Mobile Netw Appl (2015) 20:449–458
DOI 10.1007/s11036-015-0631-2

Data and Control Plane Traffic Modelling for LTE Networks

Dima Dababneh[1] · Marc St-Hilaire[1] · Christian Makaya[2]

Abstract LTE networks constitute a major evolution toward the All-IP concept; therefore, an efficient design/planning of these networks is crucial for providing various IP-based services such as VoIP and Rich Communication Services (RCS). Also, due to the heterogeneity of access technologies supported by the mobile core network, referred to as the Evolved Packet Core (EPC), the planning aspect becomes very important and is still a challenging task. Traffic modelling is an important part of the whole network planning process. Previous traffic modelling approaches tend to collect measurements regardless of how time consuming it is considered, while others find it easier to assume or predict values. The traffic model proposed in this paper considers different realistic parameters that can be used for LTE network planning and optimization. In fact, previous LTE planning work did not include all the parameters needed for the network planning process; their focus was mainly on bandwidth where other parameters were not considered. Therefore, it was necessary to come up with reasonable traffic profiles that take into consideration a variety of practical aspects such as signalling, bandwidth, busy hour session attempts and number of simultaneous EPS (Evolved Packet Systems) bearers. The traffic model proposed in this paper is considered a very important part of the network planning process for the EPC. The proposed solution is beneficial for Mobile Network Operators (MNOs) while they are in the deployment phase of 4G/LTE networks.

✉ Marc St-Hilaire
marc_st_hilaire@carleton.ca

[1] Department of Systems and Computer Engineering, Carleton University, Ottawa, Canada

[2] IBM Research, Yorktown Heights, NY, USA

Keywords Long Term Evolution (LTE) · Cellular networks · Traffic model · Network planning · Evolved Packet Core (EPC)

1 Introduction

During the recent years, 3G networks were overwhelmed by the amount of growth in services and applications such as media streaming. Hence, the fourth generation (4G) was introduced and two competing standards were proposed: WiMAX and Long Term Evolution (LTE). It eliminated circuit switching and utilized packet switching efficiently over the Internet to provide users with better performance. Despite the fact that WiMAX access technology had a short lifespan, LTE has been widely adopted by MNOs as 4G technology. LTE provides an increased bandwidth and spectrum efficiency allowing more applications to be used such as mobile online gaming, HD voice, and high quality mobile video streaming.

The LTE network architecture is presented in Fig. 1. The LTE network consists of two parts; the access network consisting of the enhanced NodeB (eNodeB or eNB) and the core network consisting of the Mobility Management Entity (MME), the Serving Gateway (SGW), the Home Subscriber Server (HSS), the Packet Data Network (PDN) Gateway (PGW) and the Policy Control and Charging Rules Function (PCRF). In order to be able to understand the traffic model, it is important to understand the functions of each network elements and the relationships between them. The SGW is the local mobility anchor that holds data when the UEs are moving between eNodeBs during handover, and it deals with the user plane. On the other hand, the MME is the control node that is responsible for the signalling between the UE and the core network. It plays a key role in initiating and maintaining the EPS bearers, managing connections, distributing the

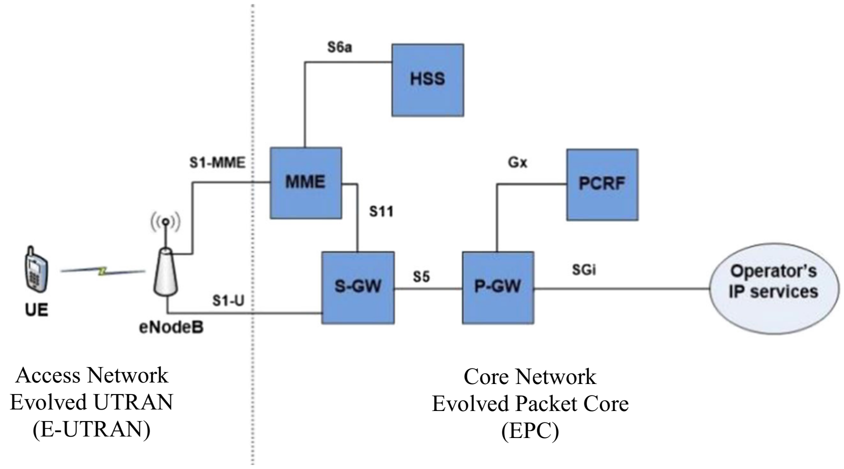

Fig. 1 Typical LTE network architecture

paging messages to the eNBs, providing security, mobility control for users in idle state, and protecting signalling integrity and ciphering. The MME and the SGW are the connecting points between the radio part and the EPC. The PGW is the interconnection point between the EPC and the external IP networks (e.g., Internet or MNO's IP services such as IP Multimedia Subsystem) and it is mainly in charge of assigning and distributing the IP addresses for the UE, besides enforcing the QoS and flow based charging. It also has the ability to work as a mobility anchor for internetworking with non-3GPP technologies like High Rate Packet Data (HRPD) (aka 1xEV-DO) and WiFi. The PGW is considered the default gateway since it performs packet filtering and lawful interception which includes analysing the signalling data in addition to the network management information. The HSS holds dynamic information to keep track of the MME identities to which users are connected. The HSS also includes data for the users' subscription such as the QoS profile and any roaming access restrictions. It also plays a role in authentication and security due to its ability to integrate the Authentication Center (AuC) which formulates security keys and authentication vectors. The PCRF and the PCEF (Policy Control and Charging Enforcement Function) are defined in the so called Policy Control and Charging (PCC) architecture. The PCEF is usually collocated with the PGW.

Understanding the connections between the elements is as important as understanding the functions of each element. The S1-U interface connects the eNB and the SGW for user plane traffic (i.e., bearers' tunnelling, inter-eNB handover), and the S1-MME interface connects the eNB with the MME for management and control processes. The S11 interface connects the MME to the SGW, the S6a interface connects the MME and the HSS, and the S5 interface connects the SGW to the PGW. Moreover, the Gx interface connects the PGW (PCEF) to the PCRF, and the SGi is the interface between the PGW and the packet data network such as Internet or IP Multimedia Subsystem (IMS).

The LTE architecture presented in Fig. 1 shows the different elements considered in the planning process, and in order to design an optimal network, sophisticated planning tools are required. One crucial input of these planning tools is the traffic model. Without a good traffic model, it is difficult to accurately plan the network. In fact, traffic measurement is a major factor for network planning and design. Traffic can be defined as the amount of data and signalling carried over a link for a given period of time. In LTE, there is a classification based on the delay sensitivity that divides LTE traffic into 4 different classes [1]: conversational class, streaming class, interactive class, and background class. The conversational class is considered the most delay sensitive since it carries real-time traffic such as Voice over IP (VoIP) and video conferencing whereas the streaming class is considered less delay sensitive, and it generally carries traffic for streaming purposes such as streaming movies. Regarding the other two types of traffic, the interactive class (e.g., web browsing) is delay insensitive but not as much as the background class since it is considered the most delay insensitive class (e.g., background e-mail downloading).

The goal of this paper is to provide network planners with a new traffic model for LTE networks. The traffic model we introduce in this paper considers different realistic parameters, unlike previous models that use time consuming or predicted measurements. More precisely, we propose a traffic modelling framework to generate reasonable traffic profiles taking into consideration various practical aspects such as signalling, bandwidth, busy hour session attempts for voice and data and number of simultaneous bearers. In addition to that, an example of how the traffic model is used in the network planning process is explained.

The rest of the paper is organized as follow. Related work is discussed in Section 2 where we present different ways that deal with traffic dimensioning. Section 3 describes the proposed traffic model for LTE networks whereas Section 4 presents the simulation results along with an example of how the traffic model can be used in the network planning process. Finally, the conclusion of the paper is provided in Section 5.

2 Related work

There are two major types of traffic: elastic traffic and real time traffic. Elastic traffic, such as web browsing and FTP, is generated by non-real time applications and carried over TCP. On the other hand, real time traffic, such as streaming, conferencing and VoIP, is very sensitive to delay and require specific requirements to be transmitted.

Considering the two different types of traffic mentioned above, Li et al. [2] propose two different models for dimensioning the bandwidth for the S1-U interface in LTE networks given the amount of traffic and the number of users in the cell [2]. The model suggested for elastic traffic is based on the M/G/R-PS (M/G/R-Processor Sharing) model, and it is used to measure the mean time or throughput for TCP flows; whereas the model suggested for real time traffic uses the M/D/1 queuing model which estimates the network delay and performance. Bandwidth was the main concern in this paper, but signalling and control parameters were not taken into consideration. The M/G/R-PS model discussed in [2] is also used in [3] to dimension the bandwidth of elastic traffic for LTE networks. This model measures the bandwidth of the eNB required to be handled by the inbound and outbound interfaces to carry elastic traffic. The model guarantees end-to-end QoS by following the theory of process sharing which characterizes the traffic at the flow level, and the two main QoS concerns to be guaranteed are throughput and delay. The model is capable of characterizing the TCP traffic assuming each user has an individual flow for Internet services. The basic M/G/R-PS model is discussed in [4]; it is applied for dimensioning mobile networks as well as ADSL networks. The elastic traffic acts like a processor sharing system because all elastic traffic flows sharing the same link share the same amount of bandwidth and other resources.

Checko et al. [5] developed a traffic model based on predicted traffic values for 2015 in order to dimension the LTE backhaul network using three capacity planning methods: i) a delay based approach, ii) a dimensioning formula-based approach, and iii) an overbooking factor-based approach. The total amount of mobile data traffic predicted for 2015 is equal to 6,253,920 terabytes (TB) resulted by different applications such as video, web-browsing, Peer-to-Peer (P2P), VoIP, Machine-to-Machine (M2M) communications, and gaming. Based on the forecasted values, the average user will transmit and receive 852 MB of data per busy hour [5]. The delay based approach allows for increasing the capacity as much as needed as long as the delay requirements are satisfied. For the formula based approach, it calculates the bandwidth needed to support a number of users based on their peak aggregated throughput. Finally, the overbooking factor-based approach takes into consideration the probability of having the connection in an active mode, and states that certain users are assigned capacities lower than the sum of their required capacities due to the fact that not all users are using all of their network resources.

Jailani et al. [6] performed a research study in Malaysia to collect data using the Network Performance Optimizer (NPO) tool; in particular they used traffic counters and indicators to perform their study. The paper provides a dimensioning approach for LTE network based on the available LTE voice traffic taking the busy hour traffic as the best representation to evaluate the network performance and perform network dimensioning. Unfortunately, the approach presented only deals with speech traffic and does not consider signalling, video or other applications.

In summary, the main focus of all previous works was related to bandwidth without considering different parameters such as signalling, Busy Hour Session Attempt (BHSA), or Evolved Packet System Bearers (EPSB). Furthermore, none of these models proposed methods to generate traffic; they basically dimensioned specific interfaces based on a given traffic. As a result, the traffic generated using the models described above may not be sufficient for the proper planning and design of LTE networks. Consequently, it is important to develop a tool that generates traffic taking into account different realistic aspects (i.e., bandwidth, EPSB, signalling, and BHSA). In the next section, we introduce a traffic model that considers different traffic parameters for LTE networks.

3 Traffic model

In order to be able to provide reasonable traffic, services are assumed to be asymmetric and the highest amount of traffic during busy hour is taken into consideration. In addition to that, there are certain important traffic parameters that need to be considered:

- *The number of subscribers*: This parameter represents the total number of subscribers that are currently served by a given eNB.
- *The number of attached subscribers (ASUB) during the Busy Hour (BH)*: This parameter represents the number of LTE subscribers that were able to have a successful connection with the PGW along with a successfully established default EPS bearer and successfully allocated IP address. BH is known to be the busiest 60 min period of the day, in which the total traffic is the maximum throughout the day.
- *Busy Hour Data Session Attempt (BHDSA)*: This parameter represents the number of data sessions attempted during the busy hour, and it is one of the main methods to measure the capacity of the network.
- *Busy Hour Voice Session Attempt (BHVSA)*: This parameter represents the number of voice sessions attempted during the busy hour.

- *Bandwidth required for bearer sessions (BW)*: This parameter characterizes the amount of data rate required for the users' services.
- *Simultaneous Evolved Packet System Bearer (SEPSB)*: This parameter shows the number of EPS bearer sessions occurring simultaneously in a busy hour. The EPSB is an established end-to-end connection between the UE and the PGW to provide the users with the Internet services they need.

The subscriber traffic profile provided by the operator and presented in Table 1 is used with the planning parameters presented in Table 2 to compute the eNBs traffic profile. It is important to note that these values can be changed to better fit the specific requirements of different mobile network operators.

The highest amount of traffic during the day occurs during the busy hour, and it is calculated using Eq. (1). The average BH usage (BHU) per subscriber is measured in bits/busy hour. Working days traffic ratio (WDR) represents a percentage of the amount of traffic that occurs during working days (WD), and busy hour traffic ratio (BHR) resembles a percentage of the amount of traffic that occurs during a busy hour.

$$BHU = \frac{MU * WDR * BHR}{WD} \tag{1}$$

There are different techniques in LTE that support voice [15, 16] such as Voice over LTE (VoLTE) and Circuit Switched Fallback (CSFB). The VoLTE initiative was announced to develop a framework that supports voice over LTE using IMS [17]. Considering the VoLTE, there are several factors that control the voice bandwidth [18]: Codec and sampling period, IP header, transmission medium, and silence suppression. Adaptive Multiple Rate (AMR) codec increases the voice capacity and it uses multiple voice encoding rates ranging from 4.75 to 12.2 kbps [19, 20], in this paper we use AMR equal to 12.2 kbps.

The average throughput of the S1-U interface per subscriber (S1UBW) is calculated using Eq. (2). The burstiness factor (*B*) is taken into consideration because the S1-U interface, connecting the eNB with the SGW, carries different types of traffic with different data rates. Burstiness [21] is defined as a representation of a group of packets with shorter gaps between other packets being handled before or after. The Handover Ratio (HO) is also considered due to the fact that the SGW is considered as the mobility anchor for inter-eNB handovers. Certain types of applications and services may require packet retransmission in case of failure; therefore Retransmission Factor (RTF) is also included in this equation. The main two parts of the equation are controlled by the Voice Activity Factor (VAF) to ensure calculating the period in which voice is active, and other periods where other data applications are being handled. In addition, voice data constant (VDC) represents the amount of data needed by the AMR codec for transmission, and since data is transmitted over IP, the IP overhead (*IPov*) is taken into account.

Table 1 Subscriber traffic profile

Average rate	25Mbps
Monthly usage (MU)	2GB/month/sub

Table 2 Planning parameters

Adaptive multiple rate [7, 8]	12.2 kbps
Mean session time (MST) [9]	180 s
Handover ratio (HO) [10]	0.4
IP overhead percentage (IPov)	50 %
Dense area attached subscriber ratio (ASR)	0.9
Active BH EPSB ratio (BR)	0.5
Average EPSB session duration (BSD) [11]	900 s
Retransmission factor	0.25
S1U utilization factor (S1UF) [12]	0.8
Working days per month (WD) [13]	22
Working days traffic ratio (WDR) [13]	0.9
Busy hour traffic ratio (BHR) [9]	0.15
Voice Activity Factor (VAF) [8]	0.5
Burstiness (*B*) [14]	0.25

$$\begin{aligned} S1UBW = {} & [(1-VAF)(BHU) + (VAF*AMR*VDC)] \\ & (1+HO)(1+RTF)(1+IPov)(1+B) \end{aligned} \tag{2}$$

BHSA, presented in Eq. (3), provides the number of session attempts during the busy hour [22], and it is calculated [23] by multiplying the number of attached subscribers by busy hour traffic intensity (ρ), which represents the amount of usage per subscriber in busy hour and then dividing the answer by the mean session duration (MST). The unit for BHSA is the number of sessions per busy hour.

$$BHSA = \frac{N_a * \rho}{MST} \tag{3}$$

BHSA is represented by two separate equations (i.e., BHVSA and BHDSA) due to the two different types of BHSA (i.e., voice and data). Equations (4) and (5) present the BHVSA and BHDSA respectively. The voice traffic intensity (ρ_v) is equal to 20, and the data traffic intensity (ρ_d) is averaged to 1.52 [24]. As noticed, the voice traffic intensity is higher than the data traffic intensity, and this is because the arrival rate for voice connections is higher than data. In addition to that, the data session can be carried over default bearers, whereas the voice session usually requires dedicated bearers which are established more often.

$$BHVSA = \frac{N_a{*}\rho_v}{MST} \tag{4}$$

$$BHDSA = \frac{N_a{*}\ \rho_d}{MST} \tag{5}$$

The result of multiplying the total number of subscribers (*N*) by the attached subscriber ratio for dense area (ASR) is the number of attached subscribers for the eNB (N_a) as shown in Eq. (6). The area covered by the eNB affects the attached subscriber ratio (e.g., highly populated areas requires higher ratio).

$$N_a = \ N{*}\ ASR \tag{6}$$

Equation (7) provides the number of simultaneous EPS (SEPSB) dedicated bearer sessions over the S1-U interface. It is the result of multiplying the number of attached subscribers (N_a) by the average duration of each data bearer session (BSD) and the active BH EPS ratio (BR) which represents the percentage of active sessions and it is controlled by the operator and the capabilities of the network. In general, the subscriber has the ability to have more than one simultaneous bearer session, and one of the advantages of this feature is that a user would be able to connect to a PDN for Internet service, and simultaneously connect for video or to another PDN (e.g., IMS) [25].

$$SEPSB\ = N_a\ {*}\ BSD\ \ {*}\ BR \tag{7}$$

Fair Usage Policy (FUP) controls the ability to access the provided services, and it is applied on users who deplete their quotas and get limited access speed as described in [26]. In fact, service providers make sure the data provided to users is unlimited; however, speed is reduced until the end of the subscriber's billing cycle. For example, the operator's policy could be switching further sessions of LTE users to 3G UMTS or 1xEV-DO network when their LTE data usage threshold has been exceeded. Normally, applications that require high bandwidth such as video streaming will be affected; nevertheless, applications such as emails and web browsing will not be highly affected due to the fact that they don't require high download rates. The total throughput of the S1-U interface is affected by the FUP. Let us define α (e.g., 10 %) as the ratio of attached subscribers that depleted their quota and had to use extra FUP data. The FUP throughput (FUPBW) can be obtained by using Eq. (8). It is the result of multiplying the number of users using the FUP feature by the amount of excess data per subscriber (*X*). The rest of the values were explained earlier and they are used to ascertain that FUPBW is in bps.

$$FUPBW = \ \frac{\alpha\ {*}\ N_a{*}\ X{*}WDR{*}\ BHR}{WD} \tag{8}$$

In fact, a number of users who depleted the quota, and got their access speeds slowed down or their amount of data usage restricted may choose to purchase extra amount of data. The Top Up (TU) feature is a significant factor that adds to the total throughput of the network. Top Up Bandwidth (TUBW) is shown in Eq. (9) assuming a user pays $20 for 1 GB of data in a month, and the rate for the number of users who use the Top Up feature is β (e.g., 20 %) of the total number of attached subscribers. It is the result of multiplying the number of subscribers using the TU feature by the excess data per subscriber (*X*).

$$TUBW = \ \frac{\beta{*}\ N_a{*}\ X{*}\ WDR{*}\ BHR}{WD} \tag{9}$$

Equation (10) is used to measure the total amount of bandwidth being carried on the S1-U interface assuming one bearer session for each attached subscriber. Taking into consideration that utilization is resulted by dividing the traffic load (bps) by the capacity (bps); where capacity is the maximum amount of load supported by the network. The S1-U total BW (S1UTBW) in Mbps is a result of adding the FUPBW and the TUBW to the amount of simultaneous EPS bearers multiplied by the average S1-U throughput and the result of the summation is divided by the interface utilization which ought to be less than 0.85.

$$S1UTBW = \frac{(SEPSB{*}S1UBW) + FUPBW + TUBW}{S1UF} \tag{10}$$

Signalling and control traffic on the S1-MME adds to the load of the network [27]. The control operations that are performed over the S1-MME include signalling for attachment and detachment, as well as EPS bearer establishment and management (over S1-MME, S11, and S5), along with authentication requests and responses. Since eNB is directly connected to the MME, the MME has a massive load of transactions per second with both the HSS and the SGW. The amount of signalling varies between elements and the number of transactions per seconds also differs based on the component, the operations, and the functions of the network element. It also varies depending on the vendor and the specifications for the specific model of the element.

The process in which HSS signalling is calculated is different than the process used for computing the signalling of the other network elements. Equations (11) and (12) show the two ways to compute the amount of signalling for the HSS, and the rest of the elements (MME, SGW, PGW, and PCRF) respectively. HSS Signalling is represented as Signalling Load per Authentication Element (SLAE), and signalling for other core elements is represented as Signalling Load per Core Element (SLCE).

The amount of transactions per second is not absolute; it simply depends on the element, the provider, and the model along with different parameters. Equation (11) presents the calculation of the total amount of signalling for the HSS, which is the result of the number of transactions per second per subscriber (TPS) multiplied by the total number of subscribers (*N*); whereas in Eq. (12), the total amount of signalling for the each element is calculated by multiplying the number of bearers during a busy hour by the number of transactions per second per bearer (TPSB).

$$SLAE = N * TPS \tag{11}$$

$$SLCE = EPSB * TPSB \tag{12}$$

The number of transactions is completely relative and depends on many factors as explained previously. In other words, these values are not standard and they are usually specific to a certain implementation or product. As an example, it can be concluded from [28] that the MME handles 9.3 transactions per second per bearer since it handles around 290,000 messages per second. For the PGW, it handles 63,000 messages per second leading to 2 transactions per second per bearer. For the SGW, the number of messages per second is 94,000 leading to 3 transactions per second per bearer. The fact that the SGW is higher than the PGW is that the latter doesn't have to deal with service request, release and paging messages. For the PCRF, as given in [11], the number of transactions per second per bearer is equal to 2. By using the Diameteriq smart signalling tool and taking into consideration the results shown in [29], the number of diameter messages per second in the HSS is equal to 5000 messages per second. Hence, it equals 6.2 transactions per second per subscriber multiplied by the total number of subscribers.

The next section will show how these equations are used in order to generate traffic profiles for eNodeBs and how this can be used in the actual planning process.

4 Traffic profile generation and application example

The example presented in this section provides the eNB traffic records obtained by using the equations in Section 3. The given traffic records take into account the different types of traffic (i.e., voice, web browsing, etc.) based on the delay sensitivity traffic classification (i.e., conversational, streaming, interactive, and background classes) as discussed earlier. The example presents different traffic profiles showing different aspects of the traffic measurements for 10 eNBs with a maximum capacity of 1500 subscribers and other 5 eNBs of higher capacity that can handle up to 10,000 subscribers. The reason why a larger scale is presented is to show how the other parameters are directly proportional to the number of subscribers. The number of subscribers that can be handled by an eNB depends on the capability and the performance of the eNB (e.g., memory, hardware, CPU, etc.). The network operators deploying the traffic model have the ability to choose a lower or a higher number of subscribers based on the need and the equipment they have.

The eNBs traffic profiles are presented in Table 3. The first column, eNB index, represents the index of the different eNBs to which the traffic belongs. The second column has the total number of subscribers covered by each eNB (i.e., subscribers whether they are active or idle), and it is randomly generated between 1000 and 1500 for the small scale (i.e., the first 10 rows) and between 5000 and 10,000 for the larger scale (the last 5 rows). Attached subscribers, presented in the third column, represent a percentage of subscribers who have a successful connection with the PGW. Busy hour session attempt is presented for data and voice in columns four and five respectively. In column six, bandwidth measured in Megabits per second (Mbps) is presented. The peak data rate for LTE varies from 5–75 Mbps on UL (uplink) and 10–300 Mbps on DL (downlink), and the effective bandwidth is usually less. In

Table 3 Traffic record for 10 ENBs with a max capacity of 1500 subscribers and 5 ENBs with a max capacity of 10,000 subscribers

eNB	N	N_a	BHDSA	BHVSA	BW(Mbps)	EPSB
1	1482	1333	40,547	533,520	27	166
2	1131	1017	30,944	407,160	20	127
3	1393	1253	38,112	501,480	25	156
4	1332	1198	36,443	479,520	24	149
5	1076	968	29,439	387,360	19	121
6	1211	1089	33,132	435,960	22	136
7	1458	1312	39,890	524,880	26	164
8	1389	1250	38,003	500,040	25	156
9	1489	1340	40,739	536,040	27	167
10	1315	1183	35,978	473,400	24	147
11	6756	6080	184,840	2,432,164	124	760
12	7914	7122	216,523	2,849,044	146	890
13	7702	6931	210,722	2,772,724	142	866
14	8075	7267	220,927	2,907,005	149	908
15	9850	8865	269,490	3,546,006	181	1108

column seven, the number of simultaneous sessions of dedicated EPS bearer is provided. As can be concluded from the table, the amount of traffic is proportional to the number of subscribers for each eNB. Compared to other eNBs, the eNB that has a higher number of subscribers has a higher number of attached subscribers, busy hour sessions attempts, simultaneous bearers and higher amount of BW. For a range of 1076 to 1489 subscribers, the number of attached subscribers varies between 968 and 1340. The number of busy hour session attempts for voice varies between 387,360 and 536,040; whereas the busy hour session attempts for data varies between 29,439 and 40,739. The amount of bandwidth used by subscribers to access services is between 19 and 27 Mbps, and the number of simultaneous bearers varies between 121 and 167. The number of busy hour session attempts for voice is higher than the number of busy hour session attempts for data, because the voice traffic intensity is higher than the data traffic intensity. For a range of 6756 to 9850 subscribers, the number of attached subscribers varies between 6080 and 8865. The number of busy hour session attempts for voice varies between 2,432,164 and 3,546,006; whereas the busy hour session attempts for data varies between 184,840 and 269,490. The amount of bandwidth used by subscribers to access services is between 124 and 181 Mbps, and the number of simultaneous bearers varies between 760 and 1108.

Once the traffic profile of each eNodeB has been generated, the next step in the planning process is to plan the rest of the network (i.e., the evolved packet core) such that it will be able to support the load generated by all the subscribers. For the rest of this example, let us assume that the 15 eNodeBs from Table 3 are randomly located on a 10 km by 10 km area and connected to the core network as shown in Fig. 2. As we can see, the 15 eNBs are connected to a single MME and a single SGW. The MME is also connected to the SGW and to one HSS. Finally, the SGW is connected to the PGW which is then connected to the PCRF.

The traffic profile from each eNB along with the capacities of the various elements is the main factor that controls the number and types of core network elements installed. As observed, the number of links between the elements also depends on the traffic parameters from all the eNBs and the network core elements. In fact, each element has a maximum capacity that it can't exceed, and in order to plan a good network, capacity constraints need to be taken into consideration.

The importance of looking into the capacity constraints of the LTE core elements goes back to the urge to identify the type and the needed number of network elements. In general, there are four different basic types of capacity constraints [30]: throughput, transactions, subscribers, and bearers. Throughput is the total amount of data load that a node can handle, and it is considered as a data plane limitation, whereas transactions are signalling messages related to the control plane. Subscribers represent the number of subscribers that can be handled by a node, and there are different types of subscribers that can be handled by a node such as active vs. idle (i.e., without ongoing media session), or activated vs. configured subscribers who are not yet activated. Moreover, the number of bearers; a bearer is a data connection and there are two types: default and dedicated bearers. The default bearer is best effort and is mandatory for any attached user, whereas the dedicated bearer is established based on the need.

A big capacity issue for the MME is the number of transactions (or control messages) as well as the number of subscribers that are attached to the MME or in MME's temporary subscriber database. The MME is also constrained by the number of BHSA and EPSB as well as the amount of bandwidth. The HSS is a database for subscribers' data and it is concerned about the control plane in particular. The connection between

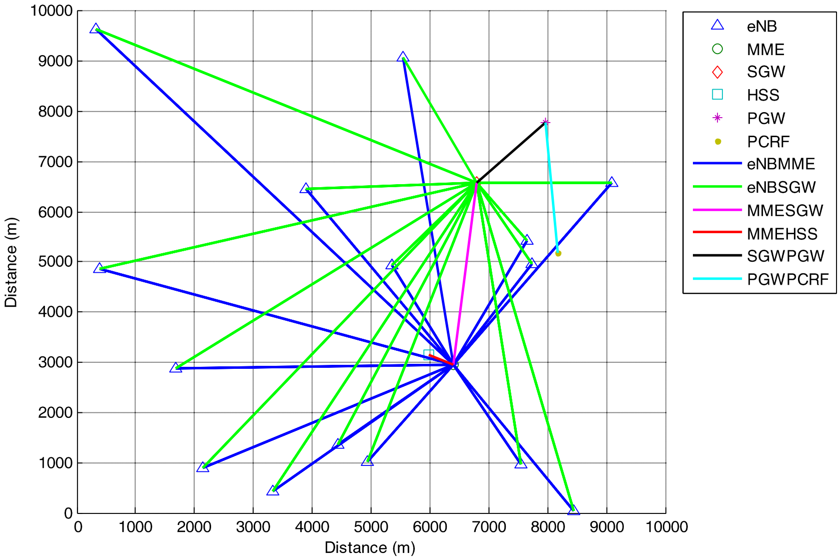

Fig. 2 LTE network planning example

the MME and the HSS is control plane only. Two legitimate concerns for the HSS are the number of transactions and the number of subscribers. The SGW works as a mobility anchor for the traffic being carried on different eNBs, and it also forwards packets between the eNB and the PGW. The SGW is focused on the user plane, and throughput is the number one limitation. Regarding the number of transactions, it is a relative issue to consider as a main constraint but in this paper we take the signalling traffic and the number of transactions into consideration. For the PGW, the number of subscribers is not considered as a big constraint since the service providers are supposed to be able to accommodate a huge number, whereas throughput is the main constraint in this context. In terms of transactions, the PGW is affected by three control operations: setup/teardown of bearers, QoS negotiation with the PCRF, and inter SGW mobility. The number of bearers is considered a limitation, and generally, it depends on the vendor.

The signalling load per authentication element, which refers to the signalling load per HSS, is presented in Table 4. The first column represents the eNB index (referring to Table 3), the second column represents the total number of subscribers connected to the eNB, and the third column represents the amount of signalling per authentication element (SLAE) which is measured in transactions per second (tps). The signalling load per authentication element is calculated by multiplying the number of subscribers by the number of transactions per subscriber per second which is equal to 6.2 as discussed in the previous section. As we can see, the number of transactions per second varies between 6671 and 9232 tps for a number of subscribers that varies between 1076 and 1489. For a larger number of subscribers (i.e., between 6756 and 9850), the number of transactions per second varies between 41,887 and 61,070.

Table 4 Signalling load per authentication element

eNB	N	SLAE (tps)
1	1482	9188
2	1131	7012
3	1393	8637
4	1332	8258
5	1076	6671
6	1211	7508
7	1458	9040
8	1389	8612
9	1489	9232
10	1315	8153
11	6756	41,887
12	7914	49,067
13	7702	47,752
14	8075	50,065
15	9850	61,070

The signalling load per core element is provided in Table 5. The first column represents the index for the eNBs, and the second column represents the number of EPSB which is considered the main factor in determining the amount of signalling and transactions per second for core elements. The signalling load per core element is calculated by multiplying the number of EPSB by the number of transactions per bearer per second. As explained in the previous section, the number of transactions per bearer per second depends on the core element. For example, for the MME the number of transactions per bearer per second equals to 9.3, the number of transactions per bearer per second for the SGW is 3, and the number of transactions per bearer per second for the PGW and the PCRF is 2. For a number of EPSB that varies between 121 and 167, the number of transactions per second for the MME varies between 1125 and 1553 as shown in the third column. The number of transactions per second for the SGW ranges between 363 and 501 as provided in the fourth column. The amount of signalling is provided for the PGW in the fifth column, and in the sixth column the number of transactions per second is given for the PCRF. The values for the PGW and PCRF signalling vary between 242 and 334 transactions per second. For a larger number of bearer that varies between 760 and 1108 bearers, the number of transactions per second for the MME is between 7068 and 10,304 tps, and for the SGW between 2280 and 3324. In addition to that, the number of transactions per second for the PGW and the PCRF ranges between 1520 and 2216. The first eNB is taken as an example with the number of EPSB equal to 166. The SLCE is calculated for the MME by multiplying the number of EPSB by the number of transactions per bearer per second which is equal to

Table 5 Signalling load per core element

eNB	EPSB	SLC (MME)	SLCE (SGW)	SLCE (PGW)	SLCE (PCRF)
1	166	1544	498	332	332
2	127	1181	381	254	254
3	156	1451	468	312	312
4	149	1386	447	298	298
5	121	1125	363	242	242
6	136	1265	408	272	272
7	164	1525	492	328	328
8	156	1451	468	312	312
9	167	1553	501	334	334
10	147	1367	441	294	294
11	760	7068	2280	1520	1520
12	890	8277	2670	1780	1780
13	866	8054	2598	1732	1732
14	908	8444	2724	1816	1816
15	1108	10,304	3324	2216	2216

9.3 resulting in 1544 transactions per second. The SLCE is calculated for the SGW by multiplying the number of EPSB by the number of transactions per bearer per second which is equal to 3 resulting in 498 transactions per second. Similarly, the SLCE is calculated for the PGW by multiplying the number of EPSB by the number of transactions per bearer per second which is equal to 2 resulting in 332 transactions per second and finally, the SLCE is calculated for the PCRF by multiplying the number of EPSB by the number of transactions per bearer per second which is equal to 2 resulting in 332 transactions per second.

In this example, the 15 eNBs are all connected to a single MME and a single SGW. As a result, the total signalling load on the MME is equal to 55,995 transactions per second, and the total load on the SGW is equal to 18,063 transactions per second. The total amount of signalling load on the PGW is equal to 12,042 transactions per second and the total amount of signalling load on the PCRF is equal to 12,042 transactions per second. This means that while planning the network, we need to make sure that the core network elements can support that load. In other topologies, the numbers of MME, SGW, PGW, and PCRF may vary based on the traffic and the resources' capacities; as a result the signalling load varies accordingly.

5 Conclusion and future work

Network planning is an essential process in today's market. Network operators' main objective is to provide subscribers with different services at reasonable prices. As a result, investing in network planning and considering realistic, real life parameters is important for operators to keep up in a competitive market. Taking different aspects of the traffic into account is an important task that provides network planners with various parameters. In this paper, we developed a traffic model considering bandwidth, signalling, voice BHSA (BHVSA), data BHSA (BHDSA), and EPSB. The aim of the model is to provide network planners with a set of realistic traffic parameters that can be used as input for the planning and design of LTE networks. The proposed parameters aim to cover different aspects of the traffic, for example signalling traffic includes the amount of transactions occurring to initiate a certain process, whereas bandwidth is used by subscribers to access services from the Internet or the PDN. In conclusion, the traffic generated is adaptable to the operators' needs and uses, each operator can deploy the proposed traffic model using their own parameters to make it compliant to their own needs. Due to the fact that operators tend to keep their information private, they can perform model validation in accordance to their needs and in compliance with their data.

As future work, we are currently developing a planning tool for the LTE EPC using the proposed traffic model as input. Parameters like signalling transactions, throughput, BHSA and EPSB will be useful to tackle the planning issue with more details.

References

1. Qiu Q-l, Chen J, Ping L-d, Zhang Q-f, Pan X-Z (2009) LTE/SAE model and its implementation in NS 2. 5th international conference on mobile Ad-hoc and sensor networks. 299-303
2. Li X, Toseef U, Weerawardane T, Bigos W, Dulas D, Goerg C, Timm-Giel A, Klug A (2010) Dimensioning of the LTE S1 interface. Wireless and Mobile Networking Conference (WMNC). 1–6
3. Li X, Toseef U, Weerawardane T, Bigos W, Dulas D, Goerg C, Timm-Giel A, Klug A (2010) Dimensioning of the LTE access transport network for elastic internet traffic. 6th IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob). 346–354
4. Lindberger K (1999) Balancing quality of service, pricing and utilisation in multiservice networks with stream and elastic traffic. In Proc. of the International Teletraffic Congress (ITC 16), Edinburgh, Scotland
5. Checko A, Ellegaard L, Berger M (2012) Capacity planning for Carrier Ethernet LTE backhaul networks. IEEE Wireless Communications and Networking Conference (WCNC). 2741–2745
6. Jailani E, Ibrahim M, Ruhani Ab Rahman (2012) LTE speech traffic estimation for network dimensioning. IEEE Symposium on Wireless Technology and Applications (ISWTA). 315–320
7. Lehpamer H (2002) Transmission systems design handbook for wireless networks. Artech House, Boston
8. (2007) Long Term Evolution (LTE): overview of LTE air-interface technical white paper. Technical white paper, Motorola
9. Cisco (2007) Traffic analysis for voice over IP
10. Mishra AR (2007) Advanced cellular network planning and optimisation 2G/2.5G/3G evolution to 4G. Wiley
11. Brown G (2010) Control plane scalability & performance requirements for 3G & LTE networks. White paper
12. Basit SA (2009) Dimensioning of LTE network description of models and tool, coverage and capacity estimation of 3GPP long term evolution radio interface. MS Thesis, Helsinki University of Technology, Finland
13. Pasandideh M R (2011) Automatic planning of 3G UMTS All-IP release 4 networks with realistic traffic. MS Thesis. Carleton University, Ottawa
14. Smith C, Collins D (2002) 3G wireless networks. McGraw-Hill, New York
15. Poikselkä M, Holma H, Hongisto J, Kallio J, Toskala A (2012) Voice over LTE (VoLTE). John Wiley & Sons Ltd, West Sussex, United Kingdom, p 243
16. Gessner C, Gerlach O (2011) Voice and SMS in LTE. Rohde & Schwarz, White Paper
17. GSMA PRD IR.92 (2010) IMS Profile for Voice and SMS. GSM Association
18. Newport Networks (2005) VoIP Bandwidth Calculation
19. Garg V (2007) Wireless communication and networking. Morgan Kaufmann
20. Holma H, Toskala A (2010) LTE for UMTS: evolution to LTE-advanced. 2nd edition. Wiley
21. Krzanowski R (2006) Burst of packets and Burstiness. Technical Report, v1.0, 66th IETF meeting

22. Chuah MC, Zhang Q (2006) Design and performance of 3G wireless networks and wireless LANs. Springer. 50
23. Cisco (2006) Sizing call center resources. Cisco IP Contact Center Enterprise Edition Releases 5.0 and 6.0 Solution Reference Network Design
24. Gu J, Ruan Y, Chen X, Wang C (2011) A novel traffic capacity planning methodology for LTE radio network dimensioning. Proc. of International Conference on Communication Technology and Application (ICCTA2011). 462–466
25. Zhao X, Wu W, Cai Z (2012) Dynamic aggregated maximum bit rate for evolved packet system non-guaranteed bit rate quality of service enforcement and network bandwidth utilization. United States Patent No. 8,218,436 B2, Research in Motion
26. (2009) Enabling mobile broadband growth – Evolved Packet Core. Technical white paper, Nokia Siemens Networks
27. Punz G (2010) Evolution of 3G networks: the concept, architecture and realization of mobile networks beyond UMTS. Springer
28. Nokia Siemens Networks (2012) Signalling is growing 50 % faster than data traffic. Nokia Siemens Networks, White paper
29. Diametriq (2012) Measuring the explosion of lte signalling traffic - a diameter traffic model. White paper
30. LTE University (2012) videos, http://lteuniversity.com/