Load Balancing in Downlink LTE Self-Optimizing Networks

Andreas Lobinger*, Szymon Stefanski†, Thomas Jansen‡, Irina Balan§

* Nokia Siemens Networks, München, Germany, andreas.lobinger@nsn.com
† Nokia Siemens Networks, Wrocław, Poland, szymon.stefanski@nsn.com
‡Technical University of Braunschweig, Braunschweig, Germany, Jansen@ifn.ing.tu-bs.de
§Interdisciplinary Institute for Broadband Technology, Ghent, Belgium, Irina.Balan@intec.ugent.be

Abstract—In this paper we present system level simulation results of a self-optimizing load balancing algorithm in a long-term-evolution (LTE) mobile communication system. Based on previous work [2][1], we evaluate the network performance of this algorithm that requires the load of a cell as input and controls the handover parameters. We compare the results for different simulation setups: for a basic, regular network setup, a non-regular grid with different cell sizes and also for a realistic scenario based on measurements and realistic traffic setup.

I. INTRODUCTION

In existing networks, parameters are manually adjusted to obtain a high level of network operational performance. In LTE the concept of self-optimizing networks (SON) is introduced, where the parameter tuning is done automatically based on measurements. A challenge is to deliver additional performance gain further improving network efficiency. The use of load-balancing (LB), which belongs to the group of suggested SON functions for LTE network operations, is meant to deliver this extra gain in terms of network performance. For LB this is achieved by adjusting the network control parameters in such a way that overloaded cells can offload the excess traffic to low-loaded adjacent cells, whenever available. In a live network high load fluctuations occurs and they are usually accounted for by overdimensioning the network during planning phase. A SON enabled network, where the proposed SON algorithm monitors the network and reacts to these peaks in load, can achieve better performance by distributing the load among neighbouring cells [10]. The load balancing algorithm aims at finding the optimum handover (HO) offset value between the overloaded cell and a possible target cell. This optimised offset value will assure that the users that are handed over to the target cell will not be returned to the source cell and thus the load in the current cell is diminished. Simulations were conducted for synthetic regular hexagonal and non regular network layout as well as a realistic network scenario.

The work has been carried out in the EU FP7 SOCRATES project [1].

II. DEFINITIONS AND METRIC

In [2] a mathematical framework for SON investigations on the downlink is defined.

The main parts for this mathematical framework are recalled below:

- A network layout by network nodes, eNodeB(s) (eNB), defining a cell c at the position $\vec{p}_c$.
- A user u located at a position $\vec{q}_u$. In dynamic simulation the user positions can be evolve over simulation time.
- A connection function $X(u)$, following that a user u is served by cell $c = X(u)$ with the constraint (according to the definition of LTE) that every user is connected exactly to a single cell.
- A pathloss mapping L, $L_{X(u)}(\vec{q}_u)$ defined by the positions of user u relative to a cell c, $c = X(u)$. The pathloss mapping takes all position-dependent channel model effects into account, which are: distance-dependent pathloss, shadowing and angle-dependent antenna patterns.
- A cell load ρ_c, which defines the ratio of used resources in LTE physical resource blocks (PRBs) to all available resources in a given cell.

With the above framework and two additional terms i.e. N as thermal noise and P_c as transmit power for a cell, we can now define and evaluate for every user, in every time-step a user specific signal to noise and interference ratio, $SINR_u$.

$$SINR_u = \frac{P_c \cdot L_{X(u)}(\vec{q}_u)}{N + \sum_{c \neq X(u)} \rho_c \cdot P_c \cdot L_c(\vec{q}_u)}. \qquad (1)$$

The evaluation complexity is reduced by the selection of a single service definition for all users, a constant bit rate (CBR) service. The CBR assumption used further on is 512kBit/s.

A. Load and throughput mapping

We assume that the best modulation coding scheme (MCS) is used for a given SINR and the highest data rate $R(SINR)$ is achievable, this can be represented by Shannon formula.

$$R(SINR_u) = \log_2(1 + SINR_u) \qquad (2)$$

However for better approximation to realistic MCS, the mapping function is scaled by a factor 0.6 and is bounded by maximum available bitrate (4.4 bps/Hz) and minimum required SINR (-6.5 dB), a detailed description of which can be found in [4]. Based on the achievable throughput at a given SINR, the demanded data rate D and the transmission bandwidth BW of one PRB (PRB bandwidth in LTE system is 180kHz),
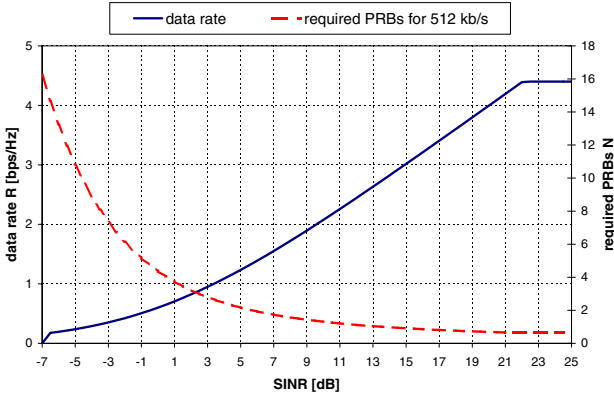
Fig. 1. Required PRBs for transmission of 512 kbps as a function of SINR.

the amount of required PRBs N can be obtained from the following:

$$N_u = \frac{D_u}{R(SINR_u) \cdot BW} \quad (3)$$

Figure 1 presents the relationship between SINR, throughput and required resources in DL transmission.

B. Virtual load and unsatisfied users metric

In this document we investigate scenarios with different load distributions including user u concentrations to overload the serving cell. The overload situation occurs when the total required number of PRBs N_u may exceed the amount of the total available resources in one cell M_{PRB}. Thus we introduce the virtual cell load, $\widehat{\rho_c}$ which can be expressed as the sum of the required resources of all users u connected to the cell c by the connection function $X(u)$ which gives the serving cell c for user u

$$\widehat{\rho_c} = \frac{1}{M_{PRB}} \cdot \sum_{u|X(u)=c} N_u. \quad (4)$$

This metric allows us to exceed 100% of cell load ($\widehat{\rho_c} > 1$) and such situations indicate a scale of overload in cell c. In overloaded cells part of the users can not be served with the required service quality and we call them as unsatisfied. The total number of unsatisfied users in the whole network (users being a sum of unsatisfied users in all cells, where number of users in cell c is represented by M_c) can be written as:

$$z = \sum_{\forall c} max\left(0, M_c \cdot \left(1 - \frac{1}{\widehat{\rho_c}}\right)\right) \quad (5)$$

Detailed explanation of the virtual load concept and unsatisfied users metrics can be found in [2].

C. Load estimation

Load balancing is achieved by handing over users from the overloaded cells to those adjacent cells which are able to accommodate additional load. After HO, the user's SINR condition at cell served by target eNB (TeNB) as well as generated load is different than in previous cell served by serving eNB (SeNB). The load transferred during LB operation should not exceed the capacity reported as available by the

TeNB. This issue should be controlled by admission and congestion control mechanisms in the TeNB. When the admission and congestion control mechanisms reject LB HO requests, there will be a significant increase in signaling overhead if the requests are repeated. We propose a prediction method to adress this for the load required at TeNB side, based on SINR estimation after LB HO, taking into account UE measurements like the reference signal received power (RSRP) and received signal strength indicator (RSSI). For simplification we do not consider additional factors related to the current load at SeNB and TeNB which may have an impact on SINR. Signal received by user u from cell c we can write as:

$$S_c = P_c \cdot L_c(\vec{q}_u) \quad (6)$$

Before LB HO the user u_1 is connected to the SeNB, $SeNB = X(u_1)$ with the strongest received RSRP signal (signal S_1 in Figure 2 a). The RSRP signal from TeNB (S_2 in Figure 2a) is a component of total interference as well as signals originated from other eNBs (represented by I in Figure 2a). After HO of user u_1 to TeNB, received signal S_1 from the previously serving SeNB now contributes to the interference signal at u_1 whereas signal S_2 from TeNB is the wanted signal (see Figure 2b).
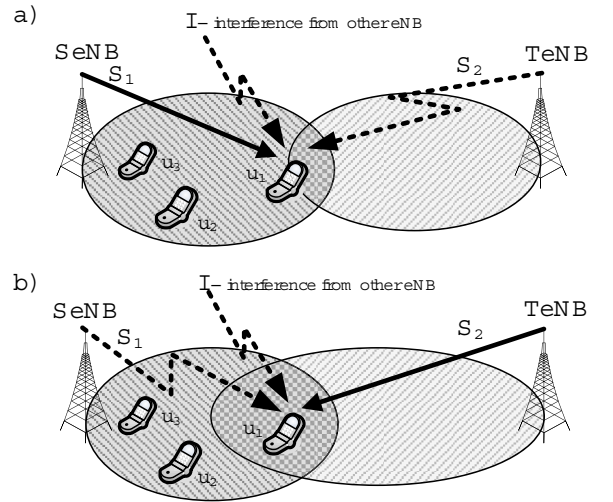

Fig. 2. Signals received by UE a) before HO, b) after HO

We assume that during the time of HO execution, the user's u_1 position $\vec{q}_{u_1}$ does not change significantly and hence we also assume no changes of received signal power by the user u_1. We can extract signals S_1 and S_2 from the interference part of $SINR_{SeNB}$ and $SINR_{TeNB}$ equations and after combining them, this relation can be written as follows:

$$SINR_{u_1,TeNB} = \frac{S_2}{\dfrac{S_1}{SINR_{u_1,SeNB}} + S_1 - S_2}. \quad (7)$$

III. Load Balancing algorithm

In this investigation we are using the parameter "cell specific offset" to force users to handover from the SeNB to a TeNB. It is important to set this parameter carefully, so as not to exceed the acceptable load at TeNB after the LB HO procedure. The main goal of the presented Algorithm 1 is to find the optimum HO offset that allows the maximum number of users to change cell without any rejections by admission control mechanism at TeNB side. Before applying the LB procedure, the SeNB needs to create a list of potential targets for HO, collect measurement reports from the served UEs as well as to collect the available resources reports from neighbouring cells. These preparation actions are included in steps 1 - 5 of Algorithm 1. For each adjusted values of HO offset T, SeNB sorts the list of the potential TeNB with respect to the number of possible LB HOs. Subsequently, for given HO offset T and cell C from the list L a load after HO, $\widehat{\rho}_c$, is estimated. If the predicted load $\widehat{\rho}_c$ does not exceed acceptable level ρ_{Thld} at TeNB, HO offset to this cell is adjusted to the T value and virtual load at SeNB ρ_{SeNB} is reduced by the amount generated by the users handed over with this offset. Algorithm works until load ρ_{SeNB} at SeNB is higher than accepted level $\rho_{Thld,SeNB}$ and HO offset is below the maximum alowed value and neighbour cells are able to accommodate additional load.

Algorithm 1 HO offset based LB algorithm

Require: List L of potential Target eNB (TeNB) for LB HO
1. collect measurements from user:; RSRP to potential TeNB,
2. group users corresponding to the best TeNB for LB HO (criterion is the difference between SeNB and TeNB measured signal quality),
3. obtain information from TeNB on available resources,
4. estimate a number of required PRBs after LB HO for each user in the LB HO group,
5. $T \Leftarrow 0$
6. **while** $\rho_{SeNB} > \rho_{Thld,SeNB} \wedge T < T_{max}$ **do**
7. $i \Leftarrow 1$
8. $T \Leftarrow T + step$
9. $L \Leftarrow sort$ (TeNB according to number of users allowed to LB HO with given T, descending order)
10.   **while** $\rho_{SeNB} > \rho_{Thld,SeNB} \wedge i \leq size(L)$ **do**
11. $C \Leftarrow L(i)$ {take next cell from list}
12.   estimate $\widehat{\rho}_c$ after HO for given T
13.   **if** $\widehat{\rho}_c < \rho_{Thld,C}$ **then**
14. $\rho_{SeNB} \Leftarrow \rho_{SeNB} - \rho_{SeNB,T}$ {update load in overloaded cell by substract handed over load}
15. $T_{C,u} \Leftarrow T$
16.   **end if**
17. $i \Leftarrow i + 1$
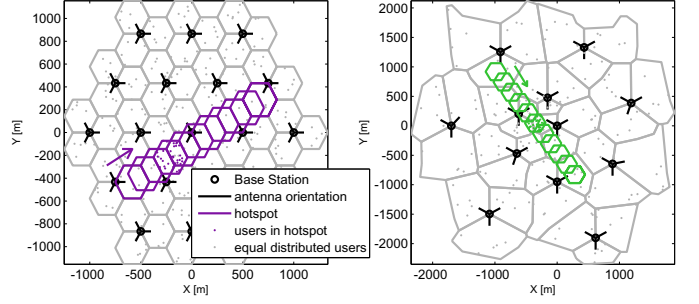18.   **end while**
19. **end while**
20. adjust HO offsets T_C

Fig. 3. Regular and non-regular simulation scenarios with hotspot path

IV. Simulation Scenarios

As already mentioned a standard LTE DL system of 10MHz bandwidth is used with the simulation assumptions in the LTE 3GPP definitions [7]. For both the synthetic and real scenario a simulation time-step of 500ms have been used. So all internal signals, evaluations and updates are carried out with averaging over this 500ms step. Also the LB algorithm is called in every time-step.

A. Synthetic scenario

Following the standard simulation assumptions a simulation setup with 19 sites in a regular hexagonal grid, 3 sectors per site and 57 cells are defined. Additionally, a non-regular grid with 12 sites, 3 sectors per site and 36 cells is used as comparison taking real network effects like different cell sizes, number of neighbor cells and interference situations into account.

For employing localized higher load in the system a simulation scenario is used here with a setup of background load in all cells with a low number of users -so they are satisfied in any position of the network- and an additional hotspot in which new, additional users are dropped over time. The hotspot area is moved over time on a path through the network as depicted in Figure 3.

The channel model is defined in closed form [7], so the pathloss mapping L is continuos. The movement of the users (whether background or dropped in the hotspot) is a constant velocity and random waypoint model.

B. Real scenario

The realistic reference scenario in the SOCRATES project is an LTE network based on the real layout of the existing 2G and 3G macro networks provided by a network operator (see [8] for more details). The scenario data includes the following information: Network configuration, Pathloss data, Clutter data, Heigth data, Traffic data, Mobility data.

An area of 72 km x 37 km is selected. The resulting network comprising of 103 sites and 309 cells is shown partly in Figure 4. For a smaller area of 1.5 km x 1.5 km within the realistic scenario the detailed microscopic mobility model SUMO (Simulation of Urban MObility) is used [11] to introduce relevant user mobility. The mobile users in this mobility model are cars and thus travel along streets. The
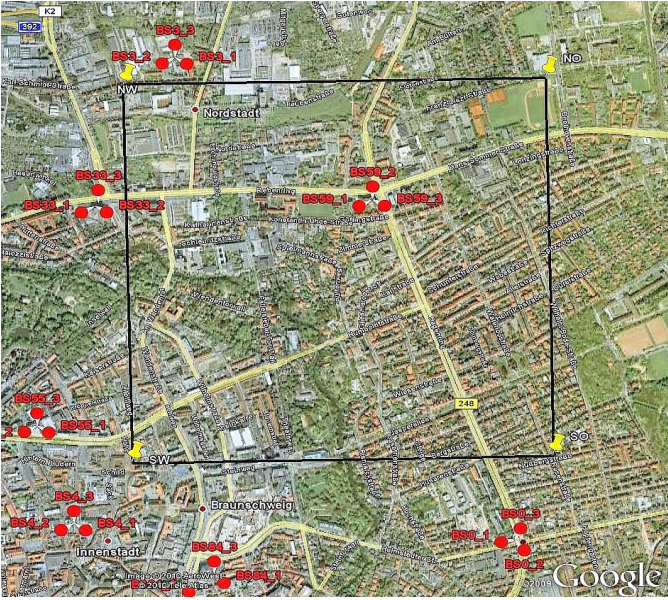
Fig. 4. Realistic Scenario, area with basestation positions

Fig. 5. Realistic Scenario, hotspot (bus) path

mobility model incorporates traffic lights, multiple lanes, user acceleration and braking, queueing in front of barricades and overtaking of slower cars. In order to generate an overload in the scenario a moving hotspot is defined by a bus with a set of users moving together along a road. The distribution of additional static users is based on traffic maps provided by the same operator.

Based on operator measurements and path predictions using a ray-tracer a pathloss mapping L is now available for background users (static + mobile) and the users in the bus (hotspot). For this pathloss mapping only the 30 most significant cells -including both the connected cell and the interfering cells- are taken into account.

In Figure 5 the path of the bus is shown in more detail. Gray shapes make an image of the city streets and magenta lines determine the theoretical cells marked by numbers. Base stations and antennas orientations are shown in the same way as on Figure 3. Bus route is tagged by the brown colour. Circle markers with the letters correspond to the points of HO of more than 50% of the users in bus. Starting from the cell #105; users in bus at point ⓐ are handovered to cell #104 it is writen as: ⓐ → 104. Following sequence describe the HO scheme of the bus users during the 10 min simulation of reference case: ⓐ → 104; ⓑ → 105; ⓒ → 103; ⓓ → 50; ⓔ → 103; ⓕ → 50; ⓖ → 99; ⓗ → 50; ⓘ → 15; ⓙ → 97; ⓚ → 15; ⓛ → 144.

V. RESULTS

A. Synthetic Scenarios

In Figures 6 and 7 timeline of the z-metric for each simulation scenario with a reference system (red curve) versus load-balancing (green curve) are depicted. The operating point here -in both scenarios- is 5 users per cell in background (equally dropped) and 40 users dropped in hotspot. The performance of
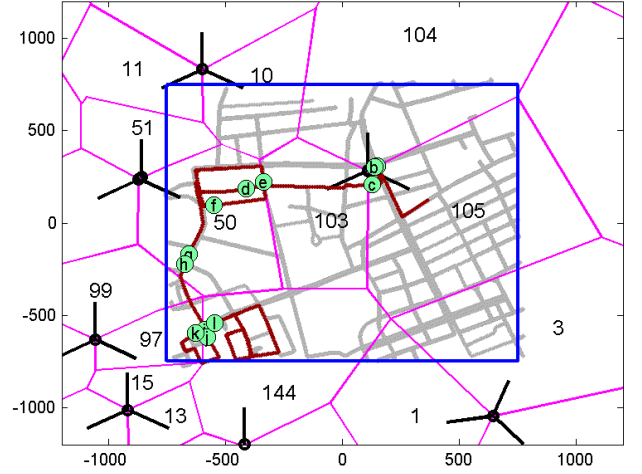
the load-balancing is -as expected- dependent on the hotspot position and on the users positions within the hotspot. The averaged unsatisfied users metric $\bar{z}$ is depicted as horizontal line. In table I some more operating points and the average results are shown as overview.
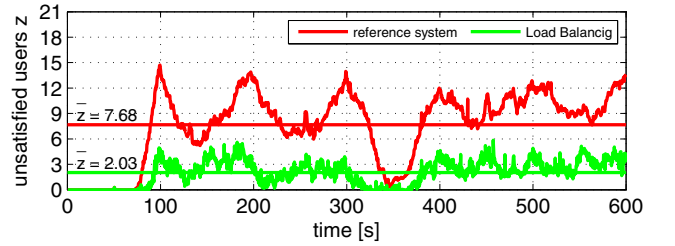

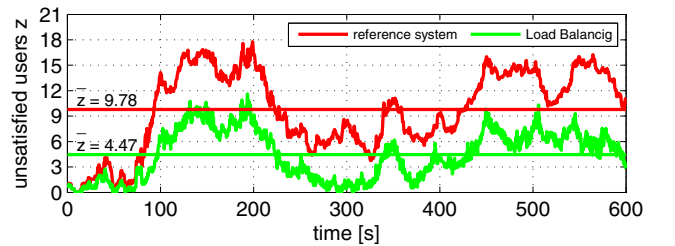Fig. 6. Load Balancing performance over time, regular network

Fig. 7. Load Balancing performance over time, non regular network

B. Real Scenario

In Figure 8 a timeline of the z-metric for the real simulation scenario with a reference system (red curve) versus load-balancing (green curve) is depicted. As described in IV-B part of the users move together on a bus along a street. Different than the operating point chosen for the synthetic scenarios here a significant background load with 80 static and 80 slow

TABLE I
SYNTHETIC SCENARIO, AVERAGE NUMBER OF UNSATISFIED USERS

scenario	$\bar{z}$ reference system		$\bar{z}$ load-balancing	
	regular	non-reg.	regular	non-reg.
users in hotspot				
30	2.8	3.2	0.3	0.7
40	7.7	9.8	2.0	4.5
50	13.6	16.8	6.4	10.6
60	22.0	23.1	15.5	17.4

moving users is used. In Figure 9 we can observe load (virtual load as defined in Equation 4) variations during the first 300s of the simulation. The markers in Figures 8 and 9 correspond to Figure 5 and the description in section IV-B. Comparing reference and load-balancing timelines we can see that the load-balancing algorithm can redistribute load to neighboring cells which significantly reduces overload in whole observed period. In table II more operating operating points and average results are shown as overview.
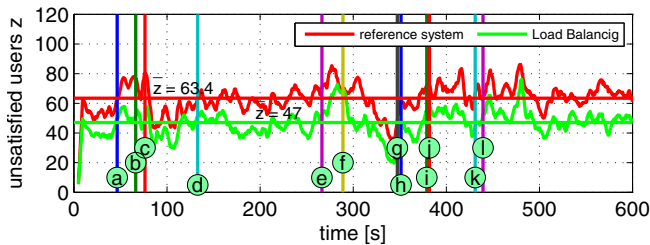

Fig. 8. Load Balancing performance over time, realistic scenario

TABLE II
REALISTIC SCENARIO, AVERAGE NUMBER OF UNSATISFIED USERS

users in bus	$\bar{z}$ reference system	$\bar{z}$ load-balancing
20	47.0	30.6
30	53.8	37.8
40	63.4	47.0
50	74.3	58.8
60	86.6	68.2

VI. CONCLUSION AND OUTLOOK

In this paper we describe a simulation campaign of a load-balancing algorithm. The algorithm evaluates the load condition in a cell and the neighbouring cells and estimates the impact of changing the HO parameters in order to improve the overall performance of the network. We propose a method for load estimation after HO would occur, which is based on SINR prediction and utilise UE measurements. The efficiency of LB algorithm was checked in simulation of synthetic networks layout as well as for a part of real network in which load situation changes dynamically.

An overall load-balancing gain is visible in all simulated scenarios in the provided timelines and tables; leading to the general conclusion that the average number of satisfied users can be improved with load-balancing. The possible gain depends on the local load situation and the available capacity
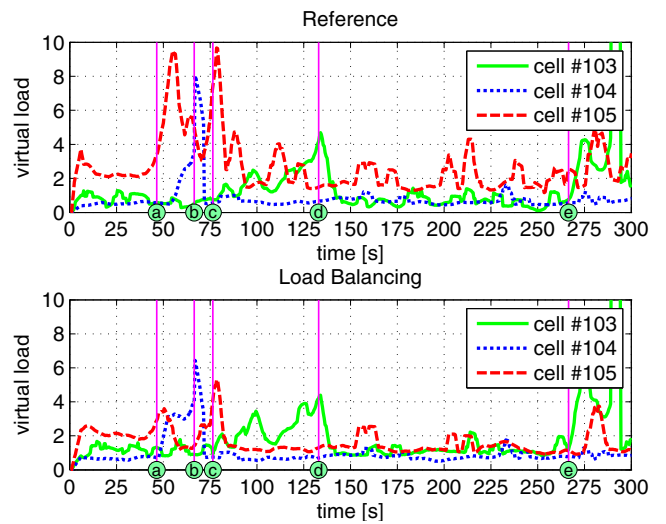

Fig. 9. Virtual load $\widehat{\rho}$ over time in seperate cells, realistic scenario

in the neighbouring cells. The algorithm relies on capacity available in the neighbouring cells around overloaded cells and the proposed load-balancing algorithm can redistribute the load by changing the HO parameters; If no capacity is available around, the network parameters and so the network performance are left unchanged.

REFERENCES

[1] SOCRATES, *Self-optimisation and self-configuration in wireless networks*, European Research Project, http://www.fp7-socrates.eu.

[2] I. Viering, M. Döttling, A. Lobinger, A mathematical perspective of self-optimizing wireless networks, *IEEE International Conference on Communications 2009 (ICC)*, Dresden, Germany, May 2009.

[3] 3GPP, *Self-configuring and self-optimizing network use cases and solutions*, Technical Report TR 36.902, available at http://www.3gpp.org.

[4] 3GPP, *3rd Generation Partnership Project;Technical Specification Group Radio Access Network;Evolved Universal Terrestrial Radio Access (E-UTRA);Radio Frequency (RF) system scenarios (Release 8)*, Technical Report TR 36.942, available at http://www.3gpp.org.

[5] M. Amirijoo, R. Litjens, K. Spaey, M. Döttling, T. Jansen, N. Scully, and U. Türke, "Use Cases, Requirements and Assessment Criteria for Future Self-Organising Radio Access Networks," *Proc. 3rd Intl. Workshop on Self-Organizing Systems*, IWSOS 08, Vienna, Austria, December 10-12, 2008.

[6] Next Generation Mobile Networks, *Use Cases related to Self Organising Network, Overall Description*, available at http://www.ngmn.org.

[7] 3GPP, *Physical Layer Aspects for evolved Universal Terrestrial Radio Access (E-UTRAN)*, Technical Report TR 25.814, available at http://www.3gpp.org.

[8] N. Scully, J. Turk, R. Litjens, U. Türke, M. Amirijoo, T. Jansen, L. Schmelz, "Review of use cases and framework II", *Deliverable 2.6 EU-Projekt SOCRATES (FP7-2007-INFSO-ICT-216284)*, March, 2009

[9] 3GPP, *Evolved Universal Terrestrial Radio Access Network (E-UTRAN); X2 Application Protocol (X2AP)*, Technical Specification TS 36.423, available at http://www.3gpp.org.

[10] U. Türke, and M. Koonert, "Advanced site configuration techniques for automatic UMTS radio network design," *Proc. Vehicular Technology conference VTC 2005 Spring*, vol. 3, pp. 1960-1964, Stockholm, Sweden, May 2005.

[11] H. Schumacher, M. Schack and T. Kürner, "Coupling of Simulators for the Investigation of Car-to-X Communication Aspects" 7th COST2100 Management Committee Meeting, TD(09)773, Braunschweig, Germany, February 2009.