On Constrained Cell Load Coupling With Applications in LTE Networks

Iana Siomina and Di Yuan

Linköping University Post Print

N.B.: When citing this work, cite the original article.

©2014 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.


http://dx.doi.org/10.1109/LCOMM.2014.2340865

Postprint available at: Linköping University Electronic Press

http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-112835
On Constrained Cell Load Coupling with Applications in LTE Networks

Iana Siomina and Di Yuan Member, IEEE

Abstract—For LTE networks, the levels of resource consumption of the cells, a.k.a. cell load factors, are coupled with each other in a non-linear manner. We establish theoretical and algorithmic results for the constrained load-coupling model. Our focus is on solution existence, i.e., whether or not a network configuration is able to support the traffic demand within the resource limit. We provide theoretical characterizations and concepts for determining feasibility. The concepts are illustrated for maximum demand scaling. Empirical insights are gained via a heterogeneous LTE network scenario.

Index Terms – Capacity, cell load, long term evolution.

I. INTRODUCTION

There is a growing interest in performance modeling of long term evolution (LTE) with orthogonal frequency division multiple access (OFDMA). The efforts in [1], [2], [3] have led to a performance model that accounts for the coupling relation among cell loads. Here, the term cell load, a.k.a. activity factor, refers to the average utilization level of the time-spectrum resource blocks (RBs) in a cell. It is an aggregation over the resource usage of the cell’s users. The load of one cell dictates the interference this cell generates to other cells, and vice versa. At the network level, the cell loads together form a vector. The term load coupling refers to the dependency relation between the load elements of the vector due to mutual interference between cells. The LTE load-coupling model exhibits a non-linear relation between the load elements. This is in contrast to Universal Mobile Telecommunications System (UMTS), where the resource (i.e., power) consumptions of cells are coupled via a linear system [4], [5].

The LTE load-coupling model has been developed in three independent works [1], [2], [3]. The focus of these studies has been on deriving the model and illustrating its use for network performance evaluation. In addition, a linear approximation of the model is given in [2]. Whereas the derivations in [1], [2] are based on using cell load for interference scaling, the authors of [3] took a queuing-theory perspective. Interestingly, the resulting models coincide. In [6], some theoretical properties of the load-coupling model are provided. On the application side, the model has been used in the context of load balancing and offloading in heterogeneous networks (HetNets), respectively, in [7] and [8].

We investigate the constrained load-coupling model, that is, a cell’s load is subject to a maximum limit representing the amount of available resource. This is in contrast to [1], [2], [3], [6] where the derivation and analysis are limited to unconstrained load coupling. A straightforward though naive way to deal with the constraint of maximum load is to solve first the unconstrained model, followed by checking whether or not any cell load exceeds the limit [7], [8]. We present significantly more efficient approaches. Thus the current study extends the results of load coupling in the existing literature.

For performance evaluation, it is important that the load-coupling model is accurate. In [3], the authors not only derived the model independently from [1], [2], but also performed system-level, dynamic simulations to assess the accuracy of the model. The results and conclusions in [3] verify that the load levels provided by the model are indeed close to those from system-level simulations, and the model is sufficiently accurate for network-level performance characterization. This result is of particular importance in the context of network configuration optimization. As there are a large number of candidate configurations to be considered, applying network-wide simulation to each of them is hardly affordable, whereas performance characterization with the load-coupling model provides a much more viable approach. Indeed, a key motivation of our study originates from network optimization. There are many parameters that should be optimized in LTE networks and HetNets, such as small cell location, range offset for user-cell association, as well as power (even though power control is not yet common for downlink in current LTE networks), and so on. For example, optimal range and location of small cells are studied in [7], [8], respectively. These optimization tasks are typically non-convex and combinatorial in nature (see, e.g., [8] for an NP-hardness proof for user association). Thus optimization algorithms often rely on searching and evaluating a large number of candidate configurations. Here, a configuration is a network-wide parameter setup, e.g., a vector representing user-cell association. For each configuration, the corresponding load-coupling model comes into play for evaluation. We provide theoretical insights as well as algorithmic notions to efficiently determine, via the load-coupling model, the maximum traffic demand that can be supported by the candidate configuration in question, and illustrate the insights for an LTE HetNet scenario. Thus our approach can act as an embedded module in network configuration optimization, which serves as the general background to the current study.

II. THE CELL LOAD-COUPLING MODEL

Denote by \( \mathcal{N} = \{1, \ldots, n\} \) the cell set, and by \( \mathcal{J}_i \) the set of demand elements of cell \( i \in \mathcal{N} \). A demand element may be a user equipment (UE), or an aggregated demand node representing multiple UEs [9]. The normalized demand of element \( j \) is denoted by \( r_j \). Denote by \( \rho = (\rho_1, \ldots, \rho_n)^T \) the

I. Siomina is with Ericsson Research, Ericsson AB Sweden (email: iana.siomina@ericsson.com).
D. Yuan is with Linköping University, Sweden (email: diyu@itn.liu.se). His work has been funded by the Swedish Foundation of Strategic Research.
cell load vector. The constrained load-coupling (CLC) model is given in a compact form below.

\[ \text{[CLC]} \quad \rho = f(\rho) \quad \text{(1a)} \]
\[ 0 \leq \rho \leq \rho^{\text{max}} \quad \text{(1b)} \]

In (1b), vector \( \rho^{\text{max}} \) is the resource limit. Function \( f(\rho) \), given below for an arbitrary cell \( i \), specifies the non-linear coupling relation. Further details are provided in [6].

\[ f_i(\rho) = \sum_{j \in c_i} \log_2 \left( 1 + \frac{r_{ij}}{\sum_{k \in C \setminus \{i\}} b_{ikj} \rho_k + c_{ij}} \right) \quad \text{(2)} \]

In (2), \( b_{ikj} \) is the ratio between the received interfering signal of cell \( k \) and that of the serving cell \( i \). Parameter \( c_{ij} \) is the noise effect over the received signal power. By (2), the load of a cell is a non-linear function of the other cells’ load levels. The elements of \( \rho \) are coupled together by equation \( \rho = f(\rho) \), of which the solution cannot be readily obtained in closed form. As the equation system is non-linear, computational approaches for linear systems do not apply. We denote by ULC the unconstrained load coupling model, i.e., \( \{ \rho = f(\rho), \rho \geq 0 \} \). Its solution, if exists, is denoted by \( \rho^* \). Its solution, if exists, is denoted by \( \rho^* \).

\[ \text{Fig. 1. A maximum-load feasibility example for two cells.} \]

Figure 1 illustrates \( f \) for a two-cell scenario, namely how \( f_1 \) grows in \( \rho_2 \) and how \( f_2 \) grows in \( \rho_1 \), denoted by \( f_1(\rho) \) and \( f_2(\rho) \), respectively. The intersection point is \( \rho^* \). A discussion of the other entities in the figure is deferred to the next section. Below we summarize some useful properties of ULC [6].

1. **Function property:** \( f_i, i \in C \), is twice differentiable, strictly concave, and monotonously increasing in \( \rho \).
2. **Solution uniqueness:** \( \rho^* \), if exists, is unique.
3. **Lower bound:** For the \( n \times n \) matrix \( H^0 \) with zero diagonal and \( H^0_{ii} = \ln(2) \), the \( \rho^0 = f(\rho^0) \) has solution \( \rho^0 \in \mathbb{R}^n_+ \), then \( \rho^0 \leq \rho^* \).
4. **Upper bound:** For any \( \hat{\rho} \in \mathbb{R}^n_+ \) and the \( n \times n \) matrix \( H \) with zero diagonal and \( H_{kk} = \frac{\partial f}{\partial \rho_k}(\hat{\rho}), i \neq k \), if \( \rho = H(\rho - \hat{\rho}) + f(\hat{\rho}) \) has solution \( \rho \in \mathbb{R}^n_+ \), then \( \rho^* \leq \rho \).

III. FEASIBILITY OF CONSTRAINED LOAD COUPLING

We assume that \( \rho^* \) exists to ULC. The straightforward approach for the feasibility of CLC is to solve ULC (e.g., via fixed-point iterations [10]), and then compare \( \rho^* \) to \( \rho^{\text{max}} \). We target determining the feasibility of CLC more efficiently.

For any \( \rho \in \mathbb{R}^n_+ \), denote by \( N^>^>(\rho), N^=^=(\rho), \) and \( N^<^<^<(\rho) \) the sets of cells for which the elements of \( \rho \) are greater than, equal to, and less than the load function values evaluated at vector \( \rho \), respectively. That is, \( N^>^>(\rho) = \{ i \in C : \rho_i > f_i(\rho) \} \); the other two sets are defined analogously. Let \( N^>=^>(\rho) = N^>^>(\rho) \cup N^=^=(\rho) \) and \( N^<^<(\rho) = N^<^<(\rho) \cup N^=^=(\rho) \).

**Definition 1:** Define \( Q \) the set of functions, such that \( q : \mathbb{R}^n_+ \rightarrow \mathbb{R}^n_+ \) is in \( Q \) if and only if for any given \( \rho \in \mathbb{R}^n_+ \), \( f_i(\rho) \geq q_i(\rho) \geq \rho_i \) for at least one \( i \in N^>^>^>(\rho) \), and \( q_i(\rho) = \rho_i \) for all \( i \in N^=^=(\rho) \), where \( \rho \) denotes the partial fixed point defined for the cells in \( N^>^>(\rho) \), i.e., \( f_i(\rho) = \rho_i \), \( \forall i \in N^>^>(\rho) \), and \( \rho_i = \rho_i \), \( \forall i \in N^<^<(\rho) \).

Note that the function definition is broad. As an example, one can set \( q_i(\rho) = f_i(\rho) \) for one or any subset of the elements in \( N^>^>(\rho) \).

**Theorem 1:** For any \( \rho \in \mathbb{R}^n_+ \) and \( q \in Q \), the following relation holds: \( \rho^* \leq \rho \) if and only if \( \rho^* \leq q_0(\rho) \).

**Proof:** The sufficiency follows immediately from that \( q(\rho) \leq \rho \). Consider necessity first for \( q_i(\rho) = f_i(\rho) \), \( i \in N^>^>(\rho) \), and assume \( \rho^* \leq \rho \). For all cells in \( N^>^>(\rho) \), \( \rho_i \leq \rho_i \). Then, by the definitions of \( Q \) and \( \rho^* \), \( f_i(\rho) = \rho_i \text{, } f_i(\rho) < \rho_i \text{, leading to } f_i(\rho) < f_i(\rho^*) \). The inequality contradicts Property 1. Hence \( q_i(\rho) \geq \rho_i^* \), and repeating the update gives a sequence converging to \( \rho \), from which the theorem follows.

**Theorem 2:** For any \( q \in Q \), the sequence \( \rho^0, \rho^1, \ldots, \rho^k \), where \( \rho^0 = \rho^{\text{max}} \) and \( \rho^k = f(\rho^{k-1}) \), \( k \geq 1 \), satisfies \( \rho^* \leq \rho^k \), \( \forall k \geq 1 \), and only if \( \rho^* \leq \rho^{\text{max}} \).

**Proof:** The result follows from applying Theorem 1 repeatedly, starting from \( \rho^{\text{max}} \).

We state two useful lemmas (see [7] for proofs).

**Lemma 1:** For any \( \rho^0 \in \mathbb{R}^n_+ \) with \( \rho^0 \geq f(\rho^0) \), the fixed-point iteration \( \rho^i = f(\rho^{i-1}) \) gives a sequence of monotonically decreasing load vectors \( \rho^0 \geq \rho^1 \geq \ldots \geq \rho^\ell \geq \ldots \geq \rho^* \), and the sequence converges to \( \rho^* \) when \( \ell \rightarrow \infty \).

**Proof:** The second case is derived similarly via Lemma 2 and Theorem 2.

**Theorem 3:** Consider \( \rho^k \) for any \( k \geq 0 \) in Theorem 2. If \( N^=^=(\rho^k) = C \), i.e., \( \rho^k \geq f(\rho^k) \), then \( \rho^* \leq \rho^{\text{max}} \). Conversely, if \( N^<^<(\rho^k) = C \), i.e., \( \rho^k \geq f(\rho^k) \), then \( \rho^* \leq \rho^{\text{max}} \).

**Proof:** Suppose the opposite, i.e., both \( N^>^>(\rho^k) \neq \emptyset \) and \( N^<^<(\rho^k) \neq \emptyset \) remain. As \( N^=^=(\rho^k) \) and \( N^<^<(\rho^k) \) will not shrink and expand, respectively, there exists \( k \), such that \( N^>^>(\rho^k) \) and \( N^<^<(\rho^k) \) do not change for any \( k \geq k \). By Definition 1, the sequence converges to \( \rho^k \), that is, the partial fixed
point defined for $N^>(p^k)$. Hence $N^>(p^k)$ is diminishing, contradicting the assumption, and the theorem follows.

Consider again Figure 1. Among $\rho^A$, $\rho^B$, $\rho^C$, and $\rho^E$, setting $\rho^A$ or $\rho^B$ as $\rho_{\text{max}}$ satisfies one of the conditions in Theorem 3. Thus $\rho^* \leq \rho^A$ and $\rho^* \leq \rho^B$. For $\rho^D$, applying $q \in Q$ in dimension $\rho_2$ arrives at $q(p^D) = \rho^D$, from which $\rho^* \leq \rho^D$ by Theorem 3. For $\rho^C$, $q(\rho^C) = \rho^C$, hence $\rho^* \leq \rho^C$. It is easily observed that, for two cells, at most one iteration of Theorem 1 suffices, giving the corollary below.

**Corollary 1:** If $n = 2$, one of the two conditions in Theorem 3 holds with $k \leq 1$ for the sequence in Theorem 2.

All results above remain valid if $q$ limits its computation to any subset (possibly one cell) in $N^>(p^k), \forall k \geq 1$. The proofs of Theorems 2 and 3 give the following corollary.

**Corollary 2:** For any $k \geq 0$, Theorems 2 and 3 remain valid for the partial update $N^>(p^k) = N^>(p^{k-1}) \cup N$ and $N^<(p^k) = N^<(p^{k-1}) \setminus N$, where $0 \subset N \subseteq \{i \in N^<(p^{k-1}) : \rho^k_i \geq f_i(p^k)\}$, provided that the latter is non-empty.

A pre-test of feasibility can be carried out, if an upper or lower bound of $\rho^*$ is at hand (cf. Properties 3 and 4). Specifically, for any $\rho \geq \rho^*$, CLC is feasible if $f(\rho) \leq \rho_{\text{max}}$, and for any $\rho \leq \rho^*$, CLC is infeasible if $f(\rho) > \rho_{\text{max}}$ for any $i \in N$. These two results follow from $\rho^* = f(\rho^*) < f(\rho)$, and $\rho^* = f(\rho^*) > f(\rho)$, where the inequalities are by Property 1.

## IV. An Application Related to Network Capacity

Given base traffic demand $r_j, j \in J_i, i \in N$, a problem of interest to network capacity is to scale up the base demand by $\kappa > 0$ and maximize $\kappa$. We denote by $f(\kappa, \rho)$ the load function with scaling parameter $\kappa$. The problem is stated below.

\[
\text{[CAP]} \quad \kappa \rightarrow \text{max} \quad \rho = f(\kappa, \rho) \quad 0 < \rho \leq \rho_{\text{max}}
\]

Denote the solution of ULC with $\kappa$ by $\rho^*(\kappa)$. The natural approach for obtaining the optimum $\kappa^*$ is to perform a bisection search on an interval $[\tilde{\kappa}, \bar{\kappa}]$ confining $\kappa^*$, utilizing the concepts in Section III repeatedly. In the bi-section search, part of the results of one trial $\kappa$ can be reused in latter trials for better efficiency. The following theorem provides a warm start for the bi-section search.

**Theorem 5:** If the feasibility of CLC under $\kappa_1$ is concluded by Theorem 3 for $\rho^*$ in the sequence in Theorem 2, then for any $\kappa_2 \leq \kappa_1$, $\rho^*(\kappa_2) \leq \rho_{\text{max}}$ if and only if $\rho^*(\kappa_2) \leq \rho^*$. 

**Proof:** Denote by $\rho^1(\kappa_1)$ and $\rho^2(\kappa_2)$ the outcomes of applying $q \in Q$ once to $\rho_{\text{max}}$ for $\kappa_1$ and $\kappa_2$, respectively. By Theorem 1, $\rho^*(\kappa_2) \leq \rho_{\text{max}}$ if and only if $\rho^1(\kappa_2) \leq \rho_{\text{max}}$. The theorem’s proof holds for $\rho^1(\kappa_1)$ as well, as $\rho^1(\kappa_1) \geq \rho^1(\kappa_2)$. Thus $\rho^1(\kappa_2) \leq \rho_{\text{max}}$ if and only if $\rho^*(\kappa_2) \leq \rho^1(\kappa_1)$. Repeating the argument establishes the theorem.

By Theorem 5, $\rho^0$ in Theorem 2 is necessarily $\rho_{\text{max}}$ only initially. It can then be improved (i.e., becomes smaller) each time CLC is concluded infeasible during the bi-section search.

Another source of reuse stems from the upper and lower bounds used in the pre-test. We have the following result.

**Theorem 6:** Assume $\kappa_1 \leq \kappa_2$. If $\rho^*(\kappa_2) \leq \rho$, then $\rho^*(\kappa_1) \leq \min(\rho, f(\kappa_1, \rho))$. If $\rho \leq \rho^*(\kappa_1)$, then $\max(\rho, f(\kappa_2, \rho)) \leq \rho^*(\kappa_2)$.

**Proof:** By (2), $\rho^*(\kappa_1) = f(\kappa_1, \rho^*(\kappa_1)) \leq f(\kappa_2, \rho^*(\kappa_1))$. Thus $\rho^*(\kappa_1)$ satisfies the condition in Lemma 2 under $\kappa_2$, giving $\rho^*(\kappa_1) \leq \rho^*(\kappa_2)$. Hence $\rho^*(\kappa_1) \leq \rho$. Setting $\rho_{\text{max}} = \rho$ and noting that the min-operator qualifies as $q$ in Theorem 1, $\rho^*(\kappa_1) \leq \min(\rho, f(\kappa_1, \rho))$. Reversing the definition of $q$ and the relations in Theorem 1 gives the second claim.

By Theorem 6, upper and lower bounds $\hat{\rho}$ and $\bar{\rho}$ can be successively improved. Suppose $[\tilde{\kappa}, \bar{\kappa}]$ is the current range, with $\tilde{\rho} \leq \rho^*(\kappa)$ and $\rho^*(\kappa) \leq \bar{\rho}$. Let $\kappa^*$ be the current trial value. If CLC is infeasible, then $\kappa^* \in [\tilde{\kappa}, \bar{\kappa}]$ and $\min(\rho, f(\kappa, \rho))$ is a valid upper bound. Note that $f(\kappa, \rho)$ is computed in the pre-test of $\kappa$ and at no extra cost. Conversely, if CLC is feasible, then $\rho$ can be updated to max($\rho(\tilde{\kappa}, f(\kappa, \rho))$).

**Algorithm 1** Maximum demand scaling

**Input:** $\rho_{\text{max}}, f(), q(), \rho \leq \rho^*(\kappa), \rho \geq \rho^*(\kappa), \kappa \geq \kappa^*, \kappa \leq \kappa^*, \epsilon_k$

1. $\rho' \leftarrow \rho_{\text{max}}$
2. repeat
3. $\kappa \leftarrow (\tilde{\kappa} + \bar{\kappa})/2, \rho^0 \leftarrow \rho'$, stat $\leftarrow$ null
4. If $f(\kappa, \rho) \leq \rho^0$ then
5. stat $\leftarrow$ feasible, $\rho \leftarrow \max(\rho, f(\kappa, \rho))$
6. If $f(\kappa, \rho) > \rho_{\text{max}}$ for any $i \in N'$ then
7. stat $\leftarrow$ infeasible, $\rho \leftarrow \min(\rho, f(\kappa, \rho))$
8. $\rho^0 \leftarrow \kappa, \rho \leftarrow 0$
9. while stat $\neq$ null do
10. $\kappa \leftarrow k + 1, \rho^k \leftarrow q(\rho^{k-1})$
11. If $N^>(\rho^k) = N$ then
12. stat $\leftarrow$ feasible, $\rho \leftarrow \max(\rho, f(\kappa, \rho)), \kappa \leftarrow \kappa$
13. If $N^>(\rho^k) = N$ then
14. stat $\leftarrow$ infeasible, $\rho \leftarrow \min(\rho, f(\kappa, \rho)), \kappa \leftarrow \kappa, \rho^k \leftarrow \rho^k$
15. until $\kappa - \tilde{\kappa} \leq \epsilon_k$
16. return $\kappa$

Putting the concepts together leads to Algorithm 1. In the description, $\epsilon_k$ denotes the precision requirement. Note that notation $q$ represents the extension of Definition 1 to $f(\kappa, \rho)$.

**V. Numerical Illustration**

We numerically investigate Algorithm 1 for the LTE HetNet scenario [11] in Figure 2 with an inter-site distance of 500 meters and wrap-around. There are 9 macro-cells. Two small cells are placed within each macro-cell area, giving 18 small cells in total. The macro-cells and small cells use directional and omni-directional antennas with output power 46 dBm and 30 dBm, respectively. There are 30 UEs per macro-cell area, out of which 2/3 are in two hotspots, and 1/3 are distributed randomly and uniformly. The system operates at 2 GHz with bandwidth 10 MHz. The propagation environment and UE distribution follow 3GPP specification with Propagation Model 1 (Okumura-Hata, urban, 8 dB standard deviation for shadow fading) and User Generation Scenario 4b. The base demand before scaling is 100 Kbps for all UEs. Four range offsets, {0, 3, 6, 9} dB, have been considered for the small cells.

The straightforward approach for determining the feasibility of CLC is to use regular fixed-point iterations to compute $\rho^*$ and check if $\rho^* \leq \rho_{\text{max}}$. This is used as the baseline method for comparison. For Algorithm 1, we use a simple choice of function $q$: Given $\rho^{k-1}, q(\rho^{k-1})$ computes $f_i(\kappa, \rho^{k-1}), \forall i \in$
the feasibility of CLC. We observe that a large offset results in many UEs being overloaded. Offset 9 dB gives the opposite effect. Thus the choice of offset and the resulting amount of UEs offloaded to small cells have a direct impact on the traffic that can be served.

If cell load exceeds 1.0, the cell is overloaded and CLC is infeasible, i.e., the traffic demand cannot be met with the time-spectrum resource available. In this case, the values above one indicate the amount of resource shortage. In Figure 4, infeasibility occurs for offsets 0 dB and 9 dB. With the baseline method, ULC is solved, before this infeasibility could be concluded. The proposed approach resolves the feasibility aspect without computing the exact load values. Hence infeasible network configurations can be more rapidly identified.

VI. CONCLUSIONS

We have investigated constrained cell-load coupling. Recall that the current study has been motivated by optimizing performance and capacity via network configuration, such as cell range and user association. Hence, even though network configuration problems are not within the scope of this letter, they represent interesting extensions for further investigations, for which the insights gained from the current study, namely how to determine if a candidate configuration results in feasible loads, will be embedded into the optimization process.

REFERENCES


