IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)
e-ISSN: 2278-2834,p- ISSN: 2278-8735.Volume 15, Issue 4, Ser. I (Jul - Aug 2020), PP 08-18
www.iosrjournals.org

CAMTEL's EPC traffic modeling to help in decision-making

Alphonse BINELE ABANA[1], Emmanuel TONYE[2]

[1](*Department of Telecommunications and Electrical Engineering, National Advanced School of Engineering/ University of Yaounde I, Cameroon)*

[2](*Department of Telecommunications and Electrical Engineering, National Advanced School of Engineering/ University of Yaounde I, Cameroon))*

Abstract:

Background*:*
The work carried out consisted of the designing models (linear and neural) in order to control the 4G traffic of the operator CAMTEL, help in decision-making in the network concerning technical and commercial actions to be taken. In this study, we are working with KPIs collected at CAMTEL's EPC (Evolved Packet Core) network on the USN and UGW entities.

Materials and Methods*:*
The linear model helped to obtain the linear multiple regression equation that specified the downlink user traffic as a function of other model variables that were all strongly correlated. It was produced using the SPSS statistical software and was validated by the Fisher test..

Results*:*
We obtained a valid model which passed the Fisher test. We were able to express the SGi downlink traffic (kbits) as a function of correlated MME and SGW/PGW KPIs. The analysis of the obtained model shows that the attached user number is negatively impacting the downlink user traffic: thus CAMTEL's network is congested at the radio side.

Conclusion:
However, this model should be reassessed regularly to allow it to always reflect the real situation as far as possible.

Key Word*: KPI; traffic; model; EPC; 4G.*

Date of Submission: 11-08-2020 Date of Acceptance: 27-08-2020

I. Introduction

The mobile telephone system is gaining considerable success, with an ever increasing number of subscribers. From the first generation of mobile networks through to the fourth generation today and migrations to the fifth generation, there is a growing demand for services. The fourth generation of mobile networks also called LTE (Long Term Evolution), is the modern technology currently marketed, it offers speeds up to 6 times higher than H + in its most modern version (4G +). The traffic of a network is the set of information flows passing through the network at any given moment. It is an important concept for bandwidth management. It is important for an operator to be able to control its traffic in order to anticipate, optimize, and make predictions on it. CAMTEL (Cameroon Telecommunications) is not on the sidelines of this trend and would like to equip itself with a mathematical model of its 4G traffic allowing him to guide its decision-making. There are many modeling or network prediction tools but we have chosen to use linear modeling. To achieve this goal, we will first present and overview of a 4G network and some KPIs then, we will recall some mathematics concepts about the linear model before proposing our traffic model.

II. 4G network overview

The 4G also known as LTE (Long Term Evolution) has brought many changes and improvements, higher throughput, improved spectral efficiency, reduced access time and compatibility with 2G/3G networks and faster mobility. The architecture of a 4G network is given in the figure below.

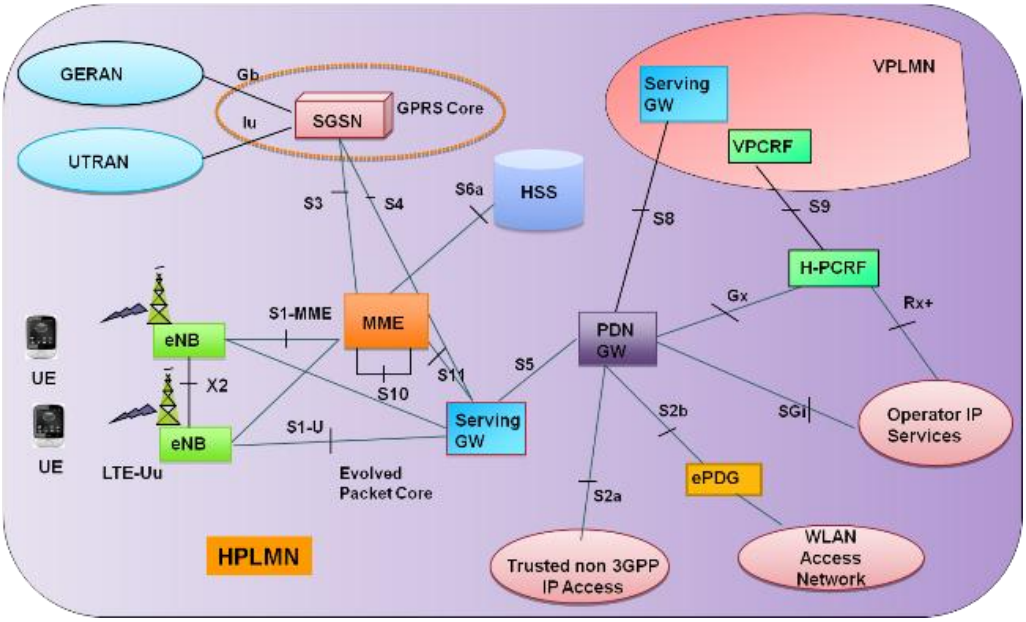

Figure 1: 4G network architecture [1]

Figure 1 shows the architecture of an LTE network with different interfaces and entities of the radio part and the core part, which will be presented below.

II.1. LTE radio Access part [3]

In the LTE access network, the antennas are called e-Node B. For operators who already have strong coverage in other technologies, the pooling of antenna sites is possible and even desirable. It is possible to replace the 2G and 3G antennas with antennas emitting at all frequencies at the same time: 2G, 3G and 4G. Thus, with a single antenna, an operator can cover a site with all the technologies. 2G had the BSC as its focal point in the access network while 3G also had the RNC as its focal point. But with LTE, the intermediate step between the antenna and the core network disappears. The antenna (eNode B) is connected directly to the core network, avoiding an intermediary and simplifying the general architecture.

II.2. LTE core network part: EPC [3][20]

4G is based on a whole new core network: the EPC. The major development of the EPC is the separation of control and usage plans for core equipment. Like NGN for the circuit, the "control" and "usage" flows will in most cases no longer be managed by the same equipment.

- The 3G SGSN has been replaced by two different functional entities: The MME (Mobility Management Entity) for the "control" plan and the SGW for the "usage" plan.
    - The MME will thus manage the sessions (authentication, authorizations, data session) and mobility (location, "paging", "hand-over") of the terminal.
    - SGW will be responsible for routing "useful" flows in the core network (data traffic, etc.).
- The 3G GGSN is replaced by a PDN Gateway (PGW). The PDN Gateway is responsible for the link with other networks (public or private), and especially with the Internet world.
- The HSS is responsible for much of the same functionality as the HLR (database of subscriber profiles, with their rights and characteristics). The HSS also includes a possible link with the IMS world, for the management of enriched voice services.
- A new element of the 4G core network is the PCRF, which enables dynamic management of billing and quality of service policy for flows (best-effort flow, "premium" flows with guaranteed bandwidth and latency, dynamic pricing according to the flow. , etc.).
- The PCEF, a functional module embedded in the PDN Gateway, applies the rules set by the PCRF.
- The IMS is a set of elements allowing multimedia services to be offered over IP networks and therefore, among other things, over the mobile network. The IMS will enable rich multimedia inter-personal services: voice over IP, video conferencing, enriched calendar, instant messaging, ringing on several terminals, etc.

II.3. Performance measurement of a 4G network

KPI (Key Performance Indicators) correspond to measures providing information on the performance of the network or of a process. KPIs are obtained using measurements made on the nodes of the network and by compiling the various data. The 3GPP 32.455 standard defines the KPIs of 4G and we have among others:

- Accessibility KPI
  - EPS Attach Success Rate
  - Dedicated EPS Bearer Creation Success Rate
  - Dedicated Bearer Set-up Time by MME (Mean)
  - Service Request Success Rate
- Mobility KPI
  - Inter-RAT Outgoing Handover Success Rate (EPS->GSM)
  - Inter-RAT Outgoing Handover Success Rate (EPS->UMTS)
  - Inter-RAT Outgoing Handover Success Rate (EPS->CDMA2000)
  - Inter-RAT Incoming Handover Success Rate (GSM->EPS)
  - Inter-RAT Incoming Handover Success Rate (UMTS ->EPS)
  - Inter-RAT Incoming Handover Success Rate (CDMA2000->EPS)
  - Tracking Area Update Success Rate 11
- Utilization KPI
  - Mean Active Dedicated EPS Bearer Utilization

II.3.1. EPS Attached Successful rate

This KPI describes the ratio of the number of successfully performed EPS attach procedures to the number of attempted EPS attach procedures for EPC network and is used to evaluate accessibility provided by EPS and network performance.
This KPI is obtained by successful EPS attach procedures divided by attempted EPS attach procedures.

$$\mathrm{EASR} = \frac{\sum_{Type} \mathrm{MM.EpsAttachSucc.}Type}{\sum_{Type} \mathrm{MM.EpsAttachAtt.}Type} * 100\%$$

II.3.2. Dedicated EPS bearer creation success rate.

This KPI describes the ratio of the number of successfully performed dedicated EPS bearer creation procedures by PGW to the number of attempted dedicated EPS bearer creation procedures by PGW and is used to evaluate service availability provided by EPS and network performance.
This KPI is obtained by successful dedicated EPS bearer creation procedures divided by attempted dedicated EPS bearer creation procedures.

$$\mathrm{DEBCSR} = \frac{\mathrm{SM.CreationPGWIniBearerSucc}}{\mathrm{SM.CreationPGWIniBearerAtt}} * 100\%$$

II.3.3. Mean dedicated bearer set-up time by MME

This KPI describes the valid time per dedicated bearer set-up procedure by MME and is used to evaluate service accessibility provided by EPS and network performance.
This KPI is obtained by the valid time per dedicated bearer set-up procedure by MME
DBSTM= SM.EstabActDedicatedEpsBearerTimeMean

II.3.4. Service request success rate.

This KPI describes the ratio of the number of successfully performed service request procedures by UE to the number of attempted service request procedures by UE and is used to evaluate service accessibility provided by EPS and network performance.
This KPI is obtained by successful service request procedures divided by attempted service request procedures.

$$\mathrm{SRSR} = \frac{\sum_{MME} \mathrm{SM.EpsServiceReqSucc}}{\sum_{MME} \mathrm{SM.EpsServiceReqAtt}} * 100\%$$

II.3.5. Inter-RAT outgoing handover success rate (EPS->GSM)

This KPI describes the ratio of the number of successfully performed outgoing handover procedures to the number of attempted outgoing handover procedures to evaluate inter RAT outgoing handover performance.

This KPI is obtained by the number of successful outgoing handover procedures divided by total number of attempted outgoing handover procedures from EPS to GSM network.

$$\text{IRATOHOSR} = \frac{\text{IRATHO.OutMMESucc.G}}{\text{IRATHO.OutMMEAtt.G}} * 100\%$$

II.3.6. Inter-RAT outgoing handover success rate (EPS->UMTS)

This KPI describes the ratio of the number of successfully performed outgoing handover procedures to the number of attempted outgoing handover procedures to evaluate inter RAT outgoing handover performance.
This KPI is obtained by the number of successful outgoing handover procedures divided by total number of attempted outgoing handover procedures from EPS to UMTS network.

$$\text{IRATOHOSR} = \frac{\text{IRATHO.OutMMESucc.U}}{\text{IRATHO.OutMMEAtt.U}} * 100\%$$

II.3.7. Inter-RAT incoming handover success rate (GSM->EPS)

This KPI describes the ratio of the number of successfully performed incoming handover procedures to the number of attempted incoming handover procedures to evaluate inter RAT incoming handover performance.
This KPI is obtained by total number of successful incoming handover procedures divided by total number of attempted incoming handover procedures from GSM network to EPS

$$\text{IRATIHOSR} = \frac{\text{IRATHO.IncMMESucc.G}}{\text{IRATHO.IncMMEAtt.G}} * 100\%$$

II.3.8. Inter-RAT incoming handover success rate (UMTS ->EPS)

This KPI describes the ratio of the number of successfully performed incoming handover procedures to the number of attempted incoming handover procedures to evaluate inter RAT incoming handover performance.
This KPI is obtained by total number of successful incoming handover procedures divided by total number of attempted incoming handover procedures from UMTS network to EPS

$$\text{IRATIHOSR} = \frac{\text{IRATHO.IncMMESucc.U}}{\text{IRATHO.IncMMEAtt.U}} * 100\%$$

II.3.9. Tracking area update success rate.

This KPI describes the ratio of the number of successfully performed tracking area update procedures to the number of attempted tracking area update procedures and is used to evaluate mobility provided by EPS and network performance.
This KPI is obtained by successful tracking area update procedures divided by attempted tracking area update procedures.

$$\text{TAUSR} = \frac{\sum_{TA}\left(\text{MM.TauInterSgwSucc} + \text{MM.TauIntraSgwSucc}\right)}{\sum_{TA}\left(\text{MM.TauInterSgwAtt} + \text{MM.TauIntraSgwAtt}\right)} * 100\%$$

II.3.10. Mean active dedicated EPS bearer utilization

This KPI describes the ratio of the mean number of active dedicated EPS bearer to the maximum number of active dedicated EPS bearers provided by EPC network, and it is used to evaluate utilization performance of EPC network.
This KPI is obtained by the mean number of dedicated EPS bearers in active mode divided by the system capacity.

$$\text{MADEBU} = \frac{\text{SM.MeanNbrActDedicatedBearer}}{\text{Capacity}} * 100\%$$

In addition to these KPIs defined by the 3GPP specifications, manufacturers measure performance indicators such as the uplink and downlink throughput on the SGi interface between the PGW and the Internet as well as the downlink and uplink traffic volume of the 'SGi interface for a given time interval.

II.4. Problem and objective

Given the extent of the LTE network, the intensity of communications and the increased demand for services by users, it is wise for CAMTEL, as for all other operators, to control its traffic. This work therefore consists of providing CAMTEL with a model of its EPC core network traffic in order to control it and therefore facilitate decision-making on the technical and managerial actions to be carried out to increase revenue.
To this end, what is traffic modeling? How reliable is it? How to proceed with traffic modeling?
To respond to the need of traffic modeling, we will propose a linear model by using adequate performance indicators.

II.5. Overview of the linear model

II.5.1. Pearson formula [19]

It measures the degree of the linear relationship between two variables. The correlation coefficient assumes a value between -1 and +1. If one variable tends to increase while the other decreases, the correlation coefficient is negative. Conversely, if both variables tend to increase, the correlation coefficient is positive. Pour les variables x et y :

$$\rho(X;Y) = \frac{\sum(X-\bar{X}).(Y-\bar{Y})}{\sqrt{\sum(X-\bar{X})^2}.\sqrt{\sum(Y-\bar{Y})^2}.} = \frac{cov(X;Y)}{\sqrt{V(X)}.\sqrt{V(Y)}}$$

We make the choice of strongly correlated variables validated by the Cohen table given below.

Autour de 0,10	Effet de petite taille	Corrélation faible
Autour de 0,30	Effet de taille moyenne	Corrélation moyenne
Plus de 0,50	Effet de grande taille	Corrélation forte

Table 1. Cohen check-list

Table 2.

II.5.2. Model presentation [10][17]

The linear model results in a mathematical equation in which we calculate the various coefficients associated with the parameters of the equation. Notice that among strongly correlated variables retained, we will express a variable called dependent variable as a function of other variables called explanatory variables. The linear model equation is in the form:

$y = a_0 + a_1x_1 + a_2x_2 + \cdots + a_nx_n + \varepsilon$ Where the a_i $i = 0,..n$ are coefficients to be determined and ε the model error; ε expresses, or summarizes, the missing information in the linear explanation of the values of y from the x_i.

To determine these coefficients, we consider the actual data collected for each time unit and we assume that the linear equation for these data is verified. Let consider a data collection of i time units. From the collected data, we assume that each $y_i$verifies the linear model equation: $y_i = a_0 + a_1x_{1i} + a_2x_{2i} + \cdots + a_nx_{ni} + \varepsilon_i$.

So we have the matrix form $y = aX + \varepsilon$

- Where X is the matrix of explanatory variables values with, in the first column, the value 1 to materialize the regression constant a_0. It dimension is $n \times p$ where n is the number of time units at which data were collected and p is the number of explanatory variables.
- y is the vector containing the real values of dependent variable we want to explain with the linear model. Its dimension is $n \times 1$
- ε is the vector determining the error of the model; its dimension is $n \times 1$.
- a is the vector of regression coefficients or constants we want to determine in order to have order model.

The ordinary least squares estimator aims to minimize or even cancel out the error or even the residuals ε_i in the model. Then the estimated residuals are the difference between the observed and estimated value of y. We have: $\varepsilon_i = y_i - \widehat{y_i}$.

The principle of least squares consists in finding the values of the parameters which minimize the sum of the squares of the residuals. The estimator which minimizes the sum of the squares of the residuals is:

$a = (X'X)^{-1}X'y$

- Where X' is the transpose of the matrix of the measured values of the explanatory variables X.
- a is the vector of the coefficients associated with the explanatory variables.
- y the measured values of the dependent variable which we want to explain.

The portion that cannot be explained by the model is symbolized by εi which represents the error made by the model for each value of y.

II.5.3. Model validation [21]

Several parameters must be considered to validate a multiple linear regression model.

Beforehand several quantities must be calculated:

- The total variability of the dependent variable, $STS = \sum i\ (y_i - \bar{y})^2$ [sum of total squares] with degrees of freedom $DF = n - 1$;
- The variability not explained by the regression, $SRS = \sum i\ (y_i - \bar{y})^2$ [sum of residual squares] with $DF = n - (p + 1)$;
- The explained variability is obtained by the difference $SES = STS - SRS$ [Sum of Estimated Squares] with $DF = p$;

- The mean squares consist of Explained Mean Square (EMS) and the Residual Mean Square (RMS). They are obtained by the ratio between the sums of the squares and the degrees of freedom: $EMS = \frac{SES}{P}$ and $RMS = \frac{SRS}{n-p-1}$.

II.5.3.1. The Fisher test

The Fisher test is significant; it is the overall significance test which consists in verifying that there is at least one relevant variable among the explanatory ones. The Fisher test statistic can be obtained by the ratio between the explained and residual mean squares read in the analysis of variance table:

$$F = \frac{EMS}{RMS}$$

It would be necessary that the probability p for F to be significant must verify: $p < 0.05$ in order to validate the model.

II.5.3.2. Coefficient of determination

Here we want to assess the adjustment of the model to see what percentage of the variables impact on the model. We will calculate the coefficient of determination $R^2 = \frac{SES}{STS}$ which represents the strength of the relationship between the DV (dependent variable) and the combination of the ND (non-depend variable) of the model. The value of R^2 must be close to 1 to validate the model.

III. Model implementation

We are working with the EPC KPI data of CAMTEL's network. We have correlated some data from the MME with that from the PGW to understand the evolution of user traffic in the CAMTEL network.

III.1. Presentation of data

III.1.1. MME KPI used

Start Time	Period (min)	NE Name	Whole System	S1 mode intra-MME S1 Handover success rate (%)	S1 mode intra-MME TAU success rate (%)	security mode command success rate (%)	S1 mode packet paging success rate (%)	S1 mode service request success rate (Network Cause) (%)	Maximum attached users (number)	Maximum attached users at ECM-CONNECTED status (number)	Maximum attached users at ECM-IDLE status (number)	S1 mode maximum bearer number (number)	Maximum PDN connection number (number)
01/01/2019 23:30:00	30	MME_YDE	MME_YDE	97,061313	99,880083	99,998208	94,158417	96,916667	7176,6875	3900,2917	3462,4792	7199,2708	7198,3958
01/02/2019 23:30:00	30	MME_YDE	MME_YDE	96,031646	99,889583	99,999104	92,511354	96,708333	7470,1042	4220,125	3437,4792	7492,1667	7491,3125
01/03/2019 23:30:00	30	MME_YDE	MME_YDE	96,654396	99,864	99,997708	91,307667	96,208333	7310,5417	4114,7708	3384,2708	7332,8333	7331,9583
01/04/2019 23:30:00	30	MME_YDE	MME_YDE	95,880542	99,839708	99,998125	91,511542	96,5	7234,4792	4107,75	3309,8333	7257,9792	7257
01/05/2019 23:30:00	30	MME_YDE	MME_YDE	96,810167	99,865729	99,999	92,906125	96,5	6917,25	3803,9167	3294,0417	6939,8333	6939,0208
01/06/2019 23:30:00	30	MME_YDE	MME_YDE	96,022063	99,890125	99,998958	93,858792	95,8125	6563,2708	3522,375	3237,2083	6584,3333	6583,6458
01/07/2019 23:30:00	30	MME_YDE	MME_YDE	96,789146	99,900583	99,996604	91,718042	95,6875	6937,5625	3882,8333	3248,2292	6958,3125	6957,4167
01/08/2019 23:30:00	30	MME_YDE	MME_YDE	97,696042	99,893042	99,995958	92,274375	95,854167	7039,1042	4003,3958	3236,9792	7059,4583	7058,6875
01/09/2019 23:30:00	30	MME_YDE	MME_YDE	97,560313	99,830333	99,997396	91,984688	95,895833	7081,2708	4029,0833	3245,7083	7101,1667	7100,5833
01/10/2019 23:30:00	30	MME_YDE	MME_YDE	97,611583	99,9015	99,998521	92,401896	95,833333	7094,2292	4009,375	3265,4375	7112,9167	7112,2292
01/11/2019 23:30:00	30	MME_YDE	MME_YDE	96,527521	99,846958	99,997042	91,284792	94,416667	7364,1667	4146,2083	3427,9375	7383,4583	7382,8125

Table 3. CAMTEL's MME sample KPI data

Here is a sample file of data collected from the MME of CAMTEL. From this file, we have the following counters:

- S1 mode intra MME S1 handover success rate (%) which provides the success rate of the S1interface handover.
- S1 mode intra MME TAU success rate (%) which provides the Tracking area update success rate at the S1 interface.
- Security mode command success rate (%) which provides the success rate of security parameters exchanges.
- S1 mode packet paging success rate (%) which provides the paging success rate.
- S1 mode service request success rate (Network Cause) (%) which provides the success rate service request on the S1 interface.
- Maximum attached users (number) which provides the peak number of attached users registered within the time unit (1h).

- Maximum attached users at ECM-CONNECTED status (number) which provides the peak number of attached users in "ECM-CONNECTED" mode registered within the time unit (1h).
- Maximum attached users at ECM-IDLE status (number) which provides the peak number of attached users in "ECM-IDLE" mode registered within the time unit (1h).
- S1 mode maximum bearer number (number) which provides the peak number of S1 bearer.
- Maximum PDN connection number (number) which provides the peak number of PDN connections.

III.1.2. PGW KPI

Start Time	Period (min)	NE Name	UGW Function	SGi downlink user traffic in KB (kB)	SGi downlink user traffic peak throughput in KB/s (kB/s)	SGi uplink user traffic in KB (kB)	SGi uplink user traffic peak throughput in KB/s (kB/s)	S1-U downlink user traffic in KB (kB)	S1-U uplink user traffic in KB (kB)	S+PGW maximum simultaneously active bearers (number)
01/01/2019 23:00:0	60	YDEUGW	YDEUGW	490772098	157478,92	77425427	28531,25	509850197	81970199	3562,8333
01/02/2019 23:00:0	60	YDEUGW	YDEUGW	520048158	165316,04	83886289	30560,208	542910812	88976812	3701
01/03/2019 23:00:0	60	YDEUGW	YDEUGW	506158487	162872,54	78235506	29020,167	527562648	83535427	3676,9167
01/04/2019 23:00:0	60	YDEUGW	YDEUGW	513060187	165591,79	78492813	28583,667	534241197	83911534	3683,1667
01/05/2019 23:00:0	60	YDEUGW	YDEUGW	490760997	157916,17	71025944	25457,75	510234611	75768049	3537,7083
01/06/2019 23:00:0	60	YDEUGW	YDEUGW	441420460	143248,63	67001589	24946,375	457684388	70682604	3335,9167
01/07/2019 23:00:0	60	YDEUGW	YDEUGW	454367300	148130,33	72888864	27413,208	471198568	76662115	3519,5

Table 4. CAMTEL's SGW/PGW sample KPI data

Here is a sample file of data collected from the PGW of CAMTEL. From this file, we have the following counters:

- SGi downlink user traffic in KB which provides the total downlink traffic of users on SGi interface between the PGW and the Internet.
- SGi downlink user traffic peak throughput which provides the peak throughput of downlink traffic on the SGi interface.
- SGi uplink user traffic which provides the total uplink traffic of users on SGi interface between the PGW and the Internet.
- SGi uplink user traffic peak throughput which provides the peak throughput of uplink traffic on the SGi interface.
- S1-U downlink user traffic which provides the total downlink traffic on the S1-U interface between the SGW and E-Node Bs.
- S1-U uplink user traffic provides the total uplink traffic on the S1-U interface between the SGW and E-Node Bs.
- S+PGW maximum simultaneously active bearers which provides the maximum simultaneous active bearers on SGW/PGW.

III.2. CAMTEL's 4G model of traffic

III.2.1. Depend variable and explanatory variables

We have chosen to express the downlink traffic on the SGi interface as a function of other strongly correlated MME and SGW / PGW variables. Thus, our depend variable is the downlink user traffic on the SGi interface. To facilitate the writing of the equation of the model, we assigned letter to each variable as given in the table below.

KPI USN		KPI UGW	
A	S1 mode intra-MME S1 Handover success rate (%)	K	SGi downlink user traffic in KB (kB)
B	S1 mode intra-MME TAU success rate (%)	L	SGi downlink user traffic in packets (number)

C	Security mode command success rate (%)	M	SGi downlink user traffic peak throughput in KB/s (kB/s)
D	S1 mode packet paging success rate (%)	N	SGi uplink user traffic in KB (kB)
E	S1 mode service request success rate (Network Cause) (%)	O	SGi uplink user traffic in packets (number)
F	Maximum attached users (number)	P	SGi uplink user traffic peak throughput in KB/s (kB/s)
G	Maximum attached users at ECM-CONNECTED status (number)	Q	SGi IP data packets discarded for error (number)
H	Maximum attached users at ECM-IDLE status (number)	R	S1-U downlink user traffic in KB (kB)
I	S1 mode maximum bearer number (number)	S	S1-U downlink user traffic in KB (kB)
J	Maximum PDN connection number (number)	T	S1-U uplink user traffic in KB (kB)
		U	S+PGW maximum simultaneously active bearers (number)
		V	GW maximum simultaneously active subscribers (number)

Table 5. Assignment of variable names to each counter

In order to determine strongly correlated variables of the system, we used the Pearson formula and we obtain the correlation table below.

RESEAU LTE	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
A	SC																					
B		SC																				
C			SC																			
D				SC																		
E					SC																	
F						SC	SC		SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC
G						SC	SC		SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC
H								SC														
I						SC	SC		SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC
J						SC	SC		SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC
K						SC	SC		SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC
L						SC	SC		SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC
M						SC	SC		SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC
N						SC	SC		SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC
O						SC	SC		SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC
P						SC	SC		SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC
Q						SC	SC		SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC
R						SC	SC		SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC
S						SC	SC		SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC
T						SC	SC		SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC
U						SC	SC		SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC
V						SC	SC		SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC

Table 6. Correlation matrix of all the variables

Notice that SC means strongly correlated. From the correlation table, we can see that the following variables are strongly correlated to our depend variable K (SGi downlink user traffic): F (Maximum attached users), G (Maximum attached users at ECM-CONNECTED status), I (S1 mode maximum bearer number), J (Maximum PDN connection number), L(SGi downlink user traffic in packets), M (SGi downlink user traffic peak throughput in KB/s), N (SGi uplink user traffic in KB), O (SGi uplink user traffic in packets), P(SGi uplink user traffic peak throughput in KB/), R (S1-U downlink user traffic in KB), S (S1-U downlink user traffic in KB), T (S1-U uplink user traffic in KB), U (S+PGW maximum simultaneously active bearers) and V (GW maximum simultaneously active subscribers).

III.2.2. SPSS tool

SPSS is a software specially designed for statistical analyzes in the social sciences. It stands for Statistical Package for Social Sciences. It is a specialized statistical data processing software. It includes several modules:

1. Basic system
2. Regression models
3. Advanced models
4. Tables (tables)
5. Exact tests (exact tests)
6. Categories
7. Trends

8. Other specialized modules.

III.2.3. Equation of the model

Once we have determined depend variable and explanatory variables, we can now estimate our equation model which has to be like the expression below:

K_i verifies the linear model equation: $K = a_0 + a_1 N + a_2 F + a_3 G + a_4 I + a_5 J + a_6 L + a_7 M + a_8 O + a_9 P + a_{10} S + a_{11} N + a_{12} U + a_{13} V + \varepsilon_i$

With $a = (a_0, a_1, \dots, a_{13}) = (X'X)^{-1} X'K$

- Where $X^{'}$ is the transpose of the matrix of the measured values of the explanatory variables X. The dimension of X is $n \times (p+1)$. With p the number of explanatory variables. Notice that the first column of X is all 1.
- a is the vector of the coefficients associated with the explanatory variables its dimension is $(p+1) \times 1$.
- K is the vector of measured values of the dependent variable which we want to explain. Its dimension is $n \times 1$. Where n is the number of data records.

To determine the vector a which represents the coefficients associated with the explanatory variables, we used the SPSS tool and the results obtained are given in the capture below.

Elements of the model		B
1	(Constante)	-7052050,466
	Maximum attached users (number)	-2111,808
	Maximum attached users at ECM-CONNECTED status (number)	-4009,714
	SGi downlink user traffic in packets (number)	1,512
	SGi downlink user traffic peak throughput in KB/s (kB/s)	166,382
	SGi uplink user traffic in KB (kB)	,128
	SGi uplink user traffic in packets (number)	-,598
	SGi uplink user traffic peak throughput in KB/s (kB/s)	-97,142
	S1-U downlink user traffic in KB (kB)	,016
	S1-U uplink user traffic in KB (kB)	-,059
	S•PGW maximum simultaneously active bearers (number)	2182,650
	GW maximum simultaneously active subscribers (number)	3396,878

Table 7. Determination of coefficients of the model

After doing the required calculations with the SPSS tool and by replacing each coefficient with its value in the model equation, we obtain the equation:

$$K_{predicted} = -7052050{,}46 + N * 0.128 - F * 2111{,}808 - G * 4009{,}714 - I * 0{,}90 - J * 0{,}878 + L * 1{,}512 + M * 166{,}382 - O * 0{,}598 - P * 97{,}142 + S * 0{,}016 - N * 0{,}059 + U * 2182{,}650 + V * 3396{,}878$$

III.3. Model validation

III.3.1. Fisher test

The SPSS tool allows us to validate our model with the ANOVA test. Notice that the SPSS tool is only available in French language. The result of this test is given below:

ANOVA[a]

Modèle		Somme des carrés	ddl	Carré moyen	F	Sig.
1	Régression	9,970E+18	11	9,064E+17	80429,385	,000[b]
	de Student	7,437E+15	660	1,127E+13		
	Total	9,977E+18	671			

Table 8. Results of statistic parameters from SPSS tool
Table 9.

From this result:

- "somme des carrés" represents the Sum of Estimated Squares (SES) for the line called "Régression". It also represents the Sum of Residual Squares (SRS) for the line called "de Student".
- ddl reprensents the Degree of Freedom (DF). It is equal to :
  - The number p of the explanatory variables for the line called "Régression" in SPSS 's perspective. In other words, it represents the number of explanatory variables which really impact the model.
  - The number $n - p - 1$ for the line called "de Student". Where n is the number of records measured.
- "Carré moyen" represents:
  - The Estimated Mean Square (EMS) for the line called "Régression".
  - The Residual Mean Square (RMS) for the line called "de Student".
- F represents the Fisher coefficient and is obtained with the formula $F = \frac{EMS}{RMS}$.
- Sig represents the significance of the Fisher test which validates the model if it is less than 0.005.

We can notice that the significance of Fisher test obtained is 0.0001^b, with $b > 1$, which is less than 0.005. So the Fisher test validates our model.

III.3.2. Coefficient of determination

The SPSS tool also allows us to calculate the coefficient of determination R^2. The table below gives the result obtained with our model.

Modèle	R	R-deux	R-deux ajusté
1	1,000[a]	,999	,999

Table 10. Coefficient of determination value

From this table, we can see that R^2 is equal to 0.999 which is very close to 1; so our model is valid.

III.4. Model analysis

Finally, we have determined the equation of traffic model of the CAMTEL's EPC network which is given by the formula:

$$K_{predicted} = -7052050{,}46 + N * 0.128 - F * 2111{,}808 - G * 4009{,}714 - I * 0{,}90 - J * 0{,}878 + L * 1{,}512 + M * 166{,}382 - O * 0{,}598 - P * 97{,}142 + S * 0{,}016 - N * 0{,}059 + U * 2182{,}650 + V * 3396{,}878$$

From this equation, we can notice that the following variables act negatively on the explained variable called downlink user traffic:

- F: Maximum attached users (number)
- G: Maximum attached users at ECM-CONNECTED status (number)
- I: S1 mode maximum bearer number (number)
- J: Maximum PDN connection number (number)
- O: SGi uplink user traffic in packets (number)
- N: SGi uplink user traffic in KB (kB)
- P: SGi uplink user traffic peak throughput in KB/s (kB/s)

So, apart from the variables O, N and P, the other (F, G, I and J) variables must normally increase with the downlink user traffic because they all represent the activity of subscribers in the network. But unfortunately, their growth decreases the downlink traffic. The model shows then a problem at the radio side because the downlink traffic must increase with the growth of active subscribers in the MME. Probably it shows that radio side is congested. So our model shows clearly that actions must be taken at the radio side to optimize the traffic.

IV. Conclusion

At the end of this paper which consisted in modeling the CAMTEL's EPC traffic, we obtained a valid model as it reacted positively to the Fisher test. The model showed that there may be a congestion problem at the radio side. Thus, it would be interesting to extend this work at the radio side.

References

[1]. ORANGE , IUT WILLIAMS JINKS, Architecture LTE, Hachette, 2020, p. 34.
[2]. ERICSSON and electronics company, Duplexage LTE, turin: ERICSSON, 2014, p. 4.
[3]. IUT/3GPP DYLAN FRONTY, La-structure-generale-dune-trame-radio-ZigBee, ENSA, 2005.
[4]. IUT/3GPP DYLAN FRONTY, La-structure-generale-dune-trame-radio-ZigBee, ENSA éd., 2010: ENSA, p. 9.
[5]. IUT/IEEE XAVIER recherche des outils: internet et services, Interfaces LTE /, ENSA, 2010, p. 76.
[6]. Technologies, Huawei, USN SYSTEM OVERVIEW, 2010, p. 6.
[7]. Technilogies, Huawei, USN 9810 SYSTEM OVERVIEW, HUAWEI, 2010, p. 2.
[8]. Technologies, Huawei, UGW 9811 SYSTEM OVERVIEW, 2010.
[9]. L. p. d. v. R. Bourbonnais et J.-C. Usunier, MODELE DE LISSAGE EXPONENTIELLE / La prévision des ventes,, Le léopard masqué Economica, 2009, p. 1.
[10]. V. O. T. B. R. P. F. Fortin, AUTOREGRESSIF METHODS Revue, HAL, (1997), p. 20.
[11]. MA Modèles stationnaires de séries chronologiques LA MOYENNE MOBILE, NEW YORK: Applied Econometric Time Series (Deuxième éd.), 2004, p. 1.
[12]. T. Dochy, Mehdi Danech-Pajouh, Yves Lechevallier, method AR, HAL éd., HAL, 2016, p. 34.
[13]. V. O. e. R. W. Schafer, ARIMA prévision de ventes, economia, 2008, p. 1.
[14]. N. n. f. p. r. C. M. Bishop, NEURONAL MODELS, Oxford, 2000, p. 45.
[15]. L. M. Y. K. . A. Cornuéjols, FILTRE / APRRENTISSAGE ARTIFICIEL, 2003, p. 16.
[16]. P. (. S. e. R. (. F.), PREMISSE MUTIPLE REGRESSION, STAT 113 , 2006, p. 7.
[17]. G. DREYFUS, Paramétre régresssion, TUNIS: Encyclopædia Universalis [en ligne], 1999, p. 4.
[18]. F. Blayo et M. Verleysen, MOD NEURONAL Les Réseaux de Neurones Artificiels, Eyrolles, 2000, p. 23.
[19]. PINDYCK (R. S.) et RUBINFELD (D. F.), cohen econemetric models, Eyrolles éd., new york, 1981..
[20]. M MOUSSAOUI, Structure trame LTE / Présentation de la technologie LTE, ENSA éd., Publié le 5 oct. 2012, p. 5.
[21]. IDIOU Ghania, Régression et modélisation par les réseaux de neurone, Mémoire pour l'obtention du diplôme de Magistère en Mathématiques, Université Mentouri, Constantine, Fasculté des Sciences Exactes., Publié le 30 Juin 2009, p. 49.

Alphonse BINELE ABANA. "CAMTEL's EPC traffic modeling to help in decision-making." *IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)* 15(4), (2020): 08-18.