protocol Type, Physical machine source and destination Port, and the identification number of the GTP and VXLAN tunnel. These 7 keys have a total length of 160 bits and they are the minimum parameters that are needed to identify a flow of a specific user in a 5G multi-tenant scenario, where the vni of VXLAN header identifies the final tenant and the GTP id allows the identification of a final user. The ternary table gives the administrator of the firewall the possibility of blocking a group of different IPs. It is noted that conventional firewalls are mainly focused on the classification of traditional IP traffic and thus inspecting the outer IP and UDP/TCP headers. In contrast, the proposed solution goes further in the traffic classification and is able to inspect the inner IP and UDP/TCP headers, which are those related to the 5G users and tenants in the Edge to Core network segment.

IV. Experimental Setup and Results

A. Experimental scenario

To demonstrate empirically the performance of the P4-NetFPGA-based 5G firewall developed, a prototyped system has been carried out in the realistic testbed depicted in 3. A user-level Application Programing Interface (API) has been also deployed to allow the control of the 5G firewall by the network administrator. The API developed uses existing SDK libraries provided by the NetFPGA-SUME project [14].

In Figure 3, the network administrator should insert the rules when some flows should be blocked. The rules can be inserted at the same time that the NetFPGA is receiving the traffic, although for this experimental setup, the rules are previously inserted before the transmission starts (steps 1 and 2 in Figure 3). To test the proposed scenario and to execute the experiments, 10 different pcaps have been created. These pcaps contain real traffic of up to 512 different 5G users where each of the users is generating 2 different flows. This traffic has been gathered from the 5G infrastructure deployed in Figure 1. These pcaps are described in Table I. The idea is to analyze the behavior of the prototyped 5G firewall, when up 1024 flows are arriving simultaneously.

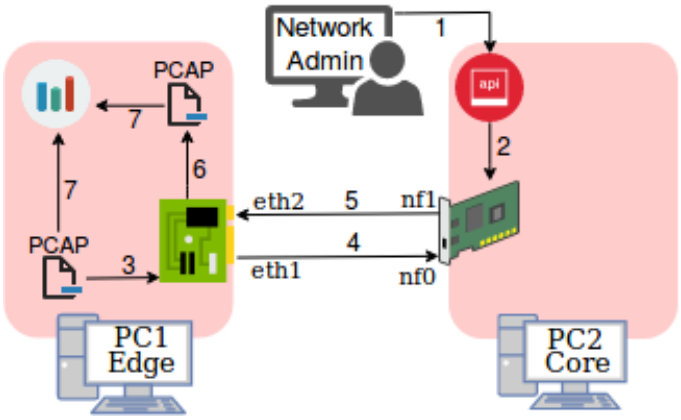

Fig. 3. Experimental testbed to evaluate empirically the behavior of the 5G Firewall developed.

TABLE I
5G NETWORK TRAFFIC GENERATED TO TEST THE PROPOSED 5G FIREWALL

Flows	Gb sent	Packets sent	Packets dropped	Duration(s)	Packets size
2 flow	0.03366	2805	1375	0.025	12000bits
3 flows	0.06732	5610	2750	0.045	12000bits
8 flows	0.13464	11220	5499	0.091	12000bits
16 flows	0.26928	22440	10997	0.184	12000bits
32 flows	0.53856	44880	21992	0.363	12000bits
64 flows	1.07712	89760	43983	0.728	12000bits
128 flows	2.15424	179520	87966	1.408	12000bits
256 flows	4.30848	359040	175931	1.834	12000bits
512 flows	8.61696	718080	351860	2.692	12000bits
1024 flows	17.23396	1436160	703719	4.699	12000bits

As can be seen in Figure 3, traffic transmission starts in PC1 (Edge) and it goes to PC2 (Core). The non-dropped traffic, by the 5G firewall, comes back to PC1. This return is only for experimental purposes, because it allows the measurement of the delay applied by the firmware developed for 5G multi-tenancy support in the NetFPGA card. Step 3 in Figure 3 shows as the different pcaps describe in the Table I are sent separately, from PC1, using tcpreplay [15]. The packets of each pcaps go through the eth1 interface to the nf0 interface in PC2 (step 4 in Figure 3). The packets are processed by the 5G Firewall and the non-dropped packets are forwarded through the nf1 interface to PC1 (step 5 in Figure 3). Once packets arrive to PC1 Edge, they are stored in a pcap file (step 6 in Figure 3), which is used later on to obtain the results (step 7 in Figure 3). In this scenario the purpose is drop only 1 of the 2 flow that each user is sending, whereas the other flow is sent to the sender to obtain the measures.

B. Results

The main aim of this experiments is to stress the scalability of the 5G firewall deployed in the testbed described in Figure 3. Empirical results are shown in Table II, Figure 4 and Figure 5. Notice that all data shown is this section has been obtained as an average of 5 executions.

Table II shows the internal delay achieved when 5G network traffic of Table I is processed by the prototyped 5G firewall. In Table II, it is observed that internal delay applied by the NetFPGA is almost insignificant and the increment suffered between the processing of 2 flows and 1024 flows is only 30 nanoseconds. So, it demonstrates that this 5G firewall has an excellent scalability because it is implemented in hardware and not via software. This data can be compared with the Intrusion Detection System (IDS) presented in [5], where the delay obtained for the classification and processing of 2 flows is nearly 5 milliseconds, increasing this value up to approximately 15 milliseconds when 512 flows are processed. It means that the tool proposed in our paper is 2493 times faster than the software tool provided by Serrano et al. [5].

TABLE II
ANALYSIS OF DELAY ACHIEVE BY THE INTERNAL PROCESSING OF THE 5G FIREWALL

Rules	Flows	Delay(ns)
1	2	5995
2	4	5995
4	8	6000
8	16	6000
16	32	6005
32	64	6005
64	128	6005
128	256	6010
256	512	6015
512	1204	6025

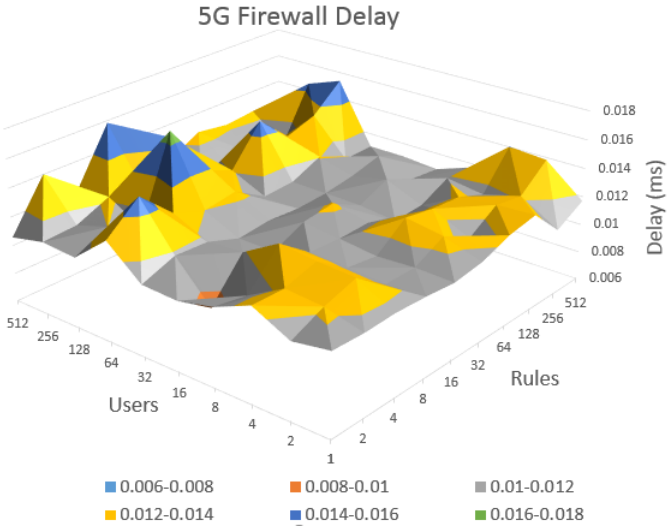

Fig. 4. Analysis of Delay achieved by the 5G Firewall prototyped

Figure 4 presents the total delay incurred in the prototyped 5G firewall system, when traffic ranges up to 1024 flows with 512 users being sending traffic at the same time. It is noted that the time implies not only the hardware delay but also the delay related to the software drivers and the captures of pcaps in the PCs into the Linux kernel to later on allow plotting the graphs. The results reveal that the 5G firewall developed does not practically incur any extra delay to the traffic processed when both the number of rules and traffic is increasing. There is a slightly increase in the delay when the NetFPGA has to process 1024 flows and it has 512 rules inserted, although it is just 3 microseconds of difference. These results demonstrate

the scalability of the prototyped 5G firewall and the efficient way to manage rules thanks to the API designed to insert or remove rules on demand.

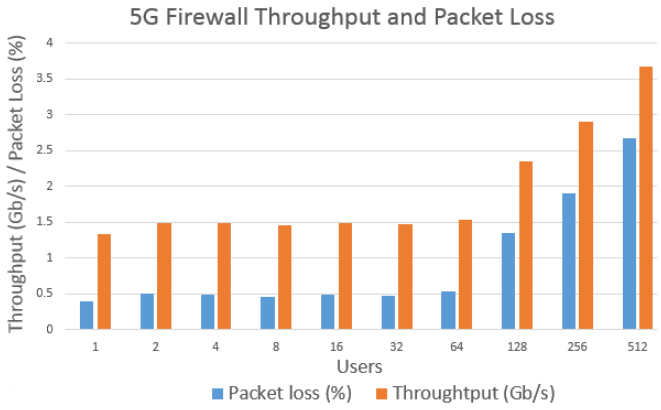

Fig. 5. Analysis of Throughput and Packet Loss achieved by the 5G Firewall prototyped

Figure 5 shows the throughput reached in the communication between PC1 and PC2. It is clear that although NetFPGA should let the exchange of traffic up to 10Gb/s, the maximum throughput reached has been 3,67 Gb/s. It is due to the quality of the network card used to send the network traffic and the lack of the NETFPGA NIC driver to deal with multiple cores. Above this speed, the network card starts unintentionally dropping packets in the transmission and thus the network performance can not be measured accurately. Figure 5 also shows the evolution of the throughput achieves with the rise of the number of users sending traffic at the same moment. Moreover, it can be observed that the percentage of packet loss is also growing following the same trend of the throughput yet always keeping an acceptable rate close to 2.5 % in the worst case scenario.

The results presented in terms of delay are really aligned with the Key Performance Indicators of the 5G-Public Private Partnership association that really want to achieve a end-to-end delay of 1 milliseconds for critical transmissions [16].

V. Conclusion

A novel firewall suitable for the Edge to Core 5G network segment in multi-tenant scenarios is proposed in this paper. The design has been prototyped by leveraging the platform of the P4-NetFPGA project. Nowadays, there is no other hardware-based solution, which allows to protect 5G multi-tenant infrastructures and mitigate cyber-attacks by selectively blocking malicious traffic. This solution is able to protect all tenants deployed in a data center, due to the fact that this 5G firewall is implemented in the border of the Core network, receiving all traffic from different Edges. Experimental results have validated the design and prototyping of the proposed 5G firewall system. Furthermore, the prototype yields good scalability even in stressed scenarios, and the end-to-end delay incurred is minimal.

Future work will investigate the performance with different NICs in the sender for higher throughput and also new FPGA solutions will be tested.

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] P. Ayuso, P. McHardy, J. Kadlecsik, E. Leblond, and F. Westphal, "The netfilter.org project," *http://www.netfilter.org*, 2014.

[2] G. YANG and S.-y. CHEN, "Research on linux firewall based on netfilter/iptables [j]," *Computer Engineering and Design*, vol. 17, p. 022, 2007.

[3] D. Cohen, "Netfilter/iptables u32," http://www.stearns.org/doc/iptables-u32.current.html, accessed: 2018-06-17.

[4] P. Salva-Garcia, J. M. Alcaraz-Calero, R. M. Alaez, E. Chirivella-Perez, J. Nightingale, and Q. Wang, "5g-uhd: Design, prototyping and empirical evaluation of adaptive ultra-high-definition video streaming based on scalable h. 265 in virtualised 5g networks," *Computer Communications*, vol. 118, pp. 171–184, 2018.

[5] A. S. Mamolar, Z. Pervez, J. M. A. Calero, and A. M. Khattak, "Towards the transversal detection of ddos network attacks in 5g multi-tenant overlay networks," *Computers & Security*, vol. 79, pp. 132–147, 2018.

[6] S. Lin, D. Zhang, Y. Fu, and S. Wang, "A design of the ethernet firewall based on fpga," in *Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 2017 10th International Congress on*. IEEE, 2017, pp. 1–5.

[7] A. Fiessler, C. Lorenz, S. Hager, B. Scheuermann, and A. W. Moore, "Hypafilter+: Enhanced hybrid packet filtering using hardware assisted classification and header space analysis," *IEEE/ACM Transactions on Networking*, vol. 25, no. 6, pp. 3655–3669, 2017.

[8] 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 *Microelectronic Systems Education, 2007. MSE'07. IEEE International Conference on*. IEEE, 2007, pp. 160–161.

[9] T. Soylu, O. Erdem, A. Carus, and E. S. Güner, "Simple cart based real-time traffic classification engine on fpgas," in *ReConFigurable Computing and FPGAs (ReConFig), 2017 International Conference on*. IEEE, 2017, pp. 1–8.

[10] R. Ricart-Sanchez, P. Malagon, P. Salva-Garcia, E. C. Perez, Q. Wang, and J. M. A. Calero, "Towards an fpga-accelerated programmable data path for edge-to-core communications in 5g networks," *Journal of Network and Computer Applications*, vol. 124, pp. 80–93, 2018.

[11] 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.

[12] P. Demestichas, A. Georgakopoulos, D. Karvounas, K. Tsagkaris, V. Stavroulaki, J. Lu, C. Xiong, and J. Yao, "5g on the horizon: Key challenges for the radio-access network," *IEEE Vehicular Technology Magazine*, vol. 8, no. 3, pp. 47–53, 2013.

[13] G. Brown, "Cloud ran and the next generation mobile network architecture," http://web.archive.org/web/20080207010024/http://www.808multimedia.com/winnt/kernel.htm, Apr 2017, accessed: 2018-06-17.

[14] 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.

[15] A. Turner, "Tcpreplay," *http://tcpreplay.synfin.net/trac/*, 2011.

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