

Impact of Traffic Geolocation Errors on Self-Organizing Network Performance

Felix Kirsten, Henrik Klessig, and Gerhard Fettweis

Vodafone Chair Mobile Communications Systems

Dresden University of Technology, Germany

Email: {felix.kirsten, henrik.klessig, fettweis}@tu-dresden.de

Andreas Hecker and Jens Voigt

Actix GmbH

Dresden, Germany

Email: {andreas.hecker, jens.voigt}@amdocs.com

Abstract—In recent years, research on flexible algorithms for self-organizing networks has been a significant part of the development of wireless network technologies, such as Long Term Evolution. In this paper, we focus on modeling specific error types related to input data, i.e., to mobile data traffic maps, and investigate their impact on self-organizing network algorithm and network performance. We develop traffic map error patterns, which are able to reproduce erroneous geolocation of mobile users, e.g., by methods like (Enhanced) Cell-ID, Observed Time Difference of Arrival, or Global Positioning System. Furthermore, we model errors that occur when call traces and performance measurement counters are used to determine user locations. We find that self-organizing network performance may be robust to systematic distortions of user and traffic distributions, provided that 1) the overall traffic in the network is estimated accurately, i.e., regions without traffic information are avoided, 2) performance measurement counters in the base stations provide additional information on imperfect geolocation data, and 3) a certain degree of correlation between original heterogeneous user distributions and distorted maps is maintained to identify traffic hot spots.

Index Terms—mobile traffic geolocation; error modeling; self-organizing network; network optimization; wireless network; performance evaluation; system modeling

I. INTRODUCTION

SON (Self-Organizing Networks) is a 3rd Generation Partnership Project (3GPP) concept for an advanced management and technology scheme that includes automated planning, optimization, and healing of cellular radio networks [1]. Based on a current network state (*input*) and aiming at achieving user-defined targets (load balancing, energy efficiency, etc.), SON adapts the network configuration (*result*) by means of flexible optimization algorithms. Algorithm performance and applicability of results depend on their input data. Hence, any deterioration in data quality affects the *optimal* result determined. To quantify this influence is the goal of our paper.

A loss in quality is similar to a degradation in accuracy. The accuracy of input data is generally limited according to 1) limitations of data sources available and 2) the reduction in granularity and precision in subsequent aggregation and estimation processes.

Data sources on the network side are performance measurement (PM) counters and call traces (CTs). PM counters comprise numerous event statistics related to network elements with temporal resolutions of hours and days [2].

CTs include detailed measurement reports [3] between user equipments (UEs) and the network elements. They are either provided from the Minimization of Drive Test NGMN/3GPP initiative [4] via a Trace Collection Entity [5] or are proprietary vendor solutions, like some PM counters.

Spatial distributions of user demands represent a type of input data that is frequently required. They are composed of information about user locations and user generated traffic volume. Aggregated to so-called *traffic maps*, they are used in reference scenarios to (re-)create user demands and traffic volumes as input for algorithms and, subsequently, their validation. In contrast to a fixed scenario with an existing traffic map, a real network collects user demands and events without explicit knowledge about a user's location. Therefore, locations have to be estimated in order to generate traffic maps.

Typical applications of localization methods [6] assign the site location (CID, Cell-ID), consider signal delay measurements (ECID, Enhanced CID), and include power measurements (AECID, Adaptive ECID) to improve results. The accuracy of these methods range from several tens of meters to several hundred meters. Other methods like OTDOA (Observed Time Difference Of Arrival) or GNSS (Global Navigation Satellite Systems) promise an even higher accuracy (10 to 60 m).

The choice of a localization method strongly depends on the data sources available and, therefore, determines the accuracy of the resulting input data. Furthermore, the data sources might be insufficient to reproduce complete knowledge about the actual traffic distribution. E.g., if measurement reports collected in CTs are only triggered by handover (HO) events, localization will be limited to HO regions, while conditions in cell centers are mostly unknown.

In this paper, we focus on quality loss of input data by means of specific error types with respect to a spatial user demand distribution. These error types are related to inaccurate localization and incomplete knowledge of network traffic. We investigate the impact of those errors on SON algorithms and the resulting network performance. As reference, we use the SON use cases Mobility Load Balancing (MLB) and Coverage and Capacity Optimization (CCO) as described in [7]. The air interface technology of the reference network is LTE (Long Term Evolution), for which SON has been a topic of discussion within 3GPP [8], [9].

The outline of this paper is as follows: Section II shows the error models to define the error characteristic of this survey. Section III describes the framework for our evaluation. Section IV presents the numerical results. Finally, Section V ends with the conclusions.

II. MODELING GEOLOCATION ERRORS

In this section, we develop two error models (one with regular characteristic, the other with respect to incomplete information) that reproduce geolocation error pattern on a reference traffic map.

A. Mobile Network and Traffic Model

Throughout the paper, we consider the downlink of a cellular network consisting of $N \in \mathbb{N}$ base stations covering a compact region $\mathcal{L} \subseteq \mathbb{R}^2$. Furthermore, we introduce corresponding cell areas $\mathcal{L}_i \subset \mathcal{L}$ with $\bigcup_i \mathcal{L}_i = \mathcal{L}$.

We assume mobile users to be spatially distributed according to some arbitrary distribution $\delta(\cdot)$ with $\int_{\mathcal{L}} \delta(u) du = 1$. Network traffic is modeled on flow level, where elastic flows represent individual data transfers of web pages, video, audio, or general data files [10]. We assume that the arrival of flow requests to the network takes place according to a Poisson process with intensity λ [11]. Flow sizes are assumed to be exponentially distributed with mean Ω . The terms λ , Ω , and $\delta(u)$ determine the traffic intensity distribution $\kappa(u) := \lambda\Omega\delta(u)$ in Mbps/km², which is represented by a spatial traffic intensity map. The reference traffic scenario chosen for our investigations represents a heterogeneous user distribution in the metropolitan area of a large North-American city and is depicted in Fig. 1 (a).

B. Accuracy Error Model [F-1]

In order to consider regular error patterns of geolocation techniques in LTE, we resort to a simple, yet realistic, model. We assume that user locations are determined independently with Gaussian error in both spatial dimensions around the actual UE position. By varying the standard deviation σ of the resulting circular Gaussian distribution between 0 to 300 m and applying it as a filter to the reference map, we are able to emulate techniques like Assisted-GPS, OTDOA, or ECID, which are typical for different localization procedures and network environments (urban density, indoor or outdoor location, antenna or satellite visibility, etc.). In the following, this error model is referred to as [F-1-GAUSS].

The resulting traffic maps show decrease in heterogeneity of user demand distribution (cf. Fig. 1 (b)). Traffic hot spots become less peaky. With standard deviation σ tending to infinity, the resulting traffic map represents a homogeneous distribution in \mathcal{L} . In the remainder of the paper, this special case is denoted as [F-1-GAUSS- ∞].

C. Incomplete Measurement Data [F-2]

In addition, we model errors originating from incomplete measurement data. We consider typical data corruption caused by a lack of UE measurement reports at locations with dominant primary cell reception, i. e., in cell centers.

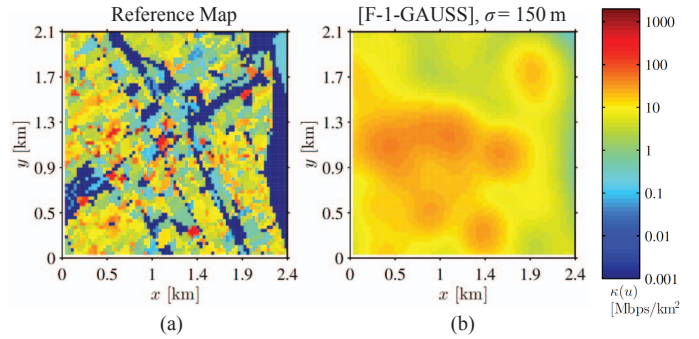


Fig. 1. (a) Reference traffic distribution $\kappa(u)$ at peak traffic hour and (b) traffic map resulting from geolocation error [F-1-GAUSS]

With this assumption, user traffic information is reduced to handover (HO) areas, where UEs are likely to perform HOs from one cell to another. To investigate different variations in the resulting fragmented maps, we introduce a HO-parameter Δp which is used as follows: We define the HO area $\mathcal{L}_{HO} \subset \mathcal{L}$ as the set of locations, where $p_{i,prim} \leq p_{j,sec} + \Delta p$ holds for all $j \neq i$. The quantities $p_{i,prim}$, $p_{i,sec}$, and Δp denote the Reference Signal Received Power (RSRP) values of the primary cell i and the second strongest cell j at location u , and the so-called HO-margin, respectively.

This method divides the cell areas \mathcal{L}_i into cell centers $\mathcal{L}_{i,ctr}$ and cell edge regions $\mathcal{L}_{i,HO}$. We assume the traffic distribution $\kappa(u)$ at cell edge locations $u \in \mathcal{L}_{i,HO}$ to be ideally and accurately determined, and consider two strategies to handle the traffic within cell centers $\mathcal{L}_{i,ctr}$:

- 1) Users at cell center locations $u \in \mathcal{L}_{i,ctr}$ cannot be localized and are, therefore, untraceable for traffic map generation. All values $\kappa(u)$ are set to 0 for $u \in \mathcal{L}_{i,ctr}$. Consequently, the overall traffic in \mathcal{L} decreases with a decrease in Δp . This error model is referred to as [F-2-EMPTY] and illustrated in Fig. 2 (a).
- 2) To obtain the overall network traffic distribution in \mathcal{L} more accurately, additional measurements may be considered (e. g., PM counters). These measurements can be used to fill the gaps in the traffic maps obtained from [F-2-EMPTY]. Traffic in $\mathcal{L}_{i,HO}$ is subtracted from the overall traffic in a cell \mathcal{L}_i , which is obtained from PM counters, and then distributed evenly over the cell center region $\mathcal{L}_{i,ctr}$. We denote this approach by [F-2-AVG]. An example traffic map is depicted in Fig. 2 (b).

The error patterns according to [F-2-EMPTY] and [F-2-AVG] represent a lack of information about the correct distribution of user traffic. For the maps depicted in Fig. 2, with $\Delta p = 3$ dB, only 33 % of the overall user traffic is localized correctly. The pattern [F-2-AVG] provides a more accurate determination of overall traffic. However, the user distribution is still assumed to be evenly distributed within the cell centers $\mathcal{L}_{i,ctr}$. With a HO margin $\Delta p = 0$ dB, the resulting traffic map exhibits homogeneous traffic distributions within each cell \mathcal{L}_i .

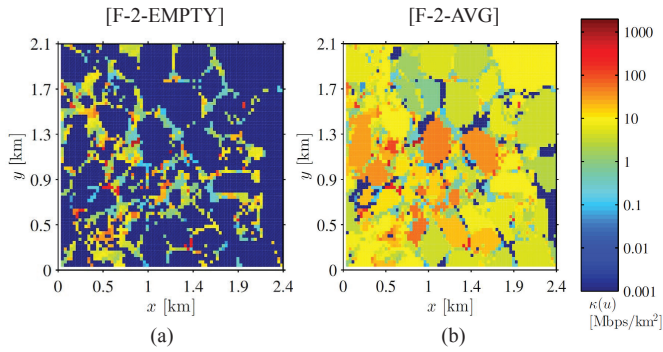


Fig. 2. Example traffic maps according to patterns (a) [F-2-EMPTY] and (b) [F-2-AVG] for $\Delta p = 3$ dB.

III. EVALUATION FRAMEWORK

The SON algorithm, which is used to characterize the impact of imperfect geolocation techniques, was initially presented in [7]. Both, the algorithm and the evaluation metrics, are based on a cell load model. This cell load model was first introduced in [12] and further developed in [13]. A brief discussion follows.

A. Cell Load Model and Network KPI Considered

We assume that data flows are subject to *average interference* conditions, which directly relate to base station resource utilizations or cell loads. Denoting the cell load vector $\eta := (\eta_1, \dots, \eta_N)^T$, the Signal-to-Interference-and-Noise-Ratio (SINR) at location $u \in \mathcal{L}$ with respect to cell i becomes

$$\gamma_i(u, \eta) := \frac{p_i(u)}{\sum_{j \neq i} \eta_j p_j(u) + N_0}, \quad (1)$$

where N_0 denotes the noise power. Further, we model the achievable rate by means of the Shannon capacity

$$c_i(u, \eta) := \min \left\{ aB \log_2 (1 + b\gamma_i(u, \eta)), c_{\max} \right\}. \quad (2)$$

The quantities B, a, b, c_{\max} denote the bandwidth, bandwidth efficiency, SINR efficiency, and the maximum rate given by the highest modulation and coding scheme for the mobile technology at hand, respectively.

Taking the (distorted) traffic density $\kappa(u)$ into account, we can define the load of a base station i as

$$\eta_i = f_i(\eta) := \min \left\{ \int_{\mathcal{L}_i} \frac{\kappa(u)}{c_i(u, \eta)} du, 1 \right\}. \quad (3)$$

Note that the cell load is given in an implicit form, which represents the interference-coupling behavior of, e.g., *frequency reuse one* network technologies such as LTE. Now, the resulting system of equations $\eta = f(\eta)$ with vector function $f(\cdot) = (f_1(\cdot), \dots, f_N(\cdot))^T$ can be solved via a fixed point iteration (see Theorem 1 in [13]).

One important network key performance indicator, which can be directly derived from the cell loads, is the average user

throughput at location u with respect to base station i

$$r_i(u, \eta) = c_i(u, \eta)(1 - \eta_i). \quad (4)$$

We use the 5th percentile of the user throughputs in order to evaluate the network and SON performance in the remainder of the paper. For further details on the cell load model and practical aspects, we refer to [7], [10], [13], [14].

B. MLB/CCO SON Algorithm

The goal of the MLB/CCO SON algorithm is to find a cell partition $\mathcal{P} := \{\mathcal{L}_1, \dots, \mathcal{L}_N\}$ and the electrical antenna tilt (etilt) configuration of all base stations, such that user throughputs are improved. The etilt configuration directly affects the receive powers $p_i(u)$ and, therefore, has a strong impact on network performance.

1) *Etilt Search*: To obtain a throughput-optimal etilt configuration, we resort to a simple, yet effective, coordinate-wise search strategy according to the taxi cab method, which is a special case of Powell's method, [15]. We choose the objective function Φ such that overloaded cells and congestion are most probably avoided if maximized, namely

$$\underset{\text{etilts}}{\text{maximize}} \quad \Phi(\eta) = \sum_i \log(1 - \eta_i). \quad (5)$$

Since the search algorithm is deterministic and limited by the maximum number of iterations through all antenna etilts, the result may not be a global but a local optimum instead. Furthermore, the result strongly depends on the input data quality and accuracy, i.e., the traffic map and, hence, on geolocation errors that may occur.

2) *User Association*: During each iteration step of the search for an optimal etilt configuration, cell areas are adjusted according to a user association rule given by

$$\mathcal{L}_i = \left\{ u \in \mathcal{L} \mid i = \underset{j}{\operatorname{argmax}} r_j(u, \eta) \right\}. \quad (6)$$

As shown in [16], choosing this throughput-optimal cell area definition is optimal with respect to the objective function (5). For further details regarding the operating principles of the SON algorithm, we refer to [7].

C. Evaluation Strategy

In order to evaluate the results, we distinguish four different types of network configuration and evaluation:

a) *[INIT]*: Network performance prior to any optimization / adaptation to traffic. This represents the initial network situation. Some cells in the network are overloaded and, therefore, congested.

b) *[REF]*: Reference network configuration and performance after the MLB/CCO SON algorithm with *undistorted* input data was triggered.

c) *[SELF]*: Network configuration / performance after the MLB/CCO SON algorithm with *distorted* input data was triggered. For network performance evaluation the *distorted* traffic map is used, such that it corresponds to the results

which are obtained if the distorted traffic map is assumed to be correct (*algorithm self-evaluation*).

d) [NET]: Network configuration / performance after the MLB/CCO SON algorithm with *distorted* input data was triggered. For network performance evaluation the *undistorted* traffic map is used, such that the results represent the real network behavior after optimization (*algorithm network evaluation*).

The evaluation scenario under consideration represents an LTE network deployed in the metropolitan area of a large North-American city. 55 base stations are irregularly deployed at 19 sites in total. Each base station employs a bandwidth of 10 MHz. For ease of computation, we numerically evaluate the SON and network performance according to the model presented before. However, the model is quite accurate in predicting real network performance, as shown in [17].

IV. NUMERICAL RESULTS

In this section, we present numerical results obtained by means of the evaluation framework explained above. We use flow throughput 5-percentiles as indicators of user satisfaction and illustrate the effects of geolocation errors according to [F-1] and [F-2] on SON performance.

A. Accuracy Error Model

The performance of the SON algorithm using the accuracy error pattern [F-1-GAUSS] is depicted in Fig. 3. Prior to any optimization, network performance is poor due to an overload situation at the peak traffic hour ([INIT]). Note that, according to our model, an overload situation in a cell implies an indefinitely growing number of users and, therefore, zero throughput in the long run. Remarkable improvements from 0 to 5.7 Mbps can be achieved, if the user distribution, i. e., the algorithm's input data, is accurately localized ([REF]). However, [NET] indicates, that the results based on increasingly distorted traffic maps show similar optimization gains. The MLB/CCO algorithm proves quite robust to inaccurate traffic measurements, i. e., less heterogeneity in spatial distribution. Interestingly, a homogeneous traffic distribution, provided by pattern [F-1-GAUSS- ∞], did not considerably lower the SON optimization gains in the scenario at hand.

The difference between the results for self-evaluation ([SELF]) and network evaluation ([NET]) reveals a self-estimation error, which is caused by user traffic that is inaccurately allocated to individual cells. The more uniform the traffic distribution among the cells, the greater is the tendency of the algorithm to underestimate network performance. This is a consequence of distributing traffic in space by the regular error pattern. The results are a degradation of the interference situation in the network and, hence, an underestimation of performance.

Additionally, standard deviations σ between 0 and 120 m caused by A-GPS, ECID, or OTDOA localization techniques, significantly improve optimization gains by up to 22 %. The reason for this gain is an increase in algorithm performance caused by the spatial traffic smoothing characteristic of the

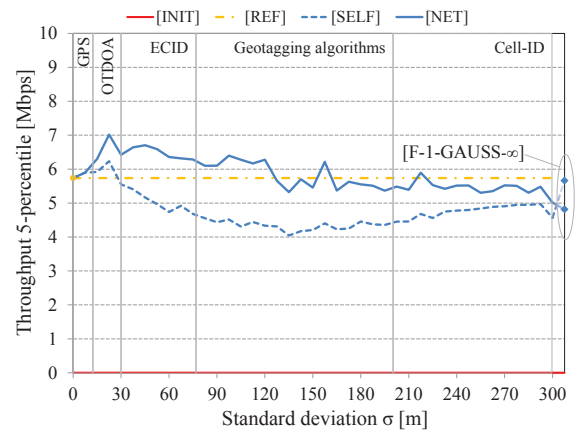


Fig. 3. User throughput 5-percentiles obtained with error pattern [F-1-GAUSS] for increasing standard deviation σ .

error pattern investigated. The SON algorithm requires a limitation of iteration steps resulting in a probability to obtain a less suitable etilt configuration. This probability is high with traffic peaks being small in terms of area, because they are more difficult to detect in a limited amount of steps. Now, this drawback is partly compensated by a redistribution of user traffic peaks. This results in a more efficient search for an optimal etilt configuration and, thereby, improves network and SON performance.

B. Incomplete Measurement Data

Fig. 4 depicts the results obtained with traffic maps distorted by error models [F-2-AVG] (evenly distributed traffic in cell centers) and [F-2-EMPTY] (cell centers not considered). Reference results ([REF]) are higher compared with the network evaluation and erroneous input ([NET]). However, [F-2-AVG] still provides acceptable results for HO margins up to 3.5 dB, which refers to 37 % of overall traffic being localized correctly. Lower margins Δp result in smaller HO areas. Therefore, less traffic is localized correctly which eventually results in lower optimization gains.

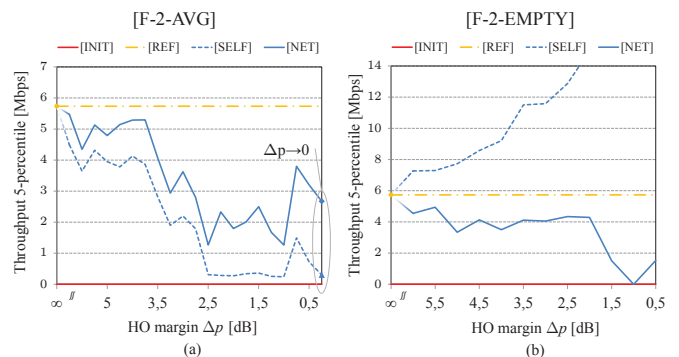


Fig. 4. User throughput 5-percentiles obtained with (a) error patterns [F-2-AVG] and (b) [F-2-EMPTY] for decreasing HO margin Δp .

Fig. 4 (a) shows that algorithm performance is underestimated with [F-2-AVG]. However, the opposite is observed when the traffic map is distorted by [F-2-EMPTY], see Fig. 4 (b). A significant overestimation is caused by disregarding traffic at the cell center areas, since cell loads are underestimated. Moreover, algorithm performance in case of pattern [F-2-AVG] is more accurate compared to the performance in case that pattern [F-2-EMPTY] is considered. This is due to a more accurate representation of absolute traffic and, hence, a more accurate prediction of cell loads.

V. CONCLUSIONS AND OUTLOOK

In this paper, we focused on modeling specific error types related to input data, more specifically to mobile data traffic maps, and investigated their impact on SON algorithm and network performance. The results presented highlight a certain robustness of SON optimization gains to systematic distortions of user and traffic distributions that are determined by methods like (Enhanced) cell-ID, Observed Time Difference of Arrival, or Global Positioning System. These optimization gains correlate with traffic map accuracy up to a certain extent. They can be achieved by three major factors:

- 1) Overall traffic in multiple cells has to be estimated accurately in absolute numbers. In this regard, performance measurement counters are an important source of information.
- 2) A certain degree of correlation between originally heterogeneous user distributions and distorted maps has to be maintained in order to locate traffic hot spots correctly and, likewise, areas of low demand.
- 3) Further, an increase in spatial traffic resolution does not necessarily yield larger optimization gains. On the contrary, performance may be increased by smoothing highly heterogeneous traffic maps for algorithms based on search strategies as in the algorithm used in this contribution.

The results presented are achieved by investigating one specific traffic and network scenario, which offers all characteristics of a typical dense urban mobile network. However, the results should be validated by further investigations utilizing different network configurations, or SON use cases, such as mobility robustness optimization or energy saving management.

ACKNOWLEDGMENT

The work presented in this paper was sponsored by the government of the Free State of Saxony, Germany, and by the European Regional Development Fund (ERDF) within the *Cool Silicon Cluster of Excellence* under contracts 31529/2794 and 29908/2794.

REFERENCES

- [1] L. C. Schmelz, H. van den Berg, et al., "Framework for the development of self-organisation methods," INFSO-ICT-216284 SOCRATES, Tech. Rep. D2.4, Sep 2008.

- [2] 3GPP, "3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Telecommunication management; Performance Management (PM); Performance measurements Evolved Universal Terrestrial Radio Access Network (E-UTRAN) (Release 12)," TS 32.425, 2013.
- [3] —, "3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification (Release 12)," TS 36.331, 2013.
- [4] —, "3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Universal Terrestrial Radio Access (UTRA) and Evolved Universal Terrestrial Radio Access (E-UTRA); Radio measurement collection for Minimization of Drive Tests (MDT); Overall description; Stage 2," TS 37.320, Mar. 2013.
- [5] —, "3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Telecommunication management; Subscriber and equipment trace; Trace concepts and requirements," TS 32.421, 2013.
- [6] A. Kangas, I. Siomina, and T. Wigren, "Positioning in LTE," in *Handbook of Position Location: Theory, Practise, and Advances*, 2012, pp. 1081–1127.
- [7] A. Fehske, H. Klessig, J. Voigt, and G. Fettweis, "Concurrent load-aware adjustment of user association and antenna tilts in self-organizing radio networks," *Vehicular Technology, IEEE Transactions on*, vol. 62, no. 5, pp. 1974–1988, 2013.
- [8] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access (E-UTRAN); Overall description; Stage 2," 3rd Generation Partnership Project (3GPP), TS 36.300, Sep. 2013.
- [9] —, "3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Self-configuring and self-optimizing network (SON) use cases and solutions," 3rd Generation Partnership Project (3GPP), TR 36.902, Mar. 2011.
- [10] S. Ben Fredj, T. Bonald, A. Proutiere, G. Régnié, and J. W. Roberts, "Statistical bandwidth sharing: A study of congestion on flow level," in *Proceedings of the 2001 conference on Applications, technologies, architectures, and protocols for computer communications - SIGCOMM '01*, vol. 31, no. 4. New York, New York, USA: ACM Press, Aug. 2001, pp. 111–122.
- [11] T. Bonald, "The Erlang model with non-poisson call arrivals," *ACM SIGMETRICS Performance Evaluation Review*, vol. 34, no. 1, p. 276, Jun. 2006.
- [12] K. Majewski and M. Koonert, "Conservative Cell Load Approximation for Radio Networks with Shannon Channels and its Application to LTE Network Planning," in *2010 Sixth Advanced International Conference on Telecommunications*. IEEE, 2010, pp. 219–225.
- [13] A. J. Fehske and G. P. Fettweis, "On flow level modeling of multi-cell wireless networks," in *2013 11th International Symposium on Modeling & Optimization in Mobile, Ad Hoc & Wireless Networks (WiOpt)*, Tsukuba City, 13-17 May 2013.
- [14] P. Mogensen, W. Na, I. Z. Kovacs, F. Frederiksen, A. Pokhariyal, K. I. Pedersen, T. Kolding, K. Hugl, and M. Kuusela, "LTE Capacity Compared to the Shannon Bound," in *2007 IEEE 65th Vehicular Technology Conference - VTC2007-Spring*, no. 1. IEEE, Apr. 2007, pp. 1234–1238.
- [15] M. J. D. Powell, "An efficient method for finding the minimum of a function of several variables without calculating derivatives," *The Computer Journal*, no. 7(2), pp. 155–162, 1964.
- [16] H. Kim, G. de Veciana, X. Yang, and M. Venkatchalam, "Distributed Alpha-Optimal User Association and Cell Load Balancing in Wireless Networks," *IEEE ACM Transactions on Networking*, vol. 20, no. 1, pp. 177–190, Feb. 2012.
- [17] A. Fehske, H. Klessig, J. Voigt, and G. Fettweis, "Flow-level models for capacity planning and management in interference-coupled wireless data networks," *IEEE Communications Magazine*, vol. 52, no. 2, February 2014.