Analysis of Limitations of Mobility Load Balancing in a Live LTE System

J. M. Ruiz-Avilés, M. Toril, S. Luna-Ramírez, Victor Buenestado, and M. A. Regueira

***Abstract*—Mobility load balancing (MLB) is a common technique to deal with the uneven traffic distribution in mobile networks. The aim of MLB is to alleviate congestion problems by sharing traffic demand among neighbor cells through the modification of handover parameters. MLB has been successfully used in legacy radio access technologies. However, in long term evolution (LTE), MLB may lead to severe network performance degradation due to the tight frequency reuse used in this technology. In this letter, a comprehensive analysis of the limitations of MLB in LTE is done based on the results of a classical MLB algorithm in a live LTE network. Field trial results confirm that MLB reduces network congestion at the expense of degrading cell-edge user performance in the uplink of congested cells and the downlink of adjacent cells receiving traffic.**

***Index Terms*—LTE, experimental and prototype results, mobile network, mobility load balance, handover.**

I. Introduction

During cellular network planning, operators design their networks based on traffic estimates. As network evolves, the matching between the spatial distribution of traffic demand and network resources becomes looser. Thus, it is common that some cells suffer from severe congestion problems, whereas their adjacent cells are underutilized. In this situation, Radio Resource Management (RRM) algorithms can be used to redistribute traffic and, hence, to improve network performance.

A powerful RRM technique for sharing traffic among cells is Mobility Load Balancing (MLB), which has been identified as one of the most relevant use cases in Self-Organized Networks (SON) [1]. MLB achieves load balancing through the modification of cell service areas by tuning handover parameters. MLB has been widely used in mature radio technologies, such as Global System for Mobile communications (GSM) [2] and Universal Mobile Telecommunications System (UMTS)[3], and has also been proposed for Long Term Evolution (LTE) [4]–[7]. However, MLB in LTE has important limitations due to the full frequency reuse (unlike GSM) and the absence of a soft-handover scheme (unlike UMTS). Re-allocating a user in any base station other than the one offering the largest signal level, as MLB does, often leads to reduced desired signal level and increased interference level, which might translate into poor link-level performance. Such a degradation of user performance may jeopardize the capacity benefit achieved by MLB at a network level.

In this letter, a comprehensive analysis of the limitations of MLB in LTE is presented. The aim is to explain when, why and by how much network performance is degraded by MLB. For this purpose, a classical MLB algorithm is tested in a live LTE network. The main contributions of this work are: a) to present the first results of a classical MLB algorithm in a live LTE network, b) to break down the negative effects of MLB on the overall network connection quality, and c) to quantify the performance gains and losses caused by MLB in a real case. The rest of the letter is organized as follows. Section II outlines the principles and limitations of MLB in LTE. Section III presents the results of a field trial of MLB in a live LTE network and Section IV presents the concluding remarks.

Manuscript received February 3, 2015; revised March 30, 2015; accepted April 24, 2015. Date of publication May 6, 2015; date of current version August 20, 2015. This work was funded by the Spanish Ministry of Economy and Competitiveness (TIN2012-36455), Optimi-Ericsson, and Agencia IDEA (Consejería de Ciencia, Innovación y Empresa, Junta de Andalucía, ref. 59288), co-funded by FEDER. The associate editor coordinating the review of this paper and approving it for publication was L. Badia.

J. M. Ruiz-Avilés, M. Toril, S. Luna-Ramírez, and V. Buenestado are with Communications Engineering, University of Malaga, Malaga 29071, Spain (e-mail: jmruiz@ic.uma.es).

M. A. Regueira is with Ericsson, Business Unit Global Services, Malaga 29590, Spain.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LWC.2015.2430345

II. Problem Formulation

In this section, the basics of MLB in LTE are first outlined. Then, a qualitative analysis of MLB limitations is presented.

A. MLB Algorithm in LTE

In LTE, a HandOver (HO) is triggered when the signal level received by a user from an adjacent cell is much larger than that of the serving cell, which can be expressed as

$$P_{rx}(j) + Ocn(i,j) - Hys(i) > P_{rx}(i) + Off(i), \qquad (1)$$

where $P_{rx}(i)$ and $P_{rx}(j)$ are the pilot signal levels from serving and neighbor cells i and j, respectively, $Ocn(i,j)$ is an additional margin defined for the adjacency (i,j) involved in the HO, and $Hys(i)$ and $Off(i)$ are the hysteresis and offset values defined for the serving cell i [8]. All variables in (1) are in dB. Basically, a user in cell i is handed over to cell j when $P_{rx}(j)$ exceeds $P_{rx}(i)$ by a margin, $Margin(i,j)$, defined as

$$Margin(i,j) = Off(i) + Hys(i) - Ocn(i,j). \qquad (2)$$

A low/high value of $Margin(i,j)$ determines how easy/difficult is to hand over a user from cell i to j, or, in terms of distance, how close/far from cell i the user is when it is handed over to cell j. Thus, the service area of a cell i is given by the value of $Margin(i,j)$ in all its adjacencies (i,j).

Traffic congestion due to lack of resources in a cell degrades end-user performance. The aim of load balancing is to change network parameters so as to reduce the service area of congested cells and thus steer traffic to neighbor cells with spare resources. In MLB, this is done by decreasing HO margins in highly congested cells, while increasing HO margins in surrounding cells with low congestion rates. Fig. 1 shows how the service areas of two unbalanced cells are modified by MLB. In the figure, j is the congested cell and i is the underutilized

2162-2337 © 2015 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

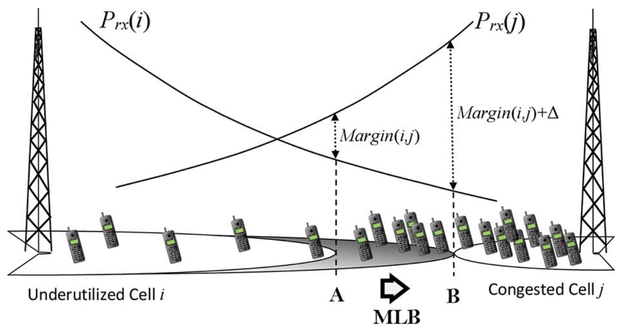

Fig. 1. Modification of cell service areas by MLB.

adjacent cell. Signal levels from both cells are superimposed, denoted by $P_{rx}(i)$ and $P_{rx}(j)$. Initially, any user moving from cell i to j is handed over at point A, where (1) is fulfilled. Then, MLB increases $Margin(i, j)$ by Δ dB. As a result, the HO point is displaced from A to B. Thus, users between A and B are now served by cell i, instead of cell j, enlarging the service area of cell i and reducing that of cell j. As a consequence, the load of cell j is reduced.

B. Limitations of MLB in LTE

For spectral efficiency reasons, full frequency reuse is configured in most live LTE networks, i.e., all cells use the whole available system bandwidth. This tight reuse causes interference problems between adjacent cells, especially in areas with large traffic demand, where MLB is applied. MLB should deal with these interference problems, since MLB steers traffic to those cells where radio collision probability is higher (i.e., those next to congested cells). In GSM, co-channel interference is mitigated by frequency planning (i.e., increasing distance to co-channel interferer) and frequency hopping (i.e., randomizing collisions), while, in UMTS, by the soft-handover scheme. These techniques are seldom used in LTE.

Interference problems are aggravated when MLB changes HO parameters. Scenario in Fig. 1 shows the case of an increase of $Margin(i, j)$ by Δ dB. Before MLB is applied, $Margin(i, j)$ is usually set to some positive value (e.g., 3 dB) in all adjacencies of the network. From Fig. 1, it is deduced that, for a user at the edge of cell i, the pilot signal level received from neighbor cell j, $P_{rx}(j)$, is $Margin(i, j)$ dB higher than that from its serving cell i, $P_{rx}(i)$. When MLB increases, the difference between pilot signal levels in the HO trigger point is also increased by Δ dB at the new cell edge.

The effect of these changes depends on radio link direction. In the *DownLink* (DL), cell-edge users at the underutilized cell i are now farther from their serving base station i and closer to the strongest interfering cell j (point B). As a result, users at B experience less desired DL signal level and higher DL interference level from cell j, which, in addition, is heavily loaded. Thus, it is expected that DL connection quality indicators get worse in cells receiving traffic (i.e., cell i in Fig. 1) when MLB is applied. Such a negative impact does not exist in the DL of the congested cell j, since cell edge is brought closer and the adjacent interfering cell is ligthly loaded. In the *UpLink* (UL), interference conditions are also deteriorated by MLB. Users reallocated in cell i (shaded area between A and B in Fig. 1) need to transmit with higher power, leading to a higher UL interference level for users in cell j. Thus, it is expected that UL connection quality indicators get worse in cells sending traffic (i.e., cell j in Fig. 1) when MLB is applied. This negative impact does not exist in the UL of the cell receiving traffic i, since former interfering users are now served by this cell.

The previous effects have a strong impact on two network performance indicators. Firstly, in the *Connection dropping ratio*, defined as the ratio between the number of dropped connections and the total number of connections. The service area of the underutilized cell i is enlarged by MLB. As a consequence, new cell-edge users in cell i (shaded area in Fig. 1) experience worse DL interference conditions. Thus, connection dropping probability increases in cells receiving traffic (i.e., cells i where $Margin(i, j)$ is increased). Similarly, in UL, MLB leads to higher UL interference levels for users in cell j, so that connection dropping probability increases for cells sending traffic (i.e., cells j where $Margin(i, j)$ is decreased). Secondly, in the *HO failure ratio*, defined as the ratio between the number of failed outgoing HOs and the total number of outgoing HO attempts. After MLB, HO from cell i to j is triggered in worse radio conditions, since signal level from the serving cell i is lower and DL interference from cell j is higher. Hence, an increase in HO failure ratio is expected in cells receiving traffic. Similarly, HO is more likely to fail in cells sending traffic due to a higher UL interference level.

The improvement in call blocking ratio obtained by MLB may not compensate for the deterioration of call dropping and HO failure ratios caused by this technique. Thus, proper monitoring of the latter performance indicators is critical to stop MLB before network connection quality degrades excessively.

III. Field Trial Analysis

In this section, a field trial of a classical MLB algorithm in a live LTE network is presented. The aim is to quantify the impact of changing HO margins for MLB. For clarity, trial setup is described first and results are presented later.

A. Trial Setup

A field trial was carried out in 2013 in an area of 713 LTE cells covering a seamless geographical area of 1000 km^2, consisting of a variety of conurbation types ranging from dense city center to rural areas. Cells include 2 carriers at 734 MHz and 2.132 GHz with 10 MHz and 5 MHz system bandwidth, respectively. Full frequency reuse is used for cells in the same carrier. Best effort data service is more than 99% of the network traffic. Network equipment was compliant to 3GPP Release 8. Base stations have 2 transmit/receive antennas and possible Transmission Modes (TM) are TM 0 (Single transmit antenna) and TM 2 (Open loop spatial multiplexing with cyclic delay diversity).

Initially, all cells and adjacencies are set with the default parameter settings $Ocn(i, j) = 0$ dB, $Hys(i) = 1$ dB and $Off(i) = 3$ dB $\forall\, i, j$. Thus, $Margin(i, j)$ is set to 4 dB for all adjacencies. Due to uneven traffic distribution, some cells in the area suffer from congestion problems with the default settings. To solve these problems, a classical MLB algorithm, similar to that proposed in [5], is executed. Once triggered, MLB changes $Margin(i, j)$ by modifying $Ocn(i, j)$, while $Hys(i)$ and $Off(i)$ remain unchanged. For stability reasons, changes in $Ocn(i, j)$ are implemented by 1 dB steps and constrained to the interval

$[-3, 3]$ dB. The hysteresis area between any pair of neighbor cells is maintained by forcing that $Ocn(i, j) + Ocn(j, i) = 0\ \forall\, i, j$.

The MLB tuning process is carried out for 3 weeks (weeks 1–3). Changes in $Margin(i, j)$ are implemented at the end of every working day. In a fourth week (week 4), HO parameters are not changed. Finally, in a fifth week (week 5), HO parameters are reverted back to the default settings. To assess MLB performance, two periods are analyzed: a) three days (Wednesday, Thursday and Friday) of week 4, to measure network performance at the end of the MLB tuning process (hereafter referred to as *optimized* network configuration), and b) the same office days in week 5, to measure network performance before MLB process (hereafter referred to as *baseline* network configuration).

For ease of analysis, cells are broken down into 6 overlapping groups. A first group, *Group R* (for receiving), consists of cells that only receive traffic through the increase of HO margins (or, equivalently, decrease of *Ocn*). A cell i is included in this group if $Ocn(i, j) \leq 0\ \forall j$ and $Ocn(i, j) < 0$ for at least one cell j, where $j \in N(i)$ and $N(i)$ is the set of neighbors of cell i. A second group, *Group R+*, is a subgroup of R comprising those cells in R with the largest decrease in *Ocn*, for which traffic absorption is more intense. A cell i belongs to $R+$ if it belongs to group R and $Ocn(i, j) \leq -2$ dB for more than 3 of its neighbor cells. A third group, *Group S* (for sending), consists of cells only sending traffic through the decrease of their HO margins (i.e., increase of *Ocn* values). A cell i is included in this group if $Ocn(i, j) \geq 0\ \forall j$ and $Ocn(i, j) > 0$ for at least one cell j, where $j \in N(i)$. A fourth group, *Group S+*, is a subgroup of cells in S having the largest increase in *Ocn*, where traffic steering is more intense. A cell i belongs to $S+$ if it belongs to group S and $Ocn(i, j) \geq 2$ dB for more than 3 of its neighbors. A fifth grouph, *Group NSNR* (for neither sending nor receiving), consists of cells neither sending nor receiving traffic. A cell i belongs to *NSNR* if $Ocn(i, j) = 0$ dB $\forall j \in N(i)$. Finally, a sixth group, *Group SR* (for sending and receiving), consists of cells receiving and sending traffic simultaneously at different adjacencies. A cell i belongs to *SR* if $Ocn(i, j) > 0$ and $Ocn(i, k) < 0$ dB for at least one j and $k \in N(i)$.

To analyze the impact of MLB on radio access network performance, two low-level indicators are collected: a) UL interference level per Physical Resource Block (PRB) on Physical UL Shared Channel (PUSCH), $I_{UL,PRB}$, as a measure of the overall UL connection quality, and b) Channel Quality Indicator, *CQI*, as a measure of DL connection quality. For the final network assessment, two key performance indicators related to network quality are used: a) the connection dropping ratio, *CDR*, calculated on a per-cell basis, and b) the HO failure ratio, *HFR*, calculated on an adjacency basis. All these indicators are collected in the before (i.e., baseline) and after (i.e., optimized) periods.

B. Trial Results

The analysis begins by checking the need for MLB. The average PRB utilization ratio in the trial area with the baseline configuration during the busy hour is 28.55%, and the 5$^{\text{th}}$- and 95$^{\text{th}}$- percentiles of PRB utilization ratio are 6.46% and 61.69%, respectively. Moreover, 47 out of 713 cells show a non-negligible number of Radio Resource Control (RRC) connection failures due to high processor load. These measurements are clear evidence of the uneven distribution of traffic demand

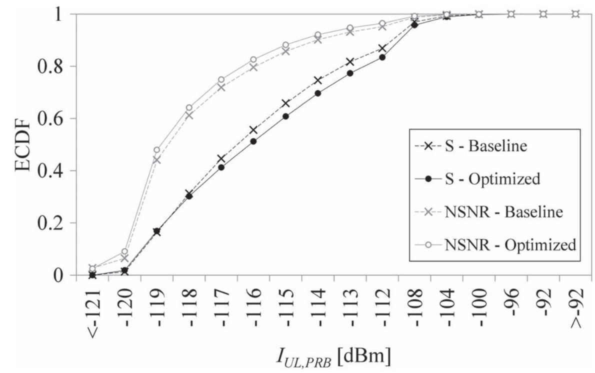

Fig. 2. ECDF of average PUSCH interference for cells sending traffic.

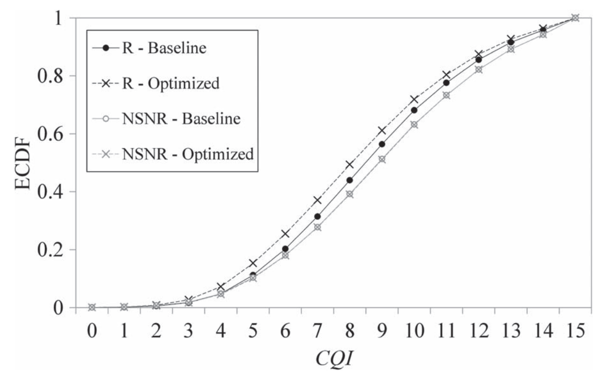

Fig. 3. ECDF of average CQI for cells receiving traffic.

in the trial area, which is the reason for executing MLB. After MLB, the number of initiated connections in cells of group R and $R+$ is increased by 3.56% and 10.59%, and decreased in group S and $S+$ by 3.52% and 15.88% respectively. Thus, it is confirmed that changes in HO margins carried out by MLB modify spatial traffic distribution. Moreover, the maximum number of active users in a cell is reduced by 40%. This reduction is clear evidence that the most congested cells in terms of users have been offloaded. As a result, the number of RRC connection failures due to high load decreases by 80%, which clearly shows the benefits of MLB.

Traffic redistribution is achieved at the expense of causing UL interference problems in cells sharing traffic. Fig. 2 shows the Empirical Cumulative Density Function (ECDF) of $I_{UL,PRB}$ for cells S before and after MLB, compared with cells *NSNR*. In the figure, it is observed that UL interference in cells S increases after MLB. Specifically, the average and 95$^{\text{th}}$-percentile of $I_{UL,PRB}$ after MLB are -115.03 and -108.79 dBm, respectively, compared to -114.61 and -108.25 dBm before MLB. Note that such a difference of less than 1 dB in the daily averages of both indicators might have a significant effect on connection quality for power-limited users in the UL.

DL connection quality is also impaired after MLB. Fig. 3 shows the ECDF of average CQI for cells R and *NSNR* before and after MLB. It is observed that the CDF moves to the left after MLB in cells R, showing that cells receiving traffic experience worse average DL channel quality. Specifically, the average *CQI* changes from 9.11 to 8.72, and so does the 5$^{\text{th}}$-percentile from 4.03 to 3.51. In contrast, CDF of CQI of cells *NSNR* (i.e., non-adjusted cells) remains the same.

The above-described impairments are reflected into key network performance indicators. Table I shows *CDR* before and

TABLE I
CONNECTION DROPPING RATIO STATISTICS

Group	R	R+	S	S+	NSNR	SR
Number of cells	261	27	86	33	29	337
$CDR_{baseline}[\%]$	0.14	0.11	0.26	0.22	0.16	0.18
$CDR_{optimized}[\%]$	0.20	0.21	0.29	0.27	0.15	0.24
$\Delta CDR[\%]$	44.4	84.8	13.1	19.4	-11.4	35.2

TABLE II
HANDOVER FAILURE RATIO STATISTICS

$Ocn(i,j)$ [dB]	-3	-2	-1	0	1	2	3
$HFR_{baseline}[\%]$	0.35	0.41	0.47	0.92	0.45	0.35	0.31
$HFR_{optimized}[\%]$	0.68	0.80	0.74	1.09	0.58	0.57	0.57
$\Delta HFR[\%]$	97.5	93.7	57.9	17.6	27.5	63.8	81.4

after MLB, $CDR_{baseline}$ and $CDR_{optimized}$, broken down by cell group. The table also includes variations of *CDR*, $\Delta CDR = (CDR_{optimized}\text{-}CDR_{baseline})/CDR_{baseline}$. It is observed that all groups changing their HO margins on average (i.e., R, R+, S, S+, and SR) endup with worse *CDR*. Cells receiving traffic (group *R*) increase their *CDR* from 0.14% to 0.20% after MLB (i.e., 44.4% increase). More importantly, it is observed that $\Delta CDR = 84.8\%$ for cells *R*+. This degradation proves the strong impact of the increase in DL interference at the edge of cells *R*. Cells sending traffic also get worse *CDR* (13.1% and 19.4% in *S* and *S*+), due to the increase of UL interference in cells *S*. Finally, cells *NSNR*, which do not modify *Ocn*, slightly improve their *CDR* by 11.4%.

Table II shows *HFR* results broken down by the value of $Ocn(i,j)$ at the end of the MLB process (recall that $Ocn(i,j) = 0 \forall\, i,j$ before MLB). In the table, a global degradation of *HFR* is observed, since all adjacencies display worse *HFR* after MLB. As expected, degradation is higher in adjacencies with larger deviations from the default 0 dB value. Specifically, adjacencies with *Ocn* modifications of -3, -2 and 3 dB show $\Delta HFR = 97.5\%$, 93.7% and 81.4% respectively. From the previous analysis, it is clear that the degradation of low-level indicators is strongly correlated with that of high-level network performance indicators. Hence, careful monitoring of low-level indicators for users about to perform (or just after having performed) a HO might be used to stop MLB, and thus keep the impairment of high-level indicators within reasonable limits.

IV. CONCLUSION

In this work, an analysis of the limitations of mobility load balancing in LTE has been carried out based on the first results of a field trial in a live LTE network. Trial results have shown that network parameter changes produced by mobility load balancing alleviate congestion problems at the expense of decreasing average *CQI* in the downlink and increasing PUSCH interference level in the uplink, especially in cells where handover margins are strongly modified. Connection dropping and handover failure ratios are increased by up to 85% and 97%, respectively, in those cells and adjacencies with the largest deviation of handover margins. These results point out the need for adding new criteria in load balancing algorithms to avoid excessive performance degradation in LTE. For this purpose, the low-level performance indicators used here to assess uplink and downlink network performance might be used to detect problems in advance, especially when measured just before/after HO is triggered.

REFERENCES

[1] Next Generation Mobile Networks (NGMN), “Recommendation on SON and O&M requirements,” Frankfurt, Germany, Tech. Rep., 2008.

[2] M. Toril and V. Wille, “Optimization of handover parameters for traffic sharing in GERAN,” *Wireless Pers. Commun.*, vol. 47, no. 3, pp. 315–336, Nov. 2008.

[3] J. Li, C. Fan, D. Yang, and J. Gu, “UMTS soft handover algorithm with adaptive thresholds for load balancing,” in *Proc. IEEE 62nd VTC*, Sep. 2005, vol. 4, pp. 2508–2512.

[4] A. Lobinger, S. Stefanski, T. Jansen, and I. Balan, “Load balancing in downlink LTE self-optimizing networks,” in *Proc. IEEE 71st VTC*, May 2010, pp. 1–5.

[5] R. Kwan, R. Arnott, R. Paterson, R. Trivisonno, and M. Kubota, “On mobility load balancing for LTE systems,” in *Proc. IEEE 72nd VTC*, Sep. 2010, pp. 1–5.

[6] S. S. Mwanje, and A. Mitschele-Thiel, “A q-learning strategy for LTE mobility load balancing,” in *Proc. IEEE PIMRC*, Sep. 2013, pp. 2154–2158.

[7] Z. Altman, S. Sallem, R. Nasri, B. Sayrac, and M. Clerc, “Particle swarm optimization for mobility load balancing son in LTE networks,” in *Proc. IEEE WCNCW*, Apr. 2014, pp. 172–177.

[8] *EUTRA; Radio Resource Control; Protocol specification*, TS 36.331 v12.3.0, 3rd Generation Partnership Project (3GPP), Sep. 2014.