Impact of Offloading on the Efficiency of Wireless Access Networks

Adel Agamy[1] · Ahmed M. Mohamed[1]

Abstract

Wireless access networks need to deliver a satisfactory level of Quality of Service (QoS) to their subscribers. The service quality can be measured through various metrics, for example the mean packet delay, average throughput, the jitter and bandwidth. Based on the type of network application, one or the other can be more important. Mobile data offloading has a crucial effect on the efficiency of wireless access networks. Offloading decreases the load on mobile networks (LTE) which frees the band to other users and consequently improves the QoS level. Also data offloading reduces the cost of the downloaded information. In this paper we assess the impact of data offloading on the efficiency of wireless access networks. We employ two popular scheduler algorithms to measure the effect of resource allocation among users. In particular, an analytical model for mixed LTE and WiFi access networks is developed. The model takes into account the role and the effect of various application demands. The model evaluates the wireless access networks behavior when some portions of network service zones are tenanted by different application types. Using the real network simulation NS3, the network performance metrics (the average throughput, delay and packet loss) are investigated to evaluate the network performance under diverse traffic loads.

Keywords Wireless Access Networks · LTE · WiFi Networks · Traffic Modeling · Performance modeling · Offloading

1 Introduction

Networks with diverse access technologies which can deliver a service using a cellular access network and capable of keeping the service when switching to another WiFi access network is called a wireless access networks [40]. The rapid increase in the use of wireless devices results in numerous challenges for wireless access networks design and implementation. Also, the various types of network applications produce different traffic types. Each application generates traffic with different characteristics which leads to different types of traffic in the network. Some applications generate traffic with fixed rate like voice calls; others like video streaming and web browsing generate traffic with variable rate. Applications like files transferring generate traffic with different burst size. These challenges should be considered in the design of the new generation of wireless access networks

due to the constraints on the delay and bandwidth requirements for different network applications [7, 11, 13, 30]. The third Generation Partnership Project (3GPP) presented the LTE characteristics [29], the novel optimized architecture for the radio access and the core networks. Many solutions have been proposed to handle these demands such as, offloading to other access networks like macrocell, microcell, picocell, femtocell and Wi-Fi hotspot. Using complementary access technologies for delivering data to mobile users in cellular network is called offloading. Offloading reduces the time of a user to be served by cellular technology and hence frees band to other users and reduces the cost of downloading data. The most significant solution is offloading to WiFi access points due to their spreading in most smartphones and deployment in wide locations. Also WiFi networks have a lower cost for both users and operators [7, 18, 37]. Through this paper we evaluate the influence of offloading to WiFi access networks under different types of application characteristics on the efficiency of the wireless access networks. The rest of paper is organized as follows. In Sect. 2, we present the related studies to this research problem. Section 3 shows the simulation model for the integration of multiple wireless access networks. Section 4 presents the analytical performance model for wireless access networks. Through

✉ Adel Agamy
a.f.agamy@aswu.edu.eg

Ahmed M. Mohamed
amohamed@aswu.edu.eg

[1] Electrical Engineering Department, Faculty of Engineering, Aswan University, Aswan, Egypt

section 5 we present the discrete event simulation results of the presented model. Section 6 concludes our results.

2 Related Work

The rapid increase in the use of wireless devices leads to a large variation of the network application demands [7]. The target of building the fifth generation access network is to support heterogeneity with multi radio access technology and various options to the users [14] and [37]. Gerhard et al. [41] investigated the performance of the physical and MAC layers of LTE cellular networks with a large class of scheduling policies. They presented a formula for user transmission rate as a function of channel conditions. Also they assumed that each user has an input queue with unlimited buffer size at the MAC layer. Using independent Poisson arrival processes they introduced an approach to get the upper bounds on all polynomial moments of the queue backlog and exponential tail bounds on the tail probability. Hussam et al. [2] introduced a resource allocation scheme for downlink LTE cellular network using queuing theory based on the channel quality information. They considered the channel quality information reporting modes in LTE where the subscriber only report the indices of their best sub bands and an actual channel quality information corresponding to the best sub bands. In Richard et al. [8] studied the problem of self-organized relays in a cellular network. They derived an optimal static closed form solution for resource sharing between base stations, relay stations links and users using Markov decision process. The influence of key system parameters has been investigated, showing the importance of relaying gain. The resources are allocated dynamically considering the stability for infinite buffers and blocking rate, also file transfer time in the presence of admission control. Wang et al. [16] introduced a scheduling scheme for the LTE cellular networks which allocates resources to users based on the queue length and the overflow probability. The model can be used for inelastic traffic and is compatible with 3GPP LTE systems. They used a discrete-time Markov-modulated Poisson process to model each user flow arrival at the queue input. They adjust the allocated bandwidth adaptively according to the average queue length and buffer overflow probabilities. Narcisse et al. [27] introduced a model with process sharing queue model for cellular LTE networks with Poisson arrival process and a round robin scheduling discipline. They studied the performance of LTE networks with different transmission zones that takes into account mobility of users between zones. Their model is based on a multiclass process sharing queue which captures mobility of users between zones through the distribution of the sojourn time a user physically stays in each zone. They assumed round robin scheduler, so resources are shared equally among

active users regardless of the user's location. Kehindeet al. [28] introduced an analytical model for estimating the LTE-based public safety network capacity for voice communications. They used a Markov birth and death process to model the network. They calculated the minimum number of channels required for different levels of users priority to achieve the grade of service. Dima et al. [9] introduced a traffic model for LTE cellular networks which considers different realistic parameters that can be used for LTE network planning and optimization. They generate the traffic profile of each eNodeB in LTE cellular network as a function of a variety of metrics such as signaling, bandwidth, busy hour session attempts and number of bearers, then they planned the LTE cellular network using 15 eNodeB. Konstantin et al. introduced in [32] a multiple servers queuing systems to analyze the subscribers blocking probability in LTE cellular network with a limited amount of resources. They derived a mathematical model for resources allocation in LTE-A cellular networks, evaluated blocking probability and average amount of occupied resources. They assumed that the subscribers arrive according to the Poisson process with rate λ and stay in server an amount of time independent of their arriving process and are exponentially distributed with rate μ. In [6] Naila et al. studied the behavior of the LTE cellular network MAC scheduler and evaluate its performance using discrete time Markov chains. The authors considered two types of traffic according to soap project, the first one is real time traffic like voice and video and the other is non-real time like FTP and HTTP applications. They used lognormal and exponential distributions for the video applications, also they model voice traffic using ON/OFF traffic model with exponential distribution for the silence and active periods with deterministic packet size and inter-arrival time. Spaey et al. [33] introduced an analytical model to study LTE cellular network time varying capacity and its influence on the admission control algorithm with no quality of service differentiation. Tony in [39] addresses the dilemma of the resource sharing by introducing a scheduling algorithm using Markov decision approach and quality of service aware. The algorithm is two-level scheduling to optimize the resource allocation and estate the amount of data sent by the traffic source. Richard et al. [25] studied the performance of the MAC performance of the LTE cellular network using cross-layer algorithm for resource optimization. The algorithm is based on queueing theory and its parameters are transmission rate, packets waiting time in the queue and the size of the queue at the MAC layer. The authors presented a closed form of the previous metrics to evaluate the performance of the algorithm by calculating the bandwidth utility. Also they use ON/OFF traffic source to generate voice data. Xi Li et al. [42] developed a model to assess the performance of the LTE cellular networks. The authors presented a mathematical model for resource scheduling in LTE networks

based on a Markov chain. Each state in the Markov chain represents the total number of subscribers associated to one MCS (Modulation and Coding Scheme). They studied the behavior of the network using two different scenarios. In the first scenario they assumed a single service class and no quality of service. In the second class they considered one quality of service class and two non-guaranteed service classes. They validated the model using fixed file size and calculated the completion time for file download under different scenarios. Li et al. [22] introduced a model for bandwidth dimensioning in LTE cellular network for S1-U interface. The authors presented the model for the elastic traffic and they used the M/G/R-PS process sharing queueing model to analyze the behavior of the system. Their main targets were the bandwidth and network delay. In [21] authors used the same previous model [22] to measure the bandwidth of the eNB required to be handled by the inbound and outbound interfaces to carry elastic traffic given the amount of traffic and the number of users in the cell. The process sharing model M/G/R-PS was discussed in [23]; where it is used for dimensioning cellular networks as well as ADSL networks. A selection mechanism called common radio resource management was presented in [36, 38] to select the used technology based on subscriber needs and the available resources between the sets of available technologies. Most of the previous studies didn't consider the network status to get the decision of the handover, they only depend on the received signal strength. The enormous data traffic in mobile networks forces the operators and users to use WiFi offloading to handle this huge traffic. The offloading techniques in previous studies were divided into two strategies, the first one lets the users free to choose the suitable technology for them. In the second strategy, the function of choosing the network technology is selected based on the operator point of view. In [1, 18, and 29], they introduced the 3GPP release11 to assist user in choosing the best WiFi network. The authors in [20] demonstrated by measurements that the WiFi offloading reduces the cellular network congestion. The author in [35] conducted a survey to estimate the mean user time spent in waiting YouTube stream and downloading a file, they found out that the user can stays "10" minutes for YouTube streaming and three to five hours waiting to download a file for the half of the set

used in their study. Many studies as in [3, 10, 15, 31] have been conducted to evaluate the impact of offloading to WiFi access point from the cost side by decreasing the volume of data used by mobile network and its efficiencyrelated to using femto cell. Also some studies [19, 20, 35] have considered the offloading to WiFi network for only decreasing the usage of mobile networks without taking into account the quality, needs and traffic types. In our proposed study we consider the needs of users through the received throughput and we also take into consideration the application traffic types. We study the behavior of the diverse wireless access networks with three types of application traffic, contentious (fixed size) and sporadic (exponential and Pareto). Also we consider the throughput metric when offloading to different traffic types.

3 Wireless Access Networks Architecture

An access wireless network which offers switching from one access network to another is named a wireless heterogeneous access network [17, 40]. The main architecture of the heterogeneous access network is assumed as shown in Fig. 1 where the topology of the network consists of five zones. All zones are fully covered by LTE cellular network technology. Each zone is also partially covered by WiFi technology. The subscribers connected to WiFi access points start their requests using WiFi technology because of their low cost. During the subscriber downloading sojourn time through WiFi network, he can move to LTE technology if the WiFi technology do not satisfy his needs. As illustrated in the flow chart in Fig. 2 which shows the activities of the user from the start of data request until the request completes. In both cases the subscribers start connecting with the strongest signal. For WiFi technology, the user connects to the closest access point and for LTE technology, he connects to the cell which have the best RSRP (Reference Signal Received Power) [4]. The handover between the cells in LTE technology is performed based on RSRQ (Reference Signal Received Quality) as in Fig. 3 where the handover can occur according to the available throughput. The parameter setting of each technology and traffic characteristics are presented in Tables 1, 2 and 3. The five zones are covered by five evolved nodesB

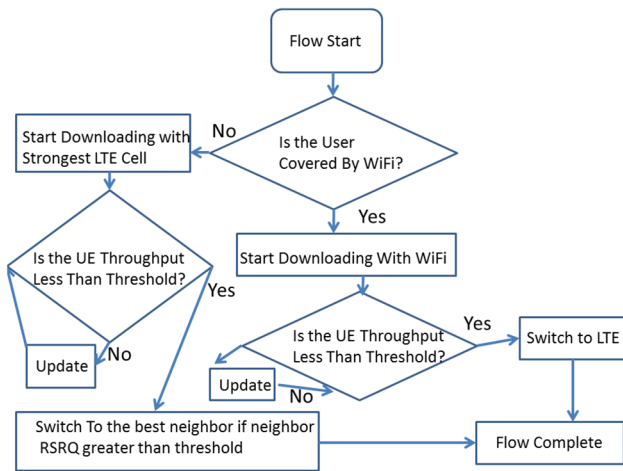**Fig. 1** Heterogeneous access network topology consisting of five zones
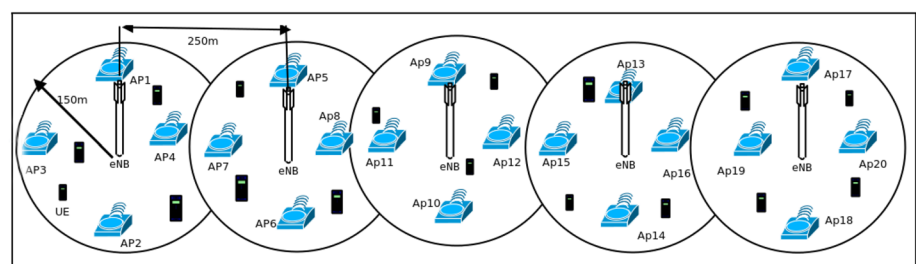
Fig. 2 User activities from the start of data request until the request completes
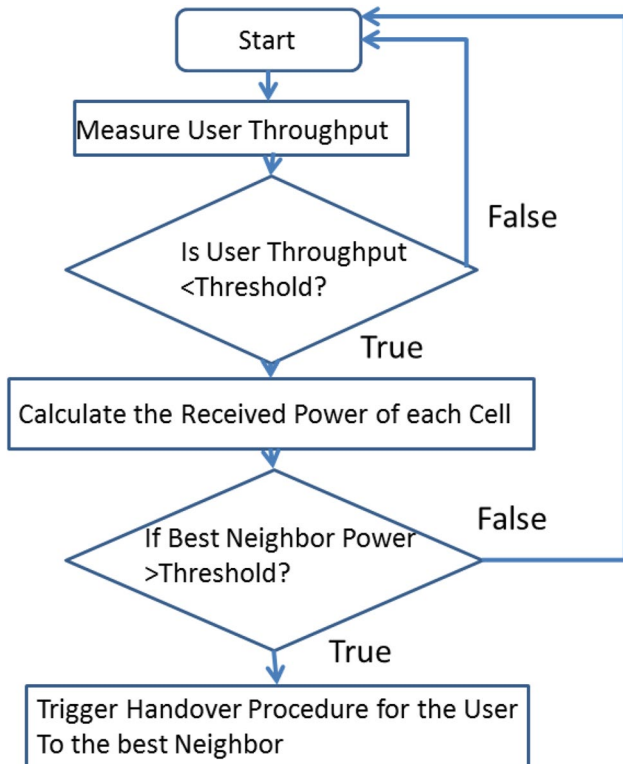

Fig. 3 Handover based throughput flow chart

(eNB). We trigger the traffic requests in the middle cell by a heavy traffic while the edge cells are triggered by normal load to study the impact of handover among the LTE cells. The users are uniformly distributed over the five zones. We assess the performance metrics (accumulative throughput, aggregated cell throughput, end to end delay, loss and Jitter) to evaluate the whole network behavior.

Table 1 LTE Parameters Setting

Frequency: 2.6 GHZ
Bandwidth: 10MHz (50RB)
eNB inter-distance: 250 m
eNB antenna hight: 25m
eNB Transmission Power: 46dBm
LTE noise figure: 5dB
User transmission Power: 24 dB
eNB Scheduler: Proportional Fair(PF)
eNB coverage radios: 150m
eNB Antenna type:Isotropic

Table 2 WiFi Parameters Setting

Frequency: 5GHz
Bandwidth: 20MHz with OfdmRate54Mbps
Adaptive Rate Fallback Algorithm
Access point Antenna: Isotropic (2.5 m height)
Transmission Power: "15" decibel
Receiver Noisefigure: "7"decibel
Receiver Gain:one decibel
Transmitter Gain : one decibel
Propagation Loss model: Friis Loss Model
WiFi Coverage Range:"70" m

Table 3 General Parameters Setting

Coverage area: "0.39" kmm^2
Minimum Throughput : "0.2"Mb/s
RSRQ threshold: 24
Total number of users: 380
user antenna : "1.5" m
number of enodeB: "5"
Access point per cell: "4"
Peak Transmission Rate : "2" Mb/s

4 The Analytical Performance Model

Most of the mobile internet traffic comes from transferring web pages, video and audio streaming, downloading video and audio files; each of them needs a different duration to reach its destination according to the status of the network congestion and the available resources. Most of the traffic inside the mobile network can be considered as a sequence of requests that can be modeled as a flow of bursts. The request rate of each user is highly depends on the network technology type, channel conditions and the set of active users competing for the network resources.

In the WiFi technology, the downloading rate of a user decreases exponentially as the number of users in the network increases and also depends on the channel quality and user distance from the access point. We consider the stream of requests as a flow of a constant or sporadic bursts through the mobile network. We use the Processor-Sharing Queue (PSQ) model where all resources shared by the set of active users equally likely as the basis of the analytical model of the wireless access network [23, 26]. We consider the continuous and sporadic arrivals as inputs for the PSQ model and finally derive an estimate of the delay and request loss probability for any general matrix exponential distribution. In all cases, we assume that all available transmitting resources have a maximum capacity of "R" bit/s.

4.1 Continuous Arrival

In the first case, we consider the continuous arrival requests with λ as the average requests arrival rate and "S" as the mean request size. The total request size by each subscriber can be given by $S_t = \lambda S$ in packets, and the network load utilization is "U" = S_t/R. Using process sharing queue model where all users equally share the available network resources so that each user can download with a rate depends on the number of active users. This rate can be obtained by dividing the total capacity by the average number of active users "R/x". Assume that "x" is the number of active wireless devices each one requests data from an access point or eNB. The size of data is exponentially distributed and we assume that the service time is exponentially distributed. So, the overall system network can be modeled as $M/M/1$ processor sharing queue. So the stationary distribution probability of the simultaneous users can be obtained from the following equation [24]:

$$\pi(x) = \pi(0)U^x, \quad For \quad x = 0, 1, 2..... \tag{1}$$

where $\pi(x)$ is the probability that there are "x" simultaneous active devices in the network. And the average number of simultaneous users can be calculated by:

$$\bar{x} = \frac{U}{1 - U}. \tag{2}$$

where "U" is the load utilization of the network.

In case that there is no restriction on the number of simultaneous users as in WiFi technology the average delay is well known and can be derived from Little's Law as in [24]:

$$MPD = \frac{\bar{x}}{\lambda * S} = \frac{1}{(R(1 - U))}. \tag{3}$$

where $\bar{x}$ come from equation 2, λ is the user arrival rate, S is the total size of requests.

If the number of simultaneous users is restricted as in LTE technology to some value "M" greater than 1, the states of stationary distribution will be limited to "M". The average delay can be obtained as follows [24]:

$$MPD = \frac{x}{\lambda(1 - P_B)s} = \frac{1 - (M + 1)P_B}{(R(1 - U))}. \tag{4}$$

where P_B is the blockage probability, "M" is the number of users that can be served simultaneously. The blockage probability can be obtained as follow [24]:

$$P_B = \frac{U^M}{1 + U + U^2 + ... + U^M}. \tag{5}$$

where "U" is the load utilization of the network.

4.2 Sporadic Arrival

In the sporadic arrival case, we consider first the traffic that can be modeled by the exponential distribution for both the active and inactive periods. Consider "N" as the number of active users. Each user is active for a random period of time then quite in the next time period for a random duration. During the active period the user requests a flow where the flow size is independent, exponentially distributed with mean "S" inactive period durations are independent, exponentially distributed with mean 1/OFF. Then the total traffic load of a single subscriber can be calculated as $S_T = OFF * S$. So the load per subscriber can be given as $\alpha = S_T/R$, and the utilization of the network by $U = N\alpha$. Now assume we have one access point or one eNB and a constant number of connected wireless devices. The resources are equally shared among the devices and this can be modeled by process sharing queueing model. Each device requests a file to download then goes to the inactive state waiting. The inactive state can be modeled as $./M/\infty$ queue. The network can be modeled as a closed queuing network that consists of two queues, an $./M/1$ processor-sharing queue and an $./M/\infty$ queue, with N customer as in Fig. 4. Then the states of stationary probability becomes [24]:

$$\pi(x) = \pi(0)\frac{N!}{(N - x)!}\alpha^x \quad ..... = 0, 1....N. \tag{6}$$

where $\pi(x)$ is the stationary distribution probability of exiting x users in the system and "N" is the total number of active users in the network. For λ the arrival rate is given by:

$$\lambda = \frac{R}{S}(1 - \pi(0)). \tag{7}$$

where R is the capacity of the network and S is the mean request size. In case of no restriction on the number of simultaneous users as in the WiFi technology the average delay can be derived from little's law [24]:
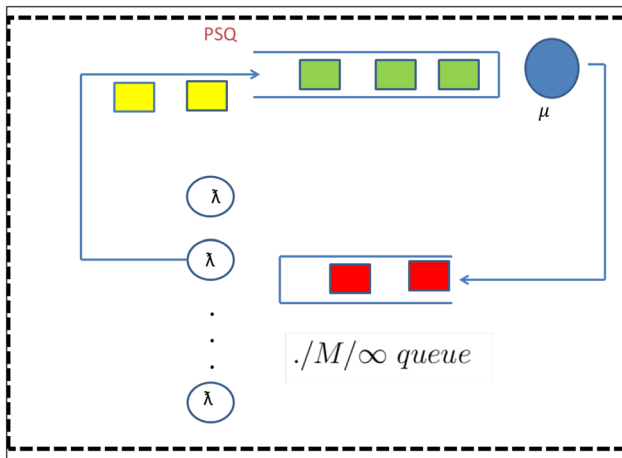
Fig. 4 The Process Sharing Queueing (PSQ) model

$$MPD = \frac{\bar{x}}{\lambda S} = \left(\frac{1}{R}\right) \frac{\sum_{x=1}^{x=N} x \frac{N!}{(N-x)!}\alpha^x}{\sum_{x=1}^{x=N} \frac{N!}{(N-x)!}\alpha^x}. \tag{8}$$

where $\bar{x}$ come from equation 2, S is the mean request size, α is the load for each wireless device and N is the number of wireless devices.

In case of restricted number of simultaneous users as in LTE technology to some constant "M" lesser than "N" in each eNB then the average delay will be [24]:

$$MPD = \frac{\bar{x}}{\lambda S} = \left(\frac{1}{R}\right) \frac{\sum_{x=1}^{x=M} x \frac{N!}{(N-x)!}\alpha^x}{\sum_{x=1}^{x=M} \frac{N!}{(N-x)!}\alpha^x}. \tag{9}$$

where "M" is the number of users that can be served simultaneously.

P_B is the blockage probability and can be derived from as follow [24] :

$$P_B = \frac{\frac{(K-1)!}{(K-1-M)!}\alpha^M}{1 + U + U^2 + \dots + \frac{(K-1)!}{(K-1-M)!}\alpha^M}. \tag{10}$$

where U is load utilization of the network and N is the total number of connected wireless devices.

In the above equations, we modeled the network performance metrics for specific cases of sporadic arrival, the contentious case (the burst degree is zero) and the exponentially ON/OFF case (as a well behaved traffic). Now we modify the model to accept any general sporadic traffic arrival distribution using the matrix exponential distribution approach as in Fig. 5. Therefore, we introduce the ME representation of the Truncated Power tail (TPT) distribution by ME function representation. In this approach, each
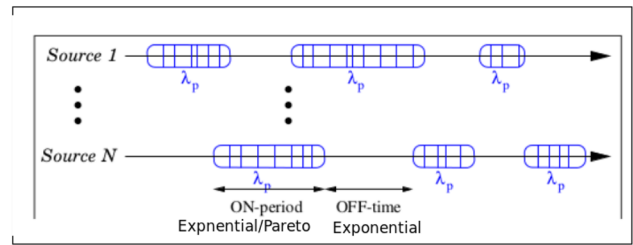

Fig. 5 The N-burst Traffic Model

distribution can be represented by two matrices P and B [24]. The reliability function of TPT distributions is:

$$R(x) := \frac{1-\theta}{1-\theta^T} \sum_{j=0}^{T-1} \theta^j \exp\frac{-\mu x}{\lambda^j} \tag{11}$$

The variable θ can be chosen from the rang $0 < \theta < 1$. For larger values of θ, more phases (T) are necessary to obtain the same heavy tailed characteristics as for lower θ. By controlling The truncation phase parameter "T", we can control the traffic type, for example when "T=1", it represents the exponential distribution and as "T" increases the tail increases (Pareto distribution) and the traffic will be heavy tailed. By using ME approach, the arrival process can be modeled as a network of states (called phases) with exponentially distributed interevent times. Any distribution can be approximated closely by using this approach [24]. Then the representation of arrival process can be defined by two matrices. As in [34] the generator matrix Q_N and the rate matrix L_N for any number of wireless devices N are defined as follows:

$$Q_N = \begin{bmatrix} z_0 & x_0 & & & \\ y_1 & z_1 & x_1 & & \\ .. & .. & .. & .. & .. \\ & & Y_{n-1} & z_{N-1} & x_{N-1} \\ & & & y_N & z_N \end{bmatrix} \tag{12}$$

where $X_i = \mu(i)I^{\oplus i} \otimes P$.
where $i = 1...N-1$
$y_i = (B\epsilon)^{\oplus i}$.
where $i = 1,...N$
$z_i = (-B)^{\oplus i} - \mu(i)I^{\oplus i}$.
where $i = 1,...N$

where B an P are the matrix exponential representation of the active period distribution, μ is the request rate, I is the identity matrix and ϵ is unity vector. The request rate μ is not a constant but it depends on the number of active users. Note that the main diagonal blocks Z_i of Q_N are quadratic but with

growing dimension for the i active sources. Consequently the other two diagonals contain non-quadratic matrices X_i and Y_i. The rate matrix L is: $L_N = \begin{bmatrix} 0 & & & & \\ & R_pI & & & \\ & & 2R_pI^{\oplus 2} & & \\ & & & & \\ & & & & NR_pI^{\oplus N} \end{bmatrix}$

where R_p is the peak transmission rate during active period and I is the identity matrix. The stationary distribution vector π can be partitioned in N+1 sub vectors according to the blocks in Q_N:

$\pi = \pi_0, \pi_1, \ldots \pi_N$

That can be computed by the following iterative approach.

$$\pi_0 = \left[\Sigma_{i=0}^{N} \frac{(\mu(i)\bar{x})^i}{i!} \right]^{-1}. \tag{13}$$

$$\pi_i = \mu(i-1)(\pi_{i-1} \otimes P)(B^{\oplus i})^{-1} . i = 1 \ldots N \tag{14}$$

Then we can treat the system as SM/M/1 queue. the SM/M/1 queue is considered as a Quasi-Birth-Death Process. By calculating the matrix R using the cyclic reduction algorithm [34]. The number of iterations is determined by the speed of convergence of $R_K \to 0$, if the eigenvalues of R are close to 1 then the case for heavy-tailed N-burst approaches the blow-up region.

$A_0 + RA_1 + R^2A_2 = 0$ where : $A_0 = L$, $A_1 = -(B + vI)$, $A_2 = vI$ and $B = Q + L$;

Then we can calculate the average delay.

$$MPD = \frac{1}{A_T}R(I - R)^{-1}\bar{\epsilon} \tag{15}$$

Also we can get the blockage probability in case of M queue size as follows [24]:

$$P_B = \frac{1}{A_T\pi(R^M L)\bar{\epsilon}}. \tag{16}$$

4.3 Hybrid Network Model with Handover

In this model, we consider the two technologies LTE and WiFi. Assuming that the bandwidth of each technology is equally shared between the active wireless devices. In both technologies the single user gets the full bandwidth of the network if he is alone in the network. Thus we can split our network to different areas according to the available resources in each region. A user in a specific area will get the full bandwidth, if he is alone in the coverage area. Each new user request has a probability "P" to be initially downloading from one of the two technologies and "1-P" probability to download from the other technology. We express the total time spent in a specific technology before handover to

another neighboring one by θ_i where i represents the technology. Also for the static user, the user can be handovered to another radio technology if the current one is not satisfying his needs (the throughput is less than threshold). The user request can complete in the initially downloading technology or some of the data is going to downloaded from a neighbor technology (handover). Users currently downloading from technology "i" have a probability P_i to download from the other technology i+1. The sum of all neighbors' handover probabilities is equal to one. Every user requests a burst of traffic to be downloaded from the access point or base station. The total amount of data downloaded from a particular technology per specific user depends on the spending time in each technology which depends on the service time of the technology and its status. We assume that the time spent in each technology is exponentially distributed random variable. The whole network can be seen as multi-class Processor Sharing queue (PSQ) model as Fig. 6. Each class in the model represents the available technology. Customers of each technology arrive to the queue according to sporadic arrival model explained earlier. We can consider the arrival request in each technology as the arrival to this technology in addition to the handover requests from the other technologies. Note that A_{wl} and A_{lw} are the average number of users handovered to the LTE technology and the average number of users handovered to WiFi technology respectively to complete their download (without finishing request). So the arrival rate of each technology can be given as:

$$\lambda_w = A_w + A_{lw}, \quad \lambda_l = A_l + A_{wl}. \tag{17}$$

where λ_w is the arrival rate at WiFi technology, λ_l is the arrival rate at LTE technology, A_{wl} and A_{lw} are the average number of users handover to the LTE technology and the average number of users handover to WiFi technology
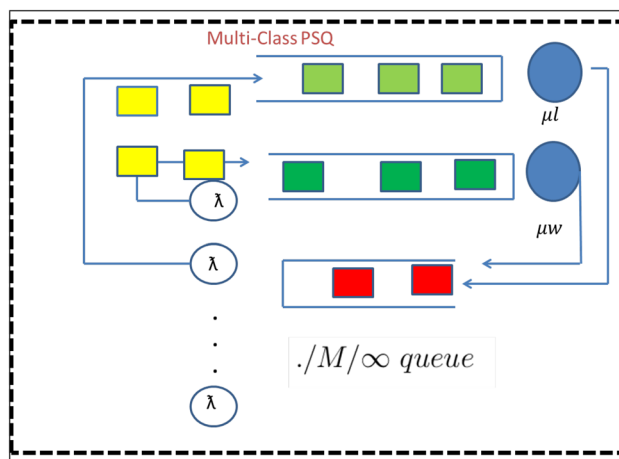

Fig. 6 The Multi-class Process Sharing Queueing Model

respectively. The service time in each technology can be assumed as above by dividing the total capacity over the total data handled by the technology:

$$\mu_w = R_w/D_w, \quad \mu_l = R_l/D_l. \tag{18}$$

where D_w and D_l are the total amount of data transferred by the active users in each WiFi and LTE technologies respectively, μ and R are the service time and data rate of each technology respectively.

Also we can compute the average number of arrival due to handover as follow:

$$A_{lw} = \lambda_l h_w, \quad A_{wl} = \lambda_w h_l. \tag{19}$$

where λ_l and λ_l are the arrival rate in each technology, h_w and h_l are the handover probability from WiFi technology and LTE technology respectively.

Then the average throughput of the user can be obtained from the standard result of multi-class processor sharing [5]:

$$\gamma_w = R_w(1 - U); \quad \gamma_l = R_l(1 - U). \tag{20}$$

where γ_w and γ_l are the average throughput of WiFi and LTE technology respectively, R_w and R_l are the data rate of WiFi technology and LTe technology.

From the time spent in each technology we can get the handover users and the total amount of data transferred by each technology as follow [5]:

$$D_W = \frac{\sigma \gamma_w}{\sigma \theta_w + \gamma_w}; \quad h_W = \frac{\sigma \theta_w}{\sigma \gamma_w + \theta_w}; \tag{21}$$

$$D_l = \frac{\sigma \gamma_l}{\sigma \theta_l + \gamma_l}; \quad h_l = \frac{\sigma \theta_l}{\sigma \gamma_l + \theta_l}; \tag{22}$$

where the σ is total volume of data requested by the wireless devices and θ is the time user spent downloading in each technology. From little's law we can compute the average time spent by the user in both technologies as follows [5]:

$$\bar{T}_t = \frac{\bar{Q}_w + \bar{Q}_l}{A}. \tag{23}$$

where $\bar{Q}_w$ and $\bar{Q}_l$ are the average numbers of active users in WiFi technology and LTE technology respectively.

We assume that the arrival at PSQ input is a combination of multiple sporadic sources (N=3 and N=4 as a case study) each one with a power tail active period distribution and exponential OFF period distribution, the integration of "N" multiple exponential distributions with rate "r" is Nr exponential distribution. Our queueing model is based on the (N-burst/M/1) PSQ model. For small values of "b", it can be approximated by a continuous arrival flow (the first case study) with rate NK while for large values of "b" the arrival process model approaches the bulk arrival that leads to the
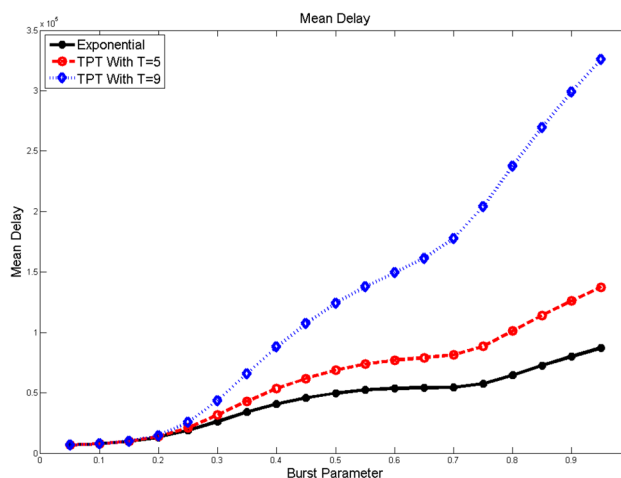

Fig. 7 The average packet delay as a function of traffic burstiness for exponential and power tail distributions
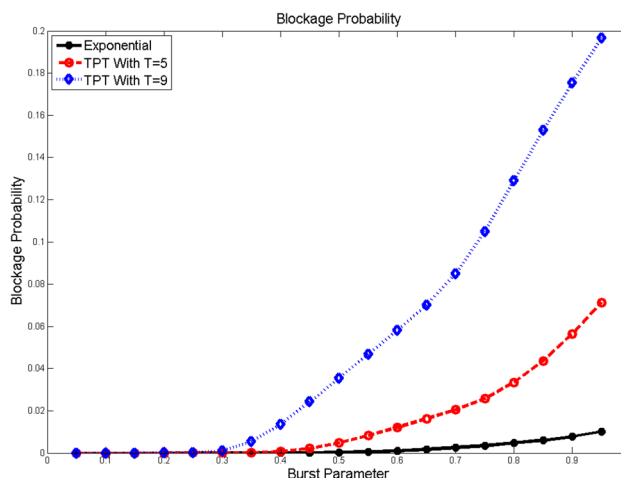

Fig. 8 The blockage probability as a function of traffic burstiness for exponential and power tail distributions

worst performance of the network. Figure 7 presents the average packet delay as a function of burstiness parameter "b". From the figure it is clear that the delay increases significantly with the increase of the burstiness parameter "b", also the delay jumps to a large value at point (b = 0.1) and then starts to increase gradually. Even for the well behaved applications (modeled by the exponential distribution), the packet delay also increases with the increase of the burstiness parameter due to the effect of the burstiness and self-similarity. The more significant packet delay occurs for the applications which follow distributions with large tail. Figure 8 presents the blockage probability as a function of the burstiness parameter "b". The figure shows how the type of the traffic affects the behavior of the loss probability. Clearly the traffic with high variance (long tails) has significant loss

probability than those with small variance (exponential distributions). Both figures show the significance of the model presented in this paper when compared to other models in the literature that use the exponential distribution for modeling the behavior of internet traffic. As our assumption of multiple users with "N=3", we have three jump points as in Fig. 7 for the delay and Fig. 8 for the blockage, but for multiple users the behavior of the delay will be more complex. The exact locations of the transitions points with multiple number of sporadic traffic can be given by the following equation:

$$b = 1 - \frac{K}{R_p} = 1 - \frac{\lambda}{NR_p} \qquad (24)$$

where b is the burst parameter and it ranges from 0 to 1, K is the arrival rate, N is the total connected devices and R_p is the peak transmission rate during active period.

5 Simulations Results

First we study the behavior of LTE technology with one eNB with two different scheduling mechanisms (Round Robin and Proportional Fair). In Fig. 9, we show the behavior of the total eNB throughput for the two popular scheduling algorithms. The settings of the LTE technology parameters are defined in Table 1 through NS3 simulator. Figure 9 shows the maximum capacity of single eNB which approximately 32 *mb/s* and this is confirmed with the standards as in [12]. As can be seen from Fig. 9 where we compare the throughput of the two scheduling algorithms, Proportional fair provides the subscriber close to the eNB with higher throughput while less throughput to subscribers away from
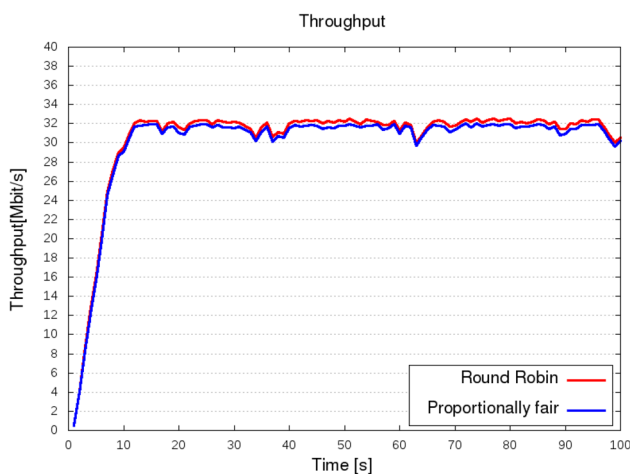
eNB due to weaker signal. Round robin assigns resources periodically to the subscribers without taking into account signal strength so it achieves the best fairness, but it would offer poor performance in throughput specially to edge users.

In Fig. 10, we show the behavior of the throughput of the whole heterogeneous access network. The throughput is presented for three different application traffic types namely fixed-size, exponential, and heavy tailed (Pareto) distributions. As we can notice from Fig. 10 the behavior of the total network throughput for applications that follow heavy tailed distributions is the worst among the three types, while the best throughput behavior is for the applications that follow the fixed-size. The behavior of exponential distribution is better than the heavy tailed distribution and close to the fixed-size distribution. There is clearly a significant degradation in the throughput between the applications that follow the power tail distribution and those that follow the exponential distribution. Figure 11 shows the instantaneous number of simultaneous users that can be served by the network. Figure 11 shows that the total number of users that can be served by the network is highly depend on the application type, where the application which follow heavy tailed distribution can serve smaller number of users when compared to the fixed-size and exponential distributions under the same network condition. The result shows the negative impact of sporadic traffic on the behavior of diverse access networks. From Figs. 10 and 11 we can conclude that the assumption of exponential distribution for nowadays Internet applications may lead to a serious overestimation of the capabilities of the access network. Figs. 12 and 13 show the behavior of the throughput and demand of cell 3. The figures show that the behavior of the throughput and demand of cell 3 is very close to the behavior of the throughput and demand
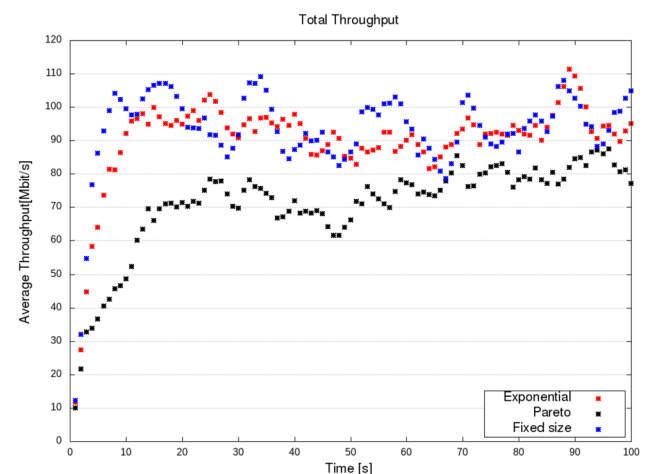

Fig. 9 Wireless access network throughput for proportional fair and round robin schedulers

Fig. 10 Wireless access network throughput for fixed-size, exponential and Pareto distributions
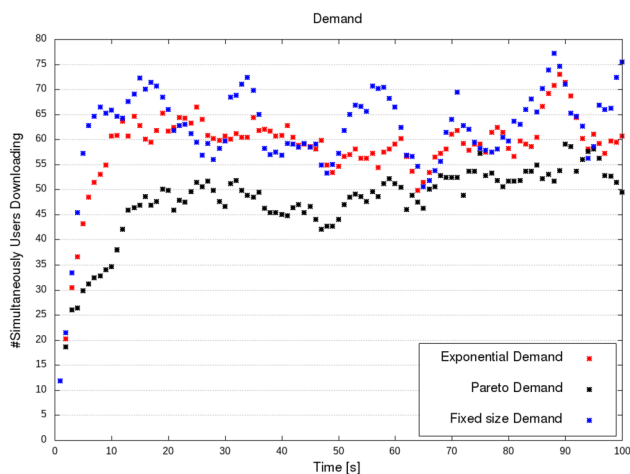
Fig. 11 The instantaneous number of simultaneous users that can be served by the network for fixed-size, exponential and Pareto distributions
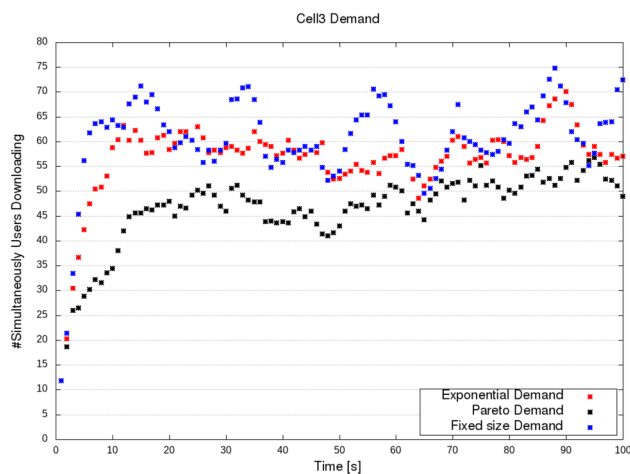

Fig. 13 The instantaneous number of simultaneous users that can be served by cell 3 for fixed-size, exponential and Pareto distributions
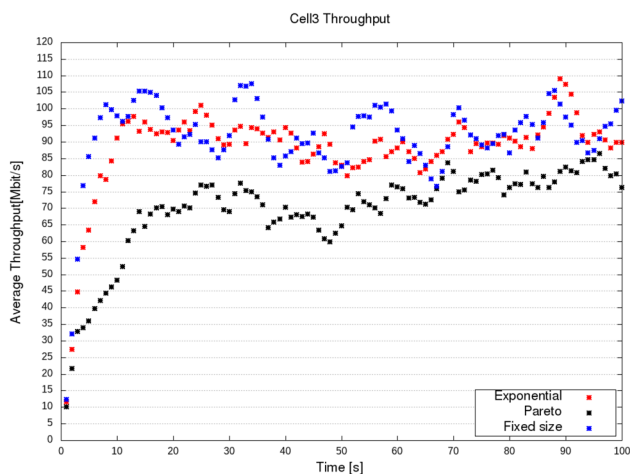

Fig. 12 Cell 3 throughput for fixed-size, exponential and Pareto distributions
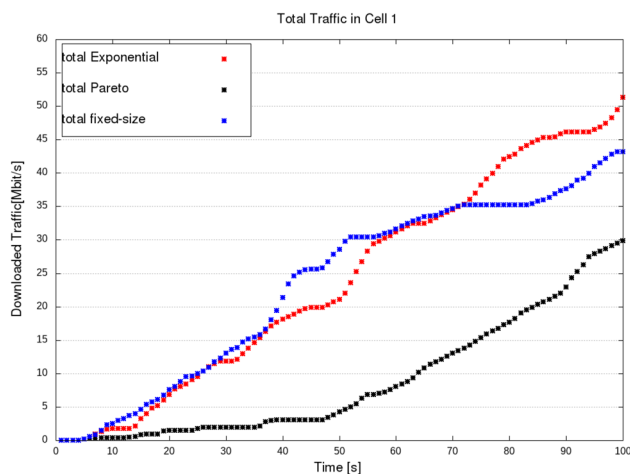

Fig. 14 The total traffic handled in the edge cell 1 for fixed-size, exponential and Pareto distributions

of the whole network. So, one loaded cell can affect significantly the behavior of the whole network. We expect that this poor effect will be higher if more cells use the sporadic traffic arrival. Figures 14 and 15 show the total traffic handled by cell 1 and cell 2. The figures show the impact of the overloaded cell 3 on the neighbor cell 2. As can be seen from Fig. 15 the total accumulative traffic handled by cell 2 is greater than the total accumulative traffic of the edge cell 1 in Fig. 14, this is due to the handover between the cell 3 and cell 2. The figures show the effect of the three different application distributions. Figures 16 and 17 show the total traffic handled by cell 4 and the edge cell 5. The results in Figs. 16 and 17 matches the same results of cell 1 and cell 2. The aggregated traffic over all

five cells is illustrated in Fig. 18. The figure shows that the total traffic served by the whole network for the fixed size application is larger than other distributions (Exponential and Pareto). Also from the Fig. 18 the least amount of traffic handled by the network is in case of the applications that follow the Pareto distribution. In Fig. 19, we can observe that the total data served by each technology (WiFi and LTE) over all five cells. The WiFi technology has a significant share of the handled traffic in the network. The figure clearly shows how the WiFi can improve the cost effective use of the network in case of using an effective offloading mechanism. So, we can conclude that the use of offloading mechanism can increase the efficiency of the wireless access networks.
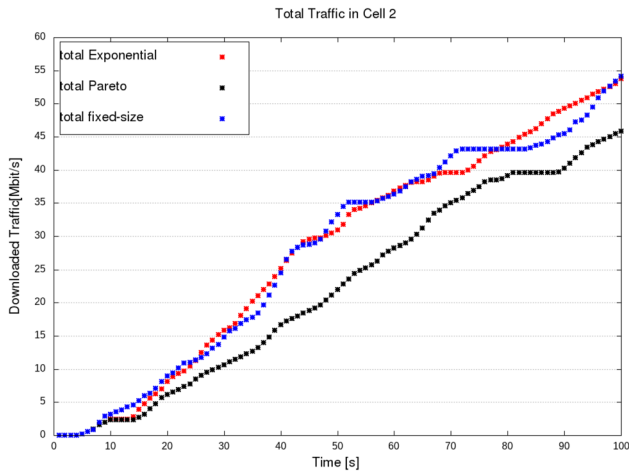
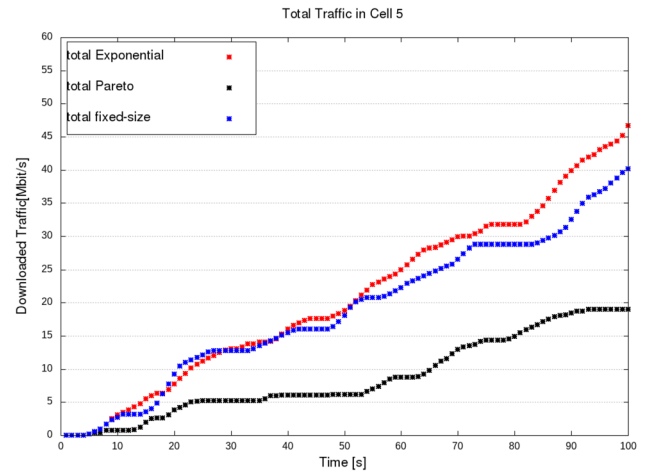**Fig. 15** The total traffic handled in the neighbor cell 2 for fixed-size, exponential and Pareto distributions

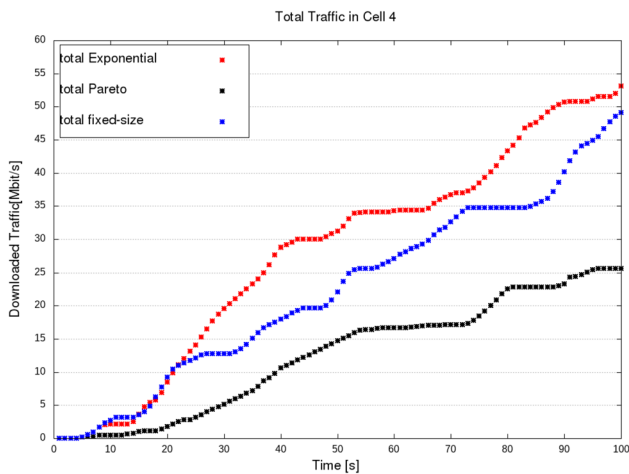**Fig. 17** The total traffic handled in the edge cell 5 for fixed-size, exponential and Pareto distributions

Fig. 16 The total traffic handled in the neighbor cell 4 for fixed-size, exponential and Pareto distributions
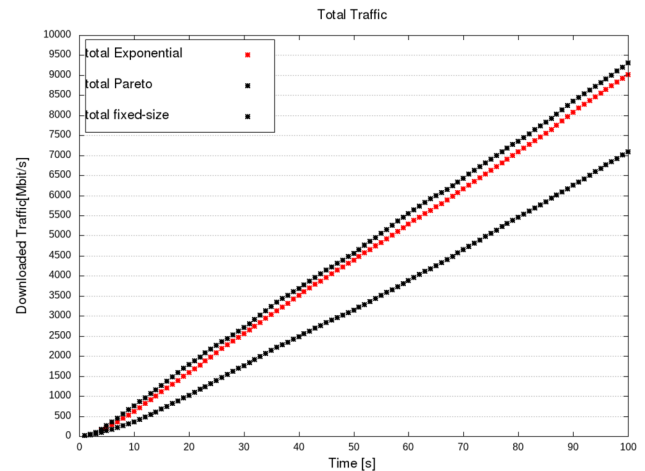

Fig. 18 Wireless access network handled traffic for fixed-size, exponential and Pareto distributions

6 Conclusions and Future work

The use of the wireless access networks is becoming more vital nowadays. Wireless access networks should deliver an acceptable level of quality of service to their users. In this paper, we developed an analytical model for the mixed LTE and WiFi access networks based on queuing theory. Our heterogeneous wireless access networks model estimates the effect of three types of user applications traffic analytically and by simulation. Through this paper we evaluate the impact of offloading and application type on the efficiency and performance of the heterogeneous

access networks. Also we study the effect of different schedulers algorithms to show how LTE technology allocate resources to the subscribers. The behavior of diverse access network evaluated analytically using PSQ queuing model with sporadic arrival. We validate the model using discrete event simulator NS3 to reflect the credibility of the model. The extension of our work is to include the analytical analysis of various connectivity options to improve the edge user connectivity and poor coverage area through device to device communications using WiFi direct and Bluetooth technologies. Also, we are going to study the impact of mixing different types of traffic on the behavior of the diverse access network.
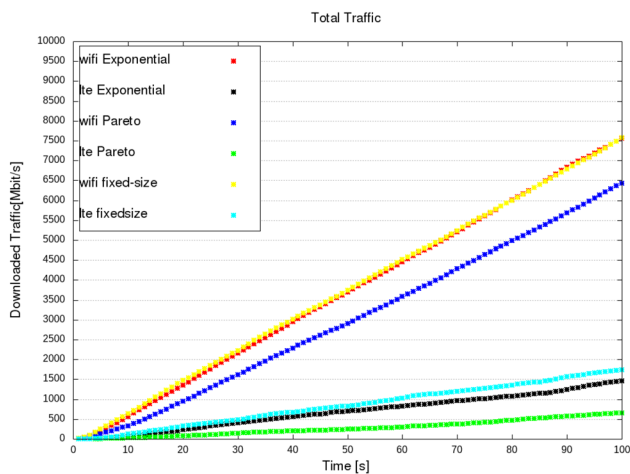
Fig. 19 Wireless access network handled traffic for each Technology

References

1. 3GPP, Architecture enhancements for non-3gpp accesses. 3rd Generation Partnership Project, Sophia-Antipolis Cedex, France. http://www.3gpp.org/ftp/Specs/html-info/23402.htm, 2015.

2. H. Ahmed, K. Jagannathan and S. Bhashyam, Queue-aware optimal resource allocation for the lte downlink with best m subband feedback, *IEEE Trans. Wireless Communication*, Vol. 14, No. 9, pp. 4923–4933, 2015.

3. A. Aijaz, H. Aghvami and M. Amani, A survey on mobile data offloading: technical and business perspectives, *IEEE Wireless Commun.*, Vol. 20, No. 2, pp. 104–112, 2013.

4. At A., ns-3 documentation. http://www.nsnam.org/doxygen/index.html.

5. B. Baynat, R. M. Indre, N. Nya, P. Olivier, A Simonian, Impact of mobility in dense lte-a networks with small cells. 2015 IEEE 81st Vehicular Technology Conference (VTC Spring), pp. 1–5, 2015.

6. N Bouchemal, N Izri, S Tohme, Mac-lte scheduler modeling and performance evaluation in lte network. In: Personal, Indoor, and Mobile Radio Communication (PIMRC), 2014 IEEE 25th Annual International Symposium on, pp. 1007–1012. IEEE (2014)

7. Z. Chen, S. Cheng, Computation offloading algorithms in mobile edge computing system: a survey. In: International Conference of Pioneering Computer Scientists, Engineers and Educators, pp. 217–225. Springer, 2019.

8. R. Combes, Z. Altman, E. Altman, Self-organizing relays in lte networks: queuing analysis and algorithms. In: Proceedings of the 7th International Conference on Network and Services Management, pp. 99–106. International Federation for Information Processing, 2011.

9. D. Dababneh, M. St-Hilaire and C. Makaya, Data and control plane traffic modelling for lte networks, *Mobile Networks and Applications*, Vol. 20, No. 4, pp. 449–458, 2015.

10. S. Dimatteo, P. Hui, B. Han, V.O. Li, Cellular traffic offloading through wifi networks. In: Mobile Adhoc and Sensor Systems (MASS), 2011 IEEE 8th International Conference on, pp. 192–201. IEEE, 2011.

11. M. O. Farooq, C. J. Sreenan, K.N. Brown, Research challenges in 5g networks: a hetnets perspective. In: 19th International Conference on Innovations in Clouds, Internet and Networks (ICIN 2016). IFIP Open Digital Library, 2016.

12. 3rd Generation Partnership Project, G.: Technical specification group radio access network: evolved universal terrestrial radio access (e- utra): User equipment (ue) radio transmission and reception. Tech. rep., TS 36.101 v8.17.0 Release 8, http://www.3gpp.org, 2008.

13. A. Gupta and R. K. Jha, A survey of 5g network: architecture and emerging technologies, *IEEE access*, Vol. 3, pp. 1206–1232, 2015.

14. F. Hantera, F. Digham, H. Tawfik, Lte-wifi offloading with realistic attributes. In: 2019 IEEE Wireless Communications and Networking Conference (WCNC), pp. 1–6. IEEE, 2019.

15. L. Hu, C. Coletti, N. Huan, I. Z. Kovács, B. Vejlgaard, R. Irmer, N. Scully, Realistic indoor wi-fi and femto deployment study as the offloading solution to lte macro networks. In: Vehicular Technology Conference (VTC Fall), 2012 IEEE, pp. 1–6. IEEE, 2012.

16. W. Ke, L. Xi, J. Hong, et al., Traffic-based queue-aware scheduling for 3g pp lte system, *The Journal of China Universities of Posts and Telecommunications*, Vol. 21, No. 2, pp. 63–68, 2014.

17. M. A. Khan, S. Leng, W. Xiang, K. Yang, Architecture of heterogeneous wireless access networks: a short survey. In: TENCON 2015-2015 IEEE Region 10 Conference, pp. 1–6. IEEE, 2015.

18. L. Kuang, T. Gong, S. OuYang, H. Gao, S. Deng, Offloading decision methods for multiple users with structured tasks in edge computing for smart cities. Future Generation Computer Systems, 2020.

19. M. Lacage, M. H. Manshaei, T. Turletti, Ieee 802.11 rate adaptation: a practical approach. In: Proceedings of the 7th ACM international symposium on Modeling, analysis and simulation of wireless and mobile systems, pp. 126–134. ACM, 2004.

20. K. Lee, J. Lee, Y. Yi, I. Rhee, S. Chong, Mobile data offloading: how much can wifi deliver? In: Proceedings of the 6th International Conference, p. 26. ACM, 2010.

21. X. Li, U. Toseef, T. Weerawardane, W. Bigos, D. Dulas, C. Goerg, A. Timm-Giel, A. Klug, Dimensioning of the lte access transport network for elastic internet traffic. In: Wireless and Mobile Computing, Networking and Communications (WiMob), 2010 IEEE 6th International Conference on, pp. 346–354. IEEE, 2010.

22. X. Li, U. Toseef, T. Weerawardane, W. Bigos, D. Dulas, C. Goerg, A. Timm-Giel, A. Klug, Dimensioning of the lte s1 interface. In: Wireless and mobile networking conference (WMNC), 2010 Third Joint IFIP, pp. 1–6. IEEE, 2010.

23. K. Lindberger, Balancing quality of service, pricing and utilization in multiservice networks with stream and elastic traffic, *Proc. ITC*, Vol. 16, p. 1999, 1999.

24. L. Lipsky, *Queueing theory: a linear algebraic approach, 2nd ed*. Springer Science & Business Media, 2008.

25. R. Musabe and H. Larijani, Low complexity cross-layer scheduling and resource allocation for voip in 3g lte, *International Journal on Advances in Telecommunications*, Vol. 6, No. 1 & 2, p. 2013, 2013.

26. R. Núnez-Queija, *Processor-sharing models for integrated-services networks*. Ph.D. thesis, PhD thesis, Eindhoven University of Technology, 2000.

27. N. Nya, B. Baynat, A ps queue model for lte macrocells taking into account mobility of users. In: Proceedings of the 2015 Workshop on Wireless of the Students, by the Students, & for the Students, pp. 44–46. ACM, 2015.

28. K. O. Olasupo, I. Kostanic and T. O. Olasupo, Analytical modeling of lte-based network capacity for public safety communications, *Universal Journal of Communications and Network*, Vol. 4, No. 3, pp. 37–46, 2016.

29. M. Olsson, P. Bleckert, M. Buchmayer, A. Norefors, J. Vikberg, Access network discovery and selection function, andsf, node distributing closed subscriber group, csg, information. US Patent 8,437,743, 2013.

30. N. Panwar, S. Sharma and A. K. Singh, A survey on 5g: The next generation of mobile communication, *Physical Communication*, Vol. 18, pp. 64–84, 2016.

31. M. H. Qutqut, F. M. Al-Turjman, H. S. Hassanein, Hof: a history-based offloading framework for lte networks using mobile small cells and wi-fi. In: Local Computer Networks Workshops (LCN Workshops), 2013 IEEE 38th Conference on, pp. 77–83. IEEE (2013)

32. K. Samouylov, E. Sopin, O. Vikhrova, Analyzing blocking probability in lte wireless network via queuing system with finite amount of resources. In: International Conference on Information Technologies and Mathematical Modelling, pp. 393–403. Springer, 2015.

33. B. Sas, E. Bernal-Mor, K. Spaey, V. Pla, C. Blondia, J. Martinez-Bauset, An analytical model to study the impact of time-varying cell capacity in lte networks. In: Wireless and Mobile Networking Conference (WMNC), 2011 4th Joint IFIP, pp. 1–8. IEEE, 2011.

34. H. P. Schwefel, L. Lipsky, Performance results for analytic models of traffic in telecommunication systems, based on multiple ON-OFF sources with self-similar behavior. Teletraffic science and engineering pp. 55–65, 1999.

35. S. Sen, C. Joe-Wong, S. Ha, J. Bawa, M. Chiang, When the price is right: enabling time-dependent pricing of broadband data. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2477–2486. ACM, 2013.

36. A. Serrador, L. M. Correia, A model to evaluate vertical handovers on jrrm. In: Personal Indoor and Mobile Radio Communications (PIMRC), 2010 IEEE 21st International Symposium on, pp. 2151–2155. IEEE, 2010.

37. M. Sheng, Y. Wang, X. Wang, J. Li, Energy-efficient multiuser partial computation offloading with collaboration of terminals, radio access network, and edge server. IEEE Transactions on Communications, 2019.

38. A. Tolli, P. Hakalin, H. Holma, Performance evaluation of common radio resource management (crrm). In: ICC 2002. IEEE International Conference on Communications, vol. 5, pp. 3429–3433. IEEE, 2002.

39. T. Tsang, Performance analysis for lte networks with markov decision process. Cyber Journals: Multidisciplinary Journals in Science and Technology, Journal of Selected Areas in Telecommunications, Vol. 3, No. 8, 2013.

40. C. X. Wang, F. Haider, X. Gao, X. H. You, Y. Yang, D. Yuan, H. Aggoune, H. Haas, S. Fletcher and E. Hepsaydir, Cellular architecture and key technologies for 5g wireless communication networks, *IEEE Communications Magazine*, Vol. 52, No. 2, pp. 122–130, 2014.

41. G. Wunder and C. Zhou, Queueing analysis for the ofdma downlink: throughput regions, delay and exponential backlog bounds, *IEEE Transactions on Wireless Communications*, Vol. 8, No. 2, pp. 871–881, 2009.

42. Y. Zaki, T. Weerawardane, X. Li, C. Görg, Lte radio schedulers analytical modeling using continuous time markov chains. In: Wireless and Mobile Networking Conference (WMNC), 2013 6th Joint IFIP, pp. 1–10. IEEE, 2013.

Adel Agamy Received his M.Sc. and B.Sc. from Electrical Engineering Department, Aswan University in 2012, 2007 respectively and the Ph.D. in computer engineering from Aswan university in 2020. He is currently an Assistant professor in Faculty of Engineering, Aswan University. His fields of interest: traffic modeling, wireless communication, cellular network and computer network.

Ahmed M. Mohamed Received the B.Sc. in electrical and computer engineering from Assiut university, Egypt, in 1994, the M.Sc. in computer science and engineering from university of Connecticut, USA in 2001 and the Ph.D. in computer science and engineering from university of Connecticut, USA in 2004. He is currently an assistant professor at electrical engineering department, Aswan faculty of engineering, Aswan University. His research interests include, performance modeling, queueing analysis, distributed systems, computer networks and operating systems.