Distributed Energy-Saving Mechanism for Self-Organizing Femto LTE Networks

Raymond Kwan
Ubiquisys Ltd. Swindon, United Kingdom
Email: raymond.kwan@ubiquisys.com

Abstract—**In this paper, a distributed power adjustment algorithm for downlink Long Term Evoluation (LTE) mobile networks is proposed, in which the Quality of Service (QoS) requirement is taken into account. By tailoring the downlink transmit power to match the aggregate service requirement, it is possible to eliminate unnecessary power, and thereby reducing inter-cell interference to the neighbouring cells. The reduction of inter-cell interference further lowers the required power to meet a certain QoS target. Simulation results have shown that such an approach can significantly reduce the energy consumption with negligible impact on the system performance.**

I. Introduction

The fourth generation (4G) cellular systems such as the Long-Term Evolution (LTE) are currently being developed in order to improve both systems performance and user data rate relative to the third generation systems. Although such systems are designed to improve systems performance and user data rate, strong emphasis is given to enhance performance of users at the cell edge. A well-known method to mitigate inter-cell interference is via what is known as fractional frequency reuse (FFR), in which mobile users in the cell-center are allocated the same frequency for all cells, whereas users at the cell-edge are allocated with a subset of frequencies that are different from those of the immediate neighbour. As a result, the inter-cell interference at the cell-edge users can be significantly reduced. Soft frequency reuse (SFR) represents a variant of FFR, in which the different power levels are associated with the edge and center sub-bands. Typically, the sub-band is allocated with a higher power in order for the user equipments (UEs) to better compensate for the larger pathloss, and thereby providing addition flexibilities. More descriptions about these methods as well as their variants can be found, for example, in [1], [2], [3] and the references therein, and will not be elaborated here.

While FFR, SFR, and many of their variants are useful techniques for interference mitigation, they suffer from important draw-backs: the subsets of frequency used for the cell-edge mobile users among neighbours need to be carefully planned, and are typically done statically during the network planning stage. Such methods are especially not suitable for femtocells, where base stations are deployed in an ad hoc manner. Also, these methods do not take into account the dynamic user traffic distributions, and thereby reducing the efficiency of the spectrum utilization. It is possible to make the allocation of power and frequency resources dynamically by allocating frequency, power, modulation and coding schemes (MCS) jointly for each user in a cell in a centralized fashion [4]. In [5], [6], gradient-based algorithms have been presented in which the frequency reuse patterns are dynamically adapted to the traffic distribution. As this approach is self-organizing among cells in a distributive manner, the tedious process of frequency planning is not required. In addition, it allows the power to be adjusted dynamically in frequency, and thereby providing an extra degree of flexibility. While this approach provides a powerful framework in dynamic interference mitigation, it does not aim to provide details regarding how Quality of Service (QoS) can be taken into account. As a result, the power allocation may not necessarily tailor to what the services actually require. Also, this approach is designed for generic Orthogonal Frequency Division Multiple Access (OFDMA) systems, and assumes a certain quantity to be communicated among base stations which is not readily supported in the LTE specification. In this paper, a modified algorithm is proposed in which QoS is taken into account during the power adjustment. Furthermore, the new exchanged quantity among base stations[1] is proposed to be more compatible with the X2 interface specified in LTE. In section II, the concept of gradient-based distributed power control is described, which is followed by our proposed happiness-based formulation in section III, and our full X2 compatible solution in section IV. This is then followed by the numerical results in section V and a conclusion.

II. Gradient-Based Distributed Power Control

Assume that we have K cells indiced with $k \in \mathcal{K} = \{1, 2, \ldots, K\}$, and J sub-bands $j \in \mathcal{J} = \{1, 2, \ldots, J\}$ in the system. Furthermore, assume that each sub-band consists of a fixed number of sub-carriers. Also, the time is assumed to be slotted, and that transmissions within each cell are synchronized, so that intra-cell interference is not present. Two generic quantities are particularly relevant to inter-cell interference coordination scheme for LTE-based systems. The first one is the concept of utility, which generally quantifies the level of satisfaction of the entity involved. Let U be a global utility function of the system, which is given by

$$U = \sum_k U_k, \qquad (1)$$

where U_k is the utility function of cell k, and is given by the sum of the user utility $U_{k,i}$ within the cell, i.e. $U_k = \sum_i U_{k,i}$.

[1]In LTE, a base station is known as the enhanced Node B, or eNB.

978-1-4673-1881-5/12/$31.00 ©2012 IEEE

The basic objective is to find a way (or ways) to maximize the global utility function U given a certain set of contraints.

The second concept is transmit power. Here, in the context of OFDMA systems such as LTE, the transmit power has the flexibility to be frequency dependent. Let $P_{k,j}$ be the power allocated in sub-band j of cell k, and the maximum power cell k can have is $\tilde{P}_k$, i.e. $P_k = \sum_j P_{k,j} \leq \tilde{P}_k$. The whole problem of inter-cell interference coordination reduces to how $P_{k,j}, \forall j$ is allocated for each sub-band in order to maximize U.

In [5], a gradient-based method has been proposed, in which the global utility is improved sub-optimally in a distributive fashion. The main idea of the proposed method is as follows:

Let $D_j(m,k)$ be the rate of change of utility U_k for cell k, with respect to the transmit power cell m which has been allocated for sub-band j, i.e.

$$D_j(m,k) = \frac{\partial U_k}{\partial P_{m,j}}, \ \forall j. \tag{2}$$

An increase in $P_{m,j}$ may potentially have a negative impact on U_k when $k \neq m$ (i.e. cell m is a neighbour cell), as such an increase would give rise to additional interference at sub-band j coming from cell m, and vice versa. On the other hand, when $k = m$, an increase of power at sub-band j would enhance the signal quality at this sub-band, and would have a positive impact on its own utility.

Note that $D_j(m,k)$ is not very useful if it is kept in the own cell. However, when it is exchanged among neighbour cells, it allows the neighbour cells to know the level of impact it incurs to the other cells when a certain power level is allocated at each sub-band. By receiving $D_j(m,k)$ from the neighbour cells, cell k would then aggregate them for each sub-band , i.e.[2]

$$D_j(k) = \sum_m D_j(k,m) \ \forall j, \tag{3}$$

including the case of $k = m$. In other words, $D_j(k)$ corresponds to the aggregate sensitivity of utility function to all cells due to the perturbation of its own transmit power at sub-band j. When $D_j(k) < 0$, a positive power increment would incur a negative impact on the aggregate satisfaction among all cells, and vice versa. The general idea proposed in [5] is for cell k to increase the power by selecting a sub-band associated the largest positive value of $D_j(k)$, and vice versa. For example, the sub-band specific sensitivity can be written as

$$D_j(m,k) = \sum_i \frac{\partial U\left(\overline{R}_{k,i}\right)}{\partial P_{m,j}} \tag{4}$$

$$= \sum_i \frac{\partial U\left(\overline{R}_{k,i}\right)}{\partial \overline{R}_{k,i}} \frac{\partial \overline{R}_{k,i}}{\partial P_{m,j}}, \tag{5}$$

where $\overline{R}_{k,i}$ is the average bit rate achieved by user i in cell k, which is a function of the average bit rate $\overline{R}_{k,i}^{(j)}$ achieved by user i at sub-band $j, \forall j$. The term $\overline{R}_{k,i}^{(j)}$, in turn, is a function of $P_{m,j}$. Typically, to compute $D_j(m,k)$, the analytical modelling of the relationship between $\overline{R}_{k,i}^{(j)}$ and $P_{m,j}$ is required. In [6], a similar method has been proposed, except that the utility function is computed based on the statistical average of the bit rate assuming a certain known fading statistics.

[2]Note that the switch of the indices m and k is to represent the fact that cell k is now the neighbour cell of each of the neighbour cell m.

III. Happiness-based Distributed Power Management

It is important to note that a utility does not necessarily provide a means to incorporate QoS into the adjustment mechanism. On the other hand, an unnecessarily high power generates an unnecessary level of interference, which would then have a "knock-on" effect to the neighbouring cells. In order to maintain a good level of satisfaction, the neighbours would require a higher power, and thereby boosting the overall background interference. The implication of the above process is important, as the idea of removing unnecessary power provides a "feedback" mechanism which eventually helps to further reduce the power requirement for a fixed QoS due to the lowering of the overall interference. It is such lowering of power requirement which translates to energy saving for the network.

One way to take QoS into account is to modify the utility function. However, such an approach can potential make the utility function more complex, and thereby complicating the sensitivity calculation. In this paper, we propose to quantify whether a user's expectation is met by a quantity known as the "Happiness Factor", $H_{k,i}$, and is given by

$$H_{k,i} = \frac{\overline{R}_{k,i}}{\tilde{R}_{k,i}}. \tag{6}$$

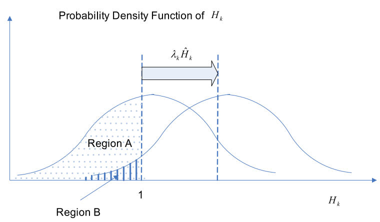
The term $\tilde{R}_{k,i}$ is the bit rate requirement which can be directly proportional to the guarantee bit rate (GBR) [7] or some function of GBR. When $H_{k,i} > 1$, the user is experiencing a bit rate that exceeds expectation, and the opposite is true when $H_{k,i} < 1$. Let $H_k^{(n)}$ be the weighted n-th moment of happiness of cell k, i.e.

$$H_k^{(n)} = \frac{1}{N_k} \sum_{i=1}^{N_k} w_{k,i} H_{k,i}^n, \tag{7}$$

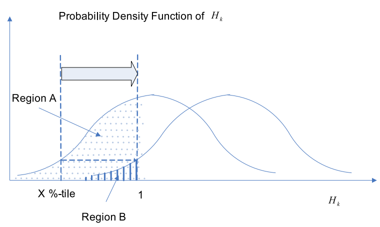
where N_k is the number of users in cell k, and $w_{k,i}$ is a cell-specific weight for user i in cell k. This weight can be used to bias the emphasis among users within the cell, and follows the constraint $\sum_{i=1}^{N_k} w_{k,i} = N_k$. As a special case when $w_{k,1} = w_{k,2} = \ldots, w_{k,N_k} = 1$, $\overline{H}_k = H_k^{(1)}$ reduces to a simple arithmetic mean. Note that $\overline{H}_k = 1$ implies that the average happiness for cell k meets the expectation. However, it also implies that some users are below expectation, while some are above. While the average is useful to quantify performance in general, a more refined approach is to provide a conservative margin to the average value such that

$$H_k = \overline{H}_k - \lambda_k \hat{H}_k, \tag{8}$$

where H_k is known as the *true* happiness, and λ_k is a scaling factor which controls the level of *conservativeness*. The quantity $\hat{H}_k$ is the weighted standard deviation of happiness within cell k, i.e. $\hat{H}_k = \sqrt{H_k^{(2)} - \left(H_k^{(1)}\right)^2}$. Fig. 1a shows the effect of offsetting the happiness factor in order to increase conservativeness for power adjustment. By offsetting the happiness factor by $\lambda_k \hat{H}_k$, the probability that the true happiness is below unity reduces from the area of region A to that of region B. A more general way to increase conservativeness is to define H_k as an X-percentile of $H_{k,i}, \forall i$. Under this definition, only X-percents of the happiness would fall below unity.



(a) Standard deviation based offset



(b) CDF based offset

Fig. 1. Happiness offset

IV. Standard X2-based Solution

A. Re-definition of sensitivity function, $D_j(k,m)$

Note that the communication of $D_j(k,m)$ among neighbour base stations requires proprietary interface, as it is not supported by the standard X2 interface [7]. While proprietary interface is a viable solution, it is more desirable to support standardized interface, especially in the case of inter-vendor compatibility. In addition, solutions based on standardized interface can be tested using standard test equipments, and thereby reducing the cost of testing and deployment.

According to [7], a Relative Narrowband Tx Power (RNTP) information element (IE) is included in the *Load Information* X2 message. For each resource block, the RNTP IE informs the neighbouring cells whether the sending cell power at such resource block is above (1), or below (0) a certain threshold (RNTP threshold). In order to make use of the X2-interface, $D_j(k,m)$ needs to be formulated in terms of what are available in X2. Let

$$D_j(k,m) = \begin{cases} -\sum_i \frac{G_i^{(m)}}{G_i^{(k)}} \rho_{j,m} & m \neq k \\ 0 & m = k \end{cases}, \qquad (9)$$

where $G_i^{(m)}$ is the path gain between the mobile i (served by cell k) and neighbour m, and $\rho_{j,m}$ is the RNTP for sub-band j. The quantity $G_i^{(m)}$ can be obtained at the mobile by measuring the Reference Signal Received Power (RSRP) [8] and the corresponding transmit power from neighbour m via the neighbour's broadcast channel. More precisely, let

$$g_i^{(m)}(t) = \frac{\hat{P}_m(t)}{P_m^{ref}}, \qquad (10)$$

where $\hat{P}_m(t)$ is the RSRP from cell m sampled at time t, and P_m^{ref} is the reference signal power from the neighbour's broadcast channel. Note that the RSRP measurements obtained at the mobile or at the network monitor mode (NMM) [9] of the base station can fluctuate due to channel fading, shadowing, etc. An average value is typically more useful in reflecting the long-term behaviour by setting

$$G_i^{(m)}(t) = \begin{cases} (1-\alpha) G_i^{(m)}(t-1) + \alpha g_I^{(m)}(t) & \text{Exp avg} \\ \frac{1}{N}\sum_{i=1}^{N} g_i^{(m)}(t-i) & \text{Blk avg} \end{cases} \qquad (11)$$

where N denotes the number of samples for blocking averaging, and α corresponds to the filter coefficient for exponential averaging. Alternatively, $G_i^{(m)}(t)$ can be an x-percentile of the samples $\left\{ g_i^{(m)}(t'),\ t' = t-1, t-2, \ldots, t-N \right\}$.

The quantity $\rho_{j,m}$ can be a reasonable aggregate of the RNTP values for each resource blocks within a sub-band[3]. For example,

$$\rho_{j,m} = \frac{1}{Q} \sum_{q=1}^{Q} \rho_{j,m}^{(q)}, \qquad (12)$$

where Q is the number of resource blocks per sub-band, and $\rho_{j,m}^{(q)}$ is the RNTP for resource block q in sub-band j from neighbour m.

B. Overall Algorithm

The overall flow diagram is shown in Fig. 2. It is important to note that $D_j(k,m)$ no longer explicitly represents the sensitivity of the utility of cell k at sub-band j with respect to the power from neighbour m. Rather, it represents the aggregate impact among mobiles in cell k due to cell m if cell k were to transmit at sub-band j. The more negative the quantity is, the more impact it has, and, therefore, cell k would further avoid its transmission at the respective sub-band.

- *S*tep 1: A sub-band j_* is selected such that $D_{j_*}(k)$ is the smallest among all j's with $P_{k,j} > 0$. Thus, this step selects the sub-band for which an increase in power would have the least beneficial effect on the cell performance.
- *S*tep 2: A sub-band j^* is selected so that a power increase would have the most beneficial effect on the cell performance. As shown in (9), in the best sub-band $D_j(k)$ would have a value of zero, and it is possible to have multiple sub-bands which would satisfy this

[3] This is due to the fact that the RNTP values provided by the X2 *Load Information* message are defined in terms of resource blocks.

criterion. In order to avoid the possibility that the power is increased in only one sub-band, the sub-band in which power might be increased later is chosen randomly and uniformly from a set of sub-band having $D_j(k) = 0$. In this way, the utilization of potentially useful sub-bands can be increased.

- *S*tep 3: If $D_{j_*} < 0$, sub-band j_* is expected to have a negative impact to the performance, and the power associated to sub-band j_* is reduced in *S*tep 4.
- *S*tep 5: If the total transmit power for cell k, P_k, is less than the maximum limit $\tilde{P}_k$ and that while $D_{j^*}(k) = 0$, it is an indication that there is room for potential power increase. Whether a power increase is justified depends on whether the cell is happy (*S*tep 6).
- *S*teps 6,8,9: If the cell is already happy, it is beneficial to reduce the power for sub-band j^* in order to 1) reduce interference to other cells, and 2) reduce energy consumption for the serving cell (*S*tep 9). Of course, if the cell is not yet happy, power to sub-band j^* is increased (*S*tep 8).
- *S*tep 7: When the condition in *S*tep 5 is not satisfied, the process checks whether the total transmit power has already reached the power limit and that it is beneficial to increase the power for sub-band j^*, i.e. $D_{j^*} = 0$. If the conditons are not satisfied, the algorithm stops and waits until the next execution begins. However, if the conditions are met, further power adjustment may still be worthwhile.
- *S*tep 10: If conditions in *S*tep 7 are satisfied, it may be beneficial for sub-band j^* to increase power. However, as the total power has already reached its limit, the only way to increase power for sub-band j^* is to make room for it by taking power from the worst sub-band j_*. If $D_{j_*}(k) < 0$, sub-band j_* would then be a good candidate for power decrease (*S*tep 12). If $D_{j_*}(k) = 0$, a new sub-band j_* is randomly selected for power decrease (*S*tep 12). The reason for such randomization is to diversify the sub-band for power decrease.
- *S*teps 13-15: After the power is reduced at sub-band j_*, the algorithm checks whether the cell is happy. If the cell is happy, power is further reduced at sub-band j^* (*S*tep 15). Otherwise, the power at sub-band j^* is increased (*S*tep 14).

V. Numerical Results

Simulations have been done to illustrate the benefit of the proposed algorithm. Although the proposed method works for both macro and small cell environments, the latter is chosen for the sake of illustration. The scenario consists of a two floor building, and each floor is a 2×2 grid of 4 rooms. Each room has a dimension of $8m \times 8m$, and a single Home eNB (HeNB) is located in the centre of the room. There are 4 connected UEs per room. The heights of each room, UE, and HeNB are 3.0m, 1.0m, and 1.5m respectively, with internal wall loss of 5 dB. The transmit power for the each HeNB and UE is 10mW, and their respective noise figures are 8 dB and 9 dB.

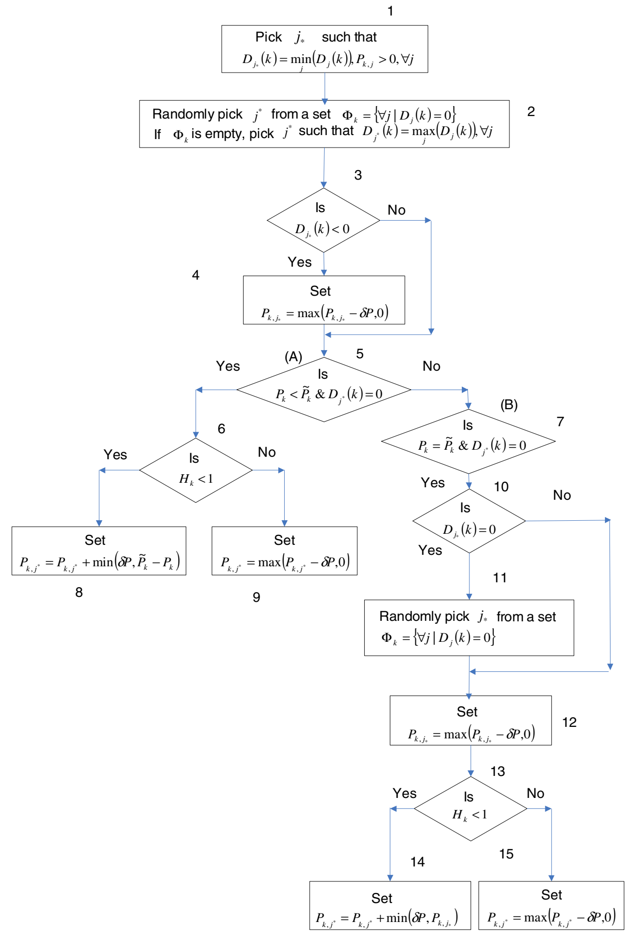


Fig. 2. X2-based Distributed Power Management Algorithm.

A full buffer traffic model and a proportional fair (PF) multi-user downlink scheduler are assumed. The propagation model is assumed to be Okumura-Hata within a large urban city. A 5MHz bandwidth is assumed. The UEs are assumed to be static, and randomly located within the room, and the Extended Padestrian A (EPA) channel model is used. The simulator is an ns-3 based platform with LTE-added functionalities[4] [10]. Three cases have been simulated, and the details are shown in Table I[5]. Let $\overline{U}_g = \sum_{k=1}^{K} \sum_{i=1}^{N_k} U_k(\overline{R}_{k,i})/K$ be the average

TABLE I
SIMULATION SCENARIOS

Case	λ	Bit rate requirement $\tilde{R}_{k,i}$ (in Bytes/TTI)
1	0.01	70
2	1.20	70
3	0.01	96

[4]These LTE-specific functionalities are based on a join project between Ubiquisys Ltd. and Centre Tecnologic de Telecommunicacions de Catalunya (CTTC).

[5]Transmission Time Interval of 1ms

global utility function, assuming $U_k(x) = \log_2(x)$. Fig. 3 shows the average user bit rate $\overline{R}_{k,i}, \forall i$ within the system for case 1. It can seen that the average performance converges well at the target. Fig. 4 shows $\overline{U}_g$ as a function of time for the three cases. It can be seen that the results for case 1 and 3 converge to the target. On the other hand, the result for case 2 stablizes slightly above the target due to the higher conservative factor as expected. This increased conservativeness improves the overall cell performance at the expense of a higher power consumption. Finally, Fig. 5 shows the average downlink transmit power per cell as a function of time for all three cases. It can be seen that significant power saving can be achieved by simply transmitting only with what is needed. In this way, less inter-cell interference is generated. The reduction in inter-cell interference translates to the reduced need to overcome interference from the neighbour, and thereby allowing further power reduction to take place. Eventually, the power converges to a level that is quite significantly lower than the original level, and yet the performance does not suffer (Fig. 4).

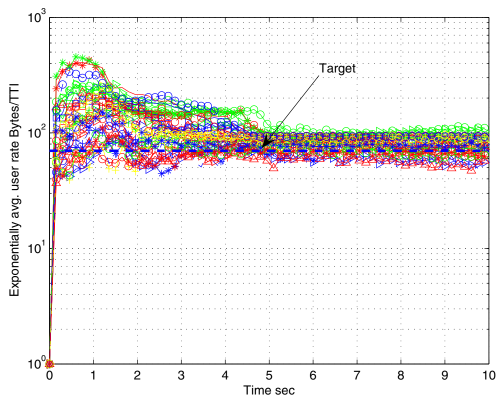


Fig. 3. Average user bit rate as a function of time.

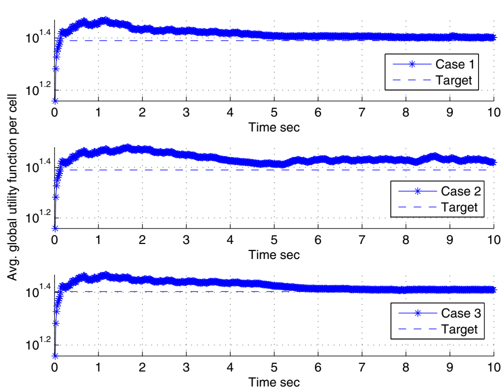


Fig. 4. Average Utility per Cell as a function of time.

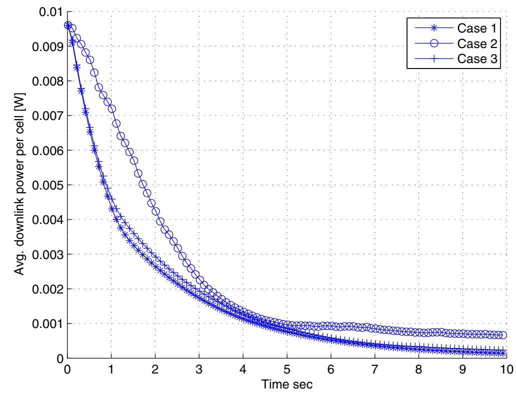


Fig. 5. Average downlink power per cell as a function of time.

VI. Conclusion

In this paper, a distributed power allocation algorithm for LTE has been proposed. It takes into account the cell-wise satisfaction level, and is compatible with the X2-interface of LTE. Simulation results have shown that, by allocating just the right amount of power to meet the target requirement, and thereby generating less inter-cell interference to the neighbours, it is possible to achieve significant energy saving.

Acknowledgment

The author would like to thank Nicola Baldo and his team in the Centre Tecnologic de Telecommunicacions de Catalunya (CTTC) for the support of the ns-3 simulator.

References

[1] R. Kwan and C. Leung, "A Survey of Scheduling and Interference Mitigation in LTE," *Journal of Electrical and Computer Engineering*, vol. 2010, no. 273486, 2010.

[2] A. Mills, D. Lister, and M. D. Vos, "Understanding Static Inter-Cell Interference Coordination Mechanisms in LTE," *Journal of Communications*, vol. 6, no. 4, pp. 312 – 318, July 2011.

[3] S. Hämäläinen, H. Sanneck, and C. Sartori, *LTE Self-organizing Networks (SON): Network Management Automation for Operational Efficiency*.

[4] D. López-Perez and G. de la Roche et al., "Interference Avoidance and Dynamic Frequency Planning for WiMAX Femtocells Networks," in *Proc. of IEEE International Conference on Communications (ICC)*, Guangzhou, 19-21 Nov. 2008, pp. 1579 – 1584.

[5] A. L. Stolyar and H. Viswanathan, "Self-organizing Dynamic Fractional Frequency Reuse for Best-Effort Traffic Through Distributed Inter-cell Coordination," in *Proc. of IEEE Infocomm.*, Rio de Janeiro, 19 - 25 April 2009, pp. 1287 – 1295.

[6] R. Combes, Z. Altman, and E. Altman, "Self-Organizing Fractional Power Control for Interference Coordination in OFDMA Networks," in *Proc. of IEEE International Communication Conference (ICC)*, Kyoto, Japan, 5-9 June 2011, pp. 1–5.

[7] TS36.423, "X2 application protocol (X2AP)," 3GPP, Tech. Rep. V8.3.0, 2008.

[8] "Physical Layer Measurement (FDD)," 3rd Generation Partnership Project, Technical Specification 3GPP TS36.214, 2010.

[9] "LTE Network Monitor Mode Specification," Small Cell Forum, Technical Document FF Tech 003 v1.01, 2010.

[10] N. Baldo, M. Requena, J. N. Guerrero, and M. Miozzo, "A new model for the simulation of the LTE-EPC data plane," in *Proc. of Workshop on ns-3 (WNS3), in conjunction with ICST SIMUTOOLS*, Sirmione, Italy, March 2012.