

# Programmable Policies for Data Offloading in LTE Network

Mojdeh Amani<sup>†</sup>, Toktam Mahmoodi<sup>†</sup>, Mallikarjun Tatipamula<sup>‡</sup>, Hamid Aghvami<sup>†</sup>

<sup>†</sup> King's College London, London, UK

<sup>‡</sup> F5 Networks, San Jose, CA, USA

(mojdeh.amani, toktam.mahmoodi, hamid.aghvami)@kcl.ac.uk, m.tatipamula@f5.com

**Abstract**—Mobile data offloading on smaller cells such as Wi-Fi comes as a natural solution to boost cellular networks capacity and keep up with the rapid increase of mobile data traffic demand. In this paper, we propose an offloading mechanism through applying the abstraction of Software-defined Networking (SDN) in the mobile backhaul to provide programmable offloading policy derivation that are aware of users and applications as well as condition of wireless network. The proposed mechanism considers the real-time network condition to derive the offloading policies and efficiently accommodate the traffic in both LTE and Wi-Fi networks. Numerical results prove that the proposed approach can significantly improve dropping rate of the incoming traffic with using more real-time and dynamic decisions for offloading.

**Index Terms**—mobile data offloading; LTE-Wi-Fi interworking; policy control; Software-defined Networks;

## I. INTRODUCTION

In 2013, mobile phones will overtake PCs as the most common Internet access device worldwide, and by 2015, over 80% of the handsets sold in mature markets will be smartphones [1]. According to forecasts, global mobile data traffic will grow 13-fold from 2012 to 2017; in particular, global mobile data traffic will grow 3 times faster than fixed IP traffic [2]. While mobile network operators will carry the bulk of Internet traffic in the future, they face significant challenges in addressing the needs of increased traffic demand. To address this demand, mobile operators invest in more network capacity. Scarcity of spectrum forces mobile operators to deploy smaller cells, and to utilize unlicensed spectrum such as Wi-Fi. Availability of built-in Wi-Fi on smartphones along with characteristic such as unlicensed spectrum makes Wi-Fi a natural solution for accommodating the increased traffic and maintaining the quality of users' connections.

To this end, mobile data offloading, which refers to the use of complementary network technologies for delivery of data originally targeted for mobile/cellular networks, has already become a key solution to address the traffic demand [3], [4], [5]. Various platforms including opportunistic communications are considered for mobile data offloading [6]. Furthermore, mobile networks must support various applications such as voice, streaming, as well as best-effort services on a single IP-based infrastructure. Each of these converged services has Quality of Service (QoS) requirements such as latency, packet loss and data rates, which must be obtained through efficient allocation of the wireless network resources and cannot be

addressed only by provisioning the network. Therefore, one of the main challenges in data offloading solution is making real-time decisions for offloading different users' flows and services/applications while taking condition of the available networks as well as the QoS needs of flows into account. In this regard, previous research works also focused on providing seamless move between different access networks and during the data offloading [7]. Using the dual stack mobile IP is also considered to enable the simultaneous use of two interfaces [8]. Although all technical aspect of the mobile data offloading from the LTE network is ready for the deployment (and in fact deployed by the operators), we lack efficient offloading techniques, and decisions criteria.

The mobile networks architecture as proposed in the 4G/LTE provides an easier management comparing to earlier technologies, by separating the signalling plane functionalities such as mobility management, policies and charging. Despite this, today's 4G/LTE architecture is not yet as flexible or programmable as can be. For example, the 4G/LTE network is capable of offering QoS guarantees and differentiated services to the user, using Policy Charging and Rules Function (PCRF) nodes, while the Policy and Charging Control (PCC) ensures users' QoS for a particular subscription and service type. Moreover, today's PCCs are aware of users and applications, and not aware of the congestion in the network.

To this end, we argue that placing the policy control closer to the wireless access and deriving the policies based on the radio network information in addition to the previously used user and application information can significantly improve the performance of mobile data offloading. Based on the concept of Software-defined Networking (SDN) [9], we propose an abstraction layer that integrates the Network Resource Management (NRM) and Radio Resource Management (RRM), and enables programmable and dynamic policy functions.

The remainder of this paper is structured as follows. Section II describes the LTE/Wi-Fi interworking architecture. How programmable policy functions can modify the interworking to enable better QoS provisioning is elaborated in Section III. Details of our policy functions and parameters that potentially affect offloading decisions are explained in Section IV. After describing the simulation scenarios in which our proposed offloading mechanism are examined, results are presented in Section V. Finally, we conclude in Section VI with a summary of the programmable data offloading policies.

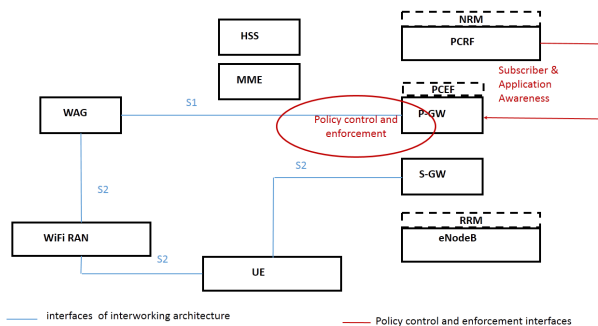


Fig. 1. 3GPP architecture for the non-3GPP IP access integration into the EPC

## II. LTE AND WI-FI INTERWORKING ARCHITECTURE

In the 4G/LTE mobile backhaul architecture, Mobility Management Entity (MME), Policy Charging and Rules Function (PCRF) and Home Subscriber Server (HSS) are participating in only signalling plane functionalities while Base Station (eNodeB), Serving Gateways (S-GW) and Packet data network Gateway (P-GW) involve in dataplane functionalities.

The 3GPP standard describes native integration of trusted and untrusted non-3GPP IP access networks into the EPC [10]. This standard considers Wi-Fi Radio Access Network (RAN) as valid an access network as any other 3GPP radio access network, and enables operators to use the standard-based EPC components for integration of different types of access networks, while also facilitating inter operability between them [8].

In this paper, we focus on trusted non-3GPP access network architecture i.e. Wi-Fi access network which is owned by the cellular operator. As it can be seen in Figure 1,  $S1$  and  $S2$  are the two interfaces that provide control and mobility support between non-3GPP access and P-GW, and forward the Wi-Fi traffic to the EPC. The  $S2$  interface provides mobility and control support between UE and P-GW over the non-3GPP access, while the  $S1$  interface, provides control and mobility support between trusted non-3GPP access and P-GW. The S-GW serves between LTE and Mobile access gateway (MAG) of Wi-Fi, and reports to the PCRF while also set the QoS parameters of bearers.

## III. SYSTEM MODEL

### A. Policy Control

In the architecture of Figure 1, the Policy and Charging Enforcement Function (PCEF) in P-GW, enforces the policies and map service data flows to bearer to be mapped to the underlying transport network. The PCRF is LTE policy manager, which takes the operator policies, network information and user's profile stored in the HSS to make decisions based on the set of pre-defined rules and functions; QoS authorization, i.e. how to treat each traffic flow, is also performed by the PCRF. Today's policy controls are aware of users and applications, and not aware of the congestion in the network [11]. While

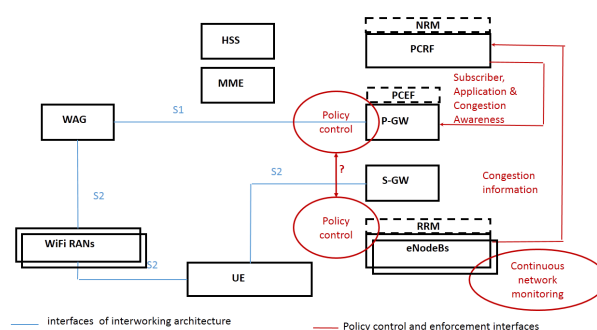


Fig. 2. Coupling of RRM and NRM to optimize utilization of both radio and network resources

having fine-grain control on various entities in the mobile network is crucial for operators to allocate their resources, and to maintain and expand the network with low cost, it is also important to employ dynamic policies, e.g. to manage the traffic and selectively offload the traffic between different access networks depending on the current network condition.

When performing traffic management, policy enforcement point can either be located at the core network or at the access network. At the core, enforcement is in the P-GW, and where the deep packet inspection (DPI) functions - similar to today's architecture in Figure 1. In this case, the service information does not have to travel at all, and the DPI engine can store subscriber information from the policy controller. Hence, the congestion and location information can be estimated. The central solution as depicted here, requires simple integration of the DPI engine with the policy controller. On the other hand, the congestion information is extremely dynamic, i.e. it changes rapidly and by the time values sent from RAN are received by the P-GW, the information is not valid any more. Various studies show that 40% of the time bad decisions were made when outdated congestion information was used.

Moreover, estimation based on the current traffic is not simply possible for the DPI engine for various reasons. (1) The policy controller lacks a feedback mechanism. Simple questions such as whether the 1 Mbps for P2P is enough or we are over penalizing, can not be answered. (2) Throughput reduction is the only indication for the the policy controller, which can be either as a result of poor coverage or congestion. Only the RAN can differentiate between these two causes of data rate drops. (3) Quite often, the RAN is shared among different operators or perhaps the operator is using multiple access points. In these cases, policy controller can not see all traffic going through the congested cells. (4) Cell capacity depends on the coverage of the individual subscribers and varies even with weather conditions. There may be an additional reserved bandwidth for future bearers with guaranteed data rate, and other similar cases that affect the total cell capacity. In other words, the cell capacity varies with time and the policy controller receives no information about the capacity variations of the congested cells.

One alternative solution that can address the above chal-

lenges, is moving the policy enforcement point to the place where congestion occurs, so that we do not need to transfer this dynamic information anywhere. In this case, the scheduler is continuously prioritizing data packets and subscriber sessions. The scheduler has perfect knowledge about the location of the user, traffic and the real congestion conditions at that location. Then, policy controller will take service information from the DPI function and it will change the QoS parameters of the subscriber bearer, such as the traffic handling priority, the maximum bit rate, the guaranteed bit rate or the QoS Class Identifier (QCI). In the LTE, there is also the possibility to create a dedicated bearer for a specific traffic flow that requires a differentiated QoS treatment at the policy enforcement point. This architecture, plotted in Figure 2, brings additional complexity to the network management. Thus, it is still not clear where exactly the policy control should be located, to benefit from additional efficiency but not dramatically being penalized by additional complexity.

### B. SDN-Controller and Mobile Data Offloading

As discussed earlier, deriving real-time offloading policies for selectively offloading different services/applications based on the dynamics of network (Figure 2) could potentially be complex. On the other hand, a programmable interface similar to what SDN offers, facilitates offloading mechanism by providing an end-to-end communication between network elements and by pushing corresponding forwarding rules to the local elements (i.e., eNodeB, and P-GW). In our considered architecture, the control-plane functionality of the gateways are decoupled and logically located as applications at the *SDN-controller*, while gateways run *local control agents*. The SDN-controller derives the offloading policy functions and rules by combining information from the RRM and PCRF applications. Radio network condition, defined by parameters such as wireless condition and traffic load, are measured frequently by the local control agents. This will allow operators to perform real-time traffic monitoring and provide per-subscriber QoS by programmable application modules in the SDN-controller and deriving forwarding rules accordingly. These policies and forwarding rules are periodically sent to the local control agents (in the access network) to be forwarded to the UE. The LTE and Wi-Fi interworking architecture including SDN-controller is depicted in Figure 3, where interactions with the local control agents are also shown. The two main parts of this architecture are as follows:

1) *SDN-Controller*: The SDN-controller in this architecture is an abstraction model that runs programmable applications modules such as RRM, and the PCRF. The PCRF application module is in possession of subscriber and application information, while RRM application collects radio access network condition such as traffic load and cell capacity. The SDN-controller combines information of these two application modules to derive a single set of policies and rules. These rules are sent periodically to the local control agents.

2) *Local control agents*: To address the challenges raised in Section III-A as well as the scalability issue, we consider

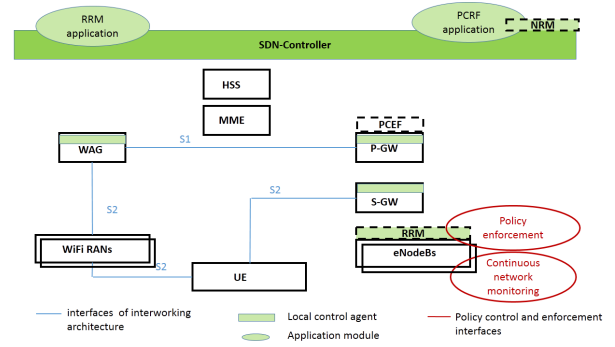


Fig. 3. Coupling RRM and NRM via SDN-controller in the non-3GPP IP access integration into the EPC

local control agents in the network gateways (P-GW, S-GW and WAG), and the RAN. These local control agents should perform measurements and some controlling actions that are authorized by the SDN-controller. For example, the agents that run on the gateways in this architecture, can measure QoS parameters such as delay and resource utilization and compare the traffic counters against the threshold and notify the SDN-controller in case of exceeding the threshold. To communicate back with the controller, an interface similar to OpenFlow [12] is required at the local agents, which also allow them to perform simple control actions such as changing the weight or priority of a queue when the traffic counter exceeds a threshold.

## IV. POLICY DERIVATION AND OFFLOADING MECHANISM

### A. Policy Derivation

We detail two offloading methods here, offline and online, where the offline method refers to the current architecture of fixed policy functions at the PCRF (Figure 1), and the online method refers to the policy decisions by the SDN-controller that are enforced via local control agents (Figure 3).

1) *offline Method*: In offline method, the PCRF identifies policies based on pre-defined set of functions and rules and depending on the list of subscribers and applications. These functions, set the *OffCap*, and the offered QoS by each network as fixed values and derive offloading policies accordingly. Threshold value of *OffCap* is the value at which offloading from LTE to Wi-Fi occurs, which is expressed in terms of percentage of the total capacity of LTE network ( $C_{lte}$ ).

Given  $\Lambda$  is the LTE network utilization,  $Q^0(wifi)$  is the QoS that can be offered by Wi-Fi network, then whether a flow is served by the LTE or Wi-Fi is decided by,

$$\begin{cases} 1 - (\Lambda < OffCap) \text{ or} \\ (OffCap \leq \Lambda < C_{lte} \text{ and } Q_i < Q^0(min)) \\ i \text{ is served by the LTE} \\ 2 - \Lambda \geq OffCap \text{ and } Q_i \geq Q^0(wifi) \\ i \text{ is offloaded to Wi-Fi} \\ 3 - Otherwise \quad i \text{ is dropped} \end{cases} \quad (1)$$

where  $Q_i$  represents a QoS parameter that has an upper bound, e.g. maximum delay that a flow can tolerate. The

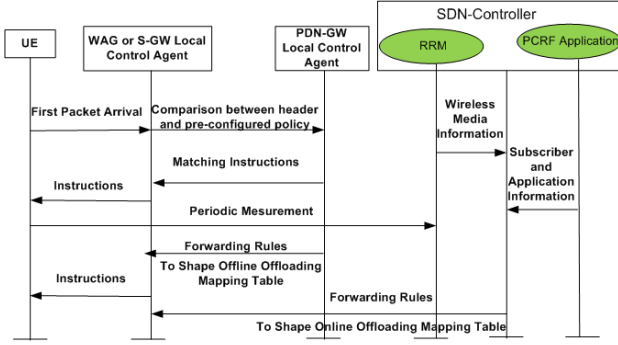


Fig. 4. SDN enabled offloading procedure

second part of line 1 in Equation (1), refers to the very delay sensitive applications (their delay requirement is less than a pre-defined minimum  $Q^0(min)$ ), such as voice, which will be served by the LTE even after utilization is beyond  $OffCap$  (and as long as the LTE network is not over congested).

2) *Online Method Scenario*: In online method, network parameters are measured frequently by local control agents in the access network and are available at the RRM application. The SDN-controller combines this information with the PCRF data and enforces policies by pushing the forwarding rules to the gateways. Combining these two application modules allow the offloading decision to be made based on the QoS requirement of that particular flow (user/application) as well as the offered QoS by each access network at any time.

As mentioned earlier, the offered QoS, e.g. the offered latency, by each network differs depending on the network condition and is noted by  $Q^c(lte)$  for the LTE and  $Q^c(wifi)$  for the Wi-Fi network. In this case, policies are similar to the offline method with the only different being the dynamic  $Q^c$  instead of the fixed values - the online policies are explained in Equation 2.

$$\left\{ \begin{array}{l} 1 - (\Lambda < OffCap) \text{ or} \\ (OffCap \leq \Lambda < C_{lte} \text{ and } Q_i < Q^0(min)) \\ i \text{ is served by the LTE} \\ 2 - \Lambda \geq OffCap \text{ and } Q_i \geq Q^c(wifi) \\ i \text{ is served by the Wi-Fi} \\ 3 - Otherwise \quad i \text{ is dropped} \end{array} \right. \quad (2)$$

where  $Q^c$  is dynamically changing, depending on the current measurement from the network.

### B. Offloading Mechanism

The policy derivation is performed in the SDN-controller, and the interworking between different modules is shown in Figure 4.

Initially UE transmits to the either or both of the Wi-Fi and LTE access network gateways. Upon arrival of the first packet of the flow, local control agents at the P-GW compares the header information with pre- configured policies and sends the match offloading instruction to the access network gateway and the UE. After connection is established with the chosen access network, the service data flows are mapped to the

bearers that match this policy. Then P-GW maps the bearers to the underlying transport layer.

The local control agents in the radio access network and the access gateways periodically collect the information such as utilization, and drop rate, from Wi-Fi and LTE access networks and report these measurements to the SDN-controller. The SDN-controller then combines these with the PCRF information, derives the policy functions and forwarding rules, and pushes them back to the local control agents. Hence, the local control agent will shape an offloading mapping table based on these rules. The updating frequency of these rules and functions can be set according to the frequency of change in the network conditions as well as the availability of communication resources between the controller and the access network.

Hence, the proposed architecture here, offers (1) **programmable** offloading policies by converting NRM and RRM to software modules; (2) **simplicity** by abstraction of the network view and simplifying the communication between network elements, (3) enhanced **efficiency** by allowing more up to date and precise congestion information to reflect on the offloading policies. In the next section, we show how such an architecture can enhance the network performance in quantifiable metrics.

## V. PERFORMANCE INVESTIGATIONS

In this paper we mainly focus on the programmable offloading policies, that is managed by a logically centralized SDN-controller (Figure 3), while details of mobility management and authentication are out of the scope of this work.

### A. Simulation Model

The simulation model considers a single cell served by an omni-directional antenna, and OFDMA downlink (general 4G/LTE model) with the total bandwidth of 20 MHz and the downlink data rate of 100 Mbit/s. The cell also covered by a number of randomly distributed Wi-Fi access networks, which their coverage area does not overlap each other, i.e. each mobile user can only get service via a single Wi-Fi access point at any location.

For the channel model, a propagation loss model has been considered similar to [13], where free space path loss and shadowing affect the received power at the UEs. The free path loss between eNodeB and UE is determined using standard radio propagation models, considering the loss ( $P_L$ ) as a function of the distance between eNodeB and UE in kilometer (denoted by  $d$ ), and defined as Equation 3.

$$P_L = 128.1 + 37.6 \times \log d. \quad (3)$$

Shadowing is modeled as a log-normal random variable with zero mean and standard deviation of 8 dB. We later compute users' received Signal-to-Interference plus Noise Ratio (SINR) based on the  $P_L$  and the log-normal shadowing to model the effect of wireless channel on throughput. We further assume, there are enough Wi-Fi resources to maintain the current Wi-Fi data rate; hence data rate that Wi-Fi access point can offer

TABLE I  
PERFORMANCE REQUIREMENT BY SERVICE CATEGORY

Service ID	Service Type	Rate (kbps)	Delay (ms)
1	VoIP	21-64	50-100
2	Video	500-700	200-300
3	Interactive gaming	300-600	100-300
4	web browsing	50-600	300-600
5	Peer-to-Peer	700-1000	300-600
6	Business services	600-800	50-100

is independent of its traffic load in the network. As Wi-Fi and LTE networks operate on different frequency bands, there is no resource partitioning or interference between these access networks. We do not consider interference between adjacent Wi-Fi access points, by relying on the Wi-Fi planning strategy, load balancing, and power control across overlapping access networks.

In addition, low mobility users are assumed. Mobile users are distributed uniformly over the cell with radius 1 km. To uniformly distribute the mobile users, a cartesian coordinate system is considered where  $x$  and  $y$  coordinates are chosen uniformly. In each round of simulation, users move from their location between 10 m to 70 m.

On the other note, all services in the LTE are provided as packet services including voice services. Hence, real-time and non real-time services are multiplexed over the air interface and core network. For modeling traffic, we define six categories of services as Voice over IP (VoIP), video, interactive gaming, web browsing, business (high priority) services and peer-to-peer. Each service category  $k \in \{1, \dots, 6\}$  is characterized with uniformly distributed random delay and data rate requirements. The range of delay and bandwidth of these service types are detailed in Table I. The proposed approach here is independent from the model used for describing incoming data, i.e. stochastic flow modeling is not required. We define a flow as a bit stream generated by the application layer. Packets waiting for transmission are stored in a queue associated to the user buffer and for each specific type of traffic. We assume infinite size queues so as to ignore the buffer overflow effects.

We denote by  $F$  set of  $n$  active flows, i.e.  $f_i \in F$ , where  $f_i$  is identified by  $d_i$ , delay requirement, and  $b_i$ , data rate requirement, of the related service category. In consecutive rounds of simulation,  $n$  is increased from 180 to 270. The background traffic load of cellular network is uniformly distributed random variable in the range of [10 Mbps, 50 Mbps], and LTE network can handle up to 100 Mbps data transmission. Users' throughput is computed depending on the current traffic load as well as the users' channel condition, and the service type of the incoming flow  $i$ . For example, flow  $i_1$  that is of service type  $k_1$ , will receive  $\alpha \times \beta \times b_{k_1}$  as their throughput, where  $\alpha$  shows how busy the network is and  $\beta$  determines the effect of user's mobility and its channel condition on throughput. The  $\alpha$  is inversely proportional to the network utilization, i.e.  $\alpha = 90\%$  with 10 Mbps background traffic and it will be decreased to 50% by increasing the background load to

50 Mbps. Using the SINR for each user and at any specific location,  $\beta$  is extracted from the curve showing BLocking Error Rate (BLER) versus the SINR of [14]-figure ten.

It is assumed that Wi-Fi network carries a traffic load itself before we offload any traffic. That load is a uniformly distributed random value between 1 Mbps and 40 Mbps for each Wi-Fi access point. The Wi-Fi is considered to be 802.11n with the maximum data rate of 60 Mbps.

## B. Numerical Results

The examined metrics here include dropping rate and offloaded traffic rate. These two figures of merit are defined as follows:

- (1) **Dropping rate** is the ratio of dropped traffic to the overloaded traffic
- (2) **Offloaded traffic rate** is the ratio of offloaded traffic to the overload traffic on the LTE.

Simulations run separately for the online and offline methods, and based on their related constraints. In both online and offline method *OffCap* is set to 60% of i.e. the threshold at which offloading from LTE to Wi-Fi is triggered. In the offline method, offered delay by Wi-Fi network is set to the fixed values of 200 ms ( $Q^0(min)$ ), independent of service category of the flow, and LTE network admits only delay sensitive flows, i.e. those flows with delay requirement less than 100 ms ( $Q^0(min)$ ). When offloading triggers, a flow is offloaded from LTE to Wi-Fi only if Wi-Fi satisfies its delay, and data rate requirements; otherwise that flow is dropped.

In online method, offloading decisions are made depending on the current condition of access network, in addition to the delay and bandwidth requirements of the flows. To consider the current condition of the access network, we assume that delay experienced by the flows increases as a result of utilization. We assume when Wi-Fi utilization increases from 30% to 70%, delay increases linearly between 100 to 300 ms. Furthermore delay experienced in LTE access network increases to 200 ms when the utilization goes beyond 70% When offloading triggers, if Wi-Fi network can accommodate delay and data rate requirement of flow  $i$ , then this flow is offloaded, otherwise if its data rate requirement can be satisfied by LTE, it will be served by LTE. Finally, a flow will be dropped by the LTE, if its delay requirement can not be met or by adding it, load goes beyond the network capacity,  $C_{lte}$ . The results in Figure 5 show the offloaded traffic rate and the dropping rate for online and offline methods. The online method shows an average 15% to 35% higher offloaded traffic rate and 10%-20% lower dropping rate comparing to the offline method. In offline method, as a consequence of fixed/pre-defined delay in Wi-Fi, less flows are offloaded to the Wi-Fi, which results in higher dropping rate. The online method, considering the dynamic of the Wi-Fi and LTE network, more flows are offloaded on Wi-Fi network, or more flows are served by the LTE, hence less flows are dropped. It should be noted that, in each round of simulation on average 40% of the traffic is dropped in offline method due to the fixed offered delay by both networks.



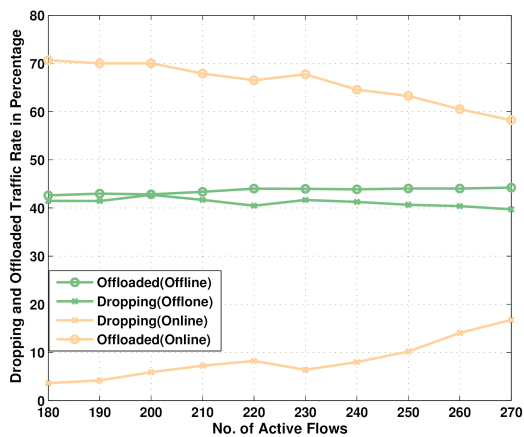


Fig. 5. Offloaded traffic rate, and dropping rate in percentage Vs. number of active flows for both online and offline policy derivations.

Clearly, by increasing the traffic load, drop rate is increased online offloading, and the gap between the offline and online dropping rate is decreasing. Figure 6 shows offloaded and dropped flows for both online and offline methods. In the online offloading, on average 20 more flows are offloaded to the Wi-Fi network and 50 less flows are dropped.

## VI. CONCLUSION

Mobile data offloading has already become a key solution to address the explosive data traffic demand over mobile networks. In this paper, a programmable policy function derivation framework through applying an abstract of SDN in the mobile backhaul is proposed. This framework consider the real-time network condition measurement in addition to the existing user and application information to control offloading policies so as to efficiently accommodate the traffic on the LTE or Wi-Fi accesses. We converting the RRM and NRM into application modules and couple them for deriving offloading policies to optimize both network and wireless resources. Our simulation results show enhanced offloading performance, i.e. enumerated with 35% increase in offloaded traffic from the LTE to Wi-Fi and 15% decrease in dropping rate. In addition to the significant performance improvement, the proposed policy control here offers further programmability for the newly added features, simplicity of implementation, and efficiency in the communication between network elements.

## REFERENCES

- [1] "Top 10 Strategic Technology Trends," Available online: [www.gartner.com/newsroom/id/2209615](http://www.gartner.com/newsroom/id/2209615), 2013.
- [2] "VNI Global IP Traffic Forecast, 2012-2017," Cisco White paper, May 2013.
- [3] A. Aijaz, H. Aghvami, and M. Amani, "A Survey on Mobile Data Offloading: Technical and Business Perspectives," *IEEE Wireless Communications Magazine*, April 2013.
- [4] K. Samdanis, T. Taleb, and S. Schmid, "Traffic Offload Enhancements for eUTRAN," *IEEE Communications Surveys and Tutorials*, vol. 14, no. 3, pp. 884–896, 2012.

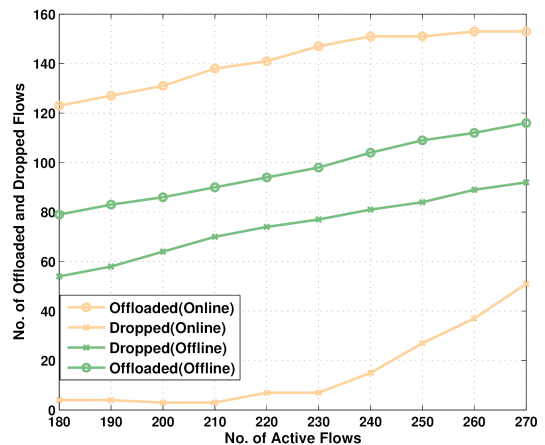


Fig. 6. Number of offloaded and dropped flows Vs. number of active flows for both online and offline policy derivations

- [5] K. Lee, J. Lee, Y. Yi, I. Rhee, and S. Chong, "Mobile Data Offloading: How Much Can WiFi Deliver?" *IEEE/ACM Transaction on Networking*, vol. 21, no. 2, pp. 536–550, April 2013.
- [6] B. Han, P. Hui, V. Kumar, M. Marathe, J. Shao, and A. Srinivasan, "Mobile Data Offloading through Opportunistic Communications and Social Participation," *IEEE Transactions on Mobile Computing*, vol. 11, no. 5, 2012.
- [7] A. De La Oliva, C. Bernardos, M. Calderon, T. Melia, and J. Zuniga, "IP Flow Mobility: Smart Traffic Offload for Future Wireless Networks," *IEEE Communication Magazine*, vol. 49, no. 10, pp. 124–132, October 2011.
- [8] "Architecture for Mobile Data Offload over WiFi Access Networks," Cisco White Paper, 2012.
- [9] L. E. Li, Z. M. Mao, and J. Rexford, "Toward Software-Defined Cellular Networks," in *European Workshop on Software-defined Networking (EWSN)*, October 2012.
- [10] "3GPP TS 23.402 V10.7.0: Architecture enhancements for non-3GPP accesses," Available online: [www.quintillion.co.jp/3GPP/Specs/23402-a70.pdf](http://www.quintillion.co.jp/3GPP/Specs/23402-a70.pdf), March 2013.
- [11] "3GPP TS 23.203 V9.3.0: Policy and Charging Control Architecture," Available online: [www.quintillion.co.jp/3GPP/Specs/23203-930.pdf](http://www.quintillion.co.jp/3GPP/Specs/23203-930.pdf), December 2009.
- [12] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, "OpenFlow: Enabling Innovation in Campus Networks," *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 2, April 2008.
- [13] "3GPP TS 25.814 V7.1.0: Physical Layer Aspects for Evolved Universal Terrestrial Radio Access(UTRA)," Available online: <http://www.3gpp.org/FTP/Specs/html-info/25814.htm>, October 2006.
- [14] G. Piro, L. Grieco, G. Boggia, F. Capozzi, and P. Camarda, "Simulating LTE Cellular Systems: An Open-Source Framework," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 2, pp. 498–513, 2011.