Efficient Tracking Area Management in Carrier Cloud

Miloud Bagaa[‖], Tarik Taleb[*] and Adlen Ksentini[§]
[‖] Department of Theories and Computer Engineering, CERIST, Algiers, Algeria. Email: bagaa@mail.cerist.dz
[*] School of Electrical Engineering, Aalto University, Finland. Email: taleb@ieee.org
[§] IRISA, University of Rennes 1, Rennes, France Email: adlen.ksentini@irisa.fr

***Abstract*—One important objective of 5G mobile networks is to accommodate a diverse and ever-creasing number of user equipment (UEs). Coping with the massive signaling overhead expected from UEs is an important hurdle to tackle to achieve this objective. In this paper, we propose three solutions that aim for finding optimal distributions of tracking areas (TAs) in the form of TA lists (TALs) and assigning them to UEs, with the objectives of minimizing two conflicting metrics, namely paging overhead and tracking area update (TAU) overhead. Two solutions favors one objective than the other. The third one incorporates a novel scheme, dubbed Fair and Optimal TAL Assignment (FOTA), based on Nash bargaining game theory. FOTA improves overall network performance minimizing overhead due to both paging and TAU messages, taking into account the behavior and mobility features of UEs. The performance of proposed schemes are evaluated via simulations and the obtained results demonstrate their feasibility and ability in achieving their design goals, improving network performance by minimizing cost associated with paging and TAU.**

I. INTRODUCTION

There is a global commitment from major mobile operators to build the 5G mobile architecture to handle the growing mobile traffic as well as to allow the deployment of new mobile services in an elastic and flexible manner. Indeed, the 5G vision is not only focusing on increasing data rate but also unveils a complete new architecture for the mobile network. Optimizing the Radio Access Network (RAN) is in the heart of the 5G vision, whereby the cloudification of RAN is envisioned as a vital solution to reduce both operational and capital expenditures for mobile operators. User equipments (UEs) are usually in the idle mode and have no call activity for a long duration. When an incoming connection is destined to a UE in idle mode, the Mobility Management Entity (MME) sends a signaling message, namely paging, to all eNodeBs to find the UE's location (cell) in the network. Accordingly, in case a high number of UEs need to be paged, a massive number of downlink signaling messages have to be transmitted, resulting in high signaling overhead and wasting scarce resources of the mobile network. To overcome this issue, the Tracking Area (TA) concept has been introduced in Release 8 of the 3GPP mobile network specifications.

The key idea beneath the TA principle consists in grouping several cells into one TA. MME keeps record of the location of UEs in idle mode at the TA granularity. Thus, when a connection setup request comes for a UE in idle mode, the UE in question is paged only within its current TA. Each time a UE moves to a new location and connects to a new cell not belonging to its current TA, the UE sends an uplink message, namely Tracking Area Update (TAU), to MME, that subsequently updates the TA of the UE. In this vein, it is worth noting that a TA is also defined as an area where the UE can move without transmitting TAU messages to MME. Recently, 3GPP introduced the Tracking Area List (TAL) concept to further render TA configuration more flexibile. The TAL concept aims at reducing the uplink signaling messages, TAU, by grouping several TAs in one TAL. In the recent Release 12 of the 3GPP specifications, MME can assign different TALs to UEs when they enter a new TA. Although assigning TALs with high number of TAs to a UE reduces the number of TAUs, it also increases the number of paging messages as the search area would be large. In this paper, we devise three solutions for managing the TALs in $5G$ network, which are: (a) F-PAGING for the solution that favors the paging overheads than the TAU signaling; (b) F-TAU for the solution that favors the TAU than the paging overheads. (c) FOTA (i.e., Fair and Optimal TAL Assignment) for the solution that uses bargaining game to ensure a fair tradeoff between TAU and paging overheads.

The reminder of this paper is organized as follows. Section II introduces some related research work. Section III presents the envisioned network model and formulates the target problem. It also describes the model we envision to reduce the signaling overhead of TAU and paging messages. The proposed solutions are presented in section IV. Section V presents the simulation setup to evaluate the performance of proposed schemes and discusses the obtained results. Finally, the paper concludes in Section VI.

II. RELATED WORK

Mitigating signaling overhead, due to UE mobility in cellular mobile networks, has attracted high attention during the last years. To further alleviate the effect of TAU messages on the network performance, 3GPP has introduced the concept of TAL in Long Term Evolution (LTE), wherein different TALs can be assigned to UEs in the same cell (eNodeB) [1, 2]. Since TALs are overlapped, the number of UEs performing TAU when crossing TA border drastically decreases. Besides reducing the number of TAU messages, TAL prevents the ping-pong effect, i.e., frequent TAU messages when a UE keeps hopping between adjacent TAs. Nevertheless, the current LTE specifications do not provide any details on how to define

TALs and allocate them to UEs. To address this open issue, many works have been proposed.

In [3], the authors proposed a solution that organizes cells into rings, where UEs in each ring use the same TAL. Solutions, proposed in [4] and [5], use the same concept as in [3] by assigning the same TAL to different UEs when visiting a cell in the network. However, all these solutions [3, 4, 5] have not fully explored the advantage of TAL against the conventional TA approach. In [2] and [6], the authors overcome this limitation by allowing UEs residing in the same cell to register with different TALs. Indeed, in [2] they proposed a solution for congestion mitigation along a railway path. On the other hand, in [6] an extension of the former work is proposed with two new aspects: i) the solution is generalized for any arbitrary network instead of only train scenarios; ii) the authors propose a new solution that handles the extenuation of paging signaling messages via TAL management.

To the best knowledge of the authors, most existing solutions focus only on the TAU overhead and ignore the paging overhead. The only research work that addressed both constraints is presented in [6], wherein the authors proposed two separate solutions, addressing the impact of TAU and paging overhead, respectively. Both solutions are based on multi-objectives optimization techniques. The first one tries to minimize the TAU overhead while setting paging as a constraint, and the second one minimizes the paging overhead while fixing the TAU overhead as a constraint. Furthermore, the proposed schemes differ from the existing solutions in the way they cope with the problem. Indeed, most existing solutions assign the same TAL i) to the same TAs in a static manner [3, 4, 5]; or ii) with the same probability [2, 6]. In contrast, as it will be detailed later, the proposed schemes dynamcially assign the same TAL to different TAs and that is at different probabilities.

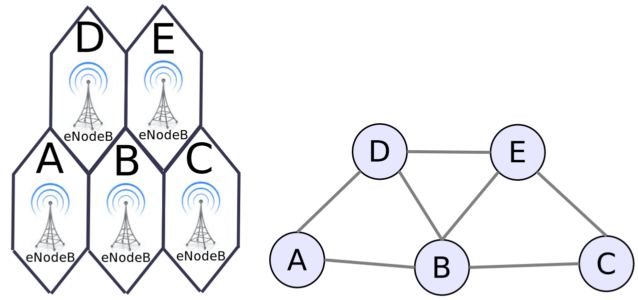
III. Envisioned network model and problem formulation

We assume that the network is subdivided into N TAs, named $\mathcal{N} = \{1, 2, \cdots N\}$. Each TA consists of a set of cells, whereby a cell is managed by an eNodeB (i.e., base station). For the sake of readability, the notations used throughout the paper are summarized in Table I.

TABLE I: Notations used in the paper.

Notation	Description
UE	User equipment.
$\mathcal{N}$	The set of TAs in the network.
η_u	The number of cells (eNodeB) in TA u.
α_u	The probability that UE u gets paged during a period D.
Γ	The set of all possible TAL in the network.
h_{uv}	The number of handover between TA u and v.

Let Γ denote the set of all possible TALs in a mobile network, and $\Gamma(A)$ be the set of possible TALs that can be assigned to TA A. Throughout this subsection, we will refer to the example depicted in Fig. 1 in order to show how Γ could be constructed. In this example, we assume that the network consists of five TAs. To form Γ, we begin by forming the



(a) (b)

Fig. 1: An example illustrating how to construct neighboring graphs G from an LTE network.

neighboring graphs G from the network as depicted in Fig. 1(b). An edge between two vertices (i.e., TA) A and B exists, if, and only if, there is a TAU possibility between them, i.e., a possibility for a UE to move from a TA to another. In the figure, there is no link between A and E since UEs cannot move from A to E without passing by another TA (i.e., B or D). Finally, $\Gamma(A)$ is formed from the neighboring graph G. Indeed, the different elements of $\Gamma(A)$ are those having all vertices of all sub-graphs of G that contain the vertex A. Thus, the vertices of a sub-graph of G that contain the vertex A are considered as one element in $\Gamma(A)$. From Fig. 1, $\Gamma(A) = \{\{A\}, \{A, B\}, \{A, D\}, \{A, B, C\}, \{A, B, D\}, \{A, B, E\}, \{A, D, E\}, \{A, B, C, D\}, \{A, B, C, E\}, \{A, B, D, E\}, \{A, B, C, D, E\}\}$. Finally, Γ is formed from different $\Gamma(i)$ as follows: $\Gamma = \bigcup_{i \in \mathcal{N}} \Gamma(i)$. Note that $\Gamma(i)$ is constituted with unordered elements; hence $\{A, B\}$ and $\{B, A\}$ are considered as the same element in Γ. From Fig. 1, $\Gamma = \{\{A\}, \{B\}, \{C\}, \{D\}, \{E\}, \{A, B\}, \{A, D\}, \{B, C\}, \{B, D\}, \{B, E\}, \{C, E\}, \{D, E\}, \{A, B, C\}, \{A, B, D\}, \{A, B, E\}, \{A, D, E\}, \{B, C, D\}, \{B, C, E\}, \{C, D, E\}, \{A, B, C, D\}, \{A, B, C, E\}, \{A, B, D, E\}, \{A, B, C, D, E\}\}$.

We assume that there are L *UEs* in the network, each of which has a specific probability to be paged (i.e., for voice call as well as for IP-based applications). We get these information by monitoring the network for a period D before starting the execution of our algorithms. We denote by $\alpha = \{\alpha_1, \alpha_2 \cdots \alpha_L\}$ the probability of paging of UEs in the network. In other words, a UE_i has a probability α_i to be called (i.e., cause a paging). Moreover, during the period D, we can deduce the number of UEs $h_{i,j}$ that moved from each TA i to another TA j. We define by $\mathcal{H}$ the matrix that represents the number of UEs that moved from different TAs. Each entry in the matrix $\mathcal{H}$ at row i and column j, denoted by $h_{i,j}$, indicates the number of UEs that moved from TA i to TA j. The value of $h_{i,j}$ can be deduced from the handover statistics of different eNodeBs or from the handover command messages sent by MME. Furthermore, each UE_i spends different times in different TAs. Let $\mathcal{M}$ denote the matrix that represents the duration spent by different *UEs* in different TAs. The lines in $\mathcal{M}$ represents the *UEs*, whereas the columns represent the different TAs in the network. The element $\mathcal{M}_{i,j}$ denotes the duration spent by UE_i in TA j.

IV. Optimal strategy for mapping between TALs and TAs

The aim of proposed solutions is to form a matrix $\mathcal{S}$, whereby the lines represent the different TAs $\mathcal{N}$ and the columns represent the TALs Γ. An element $\mathcal{S}_{i,j}$ in the matrix $\mathcal{S}$ refers to the probability that TA i assigns TAL j to different UEs. The sites (cells) belonging to the same TA i use the same line i in the matrix $\mathcal{S}$ to assign TALs to different UEs. In what follows, we present three problem formulations for optimizing TALs distribution in LTE networks. It shall be noted that the result of the three solutions is the same matrix S, however, with different elements $S_{i,j}$. In the first optimization problem, we assume that the TAU overhead is dominator and we then propose a solution to optimize the network performance that favors TAU on paging. In the second solution, we propose an optimization problem whereby the paging overhead is dominator. Finally, we introduce FOTA, which aims at capturing the trade-off between the TAU and paging overheads when assigning TALs to TAs (Fairness and Optimal Assigning of TALs to TAs), and ultimately to UEs. In FOTA, a bargaining game is used to capture the trade-off between TAU and paging.

A. Optimizing the network performance via the reduction of TAU overhead

In this solution, named F-TAU, we seek the optimal distribution of TALs by applying the min-max approach. The aim is to minimize the maximum number of TAU messages. We denote by $f(\mathcal{S})$ the function that we aim optimizing for the matrix $\mathcal{S}$. $f(\mathcal{S})$ can be formally defined as the maximum aggregate number of TAU messages sent by UEs between any two TAs in the network. In this solution, we denote by $PAGING_{max}$ the maximum number of paging messages tolerated by the network. Its value could be fixed according to the capacity of the system. Otherwise, $PAGING_{max}$ can be fixed to ∞. In this case, the optimal solution would converge to putting all TAs into the same TAL in order to reduce the TAU overhead. At this point, the optimization model which aims at reducing the TAU overhead can be formulated according to the following linear program (1):

$$\begin{cases} \textbf{min } f(\mathcal{S}) \\ \textbf{s. t.} \\ \forall i \in \mathcal{N}, \forall l \in \Gamma, \mathcal{S}_{il} \geq 0 \\ \forall i \in \mathcal{N}, \forall l \in \Gamma, \mathcal{S}_{il} \leq 1 \\ \forall i \in \mathcal{N}, \sum\limits_{l \in \Gamma} \mathcal{S}_{il} = 1 \\ \forall i, j \in \mathcal{N} \wedge i \neq j, \\ \quad c^u \times (\sum\limits_{\ell \in \Gamma(i) \wedge \ell \notin \Gamma(j)} h_{ij} \times S_{i\ell} + \sum\limits_{\ell \in \Gamma(j) \wedge \ell \notin \Gamma(i)} h_{ji} \times S_{j\ell}) \leq f(\mathcal{S}) \\ c^p \times \sum\limits_{\ell \in \Gamma} \sum\limits_{i \in \ell} ((\sum\limits_{k=1}^{L} \alpha_k M_{ki}) \times \sum\limits_{j \in \ell \wedge j \neq i} \eta_j \times \mathcal{S}_{i\ell}) \leq PAGING_{max} \end{cases} \tag{1}$$

The objective is to minimize $f(\mathcal{S})$ to reduce the TAU overhead. The first three constraints are used to ensure that each TA $i \in \mathcal{N}$ can select its TAL from $\mathcal{S}_i$ with a probability (i.e., a value between 0 and 1). The fourth constraint ensures that the TAU overhead between any two TAs i and j ($i \in \mathcal{N}$ and $j \in \mathcal{N}$) should not exceed $f(\mathcal{S})$. The number of UEs that transited from TA i (resp., j) are scaled by the variable $\mathcal{S}_{i\ell}$ (resp., $\mathcal{S}_{j\ell}$), which represents the proportional use of TAL ℓ by TAL i (resp, j). It shall be also noted that by the conditions,"$\ell \in \Gamma(i) \wedge \ell \notin \Gamma(j) \Leftrightarrow \forall i \in \mathcal{N}, \forall j \in \mathcal{N}, i \neq j, \forall \ell \in \Gamma : i \in \ell \wedge j \notin \ell$," our aim is to reduce the number of UEs moving between different TAs that do not belong to the same TALs. The last constraint ensures that the sum of all paging overhead in the network should not exceed a predefined threshold $PAGING_{max}$. For any TAL ℓ, the overhead caused by paging UEs, residing in TA $i \in \ell$, by sending paging messages to all TAs $j \in \ell \wedge j \neq i$ is $\sum\limits_{k=1}^{L} \alpha_k M_{ki}$, scaled by the number of sites η_j and a variable $\mathcal{S}_{i\ell}$ that represents the proportional use of ℓ. $\sum\limits_{k=1}^{L} \alpha_k M_{ki}$ is a constant that represents the paging overhead at TA i. Formally, $\sum\limits_{k=1}^{L} \alpha_k M_{ki}$ is defined as the sum of the probabilities of paging of each UE k scaled by its residence time in TA i.

B. Optimizing the network performance via the reduction of paging overhead

F-PAGING solution is modeled through linear program 2. The goal of F-PAGING is to optimize the network performance seeking the optimal distribution of TALs that minimizes the paging overhead. Similar to the previous solution, we apply the min-max approach. We define by $g(\mathcal{S})$ the function that we aim at optimizing for the matrix $\mathcal{S}$. $g(\mathcal{S})$ can be formally defined as the maximum paging overhead of the network. In this solution, we set the maximum amount of TAU overhead tolerated by the network to TAU_{max}. Its value could be defined according to the capacity of UEs (and also MME). Otherwise, TAU_{max} can be fixed to ∞. In this case, the optimal solution would converge to putting all TAs in a unique TAL in order to reduce the paging overhead. The linear program is formulated as follows:

$$\begin{cases} \textbf{min } g(\mathcal{S}) \\ \textbf{s. t.} \\ \forall i \in \mathcal{N}, \forall l \in \Gamma, \mathcal{S}_{il} \geq 0 \\ \forall i \in \mathcal{N}, \forall l \in \Gamma, \mathcal{S}_{il} \leq 1 \\ \forall i \in \mathcal{N}, \sum\limits_{l \in \Gamma} \mathcal{S}_{il} = 1 \\ \forall i, j \in \mathcal{N} \wedge i \neq j, \\ \quad c^u \times (\sum\limits_{\ell \in \Gamma(i) \wedge \ell \notin \Gamma(j)} h_{ij} \times S_{i\ell} + \sum\limits_{\ell \in \Gamma(j) \wedge \ell \notin \Gamma(i)} h_{ji} \times S_{j\ell}) \leq TAU_{max} \\ c^p \times \sum\limits_{\ell \in \Gamma} \sum\limits_{i \in \ell} ((\sum\limits_{k=1}^{L} \alpha_k M_{ki}) \times \sum\limits_{j \in \ell \wedge j \neq i} \eta_j \times \mathcal{S}_{i\ell}) \leq g(\mathcal{S}) \end{cases} \tag{2}$$

The objective is to minimize $g(\mathcal{S})$ to reduce the paging overhead. The first three constraints are similar to the first linear program presented in the precedent section. The fourth constraint ensures that the total number of TAU messages sent by UEs when transiting between any two adjacent TAs $i \in \mathcal{N}$ and $j \in \mathcal{N}$ should not exceed the threshold TAU_{max}. The last constraint ensures that the paging overhead in the network should not exceed $g(\mathcal{S})$.

C. Trading off TAU against paging using Nash bargaining

1) Nash bargaining model and threat value game: Nash bargaining model can be viewed as a game between two players who would like to barter goods. This model is a cooperative game with non-transferable utility. This means that the utility scales of the players are measured in non-comparable units. This model is adopted in our proposed FOTA scheme to find a Pareto efficiency between the paging and TAU overheads. In our case, the players are the paging and TAU overheads which do not use the same unity. This model is based on two elements, assumed to be given and known to the players. First, the set of vector payoffs $\mathcal{P}$ achieved by the players if they agree to cooperate. $\mathcal{P}$ should be a convex and compact set. Formally, $\mathcal{P}$ can be defined as $\mathcal{P} = \{(u(x), v(x)), x = (x_1, x_2) \in X\}$, whereby X is the set of strategies of two players, and $u()$ and $v()$ are the utility functions of the first and the second users, respectively. Second, the threat point, $d = (u^*, v^*) = (u((t_1, t_2)), v(t_1, t_2)) \in \mathcal{P}$, that represents the pair of utility whereby the two players fail to achieve an agreement. In Nash bargaining game, we aim to find a fair and reasonable point, $(\overline{u}, \overline{v}) = f(\mathcal{P}, u^*, v^*) \in \mathcal{P}$ for an arbitrary compact convex set $\mathcal{P}$ and point $(u^*, v^*) \in \mathcal{P}$. Based on Nash theory, a set of axioms are defined that lead to $f(\mathcal{P}, u^*, v^*)$ in order to achieve a unique optimal solution $(\overline{u}, \overline{v})$:

1) **Feasibility:** $(\overline{u}, \overline{v}) \in \mathcal{P}$.
2) **Pareto Optimality:** There is no point $(u(x), v(x)) \in \mathcal{P}$ such that $u(x) \geq \overline{u}$ and $v(x) \geq \overline{v}$ except $(\overline{u}, \overline{v})$.
3) **Pareto Optimality:** If $\mathcal{P}$ is symmetric about the line $u(x) = v(x)$, and $u^* = v^*$, then $\overline{u} = \overline{v}$.
4) **Independence of irrelevant alternatives:** If T is a closed convex subset of $\mathcal{P}$, and if $(u^*, v^*) \in T$ and $(\overline{u}, \overline{v}) \in T$, then $f(\mathcal{P}, u^*, v^*) = (\overline{u}, \overline{v})$.
5) **Invariance under change of location and scale:** If $T = \{(u'(x), v'(x)), u'(x) = \gamma_1 u(x) + \beta_1, v'(x) = \gamma_2 v(x) + \beta_2 \; for \; (u(x), v(x)) \in \mathcal{P}\}$, where $\gamma_1 > 0, \gamma_2 > 0$, and B_1 and B_2 are given numbers, then $f(T, \gamma_1 u^* + \beta_1, \gamma_2 v^* + \beta_2) = (\gamma_1 \overline{u} + \beta_1, \gamma_2 \overline{v} + \beta_2)$.

Moreover, the unique solution $(\overline{u}, \overline{v})$, satisfying the above axioms, is proven to be the solution of the following optimization problem:

$$\begin{cases} \textbf{max } (u(x) - u^*)(v(x) - v^*) \\ \textbf{s. t.} \\ \quad (u(x), v(x)) \in S \\ \quad (u(x), v(x)) \geq (u^*, v^*) \end{cases} \tag{3}$$

2) FOTA: Fair and Optimal TAL Assignment: We denote by $d = (TAU_{worst}, PAGING_{worst})$ the threat point of our bargaining game that resolves FOTA. In contrast to traditional bargaining game, the utility function of each player, (i.e., TAU and paging overhead) in our model, is the opposite of its cost. In other words, $(TAU_{worst}, PAGING_{worst}) \geq (f(\mathcal{S}), g(\mathcal{S})), \forall \mathcal{S} \in X$. The trade-off problem between TAU and paging overheads can be modeled as a convex optimization problem 4.

$$\begin{cases} \textbf{max } (TAU_{worst} - f(\mathcal{S}))(PAGING_{worst} - g(\mathcal{S})) \\ \textbf{s. t.} \\ \forall i \in \mathcal{N}, \forall l \in \Gamma, \mathcal{S}_{il} \geq 0 \\ \forall i \in \mathcal{N}, \forall l \in \Gamma, \mathcal{S}_{il} \leq 1 \\ \forall i \in \mathcal{N}, \sum_{l \in \Gamma} \mathcal{S}_{il} = 1 \\ \forall i, j \in \mathcal{N} \wedge i \neq j, \\ \quad c^u \times (\sum_{\ell \in \Gamma(i) \wedge \ell \notin \Gamma(j)} h_{ij} \times S_{i\ell} + \sum_{\ell \in \Gamma(j) \wedge \ell \notin \Gamma(i)} h_{ji} \times S_{j\ell}) \leq f(\mathcal{S}) \\ c^p \times \sum_{\ell \in \Gamma} \sum_{i \in \ell} ((\sum_{k=1}^{L} \alpha_k M_{ki}) \times \sum_{j \in \ell \wedge j \neq i} \eta_j \times \mathcal{S}_{i\ell}) \leq g(\mathcal{S}) \\ f(\mathcal{S}) \leq TAU_{worst} \\ g(\mathcal{S}) \leq PAGING_{worst} \end{cases} \tag{4}$$

Let $\mathcal{S}^{TAU}$ and $\mathcal{S}^{PAGING}$ be the optimal solutions of the linear programs 1 and 2, respectively. Then, we can define $PAGING_{worst}$, $PAGING_{best}$, TAU_{worst} and TAU_{best} as follows:

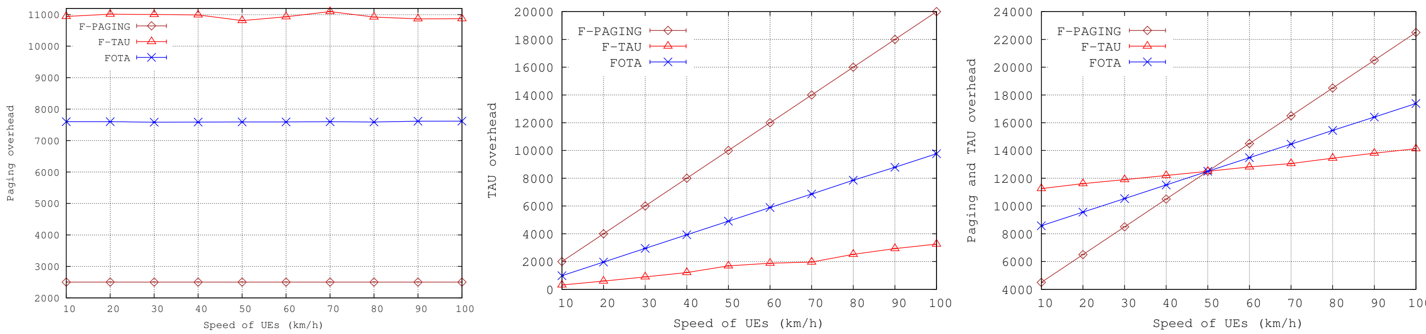
1) $PAGING_{worst} = c^p \times \sum_{\ell \in \Gamma} \sum_{i \in \ell} ((\sum_{k=1}^{L} \alpha_k M_{ki}) \times \sum_{j \in \ell \wedge j \neq i} \eta_j \times \mathcal{S}_{i\ell}^{TAU})$
2) $PAGING_{best} = c^p \times \sum_{\ell \in \Gamma} \sum_{i \in \ell} ((\sum_{k=1}^{L} \alpha_k M_{ki}) \times \sum_{j \in \ell \wedge j \neq i} \eta_j \times \mathcal{S}_{i\ell}^{PAGING})$
3) $TAU_{worst} = \max_{\forall i, j \in \mathcal{N}, i \neq j} (c^u \times (\sum_{\ell \in \Gamma(i) \wedge \ell \notin \Gamma(j)} h_{ij} \times S_{i\ell} + \sum_{\ell \in \Gamma(j) \wedge \ell \notin \Gamma(i)} h_{ji} \times S_{j\ell}^{PAGING}))$
4) $TAU_{best} = \max_{\forall i, j \in \mathcal{N}, i \neq j} (c^u \times (\sum_{\ell \in \Gamma(i) \wedge \ell \notin \Gamma(j)} h_{ij} \times S_{i\ell} + \sum_{\ell \in \Gamma(j) \wedge \ell \notin \Gamma(i)} h_{ji} \times S_{j\ell}^{TAU}))$

It is easily noticeable that $PAGING_{best} \leq PAGING_{worst}$ and $TAU_{best} \leq TAU_{worst}$. The values of $PAGING_{best}$, $PAGING_{worst}$, TAU_{best} and TAU_{worst} are obtained by updating the linear programs 1 and 2 as follows:

$$\begin{cases} \textbf{min } f(\mathcal{S}) \\ \textbf{s. t.} \\ \forall i \in \mathcal{N}, \forall l \in \Gamma, \mathcal{S}_{il} \geq 0 \\ \forall i \in \mathcal{N}, \forall l \in \Gamma, \mathcal{S}_{il} \leq 1 \\ \forall i \in \mathcal{N}, \sum_{l \in \Gamma} \mathcal{S}_{il} = 1 \\ \forall i, j \in \mathcal{N} \wedge i \neq j, \\ \quad c^u \times (\sum_{\ell \in \Gamma(i) \wedge \ell \notin \Gamma(j)} h_{ij} \times S_{i\ell} + \sum_{\ell \in \Gamma(j) \wedge \ell \notin \Gamma(i)} h_{ji} \times S_{j\ell}) \leq TAU_{best} \\ \quad TAU_{best} \leq f(\mathcal{S}) \\ c^p \times \sum_{\ell \in \Gamma} \sum_{i \in \ell} ((\sum_{k=1}^{L} \alpha_k M_{ki}) \times \sum_{j \in \ell \wedge j \neq i} \eta_j \times \mathcal{S}_{i\ell}) \leq PAGING_{worst} \\ \quad PAGING_{worst} \leq PAGING_{max} \end{cases} \tag{5}$$

$$\begin{cases} \textbf{min } g(\mathcal{S}) \\ \textbf{s. t.} \\ \forall i \in \mathcal{N}, \forall l \in \Gamma, \mathcal{S}_{il} \geq 0 \\ \forall i \in \mathcal{N}, \forall l \in \Gamma, \mathcal{S}_{il} \leq 1 \\ \forall i \in \mathcal{N}, \sum_{l \in \Gamma} \mathcal{S}_{il} = 1 \\ \forall i, j \in \mathcal{N} \wedge i \neq j, \\ \quad c^u \times (\sum_{\ell \in \Gamma(i) \wedge \ell \notin \Gamma(j)} h_{ij} \times S_{i\ell} + \sum_{\ell \in \Gamma(j) \wedge \ell \notin \Gamma(i)} h_{ji} \times S_{j\ell}) \leq TAU_{worst} \\ \quad TAU_{worst} \leq TAU_{max} \\ c^p \times \sum_{\ell \in \Gamma} \sum_{i \in \ell} ((\sum_{k=1}^{L} \alpha_k M_{ki}) \times \sum_{j \in \ell \wedge j \neq i} \eta_j \times \mathcal{S}_{i\ell}) \leq PAGING_{best} \\ PAGING_{best} \leq g(\mathcal{S}) \end{cases} \tag{6}$$

The optimization problem shown in the linear program 4 is non-convex. Using the approach proposed in [7], the



(a) paging overhead vs speed of UEs (b) TAU overhead vs speed of UEs (c) paging and TAU overheads vs speed of UEs

Fig. 2: Comparison of performances of FOTA as a function of speed of UEs

problem can be transformed to convex-optimization problem without changing the solution. The key idea is to introduce the log function which is an increasing function. Therefore, the optimization problem is reformulated as follows:

$$\begin{cases} \textbf{max} \ \log((TAU_{worst} - f(\mathcal{S}))) + \log((PAGING_{worst} - g(\mathcal{S}))) \\ \textbf{s. t.} \\ \forall i \in \mathcal{N}, \forall l \in \Gamma, \mathcal{S}_{il} \geq 0 \\ \forall i \in \mathcal{N}, \forall l \in \Gamma, \mathcal{S}_{il} \leq 1 \\ \forall i \in \mathcal{N}, \sum_{l \in \Gamma} \mathcal{S}_{il} = 1 \\ \forall i, j \in \mathcal{N} \wedge i \neq j, \\ \quad c^u \times \Big(\sum_{\ell \in \Gamma(i) \wedge \ell \notin \Gamma(j)} h_{ij} \times S_{i\ell} + \sum_{\ell \in \Gamma(j) \wedge \ell \notin \Gamma(i)} h_{ji} \times S_{j\ell} \Big) \leq f(\mathcal{S}) \\ c^p \times \sum_{\ell \in \Gamma} \sum_{i \in \ell} \Big(\big(\sum_{k=1}^{L} \alpha_k M_{ki} \big) \times \sum_{j \in \ell \wedge j \neq i} \eta_j \times \mathcal{S}_{i\ell} \Big) \leq g(\mathcal{S}) \\ f(\mathcal{S}) \leq TAU_{worst} \\ g(\mathcal{S}) \leq PAGING_{worst} \end{cases} \tag{7}$$

V. Performance evaluation

In the performance evaluation, we fix the overhead of a single TAU, c^u, to be ten times the value of c^p [8]. All solutions (i.e. FOTA, F-TAU and F-PAGING) are evaluated in terms of the following metrics:

1) *TAU overhead:* the number of TAU messages (UP-Link) generated by UEs when visiting new TALs.
2) *Paging overhead:* the number of paging packets sent from MME to find a UE's location to establish a connection.
3) *Total overhead:* the generated overhead due to both paging and TAU. The aim of this metric is to show the Pareto-efficiency between the TAU and paging overheads.

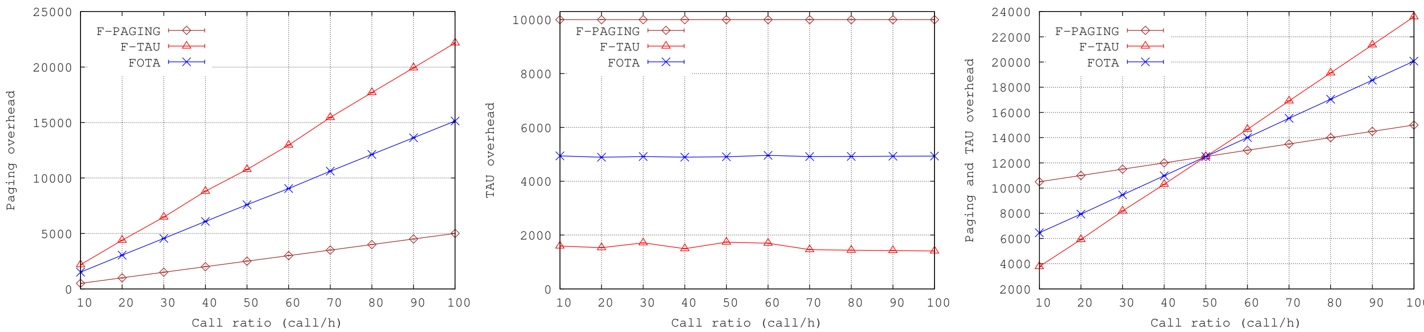
Accordingly, we have developed a simulator tool based on Matlab and CVX (a package for disciplined convex optimization and geometric programming [9]). In our simulation, the sites (i.e., eNodeBs) are randomly deployed over the simulated network. The optimization problems are solved considering different values of the maximum speed of UEs and the maximum ratio of calls of each UE in the network. The latter refers to the percentage of time that a UE is called during the simulation time. Here, the "call" term refers not only to the classical voice call but also to data connection, such as VoIP and emails. This parameter allows us to model the user activity in terms of active connections. Whereas, the maximum speed of UEs shows the impact of TAUs signaling on the different optimization problems.

We simulated two scenarios: (i) we vary the maximum speed of UEs and fix the maximum call ratio to $50 calls/h$ for each UE in the network; (ii) we vary the maximum call ratio of UEs and fix the maximum speed of UEs to $50 km/h$. For each scenario, we carry out the following three steps:

1) Gathering information on the signaling pattern from the network, by observing the network (i.e., UEs and MME) activities during one hour of time. In this step, the observed information concern the overhead of TAU and paging for each TAL.
2) Solving the optimization problems 5, 6, and 7 in order to obtain the matrix S for each solution.
3) Simulating UEs' activities for few hours. To get a fair comparison, the UEs use the same trajectory, same speed and same paging ratio in all solutions.

Figs. 2 and 3 show the resilience of FOTA, F-TAU and F-PAGING against increase in UEs' speed and call ratio, respectively. Figs. $2(b)$ and $2(c)$ show that the speed of UEs has a negative impact on TAU and total overheads, respectively. This behavior is expected as highly mobile users perform frequently handoffs between cells and ultimately generate high TAU messages. Thus, the higher the speed of UEs is, the higher the TAU overhead becomes. Further, we remark from Fig. $2(b)$ that F-TAU exhibits better performance than FOTA and F-PAGING in terms of TAU overhead regardless the speed of UEs. This is attributable to the fact that the key objective of F-TAU is to minimize TAU overhead without tacking into account the paging overhead.

Figs. $3(a)$ and $3(c)$ demonstrate that the call ratio has a negative impact on paging and total overheads, respectively. This is also predictable as highly active UEs (i.e., with high call ratios) cause high number of paging messages when they go in the idle mode and their locations are searched by the network. Moreover, from Fig. $3(a)$, we observe that F-PAGING exhibits better performance than FOTA and F-TAU in terms of paging overhead regardless the call ratio. This is intuitively due to the fact that F-PAGING is designed to optimize the paging overhead without tacking into account the TAU overhead.



(a) Paging overhead vs call ratio.
(b) TAU overhead vs call ratio.
(c) Paging and TAU overheads vs call ratio.

Fig. 3: Comparison of FOTA's performance for different values of the call ratio.

Figs. 2(c) and 3(c) illustrate the tradeoff achieved by FOTA between the two conflicting objectives, i.e; reduction of both TAU and paging overhead. They show the total overhead incurred in the three solutions and that is for different values of the UE speed and call ratio, respectively. We observe from these figures that: (i) F-PAGING exhibits better performance in terms of total (i.e., paging and TAU) overhead when the speed of UEs is below $50km/h$ or when the call ratio exceeds $50calls/h$; (ii) F-TAU exhibits better performance when the maximum speed of UEs exceeds $50km/h$ or when the call ratio does not exceed $50calls/h$; and (iii) FOTA has performance similar to that of F-PAGING when the speed of UEs is below $50km/h$ or when the call ratio exceeds $50calls/h$. It is also observed that FOTA performs similarly to F-TAU when the call ratio does not exceed $50calls/h$ or the speed of UEs exceeds $50km/h$. Indeed, the performance of FOTA is always between F-TAU and F-PAGING, depending on the UEs' speed and their activity levels (i.e., call rate). For highly mobile UEs, FOTA performs similar to F-TAU (optimal) and better than F-PAGING, whilst for highly active UEs, FOTA performs similar to F-PAGING (optimal) and better than F-TAU. FOTA always finds an optimal tradeoff between TAU and paging overhead by maintaining the total overhead near to the optimal value.

VI. Conclusion

One key vision of the upcoming 5G is to support potential numbers of users connecting to the mobile networks. An important challenge is to cope with the amount of signaling to be generated by these mobile users, particularly signaling messages due to mobility (i.e., TAU) and for connection setup (i.e., paging). In this paper, we have devised three solutions to mitigate the effect of TAU and paging signaling messages on the network. The first solution, named F-TAU, favors the TAU overhead than the paging overhead, whereas the second one, F-PAGING, favors the paging overhead than the TAU overhead. The third one, named FOTA, uses a bargaining Nash theory to find a fair tradeoff between minimizing both TAU and paging messages. Simulation results proved the efficiency of each solution in achieving its design goal. From the obtained results, it can be concluded that FOTA has the ability to adapt to any network configuration; an important feature mainly that it is all but difficult to accurately predict in advance the UEs' activities in terms of mobility and call rate.

References

[1] Mitsubishi Electric, "Tracking areas size tracking area list optimization technical report 3GPP TSG RAN WG3 meeting, , r3-071931, 2007." Tech. Rep.

[2] S. Modarres Razavi and D. Yuan, "Mitigating mobility signaling congestion in LTE by overlapping tracking area lists," in *ACM MSWIM'11*, 2011.

[3] Y. W. Chung, "Adaptive design of tracking area list in lte," in *IEEE WOCN'11*, May 2011, pp. 1–5.

[4] S. Razavi, D. Yuan, F. Gunnarsson, and J. Moe, "Exploiting tracking area list for improving signaling overhead in LTE," in *IEEE VTC'10*, May 2010, pp. 1–5.

[5] S. Razavi, Y. Di, F. Gunnarsson, and J. Moe, "Dynamic tracking area list configuration and performance evaluation in LTE," in *IEEE GLOBECOM Workshops (GC Wkshps)*, December 2010, pp. 49–53.

[6] S. M. Razavi and D. Yuan, "Mitigating signaling congestion in LTE location management by overlapping tracking area lists," *Elsevier Computer Communications*, vol. 35, no. 18, pp. 2227 – 2235, 2012.

[7] Y. Zhao, S. Wang, S. Xu, X. Wang, X. Gao, and C. Qiao, "Load balance vs energy efficiency in traffic engineering: A game theoretical perspective," in *IEEE INFOCOM'13*, April 2013, pp. 530–534.

[8] K. Kyamakya and K. Jobmann, "Location management in cellular networks: classification of the most important paradigms, realistic simulation framework, and relative performance analysis," *IEEE Transactions on Vehicular Technology*, vol. 54, no. 2, 2005.

[9] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, ser. Lecture Notes in Control and Information Sciences, V. Blondel, S. Boyd, and H. Kimura, Eds. Springer-Verlag Limited, 2008, pp. 95–110.