CrossMark

Empirical Statistical Model for LTE Downlink Channel Occupancy

Mohamed Hamid[1,2] · **Niclas Björsell**[2] · **Slimane Ben Slimane**[1]

Abstract This paper develops an empirical statistical channel occupancy model for downlink long-term evolution (LTE) cellular systems. The model is based on statistical distributions mixtures for the holding times of the channels. Moreover, statistical distribution of the time when the channels are free is also considered. The data is obtained through an extensive measurement campaign performed in Stockholm, Sweden. Two types of mixtures are considered, namely, exponential and log-normal distributions to fit the measurement findings. The log-likelihood of both mixtures is used as a quantitative measure of the goodness of fit. Moreover, finding the optimal number of linearly combined distributions using the Akaike information criterion is investigated. The results show that good fitting can be obtained by using either exponential or log-normal distributions mixture. Even though, the fitting is done for a representative case with a tempo-spatial consideration, the model is yet applicable in general for LTE and other cellular systems in a wider sense.

Keywords LTE · Cellular traffic · Channel occupancy model · Exponential mixture fitting · Log-normal mixture fitting · Akaike information criterion (AIC)

✉ Mohamed Hamid
mohamed.hamid@uia.no; mhamid@kth.se

Niclas Björsell
nbl@hig.se

Slimane Ben Slimane
slimane@kth.se

[1] Communication Systems Lab, The Royal Institute of Technology (KTH), 16440 Stockholm, Sweden

[2] University of Gävle, Kungsbacksvägen 47, 80176 Gävle, Sweden

1 Introduction

A need for different data rates in mobile broadband systems has been rapidly growing in recent years. In that regard, long term evolution (LTE) has been provided by the 3rd generation partnership project (3GPP) as a standard for packet based adaptive data rate systems [1]. LTE has been further developed to LTE advanced (LTE-A) to provide higher data rates and more spectral efficiency [2]. For robust optimization for cellular systems in general and LTE systems in particular, the traffic demand of cellular networks is needed to be modelled.

Beside resource optimization, other several optimization problems in cellular networks impose traffic modelling such as performance evaluation and billing. Among the statistics used for traffic evaluation in cellular systems is the channel occupancy which is defined as the time that a user occupies a channel in a cell while it is located in the serving area of that cell [3]. The channel usage for a cellular system is modelled as a two states Markov chain process [4]. The first state is the busy state when the channel is being assigned for a user whereas the second state is the idle state when the channel is idle.

Many studies have been carried out to characterize the cellular channel occupancy statistical distribution. In [5], it is shown that mobile telephony channel occupancy can be approximated by exponential distribution. A great advantage of the exponential distribution is the traceability in finding analytical solutions for optimization problems. Therefore, exponential distribution has been intensively used to model cellular channel occupancy, see [4] as an example. Nevertheless, many research findings concluded poor similarity between exponential distribution and empirical data [6]. One of the main disagreements between exponential distribution and empirical data is the heavy tail behaviour for the empirical channel occupancy which is not properly characterized by exponential distributions. Therefore, some heavy tail distributions are used as alternatives to model the cellular channel occupancy, among which, the log-normal distribution is found to better fit the empirical data [7, 8].

Even though many studies were carried out to model the cellular channel occupancy, non of these studies considers LTE yet. Therefore, LTE channel occupancy modelling is a topic that needed to be studied which is the main contribution of this paper. Furthermore, this paper contributes also in exploring fitting the empirical data for the cellular channel occupancy into a mixture of either exponential or log-normal distributions, combined linearly. This contribution is seen by using LTE as an example of a cellular system.

Using distribution mixture is motivated by keeping the advantageous of the ease of exponential and log-normal distributions. Hence, we can avoid using complicated distribution to model the cellular channel occupancy such as Beta and Kumaraswamy distributions [9]. Moreover, distribution mixtures are more general than single distributions and can be used to fit the data under different conditions. Consequently, the algorithms developed based on exponential and log-normal distributions of cellular channel occupancy can still be used based on their mixtures with small changes considering the linear combination of many of them.

The rest of this paper is structured as follows: Sect. 2 handles the theoretical aspects of the paper including the channel usage model and using distributions mixture to fit data. Section 3 shows the measurements setup and the fitting results. Finally, Sect. 4 concludes the paper.

2 Theory

The theoretical aspects of the paper are handled in this section. The section starts with presenting the Markov based model for the channel occupancy. Following that, distribution mixture fitting mathematical framework is introduced. Furthermore, exponential and log-normal distributions mixture fitting are studied in particular.

2.1 System Model

The LTE channel usage can be modelled as a two states Markov process. These two states are the ON state representing occupied channel state and the OFF state denoting the channel being idle. ON and OFF states temporal length are random variables (RV). Hereafter, ON and OFF temporal length are assigned the RVs x and y respectively. Figure 1 exhibits the channel usage model. The problem tackled throughout this paper is how to find statistical distributions that fit x and y.

The rest of this section provides the theoretical aspects of distributions mixture fitting in general and exponential and log-normal mixtures fitting in particular.

Without lose of generality, the RV x is considered in the coming parts of this paper. The same findings of x can be applied to y. Denote the empirical probability density function (pdf) of x as $g(x)$. $g(x)$ can be fitted with a linear combination of k pdfs as

$$g(x) \approx \sum_{i=1}^{k} p_i f(x|\mathbf{\Theta}_i), \tag{1}$$

where

$$0 < p_i < 1 \quad \forall i, \quad \sum_{i=1}^{k} p_i = 1,$$

p_i is the weight of the pdf number i, $f(\cdot)$ denotes a single pdf and $\mathbf{\Theta_i}$ is the distinct distribution parameters for the pdf number i. For the whole mixture model, $\mathbf{\Omega}$ contains all the distinct mixture parameters and defined as

$$\mathbf{\Omega} = \left[p_1, \ldots, p_k, \mathbf{\Theta}_1^T, \ldots, \mathbf{\Theta}_k^T \right]^T, \tag{2}$$

with $(\cdot)^T$ denoting the transpose.

An important notice here is that the formulation of $\mathbf{\Omega}$ given in (2) assumes that the mixture is composed of the same distribution type which is considered in this paper. The goodness of fit is judged through the log-likelihood estimator, $L(x|\mathbf{\Omega})$, found as
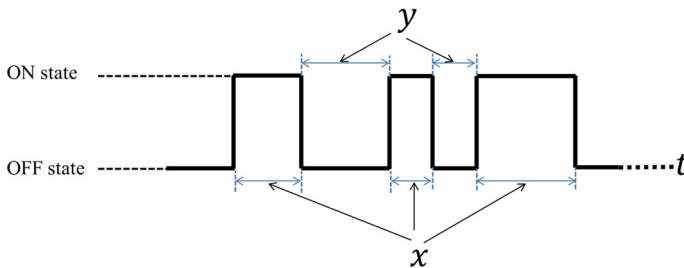

Fig. 1 Markov channel usage model

$$L(x|\mathbf{\Omega}) = \int g(x) \, \log \left(\frac{f(x|\mathbf{\Omega})}{g(x)} \right) dx. \tag{3}$$

2.2 Exponential Distributions Mixture Fitting

In [10] a linear combination of exponential pdfs is introduced to fit a heavy tail distributed data. For exponential mixture distribution, the pdf number i has a form as in (4a) while the collection of the distinct parameters, $\mathbf{\Omega}_{exp}$, is expressed in (4b).

$$f(x|\mathbf{\Theta}_i) = \lambda_i e^{-\lambda_i x}, \tag{4a}$$

$$\mathbf{\Omega}_{exp} = \begin{pmatrix} p_1 & \lambda_1 \\ \vdots & \vdots \\ p_k & \lambda_k \end{pmatrix}. \tag{4b}$$

The rest of this subsection shows how to find $\mathbf{\Omega}_{exp}$ as the essence of [10]. The process of finding $\mathbf{\Omega}_{exp}$ is a recursive procedure and starts with fitting the tail and moving backwards. Starting from the assumption that the part of the tail where $x > c_1$ can be fitted exclusively with the first exponential distribution, then

$$p_1 e^{-\lambda_1 c_1} = F^c(c_1), \tag{5a}$$

$$\sum_{i=2}^{k} p_i e^{-\lambda_i x} = 0 \quad \text{for } x > c_1. \tag{5b}$$

where $F^c(x)$ is the empirical complementary cumulative distribution function (CCDF) of x. Similarly, $p_1 e^{-\lambda_1 b c_1} = F^c(bc_1)$ where $b > 1$. Accordingly, the first pair, (λ_1, p_1) is found as

$$\lambda_1 = \frac{1}{(b-1)c_1} \, \ln \left(\frac{F^c(c_1)}{F^c(bc_1)} \right), \tag{6a}$$

$$p_1 = F^c(c_1) e^{\lambda_1 c_1}. \tag{6b}$$

Following the same idea, the pairs (λ_i, p_i) for $2 \leq i \leq k$ are found as

$$\lambda_i = \frac{1}{(b-1)c_i} \, \ln \left(\frac{F_i^c(c_i)}{F_i^c(bc_i)} \right), \tag{7a}$$

$$p_i = F_i^c(c_i) e^{\lambda_i c_i}, \tag{7b}$$

where

$$c_i = c_1 \alpha^{-(i-1)}, \quad \alpha > b,$$

$$F_i^c(c_i) = F_{i-1}^c(c_i) - \sum_{j=1}^{i-1} e^{-\lambda_j c_i},$$

$$F_i^c(bc_i) = F_{i-1}^c(bc_i) - \sum_{j=1}^{i-1} e^{-\lambda_j b c_i}$$

and

$$F_1^c(x) = F^c(x).$$

Finally the last pair (λ_k, p_k) is found as

$$p_k = 1 - \sum_{j=1}^{k-1} p_j, \tag{8a}$$

$$\lambda_k = \frac{1}{c_k} \ln \left(\frac{p_k}{F_k^c(c_k)} \right). \tag{8b}$$

The values of c_1, b, and α are user defined and the reader is referred to [10] for more details on how to set them.

2.3 Log-Normal Distributions Mixture Fitting

In this paper, log-normal distributions mixture is used to improve the goodness of fit for cellular channel occupancy compared to a single log-normal distribution. In [11] a mixture of normal distribution is used to fit a specific data. To deal with the monotonicity behaviour of the measured cellular channel occupancy, log-normal mixture can be used instead of normal mixture. The pdf number i and the collection of distribution distinct parameters, Ω_{lgn}, in a log-normal mixture are shown in (9a) and (9b) respectively.

$$f(x|\mathbf{\Theta}_i) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\ln(x) - \mu)^2}{2\sigma^2}}, \tag{9a}$$

$$\mathbf{\Omega}_{lgn} = \begin{pmatrix} p_1 & \mu_1 & \sigma_1 \\ \vdots & \vdots & \vdots \\ p_k & \mu_k & \sigma_k \end{pmatrix}. \tag{9b}$$

$\mathbf{\Omega}_{lgn}$ can be found using Newton Raphson optimization method by solving the equation

$$L(x|\mathbf{\Omega}) = 0. \tag{10}$$

Starting from an initial guess of $\mathbf{\Omega}_{lgn}^{(1)}$, then $\mathbf{\Omega}_{lgn}^{(i+1)}$ is updated as

$$\mathbf{\Omega}_{lgn}^{(i+1)} = \mathbf{\Omega}_{lgn}^{(i)} - \mathbf{H}^{-1}\left(\mathbf{\Omega}_{lgh}^i\right) L\left(x|\mathbf{\Omega}_{lgn}^{(i)}\right), \tag{11}$$

where $\mathbf{H}(\cdot)$ denotes the Hessian matrix. As the Hessian matrix is needed to be updated every iteration, then the stopping criterion is the convergence of $\mathbf{H}$.

2.4 Optimizing the Number of Distributions

To optimize the value of k, Akaike information criterion [12] is used. AIC is a statistical model identification used to optimize the model order [12]. AIC is calculated considering the log-likelihood penalized by the number of independent model parameters. AIC is obtained using

$$AIC(x; \mathbf{\Omega}, N) = \underbrace{-2L(x|\mathbf{\Omega})}_{\text{log likelihood}} + \underbrace{2N}_{\text{Parameters penalty}}, \tag{12}$$

where N is the model order defined as the number of independent model parameters. The optimal model order is found by minimizing the value of AIC in (12). For the exponential mixture distribution, each pair i where $1 \leq i \leq (k-1)$ represents a single independent parameter while the last pair (λ_k, p_k) is fully dependant on the other pairs. Hence, the exponential distributions mixture has $(N = k-1)$ independent parameters. Accordingly, the optimal model order for exponential mixture distribution, k_{AIC}^{exp}, is found as

$$k_{AIC}^{exp} = \underset{k}{\operatorname{argmin}} \left(-2L(x|\mathbf{\Omega}_{exp}) + 2(k-1) \right). \tag{13}$$

For the log-normal distribution mixture, with Newton Raphson method, there are $N = 3k$ independent parameters as all the components of $\mathbf{\Omega}_{lgn}$ are independent. Therefore, the optimal model order for log-normal distributions mixture, k_{AIC}^{lgn} is determined as

$$k_{AIC}^{lgn} = \underset{k}{\operatorname{argmin}} \left(-2L(x|\mathbf{\Omega}_{lgn}) + 6k \right). \tag{14}$$

3 Measurements

3.1 Measurements Setup

The empirical downlink LTE traffic is obtained through a measurement campaign performed in an indoor location in Kista, Stockholm, Sweden. The measurements location has a GPS coordinates of $59°24'19.13''N$, $17°56''56.12'E$. The measurement area is densely occupied by offices with a shopping mall and residential buildings in the surroundings. A google map of the measurement location is shown in Fig. 2.

For robust measurements, a real time spectrum analyser (RTSA) is used to collect the data. The data is fed to the RTSA through a wideband tunable antenna. Figure 3 exhibits the measurements setup. Since different channels experience different loads at different times, the measurements are treated in time spans of 2 h. Hereafter, the findings for an LTE downlink traffic channel will be discussed as a representative case. The results for the other channels and systems are similar with different parameters. The presented results are for the measurements carried out for a 1.4 MHz channel lies between 2650.6 and 2652.0 MHz during the period: Wednesday, 2013/10/02 09:00 am to 11:00 am.

3.2 Fitting Results

Before diving into the fitting results, it is important to note that the LTE load on the measurements area changes with time, This changes are depicted by the obtained values of the duty cycle through a week of measurements shown in Fig. 4. Even-though, different loads are experienced at different times, yet the fitting procedure is the same and the findings are similar with different values. Hereafter, the results for the period Wednesday, 2013/10/02 09:00 am to 11:00 am are shown as an example of the results.

Figures 5 and 6 show the empirical distribution and the fitted exponential and log-normal mixtures respectively. Both Figs. 5 and 6 illustrate how the fitted mixtures of
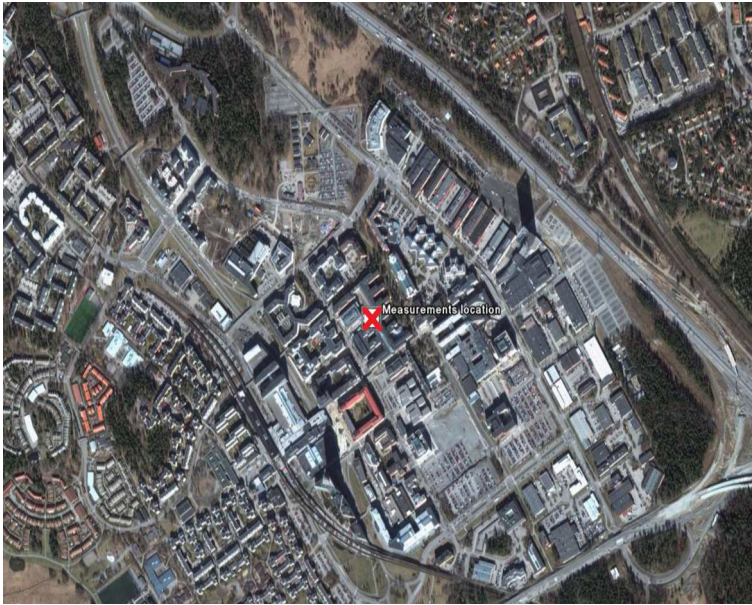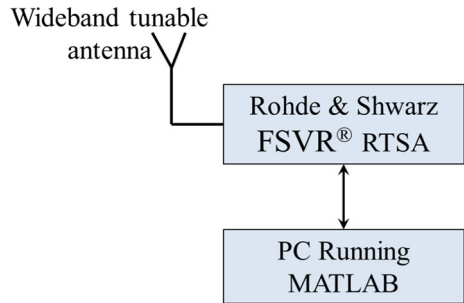
Fig. 2 The measurements location

Fig. 3 The measurements setup

Wideband tunable
antenna

Rohde & Shwarz
FSVR® RTSA

PC Running
MATLAB

exponential or log-normal distributions approach towards the empirical distribution with the change of k. A quantitative evaluation is obtained by means of the log likelihood estimation which is provided in Fig. 7.

As it is shown in Fig. 5, the lower values of k make the exponential mixture to fit the tail with poor fitting for the lower values of x. In contrast, increasing k improves fitting the lower region of x. This is explained as follows; as the first pair (λ_1, p_1) always characterizes the tail beyond c_1, then there is always a guarantee that all the values greater than c_1 are well fitted, depending on the obtained values of (λ_1, p_1) and the value of k, rest pairs (λ_i, p_i) are obtained and the last pair (λ_k, p_k) is fully dependant on the previous obtained pairs. Therefore, when k increases the part that is characterized by (λ_k, p_k) decreases. However for very large values of k a point where the property expressed in (5) is not held which makes the recursive fitting procedure for the remaining pairs inapplicable any longer. Therefore, there is a crossover point when the log-likelihood estimation starts to
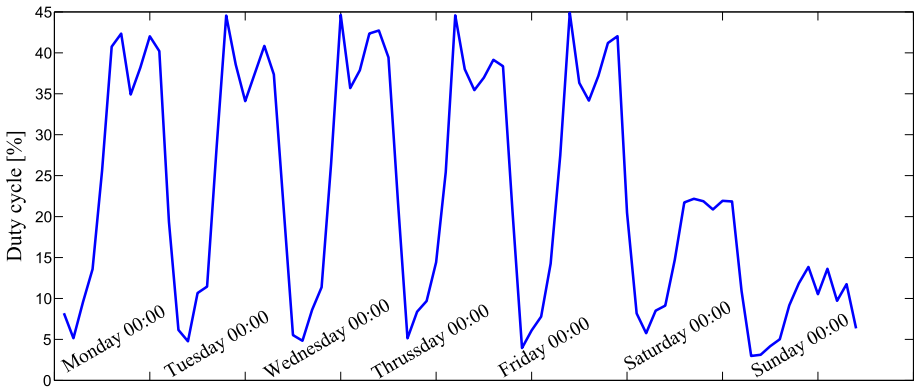
Fig. 4 Duty cycle through a week of measurements, from Monday, 2013/09/30 to Sunday, 2013/10/06
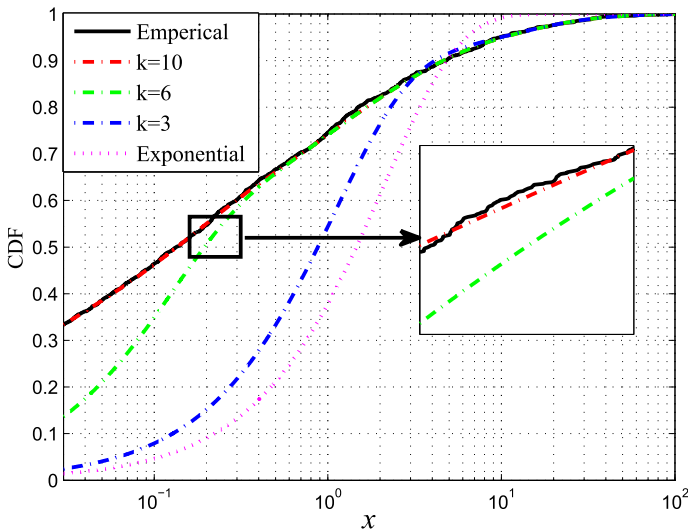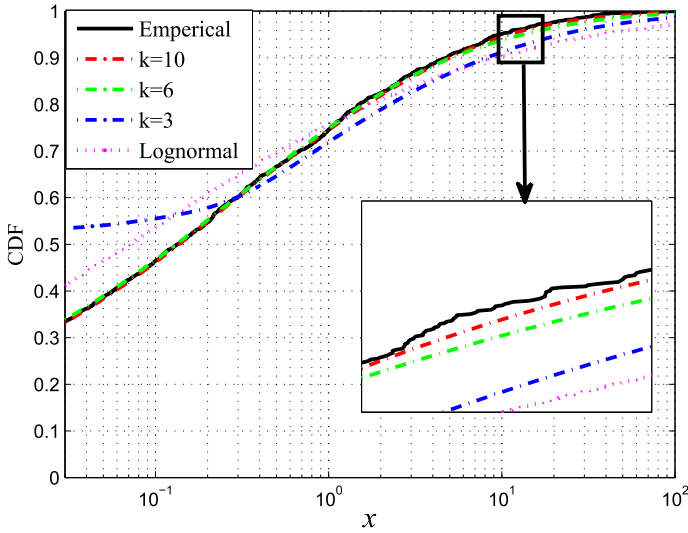

**Fig. 5** The empirical and fitted CDF for exponential distributions mixture with different values of k

degrade with the increase of k as shown in Fig. 7. For the lognormal mixture the higher the k, the better the fitting as the log-likelihood curve exhibited in Fig. 7.

Figure 7 depicts the obtained AIC for both exponential and log-normal distributions mixtures when k changes. According to the figure, for the exponential distributions mixture the optimal model order is 7 while for log-normal distribution mixture the optimal model order is 4. The difference in the model order between the two mixtures is explained by the influence of the parameters penalty function. As shown in (13) and (14) the log-normal mixture AIC is penalized more than the exponential mixture AIC. Moreover, for the same reason in the case of exponential distributions mixture, the AIC curves follow the log-likelihood curve. On the other side, for the log-normal distributions mixture, the AIC and

Fig. 6 The empirical and fitted CDF for log-normal distributions mixture with different values of k
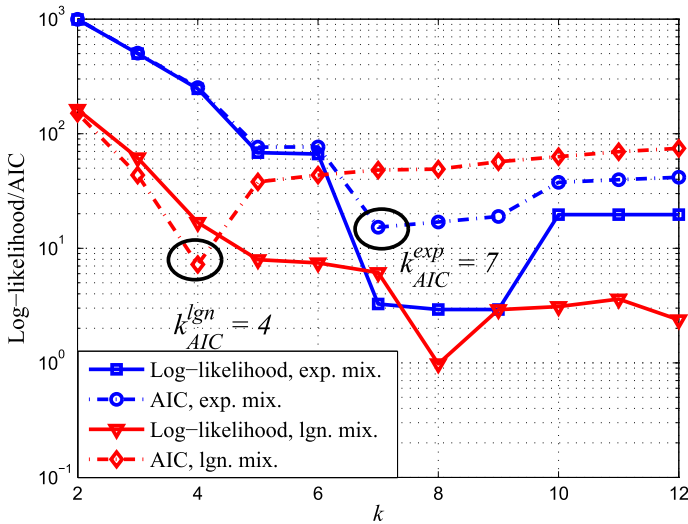
Fig. 7 Distribution log-likelihood and the AIC for exponential and log-normal distribution mixtures

the log-likelihood curves have different tendencies as the parameters penalty function impacts more in the AIC values.

The obtained distinct mixture parameters matrices for the optimal exponential and log-normal mixtures, Ω_{exp} and Ω_{lgn} are shown respectively below.

$$\boldsymbol{\Omega}_{exp} = \begin{pmatrix} 0.04 & 0.04 \\ 0.08 & 0.11 \\ 0.20 & 0.47 \\ 0.16 & 1.75 \\ 0.23 & 8.71 \\ 0.12 & 54.24 \\ 0.17 & 103.85 \end{pmatrix}.$$

$$\boldsymbol{\Omega}_{lgn} = \begin{pmatrix} 0.33 & 0.67 & 1.53 \\ 0.21 & -6.38 & 2.59 \\ 0.40 & -2.32 & 1.64 \\ 0.06 & -9.84 & 5.05 \end{pmatrix}.$$

As explained by (4b), the first column of $\boldsymbol{\Omega}_{exp}$ is the probabilities of the different exponential distributions with their corresponding values of λ in the second column. Similarly, as in (9b) the probabilities of the log-normal distributions are placed in the first column of $\boldsymbol{\Omega}_{log}$ with the corresponding values of the means and the standard deviations in the second and third columns respectively.

4 Conclusions

An empirical statistical model for the downlink LTE channel occupancy is introduced in this paper. The introduced model is based on using a linear mixture of exponential or log-normal distributions. The exponential and log-normal distributions mixture can better characterize the downlink LTE channels occupancy compared to the single exponential and log-normal distributions. Akaike information criterion is used to optimize the number of the exponential or log-normal distributions composing the mixture. Log-normal mixture Akaike information criterion is affected more by the model order compared to the exponential mixture. The model is a general statistical model and can be used for other cellular systems.

References

1. Astely, D., Dahlman, E., Furuskar, A., Jading, Y., Lindstrom, M., & Parkvall, S. (2009). LTE: The evolution of mobile broadband. *IEEE Communications Magazine*, *47*(4), 44–51.
2. Ghosh, A., Ratasuk, R., Mondal, B., Mangalvedhe, N., & Thomas, T. (2010). LTE-advanced: Next-generation wireless broadband technology [invited paper]. *IEEE Wireless Communications*, *17*(3), 10–22.
3. Yavuz, E., & Leung, V. (June, 2007). Modeling channel occupancy times for voice traffic in cellular networks. In *IEEE international conference on communications (ICC)* (pp. 332–337).
4. Hamid, M., Mohammed, A., & Yang, Z. (2010). On spectrum sharing and dynamic spectrum allocation: MAC layer spectrum sensing in cognitive radio networks. In Int. *Conference on Communications and Mobile Computing (CMC)*, *2*, 183–187.

5. Hong, D., & Stephen, S Rappaport. (1986). Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures. *IEEE Transactions on Vehicular Technology*, 35(3), 77–92.
6. Guerin, R. (1987). Channel occupancy time distribution in a cellular radio system. *IEEE Transactions on Vehicular Technology*, 36(3), 89–99.
7. Wellens, M., Riihijrvi, J., & Mhnen, P. (2009). Empirical time and frequency domain models of spectrum use. *Physical Communication*, 2(12), 10–32. Cognitive radio networks: Algorithms and system design. http://www.sciencedirect.com/science/article/pii/S1874490709000299.
8. Yavuz, E., & Leung, V. C. M. (2006). Computationally efficient method to evaluate the performance of guard-channel-based call admission control in cellular networks. *IEEE Transactions on Vehicular Technology*, 55(4), 1412–1424.
9. Lopez-Benitez, M., & Casadevall, F. (2011). Empirical time-dimension model of spectrum use based on a discrete-time markov chain with deterministic and stochastic duty cycle models. *IEEE Transactions on Vehicular Technology*, 60(6), 2519–2533.
10. Feldmann, A., & Whitt, W. (1997). Fitting mixtures of exponentials to long-tail distributions to analyze network performance models. *Performance Evaluation*, 31, 245–279.
11. Du, J. (July, 2002). Combined algorithms for fitting finite mixture distributions. Master's thesis, McMaster University.
12. Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723.

Mohamed Hamid received the B.Sc. degree in electrical engineering from University of Khartoum, Sudan in 2005, the M.Sc. degree from Blekinge Institute of Technology (BTH), Karlskrona, Sweden in 2009 and the Ph.D. degree from KTH The Royal Institute of Technology, Stockholm, Sweden, in 2015. In 2016 he joined the University of Agder, WISENET lab where he is currently working as a Postdoctoral researcher. His research interests are cognitive radio systems, spectrum sensing, geo-location database based spectrum opportunities spectrum cartography, device-to-device communications and traffic modeling.

Niclas Björsell received his B.Sc. in Electrical Engineering and his Lic. Ph. in Automatic control from Uppsala University, Sweden in 1994 and 1998, respectively; he received his Ph.D. in Telecommunication from the Royal Institute of Technology, Stockholm, Sweden, in 2007. In 2012 he was appointed Docent in Telecommunications at the Royal Institute of Technology, Stockholm. For more than 25 years he has hold positions in the academy as well as in industry. He has 20 years of experience from research and development projects; both national and international and mainly in collaborations between industry and the academy. Between 2006 and 2009 he served as the head of Division of Electronics at the Department of Technology and Built Environment, University of Gävle, Sweden. He is currently Associate Professor at the University of Gävle and Guest Professor at the Vrije Universiteit Brussel, Belgium. He is program director for the study program in Automation Engineering. He has published more than 60 papers in international peer-review journals and conferences, and his research interests include radio frequency measurement technology, analog-to-digital conversion, non-linear systems, cognitive radio and automation. Future research may well be closely related to industrial research and development. Dr. Björsell is Associate Editor of IEEE Transactions on Instrumentation and Measurement, reviewer for several journals and conferences, and active member of the Waveform Generation, Measurement and Analysis Committee (IEEE, Instrumentation and Measurement).

Slimane Ben Slimane received his B.Sc. degree in electrical engineering from the University of Quebec at Trois-Rivieres, Canada, in 1985, and his M.Sc. and the Ph.D. degrees, both from Concordia University, Montreal, Canada, in 1988 and 1993, respectively. During the period 1993–1995, he worked as a research associate and part-time instructor at Concordia University. He is currently an associate professor in the Radio Communication Systems group, department of Communication Systems (COS), the Royal Institute of Technology (KTH), Stockholm, Sweden. His research interest is in the area of wireless communications, with special emphasis on digital communication techniques for fading channels, error control coding, cooperative communications, spread spectrum communications, multicarrier transmission techniques, and cognitive radio.