Telecommun Syst (2016) 61:675–694
DOI 10.1007/s11235-015-0061-2

The performance of network-controlled mobile data offloading from LTE to WiFi networks

Desta Haileselassie Hagos[1]

Published online: 13 May 2015

Abstract Global mobile data traffic consumption and usage continues to increase rapidly leading to congested networks. Currently, cellular networks are overloaded with mobile data traffic due to the rapid growth of mobile broadband subscriptions and the increasing popularity of diversified applications for smartphones with multiple wireless interfaces and the flat-rate pricing model of cellular networks. One possible practical solution to alleviate this problem is the offloading of mobile data traffic from the primary access technology to the WiFi infrastructure to gain extra capacity and improve the overall network performance. As the strategy what and when to offload data is non-trivial, it is of vital importance to develop novel algorithms to guide this process. This paper addresses solutions for network-controlled WiFi offloading in Long Term Evolution (LTE) cellular networks when performance needs exceed the capability of the LTE access. It then compares the performance of each access technology using different network performance metrics. In detail, an optimized signal-to-noise ratio-threshold based handover solution and extension to the 3rd Generation Partnership Project standard for Access Network Discovery and Selection Function (ANDSF) framework for WiFi offloading is proposed. Our simulation results have shown that ANDSF discovery can be used to control the amount of WiFi offloading.

Keywords LTE · IEEE 802.11 · ANDSF · Access selection · WiFi access points · Offloading · Multi-access · Seamless handover

✉ Desta Haileselassie Hagos
desta161921@gmail.com; desta@unik.no;
destahh@ifi.uio.no

1 UNIK - University Graduate Centre, Oslo, Norway

1 Introduction

The adoption of mobile communication technology has evolved rapidly in recent years due to the increasing demands for higher data rates and higher quality services. Currently mobile users are able to download content (be it audio, video and data) from the Internet (e.g., YouTube) or generate content to upload to remote mobile cloud services (e.g., Facebook Mobile, Google plus, etc.) with their smartphones. The combination of smartphones equipped with high computing powers, tablets, netbooks and cellular mobile networks are rapidly growing in very large numbers and as a result, this has created an exceptional demand for ubiquitous connectivity and quality of rich digital content and applications. Mobile broadband traffic communicated over cellular networks has seen exponential increases, and a recent report from Ericsson which is a representative base for calculating world total data traffic in cellular networks predicts mobile data traffic to grow tenfold between 2011 and 2016 [1].

The dramatic increase of mobile data traffic is a major concern for network operators with limited radio spectrum. Fortunately, there is a real opportunity of offloading the data traffic to other networks. Here, offloading or Mobile data/cellular traffic offloading refers to using alternative network technologies for delivering data originally targeted for, e.g., a cellular network when it becomes saturated. It helps to ensure optimal usage of available radio resources and load-balancing among available radio accesses. To meet the requirements of future data-rich applications with improved multimedia, future wireless networks are expected to combine multiple access technologies. Recently, some cellular broadband network operators, including AT&T, T-Mobile, Vodafone, and Orange, are utilizing WiFi networks as an alternative access network technology worldwide [2]. At the same time, there are already various data traffic offloading solutions and applications proposed from the industry. Projection data released in the Cisco Visual Networking Index [3] white paper, estimates that over 800 million terabytes of mobile data traffic can be offloaded which would be 72 % of the total data traffic in 2015. It is also reported that global mobile data traffic grew 2.3-fold in the year 2011 [3]. It is also predicted that the total mobile data traffic will be 18-fold between the years 2011 and 2016 [3].

At the technical level, Evolved Packet Core (EPC) provides interworking functionality between 3GPP and non-3GPP (both trusted and non-trusted) accesses according to the 3GPP standards [4,5]. To allow mobile devices to know when, how and where to select an access network discovery support functions and usage scenarios can be found in [5]. The main drivers for the interworking of the accesses are to reduce the load on the cellular network (i.e. offload LTE network) and supplement 3G[1] access coverage. When several radio accesses are available, the assignment and the handover of users among these accesses, session management with the flows and much more become the fundamental problems. While there has been only a limited number of studies so far to mitigate this critical problems, the performance issue has been even more neglected.

In this paper, our main focus is to address how to overcome the mobile network congestion by offloading a portion of the data traffic to complementary access networks, i.e. by using WiFi whenever there is data congestion in the LTE cellular networks. Mobile data offloading has become a main concern for cellular network operators dealing with network congestion. Our offloading strategies are compared to steer WiFi offloading to increase the combined network performance of LTE and WiFi access connected to the core network with at least the baseline case of having all the traffic in LTE. In this paper, we have also motivate the reasons why ANDSF is a 3GPP choice for network selection in heterogeneous networks. Accordingly, we devise and implemented three ANDSF offloading algorithms as per the specified standard [6]. In the 3GPP standard, it is stated that the ANDSF can trigger the radio cell of a User Equipment (UE) for its discovery information using the geographical location of latitude, longitude, and/or altitude (radius) coordinates. We performed extensive simulations to evaluate the algorithms considered. An optimized SNR-threshold based handover solution for WiFi offloading is proposed.

This paper makes the following contributions : (1) We address the growth of mobile data traffic. In particular, we have considered the possible approaches to alleviate the data traffic congestion, i.e. network-controlled offloading traffic to WiFi network to reduce the load on the LTE network. The viable and cost-efficient approach we considered takes advantage of exploiting heterogeneous access networks. The main advantages include, an improved user experience, the ability to actively offload data traffic from the cellular network, and to ensure that operator-hosted services are available over the WiFi network. (2) We devise and implemented novel offloading algorithms that decides when to move flow(s) between LTE and WiFi access networks. (3) Comparing the performance of the considered offloading algorithms. These offloading algorithms are evaluated and compared to steer WiFi offloading to increase the combined network performance of LTE and WiFi access technologies connected to the EPC with at least the baseline case of having all the traffic in LTE. (4) Understanding how these offloading algorithms fit into the existing standards is part of the scope of thi work. (5) We performed an extensive performance evaluation simulations to evaluate the algorithms considered covering all of the proposed scenarios. An optimized SNR-threshold based handover solution for WiFi offloading is proposed.

The rest of this paper is structured as follows. Section 2 presents a target scenario on how and when data traffic offloading can be applied. Section 3 presents the considered IP flow mobility use cases. Section 4 presents the network access selection methods used. Section 5 provides an in-depth explanation and summary of the models and assumptions used. Section 6 presents the numerical results. Section 7 addresses related work in the area of data traffic offloading solutions. Section 8 describes the proposed optimized handover solution and proposes how the 3GPP standard could be improved. We finally draw the most important conclusions of the paper and point out future work in Sect. 9.

2 Target scenario

We assume a 3GPP network operator controlled scenario as it is shown in Fig. 1. Our overall assumption is that we have a network operator in charge of both the LTE and WiFi networks. This helps the network operator to have control over WiFi traffic and to ensure a better customer experience and high lever of mobile network performance across the available networks. This means that, for example, UE's IP address allocation, access to general IP services as well as network features like security, charging, Quality of Service (QoS) and policy control can be made independent of the access technology. And therefore migration between LTE and WiFi is easily possible. Data session managements are seamlessly handed off by the Packet Data Network Gateway (PDN GW).[2]

[1] Throughout this paper, we use '3G' to refer to a cellular network (For example, LTE).

[2] PDN GW is the point of interconnect between the EPC and the external IP networks. These networks are called Packet Data Network (PDN). The PDN GW routes packets to and from the PDNs and it also performs various functions such as IP address or policy control and charging.

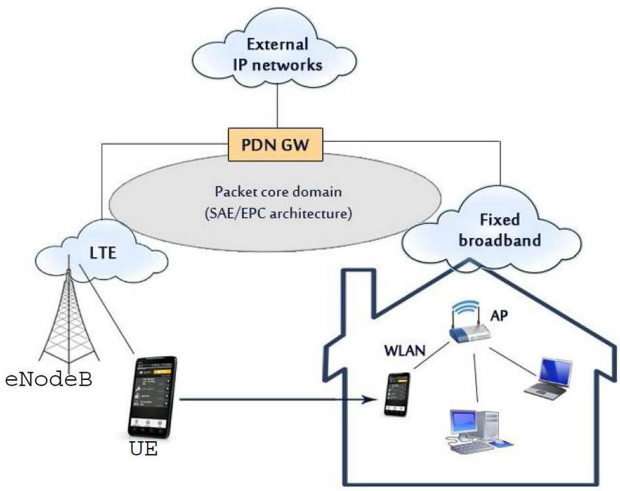

Fig. 1 Target scenario of the work

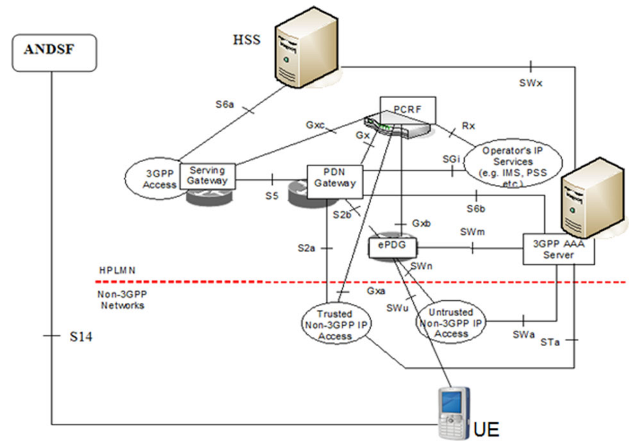

Fig. 2 ANDSF integration to the core network

Figure 1 shows the baseline scenario when a user moves from LTE network to a WiFi network coverage. We can think of this scenario as an IP flow mobility service: we carry a UE which can access among other technologies, LTE and WiFi. We are connected to LTE network and move indoors, into coffee shops or big shopping malls. There we have a fixed broadband connection connected to a IEEE 802.11-capable Access Point (AP). Depending on preferences, the UE in this situation may switch access from LTE to WLAN which is the main objective of this paper. Whenever WiFi access network is available nearby, the operator can choose to offload some or all of its traffic depending on the QoS requirement through a WiFi AP.[3] Selectively offloading traffic to approved WiFi networks gives mobile operators an opportunity to increase their total network capacity to meet rising traffic demands and a way to extend network coverage and capacity to WiFi networks.

To enable such an approach ANDSF is a suitable basis as it is a 3GPP approach for controlling handover operation between 3GPP and non-3GPP access networks. Release-8 of 3GPP [5] has specified the ANDSF framework through which the network operator can provide a list of preferred access networks with inter-system mobility policies. Figure 2 shows how the ANDSF is integrated in to the core network. As a UE moves across a heterogeneous network environment, it has to discover other radio accesses available in its vicinity. For example, a UE using 3GPP radio access needs to discover when WiFi access becomes available and possibly trigger a handover based on a predefined operator policies, or when the radio signal from its serving 3GPP cell starts to get weak. When a user changes connection to another network leaving the service area of its current serving network, handover needs to be executed seamlessly such that ongoing service sessions are not interrupted. Hence, the purpose of the ANDSF is to assist user devices to discover access networks in their vicinity and to provide rules so as to prioritize and manage connections to all networks. Based on operator's policies, the IP flows are routed differently when either the core network is congested or when the current serving cell is overloaded. Different services with different characteristics in terms of QoS requirements and bandwidth, for example a web browsing session and non-conversational video streaming session will be offloaded to the non-3GPP WiFi network. This helps to balance the load and relieve the traffic issue of the 3G access network usage and to guarantee optimal usage of the available radio accesses. It also increases the end-user throughput for IP flows with high throughput requirements.

As we know, in a very simplistic case, the mobile terminal can discover neighboring cells with no assistance from the network by periodically conducting a radio scanning in the background. Although this is very simple and does not require any modifications in the network. However, some of the major problems of such a simplistic approach are that:

- *Battery consumption can increase considerably, especially when we demand fast discovery.*
- *The information discovered about the neighbor cells is only limited.*
- *The UE needs to have two receivers working in parallel (one dedicated to scanning and another for ongoing communications).*
- *Long delay from that a terminal enters a region where handover would be beneficial until the blind scanning discovers the available access.*

This drives the need for network-assisted access network discovery and this is the main reason why ANDSF is one algorithm evaluated in this work. In addition to the previously stated benefits, ANDSF features the following advantages:

[3] Throughout this paper, we use the word 'AP' to refer to a WiFi AP.

- *Discovering information about neighboring networks.*
- *Dynamic construction of the discovery database, information repository function.*
- *Determining what information to collect and provide to mobiles.*
- *Includes validity conditions to the information provided (i.e. indicate whether the provided information policies are valid or not).*

2.1 Why WiFi as a practical offloading technique?

This paper addresses WiFi offloading as a solution to the exploding future growth of mobile broadband data traffic in the deployed LTE networks thereby using WiFi as an alternative access network technology. The reason why traffic offloading by WiFi is considered to be a viable solution of mobile data traffic explosion and why it is the focus of this work is that because there is a lot of unlicensed WiFi spectrum already existing with very large number of compatible devices in which network operators can make use of without much financial burden in practice. There are already millions of installed WiFi networks mainly in congested areas such as universities/colleges, airports, hotels and city centers and the number is growing very rapidly. WiFi works on the unlicensed spectrum, i.e. there is no monetary cost for it and causes no interference with cellular networks. This helps to simplify the complexity as well as cost of managing and deploying a WiFi network. Network operators as well as users can quickly and easily install WiFi APs with very low costs. In this case, network operators can provide services that take the advantage of WiFi both indoor and outdoor environments and so that it increases revenue and capacity through subscriber retention and increased market share. Therefore, it is very viable to exploit the existing cost efficient heterogeneous access network infrastructures to supplement 3GPP access technology coverage.

Existing major cellular traffic offload solutions [7–9] have shown that it is possible to offload a huge portion of mobile data traffic to WiFi access networks by allowing the users delay their delay-tolerant data (For example, cloud services, software downloads, movies, .etc.), and upload/download data whenever they have a nearby WiFi AP within a predefined delay deadline. In [7], it is predicted that about 60–80 % of mobile data traffic can be reduced when there is about 30 min to 1 h delay for human mobility, and when a delay of around 10 min is allowed for vehicular mobility [8]. In addition to this, it is also predicted in [9] that more than 80 % of news data can be prefetched on a random model. Such an offloading technique is called *delayed WiFi offloading*. In this technique, each data transfer is given a "deadline"? when it must be sent out. It then sends the data piece by piece as a user enters and exits different WiFi networks. However, it if the data is not sent out before predefined deadline, it is finished using the cellular network. There is also another offloading technique called *on-the-spot WiFi offloading* which uses spontaneous connectivity to WiFi and transfers data on the spot. This means, when a user leaves the WiFi network coverage area, offloading immediately ends and unfinished data will then be transferred through the cellular network.

Another two possible approaches to alleviate mobile data traffic congestion but not the focus of this work are deploying bandwidth limit and scaling the network capacity. The other approach is to optimize by changing the policy control on existing networks. However, optimization requires intensive packet inspection and correlation, by isolating the heavy data users. The last approach is scaling the network which involves upgrading the network by building out more towers and base stations. However, this approach demands huge cost for network operators and therefore it is not viable.

3 IP flow mobility: use cases and possible scenarios

One of the promising evolutions of mobile technologies is combining different existing wireless access networks so as to offer access to services while on the move, at any place anytime which is one motivation of Next Generation Networks. These access technologies are integrated to complement each other in terms of coverage area, mobility support, higher bandwidth, and operation cost. From the user's point of view, the main reason behind data traffic offloading is that it allows them to receive the services when they are out of cellular network coverage or else when they want to receive some of the services at a lower cost even when they are within the cellular network coverage. As a result of this, recent mobile devices are integrated with multiple network interfaces and users want to stay connected to the network anytime anywhere. A baseline architecture for IP flow mobility within EPS is specified in [5]. This solution is technically based on Dual-Stack Mobile IPv6 (DSMIPv6) [10] which provides IP address preservation and session continuity when moving IP flows from one access network to the other, and it is applicable to both the 3GPP and non-3GPP infrastructure architectures. A technique which provides proactive secured handover when a mobile user moves between heterogeneous access networks is described in [11]. In heterogeneous access network environments, a network selection that scales for services is very crucial, which is required to achieve seamless mobility, support quality of QoS enhancement and load balancing. And hence, an architecture where application service providers and network service providers define service levels to be used by a mobile node and its user is proposed in [12]. As it is specified in [13], the main goal of the EPC is to provide seamless service continuity for multi-mode terminals as these terminals move from one radio access technology to another.

In this section, we will describe detailed multi-access scenarios for network discovery and selection procedures, where the mobile terminal is connected to the EPC via different access networks simultaneously with service continuity over multiple accesses, sending and receiving different IP flows when the UE is under the coverage of both 3GPP and non-3GPP access networks, and redistribution of IP flows when the non-3GPP access is no longer available. These accesses are interconnected to the EPC via the PDN GW.

3.1 Use case 1

We considered a scenario which represents a seamless IP flow mobility services, with IP flows belonging to different or the same applications being moved seamlessly between a 3GPP and non-3GPP accesses. By seamless IP flow mobility, it means that the experience of using a service is unaffected while being mobile. Such kind of scenario allows the network operator to indicate how the IP flows are routed through the available access networks and to selectively offload some traffic (e.g. best effort traffic) to WiFi network while using LTE network for other traffic (e.g. traffic with specific QoS requirements). For example, on the way home from office, a user might only have 3GPP access. Let us assume that the user is simultaneously accessing different services with different characteristics in terms of QoS requirements and bandwidth, for example a web browsing session and a video telephony call consisting of conversational voice and non-conversational video streaming session. When the user reaches home, his/her device selects non-3GPP access (e.g. domestic WiFi hotspot) and based on his/her personal preferences, requirements of applications, etc., some of his/her currently running services will be switched over to the non-3GPP interface so as to load the balance, to guarantee optimal usage of the available radio access and increase the end-user throughput for IP flows with high throughput requirement. Some of the flows which the user is using may be from the same application. Based on operator's policies, the user's preferences, the characteristics of the application and the accesses, the IP flows are routed differently; as an example, the audio media (conversational voice which is the hard real-time) of the Video Telephony call and the video streaming are routed via 3GPP access, while the soft real-time conversational video (live streaming) of the Video Telephony, the P2P download (best effort) and media file synchronization are routed through the non-3GPP access [14]. In the middle of the IP sessions, the user's device might automatically start a non-real time streaming of FTP file synchronization with a backup server (best effort) via the WiFi access system. Due to the huge amount of traffic, WiFi becomes congested and therefore the non-conversational video streaming session doesn't get the required level of QoS treatment. This initiates the IP flow to move back to the 3GPP access. Later on, when

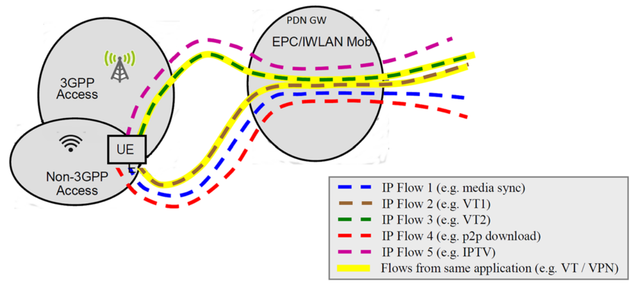

Fig. 3 Routing of different IP flows through different accesses

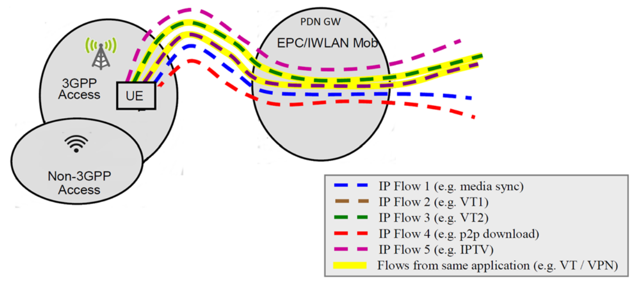

Fig. 4 Switchover of all IP flows from non-3GPP to 3GPP

the FTP file synchronization is done, the non-conversational video streaming session will be moved back to the WiFi network.

The following scenario considers, switchover of all IP flows when the UE goes out of the coverage of the access network. After a while, let's assume that the user moves out of home and loses the WiFi connectivity. Initiated by this event, all the IP flows need to be moved to the 3GPP access since it is the only access available. Figure 4, shows how the IP flows are redistributed when the non-3GPP connectivity is no longer available.

Later on, let's assume that the user goes back home or moves to another area where both the 3GPP and non-3GPP coverage are available. Initiated by this event, the video media of the Video Telephony, the P2P download and the media file synchronization are moved back to the WiFi connectivity and as result of this the scenario depicted in Fig. 3 is restored.

3.2 Use case 2

Let's again assume that the user has an online VoIP session (conversational voice) combined with video (conversational video) session with his friends. During the multimedia session the user browses web (best effort) and occasionally watches video clips (Non-conversational video streaming). Based on the network operator's policy the VoIP flow and conversational video are routed via 3GPP access, while the non-conversational video and best effort IP flows are routed via the non-3GPP. And let's say that his/her device starts ftp

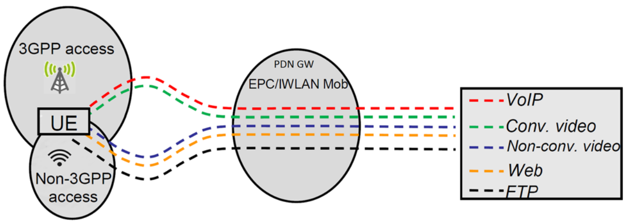

Fig. 5 Splitting of IP flows based on operator's policies

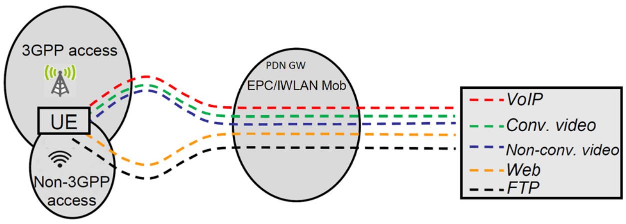

Fig. 6 Movement of one IP flow due to network congestion

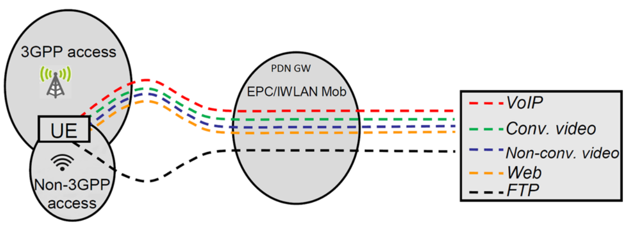

Fig. 7 Further mobility of IP flows due to network congestion

file synchronization with a backup server (best effort) via the WiFi access as it is depicted in Fig. 5.

Because of the FTP file synchronization, the non-3GPP access network becomes congested and the non-conversational video flows are moved back to the 3GPP access network as it is depicted in Fig. 6.

We can also selectively remove an IP flow from the active PDN Connections when the UE is under the coverage of both 3GPP and non-3GPP access networks and has simultaneous active IP sessions via both access systems. In this case, the UE moves all traffic associated with one access to another access and disconnects form one access (for example, due to loss of coverage or by an explicit detach). For example, let's assume that the HTTP server response time for the web browsing (best effort) is detected to have increased; also the best effort web browsing is moved back to the 3GPP access network. Figure 7 shows when we have only the FTP file synchronization is left to non-3GPP access network.

Finally, when the FTP file synchronization completes, the non-conversational video and web browsing are moved back to non-3GPP access as it is shown in Fig. 8.

Based on the IP flow use cases discussed previously, we can distinguish scenarios where the UE is capable of routing different simultaneously active PDN connections connected to the Evolved Packet System (EPS) through different

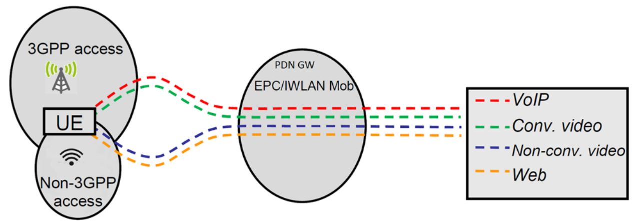

Fig. 8 Sharing of IP flows after network congestion is over

accesses systems the UE can stay simultaneously connected with. Possible scenarios could be, adding an IP flow to an existing one via different accesses, removing an IP flow established via different accesses, IP flow mobility between accesses (when both interfaces are active simultaneously or when only one interface is active) and switchover of all IP flows from one access network to another. The UE selects the access system where to route a specific PDN connection based on the user preferences and the inter-system mobility policies statically pre-configured by the network operator on the UE or provided by the ANDSF entity of the EPS [14]. A more thorough discussion of this is available in [15,16].

4 Network access selection methods

Next generation wireless heterogeneous networks are characterized by the co-existence of multi-access wireless networks utilizing different access networks which complement each other in terms of offered actual throughput and operational costs, e.g., LTE and WiFi. In such networks access selection is a fundamental problem. Access selection principle does not only affect the performance of the overall network, but through its required input parameter set, potentially originating from different access network technologies, also determines what network architecture solutions are called for.

Network access selection may be done on different time scales, based on different parameters, and with different objectives in mind, e.g. to maximize capacity or some form of user quality measure. But in this paper, the main goal is for WiFi offloading when the LTE network is congested or when the current serving cell is overloaded, and access selection principles based on different input parameters are evaluated. The access selection is instantaneous, i.e., the time-scale aspect is neglected and the input parameter set is varied. Generally, an access selection principle can be defined as a function *f*, which is based on its set of input parameters P_i selects an (or set of) access network(s)/decision variable(s) $access_i$ for each user index i:

$$access_i = f(P_i) \tag{1}$$

It is of great interest to limit the set of required input parameters P_i. This is mainly because this set must be made available across different access networks, and potentially across different network operators, which is a challenging task from protocol and an architecture point of view. In this paper, access selection algorithms given in Eqs. 2, 7 and 8 are studied for the evaluation of the proposed strategies. The first algorithm considers only the SNR information for access selection decision and therefore it is traffic load-independent:

$$access_i = \begin{cases} WLAN\ if(SNR_{WLAN^i} \geq SNR_{min}) \\ LTE \quad \text{otherwise.} \end{cases} \tag{2}$$

A mobile device i thus selects WLAN if the SNR from the best WLAN AP equals or exceeds the threshold SNR_{min}. Algorithm 1, the so called "WiFi if coverage", is realized by setting $SNR_{min} = 0$ dB.

A second alternative (Algorithm 2) is to set SNR_{min} to a higher value so as to balance relative loads, i.e. so that $U_n/C_n = U_m/C_m$, for all n, m, where U_n and C_n are the number of users and capacity of radio access technologies n respectively. A further alternative, is to use a WLAN system load dependent SNR_{min}. Here, SNR_{min} is set to 0dB for low WLAN traffic loads, and increased with system load to maintain a maximum system load just below the AP capacity. These algorithms will be evaluated in Sect. 6.

5 Models and assumptions

This section discusses a summary of the general models and assumptions used in the evaluation, of which a subset of default simulation parameters are listed in Table 1.

5.1 Radio access network simulator

Due to the complexity of the problems and scenarios studied, a Monte-Carlo static MATLAB-based multi-cell radio access network simulator model which separately derives Downlink (DL) and Uplink (UL) has been used for the evaluations. The simulated system consists of 21 cells and a wrap-around technique is used to avoid border effects and to enable faster simulation run times.

5.2 Traffic model

A traffic model named Equal Buffer is used in this paper. This traffic model assumes Poisson processes for calculating user arrivals that different arrival intensities (probability of arrival per time unit) may be given for different services. This model is based on these assumptions; the same amount of information has to be transmitted for every active user.[4] But due to different transmission bitrates (channel conditions), and thus users consume different amount of radio resources. The transmission time will vary by users and slow users will hold radio resources for a longer time and thus will have a large impact on cell performance. For example, cell-edge users[5] may then consume most of the resources.

5.3 User behavior models

We assume a heterogeneous user behavior with spatial traffic hotspots, covered by WLAN APs. We also assume a network consisting of cellular hotspots and WiFi APs, where N users are served by the cellular network. Users are always guaranteed to be under the coverage of a cellular network, but not necessarily under the coverage of a WiFi AP (depending on the coverage availability). Users are positioned in dense hotspot areas and are modeled by the parameter with a probability (for a UE to be in a hotspot) of $P_{hotspot}$ and in the remaining cellular area with probability of 1-$P_{hotspot}$. The users are always guaranteed to be under a cellular network coverage but not necessarily of a WiFi AP. It is also assumed that all hotspot areas have the same value of $P_{hotspot}$. A scenario with $P_{hotspot}$ = 80 % and uniform user distribution where $P_{hotspot}$ = 0 which puts fewer numbers of users to WiFi networks are studied. The hotspots have a radius of 50m, within which 95 % of the hotspot users are located and it is also assumed that all hotspot areas have the same value of $P_{hotspot}$. The hotspots together cover about 5 % of the system area, which results in log-scale user density standard deviations of 0.4 and 0.6 for $P_{hotspot}$ = 50 % and $P_{hotspot}$ = 80 % respectively. We assumed that there are 100 hotspots per square kilometer (km^2), and a subscriber density/km^2 of 5400. Hotspot areas, in which WLAN APs are placed half radius away from the cellular base station of its cellular cell. It is of course possible to study different sizes and positions of hotspots.

The total number of users in the system can be varied but as a reference, the value of 1000 active users per cellular cell which corresponds to about 110 users per km^2 moving at a speed of 3 km/h is assumed. Users are assumed to generate an average traffic of 100 kbps each, regardless of position.[6] Users are also assumed to require only data-like services. The radio channel between each base station and terminal antenna

[4] An active user is a terminal that is registered with a cell and is using or seeking to use air link resources to receive and/or transmit data within a short time interval (e.g., within 100 ms).

[5] Cell-edge user : is a user with a bitrate close to a pre-defined percentile of the bitrate distribution.

[6] With the system models considered it is mainly the aggregate traffic per AP that affects the performance. Therefore, a hotspot probability of X % with equal traffic generation per user could also roughly be said to represent a hotspot probability of Y % but with k times the traffic generation per user in the hotspot fulfilling the relationship kY/(1-Y)=X. Similarly, if there are N hotspots per cellular cell rather than 1, this would on average divide the traffic per hotspot with N, which could be approximated with a correspondingly reduced hotspot probability.

Table 1 A summary of the default simulation parameters

Traffic models	
User distribution	Random and uniform
User position probability ($P_{hotspot}$)	80 % (positioned in the hotspot area)
	1-$P_{hotspot}$ (probability to be in the remaining cellular areas).
Terminal speed	3 km/h
Data generation	Equal buffer (EB)
Fraction of traffic generated in DL	75 %
Minimum data rates (Mbps) in the DL & UL	1, 0.5 in (Mbps)
Radio network models	
Distance attenuation	L = 35.3 + 37.6 * log(d), d = distance in meters
Shadow fading	Log-normal, 8 dB standard deviation, 100 m correlation distance
Path-loss	Exponential (r^{α}), $\alpha = 3.52$
Propagation model (Path-loss as a funciton)	$L(d) = \beta + 10 * \alpha * \log_{10}{}^{(d)}$, where d is distance in meters
Multipath fading	LTE SCM, Suburban Macro
Cell layout (system size)	Hexagonal grid, 3-sector sites, 21 sectors in total with wrap around
Inter Site Distance (ISD)	500 m or 800 m
Cell radius	$\sim$ 166 m or $\sim$ 266 m
Subscriber density/km^2	5400
Number of hotspots per km^2	100
Hotspot radius	50
System models	
Spectrum allocation	10 MHz (50 resource blocks) 180 KHz (1 resource block)
Maximum UE antenna gain	15 dBi
Max UE output power	250 mW into antenna
Modulation and coding schemes supported	64QAM, 16QAM, QPSK and 3GPP turbo codes
Number of base station antennas	2 per cell with 10-wavelength spacing
Number of mobile station antennas for reception	2 per UE with half-wavelength spacing
Number of mobile station for transmission	1
Scheduling algorithm	Round-robin
Receiver	MMSE [17] with 2-branch receive diversity
Miscellaneous	Downlink and Uplink

pair is calculated according to the propagation and fading models. Furthermore, the output power is set sufficiently high so as to avoid coverage problems, rendering the radio link quality in the system limited by interference rather than noise. As an underlying assumption for the mobility issues, dual-mode mobile terminals capable of processing LTE networks and WLAN air interface is presumed.

Our scenarios are more appropriate for deployment in places in which a network operator is required to support highly dense populations in a given location with high population density like universities/colleges, malls/shops, airports etc. where a network operator is required to deploy large-scale WiFi APs as it is depicted in Fig. 9 so that users can have instant access to the available broadband access, and at the same time a user can be offloaded to WiFi as per the predefined operator policies.

Figure 9 depicts when we have a multi-access network deployed with less number of WiFi APs as compared to hotspots per km^2. This simulated layout is from the default parameters considering deployment of less APs by a network operator than hotspots.

Figure 10 shows when we have a multi-access network deployed with almost the same number of WiFi APs as hotspots per cell. In this case, we assumed that there are 7 Hotspots and 6 base stations per cell are deployed. And the layout shown in Fig. 10 reflects the whole evaluation assumption of our paper.

5.4 Radio network models

Radio propagation models are selected to model macro scenarios of an urban environment. The radio network model

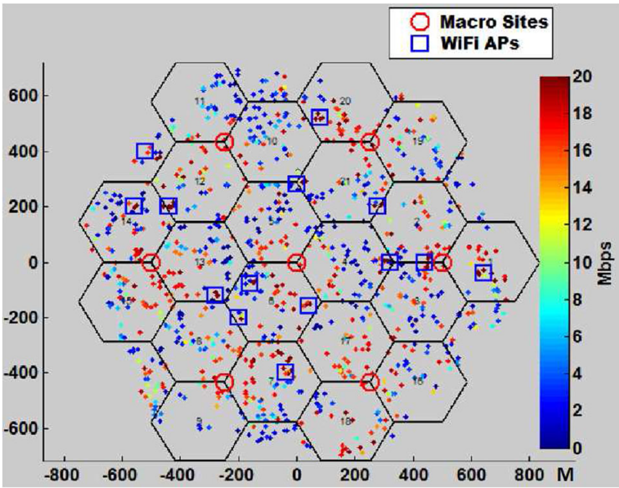

Fig. 9 Multi-access network layout with less WiFi access points per km^2

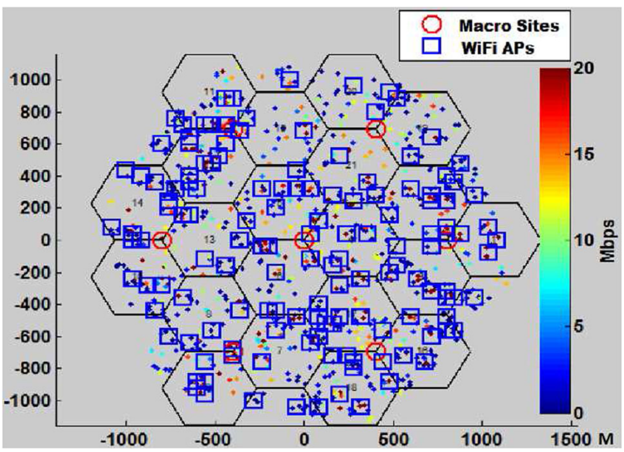

Fig. 10 Multi-access network layout with the same number of WiFi access points and hotspots per km^2

and scenarios follows the Heterogeneous Network (HetNet) scenario as described in 3GPP for LTE-Advanced [18]. Simple radio network models for both cellular networks and WLANs are used to realize the multi-access networks. A regular hexagonal grid of 7 3GPP urban macro base station sites[7] with three sectors (cells) per site and carrier bandwidth of 10 MHz which consist of 50 resource blocks is considered. Cellular base stations with Omni directional antennas are deployed at the center of the cell. For the WLAN APs, it is assumed that they are deployed at the center of the hotspots and the default positions of the hotspots are half radius away from the base station of the host cell.

Cellular propagation is exponentially modeled by a distance dependent path-loss with a constant exponent of 3.52

[7] A cell site or simply site is a base station or the geographical location of a base station which is equipped with transmission and reception equipment.

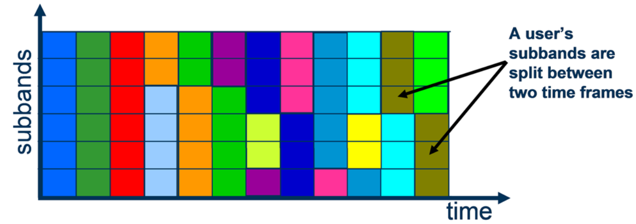

Fig. 11 Assumed scheduling algorithm

and lognormal shadow fading with standard deviation of 8dB at a correlation distance of 100 m is assumed. Whereas for WLAN, a “Keenan-Motley” propagation model assuming line-of-sight up to 60 m, followed by a constant attenuation of 0.3 dB/m is used. Log-normal shadow fading with a standard deviation of 3 dB is also assumed for WLAN. Multi-path fading is implicitly modeled through the utilized link-level performance. The base station's antenna gain is 15 dBi and the maximum transmit power from a base station in a macro cell is 46 dBm. In this evaluation 64-QAM, 16-QAM and QPSK modulation and coding schemes are supported [19].

5.5 Assumed scheduling algorithm

A Round-Robin (RR) scheduling algorithm which takes place in the eNodeB (LTE base station, i.e where mobile devices communicate through) is modeled. It assumes and handles all active users with some priority. With RR scheduling every user is allocated full bandwidth in the DL and the scheduler iterates over users. However, in the UL the bandwidth allocation per Time Transmission Interval is limited by the maximum transmit power (Fig. 11).

5.6 Propagation models: path-loss and fading

The distance attenuation between two nodes is determined using standard radio propagation models with the path loss (L) as a function of the distance d in meters as defined by the equation given below.

$$L(d) = \beta + 10 * \alpha * \log_{10}(d) \qquad (dB) \qquad (3)$$

The simulator uses the ray-based 3GPP Spatial Channel Model Extension [20] to model the multipath fading propagation in the system. Slow fading is modeled as a log-normal random variable with zero mean and standard deviation of 8 dB. In addition to path loss due to distance, a transmitted signal will be attenuated by objects blocking the line-of-sight path between transmitter and receiver. This attenuation is referred to as shadow fading and is usually modeled as a lognormal distribution, Where g_$shadowing$ is the shadow fading with zero-mean Gaussian random variable and with a variance of δ^2.

$$\log(g\_shadowing) \sim N(0, \delta) \quad (4)$$

In the simulator, the following path-loss models are available for macro scenarios; Modified *Okumura-Hata* (non-3GPP) path-loss model parameters which are calculated for four environments: urban (default), suburban, rural and open environments. Free-space path-loss models with a frequency f in Hz is modeled as

$$attconst = 20 * \log\_10(^{4*\pi*f/3e^8}) \quad (5)$$

5.7 LTE and IEEE 802.11a/802.11b

Relatively simple models of LTE and WLAN are used. To ensure reasonable accuracy, benchmarking of the single access results achieved with these models, summarized in Figs. 9 and 10 to previously established results have been made. LTE detailed simulation uses average SINR, where no fast fading is included and Interference Rejection Combining (IRC) gain is added to average SINR. More about IRC and its algorithm implementation can be found in [21]. For IEEE 802.11a/11b, one AP with a power of 100 mW is deployed at the center of each hotspot. And therefore, co-channel interference is not included. The average packet size is assumed to be 1000 bytes in order to model one UL TCP acknowledgment for every second DL frame of 1500 bytes. Each WLAN user is assigned a link rate, which yields the shortest expected time of transmission, assuming geometrical distribution for the number of trials. Based on users' SNR values, corresponding *radio link bitrates* and desired link utilization levels are calculated. Then, based on these and assuming a round-robin scheduler, resulting link utilization levels are derived. The difference with the LTE model is that active radio link bitrate of each user is scaled by a mapped value of the utilization factor of the AP that the user is connected to. The utilization factor of an AP is obtained by summing up the link utilization of the users connected to that AP:

$$\eta_j = \sum_{AP_j} \rho_i \quad (6)$$

where η_j is the utilization factor for the users of the AP j. Users are selected in terms of their link quality and the number of selected users are adjusted to the capacity limit of the AP. The utility factor is done using a table that is based on the simulation results found in [15].

5.8 Evaluation methodology

Simulation methodology has been employed throughout the evaluations, and results for both UL and DL are independently derived. To reach satisfactory accuracy and realistic results a number of simulations by increasing the number of points and different parameter settings (for example, LTE Inter Site Distance (ISD), WiFi APs, distance thresholds, load, hotspots, etc) are run for each studied hypothetical scenarios. Users are randomly placed in the network area according to the user behavior modeling over the system area as it is described earlier and access selection is done afterwards with the identified access selection methods. The simulation methodology is based on the following iterative simulation loop.

- *Generate a network It generated a regular hexagonal layout as it is previously explained and it optionally mitigates the network border effects by wrap-around (default).*
- *Distribute UEs randomly with a uniform distribution.*
- *Schedule users randomly. For a given load, generate interferers for each user.*
- *Calculate Signal to Interference plus Noise Ratio (SINR) for each user (SINR per antenna and per stream after combining).*
- *Calculate the bitrate for each user and apply SINR-to-bitrate mapping using the mutual information model of* [22]. *This gets the combined bitrate per user when the user is scheduled.*
- *Calculate the achievable user throughput, cell throughput and other performance metrics.*

Terminology The terms definition used in the next sections of the paper are given here. In the plots we use terminologies 11a to denote for IEEE 802.11a, LPN to denote Low Power Node (WiFi), Macro to denote for LTE.

- *Alternative0 denotes WiFi if Coverage*
- *Alternative1 denotes ANDSF rule based on Cell ID*
- *Alternative2 denots ANDSF rule based on Position*
- *Alternative3 denotes ANDSF rule based on Cell ID and position*
- *Combined denotes LTE and WiFi*

6 Numerical results

This section presents the detailed simulated numerical performance results based on the offloading algorithms considered for evaluations and hypothetical scenarios analyzed in this paper. Simulation graphs for scenarios with combined LTE and IEEE 802.11 systems are included. The numerical results obtained with the proposed algorithms are naturally subject to the assumptions and modeling applied.

6.1 Algorithm 1: WiFi if coverage

In this algorithm, the user connects to the best WiFi AP if his/her SNR is greater than SNR_{min}, i.e. select WiFi when-

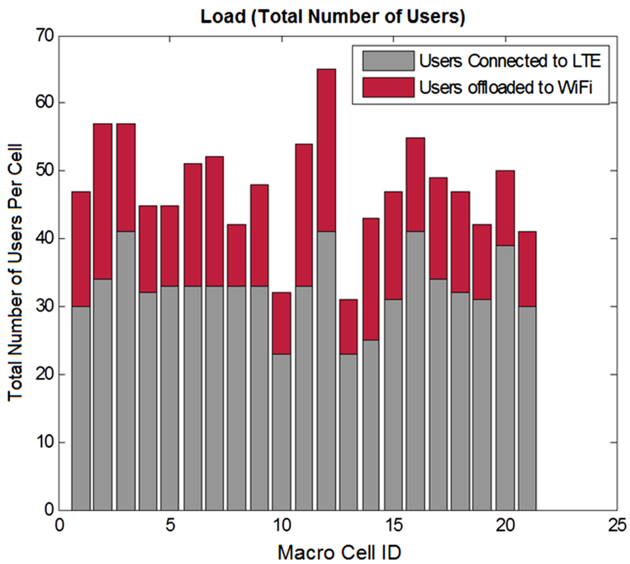

Fig. 12 Average number of users connected to LTE and WiFi using Algorithm 1

ever there is WLAN coverage. As a result, in the simulation this algorithm is realized by setting the SNR_{min} = 0 dB and setting the SNR threshold to 0 dB could be interpreted as "WiFi if Coverage" principle (Fig. 12).

Using this algorithm, an average of 31.5 % of users can be offloaded to the available WiFi access network while the remaining 68.5 % average numbers of users can use the LTE network on the available macro cell ID based on the topology depicted in Fig. 10.

6.2 Algorithm 2: fixed SNR threshold

As it is explained in the last two paragraphs of Sect. 4, the threshold is set to 0 dB by default which could be interpreted as *WiFi if Coverage* algorithm. The threshold could also be set to a higher value so as to balance the relative loads. With a fixed SNR threshold from WLAN, a user selects the best WLAN if the SNR from the best WLAN AP equals or exceeds the predefined threshold SNR_{min}; otherwise the user will be directed to the cellular network. This algorithm is the same as the *WiFi if coverage* algorithm at the low traffic loads and hence the SNR_{min} is a function of the WLAN load. However, this algorithm is load adaptive, and this means how often a user selects the access network can be triggered by changing the load. Since only one user is served at a time in WLAN technology, there is no interference between users. And therefore, SNR instead of SINR is used in our evaluation for a better signal quality between base station and mobile terminal. The only parameter needed from this algorithm is the SNR information from WLAN and then the network access selection decision will be made according to the following expression:

$$access_i = f(\{SNR_i^1, load_i^1\}) \tag{7}$$

and then this could again be expressed as in the following where the SNR_{min} is a function of the WLAN load.

$$access_i = \begin{cases} WLAN \ \ SNR_i^1 \geq SNR_{min}^1(load_i^1) \\ LTE \quad \text{otherwise} \end{cases} \tag{8}$$

where SNR_i^1 is the SNR of user i from WLAN and it is a function of the WLAN load. The load and the SNR_{min} are directly proportional at each WLAN access point. When the load increases the SNR_{min} is also increases at each individual WLAN access points so as not to have overloaded WLANs. When load is high, the algorithm sorts the users in terms of their link quality and selects the best ones. At the same time the number of selected users is adjusted to the capacity limit of WLANs. In real time, users' access network selection can be initiated by changing the load or the SINR. Another approach deployed in this work due to the use of a static simulator is access network reselection after fixed time intervals.

6.3 Algorithm 3: ANDSF models

ANDSF is a new EPC entity defined in the 3GPP standard [5] that contains data management and control functionality to provide necessary access network discovery and selection assistance to the UE as per the operator's policy. It allows mobile device users to discover and select from a list of nearby access networks, determine AP traffic load and their capabilities, and to connect to the available networks based on the user preferences and predefined network operator policies.

6.3.1 ANDSF model based on Cell-ID

(I) *Discovery* WiFi APs are discovered if they are located in the geographical area corresponding to the current macro cell ID. The geographical area of a macro cell is a circle circumscribing the hexagon in the regular macro cell plan where the the cell radius, (r = $\frac{ISD}{3}$) in the tri-sector antenna as it is shown in Fig. 13. Both suburban and urban scenarios can be modeled with this cell configuration.

(II) *Connection/offloading/handover* In this case of the ANDSF model, the mobile connects to the strongest coverage of the discovered WiFi APs provided it has SNR value of greater than zero (i.e. $SNR > 0$).

Using this offloading algorithm, the simulation result as it is depicted in Fig. 14 shows that an average of 75.7 % of users can use the LTE network on the available macro cell ID while the remaining 24.3 % average of users can be offloaded

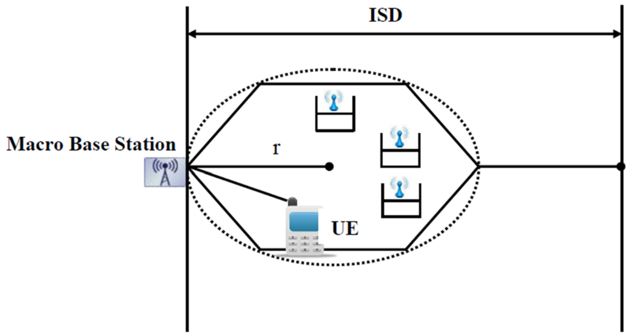

Fig. 13 Macro-cell with tri-sector antenna

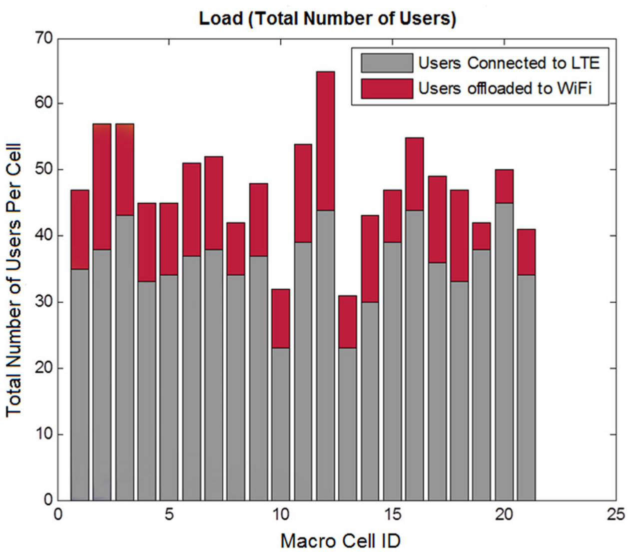

Fig. 14 Average number of LTE and WiFi users, using the ANDSF rule-based on cell ID

to the available WiFi access network based on the network operator's predefined policies.

6.3.2 ANDSF model-based on position

(I) *Discovery* In this case of the ANDSF model, WiFi Access Points are discovered if they are physically located close enough to the mobile user, which is the UE.
(II) *Connection/offloading/handover* The user connects to the strongest coverage of the discovered WiFi APs nearby provided that the mobile is located within the distance threshold predefined by the network operator (i.e. if it is in the range of the radius specified which in this case the radius is for e.g. 200 m.

As it is depicted in Fig. 15, this offloading algorithm allows more users to be connected to LTE than being offloaded to WiFi when we set the distance threshold for the network discovery very much less than the radius of the hotspot. This algorithm of course varies on the value of the network discovery distance threshold, i.e. the larger the network discovery

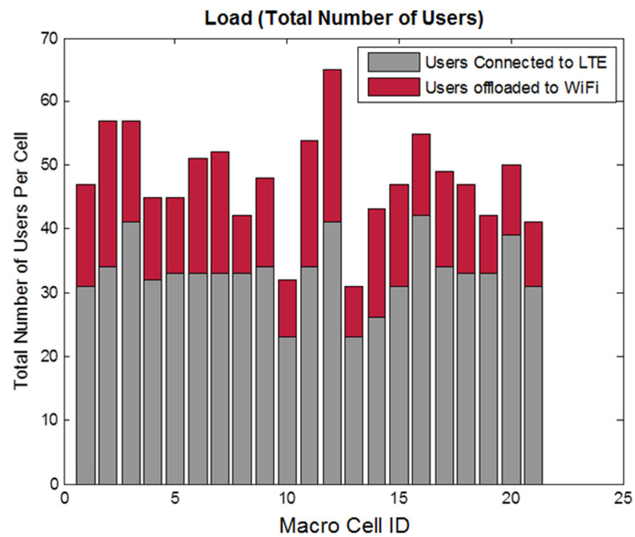

Fig. 15 Average Number of LTE and WiFi Users when the D_{thr} for the discovery = 200, SNR = 0

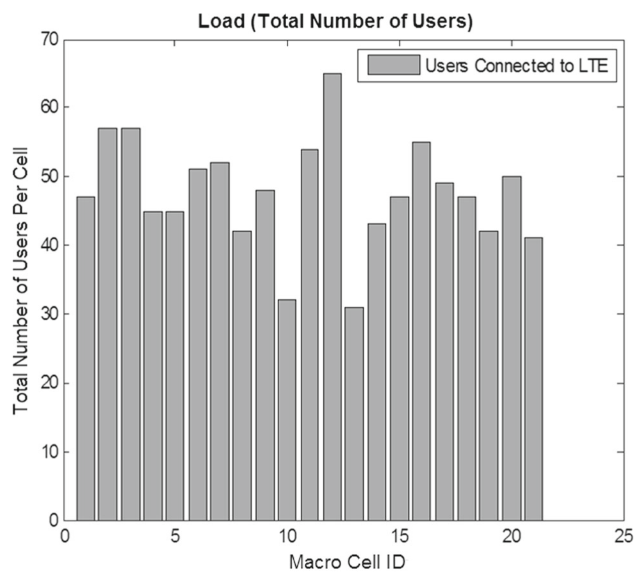

Fig. 16 All users are connected to LTE when D_{thr} = 0

distance threshold the more users will be offloaded to WiFi. In this case a load (monthly data volume per user, i.e. vectors allowed) of 3GB/Month. For some reason, if the network operator wants to block all users from being offloaded to the available WiFi access networks and to stay connected to LTE, we only need the distance threshold for the discovery to set it to 0 (zero) (Fig. 16).

Figure 17 shows the 10th percentile bitrate for the number of users connected to the WiFi network versus User Throughput (Mbps) in a random user distribution (i.e. $P_{hotspot}$ = 80 %). In this scenario, an LTE with smaller ISD of 500m is assumed. Keeping the D_{thr} constant (200 m) and varying the SNR value (e.g. 30, 15, 10, 6, 4 and 2 dB) gives better

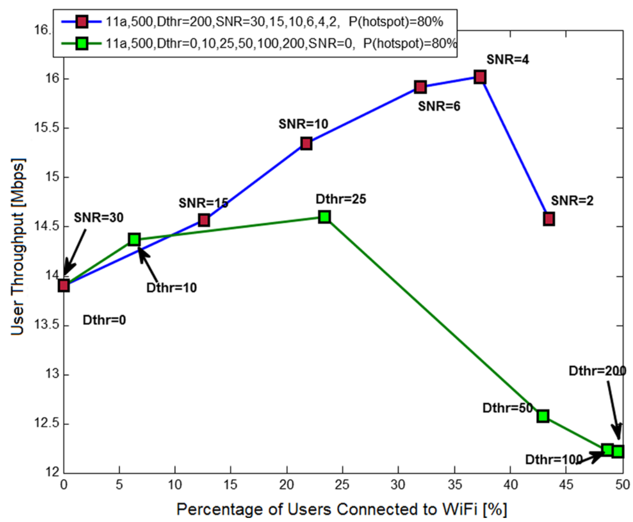

Fig. 17 Random user distribution with smaller ISD and the same number of hotspots and WiFi access points per cell

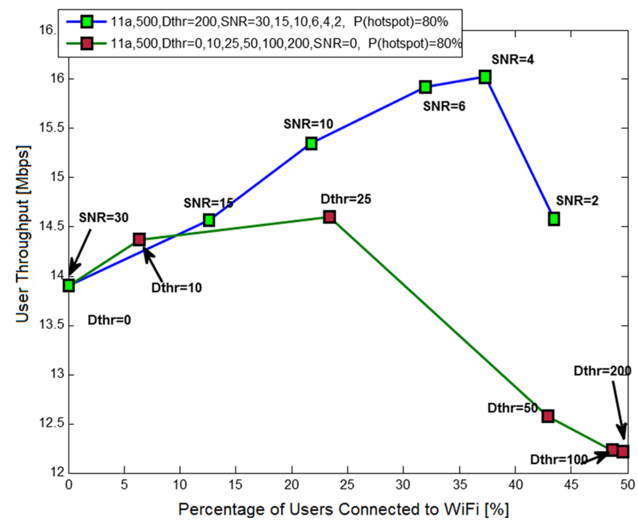

Fig. 19 Random user distribution with larger ISD and the same number of hotspots and WiFi access points per cell

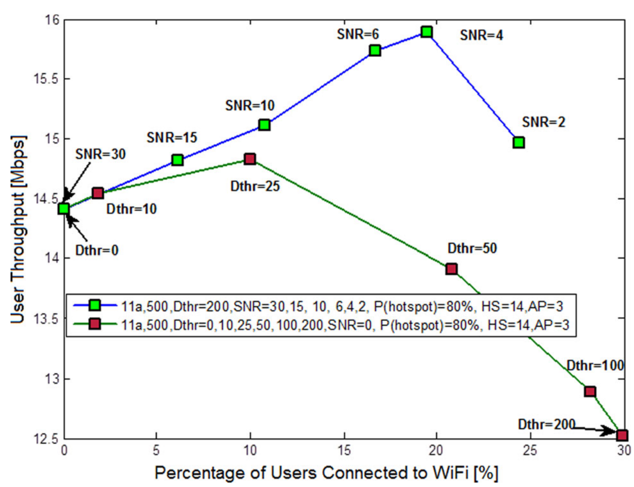

Fig. 18 Random user distribution with smaller ISD and more number of hotspots per cell

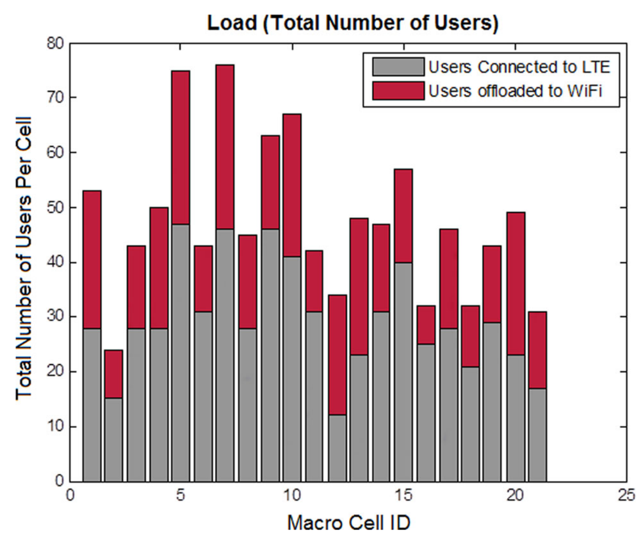

Fig. 20 Average number of users connected to LTE and WiFi when the D_{thr} = 200, SNR = 0

user throughput than varying the network discovery distance threshold (D_{thr}) while setting the SNR value to a threshold of 0 dB. This holds true where users are randomly placed in a network with the same number of hotspots and WiFi access points per cell. And Fig. 18 depicts the 10th percentile bitrate when we deploy more base stations/km^2 than WiFi APs/km^2 in a random user distribution.

Figure 19 depicts when we have a scenario with larger LTE ISD of 800 m), for random distribution of users, varying the D_{thr} (0, 10, 25, 50, 100 and 200 m) while keeping the SNR to a fixed value of 0 dB which gives better user throughput and it is valid for the majority of our simulation results. It is quite obvious that the larger the LTE ISD, the lesser the user bitrate will be (Fig. 20).

6.3.3 ANDSF model-based on cell-ID and position

This ANDSF model combines both the ANDSF models described above. The following figure shows the average number of users connected to LTE and offloaded to WiFi by applying this offloading algorithm when the distance threshold for the network discovery is set to 200 m while the SNR is set to a threshold value of 0.

Figure 21 shows the user throughput percentiles (10th, 50th or mean and 90th) versus traffic load (traffic volume per user) characteristics in the UL and DL for the access technologies considered in this work. Clearly these depend on both system load and signal quality. And therefore, in order to select the access technology that maximizes user bitrate,

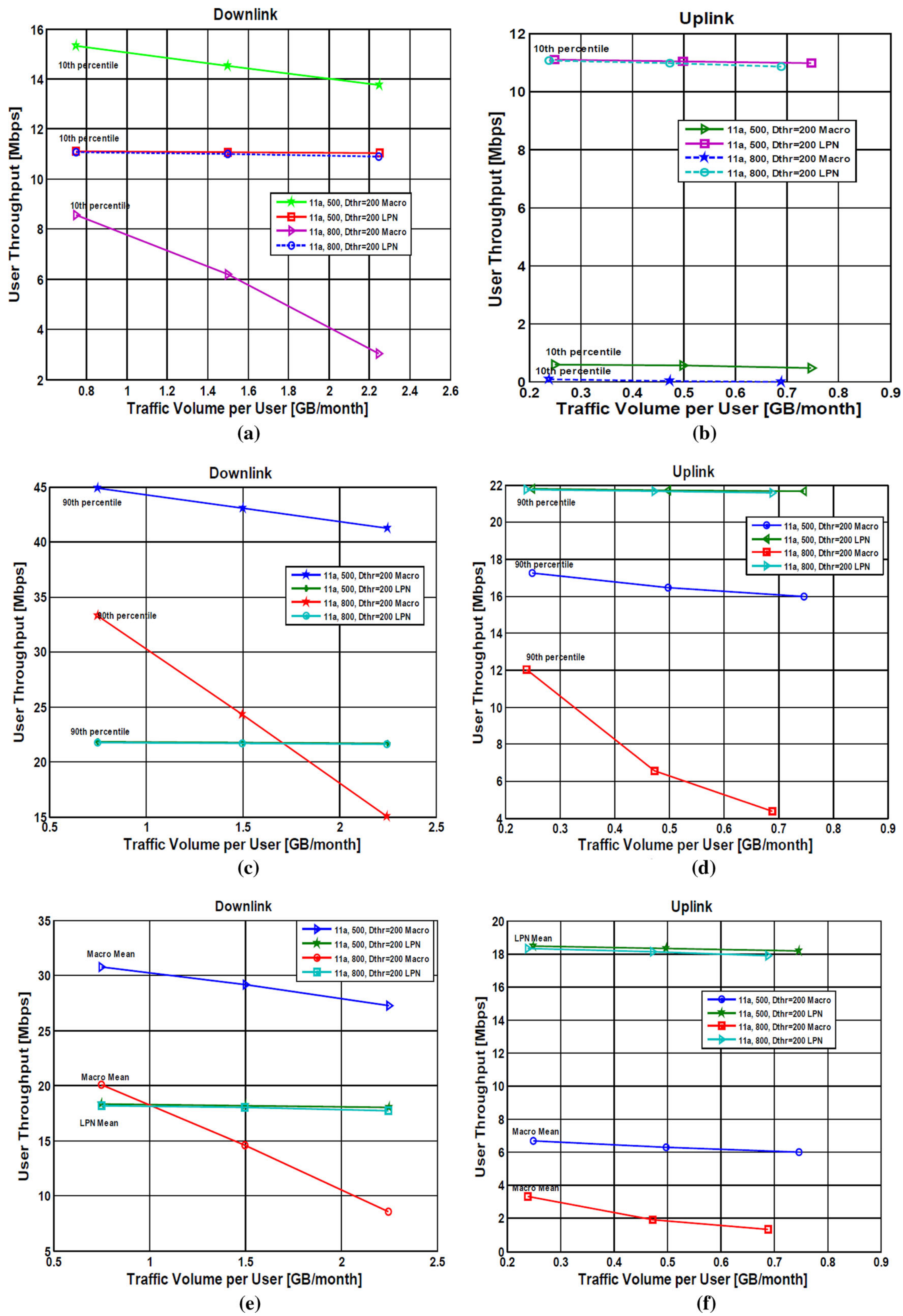

Fig. 21 Mean, 10th and 90th percentile comparison between LTE and 802.11a in the DL/UL when LTE ISD = 500 & 800 m, SNR = 0 dB, D_{thr} = 200 m, **a** downlink 10th percentile comparison, **b** uplink 10th percentile comparison, **c** downlink 90th percentile comparison, **d** uplink 90th percentile comparison, **e** downlink mean comparison, **f** uplink mean comparison

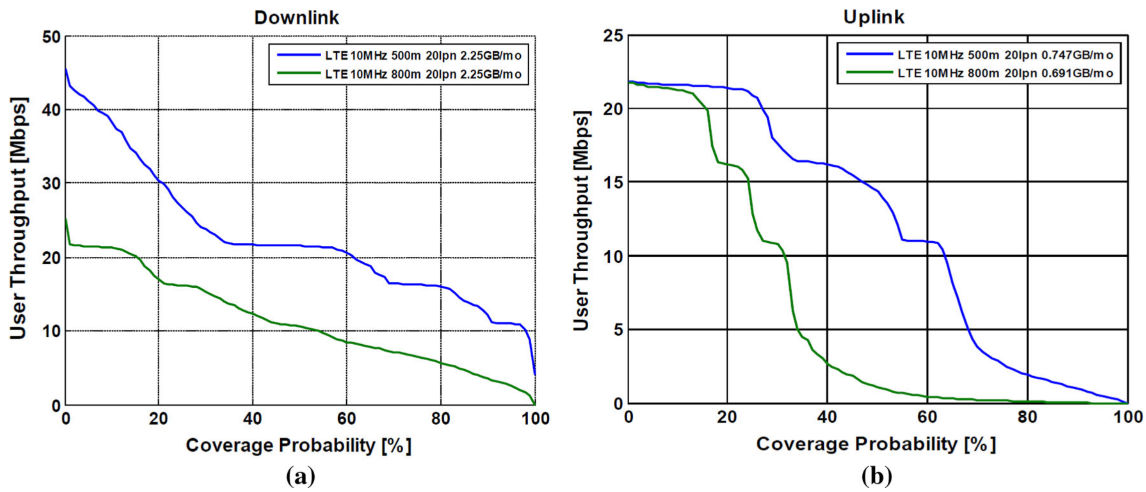

Fig. 22 DL and UL performance comparison between LTE and WiFi when LTE ISD = 500 m & 800 m, SNR = 0 dB, **a** LTE & WiFi DL performance, **b** LTE & WiFi UL performance

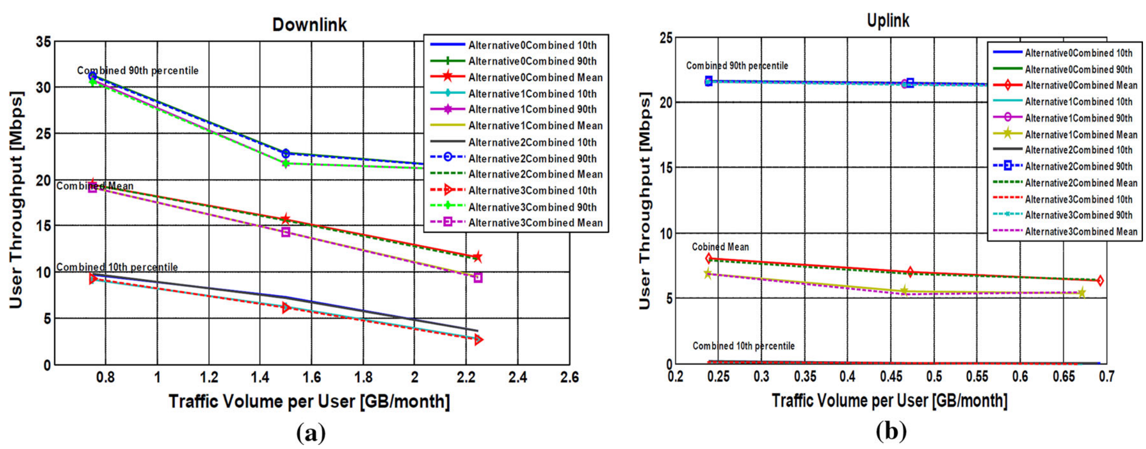

Fig. 23 LTE and WiFi combined user throughput in the DL and UL, **a** LTE & WiFi combined user throughput in the DL, **b** LTE & WiFi combined user throughput in the UL

both individual signal quality and system load need to be known and we should also note that the spectrum allocation differs between the access technologies. Figure 21a–f shows that with smaller LTE ISD (500 m) and fixed distance threshold (D_{thr}) of 200 m, LTE (Macro site) offers better mean, 10th and 90th percentile user throughput in the DL. Whereas with larger LTE ISD (800 m) WiFi APs offer a better mean, 10th and 90th percentile user throughput than LTE in the DL. When it comes to the UL; WiFi APs offer a better mean, 10th and 90th percentile user throughput with both smaller and larger LTE ISDs. A macro site (LTE) with smaller LTE ISD relatively yields a better user throughput in the UL.

Figure 22 shows the combined simulated performance results for the DL and UL with different ISDs where the hotspot probability is 80 %. It is assumed that the fraction of traffic generated in the DL is 75 % and the remaining 25 % traffic is generated in the UL with a monthly data volume load per user of 3 GB/Month. Figure 22a shows that the network can offer a maximum user throughput of around 25 Mbps in the DL with a larger LTE ISD. Whereas, with a smaller LTE ISD a user throughput of around 45 Mbps can be achieved in the DL as it is shown in Fig. 22b.

In this section a detailed simulated performance results of the offloading algorithms considered are presented. As it was for the previous plots, the hotspot probability for the following plots is also 80 %. Figure 23a and b shows how the different algorithms considered for evaluation of our work distribute the traffic load on the cellular network (LTE) and WLAN subsystems. As a reference the single-access capacities are also shown.

As it is depicted in Fig. 23a and b, alternative 2 (which is the ANDSF algorithm based on Position) achieves the best combined 90th percentile bitrates in both DL and UL scenarios. It is seen that with the 80 % hotspot probability assumed in this scenario, all algorithms result in the

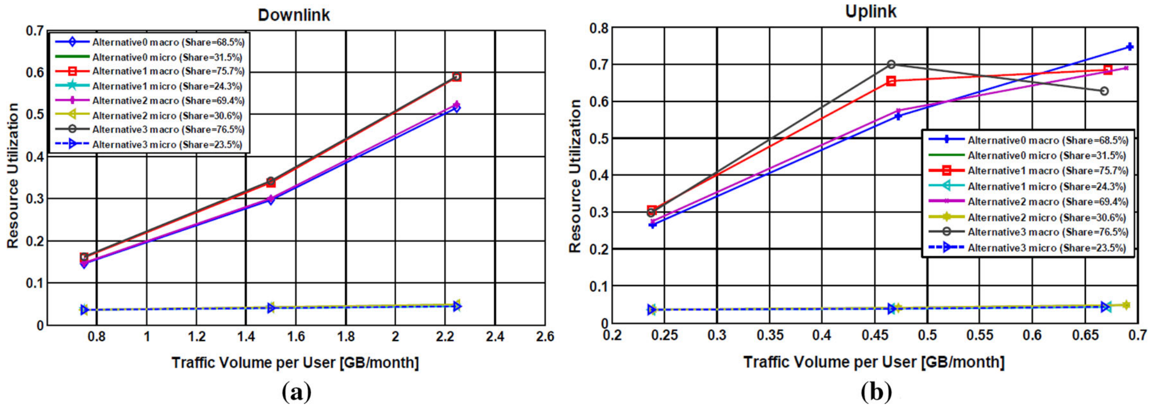

Fig. 24 LTE and WiFi resource utilization in the DL and UL, **a** LTE & WiFi resource uitilization in the DL, **b** LTE & WiFi resource utilization in the UL

same traffic distribution. This is because when $P_{hotspot} < C_{WLAN}/(C_{WLAN} + C_{Cellular})$, the cellular network reaches its capacity limit before WLAN even if all hotspot users are allocated to WLAN. In this case alternative 2 reduce to *WiFi if coverage*, and no bitrate gain is achieved, and therefore as with the low WLAN load, WLAN is typically the best alternative. For high loads, better mean values are however achieved with alternative 0 and alternative 2. It is also worth noting that with algorithm 3 acceptable bitrates are achieved for subsystem traffic loads beyond the single-access capacity limits. This multi-access related capacity gain stems from the fact that the access selection principle allocates users to the subsystem in which they consume the least radio resources. In this scenario it is seen that alternative 3, *ANDSF rule based on Cell ID and position*, dramatically overloads WLAN, already for relatively low traffic loads, which results in poor bitrates. Figure 24 presents the DL and UL utilization of LTE/WiFi assuming that the monthly data volume per user is 3 GB/Month which can be represented as scalar or vector for capacity evaluation of 1, 2. The higher the load the higher is the resource utilization and it is directly proportional with the given value of LTE ISD. We have only one share of the access networks for the given loads both in the DL and UL due to the fact that the same number of users is assumed in our simulator for both utilizations.

6.4 Average cell discovery cost

This algorithm measures the average number of APs discovered when the UE scans the available WiFi access networks as it is shown in Table 2. Access network discovery received by the ANDSF is used in order to scan for certain accesses only in specific area; in this way the energy consumption of the UE can be reduced and lead to a long battery lifetime. But this is only part of the problem, because to blindly scan for APs takes a lot of time and as result of this the mobile device may not discover it in time. This means, handover would be beneficial until the blind scanning discovers the available access.

Table 2 Average cell discovery cost

Algorithms	Average cell discovery cost	
	ISD = 500, D_{thr} = 200	ISD = 800, D_{thr} = 200
Algorithm 1	140 APs	140 APs
Algorithm 2	140 APs	140 APs
Algorithm 3 (ANDSF Models)		
Model-1	7.464 APs	7.775 APs
Model-2	10.486 APs	14.645 APs
Model-3	5.203 APs	2.936 APs

As an extreme scenario, when we set the D_{thr} to a value of 0 or higher SNR (e.g. 30 dB) for the ANDSF algorithms, the cell discovery cost will be 0. This means no user is allowed to be offloaded to the available WiFi networks. In this scenario the network operator can block users from being offloaded to and stay connected to LTE for some reason.

6.5 Summary of simulation results

In this study the networks are assumed to support data-like services. Other types of services, such as voice and also a combination of different services could also be studied. The performance of the network depends on the value of LTE ISD and it is assumed that the fraction of traffic generated in the DL is 75 % and the remaining 25 % traffic is generated in the UL with a monthly data volume load per user of 3GB/Month. Our detailed simulation results of LTE and IEEE 802.11a/11b have shown that, when we have a smaller LTE ISD it is always best to stay connected to LTE cellular network. Whereas when we set the ISD to a larger value, it is seen that the bitrate offered by WiFi is typically higher than that of LTE and hence, it is always best to connect to WiFi.

In this case users can benefit from being offloaded to the available WiFi access networks. A simple *WiFi if coverage* access selection principle may thus be expected to perform reasonably well. With smaller ISD, it is also seen that variable SNR-based access network selection yields better user throughput than keeping the Network Discovery distance threshold fixed (D_{thr} = 200 m) in both uniform and random user distributions. This also applies for scenarios where we have either the same number of WiFi APs as base stations deployed in a network or more number of base stations than APs. As a reference the single-access capacities are also shown. Clearly these depend on both system load and signal quality. And therefore, in order to select the access technology that maximizes user bitrate, both individual signal quality and system load need to be known and we should also note that the spectrum allocation differs between the access technologies.More simulation results in terms of mean, 10th and 90th percentiles for user throughput versus traffic load characteristics per user in the UL and DL for LTE cellular network and WLAN subsystems are briefly discussed in [15].

With the assumed hotspot probability of 80 % , it is seen that all algorithms result in the same traffic distribution. This is because when $P_{hotspot} < C_{WLAN}/(C_{WLAN} + C_{Cellular})$, the cellular network reaches its capacity limit before WLAN even if all hotspot users are allocated to WLAN. In this case no bitrate gain is achieved, and therefore as with the low WLAN load, WLAN is typically the best alternative. For high loads, better mean values are however achieved with *WiFi if Coverage*. It is also worth noting that with algorithm 3(ANDSF model) acceptable bitrates are achieved for subsystem traffic loads beyond the single-access capacity limits. This multi-access related capacity gain is due to the fact that the access selection principle allocates users to the subsystem in which they consume the least radio resources. It is also seen that ANDSF model based on Cell ID and position dramatically overloads WLAN, already for relatively low traffic loads which results in poor bitrates. DL and UL utilization of LTE/WiFi with monthly data volume per user of 3GB/Month which can be represented as scalar or vector for capacity evaluation of 1, 2 is assumed. The higher the load the higher is the resource utilization and it is directly proportional with the given value of LTE ISD. We have only one share of the access networks for the given loads both in the DL and UL due to the fact that the same number of users is assumed in our simulator for both utilizations. In addition to the simulation results discussed above, a more thorough discussion of several additional combined scenarios have been evaluated and are briefly presented in [15].

Concerning access combinations, scenarios including IEEE 802.11a/b as well as LTE indicate that it is the relationship between the cellular and WLAN capabilities that matters. The *better* the WLAN system the better the *WiFi if coverage* algorithm. Similarly, the better the cellular system, the higher the user throughput and the higher the SNR value the lesser the WiFi share will be. Higher SNR value directs users to the cellular network and makes the WiFi network less loaded. As a result, fewer users connect to WiFi where they experience high bitrates. Varying hotspot probability and propagation conditions also affects the SNR_{min} that yields balanced traffic loads. In summary, signal strength can be considered as a basis access selection input, and may also be sufficient in many scenarios. System load is motivated as an additional input in scenarios with high bitrates offered by the LTE network and strongly heterogeneous traffic load, and so that the traffic load per WLAN AP is similar to that of the cellular access points. Finally, favorable mean bitrate results have been derived through an extensive simulation.

7 Related work

In this section, we briefly review the related work on mobile data traffic offloading. We have studied the existing major cellular traffic offload solutions related to efforts in the current 3GPP standards and mobile industry initiatives. Among the existing primary offloading solutions used by the industry are traffic offloading via deployment of Femtocels for indoor cellular environments whereby facing the technical challenges found in [23] and via WiFi APs. An investigation of handover procedure in LTE-based Femtocell network is given in [24] to alleviate the problem of wireless systems in situations where we have bandwidth restrictions and coverage. There are also delay-tolerant approaches to traffic offloading where the delay is usually caused by intermittent connectivity [25] and delay the delivery of information over cellular networks and exploit opportunistic communications and WiFi networks to offload mobile data traffic [26–28]. There is also another approach which has proposed a scheme called Wiffler [8] to augment mobile 3G networks using WiFi for delay-tolerant applications. This approach is used for offloading of mobile 3G traffic to WiFi networks mainly for vehicular networks. In [29], a framework is modeled to study the economic aspects of WiFi offloading and its benefit for both the provider and users. Similarly, a work [30] examines the viability of WiFi offloading through a variety of aspects.

The 3GPP standard [31] provides an architecture which allows easy interworking between cellular systems and WiFi hostspots. Such interconnected technologies allow full mobility of the user in the available access networks and enable dynamic management policies that ensure end-to-end secure transmission of data and services across the heterogeneous networks. A cooperation architecture and mobility management functionalities and the rules that are applied for the support of mobility within the heterogeneous networks is described in [20]. An enhanced offload solution is also introduced in 3GPP work item in [32], which makes use of a

Table 3 Proposed ANDSF database organization

UE_Location 3GPP(cell_ID) Other	Access type = WiMAX	Access type = WiFi
Locn_ l Cell_Id = Cell_ l	NSP_ID = NSP_1 NAP_ID = NAP_1 NAP_ID = NAP_2 NSP_ID = NSP_2 NAP_ID = NAP_2 NAP_ID = NAP_3	SSID = WiFil, BSSID = BSl, SNR l SSID = WiFi2, BSSID = BS2 , SNR2
Locn_2 Cell_Id = Cell_2	NSP_ID = NSP_2 NAP_ID = NSP_3	Not available
Locn_3 Cell_Id = Cell_3	Not available	SSID = WiFil, BSSID = BS3, SNR3 SSID = WiFi4, BSSID = BS4, SNR4
……………………	…….	…………
Locn_n Cell_Id = Cell_n	NSP_ID = NAP_n NAP_ID = NAP_2	SSID = WiFi6, BSSID = BS5, SNR5

client–server based protocol DSMIPv6 [10] to achieve seamless handover solution between cellular networks and WiFi has recently been standardized. It supports IP session continuity for IPv4 and IPv6 sessions and the signaling needed to carry routing filters when the UE is connected to multiple accesses simultaneously are specified in [33]. The handover procedures can be applied to a single or multiple IP flows belonging to the same PDN connection. This means that some IP flows of one PDN connection can be routed via one access network while simultaneously some IP flows of the same PDN connection can be routed via another access network.

As per the current 3GPP standards, it is possible to provide a non-seamless WiFi offloading by assigning the UE a separate IP address specific to the WiFi connection. However, a seamless handover is needed in other cases, for example when paid service subscriptions are involved. In [34], a solution is addressed to the challenging research question as to which user should be served by which access network and when to conduct a handover in a heterogeneous mobile network environments. Similarly, in a previous work seamless handover scheme for mobile IPv6 is proposed in [35]. As a best offload solution, the 3GPP [36] has produced a work item function called SIPTO (Selected IP Traffic Offload) which enables IP Flow Mobility for an operator to seamlessly offload selective IP traffic while supporting simultaneous cellular and WiFi access networks. This function function enables an operator to offload certain types of traffic at a network node close to that UE's point of attachment to the access network.

These all studies have limitations, and more importantly the performance aspect is neglected; and the strategies of the big vendors that do offloading are not publicly available. We have considered the performance as a key issue in mind and to make our offloading algorithms publicly available. Our intent in this paper is to propose a novel handover solution as an extension to the 3GPP standard.

8 Proposal

The challenge with ANDSF access selection in the current 3GPP standard is that it only allows prioritized access selection, that is an access network is only selected if no other access network having a higher priority is available. The ANDSF selection rules have a number of conditions where one or more access networks to be used whenever a selection rule is active. When the ANDSF policy selection rules identify an available network, the highest priority rule becomes an active selection rule and network reselection is performed. But this doesn't take network performance into account and it doesn't seem to be applicable in some situations say for example in trees/forest environments where radio signal attenuation easily occurs or when there is damage on buildings and other obstructions which result in loss of radio signal during mobility. And our system evaluation has demonstrated that this is non-optimal for heterogeneous wireless environments.

One possible solution to the prioritized access network selection problem is, if we have a conforming UE we may be able to tell it *"Connect to this AP only if the signal strength is greater than SNR-threshold"*. We could say, choose any AP with the strongest SNR which is not possible in the current 3GPP standard. When we have WiFi APs deployed in a distributed wireless environment, for some reason the user throughput could be good for some of the users and it could also be terribly bad for other users and this problem can be effectively solved by adding SNR.

Based on our evaluation results we propose that the SNR-threshold should be included in the current ANDSF database organization which the ANDSF features for storing both the Access Network Discovery information and the inter-system client mobility policy parameters [37]. Accordingly, the SNR-threshold could be set with the coverage mapping information for the available non-3GPP access networks based on the 3GPP cell Identifiers after the ANDSF network discovery is performed as it is shown in Table 3. Because we believe that for a network access selection to perform well, it requires both signal quality and of course system load information as an input parameters and this in return helps for an efficient offloading decision.

9 Conclusion

This paper addresses solutions for network-controlled WiFi offloading in LTE cellular networks when performance needs exceed the capability of the LTE access. With the proposed models and assumptions, our system evaluations have demonstrated that offloading users from LTE to WiFi networks reduces demand on the LTE network without affecting user performance, and hence, it is encouraging to offload mobile data traffic and reduces the load on licensed cellular spectrum of mobile network. When the hotspot probability increases beyond $C_{WLAN}/(C_{WLAN}+C_{Cellular})$,[8] the *WiFi if coverage* algorithm may yield overloaded WLAN and poor user bitrates at high traffic loads. This can be avoided by using a load-based SNR-threshold gradually balancing the load between the access networks. Our simulation results have shown that it is possible to control the network by changing the network discovery distance threshold to control the amount of WiFi offloading. Since WiFi is greatly available through various hotspots and at home, and is also in a number of LTE/3G devices, it offers the capacity to become a seamless offloading solution of LTE.

In scenarios with moderate hotspot probability, yielding a hotspot load within the WLAN capacity, the *WiFi if coverage* algorithm is sufficient. But, in some scenarios with high traffic load concentrated to the LTE networks in the same order as WLANs or hotspots, different conclusions may however be reached with more complex principles taking traffic load into account. In network scenarios with very high hotspot traffic load, the LTE network may offer higher user throughputs. This can be taken into account by introducing some form of load analysis and admission control for the WLAN network to prevent it from being overloaded. A user can discover multiple access points in a heterogeneous wireless networks. For various reasons these access points may have different capabilities. Hence, we can measure the capability of an access point so as to select the best wireless network connections available to use. A proposed solution of traffic load metrics, Running Variance Metric and a Relative Network Load which are used for measuring the traffic load of access points in wireless access networks is discussed in [38]. In another study [39], a resource allocation scheme in cellular networks using a cooperative game theoretic approach at the network level while utilizing simultaneous use of available radio interfaces at the device level is proposed. As it is observed from our simulation results for a combined LTE and IEEE 802.11, higher user throughputs are offered in the LTE network for more combinations of traffic load. It is thus not obvious that WLAN is always the best alternative. It is also seen that signal strength-based network access selection yields quite satisfactory results in many scenarios.

An additional parameter of great importance not discussed in this single-service work is the service type. In this work, networks are assumed to support only data-like services. Other types of services, such as voice and also a combination of different services could also be studied. The results of this work would for example be very different for voice services, and as a result it might be possible to arrive at different conclusions. Because, WLAN capacity for voice-like services is much lower than for data-like services. A more realistic user behavior model could also be employed. The robustness of the results can be analyzed by varying some system parameters such as propagation models, hotspot radius, hotspot position etc.

In future work, using dynamic simulator the effect of those issues on the network performance can be investigated. For example analysis in dynamic simulator which models connections over time and probably individual packets, different solutions for measuring network performance in LTE and WiFi access networks connecting to the core network can be evaluated by taking the techniques of interest like *Kalman Filters* and *Exponential Averaging* algorithms using network performance indicators like roundtrip delay variations (jitter) and packet drops. Future work will also include further integration of the proposed solutions into upcoming 3GPP standards.

Acknowledgments The author would like to thank the anonymous reviewers for their insightful comments and valuable inputs.

References

1. Ericsson. (2011). "Ericsson Research, Annual Report 2011, Traffic and Market Data Report on the Pulse of the Networked Society," tech. rep., Ericsson, November.
2. Handa, A. (2009). "Mobile Data Offload for 3G Networks. white paper," tech. rep., IntelliNet Technologies Inc, October.

[8] Where $\mathbf{C}_n$ is the capacity of the network **n**.

3. Cisco Systems Inc. (2011). "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2010–2015," white paper, Cisco Systems Inc., February.
4. "3GPP TS 23.401: GPRS Enhancements for E-UTRAN Access (Rel. 10)," June 2011.
5. "3GPP TS 23.402 V10.3.0, Architecture Enhancements for Non-3GPP Accesses (Rel. 10)," March 2011.
6. "3GPP TS 24.312: Access Network Discovery and Selection Function (ANDSF) Management Object (MO) (Rel. 10)," June 2011.
7. Lee, K., Lee, J., Yi, Y., Rhee, I., & Chong, S. (2010). Mobile data offloading: How much can wifi deliver? *SIGCOMM Computer Communication Review*, *40*, 425–426.
8. Balasubramanian, R. M. A., & Venkataramani, A. (2010). Augmenting Mobile 3G Using WiFi. In *Proceedings of MobiSys 2010*, pp. 209–222.
9. Hultell, J., Lungaro, P., & Zander, J. (2005). Service provisioning with ad-hoc deployed high-speed access points in urban environments. In *IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications, 2005. PIMRC 2005*, vol. 3, pp. 2019–2023.
10. "IETF RFC 5555: Network Working Group, Mobile ipv6 support for Dual Stack Hosts and Routers," June 2009.
11. Dutta, A., Madhani, S., & Zhang, T. (2006). Network discovery mechanisms for fast-handoff. In *3rd International Conference on Broadband Communications, Networks and Systems, 2006. BROADNETS 2006.*, pp. 1–11.
12. Åhlund, C., Wallin, S., Andersson, K., & Brannstrom, R. (2008). A service level model and internet mobility monitor. *Telecommunication Systems*, *37*(1–3), 49–70.
13. "3GPP TS 24.312: Service requirements for the Evolved Packet System (EPS) (Rel. 11)," June 2011.
14. "3GPP TS 23.861 V1.3.0: Multiple access PDN connectivity and IP flow mobility (Rel. 9)," September 2009.
15. Hagos, D. H. (2012). *The performance of WiFi Offload in LTE Networks: Long term evolution*. Germany: LAP Lambert Academic Publishing.
16. Hagos, D. H., & Kapitza, R. (2013). Study on performance-centric offload strategies for lte networks. In *IEEE Wireless and Mobile Networking Conference (WMNC), 2013 6th Joint IFIP*, pp. 1–10.
17. Winters, J. (1984). Optimum combining in digital mobile radio with cochannel interference. *IEEE Journal on Selected Areas in Communications*, *2*(4), 528–539.
18. "3GPP TR 36.814 Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA Physical Layer Aspects, Rel. 9," February 2009.
19. "3GPP TS 36.213 Evolved Universal Terrestrial Radio Access (E-UTRA), (Rel. 10)," June 2012.
20. Tragos, E., Mihovska, A., Mino-Diaz, E., Karamolegkos, P., Vlacheas, P., & Luo, J. (2009). ccess selection and mobility management in a beyond 3g ran: The winner approach. *Telecommunication Systems*, *42*(3–4), 165–177.
21. Karlsson, J., & Heinegard, J. (1996). Interference Rejection Combining for GSM, Universal Personal Communications. In *IEEE International Conference*, vol. 1, pp. 433–437, October.
22. Brueninghaus, K., Astely, D., Salzer, T., Visuri, S., Alexiou, A., Karger, S., & Seraji G. -A. (2005). Link performance models for system level simulations of broadband radio access systems, vol. 4, pp. 2306–2311, September.
23. Chandrasekhar, J. G. A. V., & Gatherer, A. (2008). Femtocell networks: A survey. *IEEE Communications Magazine*, *46*(9), 59–67.
24. Ulvan, A., Bestak, R., & Ulvan, M. (2013). Handover procedure and decision strategy in LTE-based femtocell network. *Telecommunication Systems, 52*(4), 2733–2748.
25. Roland Fall, K. (2003). A delay-tolerant network architecture for challenged internets, pp. 27–34, August.
26. Han, B., Hui, P., Kumar, V., Marathe, M., Shao, J., & Srinivasan, A. (2012). Mobile data offloading through opportunistic communications and social participation. *Mobile Computing on IEEE Transactions*, *11*, 821–834.
27. Han, B., Hui, P., & Srinivasan, A. (2010). Mobile data offloading in metropolitan area networks. *SIGMOBILE Mobile Computing and Communications Review*, *14*, 28–30.
28. Dimatteo, S., Hui, P., Han, B., & Li, V. O. (2011). Cellular traffic offloading through wifi networks. In *IEEE 8th International Conference on Mobile Adhoc and Sensor Systems (MASS)*, 2011, pp. 192–201.
29. Lee, J., Yi, Y., Chong, S., & Jin, Y. (2013). Economics of wifi offloading: Trading delay for cellular capacity. In *IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2013, pp. 357–362.
30. Liu, S., & Striegel, A. (2012). Casting doubts on the viability of wifi offloading. In *Proceedings of the 2012 ACM SIGCOMM Workshop on Cellular Networks: Operations, Challenges, and Future Design*, Cell Net '12, New York, pp. 25–30.
31. "3GPP TS 23.234: System to Wireless Local Area Network (WLAN) Interworking; System Description (Rel. 10)," March 2011.
32. "3GPP TS 23.327: Mobility between 3GPP-Wireless Local Area Network (WLAN) Interworking and 3GPP Systems (Rel. 10)," March 2011.
33. "IETF RFC 5648: Multiple Care-of Addresses Registration," October 2009.
34. Fouquet, M., Hoene, C., Schläger, M., & Carle, G. (2011). Data collection in future mobile networks. *Telecommunication Systems*, *48*(3–4), 289–300.
35. Chen, W.-M., Chen, W., & Chao, H.-C. (2009). An efficient mobile ipv6 handover scheme. *Telecommunication Systems*, *42*(3–4), 293–304.
36. "3GPP TS 23.829: Local IP Access and Selected IP Traffic Offload (LIPA-SIPTO (Rel. 10)," March 2011.
37. "3GPP TS 24.302: Access to the 3gpp Evolved Packet Core (EPC) via non-3GPP access networks (Rel. 10)," April 2011.
38. Åhlund, C., Brannstrom, R., & Zaslavsky, A. (2006). Traffic load metrics for multihomed mobile IP and global connectivity. *Telecommunication Systems*, *33*(1–3), 155–185.
39. Khan, M., Toker, A., Sivrikaya, F., & Albayrak, S. (2012). Cooperation-based resource allocation and call admission for wireless network operators. *Telecommunication Systems*, *51*(1), 29–41.

Desta Haileselassie Hagos received his B.Sc. from Mekelle University in 2008. He obtained his M.Sc. Degree in Mobile Systems from Luleå University of Technology, Department of Computer Science Electrical and Space Engineering in 2012. Currently, he is a Ph.D. student at the University Graduate Centre (UNIK) of Oslo. His research interests are Software-Defined Networking (SDN), Network Virtualization and Cloud Computing.