

Energy Based Clustering Method to Estimate Channel Occupation of LTE in Unlicensed Spectrum

Daniel Malafaia¹(✉), José Vieira^{1,2}, and Ana Tomé²

¹ Instituto de Telecomunicações, Universidade de Aveiro, Aveiro, Portugal
danielmalafaia@ua.pt

² IEETA, Universidade de Aveiro, Aveiro, Portugal

Abstract. In this article we propose a subtractive histogram clustering method to estimate the number of LTE users based on the uplink energy probability density distribution sensed by an RF front-end. The energy of the signal is estimated and its histogram is analyzed to determinate the number of different distributions. As the energy estimation of the sensed LTE uplink can be modeled by a Gaussian mixture, this allow us to have *a priori* information that help us to determine the number of users. The lowest value Gaussian distribution can be used to accurately estimate the noise floor and the remaining distributions allow us to estimate the number of LTE users and their received power.

Keywords: Multiuser detection · 4G mobile communication · Detection algorithms · Radio spectrum management

1 Introduction

Unlicensed spectrum is one of society's most valuable resources. Certified devices can operate in it, under the legal regulations and radiated power limits without the obligation of paying license fees and with minimal license administration. These advantages make it very attractive model for low power and short range communications.

There is a current trend of implementing telecommunication protocols that used to be confined to licensed frequency bands, in the unlicensed spectrum. One of these implementations is Long-Term Evolution (LTE) in the unlicensed spectrum [8]. LTE on unlicensed spectrum allows an easier proliferation of smaller base stations with better spectral efficiency, easier network integration and better traffic load management. The critical aspect of LTE in unlicensed band is to ensure that it can co-exist with current access technologies such as WiFi. A fundamental necessity will then be the implementation of coexistence mechanisms that permits the best channel selection scenario for every protocol using the unlicensed spectrum.

In order for a communication protocol to take an informed decision of which band of the unlicensed spectrum to use, it should be able to determine the

channel occupation. The dynamical allocation of unlicensed channels must take into account the number of users that in each moment is using each channel.

The LTE uplink of the standard air interface transmission scheme is based on a Single-Carrier Frequency Division Multiple Access (SC-FDMA) which converts a wideband channel into a set of sub-channels called resource blocks [2]. For each User Equipment (UE) a number of resource blocks are attributed according to the user's needs.

The energy of the LTE uplink air interface channel can be used to detect the presence of a UE in an spatial region without the need for signal demodulation. It is expected that different users will be received with different powers levels, due to path loss and multi-path effect [9]. Therefore, it is possible to discriminate each individual LTE uplink UEs by the received energy level and estimate the numbers of users in the channel.

For the noise floor, in order to obtain a good estimation, we need to ensure that we have a time-slot where no users are transmitting. In the LTE uplink it should be expected in each uplink frame, free Resource Elements (REs). These free time/frequency resources can be used to dynamically obtain an estimation of the noise floor.

As the LTE uplink is a shared medium, at a given moment an RE can be either free or occupied. When none of the users has the RE allocated, then the slot is free and only the noise floor will be present (produced by the environment and by the RF front-end itself). When a user transmits in a RE the transmitted signal will have an approximately constant power. The PDF of a resource block will then be a mixture of both noise and the users PDF.

By using an energy estimator, that averages a large number of samples, then by the central limit theorem, more specifically De Moivre Laplace theorem [4], the histogram of each transmitting user will converge toward a Gaussian distribution. Therefore the sum of all users signal Gaussian distributions is considered a Gaussian Mixture Model (GMM). The lowest mean component will be the noise floor and the other components of the mixture are related with the signals transmitted by different users.

2 Proposed Method to Determine the Number of Users

In order to determine the number of users, their power level and the channel occupation we propose a subtractive histogram method. The proposed subtractive histogram method is a clustering technique capable of determining the number of transmissions, during a given time interval, based on an energy estimation analysis of a multiple user RF shared medium. The method requires the energy, of the sampled RF data, to be estimated as explained in Subsect. 2.1. The histogram of the energy estimation is then analysed as a GMM and its components are extracted with a subtractive histogram as explained in Subsect. 2.2.

2.1 Energy Estimation

The RF acquired data is sampled in In-phase and Quadrature (I/Q) and its energy is estimated in order to obtain its energy modeled by a Gaussian distribution as seen in Fig. 1.

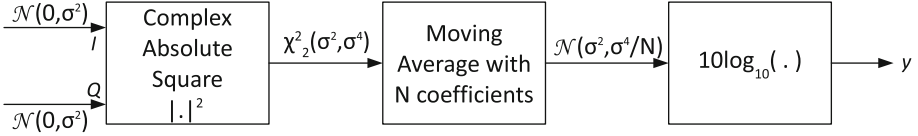


Fig. 1. Functional diagram of the energy estimator.

The complex I/Q data is assumed to be from a single user and can be approximately modeled by a Gaussian distribution with zero mean and a variance given by σ^2 . The I/Q signal is squared in order to obtain its instant energy, resulting in a signal with a Chi-Square distribution of second order. By analyzing the squared signal with an energy estimator of order N it is obtained, as previously mentioned, an approximate Gaussian distribution. The energy estimation is then converted to a dB scale.

For a moving average of high order N and assuming that the moving average energy output is concentrated around its mean, it can be demonstrated that the energy estimator output y , has the following statistical properties [5]:

$$E[y] \approx 10\log_{10}(\sigma^2) - \frac{10}{2N\ln(10)} \tag{1}$$

$$var[y] = v \approx \frac{100}{N\ln(10)^2}. \tag{2}$$

For a estimation using a large N , the mean value of the estimation is approximately the dB of the input signal energy and its variance will be mainly a function of the estimation order N . The energy estimation of a single user transmission will then be $y \sim \mathcal{N}(\sigma_x^2, \frac{\sigma_x^4}{N})$. In a scenario where the number K of users are transmitting in the RF channel in a time multiplexing way, the energy y can be model as a GMM with $K + 1$ components that includes the background noise.

2.2 Subtractive Histogram Method

Most of the existing algorithms available in the literature to get the parameters of each component in a GMM assumes that the number of components K is known [1]. In this work this number is unknown and may change with time. To overcome this problem we propose a greedy algorithm that iteratively search for the signals with higher probability and stops with the background noise component. To analyse the channel PDF we start by evaluating histogram of y .

In the obtained histogram, the received signal from each user will be a Gaussian component with the average energy, in dB, being its mean value. The histogram will then be a mixture of at most K Gaussian distributions, where K is the number of different UE received transmissions with unique power plus the noise floor.

The subtractive method iteratively searches, one by one, the Gaussian distribution components in the histogram of the y energy. Searching by the maximum value of the histogram and then using it as a starting value for a Gaussian fit. The method obtains then the mixture model that will better fit the y histogram and also estimates the number K of Gaussian distributions in the GMM.

The method iterates till most of the histogram y , that is defined as 90%, is fitted. The first step is to detect the maximum value on the histogram under evaluation, this will give us the user signal that occurs more often, or the most frequent contribution in the observation for the GMM. After determining the peak value, the coefficient of determination R^2 is obtained,

$$R^2 = 1 - \frac{\sum_{i=1}^n (h_i - \hat{h}_i)^2}{\sum_{i=1}^n (h_i - \bar{h})^2} \quad (3)$$

where n is the number of bins of the histogram being evaluated, h_i is the Gaussian distribution PDF for each i energy value, \bar{h} is its mean and \hat{h}_i is the y histogram [6]. The R^2 is calculated for multiple fitting Gaussian distributions, with a fixed v variance and a variable M mean. The tested interval for M is given by:

$$M \in [p - v, p + v] \quad (4)$$

where p is the detected peak value of the histogram. The value of M that generates the distribution with the closest fit to the original PDF (having a higher R^2) is selected and defined as the mean value that generates the closest Gaussian distribution fit for that energy interval. For this goodness of fit evaluation, it is defined an interval of three times the theoretical standard variation, where 97.7% of the Gaussian distribution energy is concentrated [3]. This interval around the detected peak is then used for analyzing the R^2 .

The difference between the calculated fit and the histogram data in that interval is calculated, if it is an acceptable fit, defined as R^2 being higher than a threshold value of 0.7, then the Gaussian is identified as an transmitting user or the noise floor. If the determined R^2 is low, then the analyzed data does not represent a Gaussian of the GMM and the obtained peak value position is ignored. Nevertheless the calculated Gaussian distribution, that fits the detected peak, is subtracted from the original histogram. The value of \hat{h}_i is updated by removing the Gaussian fit in the following manner:

$$\hat{h}_i^{(k)} \leftarrow \hat{h}_i^{(k-1)} - h_i^{(k-1)}, \quad (5)$$

where k is the iteration number. The algorithm will iterate and detect the next maximum value position till $\hat{h}_i^{(k)}$ reaches the majority of the y histogram data is fit by the method. After k iterations, if the remaining data is less than 10% of the original histogram data then the method ends.

3 Experimental Setup

To validate the proposed algorithm with real signals we performed a test using LTE FDD signals. These are composed by three Physical Uplink Shared Channel (PUSCH) symbols of the type TS36.104 Uplink FRC A3-3 [7] that occupy 15 Resource Blocks (RB) with a total of 3 MHz bandwidth. Each symbol is generated with an individual UE, a unique Radio Network Temporary Identifier and a data block containing random data. The modulation used is QPSK and the symbols are scheduling so that the three users are allocated in three sequential resource blocks on the same frequency band.

The complete frame used in this experiment as the following composition:

1. A blank resource block that has the same length as the PUSCH subframe;
2. The first user PUSCH subframe;
3. The second user's PUSCH subframe with 4 times higher amplitude than the first user;
4. The third user with a PUSCH subframe with 16 times higher amplitude than the first user.

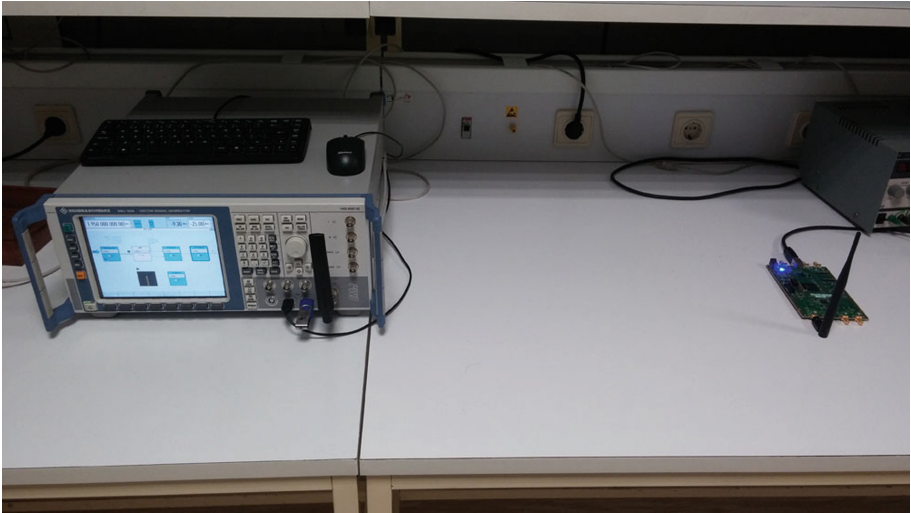


Fig. 2. The experimental setup to simulate multiple UE uplink transmissions.

The generated frame is transmitted by a Rohde & Schwarz SMJ100A at -25 dBm, centered at channel 1 Uplink (1950 MHz) and is intended to simulate the usage of the LTE uplink by three different UEs. The data will be received by the detection unit, an Ettus USRP B210 connected through USB3 to a personal computer as seen in Fig. 2. The B210 is configured to operate at 4MSPS in the same central frequency as the LTE signal is transmitted. The received

signal has 8900 samples and is analyzed with an energy estimator of order 50. This energy estimator has a theoretical variance of 0.38 given from Eq. 2. The output of the power estimator can be seen in Fig. 3. This data is then analysed using a histogram with a thousand bins and the algorithm from Sect. 2 is used with a threshold of 10% of the total energy of the histogram. Distributions with a goodness of fit, $R^2 < 0.7$ are ignored.

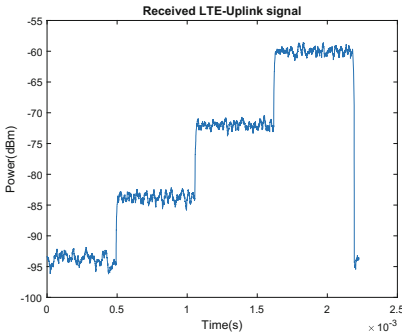


Fig. 3. Energy estimation of the received LTE-Uplink signal.

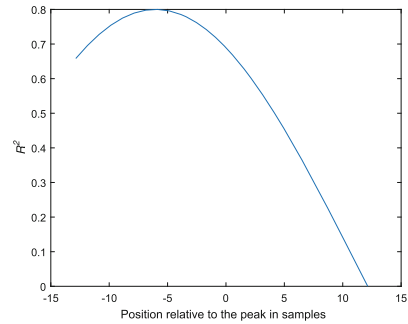


Fig. 4. Variation of the coefficient of determination around the first detected peak bin.

The first histogram peak value, is detected at -72 dBm and the fit of the mean value is illustrated in Fig. 4. The curve obtained by the coefficient of adaptation, shows that the first detected peak is approximately six bins away from the local optimum, this allows for a better fit than only use the detected maximum as the mean for the fitting Gaussian. This fit is then removed from the original PDF and a new iteration is run. This iterations of the applied algorithm can be seen in Fig. 5.

After detecting the four Gaussian distributions present on the signal segment, the remaining data represents 7.8% of the occurrences from the original generated histogram. The calculated R^2 values for the various detected distributions are the following: 0.8; 0.86; 0.92 and 0.86. As the remaining histogram data is below the defined threshold of 10%, the algorithm concludes the iterations. If that wouldn't be the case, then the next distribution that would be detected, by ignoring the data remaining threshold, would reduce the data left in the histogram to 6.6%. This detection would have a coefficient of determination of 0.04 that would be ignored by the method due to the low R^2 .

From the obtained distributions it is then possible to have a power estimation for the noise floor, that is given by the lowest detected power distribution. The other three detected distributions refer to the number of transmitting UE present on the shared medium.

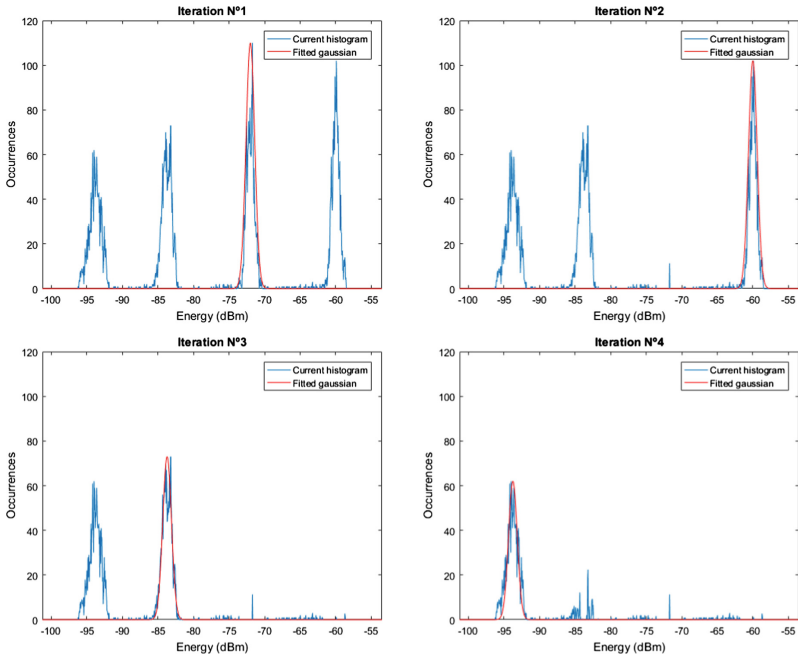


Fig. 5. Iterations of the subtractive histogram method for the LTE-Uplink data.

4 Conclusion

The subtractive histogram method proposed in this article was able to correctly determine the number of users and their power in a LTE channel using simulated PUCSH transmitted data using a signal generator and received in a software defined radio kit. This is done by analysing the data with an energy estimation method and an analysis based on fixed variance Gaussian distribution model. The method was also able to give a power estimation for the noise floor obtained in the detection unit.

Acknowledgment. This work was funded by the European Defence Agency under the pilot project grant with agreement number PP-15-INR-02_02_SPIDER.

References

1. El-Zaart, A.: Expectation-maximization technique for fibro-glandular discs detection in mammography images. *Comput. Biol. Med.* **40**(4), 392–401 (2010)
2. Holma, H., Toskala, A.: *LTE for UMTS - OFDMA and SC-FDMA Based Radio Access*. Wiley, Hoboken (2009)
3. Katayama, T., Sugimoto, S.: *Statistical Methods in Control & Signal Processing*. CRC Press, Boca Raton (1997)

4. Loève, M.: Fundamental limit theorems of probability theory. *Ann. Math. Stat.* **21**(3), 321–338 (1950)
5. Malafaia, D., Vieira, J., Tomé, A.: Adaptive threshold spectrum sensing based on EM algorithm. *Phys. Commun.* **21**, 60–69 (2016)
6. Renaud, O., Victoria-Feser, M.P.: A robust coefficient of determination for regression. *J. Stat. Plann. Infer.* **140**(7), 1852–1862 (2010)
7. Technical Specification: LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) conformance testing (3GPP TS 36.141 version 10.1.0 Release 10). Technical report (2011)
8. Syrjäälä, V., Valkama, M.: Coexistence of LTE and WLAN in unlicensed bands: full-duplex spectrum sensing. In: Weichold, M., Hamdi, M., Shakir, M.Z., Abdallah, M., Karagiannidis, G.K., Ismail, M. (eds.) *CrownCom 2015. LNICST*, vol. 156, pp. 725–734. Springer, Cham (2015). doi:[10.1007/978-3-319-24540-9_60](https://doi.org/10.1007/978-3-319-24540-9_60)
9. Zuniga, M., Krishnamachari, B.: Analyzing the transitional region in low power wireless links. In: *2004 First Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks, IEEE SECON 2004*, pp. 517–526 (2004)