ERCIM "ALAIN BENSOUSSAN"
FELLOWSHIP PROGRAMME

ERCIM
European Research Consortium for Informatics and Mathematics

Scientific Report

First name / Family name	STAVROS/NTALAMPIRAS
Nationality	GREEK
Name of the *Host Organisation*	NATIONAL RESEARCH COUNCIL OF ITALY
First Name / family name of the *Scientific Coordinator*	MARCO/FIORE
Period of the fellowship	01/04/2017 to 30/11/2017

I – SCIENTIFIC ACTIVITY DURING YOUR FELLOWSHIP

Cellular networks have become an important part of the everyday lives of the major part of the population. They include wireless communication links enabling services based on the exchange of various information types among mobile devices. Learning, and subsequently, predicting traffic patterns in a service-depended way is a research field which, even though appealing, has not received a lot of attention by the scientific community.

Most research attempts [1], [2] focus on optimizing the network utilization under the assumption that important elements, including traffic load are a-priori known. However, in real-world applications such information might not always be available. On top of that, having the traffic predicted on a daily, hourly or even minutely basis could be the cornerstone for optimizing energy savings [3], opportunistic scheduling [4], and network anomaly detection [5]. Interestingly, such technologies could potentially contribute to the prompt diagnosis of faults and/or cyber-threats affecting various network components by evaluating the discrepancy between observed and predicted variables. Overall, having available accurate knowledge regarding the future traffic, enables the dynamic management of both designing and operational issues opening the path to traffic-aware cellular networks.

The proposed method was designed having in mind the following requirements: *a*) service independence, i.e. the method is able to operate on datastreams representing diverse services, i.e. video streaming, messaging, etc., *b*) geographical area in- dependence, i.e. the method does not take under consideration any a-priori knowledge regarding the

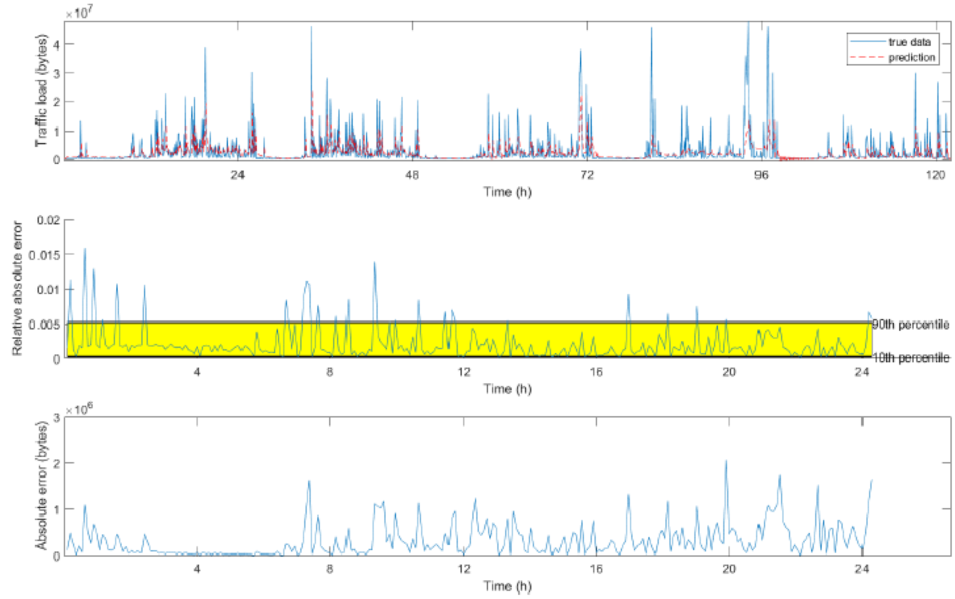

Fig. 1 Prediction results w.r.t the gaming services recorded in the area with id 13063 of Marseille.

characteristics of the geographical area that is to be applied, *c*) data availability, i.e. the method is able to operate in a limited data environment, and *d*) does not make any assumptions regarding the network topology nor the communication routing protocol.

The method exploits the functional characteristics of the network in an unsupervised manner by considering the correlation exhibited among the acquired datastreams. More specifically the algorithm automatically selects the *c* strongest correlated traffic loads with the one we wish to model. Subsequently, these *c* datastreams comprise the inputs to a multiple-input single-output (MISO) autoregressive moving average modelling scheme. The considered dataset includes data coming from five weekdays (Monday-Friday) regarding two cities in France (Marseille and Paris). The method was applied to all present communes, while results are reported using various figures of merit capturing

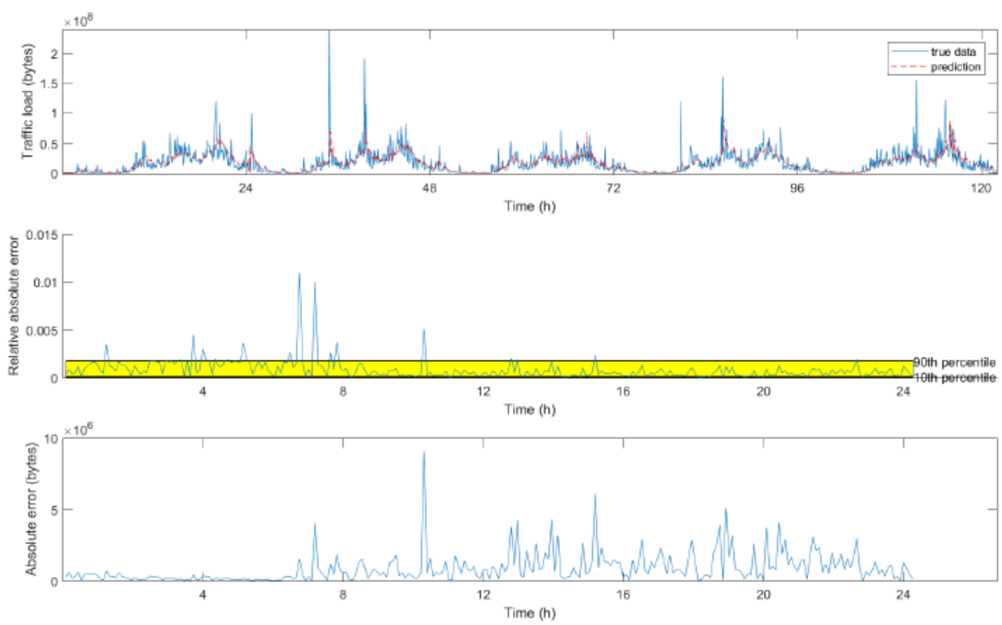

Fig. 2 Prediction results w.r.t the gaming services recorded in the area with id 75109 of Paris.

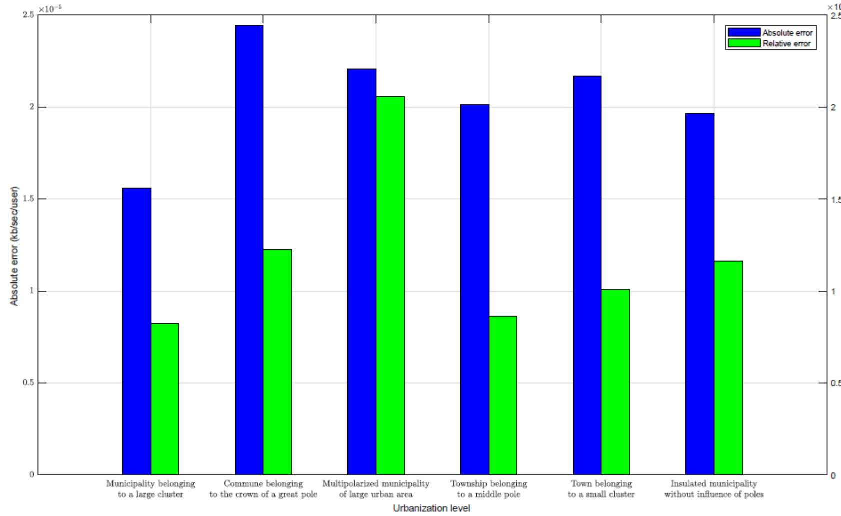

Fig. 3 Prediction results w.r.t the gaming services recorded in the area with id 75109 of Paris.

different aspects of the problem. Prediction results are then correlated with several demographic factors, such as urbanization level, amount of traffic, user age, income, etc. Finally, the performance of the proposed method was compared with other methods commonly used for time-series prediction and specifically for traffic load prediction at the service level.

Application examples

This subsection includes descriptive examples regarding the application of the proposed prediction algorithm. More specifically the following Figures 1 and 2 demonstrate prediction results associated with areas of Marseille and Paris.

Results

At this stage of the analysis, we aimed at understanding the different variables that may affect the prediction quality. Firstly, we examined the way the errors vary according to the urbanization level of each commune. The considered communes include the following urbanization levels:

- Municipality belonging to a large cluster,
- Commune belonging to the crown of a great pole,
- Multipolarized municipality of large urban area,
- Township belonging to a middle pole,
- Town belonging to a small cluster, and
- Isolated municipality without influence of poles.

The next Figure 3 shows how the prediction errors alter according to the urbanization level. We observe the absence of correlation with the urbanization level.

Comparison with the state of the art

The comparative results are demonstrated in the following Figure 4. The results are in the form of MSE, MAE, and logloss. We also employed an ARMA model in the comparative evaluation. There, we make the following observations:

- the effect of total traffic load is more evident in the MISO approach which may be due to the fact that it uses more than one datastreams for modelling and prediction,
- DL provides the worst performance among the considered methods which is

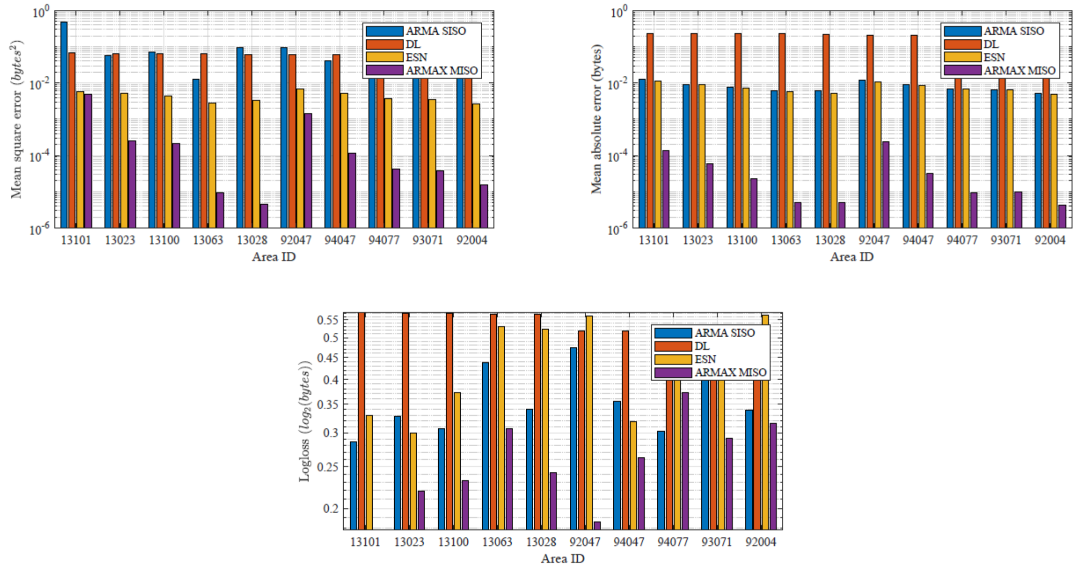

Fig. 4 Comparative results w.r.t prediction performance in 10 communes of diverse traffic load levels. The prediction methods are ARMA/DL/ESN/MISO- ARMAX

may be due to the limited amount of data for training,

- the MISO approach provides the best results w.r.t all metrics, while ARMA and ESN methods achieve inferior performance.

II – PUBLICATION(S) DURING YOUR FELLOWSHIP

Stavros Ntalampiras and Marco Fiore, "Prediction of cellular network traffic at the service level", *IEEE WoWMoM 2018* (pending).

III – ATTENDED SEMINARS, WORKHOPS, CONFERENCES

Talk "Model-free fault diagnosis for large scale cyber physical systems", at Politecnico di Torino, Italy, 16/5/2017.

IV – RESEARCH EXCHANGE PROGRAMME (REP)

Not implemented.

References

[1] J. Tang, G. Xue, and W. Zhang, "Cross-layer optimization for end-to-end rate allocation in multi-radio wireless mesh networks," *Wireless Networks*, vol. 15, no. 1, pp. 53–64, Jan 2009. [Online]. Available: https://doi.org/10.1007/s11276-007-0024-y

[2] P. Thulasiraman, J. Chen, and X. Shen, "Multipath routing and max- min fair qos provisioning under interference constraints in wireless multihop networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, no. 5, pp. 716–728, May 2011.

[3] R. Li, Z. Zhao, X. Zhou, and H. Zhang, "Energy savings scheme in radio access networks via compressive sensing-based traffic load prediction," *Trans. Emerg. Telecommun. Technol.*, vol. 25, no. 4, pp. 468–478, Apr. 2014. [Online]. Available: http://dx.doi.org/10.1002/ett.2583

[4] U. Paul, M. M. Buddhikot, and S. R. Das, "Opportunistic traffic scheduling in cellular data networks," in *2012 IEEE International Symposium on Dynamic Spectrum Access Networks*, Oct 2012, pp. 339– 348.

[5] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, "Network anomaly detection: Methods, systems and tools," *IEEE Communications Surveys Tutorials*, vol. 16, no. 1, pp. 303–336, First 2014.