

Proportional and Preemption-enabled Traffic Offloading for IP Flow Mobility: Algorithms and Performance Evaluation

Yi Ren, Chih-Wei Tung, Jyh-Cheng Chen, *Fellow, IEEE*, and Frank Y. Li, *Senior Member, IEEE*

Abstract—IP Flow Mobility (IFOM) enables a user equipment to offload data traffic at the IP flow level. Although the procedure of IFOM-based flow offloading has been specified by 3GPP, how many IP flows should be offloaded and when offloading should be performed are not defined. Consequently, IP flows may be routed to a target access network which has a strong signal strength but with backhaul congestion or insufficient access capability. In this paper, we propose two algorithms, referred to as proportional offloading (PO), and proportional and preemption-enabled offloading (PPO), respectively, for IP flow offloading in hybrid cellular and wireless local area networks. The PO algorithm decides an optimal proportion of IP flows which could be offloaded by considering available resources at the target access network. In the PPO algorithm, both service continuity and network utilization are taken into consideration. Furthermore, a detailed analytical model is developed in order to evaluate the behavior of the proposed algorithms. The results show that by dynamically adjusting the percentage of traffic flows to be offloaded, PO can reduce blocking probability and increase resource utilization. PPO further improves the performance at the cost of slightly higher offloading overhead.

I. INTRODUCTION

To tackle the ever-growing capacity problem in cellular networks caused by exponentially growing data traffic nowadays, mobile operators are actively seeking for alternative offloading solutions. Data offloading techniques, which allow re-direct data traffic from a cellular network to another access network (AN), have been widely deployed. For example, operators may offload data traffic from a congested or close-to-congestion evolved universal terrestrial radio access network (E-UTRAN) to an IEEE 802.11 wireless local area network (WLAN), considering the ubiquity of Wi-Fi access points (APs) worldwide. In brief, there are two cases for Wi-Fi offloading. They are: (1) a user equipment (UE) offloads its whole traffic volume to a WLAN AN; and (2) a UE is connected to two access networks simultaneously and only a portion of its traffic is offloaded to

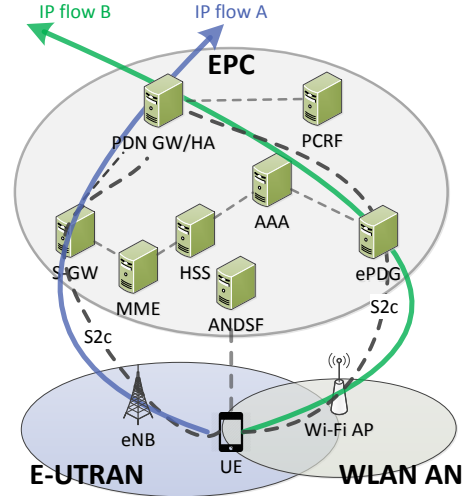


Fig. 1. A simplified IFOM architecture. Traffic from a UE is split into IP flow A and IP flow B, transmitted via E-UTRAN and/or WLAN AN respectively. The right-most dashed line indicates that IP flow B is switched from an eNB in E-UTRAN to a Wi-Fi AP.

the WLAN AN¹. *Internet protocol (IP) flow mobility (IFOM)* is one of the solutions for the latter case and it is specified by 3GPP [1].

IFOM defines a fine-grained and flexible mechanism by enabling a UE to offload data traffic *seamlessly at the IP flow level*. Based on dual stack mobile IPv6 (DSMIPv6), IFOM allows a UE to selectively re-route parts of its IP flows to another radio access network (RAN), while keeping other ongoing connections or data flows in the current RAN. Fig. 1 illustrates a simplified IFOM architecture. While a dual-mode UE has a connection consisting of several flows (e.g., video stream, delay tolerant traffic, etc.), the UE can keep IP flow A (a real-time flow, for example) staying in E-UTRAN while offloading IP flow B (such as a best effort flow) to a WLAN AN.

However, IFOM does not specify any techniques on *when and how many flows should be re-routed to a WLAN*. Nor is it clear on which criterion or parameter should be adopted to make an offloading decision. Offloading mobile traffic aggressively² from E-UTRAN to WLAN may degrade user experience and network performance. For instance, a Wi-

¹These WLANs in this study are assumed to be managed by the same operator that controls the E-UTRAN so that the operator is able to monitor the traffic of both access networks and to make decision for load balancing.

²It is meant to perform offloading as long as the signal strength at the target network is high enough.

Manuscript received November 16, 2017; revised June 23, 2018.

Y. Ren is with the School of Computing Science, University of East Anglia (UEA), Norwich, UK. E-mail: e.ren@uea.ac.uk

C.-W. Tung is with Mozilla Corp. Taipei, Taiwan, R.O.C. E-mail: shes050117@gmail.com

J.-C. Chen is with the Department of Computer Science, National Chiao Tung University, Hsinchu, Taiwan, R.O.C. E-mail: jcc@cs.nctu.edu.tw

F. Y. Li is with the Department of Information and Communication Technology, University of Agder (UiA), N-4898 Grimstad, Norway. E-mail: frank.li@uia.no

Fi AP with excellent signal strength may suffer from heavy traffic load or have less effective capacity. On the other hand, passive offloading³ cannot efficiently alleviate congestion in E-UTRAN and may lead to resource under-utilization at the target AN. Therefore, a more adaptive and effective offloading policy to diminish backhaul or target network congestion is needed.

In this paper, we propose two traffic offloading algorithms, referred to as proportional offloading (PO) and proportional and preemption-enabled offloading (PPO), respectively, for hybrid E-UTRAN and Wi-Fi networks. PO adaptively offloads IP flows by considering the available resources at the target network. PPO, instead, allows seamless offloading by preempting certain flows in order to maximize network utilization. Moreover, we specify the time to re-route IP flows in PPO. By doing so, the probability for maintaining IP services becomes higher and the network utilization is increased with moderate signaling overhead. The proposed PO and PPO algorithms distinguish themselves from the existing ones with respect to the following two aspects: 1) while most of the previous studies apply an on-the-spot offloading policy and offload all traffic flows from an AN to another AN, our algorithms offload only a certain percentage of these flows; and 2) offloading failures, which may happen if the available resource at the target AN is insufficient, are often overlooked in other studies but are considered before offloading decision making in our algorithms. Furthermore, our proposed algorithms provide operators with guidelines to offload traffic with an appropriate percentage by jointly considering three network performance parameters.

In brief, the main contributions of this paper are summarized as follows.

- Two algorithms are proposed for IP flow offloading in hybrid cellular and wireless networks. While PO answers the question on *how many* flows should be re-routed to a target AN, PPO further specifies *when* the flows should be re-routed to the target AN.
- A performance evaluation model to quantitatively analyze the performance of such traffic offloading algorithms is developed. Accordingly, a step-by-step analysis is performed to evaluate and compare the proposed two algorithms.
- Extensive discrete-event simulations are carried out in order to validate the analytical model. The correctness of the analytical model is verified via simulations under various traffic conditions.

The rest of this paper is organized as follows. Section II and Section III provide an overview of related work and the IFOM offloading mechanism, respectively. After the proposed offloading algorithms are presented in Section IV, the analytical model is discussed in Section V. The numerical results are illustrated in Section VI. Section VII concludes this paper. Furthermore, the derivation of the performance expressions is presented in an appendix.

³Offloading is performed only due to insufficient resource at the serving network, conditioned on high enough signal strength at the target network.

II. RELATED WORK

Traffic offloading algorithms have been extensively studied recently [2]–[10]. The authors of [2] proposed an integrated offloading algorithm by taking both signal quality and network load into consideration. In [3], an offloading mechanism through the abstraction of software-defined networking (SDN) in the mobile backhaul was proposed in order to provide programmable offloading policies in wireless networks. In [4], the authors proposed a dynamic offloading algorithm via a policy in the access network discovery and selection function (ANDSF). However, the studies presented in [2]–[4] do not take network information into consideration, nor are they possible to offload traffic flows during an ongoing session. These drawbacks may either lead to possible backhaul congestion when accessing an AP, or make it difficult to adjust traffic flows adaptively.

Moreover, several studies [5]–[7] investigated radio access technology (RAT) selection, enabling to offload traffic by finding an optimally connected network. A survey on mathematical modeling for network selection was presented in [5]. In [6], two game-theory-based approaches were proposed for the purpose of load balancing between worldwide interoperability for microwave access (WiMAX) and Wi-Fi. A general and tractable model to analyze inter-RAT offloading was developed in [7]. However, these studies [5]–[7] focused merely on on-the-spot offloading without considering dynamic traffic switching.

Considering the contention-based nature of Wi-Fi access, the authors of [8] proposed a network-assisted user-centric Wi-Fi offloading model to maximize per-user throughput by utilizing available network information in a heterogeneous network. However, the authors of [8] considered only users who can access both Wi-Fi and cellular networks and focused on throughput maximization from users' perspective. Therein, network utilization and channel occupancy from UEs with Wi-Fi only access were not considered. The work in [9] introduced a hybrid method using both traffic offloading and resource sharing approaches to offload traffic to Wi-Fi APs. The authors in [10] investigated the throughput-power tradeoff in Wi-Fi/cellular network offloading. A utility function is well defined to reconcile the offloading problem as a game. An incentive mechanism is proposed to encourage UEs with enough energy to use cellular networks. In contrast, in this paper we consider *dynamic switching of ongoing flows, traffic load in both access networks, network utilization*, as well as mathematical modeling for traffic offloading in hybrid cellular and Wi-Fi networks.

Recently, delay tolerance offloading for cellular networks with low traffic occupancy has received extensive attention [11]–[15]. The authors in [11] proposed a heuristic approach to delay to-be-offloaded data traffic flows. In [12], the authors proposed to use the potential cellular bytes savings from a delay-tolerant offloading scheduler. These two studies were performed based on the assumption that the offloaded traffic had certain level of delay tolerance. In [13], the authors addressed the research gap and proposed four algorithms to dynamically and adaptively deduce an application's delay tol-

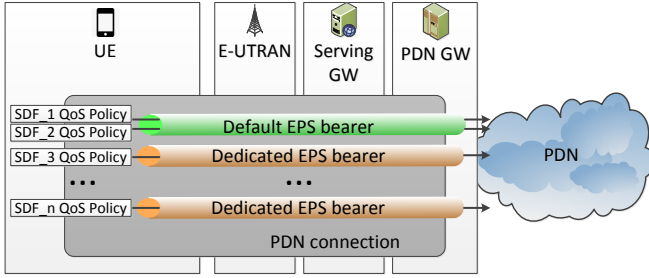


Fig. 2. A simplified illustration of a PDN connection with multiple IP flows.

erance. The work in [14] introduced a contract-based incentive mechanism for delayed traffic offloading in cellular networks. The incentive motivates users to leverage their delay and price sensitivity in exchange for service cost. The authors in [15] proposed a queueing analytic model for delayed offloading. A set of performance metrics, such as mean delay, offloading efficiency, were derived. Moreover, the model was validated using a range of realistic scenarios and real data traces. However, these studies [11]–[15] addressed traffic offloading from an application’s point of view. The proposed algorithms therein are not applicable to the scenarios in the studies which require IP-level flow offloading during a session based on IFOM.

III. BACKGROUND OF IFOM

In this section, we introduce the essentials of IFOM.

A. IFOM Architecture

Fig. 1 illustrates a simplified architecture of IFOM for a hybrid E-UTRAN and WLAN AN network⁴. Generally, the IFOM architecture can be divided into three parts: evolved packet core (EPC), E-UTRAN, and WLAN AN. In particular, the EPC consists of packet data network (PDN) gateway (GW), serving gateway (S-GW), mobility management entity (MME), policy and charging rules function (PCRF), ANDSF, evolved packet data gateway (ePDG), home subscriber server (HSS), authentication, authorization, and accounting (AAA) server, etc.

The PDN GW is a gateway which provides connectivity between UEs and an external PDN. An interface referred to as “S2c” is defined to support the communication between a PDN GW and a UE [1]. S-GW handles user data functions such as routing and packet forwarding to the PDN GW. MME is responsible for UE mobility, e.g., the procedures for paging and tagging idle mode UE. PCRF is a policy and charging control element supporting policy enforcement, service data flow detection, and flow-based charging. ANDSF provides information for UEs about connectivity to E-UTRAN and WLAN AN. While a UE attempts to access a network, ANDSF will help the UE to decide which network to connect to and provide routing policies to the UE. ePDG is responsible for securing the data transmission between the EPC and UEs for untrusted WLAN access. HSS is a master user database

⁴In this study, we use E-UTRAN and WLAN AN as examples for 3GPP access network and non-3GPP access network, respectively.

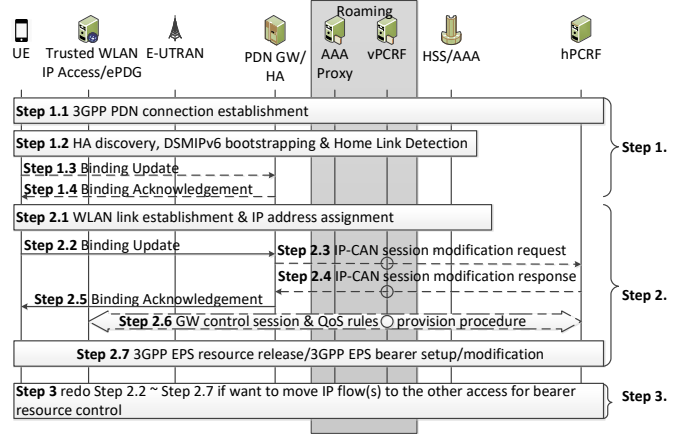


Fig. 3. IFOM offloading procedure for IP flow handover from E-UTRAN to WLAN AN based on 3GPP TS 23.261 [1].

that contains subscriber-related information and manages subscriber identities. The AAA server is a server that handles user requests for accessing to network resources and provides authentication, authorization, and accounting services.

B. IFOM Offloading Procedure

IFOM enables a UE to offload data *seamlessly* at the IP flow level. Considering that multiple flows co-exist from a UE to the EPC, the UE can route IP flows through the available access networks and to selectively offload part of the traffic to a WLAN AN. Fig. 1 illustrates an example that a UE accesses the EPC via two parallel IP flows, where flow A with a higher quality of service (QoS) requirement (e.g., a voice over IP (VoIP) session) goes through the E-UTRAN while flow B with a lower QoS requirement (e.g., a non-real-time (NRT) session) is offloaded to a WLAN AN.

The procedure for IFOM offloading is explained below.

1) *PDN connection*: When a UE accesses the Internet with an IP address via the Long-Term Evolution (LTE) network, a PDN connection is established. Within a PDN connection, there exists one or more IP flows which are classified into multiple *service data flow (SDF)* traffic flows. Fig. 2 shows the relationship between a PDN connection and its IP flows. Within a PDN connection, all evolved packet system (EPS) bearers share the same UE IP address.

2) *Binding cache*: To offload an IP flow to another wireless network, the UE needs to obtain a new IP address in the target network. Binding cache is a way regarding how a newly obtained address is associated with the existing IP address. IFOM adopts DSMIPv6 [16] to manage such an IP flow mobility procedure in order to change/update its point of attachment (PoA). Before a UE hands over from an E-UTRAN to a WLAN AN, the ANDSF provides a list of the available access networks to the UE. The UE then attempts to connect to one of the available access networks as the target network. Once a new IP address is assigned to the UE, the UE will register it as a care-of-address (CoA) with its home agent (HA). In this way, a binding is established between the target network and the network with existing flows. To support multiple IP flows, different CoAs may be used.

3) *LTE bearer resource control*: When a UE accesses a service through the LTE network, the PCRF enables centralized control to ensure that the service sessions are provided with sufficient radio resources and appropriate QoS. When the available resource is not sufficient to satisfy the service requested, the PCRF may reject the request. When the traffic load in the cellular network becomes heavier, a UE may prefer to offload its traffic to a WLAN AN. At the same time, the cellular network operators may not be willing to offload too many flows due to the consideration of achieving higher network utilization or to get more revenue from end users. So a routing policy has to be designed in order to decide when and how to offload an IP flow.

A step-by-step procedure for IFOM IP flow offloading, including PDN connection, binding cache, and bearer resource control, is illustrated in Fig. 3 Steps 1~3. For more details, please refer to [1].

C. Routing Policies in ANDSF

A routing policy is an operator's choice and it is a network-based decision by the ANDSF. A UE will then follow the routing policy decided by its operator to select a RAN and route its traffic flows correspondingly. Although an offloading and routing policy framework has been provided by 3GPP [17], it does not suggest any offloading policies.

To design a routing policy, the ANDSF needs to collect RAN assistance information [18], including cellular access thresholds, WLAN access thresholds, offload preference indication values, backhaul data rate for both uplink and downlink, channel utilization in IEEE 802.11 [19], etc. Before deciding to change the PoA for a specific IP flow, the ANDSF needs also to check whether the flow is re-routable (RR) or not⁵. A flow may be non-re-routable (NRR) if the UE is only covered by one RAN or it does not have multiple radio interfaces for connections with different RANs.

IV. THE PROPOSED PO AND PPO ALGORITHMS

In this section, we first present our design goals and then propose two flexible and fine-grained offloading algorithms based on IFOM, referred to as PO and PPO, respectively. While PO focuses on offloading an appropriate percentage, θ where $0 \leq \theta \leq 1$, of traffic flows to reduce blocking probability and increase network utilization, PPO allows dynamic switching of the PoA for *ongoing* flows in order to further improve network utilization of the whole network. The network resources considered in this study are radio resources at the backhaul of a RAT and a WLAN AN, respectively. Note that in the considered scenarios for offloading, a traffic flow is typically offloaded from the E-UTRAN to a WLAN, but a flow may be re-directed back to the E-UTRAN when necessary. Moreover, both PO and PPO are executed at the routing policy server, i.e., the ANDSF shown in Fig. 4. To make offloading decisions which reflect precisely the dynamic feature of traffic variations over time, the ANDSF collects the

statistic information on traffic load and resource utilization from the serving PDN GW, eNB, and WLAN AP periodically and makes proper decisions accordingly.

A. Design Goals

Three performance metrics are defined in this study as presented below. Since the value of θ has impact on all these three parameters, we include it in the notations.

- **Access blocking probability, $P_b(\theta)$** : This is the probability that a request of an IP flow cannot get access to the network, either in the E-UTRAN or in the WLAN. The IP flow considered can be either an RR or an NRR flow. While an NRR flow may be rejected due to lack of resources at the serving RAN, an RR flow may be rejected during the registration phase if there are not enough resources at the target RAN.
- **Network utilization, $U(\theta)$** : The resource utilization of a RAN is a dimensionless parameter represented by the ratio between the number of resource units occupied by ongoing flows and the total number of resource units in the network.
- **Offloading overhead, $\alpha(\theta)$** : This is the additional signaling overhead caused by an offloading algorithm. It is calculated as the ratio between the number of offloading requests and the total number of arrivals to a network.

Bearing the above definitions in mind, we reiterate the design goal for an offloading algorithm as to increase $U(\theta)$ while keeping $P_b(\theta)$ and $\alpha(\theta)$ as low as possible.

B. Proportional Offloading (PO)

As mentioned earlier, an aggressive offloading algorithm offloads UE traffic as long as the signal strength in the target AN is high enough for channel access, whereas a passive offloading algorithm offloads UE traffic only when no resource is available at the serving AN and at the same time the signal strength in the target AN exceeds a pre-defined threshold. However, no traffic condition is considered in these algorithms. In contrast, our algorithms make an offloading decision not only depending on signal strength but also by taking traffic load and available resources at both E-UTRAN and WLAN into consideration.

A basic rule for PO is that a certain percentage of the arrival IP flows should be offloaded to the WLAN while keeping the rest in the E-UTRAN in order to achieve a low blocking probability and high network utilization. When a new IP flow arrives to the E-UTRAN, the routing policy informs the UE whether this flow should be offloaded to the WLAN or not, depending on the rule. The offloading decision on the percentage of IP flows to be offloaded is determined by the ANDSF server. To reduce signaling cost, the information of traffic load and resource utilization is collected periodically by the ANDSF server rather than the UEs. Such information can be collected from the serving PDN GW, eNodeB, or WLAN AP. Based on the collected information, the ANDSF server decides the percentage of traffic offloading and informs each UE on how to route or re-route its traffic flows according to the decided routing policy.

⁵Since both access networks are managed by the same operator, it is feasible to check the connectivity as well as coverage information for each UE.

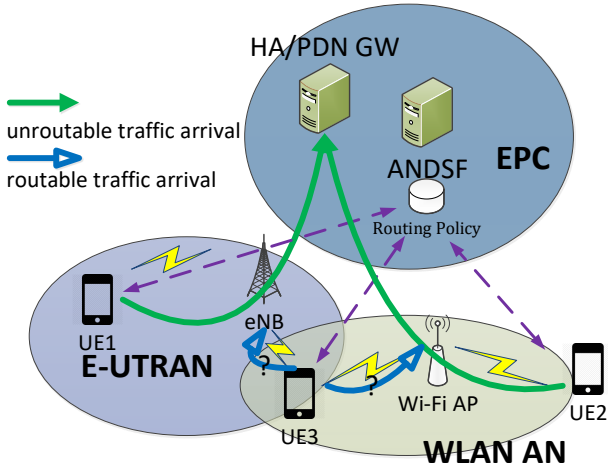


Fig. 4. An illustration of hybrid E-UTRAN and WLAN.

Consider a hybrid network composed of E-UTRAN, WLAN AN, and an overlapped area of both networks, as shown in Fig. 4. A UE may access the EPC through E-UTRAN or/and Wi-Fi RATs. For traffic offloading, a UE follows the routing policy it receives from the ANDSF server. Inside this hybrid RAN, the IP flows to and from UEs are classified into RR and NRR IP flows. As shown in the figure, the flows connecting UE1 and UE2 are non-re-routable since they are covered only by one RAN, whereas the flows from UE3 are re-routable since UE3 is connected to both RANs. If an arriving flow is NRR, the routing policy can only direct it to its serving RAN⁶. If the arriving IP flow is re-routable, the routing policy will inform the UE whether this flow should stay in the E-UTRAN or should be offloaded to the WLAN AN. Hence, the offloading decision is applicable to arriving IP flows carried by a UE with access opportunities to both the E-UTRAN and the WLAN AN.

In Algorithm 1, we illustrate the procedure of the proposed PO algorithm. Three performance parameters, i.e., network utilization, $U(\theta)$, access blocking probability, P_b , and offloading overhead, $\alpha(\theta)$, are considered in our study. Their expressions will be presented in Sec V. In order to make a proper and balanced decision based on these three performance parameters, we have introduced three weight factors, w_1 , w_2 , and w_3 which could be configured according to an operator's preference. However, to make our algorithms more flexible, we do not suggest any concrete values for these weight factors due to the consideration that such a value should be determined by a mobile operator when making their offloading policies.

Discussion: PO provides guidelines to identify an appropriate percentage of traffic load for offloading, but it allows offloading merely at the beginning of flow access. When an NRR flow arrives, an ongoing flow would probably be forcibly terminated if no resources are available at the serving AN. As commonly understood from a QoS point of view, to forcibly terminate an ongoing flow is more annoying for an end user

⁶A serving RAN is the radio access network a UE is currently connected to, whereas a target RAN is the radio access network to which an IP flow attempts to join via offloading.

Algorithm 1 Selecting the percentage of re-routable IP flows to be offloaded in PO.

Input: λ_e, λ_w : Mean value of NRR IP flow arrival rate from E-UTRAN and WLAN AN, respectively

Input: λ_m : Mean value of re-routable IP flow arrival rate from E-UTRAN and WLAN AN overlapped area

Input: C_e, C_w : Resource capability of E-UTRAN and that of WLAN AN.

Input: μ_e, μ_w : Mean value of UE residence time in E-UTRAN cell and WLAN cell, respectively

Input: T : Evaluation interval

Output: θ : The percentage of re-routable IP flows to be offloaded

Ensure: The signal strength in the target RAN is enough to access

- 1: **set** $t := T$
- 2: **while** $t \geq 1$ **do**
- 3: Calculate $U(\theta)$ using (3)
- Calculate $P_b(\theta)$ using (1)
- Calculate $\alpha(\theta)$ using (7)
- $\arg \max_{\theta} \{w_1 U(\theta) + w_2 P_b(\theta) + w_3 \alpha(\theta)\}$, where w_1, w_2 , and w_3 are weight factors.
- Offload θ percent IP flows to WLAN AN
- set** $t := t - 1$
- 4: **end while**

than being rejected before the session is established. It could also further improve network utilization if we could dynamically offload traffic on-the-fly. These observations motivate us to develop another offloading algorithm which allows UEs to switch the PoA of an *ongoing* flow.

C. Proportional and Preemption-enabled Offloading (PPO)

As traffic load increases, more IP flows may be offloaded to WLAN. Considering that if there are additional incoming traffic flows, either RR or NRR, to the WLAN AN, the blocking probability could be high. The PO algorithm presented above performs traffic offloading to reduce blocking probability and increase network utilization merely for E-UTRAN. In contrast, PPO may re-route IP flows from a WLAN to an E-UTRAN in order to increase resource utilization of the whole network. It also specifies the time⁷ for a UE to perform PoA switching based on IFOM. The proposed PPO is designed to dislodge the negative effect of offloaded traffic on NRR IP flows in the WLAN AN. When an NRR IP flow arrives to the WLAN AN and all available resources are occupied, the ANDSF will not block the request. Instead, it chooses an ongoing re-routable IP flow and re-directs it to the E-UTRAN so that both flows can be kept in the hybrid network. In a nutshell, the PPO scheme allows dynamic switching of the PoA for *ongoing* flows so that the network utilization for the whole network including both E-UTRAN and WLAN is improved.

Fig. 5 shows the procedure of the proposed PPO. After the PCRF receives the request to add an NRR IP flow to the WLAN AN which does not have enough available resource (Step 2.3), it checks whether there are re-routable IP flows in the WLAN AN or not. If there is one re-routable IP flow, the PCRF selects it as a *victim* flow and sends a request to the PDN GW for PoA switching to E-UTRAN (Step 2.5). If there

⁷The time instant of the *session modification request* is initiated, which is illustrated as Step 2.3 in Fig. 5.

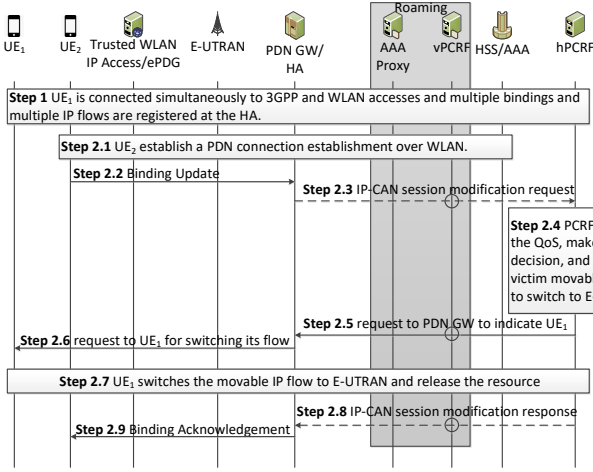


Fig. 5. A message flowchart for PPO offloading: IP flows may be dynamically re-directed back to E-UTRAN based on 3GPP TS 23.261 [1].

are multiple re-routable flows, one of them will be selected randomly as the victim flow. The PDN GW requests to the UE via the S2c interface (Step 2.6). If the PoA switching is successful, the released resource in the WLAN AN will be re-allocated to accommodate the arriving NRR IP flow. By adding slight signaling overhead, only an additional decision and two request flows (shown in Step 2.4 ~ Step 2.6), to the current IFOM message flowchart, the proposed PPO algorithm can offload traffic flows dynamically. In this way, the network utilization is improved while the access privilege of NRR IP flows is still guaranteed.

Discussion: The procedure for PoA switching may lead to extra delay. To ensure seamless service continuity based on PPO, we apply PoA switching to NRR flows only. Furthermore, additional signal overhead may be imposed which is the cost for higher network utilization and lower blocking probability. In the next section, we propose an analytical model to analyze the performance of such traffic offloading algorithms.

V. PERFORMANCE ANALYSIS

In this section, we develop a framework to analyze the performance of traffic offloading algorithms for IFOM. Following a common practice used in communication network modeling [20]–[22], the arrivals of IP flows to both E-UTRAN and WLAN are assumed to be Poisson processes. We denote λ_m , λ_e , and λ_w as the arrival rate for RR IP flow, NRR IP flows to the E-UTRAN, and NRR IP flows to the WLAN, respectively. Let θ be the percentage of re-routable IP flows to be offloaded. The arrival rate for re-routable IP flow registrations to the WLAN and the E-UTRAN becomes $\theta\lambda_m$ and $(1 - \theta)\lambda_m$, respectively. The notations adopted in our analysis are listed in Table I. In the followings, we conduct performance analysis for both PO and PPO.

A. Blocking Probability $P_b(\theta)$

For PO, forced termination of an ongoing IP flow will not occur since it offloads IP flows only at the beginning of access. Let p_{bw}^{PO} and p_{be}^{PO} represent the probabilities that the WLAN

TABLE I
LIST OF NOTATIONS

Notation	Explanation
$U(\theta)$	The resource utilization
$\alpha(\theta)$	The additional signaling overhead caused by an offloading algorithm
λ_w/λ_e	Mean value of the NRR IP flow arrival rate to WLAN AN/E-UTRAN respectively
λ_m	Mean value of the RR IP flow arrival rate to the E-UTRAN and WLAN AN overlapped area
λ_{tr}	Mean value of the arrival rate for transferring IP flows
λ_{eff}	The effective arrival rate seen by the system
t_f	IP flow lifetime
$t_{er}/t_{e,i}/t_{wr}/t_{w,i}$	IP flow residence time in the E-UTRAN/ <i>i</i> -th E-UTRAN cell/WLAN AN/ <i>i</i> -th WLAN AN cell
$\tau_{e,i}/\tau_{w,i}$	The residual life of $t_{e,i}/t_{w,i}$
$\tau_{fe,i}/\tau_{fw,i}$	The residual life of an IP flow after UE visited the <i>i</i> -th E-UTRAN/WLAN AN cell
$\tau_{e,i}^*$	The age of $t_{e,i}$
$1/\mu$	Mean value of IP flow lifetime t_f
$1/\mu_e(1/\mu_w)$	Mean value of the IP flow lifetime in the E-UTRAN (WLAN) cell
$1/\eta_w(1/\eta_e)$	Mean value of the resident time in the <i>i</i> -th WLAN cell $t_{w,i}$ (E-UTRAN cell $t_{e,i}$)
C_w/C_e	WLAN AN / E-UTRAN resource capacity
ρ_e^{PO}/ρ_e^{PPO}	E-UTRAN cell resource utilization for PO/PPO respectively
ρ_w^{PO}/ρ_w^{PPO}	WLAN AN cell resource utilization for PO/PPO respectively
P_b^{PO}/P_b^{PPO}	The probability that an IP flow request cannot get access to the network for PO/PPO respectively
p_{be}^{PO}/p_{be}^{PPO}	The probability of IP flows being blocked in the E-UTRAN cell for PO/PPO respectively
p_{bw}^{PO}/p_{bw}^{PPO}	The probability of IP flows being blocked in the WLAN AN cell for the PO algorithm and the PPO algorithm respectively
p_{bh}^{PPO}	The probability of high priority IP flows being blocked in the WLAN AN cell for the PPO algorithm
p_f^{PPO}	The probability of failure when transferring re-routable IP flows from the WLAN AN to the E-UTRAN cell
L_e/L_w	The number of IP flows accessing through the E-UTRAN/WLAN AN
ρ_e/ρ_w	The resource utilization in the E-UTRAN/WLAN AN
ρ_e^{PO}/ρ_w^{PO}	The effective network utilization for the E-UTRAN/WLAN AN

AN and the E-UTRAN block the request of an arriving IP flow, respectively.

According to PO, there are four reasons that an IP flow cannot access the network. They are: (1) an arriving NRR flow to the WLAN AN is blocked; (2) an arriving NRR flow to the E-UTRAN is blocked; (3) an arriving RR flow which should be kept to the E-UTRAN (i.e., not offloaded) is blocked; and (4)

an arriving RR flow which should be offloaded to the WLAN is blocked. Therefore, the blocking probability of PO for a given offloading percentage θ , $P_b^{PO}(\theta)$, is obtained as:

$$P_b^{PO}(\theta) = \frac{p_{bw}^{PO}\lambda_w + p_{be}^{PO}\lambda_e + p_{be}^{PO}(1-\theta)\lambda_m + p_{bw}^{PO}p_{be}^{PO}\theta\lambda_m}{\lambda_w + \lambda_m + \lambda_e}. \quad (1)$$

For PPO, there is another reason that a flow would be blocked in addition to the same four reasons mentioned above for PO. That is, an ongoing RR flow which is expected to switch its PoA from the WLAN to the E-UTRAN is blocked, with a probability of p_f^{PPO} . Let p_{be}^{PPO} , p_{bh}^{PPO} , and P_{bw}^{PPO} be the probabilities that the E-UTRAN blocks an arriving flow request, the probabilities that the WLAN AN blocks a request from a higher priority (i.e., NRR) and a lower priority (i.e., RR) IP flow, respectively. The blocking probability for PPO, $P_b^{PPO}(\theta)$, is derived as:

$$P_b^{PPO}(\theta) = \frac{p_{bh}^{PPO}\lambda_w + (p_f^{PPO} + p_{bw}^{PPO}p_{be}^{PPO})\theta\lambda_m}{\lambda_w + \lambda_m + \lambda_e} + \frac{p_{be}^{PPO}\lambda_e + p_{be}^{PPO}(1-\theta)\lambda_m}{\lambda_w + \lambda_m + \lambda_e}. \quad (2)$$

B. RAN Resource Utilization $U(\theta)$

Let L_e and L_w be the number of IP flows accessing through the E-UTRAN and the WLAN AN, respectively. We assume that each flow occupies the same amount of resource, as one resource unit, and consider the amount of resource units at the WLAN AN and the E-UTRAN as C_w and C_e units. The resource utilization of the hybrid RAN consisting of an E-UTRAN and a WLAN AN, $U(\theta)$, then can be easily computed as follows:

$$U(\theta) = \frac{L_e + L_w}{C_e + C_w}. \quad (3)$$

Denote by ρ_e and ρ_w the resource utilization in the E-UTRAN and the WLAN AN, respectively. Since no queue is considered in our performance evaluation model, the number of resource units occupied in each RAN is equal to the product of its resource utilization and the total resource units in that network. Therefore, we can rewrite (3) as:

$$U(\theta) = \frac{C_e\rho_e + C_w\rho_w}{C_e + C_w}. \quad (4)$$

Accordingly, the RAN resource utilization for PO is expressed as:

$$U^{PO}(\theta) = \frac{\rho_e^{PO}(\theta)C_e + \rho_w^{PO}(\theta)C_w}{C_e + C_w}, \quad (5)$$

and for PPO, it becomes:

$$U^{PPO}(\theta) = \frac{\rho_e^{PPO}(\theta)C_e + \rho_w^{PPO}(\theta)C_w}{C_e + C_w}, \quad (6)$$

where ρ_e^{PO} (or ρ_e^{PPO}) and ρ_w^{PO} (or ρ_w^{PPO}) represent the resource utilization in the E-UTRAN and the WLAN AN for PO or PPO, respectively.

TABLE II
SIMULATION PARAMETER CONFIGURATION

Parameter	Value
$1/\mu$	600 sec
η_e	2μ
$1/\eta_w$ for WLAN	300 sec
C_w for WLAN	5 units
λ (light)	$0.2C\mu$
λ (heavy)	$2C\mu$
η	$0.1\mu \sim 10\mu$
η_w	2μ
$1/\eta_e$ for E-UTRAN	300 sec
C_e for E-UTRAN	10 units
λ (medium)	$C\mu$

C. Offloading Overhead $\alpha(\theta)$

The offloading overhead depends heavily on the number of offloading requests between the E-UTRAN and the WLAN. When θ increases, there are more IP flows attempting to access the WLAN AN, leading to a higher blocking probability in the WLAN AN. If a request for adding an IP flow is blocked by the WLAN AN, the UE with an arriving flow will send another request to the E-UTRAN for possible access. We regard these requests as offloading overhead. Thus, the offloading overhead for PO, $\alpha^{PO}(\theta)$, is calculated as:

$$\alpha^{PO}(\theta) = \frac{p_{bw}^{PO}(1-\theta)\lambda_m}{\lambda_w + \lambda_m}. \quad (7)$$

In PPO, when an event to transfer an ongoing re-routable flow to the E-UTRAN happens, it also causes signaling overhead. Therefore, the offloading overhead in PPO, $\alpha^{PPO}(\theta)$, is derived as:

$$\alpha^{PPO}(\theta) = \frac{p_{bw}^{PPO}(1-\theta)\lambda_m + (p_{bw}^{PPO} - p_{bh}^{PPO})\lambda_w}{\lambda_w + \lambda_m}. \quad (8)$$

In the appendix, we will further derive the detailed expressions for the above three parameters.

VI. SIMULATIONS AND NUMERICAL RESULTS

The proposed PO and PPO algorithms are implemented in ns-2 [23], version 2.35, based on the network scenario shown in Fig. 4. The parameters are configured as illustrated in Tab. II unless otherwise stated. The results reported in this section are the average values obtained over 1,000,000 simulation time. The confidence intervals are so small that they overlap with the lines (the results by analysis) and the points (the results by simulations). The solid and dash-dot lines shown in the figures refer to the results obtained from PO and PPO, respectively.

In the following subsections, we illustrate the performance of PO and PPO with respect to blocking probability, resource utilization and offloading overhead as a function of traffic intensity and offloading percentage. Note that $\theta = 100\%$ and $\theta = 0\%$ in each figure correspond to aggressive and passive offloading policies which lead to un-adjustable performance.

Using PO and PPO, we can adjust the value of θ properly in order to achieve satisfactory performance.

The sub-figures in Figs. 6, 7, and 8 show the results obtained under four different traffic intensity levels:

- Sub-figure (a): Light traffic in both E-UTRAN and WLAN (with $\frac{\lambda}{C\mu} = 0.2$).
- Sub-figure (b): Heavy traffic in E-UTRAN (with $\frac{\lambda}{C\mu} = 2$) and light traffic in WLAN. More IP flows can be accommodated in the WLAN AN.
- Sub-figure (c): Light traffic in E-UTRAN (with $\frac{\lambda}{C\mu} = 2$) and heavy traffic in WLAN.
- Sub-figure (d): Medium traffic in both E-UTRAN and WLAN (with $\frac{\lambda}{C\mu} = 1$). In this case, both cells can accommodate their NRR IP flows, but the offloaded RR IP flows may cause cell congestion.

A. Effects of θ and λ_m on $P_b^{PO}(\theta)$ and $P_b^{PPO}(\theta)$

Fig. 6 plots $P_b^{PO}(\theta)$ and $P_b^{PPO}(\theta)$ against θ (the percentage of offloading traffic) and λ_m (the arrival rate for re-routable flows). When the intensity of NRR traffic in both the E-UTRAN and the WLAN AN is low ($\frac{\lambda_e}{C_e\mu_e} = 0.2 < 1$, $\frac{\lambda_w}{C_w\mu_w} = 0.2 < 1$), UEs can access the network with a low blocking probability. With a higher arrival rate of RR traffic (λ_m), however, the blocking probability increases. Offloading will then help to reduce blocking probability. For example, when $\lambda_m = 10\mu_w$, $P_b^{PO}(\theta)$ decreases as θ increases. This is because a certain amount of arriving flows are offloaded. When $\lambda_m = \mu_w$, there are few re-routable IP flows in the WLAN AN. When more IP flows attempt to access the WLAN AN (due to a higher θ), some NRR IP flows may be blocked, leading to a slightly increased blocking probability.

When $\lambda_m = 5\mu_w$, $P_b^{PO}(\theta)$ does not behave monotonically, reaching a minimum value when $\theta \approx 40\%$. The reason is that offloading re-routable IP flows reduces the access opportunity for the NRR IP flows in the WLAN AN but increases the access probability of re-routable traffic. As the result of a decreased P_b^{PO} for re-routable traffic and an increased P_b^{PO} for NRR traffic, an optimal value for θ is reached. On the other hand, PPO is able to protect the NRR IP flows in the WLAN AN by re-directing certain amount of RR flows back to the E-UTRAN, leading to a monotonically decrease of P_b^{PPO} in this case.

Fig. 6(b) shows the result when the traffic intensity in the E-UTRAN is high and the WLAN AN is low ($\frac{\lambda_e}{C_e\mu_e} = 2 > 1$, $\frac{\lambda_w}{C_w\mu_w} = 0.2 < 1$). No matter how many RR IP flows arrive, offloading IP flows always reduces the blocking probability with both PO and PPO. The larger the θ , the lower the $P_b^{PO}(\theta)$ and the $P_b^{PPO}(\theta)$. Since no significant benefit is achieved by PPO in this case, employing PO is recommended considering its lower offloading overhead (to be presented later).

With light traffic intensity in the E-UTRAN and heavy traffic in WLAN (configured as $\frac{\lambda_e}{C_e\mu_e} = 0.2 < 1$, $\frac{\lambda_w}{C_w\mu_w} = 2 > 1$), the $P_b^{PO}(\theta)$ grows linearly when θ increases as shown in Fig. 6(c). On the other hand, PPO transfers certain amount of IP flows back to the lightly-loaded E-UTRAN, resulting in both a higher access opportunity in the E-UTRAN and a

lower blocking probability in the WLAN AN. Hence, the total blocking probability is decreased with a higher θ .

Furthermore, it is interesting to observe that when $\lambda_m = 5\mu_w$, lower $P_b^{PO}(\theta)$ and $P_b^{PPO}(\theta)$ are achieved in comparison with under the traffic conditions of both $\lambda_m = \mu_w$ and $\lambda_m = 10\mu_w$. This is because the available resource in WLAN is very low ($C_w = 5$ versus $C_e = 10$) and the WLAN AN could be easily congested when more NRR flows arrive. With more arriving RR flows ($\lambda_m = 5\mu_w$), many of them could be re-allocated to the E-UTRAN, leading to a low P_b for both PO and PPO. When the number of RR flows further increases, however, no enough resources are available in the E-UTRAN either. The P_b then increases again.

Fig. 6(d) illustrates that the blocking probability decreases slightly when more flows are offloaded for both PO and PPO with medium traffic load in both the E-UTRAN and the WLAN AN. This means that when both networks are busy, offloading traffic does not reduce blocking probability to a large extent.

In summary, PPO reduces blocking probability more significantly than PO does under most traffic conditions. For a given traffic condition, an appropriate θ value needs to be configured.

B. Effects of θ and λ on $U_b^{PO}(\theta)$ and $U_b^{PPO}(\theta)$

Fig. 7 shows how the RAN resource utilization, $U_b^{PO}(\theta)$ and $U_b^{PPO}(\theta)$, varies when θ and λ_m change. In three out of the four traffic conditions, the utilization increases when θ increases for both PO and PPO as shown in Figs. 7(a), (b), and (d). In these three cases, the traffic intensity in the WLAN AN is light or medium. Thus, offloading more flows to WLAN would lead to higher resource utilization. The larger the θ , the higher the $U^{PO}(\theta)$ and the $U^{PPO}(\theta)$.

However, when the WLAN AN is already heavily loaded, increasing θ will result in lower resource utilization if PO is employed. The reason is that offloading IP flows towards an already heavily loaded target network will make it quickly congested. On the other hand, when PPO is employed, some of those would-be-rejected flows by the WLAN AN are transferred back to the E-UTRAN. Thus, the resource utilization keeps at a quite stable level under PPO.

In short, from the perspective of maximizing RAN resource utilization, offloading more IP flows to the WLAN AN would be beneficial as long as the target network is not close to congestion. PPO keeps the utilization stable even when the WLAN AN is busy. When the traffic load in the WLAN AN is low or medium, more flows should be offloaded when more re-routable flows arrive in order to achieve higher resource utilization. When few re-routable flows arrive, however, little benefit could be achieved by traffic offloading.

C. Effects of θ and λ on $\alpha^{PO}(\theta)$ and $\alpha^{PPO}(\theta)$

Finally, we look at the offloading overhead $\alpha^{PO}(\theta)$ and $\alpha^{PPO}(\theta)$ in Fig. 8 against θ and λ_m . As expected, higher overhead is introduced with both PO and PPO when more flows are offloaded. This is because that a higher number of registrations are needed when more flows are offloaded. With PPO, this number will further increase due to the fact that

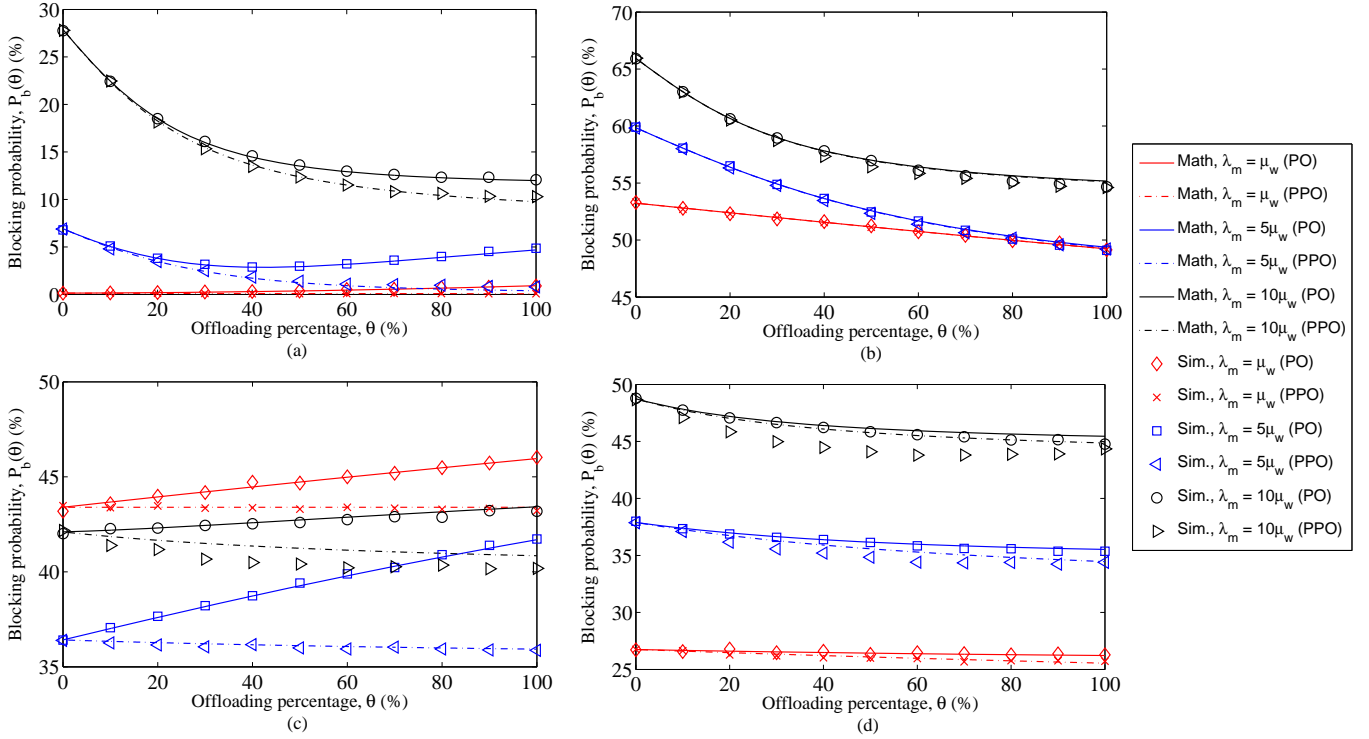


Fig. 6. Blocking probability $P_b^{PO}(\theta)$ and $P_b^{PPO}(\theta)$ with different re-routable flow arrival rate λ_m and offloading percentage θ .

a portion of the flows are re-routed back to the E-UTRAN. As a consequence, the offloading overhead caused by PPO is always higher than that of PO.

Recall that lower blocking probability and higher resource utilization can be achieved by PPO as presented earlier in this section. It is therefore of interest to investigate further the tradeoff between the benefits and the cost when employing PPO. For example, when the arrival rate of re-routable IP flows is high ($\lambda_m = 10\mu_w$), offloading 50% of arriving IP flows will lead to high $U(\theta)$ (80% or higher) and moderate $P_b(\theta)$ at a cost of around 20% of $\alpha(\theta)$. When the arrival rate of re-routable IP flows is low ($\lambda_m = \mu_w$), offloading or not does not have significant impact on $P_b(\theta)$ and $U(\theta)$. Thus, offloading passively (i.e., $\theta = 0\%$) is recommended considering its zero offloading overhead. When a medium level of re-routable flows arrive, e.g., when $\lambda_m = 5\mu_w$, an appropriate θ value needs to be configured based on the system requirement. If a mobile operator can tolerate high offloading overhead, allocating all of the IP flows to the WLAN AN can maximize the RAN resource utilization. On the other hand, if the mobile operator intends to minimize blocking probability, offloading a lower percentage, e.g., 30%, of IP flows to the WLAN AN is a better policy with low offloading overhead.

To summarize, PO defines a simple yet effective offloading policy which can reduce blocking probability for network access and raise RAN resource utilization. PPO further reduces $P_b^{PPO}(\theta)$ and increases $U^{PPO}(\theta)$ even when the WLAN AN is busy, at a cost of a slightly higher level of offloading overhead. Since PPO performs properly under different traffic conditions, it is a more efficient and favorable option for traffic offloading in hybrid cellular networks. With a higher

percentage for traffic offloading, the blocking probability goes monotonically downwards and the resource utilization becomes gradually higher or keeps stable while the signaling overhead grows monotonically. Therefore, to design a suitable offloading policy, a mobile operator needs to identify an appropriate θ that leads to an optimal level of $P_b(\theta)$ and $U(\theta)$ at a tolerable level of $\alpha(\theta)$.

VII. CONCLUSIONS AND FUTURE WORK

In this paper, we investigate the IFOM mechanism that enables a UE with multiple IP flows to switch a specific IP flow between E-UTRAN and WLAN. When a UE accesses a new IP service or switches its IP flows to another RAN, it needs to send a Binding Update to the serving PDN GW and requests for IP bearer establishment. If the resource is not enough, the request would be blocked. More specifically, we propose PO to reduce blocking probability and improve RAN resource utilization. However, PO may limit the access opportunities for WLAN-only UEs due to the offloaded traffic flows. Therefore, we further propose PPO as an advanced algorithm to re-direct re-routable IP flows back to the E-UTRAN in order to improve the access opportunities of those WLAN-only UEs and increase resource utilization for the whole network.

As a candidate solution for network based IFOM (NBI-FOM) [24], we propose an IP flow offloading procedure by adding a step with a decision and two requests via the S2c interface. Furthermore, we have developed an analytical model and performed extensive simulations to validate the mathematical model and evaluate the performance of the proposed PO and PPO algorithms with respect to blocking

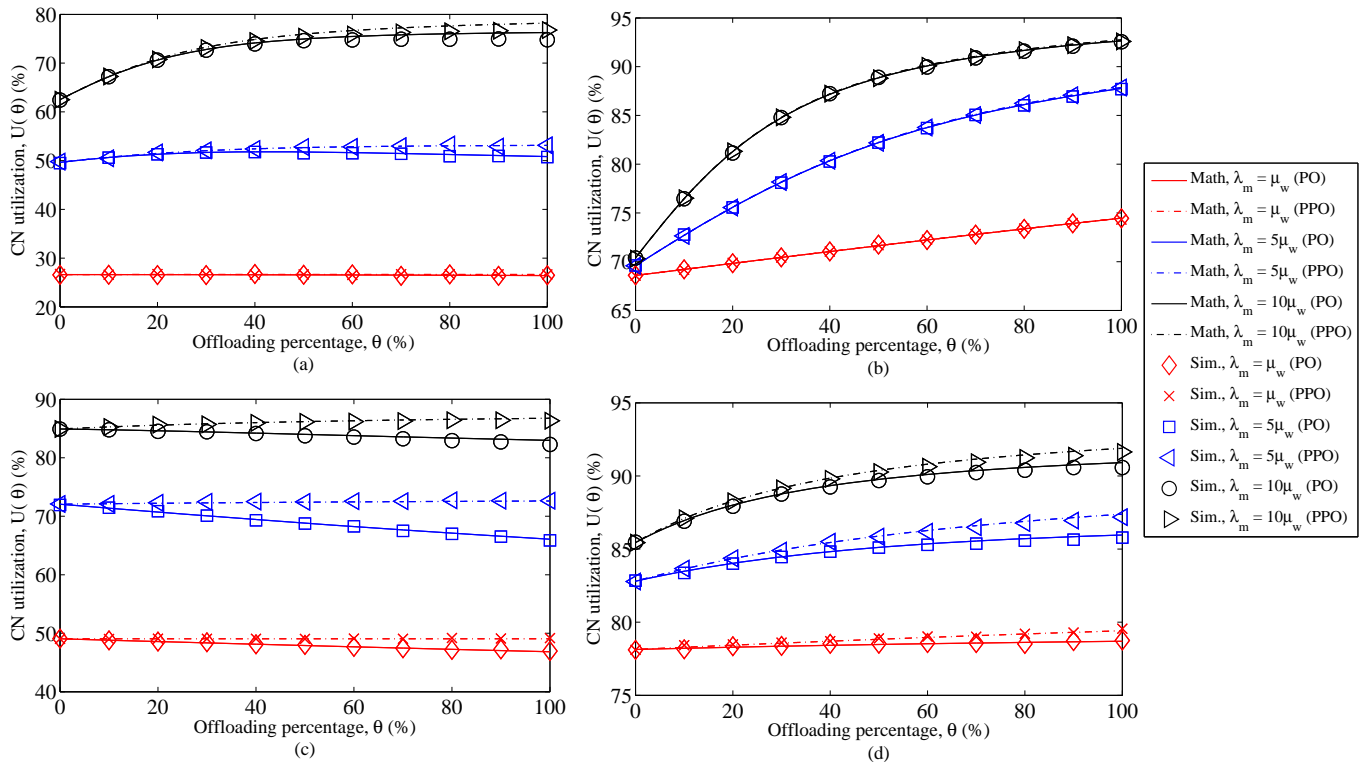


Fig. 7. Network utilization $U_b^{PO}(\theta)$ and $U_b^{PPO}(\theta)$ with different re-routable flow arrival rate λ_m and offloading percentage θ .

probability $P_b(\theta)$, RAN resource utilization $U(\theta)$, and offloading overhead $\alpha(\theta)$. Our study indicates that the θ value should not be kept as a constant in an offloading policy. By dynamically adjusting θ to an appropriate value, PO can reduce blocking probability and increase resource utilization. PPO further improves the performance with an acceptable cost of signaling overhead.

For future work, we plan to explore the impact of medium access control protocols on the performance of both PO and PPO, especially in terms of latency. Besides, it is interesting to study the impact of PO and PPO on core networks.

APPENDIX: DERIVATION OF PERFORMANCE PARAMETER EXPRESSIONS

In this appendix, we further derive the expressions for the three parameters defined in Section V. To do so, we need to calculate the arrival rates and service rates first.

A. Arrival Rates and Service Rates

Consider the behavior of a UE in the overlapped area of a hybrid E-UTRAN and WLAN AN. Let the IP flow lifetime be t_f , and the residence time in the E-UTRAN and the WLAN AN be t_{er} and t_{wr} , respectively. Assume that t_f , t_{er} , and t_{wr} are exponentially distributed with a mean value of $1/\mu$, $1/\eta_e$, and $1/\eta_w$, respectively.

Fig. 9 illustrates an event (such as arrival, handover, departure, etc.) diagram for an IP flow from a mobile UE. The arrivals to the E-UTRAN can be generally classified into three cases. That is, (1) initial access (occurred at t_1); (2) handover

access (occurred at t_5 , t_8 , t_9); and (3) re-directed access since a WLAN AN is fully occupied and the flow is switched back to the E-UTRAN (occurred at t_4 , t_7). At the period between t_2 and t_4 (also between t_6 and t_7), the UE is attached to both the E-UTRAN and the WLAN AN. Thus, during this period it is regarded as a re-routable IP flow. In the rest of the time, the PDN GW considers this IP flow as an NRR traffic flow. The other times shown in Fig. 9 are as follows. $t_{e,i}$: the residence time in the i -th E-UTRAN cell; $t_{w,i}$: the residence time in the i -th WLAN AN cell; $\tau_{e,i}$ (or $\tau_{w,i}$): the residual duration of $t_{e,i}$ (or $t_{w,i}$); and $\tau_{e,i}^*$: the age of $t_{e,i}$.

To find the service rate, we consider the initial and handover cases. Denote the remaining IP flow lifetime as $\tau_{fe,i}$ after the UE visited the i -th E-UTRAN cell. Based on the memoryless property of the exponential distribution, the probability density function (pdf) of the residence time in any E-UTRAN cell $f_e(t)$ becomes:

$$f_e(t) = \int_{\tau_{se,i}=t}^{\infty} f_s(\tau_{se,i}) f_{er}(t) d\tau_{se,i} + \int_{t_{er}=t}^{\infty} f_{er}(t_{er}) f_s(t) dt_{er} = (\mu + \eta_e) e^{-(\mu + \eta_e)t}, \quad (9)$$

with a mean value

$$\begin{aligned} \frac{1}{\mu_e} &= E[t_e] = \int_{t=0}^{\infty} t f_e(t) dt \\ &= \int_{t=0}^{\infty} t (\mu + \eta_e) e^{-(\mu + \eta_e)t} dt = 1/(\mu + \eta_e). \end{aligned} \quad (10)$$

Similarly, the residual time of an IP flow in the WLAN AN is $\tau_{fw,i}$ and its residence time in the WLAN AN is t_{wr} .

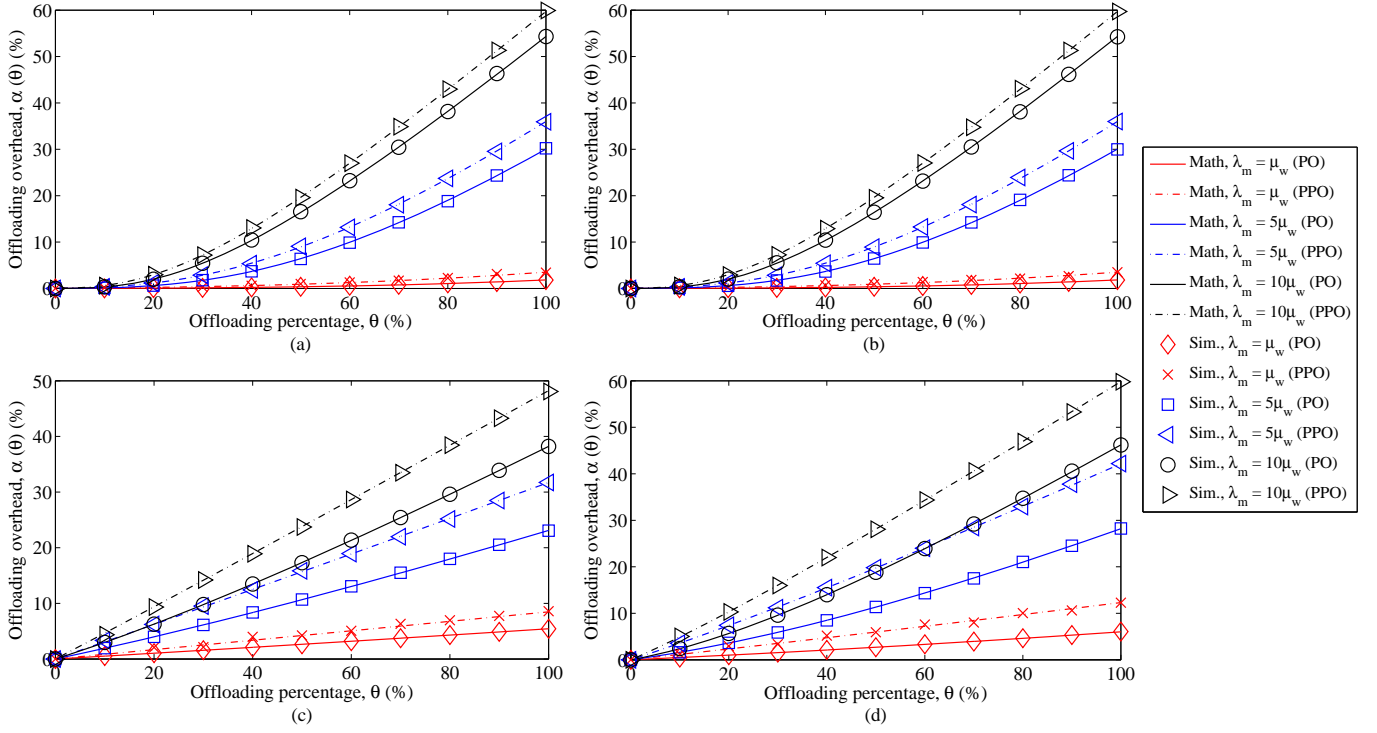


Fig. 8. Offloading overhead $\alpha^{PO}(\theta)$ and $\alpha^{PPO}(\theta)$ with different re-routable flow arrival rate λ_m and offloading percentage θ .

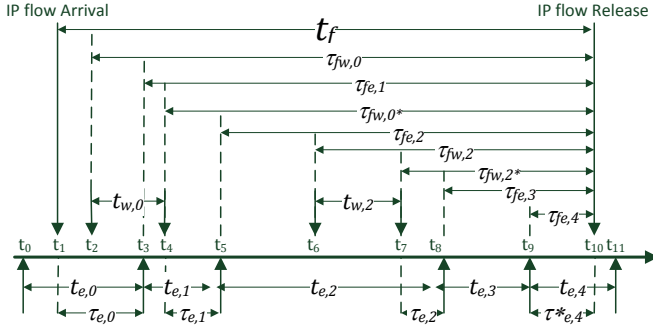


Fig. 9. Event diagram for an IP flow arrival.

Correspondingly, the pdf of the residence time in any WLAN AN cell $f_w(t)$ is expressed as:

$$\begin{aligned}
 f_w(t) &= \int_{\tau_{sw,i}=t}^{\infty} f_s(\tau_{sw,i}) f_{wr}(t) d\tau_{sw,i} \\
 &+ \int_{t_{wr}=t}^{\infty} f_{wr}(t_{wr}) f_s(t) dt_{wr} \\
 &= (\mu + \eta_w) e^{-(\mu + \eta_w)}
 \end{aligned} \quad (11)$$

with a mean value

$$\begin{aligned}
 \frac{1}{\mu_w} &= E[t_w] = \int_{t=0}^{\infty} t f_w(t) dt \\
 &= \int_{t=0}^{\infty} t (\mu + \eta_w) e^{-(\mu + \eta_w)} dt = 1/(\mu + \eta_w).
 \end{aligned} \quad (12)$$

Consider now the case when a flow is re-directed. The residual time of the residence time ($\tau_{e,1}$ and $\tau_{e,2}$) in the E-UTRAN cell has the same distribution and the mean value as the residence time $t_{e,i}$. Based on (10) and (12), we have

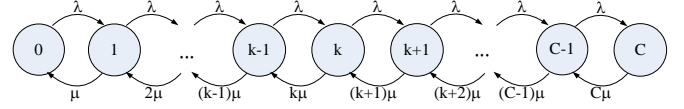


Fig. 10. State transition diagram for the considered wireless network.

$\mu_w = \mu + \eta_w$ for the WLAN cell, and $\mu_e = \mu + \eta_e$ for the E-UTRAN cell.

B. Derivation of $P_b^{PO}(\theta)$ and $P_b^{PPO}(\theta)$

As discussed above, the E-UTRAN or WLAN will block an IP flow only if all the resource units they have are occupied. Based on the *Poisson arrivals see time average (PASTA)* property [25], the Poisson arrival can see the probability the same as time average. Therefore, to find out the blocking probability, we need to derive the probability that all the resource units in the network are occupied on time average. Given the arrival and service rates as λ and μ respectively and considering that the network has capacity C and no line is allowed to form, this network can be modeled as an $M/M/C/C$ system.

Fig. 10 illustrates the state transition diagram of the stochastic process where state k represents that there are k IP flows in the system. The state transition probability is derived as:

$$\pi_k = \pi_0 \left[\frac{\lambda^k}{(k!) \mu^k} \right], \quad \pi_0 = \left[1 + \sum_{j=1}^C \frac{\lambda^j}{(j!) \mu^j} \right]^{-1}, \quad \text{for } 0 < k \leq C. \quad (13)$$

From Erlang's loss formula [25], we obtain the blocking probability as follows:

$$p_b = \pi_0 \left[\frac{\lambda^C}{(C!) \mu^C} \right], \text{ where } \pi_0 = \left[1 + \sum_{j=1}^C \frac{\lambda^j}{(j!) \mu^j} \right]^{-1}. \quad (14)$$

To find out the blocking probability for different types of users in the E-UTRAN or WLAN AN, we need to know λ , μ , and C in each network. Let μ_w and μ_e be the service rate in the WLAN AN and the E-UTRAN, respectively. For PO, the arrivals in the E-UTRAN is composed of the NRR traffic, the RR traffic staying in the E-UTRAN, and the RR traffic offloaded to the WLAN AN but blocked by the WLAN AN. Thus, the arrival rate in the E-UTRAN becomes $\lambda_e + (1 - \theta)\lambda_m + p_{bw}^{PO}\theta\lambda_m$. Based on (14), p_{be}^{PO} is derived as follows:

$$\begin{aligned} p_{be}^{PO} &= \left\{ \frac{[\lambda_e + (1 - \theta)\lambda_m + p_{bw}^{PO}\theta\lambda_m]^{C_e}}{(C_e!) \mu_e^{C_e}} \right\} \\ &\times \left\{ 1 + \sum_{j=1}^{C_e} \frac{[\lambda_e + (1 - \theta)\lambda_m + p_{bw}^{PO}\theta\lambda_m]^j}{(j!) \mu_e^j} \right\}^{-1} \\ &= \left\{ \frac{[\lambda_e + (1 - \theta)\lambda_m + p_{bw}^{PO}\theta\lambda_m]^{C_e}}{(C_e!) \mu_e^{C_e}} \right\} \\ &\times \left\{ \sum_{j=0}^{C_e} \frac{[\lambda_e + (1 - \theta)\lambda_m + p_{bw}^{PO}\theta\lambda_m]^j}{(j!) \mu_e^j} \right\}^{-1}. \quad (15) \end{aligned}$$

Similarly, the arrivals to the WLAN AN consist of the arriving NRR traffic to the WLAN and the offloaded RR traffic from the E-UTRAN. Thus, the arrival rate in WLAN becomes $\lambda_w + \theta\lambda_m$. Based on (14), the blocking probability for PO in the WLAN AN is:

$$\begin{aligned} p_{bw}^{PO} &= \left[\frac{(\lambda_w + \theta\lambda_m)^{C_w}}{(C_w!) \mu_w^{C_w}} \right] \left[1 + \sum_{j=1}^{C_w} \frac{(\lambda_w + \theta\lambda_m)^j}{(j!) \mu_w^j} \right]^{-1} \\ &= \left[\frac{(\lambda_w + \theta\lambda_m)^{C_w}}{(C_w!) \mu_w^{C_w}} \right] \left[\sum_{j=0}^{C_w} \frac{(\lambda_w + \theta\lambda_m)^j}{(j!) \mu_w^j} \right]^{-1}. \quad (16) \end{aligned}$$

Recall that $\mu_w = \mu + \eta_w$ and $\mu_e = \mu + \eta_e$. Inserting (15) and (16) into (1) yields:

$$P_b^{PO}(\theta) = \frac{p_{bw}^{PO}\lambda_w + p_{bw}^{PO}p_{be}^{PO}\theta\lambda_m + p_{be}^{PO}\lambda_e + p_{be}^{PO}(1 - \theta)\lambda_m}{\lambda_w + \lambda_m + \lambda_e}. \quad (17)$$

For PPO, the arrivals to the E-UTRAN consist of the NRR IP flows to the E-UTRAN, the RR IP flows, and the transferred IP flows from the WLAN AN to the E-UTRAN (at a rate of λ_{tr}). The arriving re-routable IP flows can either access the E-UTRAN at first or be offloaded to the WLAN AN (note that they could also be blocked). Thus, the arrival rates in the E-UTRAN are λ_e , $(1 - \theta)\lambda_m$, $\theta\lambda_m p_b^{PPO}$, and λ_{tr} , respectively.

Accordingly, p_{be}^{PPO} is derived as follows:

$$\begin{aligned} p_{be}^{PPO} &= \left\{ \frac{[\lambda_e + (1 - \theta)\lambda_m + p_{bw}^{PPO}\theta\lambda_m + \lambda_{tr}]^{C_e}}{(C_e!) \mu_e^{C_e}} \right\} \\ &\times \left\{ \sum_{j=0}^{C_e} \frac{[\lambda_e + (1 - \theta)\lambda_m + p_{bw}^{PPO}\theta\lambda_m + \lambda_{tr}]^j}{(j!) \mu_e^j} \right\}^{-1}. \quad (18) \end{aligned}$$

To increase the total resource utilization and to avoid congestion in the WLAN AN, PPO lets an arriving NRR IP flow preempt one of the RR IP flows when all the resource units in the WLAN AN are occupied, and transfers the preempted IP flow back to the E-UTRAN. Thus, we obtain p_{bh}^{PPO} as:

$$\begin{aligned} p_{bh}^{PPO} &= \left[\frac{\lambda_w^{C_w}}{(C_w!) \mu_w^{C_w}} \right] \left[1 + \sum_{j=1}^{C_w} \frac{\lambda_w^j}{(j!) \mu_w^j} \right]^{-1} \\ &= \left[\frac{\lambda_w^{C_w}}{(C_w!) \mu_w^{C_w}} \right] \left[\sum_{j=0}^{C_w} \frac{\lambda_w^j}{(j!) \mu_w^j} \right]^{-1}. \quad (19) \end{aligned}$$

The IP flows arriving to the WLAN AN include both the arriving re-routable IP flows and the offloaded flows at a total rate of $\lambda_w + \theta\lambda_m$. Thus, we have p_{bw}^{PPO} as:

$$\begin{aligned} p_{bw}^{PPO} &= \left[\frac{(\lambda_w + \theta\lambda_m)^{C_w}}{(C_w!) \mu_w^{C_w}} \right] \left[1 + \sum_{j=1}^{C_w} \frac{(\lambda_w + \theta\lambda_m)^j}{(j!) \mu_w^j} \right]^{-1} \\ &= \left[\frac{(\lambda_w + \theta\lambda_m)^{C_w}}{(C_w!) \mu_w^{C_w}} \right] \left[\sum_{j=0}^{C_w} \frac{(\lambda_w + \theta\lambda_m)^j}{(j!) \mu_w^j} \right]^{-1}. \quad (20) \end{aligned}$$

As shown in Fig. 9, IP flow transferring may have two cases: (1) IP flows handed over to the original E-UTRAN (occurred at t_7 in Fig. 9); and (2) IP flows handed over to the next E-UTRAN cell (occurred at t_4 in Fig. 9). Transferring happens when all the resources in the WLAN AN are occupied and at least one ongoing flow belongs to the NRR service type. Thus, the arrival rate of the transferred IP flows λ_{tr} is derived as follows:

$$\lambda_{tr} = p_{bw}^{PPO}\lambda_w - p_{bh}^{PPO}\lambda_w = (p_{bw}^{PPO} - p_{bh}^{PPO})\lambda_w. \quad (21)$$

From (19) and (20), (21) is rewritten as:

$$\begin{aligned} \lambda_{tr} &= (p_{bw}^{PPO} - p_{bh}^{PPO})\lambda_w \\ &= \left\{ \left[\frac{(\lambda_w + \theta\lambda_m)^{C_w}}{(C_w!) \mu_w^{C_w}} \right] \left[\sum_{j=0}^{C_w} \frac{(\lambda_w + \theta\lambda_m)^j}{(j!) \mu_w^j} \right]^{-1} \right. \\ &\quad \left. - \left[\frac{\lambda_w^{C_w}}{(C_w!) \mu_w^{C_w}} \right] \left[\sum_{j=0}^{C_w} \frac{\lambda_w^j}{(j!) \mu_w^j} \right]^{-1} \right\} \lambda_w. \quad (22) \end{aligned}$$

After obtaining λ_{tr} , p_{be}^{PPO} can be calculated from (18). However, the transfer of an ongoing re-routable flow may fail if there are not enough available resource units in the E-UTRAN. The failure probability can be derived as:

$$\begin{aligned} p_f^{PPO} &= p_{f|transfer}^{PPO} Pr\{\text{data transfer happened}\} \\ &= p_{f|transfer}^{PPO} \frac{(p^{PPO} p_{ob} - p^{PPO} p_{oh})\lambda_w}{\theta\lambda_m}, \quad (23) \end{aligned}$$

and it is equivalent to p_{be}^{PPO} as:

$$p_{f|transfer}^{PPO} = p_{be}^{PPO}. \quad (24)$$

Based on (18) and (24), (23) can be rewritten as:

$$\begin{aligned} p_f^{PPO} &= p_{f|transfer}^{PPO} \frac{(p_{bw}^{PPO} - p_{bh}^{PPO})\lambda_w}{\theta\lambda_m} \\ &= p_{be}^{PPO} \frac{(p_{bw}^{PPO} - p_{bh}^{PPO})\lambda_w}{\theta\lambda_m}. \end{aligned} \quad (25)$$

Finally, based on (19), (20), (18), (25), $\mu_w = \mu + \eta_w$, and $\mu_e = \mu + \eta_e$, (2) can be rewritten as:

$$\begin{aligned} P_b^{PPO}(\theta) &= \frac{p_{bh}^{PPO}\lambda_w + (p_f^{PPO} + p_{bw}^{PPO}p_{be}^{PPO})\theta\lambda_m}{\lambda_w + \lambda_m + \lambda_e} \\ &+ \frac{p_{be}^{PPO}\lambda_e + p_{be}^{PPO}(1-\theta)\lambda_m}{\lambda_w + \lambda_m + \lambda_e}. \end{aligned} \quad (26)$$

C. Derivation of $U^{PO}(\theta)$ and $U^{PPO}(\theta)$

Consider the $M/M/C/C$ system with arrival rate λ , service rate μ and resource capacity C resource units, respectively. From [25], the resource utilization can be derived as:

$$\rho = \frac{\lambda}{C\mu}. \quad (27)$$

Considering the fact that the blocked arrivals cannot join the system again, we need to adjust the arrival rate based on the finite queue. From the PASTA property [25], the effective arrival rate seen by the server becomes $\lambda(1-p_b)$ where p_b is the blocking probability. Denote such an arrival rate as λ_{eff} . The effective resource utilization in our model, which is obtained based on (27), is expressed as:

$$\rho_{eff} = \frac{\lambda_{eff}}{C\mu} = \frac{\lambda(1-p_b)}{C\mu}. \quad (28)$$

Based on (3) and (28), the RAN network resource utilization for PO is expressed as:

$$U^{PO}(\theta) = \frac{C_e\rho_e^{PO} + C_w\rho_w^{PO}}{C_e + C_w}, \quad (29)$$

where ρ_e^{PO} and ρ_w^{PO} stand for the effective network utilization for the E-UTRAN and the WLAN AN, respectively.

Applying $\mu_w = \mu + \eta_w$, $\mu_e = \mu + \eta_e$, (15), and (16) into (29), we have:

$$\begin{aligned} U^{PO}(\theta) &= \frac{[\lambda_e + (1-\theta)\lambda_m + p_{bw}^{PO}\theta\lambda_m](1-p_{be}^{PO})}{\mu_e(C_e + C_w)} \\ &+ \frac{(\lambda_w + \theta\lambda_m)(1-p_{bw}^{PO})}{\mu_w(C_e + C_w)}. \end{aligned} \quad (30)$$

Similarly, based on (3) and (28), the resource utilization for

PPO is obtained as:

$$\begin{aligned} U^{PPO}(\theta) &= \frac{C_e\rho_e^{PPO} + C_w\rho_w^{PPO}}{C_e + C_w} \\ &= \frac{[\lambda_e + (1-\theta)\lambda_m + p_{bw}^{PPO}\theta\lambda_m](1-p_{be}^{PPO})}{(C_e + C_w)(\mu + \eta_e)} \\ &+ \frac{[(p_{bw}^{PPO} - p_{bh}^{PPO})\lambda_w](1-p_{be}^{PPO})}{(C_e + C_w)(\mu + \eta_e)} \\ &+ \frac{(\lambda_w + \theta\lambda_m)(1-p_{bw}^{PPO})}{(C_e + C_w)(\mu + \eta_w)}, \end{aligned} \quad (31)$$

where ρ_e^{PPO} and ρ_w^{PPO} stand for the effective network utilization for the E-UTRAN and the WLAN AN, respectively.

Applying (20) and (18) into (31), we obtain:

$$\begin{aligned} U^{PPO}(\theta) &= \frac{[\lambda_e + (1-\theta)\lambda_m + p_{bw}^{PPO}\theta\lambda_m](1-p_{be}^{PPO})}{(C_e + C_w)(\mu + \eta_e)} \\ &+ \frac{(p_{bw}^{PPO} - p_{bh}^{PPO})\lambda_w(1-p_{be}^{PPO})}{(C_e + C_w)(\mu + \eta_e)} \\ &+ \frac{(\lambda_w + \theta\lambda_m)(1-p_{bw}^{PPO})}{(C_e + C_w)(\mu + \eta_w)}. \end{aligned} \quad (32)$$

D. Derivation of $\alpha^{PO}(\theta)$ and $\alpha^{PPO}(\theta)$

For the calculation of offloading overhead shown in (7) and (8), all the necessary parameters have been derived above. Thus, by keeping $\mu_w = \mu + \eta_w$ and $\mu_e = \mu + \eta_e$ in mind and inserting (18) into (7), the offloading overhead for PO is shown as following:

$$\begin{aligned} \alpha^{PO}(\theta) &= \frac{p_{bw}^{PO}(1-\theta)\lambda_m}{\lambda_w + \lambda_m} \\ &= \left[\frac{(\lambda_w + \theta\lambda_m)^{C_w}}{(C_w!)(\mu + \eta_w)^{C_w}} \right] \left[\frac{(1-\theta)\lambda_m}{\lambda_w + \lambda_m} \right] \left[\sum_{j=0}^{C_w} \frac{(\lambda_w + \theta\lambda_m)^j}{(j!)(\mu + \eta_w)^j} \right]^{-1}. \end{aligned} \quad (33)$$

When an IP flow is re-directed to the E-UTRAN under PPO, it also causes additional signaling overhead. Applying $\mu_w = \mu + \eta_w$, $\mu_e = \mu + \eta_e$ and inserting (19), (20), and (18) into (8), the offloading overhead for PPO is obtained as:

$$\begin{aligned} \alpha^{PPO}(\theta) &= \frac{p_{bw}^{PPO}(1-\theta)\lambda_m + (p_{bw}^{PPO} - p_{bh}^{PPO})\lambda_w}{\lambda_w + \lambda_m} \\ &= \left[\frac{(\lambda_w + \theta\lambda_m)^{C_w}}{(C_w!)(\mu + \eta_w)^{C_w}} \right] \left[\sum_{j=0}^{C_w} \frac{(\lambda_w + \theta\lambda_m)^j}{(j!)(\mu + \eta_w)^j} \right]^{-1} \left[\frac{(1-\theta)\lambda_m}{\lambda_w + \lambda_m} \right] \\ &+ \left[\frac{\lambda_w}{\lambda_w + \lambda_m} \right] \left\{ \left[\frac{(\lambda_w + \theta\lambda_m)^{C_w}}{(C_w!)(\mu + \eta_w)^{C_w}} \right] \left[\sum_{j=0}^{C_w} \frac{(\lambda_w + \theta\lambda_m)^j}{(j!)(\mu + \eta_w)^j} \right]^{-1} \right. \\ &\quad \left. - \left[\frac{\lambda_w^{C_w}}{(C_w!)(\mu + \eta_w)^{C_w}} \right] \left[\sum_{j=0}^{C_w} \frac{\lambda_w^j}{(j!)(\mu + \eta_w)^j} \right]^{-1} \right\}. \end{aligned} \quad (34)$$

ACKNOWLEDGMENT

The authors are very grateful for the constructive comments from the anonymous reviewers that have significantly improved the quality of this paper. This work was supported in part by the Ministry of Science and Technology of Taiwan under grant numbers MOST 106-2221-E-009-046-MY3,

106-2221-E-009-047-MY3, 106-2218-E-009-016, and the EU Erasmus+ Global Program (Project Number: 2017-1-NO01-KA107-034077) for cooperation between National Chiao Tung University (NCTU), Taiwan and University of Agder (UiA), Norway.

REFERENCES

- [1] 3GPP TS 23.261 V14.0.0, *Universal mobile telecommunications system (UMTS); LTE; IP flow mobility and seamless wireless local area network (WLAN) offload*, Std., Mar. 2017.
- [2] S. Ranjan, N. Akhtar, M. Mehta, and A. Karandikar, "User-based integrated offloading approach for 3GPP LTE-WLAN network," in *Proc. 20th National Conference on Communications (NCC)*, Feb. 2014, pp. 1–6.
- [3] M. Amani, T. Mahmoodi, M. Tatipamula, and H. Aghvami, "Programmable policies for data offloading in LTE network," in *Proc. IEEE ICC*, Jun. 2014, pp. 3154–3159.
- [4] F. Malandrino, C. Casetti, and C.-F. Chiasserini, "LTE offloading: When 3GPP policies are just enough," in *Proc. IEEE/IFIP WONS*, Apr. 2014, pp. 1–8.
- [5] L. Wang and G.-S. Kuo, "Mathematical modeling for network selection in heterogeneous wireless networks - a tutorial," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 271–292, 1st Quart. 2013.
- [6] K. Khawam, M. Ibrahim, J. Cohen, S. Lahoud, and S. Tohme, "Individual vs. global radio resource management in a hybrid broadband network," in *Proc. IEEE ICC*, Jun. 2011, pp. 1–6.
- [7] S. Singh, H. Dhillon, and J. Andrews, "Offloading in heterogeneous networks: Modeling, analysis, and design insights," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2484–2497, May 2013.
- [8] B. H. Jung, N.-O. Song, and D. K. Sung, "A network-assisted user-centric WiFi-offloading model for maximizing per-user throughput in a heterogeneous network," *IEEE Trans. Veh. Technol.*, vol. 63, no. 4, pp. 1940–1945, May 2014.
- [9] Q. Chen, G. Yu, H. Shan, A. Maaref, G. Y. Li, and A. Huang, "Cellular meets WiFi: Traffic offloading or resource sharing?" *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3354–3367, May 2016.
- [10] Q. Fan, H. Lu, P. Hong, and Z. Zhu, "Throughput-power tradeoff association for user equipment in WLAN/cellular integrated networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3462–3474, Apr. 2017.
- [11] A. Balasubramanian, R. Mahajan, and A. Venkataramani, "Augmenting mobile 3G using WiFi," in *Proc. ACM MobiSys*, 2010, pp. 209–222.
- [12] K. Lee, J. Lee, Y. Yi, I. Rhee, and S. Chong, "Mobile data offloading: How much can WiFi deliver?" *IEEE/ACM Trans. Netw.*, vol. 21, no. 2, pp. 536–550, 2013.
- [13] O. B. Yetim and M. Martonosi, "Dynamic adaptive techniques for learning application delay tolerance for mobile data offloading," in *Proc. IEEE INFOCOM*, 2015, pp. 1885–1893.
- [14] Y. Li, J. Zhang, X. Gan, L. Fu, H. Yu, and X. Wang, "A contract-based incentive mechanism for delayed traffic offloading in cellular networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 8, pp. 5314–5327, Aug 2016.
- [15] F. Mehmeti and T. Spyropoulos, "Performance modeling, analysis, and optimization of delayed mobile data offloading for mobile users," *IEEE/ACM Trans. Netw.*, vol. 25, no. 1, pp. 550–564, Feb 2017.
- [16] IETF RFC 5555, *Mobile IPv6 support for dual stack hosts and routers (DSMIPv6)*, Std., Jun. 2009.
- [17] 3GPP TS 24.312 V 14.1.0, *Universal mobile telecommunications system(UMTS); LTE; Access network discovery and selection function (ANDSF) management object (MO)*, Std., Jun. 2017.
- [18] 3GPP TS 23.402 V15.1.0, *Universal mobile telecommunications system (UMTS); LTE; Architecture enhancements for non-3GPP accesses*, Std., Sep. 2017.
- [19] IEEE Std 802.11-2012, *Part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications*, IEEE Std., May 2015.
- [20] K.-H. Chen and J.-C. Chen, "Handoff failure analysis of adaptive keep-alive interval (AKI) in 3GPP generic access network (GAN)," *IEEE Trans. Wireless Commun.*, vol. 10, no. 12, pp. 4226–4237, Dec. 2011.
- [21] Y. B. Lin, R. H. Liou, Y. C. Sung, and P. C. Cheng, "Performance evaluation of LTE eSRVCC with limited access transfers," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2402–2411, May 2014.
- [22] P. Si, H. Yue, Y. Zhang, and Y. Fang, "Spectrum management for proactive video caching in information-centric cognitive radio networks," *IEEE J. Sel. Area. Comm.*, vol. 34, no. 8, pp. 2247–2259, Aug. 2016.
- [23] The network simulator - ns-2. [Online]. Available: <http://www.isi.edu/nsnam/ns/>
- [24] 3GPP TS 23.161 V14.0.0, *Network-based IP flow mobility (NBIFOM); Stage 2 (Release 14)*, Std., Mar. 2017.
- [25] S. M. Ross, *Introduction to Probability Models*, 10th ed. Academic press, 2009.



Yi Ren (S'08-M'13) received the Ph.D. degree in information communication and technology from the University of Agder, Norway, in 2012. He was with the Department of Computer Science, National Chiao Tung University (NCTU), Hsinchu, Taiwan, as a Postdoctoral Fellow, an Assistant Research Fellow, and an Adjunct Assistant Professor from 2012 to 2017. He is currently a Lecturer in the School of Computing Science at University of East Anglia (UEA), Norwich, U.K. His current research interests include Internet of Things (IoT) and 5G Mobile Technology: security, performance analysis, protocol design, radio resource allocation, mobile edge computing, WiFi & Bluetooth Technology, 3GPP, LTE, Software Defined Networking (SDN), Network Function Virtualization (NFV), etc. He received the Best Paper Award in IEEE MDM 2012.



Chih-Wei Tung received the master degree in Institute of Computer Science and Engineering from the National Chiao Tung University (NCTU), Hsinchu, Taiwan in 2015. He got his bachelor degree in the Department of Computer Science in NCTU. He is currently a platform engineer in Mozilla Berlin office.



Jyh-Cheng Chen (S'96-M'99-SM'04-F'12) received the Ph.D. degree from the State University of New York at Buffalo in 1998.

He was a Research Scientist with Bellcore / Telcordia Technologies, Morristown, NJ, USA, from 1998 to 2001, and a Senior Scientist with Telcordia Technologies, Piscataway, NJ, USA, from 2008 to 2010. He was with the Department of Computer Science, National Tsing Hua University (NTHU), Hsinchu, Taiwan, as an Assistant Professor, an Associate Professor, and a Full Professor from 2001 to 2008. He was also the Director of the Institute of Network Engineering with National Chiao Tung University (NCTU), Hsinchu, from 2011 to 2014. He has been a Faculty Member with NCTU since 2010. He is currently a Chair Professor with the Department of Computer Science, NCTU. He is also serving as the Convener, Computer Science Program, Ministry of Science and Technology, Taiwan.

Dr. Chen received numerous awards, including the Outstanding Teaching Awards from both NCTU and NTHU, the Outstanding I.T. Elite Award, Taiwan, the Mentor of Merit Award from NCTU, the K. T. Li Breakthrough Award from the Institute of Information and Computing Machinery, the Outstanding Professor of Electrical Engineering from the Chinese Institute of Electrical Engineering, the Outstanding Research Award from the Ministry of Science and Technology, the best paper award for Young Scholars from the IEEE Communications Society Taipei and Tainan Chapters, and the IEEE Information Theory Society Taipei Chapter, and the Telcordia CEO Award. He is a Distinguished Member of the Association for Computing Machinery (ACM). He was a member of the Fellows Evaluation Committee, IEEE Computer Society.



Frank Y. Li received the Ph.D. degree from the Department of Telematics (now Department of Information Security and Communication Technology), Norwegian University of Science and Technology (NTNU), Trondheim, Norway. He worked as a Senior Researcher at UniK—University Graduate Center (now Department of Technology Systems), University of Oslo, Norway before joining the Department of Information and Communication Technology, University of Agder (UiA), Agder, Norway, in August 2007 as an Associate Professor and then

a Full Professor. During the past few years, he has been an active participant in several Norwegian and EU research projects. He is listed as a Lead Scientist by the European Commission DG RTD Unit A.03—Evaluation and Monitoring of Programmes in Nov. 2007. Dr. Li's research interests include MAC mechanisms and routing protocols in 5G mobile systems and wireless networks and the Internet of Things; mesh and ad hoc networks; wireless sensor networks; D2D communication; cooperative communication; cognitive radio networks; green wireless communications; reliability in wireless networks; QoS, resource management and traffic engineering in wired and wireless IP-based networks; analysis, simulation and performance evaluation of communication protocols and networks.