Modelling RACH Arrivals and Collisions for Human Type Communication

***Abstract*—The paper proposes an analytical model to evaluate the collision probability on the Random-Access CHannel (RACH) in LTE systems as a function of the number of User Equipments (UEs), the number of available preambles and the Inter–arrival times of the RACH Requests (IRR) of the average user. The model for the IRR of the average user is obtained from real traffic data captured at the eNodeB of a mobile operator, and is derived by emulating the RRC state machine for different Radio Resource Control Inactivity Timer (RRC_{IT}) settings. The results of the presented study suggest that when RRC_{IT} is set to a few seconds a mixture model is more accurate than the Poisson hypothesis both in modelling the IRR and in estimating the RACH performance.**

***Index Terms*—RACH Collision Probability, Random Access Opportunity (RAO), Mixture Model, RRC State Machine, Radio Resource Control (RRC) Inactivity Timer**

I. Introduction

In LTE systems, an idle UE must access the RACH channel to synchronize with the network before initiating any data transfer. Upon a successful RACH procedure, radio resources are assigned to the UE that switches from the Radio Resource Control (RRC) Idle State to the RRC Connected State. Conversely, the opposite transition to the RRC Idle State is regulated by the RRC Inactivity Timer (RRC_{IT}). In the RRC Idle state, the RF circuitry has a near zero energy consumption. Hence, lower RRC_{IT} values reduce the device energy consumption at the cost of increasing the signaling load. As a result, other than its own generated traffic, the number of RACH accesses performed by an UE is highly influenced by the value of RRC_{IT}. Today, eNodeBs can serve up to a few thousands of UEs per cell. Hence, the traffic generated by the RACH access procedures of the serviced UEs may deteriorate the performance of the RACH channel. 3GPP [1] defines six key performance indicators to assess new proposals of RACH operations. Theoretical results allow to evaluate such performance parameters by exploiting the Poisson hypothesis on the inter–arrival times of the RACH requests [2].

To the best of our knowledge, few works have modeled the RACH request traffic starting from actual data acquired on a commercial eNodeB. The results of this study i) show the inaccuracy of the Poisson hypothesis in modelling the offered RACH traffic, ii) point out that for RRC_{IT} set to a few seconds, the Poisson hypothesis provides a conservative estimation of RACH performance. This last result should stimulate the research for revising some optimization procedure on the RACH channel based on the Poisson hypothesis.

The rest of the document is organized as follows. After a short description of the RACH procedure in section II, the model for RACH inter–arrival times is given in section III. RACH collision probabilities are derived in section IV and assessed in section IV-A. Finally, section V concludes the paper with final remarks.

II. Background on RACH access procedure

The random access procedure for requiring radio resources is detailed in [3]. Our analysis focuses on the random access procedure triggered when an UE in the RRC Idle state has to receive/send a packet, i.e. when a transition to RRC Connected is needed. To this aim, the UE starts the random access procedure by sending the Random Access CHannel Preamble (RACHP) on the Physical RACH (PRACH). The contention-based operation of the RACH is based on (multi–channel) ALOHA–type access, where each UE transmits its preamble in the first available random–access slot. As detailed in [4], the “PRACH Configuration Index” (*PrachConfigIndex*) defines when the UEs can send a RACH request. To distinguish UEs sending RACH requests simultaneously within the same timeslot, each UE randomly selects a RACHP out of a pool of available RACHPs. The maximum pool dimension is 64, but it can be restricted by the eNodeB to reserve some preambles for contention free procedure (e.g., during handover). In summary, the number of Random Access Opportunities (RAOs) per second available in an LTE cell depends on the *PrachConfigIndex* and on the number of RACHP reserved by the eNodeB for contention based procedures. When two or more UEs simultaneously attempt to access the RACH channel by using the same RACHP, a collision occurs and a retransmission mechanism with backoff algorithm is started.

The preamble collision probability is one of the performance indicators defined in 3GPP [1] to be considered in evaluating the goodness of new proposals that intend to improve the operation of RACH. In reference to [1], we elaborate on both the definitions of collision probability given in section 6.3 ($P_c^{(RAO)}$) and in ANNEX B ($P_c^{(HTC)}$).

III. Modelling inter–RACH times

The first goal of our analysis is to derive a statistical model for the inter–arrival times of the UE RACH requests. In general, such values depend on the configuration of the eNodeB (specifically, in terms of RRC_{IT}) and on the traffic generated by end-users. Our starting point is the traffic data acquired by passive measurements carried out at a commercial eNodeB of a mobile operator.

The monitored eNodeB is operative in the bandwidth of 1800 MHz, and is located in a business area of Turin, Italy. Several one week long measurement sessions have been carried out in different periods of the years 2013–2014. The different sessions mostly spot the same qualitative results. The

TABLE I
STATISTICAL PARAMETERS OF THE IRR_j TIMESERIES

Dataset	Size	Mean (s)	Covariance (s^2)	Coef. Var.
IRR_2	22654	47.46	1.98e+04	2.96
IRR_5	13424	80.03	3.16e+04	2.22
IRR_{10}	9106	117.73	4.48e+04	1.80

simulation results refer to the most recent dataset, namely the one acquired in February 2014. During each session, the traffic flowing through the S1 interface is captured. Out of the available data, in this study we focus on the packet traffic of each active user. Upon the necessary anonymization procedures and after stripping out privacy–sensitive data, the acquired traffic dump consisted of a list of entries each one reporting the IP source and destination addresses of each observed packet along with its timestamp. The actual information on the RACH channel was not used since the RRC_{IT} was set by the carrier without accounting for energy aspects.

To obtain the inter–arrival times of RACH requests for each user, an emulator of the RRC state machine has been developed. For each user (say, user i), the input of the emulator is the set of the packet inter–arrival times associated with it. Taking into account the RRC_{IT} setting, the output of the emulator is the timeseries representing the *Inter–arrival times between successive RACH Requests* (IRR) of the user i. We emulated the RRC state machine for different RRC_{IT} values. The RRC_{IT} settings considered in the study (i.e. $RRC_{IT} = 2, 5, 10$ s) have been chosen taking into account the energy savings results of [5]. The timeseries IRR_j, $j = 2, 5, 10$, is obtained by merging the IRR obtained for all users i when the RRC_{IT} is set to j. The timeseries IRR_j represent the observations of inter–arrival times of the triggered RACH procedures for the average user. Some statistical properties of these timeseries are shown in Table I. Obviously, with the same input traffic, the lower the RRC_{IT} values, the higher the number of triggered RACHs (hence, the lower the mean of IRR values) generated by the emulator of the RRC state machine. The coefficient of variation (CV), which represents a standardized measure of dispersion of a probability distribution, is defined as the ratio of the standard deviation to the mean. Table I shows CV values higher than 1 in all of the datasets. Hence, the exponential distribution (having $CV = 1$) seems to be inadequate. Different alternative distribution families could be considered (e.g., Gamma, Cox, etc.) to obtain $CV > 1$. Previous analyses, such as [6], show that LTE traffic is generated by a set of applications/services that have different behaviors, both at session level and at packet level. Then, we assume that the resulting IRR timeseries is obtained from a sum of behaviors, each one associated with the different applications/services used by the average user. The IRR generated by each application/service can be modeled by an exponential distribution with its characteristic rate. This assumption suggests to model IRR timeseries by i.i.d. random variables with density equal to a mixture of exponentials:

$$f(x) = \sum_{c=1}^{C} \alpha_c \lambda_c e^{-\lambda_c x} \tag{1}$$

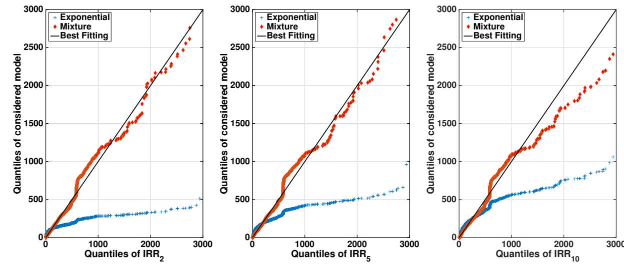

Fig. 1. QQ curves for $RRC_{IT} = 2, 5, 10$s (Left to Right)

TABLE II
ESTIMATED PARAMETERS OF THE MIXTURE MODEL

Dataset	α_c	λ_c
IRR_2	0.7782, 0.0955, 0.0102, 0.1160	0.0804, 0.0192, 0.0013, 0.0046
IRR_5	0.1935, 0.0149, 0.7916	0.0043, 0.0013, 0.0334
IRR_{10}	0.0291, 0.2269, 0.7441	0.0015, 0.0039, 0.0185

where C is to the number of components in the mixture, α_c is the mixing coefficients (with $\sum_{c=1}^{C} \alpha_c = 1$), and λ_c is the parameter of the c-th exponential component. Given that such different behaviors can be clustered, we assume that the number of components of this model should not be too high.For each considered RRC_{IT} value, the Bayesian model selection approach has been applied to estimate the mixture parameters of timeseries IRR_j, with $j = 2, 5, 10$. In particular, we referred to the algorithm presented in [7], where the prior distribution assumes that α_c and λ_c are independent. The symmetric Dirichlet distribution is used as a prior of α_c, whereas the gamma distribution $\Gamma(a_0, b_0)$ is selected as the conjugate prior distribution on λ_c. As suggested in [7], a_0 is set to a small value (we set $a_0 = 0.1$) and b_0 is chosen so that the prior mean matches the mean $\overline{Y}$ of the data, i.e., $b_0 = a_0 * \overline{Y}$. The Markov Chain Monte Carlo (MCMC) algorithm described in [8] has been used for the estimation of the model parameters for different number of components. To estimate the number of components, the Bayes Information Criterion (BIC) technique (see [8] for details) has been used. The results indicate that $C = 3$ is the optimal value for IRR_5 and IRR_{10}, whereas 4 components are needed for the IRR_2.

The estimated parameters are summarized in Table II. In addition, Table II shows that more than 74% of samples of all timeseries can be modeled with exponential distribution with a mean value around 50 s or less (λ_c are around 0.02 s^{-1} or higher), whereas the remaining observations are in the range of 250 s (i.e., λ_c around 0.004 s^{-1}) and 750 s ((i.e., λ_c around 0.0013 s^{-1}).

Figure 1 shows the quantile–quantile (QQ) plot for the different timeseries. Each subfigure reports the QQ curves obtained by comparing the actual dataset to the “Exponential” model, and to the “Mixture” model. In the first case, the parameters are set according to the mean values reported in Table I, while in the second case according to Table II. To immediately visualize the quality of the fitting results, the reference “Best Fitting” curve is also reported. The figure shows that for the IRR_2 and IRR_5 timeseries, the points of the curve “Mixture” lay very close to the “Best Fitting” curve, whereas they deviate for high quantile values in the case of IRR_{10}. Conversely, we observe that the points of the

"Exponential" curve are always far from the "Best Fitting" curve.

IV. RACH Collision Probability

In this section we derive the probability of collision for RACH procedures initiated by different stations under a quite general RACH request model. The result is then specialized for the mixture model (1) and used next in the performance evaluation section IV-A. To this aim, let us first indicate the number of stations in the RRC Idle state with N and the number of available preambles with k. In addition, let T be the duration of each timeslot. At each timeslot, any station i (with $i \leq N$) in RRC Idle state may place a RACH request by requesting a given preamble j, with $j \leq k$.

Let $A_i(n) \in \{0,1\}$ be the number of requests placed by station i at time nT and with $A_{i,j}(n) \in \{0,1\}$ the number of requests placed by station i by selecting preamble j.

Assume now that stations in RRC Idle state may each independently place a request with probability p, i.e. $\Pr\{A_i(n) = 1\} = p$ and that preambles are selected uniformly at random with probability $1/k$. As such: $\Pr\left\{A_{i,j}(n) = 1\right\} = \frac{p}{k}$. Consider now the total number of requests $X_j(n)$ to preamble j at time nT: $X_j(n) = \sum_{i=1}^{N} A_{i,j}(n)$. At each time n, and for each preamble, RACH collision occurs whenever $X_j(n) \geq 2$. The random variable $X_j(n)$ has binomial distribution with parameter p/k, and the number of RACH request failures $L_j(n)$ is:

$$L_j(n) = \begin{cases} 0 & \text{if } X_j(n) < 2 \\ X_j(n) & \text{otherwise} \end{cases} \tag{2}$$

Note that, since we assume the system at the statistical equilibrium, the dependence on time can be safely omitted. Hence, the mean number of RACH collisions to each single preamble is $\mathbb{E}\left[L_j\right] = \frac{Np}{k}\left(1-\left(1-\frac{p}{k}\right)^{N-1}\right)$, and the overall RACH collision probability $P_C^{(HTC)}$ is:

$$P_C^{(HTC)} = \frac{\mathbb{E}\left[L_j\right]}{\mathbb{E}\left[X_j\right]} = 1-\left(1-\frac{p}{k}\right)^{N-1} \tag{3}$$

From the point of view of an external observer, instead, it is easy to prove that the collision probability of a generic RAO is given by:

$$\begin{aligned} P_C^{(RAO)} &= 1-\frac{\text{\# of idle RAOs in } \tau \text{ s.}}{\text{\# of RAOs in } \tau \text{ s.}}-\frac{\text{\# of successful RAOs in } \tau \text{ s.}}{\text{\# of RAOs in } \tau \text{ s.}} \\ &= 1-\left(1-\frac{p}{k}\right)^{N}-\frac{Np}{k}\left(1-\frac{p}{k}\right)^{N-1} \end{aligned} \tag{4}$$

Equations (3) and (4) assume the probability p of a RACH request for any idle station at each timeslot is known. Hence, the next step is to analytically find the value of p.

In the previous section, we assumed the process of RACH requests to be a point process $\{t_1, t_2, \dots\}$ in which the inter–arrival times $\{X_1, X_2, \dots\}$ form a sequence of i.i.d. random variables each of them with probability density function given by (1). More generally, by indicating the common pdf with $f_X(x)$ and the common probability distribution with $F_X(x)$, the overall arrival process is a *renewal process* [9] and the arrival events are called *renewals*. Under this assumption, finding the value of p is equivalent to computing the probability that one renewal occurs in the generic time interval $((n-1)T, nT]$. Notice that, depending on the renewal distribution, the inter–arrival times can be arbitrarily small so that more than one renewal event may occur in the same timeslot. However, this phenomenon is not possible in practice and, as it will be shown, its probability becomes analytically negligible for physically reasonable model parameters and small timeslots.

In order to compute p, let us consider a generic timeslot that begins at time t. As the beginning of the timeslot is asynchronous with the arrival process, the first renewal following time t occurs after time V, while the second, third, etc. occur after time $V+X_2$, $V+X_2+X_3$, $V+X_2+X_3+\dots$. The time V is called *forward recurrence time*, while $X_2, X_3, \dots\dots$ are usual renewal times and the resulting renewal process is referred to as *modified*. When t is large enough so that the system has been running for long, it can be proven [9] that the forward recurrence time has pdf $f_V(x) = \frac{1-F_X(x)}{\mathbb{E}[X]}$. Under the previous hypothesis, the overall process is called *equilibrium renewal process* and the probability p of at least one arrival in the timeslot of length T is:

$$p = 1-\Pr\{V > T\} \tag{5}$$

The RACH collision probability and the RAO collision probability are then readily obtained by substituting (5) in (3), and (4), respectively. Finally, expression (5) can be specialized when the pdf of renewals is that of (1). Indeed, in this case,

$$p = 1-\frac{\sum_{c=1}^{C}\frac{\alpha_c}{\lambda_c}e^{-\lambda_c T}}{\sum_{c=1}^{C}\frac{\alpha_c}{\lambda_c}} \tag{6}$$

So far, the value of p has been computed as the probability of the complementary event of observing zero arrivals in a timeslot. For the sake of completeness, the probability of having *exactly* one arrival in a timeslot is given by:

$$\begin{aligned} &\Pr\{N(T) = 1\} = \Pr\{V \leq T\} - \Pr\{V+X \leq T\} = \\ &= \frac{\left(T\sum_{c=1}^{C}\alpha_c^2 e^{-\lambda_c T}+\sum_{c=1}^{C}\sum_{k\neq c}^{C}\frac{\alpha_c\alpha_k\left(e^{-\lambda_k T}-e^{-\lambda_c T}\right)}{\lambda_c-\lambda_k}\right)}{\sum_{c=1}^{C}\frac{\alpha_c}{\lambda_c}} \end{aligned} \tag{7}$$

and it is easy to prove that, under practical parameters configuration such as the ones of Table II, the difference between the numbers obtained by equations (7) and (6) is in the order of 10^{-4}.

A. Performance Evaluation

The accuracy of the proposed model is assessed by simulating a cell with 1000 UEs, each one independently placing RACH requests with inter–arrival times picked randomly from the IRR_j timeseries. Simulation runs are carried out with different RRC_{IT} settings (namely, 2, 5, and 10s), and the performance parameters (namely, $P_C^{(HTC)}$ and $P_C^{(RAO)}$) are estimated over an observation time of 1000 s. Figure 2 shows the average values of the $P_C^{(HTC)}$ and the $P_C^{(RAO)}$ along with the 95% confidence interval estimated over 10 simulation runs for the three RRC_{IT} configurations. Moreover, the figure reports the collision probability curves, $P_C^{(HTC)}$ and the $P_C^{(RAO)}$, derived in (3) and (4) respectively, using the values

TABLE III
SIMULATION RESULTS - $P_c^{(MTC)}$

Dataset and k	Mixture	Poisson	Simulation
$IRR_2 - K = 10$	0.0446	0.1113	0.0477
$IRR_2 - K = 30$	0.0161	0.0410	0.0172
$IRR_2 - K = 50$	0.0100	0.0251	0.0110
$IRR_5 - K = 10$	0.0260	0.0226	0.0262
$IRR_5 - K = 30$	0.0093	0.0081	0.0090
$IRR_5 - K = 50$	0.0055	0.0049	0.0060
$IRR_{10} - K = 10$	0.0171	0.0174	0.018
$IRR_{10} - K = 30$	0.0061	0.0062	0.0058
$IRR_{10} - K = 50$	0.0037	0.0038	0.0044

of p computed for the renewal times IRR_j modeled as in section III (i.e., p is set according to relation (6)). The results show an excellent adherence of the theoretical model with respect to the simulation outcomes as the analytical curves lay almost always within the 95% C.I. of simulations for all RRC_{IT} configurations. For clarity of visualization, we do not report the analytical curves obtained with the Poissonian model of the IRR_j, obtained using the relations shown in [2]. The estimated $P_c^{(MTC)}$ for some k values are summarized in Table III. The results refer to the theoretical estimation carried out with the proposed model (column `Mixture`) and the Possonian one (column `Poisson`). Such results are compared to the average $P_c^{(MTC)}$ estimated by simulation when using the actual IRR_j sequence (column `Simulation`) as input.

The results clearly show the higher accuracy of the mixture model with respect to the Poisson model for $RRC_{IT} = 2s$. On the contrary, for $RRC_{IT} = 5, 10s$ the mixture model provides marginal improvements of the accuracy of the performance estimation. Similar conclusions can be obtained when the $P_c^{(RAO)}$ is analyzed. The proposed theoretical model does not account for the backoff mechanism, i.e. it implicitly assumes that if two or more RACH requests collide, no successive attempts will be made. To evaluate the impact of this assumption on the performance evaluation accuracy, we modify the simulation tool to account for the backoff mechanism. We vary the Backoff Indicator (BI) in the range $[20, 320]$ ms, and set a maximum number of retransmissions to 10. The simulation results obtained loading the simulator with the IRR_j dataset and the P_c^{HTC} curves obtained from (3) are shown in figure 3. We report only the results for the border values of the tested BI range, i.e. 20 and 320 ms. The results obtained with the other tested BI values are within these two curves. The figure shows that the accuracy of the proposed model is slightly affected for $k < 10$ only. Again, for sake of visualization, we do not report the results for $P_c^{(RAO)}$, which lead to similar observations. Furthermore, we observed that only for $k < 3$ a small number of RACH requests (less than 10%) exceeded the maximum number of retransmissions and thus were not acknowledged by the eNodeB.

V. CONCLUSION

Upon the emulation of the RRC state machine with actual traffic and different RRC_{IT} settings, the paper presents a procedure to model the inter–arrival times of RACH requests by means of a mixture of exponential distributions. This procedure and the proposed analytical model provide an accurate

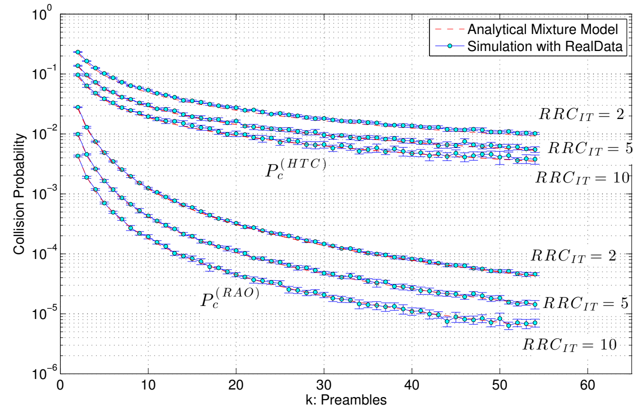

Fig. 2. Analytical model vs. simulation results - $P_C^{(HTC)}$ and the $P_C^{(RAO)}$ for different RACHP values, k

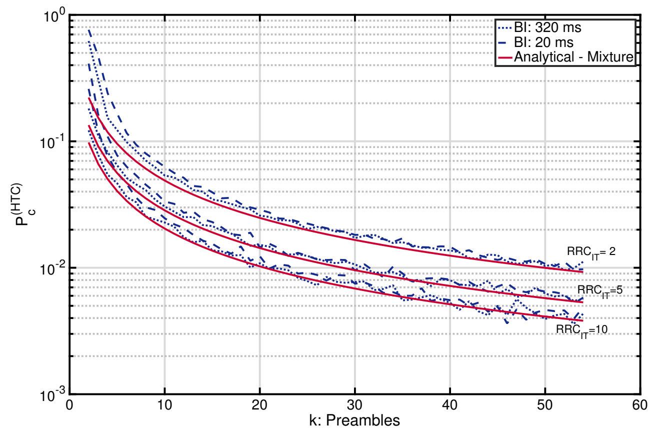

Fig. 3. Analytical model vs. simulation results - $P_C^{(HTC)}$ as a function of k, for different BI settings

estimation of the $P_C^{(HTC)}$ and the $P_C^{(RAO)}$ as a function of the number of UEs and the number of available preambles, for different RRC_{IT} settings.

ACKNOWLEDGMENT

This work has been partially supported by the PRA 2016 research project 5GIOTTO funded by the University of Pisa.

REFERENCES

[1] 3GPP TR 37.868 “RAN improvements for machine-type communication” v. 1.0.0, Oct. 2011.
[2] Cheng, R.-G. et al. “RACH collision probability for machine-type communications”, IEEE Vehicular Technology Conference, 2012.
[3] 3GPP TS 36.300 “E-UTRA and E-UTRAN; Overall description; Stage 2”, September 2014.
[4] 3GPP TS 36.211 “Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation”, September 2014.
[5] Foddis, G. et al. “Traffic Analysis for Signalling Load and Energy Consumption Trade-off in Mobile Networks”, Proc. of IEEE ICC 2015.
[6] Foddis, G. et al. “LTE traffic analysis and application behavior characterization”. Proc. European Conf. on Networks and Communications, 2014.
[7] Wagner, H. “Bayesian analysis of mixtures of exponentials”. Journal of Applied Mathematics, Statistics and Informatics 3, pp. 165-183, 2007.
[8] Frühwirth-Schnatter, S., “Finite Mixture and Markov Switching Models”, Springer, 2006.
[9] Cox, D.R., “Renewal theory”, Methuen’s monographs on applied probability and statistics, 1962.