Measurement Based Assessment of Spatial and Temporal Correlation of Base Station Load

Panagiotis Matzoros
OTEAcademy S.A.
Athens, Greece
pmatzoros@oteacademy.gr

Eduard A. Jorswieck
Communication Laboratory
TU Dresden, Germany
eduard.jorswieck@tu-dresden.de

George Agapiou
OTE S.A.
Athens, Greece
gagapiou@otereseacrh.gr

***Abstract*—For the revolution way from the current networks to the 5G era and beyond a set of stringent requirements and key performances should be meet. These define the need for more efficient procedures of designing and upgrading the Radio Access Network. In this paper we assess large scale measurements from an existing LTE network of a dense urban area (city) by focusing on the user load. In particular, we examine the spatial correlation of two clusters with different location characteristics by using a mathematical measure based on majorization. Finally we consider the temporal auto- and cross-correlation of the base stations in different locations. The assessment explains certain user and mobility behavior representative for this type of urban environment.**

Keywords—5G; User Load; LTE; Spatial and Temporal Correlation; RAN; Majorization;

I. Introduction

The continuous growing of the mobile data traffic at an exponential rate [1], the demand for more and better services from the users and in front of the new ecosystem that will be introduced by the 5th generation (5G) of the mobile networks, demands efficient ways of deploying and upgrading the current infrastructure. Moreover it gives the opportunity to better understand and measure the limits of the current deployment, to identify critical variables that affect the network capabilities and create efficient procedures of designing the future networks based on the current experience of the LTE access network. These opportunities provide a tremendous advantage to successfully meet the key performance indicators (KPIs) for the 5G era.

The work in this paper aim to benefit from the current LTE network deployment of a city in Greece by exploiting the opportunity to understand the behavior of the users equipment activity, to identify possible limits of the current deployment, to extract valuable information of how a fully operated network behaves regarding the users load and to formulate and introduce measure technics which aspire to help on designing the future networks.

A. *State of the Art*

There are several works in the past that study the dynamics and characteristics of a radio access network and some of them using data sets from actual networks [2], [3] and [4]. An analysis of a LTE RAN is reported in [2] which focuses primarily to the analysis of the users behavior and throughput of the access network during one day. The users activity distributions show the differences of peak hours in two different base stations and the common periodicity among different BS. As a step forward, it is important to expand the scale of the analysis in terms of the number of the base stations and in terms of time. In this way we can study not only the periodicity of the BS but also study if there is correlation between them in terms of distance and time. In [3] a large data set of user data traffic from a 3G network was used and with this data set a statistical analysis has been performed regarding the throughput. This analysis shows the potentials of large scale data sets and the distribution functions of the user data traffic. Finally, in [4] the spatial and temporal variations of the users load (in terms of data rate) in a 3G cellular network were examined and the importance of this in networks planning highlighted.

B. *Problem Statement and contributions*

The problem formulation is based on the measurements in the LTE access network in an urban environment. The selected area is the Heraklion city in Crete Island in Greece and the radio access network is from the leading operator in Greece. The measurements have been taken from 27 Base Stations (BS) for a week. The Fig. 1 represents the users equipment activity from 3 base stations during 5 weekdays.

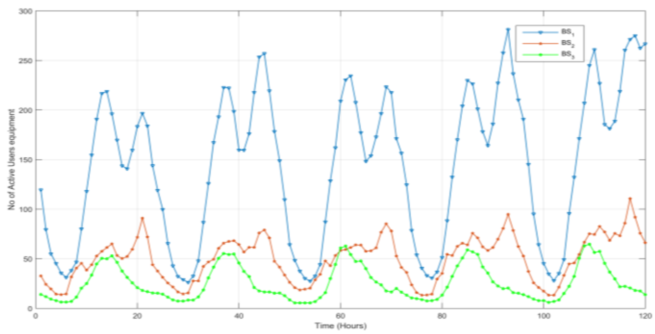

Fig. 1. The time series of users equipment activity from 4 BS in 5 working days (from Monday to Friday).

In this work we use the data from the above measurements, the users equipment activity (or user load), in order to address the following questions. Firstly, what is the impact of inter-BS distance on user load correlation? Secondly, to provide a more detailed statistical analysis and quantify temporal auto- and cross-correlation of the load. Thereby, we aim to explain the spatial correlation results. The results of the mathematical and analytical approach define our important contribution points:

- Introduce a mathematical measure based on majorization to compare different BS clusters and the inter-BS correlation of long-term user load.
- Numerical assessment and detailed discussion of auto- and cross- correlation for representative BS pairs from measurement campaign.

We discuss the geographical meaning of the inter-BS correlation measure. Furthermore, the numerical results illustrate the specific behavior and properties of Heraklion City, which can be generalized to other European cities of similar size.

The results of our analysis can be applied to the following relevant and important challenges:

- The planning and design of future Radio Access Networks in area with common characteristics.
- The extension of the current LTE network in cities with similar size and similar user behavior.
- The upgrading of the access network infrastructure in order to achieve to increase the data rate and the provided quality.
- The design of efficient algorithms to support load balancing and to save energy and costs while maintaining user satisfaction.

The paper is organized as follows: First, we provide details on the measurement setup and state the majorization relation as preliminaries in Section II. Next, we perform the spatial correlation analysis in Section III. In Section IV, the temporal/spatial correlation is studied and discussed. Finally, in Section V, the paper is concluded.

II. Preliminaries

A. Measurement Setup

This section describes the main characteristics of the LTE network and the measurement methodology. Firstly, the LTE network is composed by 32 Base stations and 150 cells. This means that one base station is able to serve up to 3 cells because of the three sections which each one has 120° angle of coverage. In addition, in many cases carrier aggregation is performed. Carrier aggregation is used in order to increase the bandwidth and thereby increase the data rate Fig. 2.

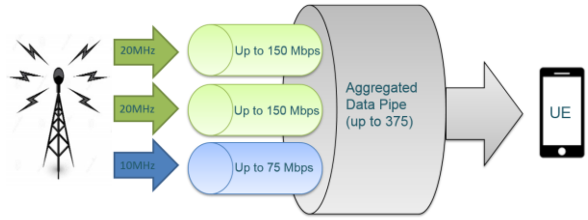

Fig. 2. Three carrier aggregation in downlink transmission.

In the specific network both, two carrier aggregation up to 300Mbps (20 MHz+20 MHz bandwidth) or 225Mbps (20MHz + 10MHz bandwidth), and three carrier aggregation, which give the ability to offer up to 375Mbps (20 MHz+20 MHz+10 MHz bandwidth), is applied. A base station which serves three directions and all of them have three carrier aggregation is able to have in total 9 cells (one for each direction and carrier) as shown in Fig. 3.

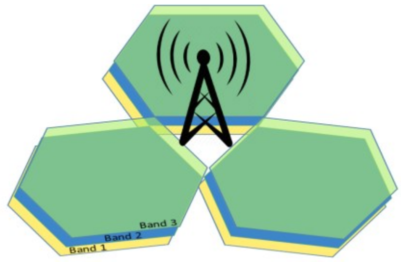

Fig. 3. Three carrier aggregation BS with total 9 cells and 375Mbps throughput capability.

The measurement was performed on 5 weekdays (Monday to Friday) for all the BSs. We take traces every hour and we focus on the users equipment activity. So every hour from all BSs we measure the aggregate average of the number of users equipments that are active (data exchange) in the network. By the term aggregated we mean the total number of users equipment in the cell (certain direction and carrier).

B. Majorization

This section provides the basic definitions for majorization. Note that parts of this section are discussed in more detail in [5]. However, the section is self-contained such that no further reference is needed for the basic understanding. Rigorous proofs can be found in [6] and [5].

There are many ways to define a partial order on vectors, e.g., majorization. In this section, we take the intuitive way and motivate the order by its operational meaning, which we use in the applications later. In order to assure that the concept of majorization is well understood, we present two equivalent definitions and discuss their relationship.

Definition 1: Let $a, b \in \Re_+^n$. We say that the vector a majorizes the vector b and write $a \succ b$ if there exists a doubly-stochastic matrix P such that $b = Pa$.

A doubly-stochastic matrix is a matrix with non-negative entries and row and column sum equal to one, i.e.

$$\sum_{i=1}^{n} P_{ij} = 1, 1 \le j \le n, \text{ and } \sum_{j=1}^{n} P_{ij} = 1, 1 \le \mathrm{i} \le \mathrm{n} \quad (1)$$

Note that by the definition, it follows automatically that the sum of all components of both vectors is equal. So this definition explains the intuition that one vector is *more spread out* than another vector if it is majorized by the other vector.

The second equivalent definition needs the elements of the two vectors ordered in non-increasing order, but it is the easiest to understand and to discuss.

Definition 2: Two vectors $a, b \in \Re_{+}^{n}$ fulfill the majorization inequality $a \succ b$ if after sorting the entries of each vector into non-increasing order, the sum of all entries is equal and the sum of the entries in every prefix of a is no less than the corresponding sum in b, i.e.

$$\sum_{k=1}^{m} a_k \ge \sum_{k=1}^{m} b_k, 1 \le m < n, \text{ and } \sum_{\mathrm{k}=1}^{\mathrm{n}} a_{\mathrm{k}} = \sum_{\mathrm{k}=1}^{\mathrm{n}} b_{\mathrm{k}} \quad (2)$$

Note that majorization is anti-symmetric, transitive but *not complete* (for vectors with more than two components). One counter example of vectors which cannot be compared by majorization is given by $a = [7 \quad 3 \quad 1]$ and $b = [6 \quad 5 \quad 0]$. Therefore majorization is a partial order.

However, we can always compare the extreme vectors $\gamma = [\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n}]$ and $\upsilon = [1,0,...,0]$ with an arbitrary vector b as follows $\gamma \succ b \succ \upsilon$.

Majorization is best illustrated by an example (from [5]). Note that the application of majorization in [5] was in spatially correlated multi-antenna channels. Here, we will apply majorization to compare different time-series of user load data from different base stations.

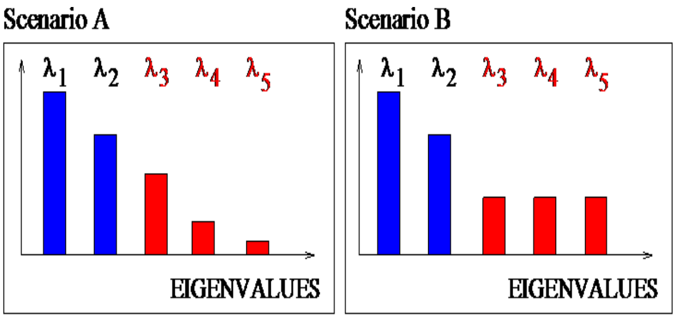

Fig. 4. Majorazition example

Let us compare eigenvalues of covariance matrices. Assume that we have two covariance matrices of size 5x5 and we want to understand which one shows stronger correlation. We propose to compute the eigenvalues of the two covariance matrices for scenario A and B and use majorization to compare them.

For the illustration above, the eigenvalues of the covariance matrix in Scenario A are given by α and the eigenvalues of the covariance matrix in Scenario B by β. Clearly, it holds that $\alpha \succ \beta$ which fits well to our intuition that the scenario A is stronger correlated, in particular looking at the three smallest eigenvalues.

In the following, we will apply this measure of correlation to the time series data from the base station measurement campaign.

III. Spatial Correlation Analysis

It is known that there exist differences between the theoretical way of how the access network is deployed and how actually the BS are deployed. In theory the BSs have constant inter-BS distance in order to be able to offer services in uniform way. However, the actual network is deployed steadily and is denser, many base stations have less distance, at areas with high capacity demand and BS are more scattered at areas with less capacity demand. This deployment offers 100% LTE coverage but it is service oriented and cost and energy efficient.

In order to motivate the spatial correlation analysis and our measure of spatial correlation, we form two different clusters, cluster N consists of 4 Base Stations i={1…4} where the inter BS distance is small and the cluster F which consists of 4 base stations j={1…4} with high inter BS distance.

To define the cross correlation matrix, first we normalize the users $U_i(t)$ where i is related to the BS and t is the time t=(1…120) hours, (24 hours for 5 days).

$$\tilde{U}_{i,j}(t) = \frac{U_{i,j}(t) - \overline{U}_{i,j}(t)}{\sqrt{\sum_{t=1}^{120} U_{i,j}(t)^2}} \quad (3)$$

Then we construct the matrix: T: [4x120].

$$\tilde{T} = \begin{bmatrix} \tilde{U}_1^1 & \tilde{U}_1^2 & \cdots & \tilde{U}_1^{120} \\ \tilde{U}_2^1 & \tilde{U}_2^2 & \cdots & \tilde{U}_2^{120} \\ \tilde{U}_3^1 & \tilde{U}_3^2 & \cdots & \tilde{U}_3^{120} \\ \tilde{U}_4^1 & \tilde{U}_4^2 & \cdots & \tilde{U}_4^{120} \end{bmatrix} \quad (4)$$

The cross-correlation matrix ($\tilde{C}$) can be computed by the multiplication of the matrix $\tilde{T}$ with its transpose $\tilde{T}^T$.

$$\tilde{C} = \tilde{T} \cdot \tilde{T}^T \quad (5)$$

The cross correlation matrix of the cluster N and cluster F, respectively, are:

$$\tilde{C}_N = \begin{bmatrix} 1 & 0.8738 & 0.8247 & 0.7915 \\ 0.8738 & 1 & 0.8695 & 0.9318 \\ 0.8247 & 0.8695 & 1 & 0.9268 \\ 0.7915 & 0.9318 & 0.9268 & 1 \end{bmatrix} \quad (6)$$

$$\tilde{C}_F = \begin{bmatrix} 1 & 0.4329 & 0.6252 & 0.8837 \\ 0.4329 & 1 & 0.9012 & 0.6270 \\ 0.6252 & 0.9012 & 1 & 0.6965 \\ 0.8837 & 0.6270 & 0.6965 & 1 \end{bmatrix} \quad (7)$$

Subsequently, to compare the clusters in terms of correlation, we apply our majorization measure on the eigenvalues of the two correlation matrices. Therefore, we compute the eigenvalue decomposition

$$[U, D] = eig(\tilde{C}) \quad (8)$$

The eigenvalues are $\lambda = diag(D)$ (9) and $U \cdot U^H = I$ is a unitary matrix.

From the equations (8) and (9), we compute the eigenvalues for the cluster N and cluster F, as follows and as shown in the diagram in the Fig. 5.

$$\lambda_N = \begin{bmatrix} 3.6104 \\ 0.2328 \\ 0.1246 \\ 0.0321 \end{bmatrix} \text{ and } \lambda_F = \begin{bmatrix} 3.0890 \\ 0.7324 \\ 0.1389 \\ 0.0397 \end{bmatrix}$$

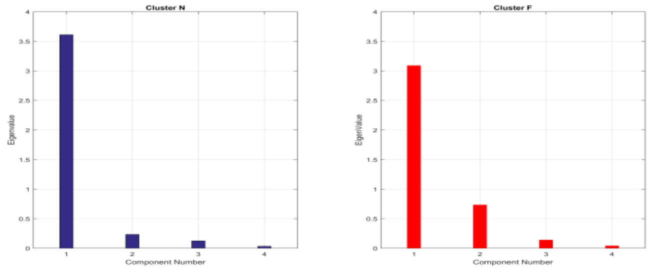

Fig. 5. Eigenvalues of cluster N and cluster F.

From the numbers and from the diagram, we can observe that the vector of eigenvalues of the correlation matrix of the closer BSs majorizes the eigenvalues of the correlation matrix of the more distant BS cluster, i.e.,

$$\lambda_N \succ \lambda_F,$$

since the sum of the two largest eigenvalues of N is 3,8432 and the sum of the two largest eigenvalues of F is 3,8214. Additionally, the sum of the three largest eigenvalues of N is 3,9678 while the sum of the three largest eigenvalues of F is 3,9603. Therefore, they fulfill the majorization order.

By formulating the cross corellation matrix we are able to observe the high spatial correlation of the base stations in the cluster N compered to the correlation of base stations in cluster F. The application of majorization verifies that cluster N majorizes the cluster F and safely can be used as correlation meusure prcedure to indetify the spatial correlation of different clusters of a network deployed at the same area. However, during this work we have to highlight that the difference of spatial correlation between the two clusters is not remarkable. This is independed from the base station selection but it is highly depended from the users behavior during the day. The users equipment activity traces have the same periodicity across the time (5 days). From this study safely we can state that the location of a base station has impact also from the users behavior and characteristics.

This motivates the following definition of "closer" or "nearer" BS cluster:

Definition 3: The BS cluster A with correlation matrix eigenvalues λ_A is said to be "closer" or "nearer" than the BS cluster B with correlation matrix eigenvalues λ_B, if it holds

$$\lambda_A \succ \lambda_B .$$

Still there remains the question why the spatial correlation between the far distance BS in the cluster F is relatively high. In order to answer this question, we have to take a closer look at the temporal/spatial correlation in the next Section.

IV. Temporal correlation analysis

In the previous section we have examined the spatial correlation of two clusters. The aim of this section is to measure the correlation of three base stations over time (and space). As input data we have used the same measurement setup from the LTE access network. One base station (U_1) is our reference point and it is located to the city center (shops, malls, office, etc.), from the other two U_2 has very low inter-BS distance and the U_3 has very high inter-BS distance (area with houses and apartments).

Firstly, we compute (Fig. 6) the auto-correlation function (10) for the 3 Base stations U_i with i {1, 2, 3} over the time T which is 120 hours (5 days).

$$C(k) = \frac{1}{T-1} \sum_{t=1}^{T-k} (U_i(t) - \overline{U}_i)(U_i(t+k) - \overline{U}_i) \quad (10)$$

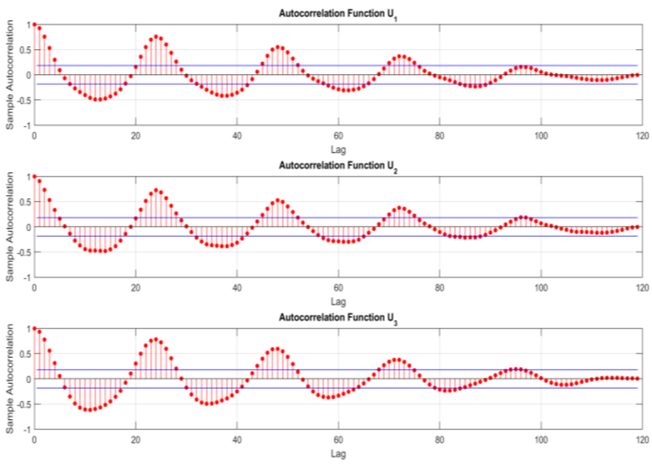

Fig. 6. Auto-correlation functions of the 3 Base stations

It can be observed that the 5 weekdays show strongly correlated behaviour of the user load. The periodicity of the auto-correlation function is about 24 hours and the minimum correlation occurs for a lag of 12 hours (as expected). All three BS show very similar auto-correlation properties.

Furthermore, in order to indentify the cross-correlation between the base stations we formulate the cross-correlation functions for two pairs. The first pair are the BSs U_1 and U_2 and the second pair are the BSs U_1 and U_3.

$$C_{U_1,U_2}(k) = \frac{1}{T}\sum_{t=1}^{T-k}(U_1(t) - \overline{U}_1)(U_2(t-k) - \overline{U}_2) \quad (11)$$

$$C_{U_1,U_3}(k) = \frac{1}{T}\sum_{t=1}^{T-k}(U_1(t) - \overline{U}_1)(U_3(t-k) - \overline{U}_3) \quad (12)$$

The figure below (Fig. 7) illustrates the cross-correlation of the two pairs.

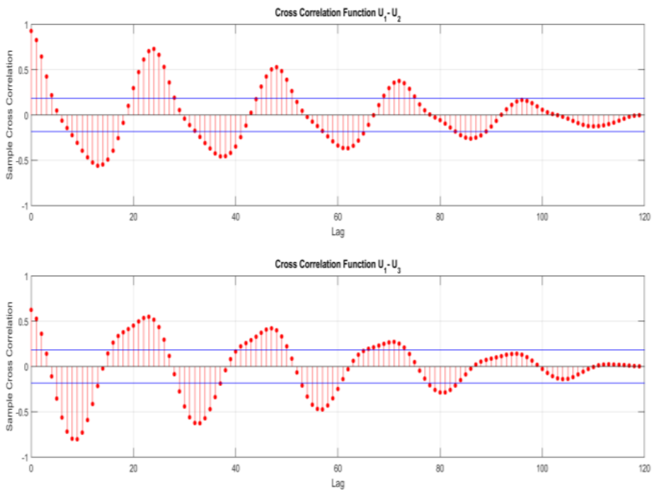

Fig. 7. Cross-correlation function of U_1-U_2 and U_1-U_3.

The above figures represent the higher correlation between the BS with small distance compared to the BS pair with larger distance. This study also verifies the impact of distance between 2 BS over the time.

However, as in the previous section we have to highlight that the difference it is not so remarkable. This can be explained as follows: From the cross correlation matrix we can see that we have a shift of one hour. This is again result of the users common behavior across the entire city. This shift of the peak illustrates the mobility of the people from the city center to the residence area. But this one hour shift does not lead to uncorrelated spatial time series - instead the one hour shift is well within the peak of the cross-correlation function. Therefore, the spatial correlation between the city center BS and the residence area BS is still large and we can conclude that all BS in the network are breathing with more or less the same rythm.

V. CONCLUSIONS

This work is a study in measurement analysis of BS load in terms of UEs activity in a large scale LTE network of a Greek city. Through this study we are able to identify the periodicity of UEs activity, the spatial correlation of two clusters with different inter-BS distance, to introduce the majorization as a spatial correlation measure and the temporal correlation of 3 BS with different inter-BS distance. The above results and observations can assist the designing of the future networks (5G) in a dense urban area. Furthermore, they can be used to improve the coverage and the data rate of the existing network. Finally, this offers the opportunity for cost and energy efficiency through the examinations of different dynamic approach of BS configuration across the time and space.

ACKNOWLEDGMENT

The work has received funding from the European Union Horizon 2020-MCSA-ITN-2015 Innovative Training Networks (ITN) under grant agreement No 641985 (5GWireless).

REFERENCES

[1] Ericsson Mobility Report. Ericsson, 2017.

[2] P. Matzoros, C. Tsirakis, G. Agapiou, "Radio Access Network Bahavior in Urban Enviroment", 2017 *14th International Symposium on Wireless Communication Systems (ISWCS'17), 2017*.

[3] E. Nan, X. Chu, W. Guo and J. Zhang, "User data traffic analysis for 3G cellular networks", 2013 *8th International Conference on Communications and Networking in China (CHINACOM), 2013.*

[4] U. Paul, A. Subramanian, M. Buddhikot and S. Das, "Understanding traffic dynamics in cellular data networks", *2011 Proceedings IEEE INFOCOM, 2011.*

[5] Jorswieck, E., & Boche, H. (2007). Majorization and Matrix Monotone Functions in Wireless Communications. *Foundations and Trends in Communications and Information Theory,* 3(6), 553-701.

[6] Marshall, A.W., & Olkin, I. (1979). Inequalities: Theory of Majorization and Its Application, *Mathematics in Science and Engineering Vol. 143, Academic Press, Inc. (London) Ltd.*