Flexible Cross-Technology Offloading using SDN

Christos Zarafetias, Nikos Makris, Apostolos Apostolaras, Thanasis Korakis and Leandros Tassiulas

Abstract — Ultra-dense networks are expected to be the vehicle for better network coverage featuring higher speeds, augmented network capacity and more per cell end-users served. Towards meeting these stiff requirements, cells of varying coverage (pico-/femto-/macro- cells) are being deployed, especially in densely populated urban areas. Considering the high availability of such equipment, as well as the low deployment and maintenance costs of WLAN networks, an ecosystem rich in heterogeneity may assist in extending the overall network capacity, thus enabling throughput-hungry services to be offered over contemporary networks. The solution that is considered in this paper is offloading macro-cell users to the available wireless networks in a geographical area, while meeting their demands for the downlink channel. We model our problem and present an applied framework that facilitates cross-technology offloading. Our solution is making use of Software Defined Networking, thus enabling the very rapid and low latency network switching. We experimentally evaluate and benchmark our framework in a real testbed setup, and present our findings.

Keywords— Offloading, LTE, WiFi, Testbed, NITOS.

I. INTRODUCTION

Since the introduction of high bandwidth mobile networks (e.g. 3G/4G), we have been witnessing an unprecedented growth in the mobile network traffic. This constant growth is mandated by novel applications that are making use of network connections to deliver their services (e.g. UHD video, IoT data collection, etc.). Towards surpassing current technological limitations, mobile operators have been deploying cells of smaller size (e.g. pico-/femto-cells) in order to offload part of the exchanged traffic. Advances in WiFi networks, especially with the IEEE 802.11ac/ad amendments, render them excellent candidates for transferring part of the offloaded mobile traffic through high speed wireless links, with reduced energy consumption [1]. As a matter of fact, the strategy of mobile network offloading seems to be yielding with reduced energy consumption [1]. As a matter of fact, the strategy of mobile network offloading seems to be yielding with reduced energy consumption [1].

51% of the total mobile data traffic was offloaded onto the fixed network through WiFi or femtocells in 2015.

Offloading to operator cells is a rather straightforward procedure, with the signaling being entirely handled by the core network and thus the operator, and presenting the end user with a seamless experience in many cases. On the other hand, WiFi offloading is a more challenging process, especially when considering the existence of several open hotspots that are beyond the operator's control and management. According to [3], WiFi offloading is categorized in the following three main classes, depending on the operator control over the cell:

• Unmanaged data offloading: the mobile user's data is transparently moved to a WiFi network whenever there is one in range. This is the scheme that is currently used by many smartphone operator systems.
• Managed data offloading: an intelligent session aware gateway is placed between the user's WLAN connection and the Internet.
• Integrated data offloading: WLAN and mobile networks are bridged through the core network whereas signaling for both networks is managed by the operator (e.g. by using the 3GPP I-WLAN [4]).

Through the last two approaches, the mobile operators can still control their subscribers over the WLAN network as well, and even deliver subscribed content to them. Nevertheless, most of contemporary deployments make use of the first approach. Additionally, operators may consider a significant CAPEX and OPEX reduction when offloading clients, especially when considering the plethora of WLAN deployments. A revolutionary idea for mobile network operators is to, apart from deploying expensive proprietary WiFi networks, lease the idle network capacity of residential users in order to offload cellular traffic on the spot and on demand [5] [6]. This approach can be further extended by organizing the candidate WLAN cells in a mesh setup [7].

Nevertheless, employing a WLAN network for offloading mobile network traffic poses new challenges towards providing guaranties to the end-users on the Quality of Service and overall Quality of Experience. This is ought to the different channel access schemes that WiFi is employing (CSMA/CA), contrary to the licensed operation of mobile networks. In this work, we initially formulate the offloading problem, when considering an LTE network as the main mobile infrastructure, and a WiFi mesh network for offloading the traffic on. Inspired by the theoretical
formulation, we come up with a fully deployed scheme, which is making use of Software Defined Networking (SDN) in order to preserve the established connections in the downlink channel. Through continuous monitoring of the wireless channel utilization, we are able to deduce the network performance of the end-clients and dynamically re-allocate the clients to the mobile or the WLAN networks. The solution is verified and evaluated in a real testbed setup, namely the NITOS wireless testbed.

The rest of the paper is organized as follows: Section II is providing some background information on former research in the field. Section III is providing our network model of the offloading problem. Section IV is presenting our experimental architecture and our contributions to deliver a fully fledged solution for evaluating offloading policies. Finally, in Section V we showcase our experimental findings and in Section VI we conclude.

II. RELATED WORK

The utilization of WiFi networks for mobile offloading has been investigated in several works [8] - [9], in terms of the desired network capacity needed by the WiFi network to transfer efficiently the mobile network traffic. Mobile network offloading may be highly advantageous in terms of the perceived Quality of Experience, delivered network capacity and energy consumption. In [10], the energy consumption benefits from offloading mobile traffic to WiFi Access Points are presented. Nevertheless, cells that are used for offloading traffic to, need to be carefully planned in terms of deployment, especially when operator deployed cells are used, in order to maximize the offloading benefits. Authors in [11] investigate the potential deployment strategies for mobile operators towards enabling the offloading of mobile traffic.

Apart from the operator deployed cells, existing WiFi deployments may be used for traversing the operator offloaded mobile traffic, as they are in abundance and densely deployed in urban scenarios. These aspects have been discussed in [5]-[7]. Following these works, authors in [12] introduce the network economics theory in the offloading process and in a similar setup used for the WiFi mesh network. Through the modeling of the offloading process as a non-cooperative game, how much traffic should each AP offload for each BS the corresponding payment of each BS to network to another relies to policies enforced by the network operator. Authors in [15] introduce a mechanism which considers the real-time network conditions to derive the offloading policies and efficiently accommodate the traffic in both LTE and Wi-Fi networks. Similarly, authors in [16] propose a collaborative scheme among the network users to report the network performance and thus select the appropriate technology to offload their traffic to.

In this work, we model the LTE network channel with respect to maximizing the throughput allocation for the base stations, and determining the policy for offloading any throughput-hungry clients to a WiFi mesh setup that is meeting their requirements for the downlink channel.

The work is inspired from the [7], [17] and is porting the system model to a real testbed setup, that can reproduce experimental results. Through the application of an SDN based scheme, we are able to transparently switch technologies, inspired from our former work in [18]. We also provide monitoring of the traffic that is exchanged via the two setups (macrocell or WiFi mesh), thus enabling different pricing schemes to be applied, similar to [19].

III. SYSTEM MODEL

Our model stems from the theoretical models in [7], [17].

**eNB operation:** The downlink operation of one macrocellular base station (eNB) for a time period of $T$ subframes is considered. A set $N$ of $N$ users exists within that cell. Every user $n \in N$ requests a data content of $D_n \geq 0$ bytes to download during this period. At the same time, some of the mobile users may be positioned in range with one or more WiFi access points (APs), while some others may not be covered by any AP. The base station can initiate connections to each mobile users for servicing their traffic requests. Each traffic request is related to a bearer that has specific characteristics in terms of bandwidth and throughput. A set $M$ of $M$ bearers can be allocated to all users in each subframe $t = 1, 2, \ldots, T$. The value of $M$ depends on the available spectrum and the allocated resource blocks, and is related to the available capacity of the physical channel/bearer.

In the beginning of each time period, the eNB devises the bearer assignment policy for serving his users. We consider that the system is quasi-static. In such kind of systems, the channel and the number of users remain invariant during the current time period, and change over the periods. Let $x_{n, m}(t) \in \{0, 1\}$ denote whether a bearer $m \in M$ is allocated to user $n \in N$ during subframe $t$. Hence, the scheduling policy of the base station consists of the bearer assignment vector $x = (x_{n, m}(t))$: $n \in N$, $m \in M$, $t = 1, \ldots, T$, such that to satisfy each user requirement for downloading the traffic request of $D_n \geq 0$ bytes during this period $T$. Notice that this policy is invoked by the eNB so as to serve the associated to it users and hence satisfy their requests. At the same time, relying on this policy, the network operator determines which users cannot be sufficiently served by the eNB and hence should be offloaded.

In order to understand the servicing for each user $n$, we first need to analyze how the operator devises the resource allocation policy for serving the mobile users (or UEs) i.e.
which bearers should serve. The problem comprises the well-known knapsack problem where given a set of bearers each with different throughput requirements, the eNB determines which bearer should be initiated and hence served, so that the total throughput (aggregate bearer throughput) is maximized and is less than or equal to the throughput equivalent for the eNBs available bandwidth configuration.

In this context, the problem of the operator is to maximize the throughput allocation for the base station in order to service most of the users while ensuring the data delivery constraints for the users that it serves. This can be written as follows:

\[
\max_{x} \sum_{t \in [1,T]} \sum_{m \in M} \sum_{n \in N} g_{n,m}(t) x_{n,m}(t)
\]

\[
s.t. \sum_{t \in [1,T]} r_n(t) T_0 \geq D_n \forall n \in N.
\]

\[
x_{n,m}(t) \in (0,1) \forall n \in N, \forall m \in M
\]

where \( T_0 = 1 \) ms is the duration of the subframe, \( g_{n,m}(t) \) is the bearer achievable rate and \( r_n(t) \) is the actual measured rate by the user. We assume that this problem has a feasible solution [20] denoted \((x^*)\). The benefits from offloading the traffic of a user \( n \in N \), can be calculated by taking into account the non-servicing or poor servicing cost of the eNB as it wouldn’t be able to provide sufficient QoS guarantees to some users.

In this context, every mobile user is described by its service requests, and whether she is covered or not by a mesh AP. The eNB sorts the users in a decreasing order and selects the most data hungry that are eligible, i.e., within the coverage area of one AP. We emphasize here that this is a greedy method for determining the most resource consuming users as it leverages the results/policy that the eNB has to devise for serving its mobile users.

IV. EXPERIMENTAL SETUP

In this section we present the experimental setup, able to transform the analytic system model to real applications running over a testbed environment. We select to execute our experiment over the NITOS testbed, due to the wide heterogeneity of resources that it provides [21]. NITOS is a heterogeneous testbed located in the premises of University of Thessaly, in Greece. It offers a very rich experimentation environment with resources spanning from commercial LTE (eNodeBs, EPC and UEs), to open source WiFi and Software Defined Radio platforms. Detailed characteristics of the nodes and the LTE components that we utilize for the execution of the offloading experiment are given in Table I.

Although the testbed is providing us with the integral components of realizing the offloading experiment, we have to significantly extend some of the given functionality in order to support the key offloading functions. The components that need to be either extended or developed in order to support the offloading process based on the aforementioned model are summarized in the following subsections. The overall architecture of the framework is depicted in Figure 1, and we refer to it as the FLOW offloading framework hereafter.

A. WiFi Access Gateway (WAG)

The WiFi Access Gateway (WAG) is serving the role of the actual gateway of the WiFi mesh network that is used for offloading the LTE clients. Although the implementation of such a device would seem straightforward, in our framework we differentiate the traffic that is generated from the offloaded clients in order to meet some minimum requirements paved by a Service Level Agreement (SLA) that they have with the network provider, regarding the minimum achieved goodput on their downlink channel. To this aim, we employ the Hierarchical Token Buffers (HTB) for traffic shaping services. The functionality and recording of the data exchange volume is taking place inside the controller functions (see section IV-C) that we employ in order to realize the offloading scheme. Moreover, in order for the operator to control the offloaded users, we employ a dedicated tunnel connection from the WAG to the EPC, used as the network gateway. Hence, the overall offloaded scheme is a hybrid approach between a managed and integrated data offloading process. The operator still has control over the subscriptions, nevertheless the LTE protocol amendments defined for the interoperability between WiFi and LTE networks are not utilized, but simpler schemes are used.

B. PDN Gateway (PGW)

The LTE PDN Gateway (PGW) interface is in-charge of terminating the SGi interface towards the PDN. In the case of multiple PDNs, more than one network interface will be available per each PDN for the UEs of the network, depending on the Access Point Names (APNs) used. Beyond the PGW interface, and for addressing LTE network clients, the GPRS Tunneling Protocol (GTP) is employed. With our framework, we extend the functionality of the PGW in order to enable the operation of our offloading scheme. We employ an Open vSwitch (OvS) bridge [22] which enables the interconnection of two different entities, the tunnel end-point of the WiFi mesh network and the respective GTP interface for each PDN in the PGW of the LTE network. Routing, connection monitoring and data manipulation in order to appropriately relay traffic to the LTE network/Internet is done via a dedicated controller which we present in the following subsection.

<table>
<thead>
<tr>
<th>Network Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>LTE RAN</td>
<td>ip.access LTE245F</td>
</tr>
<tr>
<td>LTE mode</td>
<td>FDD Band 7</td>
</tr>
<tr>
<td>LTE frequency</td>
<td>2600 MHz (DL)</td>
</tr>
<tr>
<td>No RBs</td>
<td>50</td>
</tr>
<tr>
<td>LTE EPC</td>
<td>SiRRAN LTEnet</td>
</tr>
<tr>
<td>UE</td>
<td>Cat. 4 LTE USB Dongles, Huawei E3272</td>
</tr>
<tr>
<td>WiFi Chipset</td>
<td>Qualcomm Atheros AR9380</td>
</tr>
</tbody>
</table>
C. FLOW network controller

The FLOW offloading framework has been designed in order to coordinate the interplay between the extended WAG and PGW elements. By employing OvS, we bridge the heterogeneous RAN tunnel end-points and through a network controller service we are able to select the respective RAN from the provider’s perspective. The policies that we implement for offloading clients are based on the current load of each LTE cell and some predefined SLAs that each client has contracted with the provider. Prioritization and QoS for the LTE network is provided by the LTE specification, through the utilization of specific data bearers per each UE. Nevertheless, and in the context of predefined SLAs per each client, when a user is offloaded to the WiFi mesh network, similar prioritization has to take place. Hence, the FLOW network controller utilizes different system queues for scheduling the transmissions of each user, with a predefined priority. The controller is in charge of allocating the traffic of each offloaded client to the specific queue meeting the predefined SLA. Based on the QCI parameters per UE in LTE, we schedule the transmissions of the respective data to the WAG and then the rest of the WiFi mesh network (for the downlink channel).

D. FLOW network controller

The PCC unit is in charge of applying the proper control of the policies and charging of the clients per subscriber basis, and based on the QoS class that they belong. By extracting low level statistics from the network bridging function, we are able to monitor in real time the data volume that each UE has exchanged over the LTE/WiFi network. Moreover, by actively monitoring the data that sent over the WAG, the mobile operator may be aware which WiFi nodes are used and the amount of traffic that is traversed through them to the end users.

V. EXPERIMENTAL RESULTS

This section details our experiments and findings. The overall FLOW framework manages the offloading procedure based on an existing SLA that each UE has contracted with the network operator. These SLAs mirror the minimum capacity for the DL channel that has to be served for each UE. For a proof-of-concept setup with the FLOW framework, we use the SLAs denoted in table II for each client of the NITOS testbed that we use for our experiments. We employ eight nodes in total, configured as follows:

XXX
1) 3 nodes are serving as the WiFi mesh setup in a 3-hop manner. The mesh network is configured to operate in the 5GHz band of IEEE 802.11n, in an entirely interference-free environment. The nodes are configured in a manner that they cannot communicate with each other, using Layer 2 firewalls for blocking the traffic. One of them is set to operate as the only access node in the area, whereas a second node is a traffic relay and a third one is hosting our WAG software. The measured reference end-to-end throughput for this setup is approx. 60Mbps.

2) 5 nodes are used as the potential network UEs. The UEs feature two interfaces, an IEEE 802.11n compatible card and an LTE USB dongle. On the nodes, we install our framework that is responsible for parsing the messages sent by the interfaces and configuring the correct routing rules, based on the selected network.

3) The LTE network, consisting of the eNodeB and the EPC. On the EPC we create a tunnel termination point for the WAG and install the core of the FLOW framework. In order to validate our scheme, we initially connect all the clients on the LTE network. The network controller is checking the exchanged traffic over the network, and in case that the served capacity is not sufficient for the attached UEs attached, the controller selects to offload the most data-hungry client to the WiFi mesh setup. We set the total served capacity from the FLOW controller to be 60Mbps; this means that if a client's SLA generates traffic that will exceed the aggregate traffic to over 60Mbps, the offloading framework is triggered and one of the clients will be transferred to the WiFi network. Since this is the measured capacity of the WiFi 3-hop setup as well, subsequently the FLOW framework will be able to double the served network capacity used for our experiments. We use the iPerf application to generate traffic on the DL channel using UDP datagrams, and report the collected measurements to an OML server [23].

<table>
<thead>
<tr>
<th>NITOS Nodes</th>
<th>SLA to be met (DL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>node065</td>
<td>25 Mbps</td>
</tr>
<tr>
<td>node069</td>
<td>10 Mbps</td>
</tr>
<tr>
<td>node079</td>
<td>5 Mbps</td>
</tr>
<tr>
<td>node085</td>
<td>20 Mbps</td>
</tr>
<tr>
<td>node091</td>
<td>15 Mbps</td>
</tr>
</tbody>
</table>

Figure 2 depicts our experimental results. Based on the SLAs that we use to evaluate the framework, we expect that the first three clients are served correctly from the LTE network, as they require an aggregate traffic of 40Mbps. However, when the fourth client is introduced to the network and starts downloading traffic equal to 20Mbps (at around 40secs), the most data-hungry node to that moment (node065)
is offloaded to the WiFi network. The offloading process is deduced by the pikes reported in the throughput graphic (Figure 2a), as well as the total number of datagrams that are sent (Figure 2c); as the traffic generator tries to push more traffic over the network, instantly no traffic is sent and subsequently almost the double amount of datagrams is pushed to the network. Similar patterns can be seen for the jitter as well, with small pikes being present at the time when the offloading process takes place. As we start to remove clients from the network and as soon as the capacity of the network is again sufficient for serving the UEs, node065 is again reallocated to the LTE network (at around 115secs). As we see, the jitter values for the times that the client is switching technologies are not affected, thus proving that our framework is capable of switching technologies in a seamless manner and with almost no impact on the user-perceived Quality of Experience.

VI. CONCLUSION

In this paper, we presented an SDN based framework for facilitating the offloading process from LTE to WiFi mesh networks. We presented a model for our system, and our approach on the offloading policy that we adopt. We detailed the needed extensions in terms of framework design and implementation, and showcased a proof-of-concept experiment that is able to meet the SLAs of the cell users by using a 3-hop 802.11n mesh setup, that is adding up to the network capacity of around 100%. The augmented network capacity is depending on the mesh network used, as a 2-hop IEEE 802.11g mesh would only add about 30% to the total network capacity, whereas a 2-hop IEEE 802.11n mesh would add over 150%.

In the future we foresee the further development of the framework to manage interference results and network throughput estimates for the WiFi network part, that is prone to external interference and may result to a degraded quality of experience to the end users. The proposed framework creates fertile ground on providing a fully-fledged solution for network management from a central authority, allowing novel approaches on the offloading policy that we adopt. We detailed the implementation, and showcased a proof-of-concept.