Forecasting mobile service demands for anticipatory MEC

Stavros Ntalampiras[†‡] and Marco Fiore[†]

[†]CNR-IEIIT
Corso Duca degli Abruzzi 24
10129 – Torino, Italy

[‡]Department of Computer Science
University of Milan
Via Comelico 39
20135 – Milano, Italy

***Abstract*—The accurate estimation of future traffic loads is a key enabler for anticipatory mobile networking. In this paper, we investigate the prediction of the traffic generated by different mobile service classes over base station clusters, at an order-of-minute granularity and using relatively short historical data. This scenario is relevant to mobile edge computing (MEC), where resources need to be orchestrated for individual services separately across multiple base stations, at fairly long timescales. To address the prediction problem, we propose a novel forecasting model based on an autoregressive multiple-input single-output (MISO) approach, where the inputs are collected from regions exhibiting strong correlations in the offered load of a specific mobile service. Experiments on real-world data collected in an operational 3G/4G network demonstrate the effectiveness of our model, which attains average relative errors between 0.4% and 5% when forecasting 5-minute-aggregate traffic of individual mobile service classes.**

I. Introduction

Reconfigurability will be the main paradigm shift set forth by 5G with respect to previous generations of mobile networks. Softwarization of the infrastructure and virtualization of functions will decouple the data and control planes, finally offering operators with a means to control system resources dynamically. Reconfigurability will be enacted in an end-to-end fashion, from radio access to core network gateways. A range of solutions will concur to realize this objective, including cognitive radio and spectrum sharing, cloud radio access networks, mobile edge computing, and virtualized network cores. Ultimately, this novel flexibility will provide the technical support for cognitive network management, *i.e.*, the automated data-driven orchestration of network resources envisioned to mark 5G according to industrial and innovation actors in the mobile communication ecosystem [1], [2].

Cognitive network management will require an accurate and near-real-time prediction of subscribers' mobility and requests, so as to perform a timely reconfiguration of resources according to fluctuations in the user demand [3]–[5]. This calls for a whole range of effective mobile traffic forecasting models, each operating at a specific granularity and timescale: for instance, radio resource block scheduling at individual antennas will rely on session-level millisecond-latency predictors [6], whereas virtual machine or container re-deployment in the cloudified core network will need anticipating macroscopic profiles on, *e.g.*, a hourly basis [7].

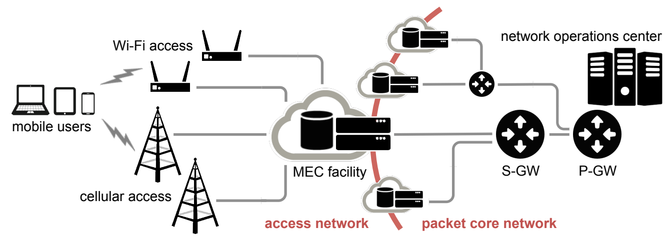


Fig. 1. System model. Each of many MEC facilities deployed at the mobile network edge controls the virtualized resources of subscribers associated to a subset of access facilities through, *e.g.*, cellular or Wi-Fi technologies.

Problem statement. We focus on traffic forecasts for the orchestration of resources in mobile edge computing (MEC) facilities deployed in proximity of the access network. Fig. 1 illustrates our target scenario. There, each MEC facility hosts virtualized personal environments that facilitate the fruition of mobile services by subscribers associated to a set of base stations. In this context, predicting the offered load at each MEC facility enables the preemption of computation and storage resources, and paves the way for a more effective allocation and migration of virtual machines or containers among MEC facilities. The milieu above imposes the following specifications to the forecast process.

(i) The resource orchestration occurs at the level of base station clusters under control of a same MEC facility. A consistent spatial granularity must be considered in the prediction model, which shall thus be tailored for aggregate traffic loads accommodated by multiple base stations.

(ii) Much of the advantage of MEC-level cognitive networking is based on the orchestration of resources allocated to separate types of mobile services, hence the prediction model shall be able to anticipate traffic demands on a per-service-class basis. The task is not trivial: as exemplified in Fig. 2, the temporal dynamics of the traffic offered by different applications at a single MEC facility can be very diverse, yet the forecasting model must perform well with all service types.

(iii) Conserving measurement data at MEC facilities is expensive, as it requires additional dedicated storage. It also raises privacy concerns, as upcoming data protection regula-

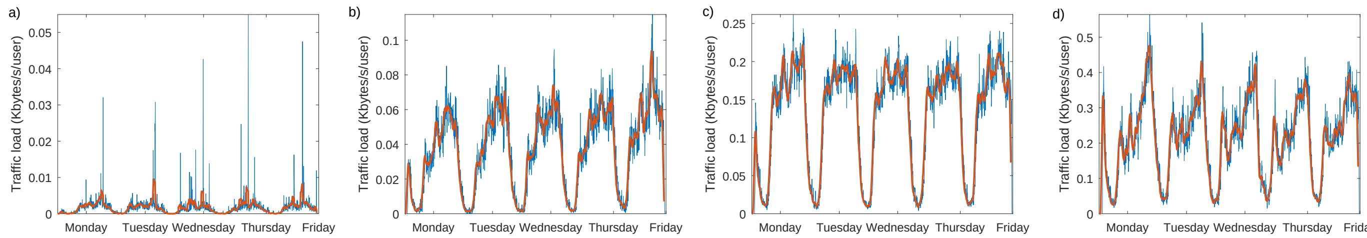


Fig. 2. Weekly time series of the offered traffic at one MEC facility, for four representative mobile services: (a) gaming, (b) messaging, (c) social network, and (d) video streaming. Smoothed time series are superposed to the instantaneous ones, so as to highlight the diverse dynamics of peaks in the demands.

tions will limit the amount of time during which operators can legitimately keep subscriber-generated information [8]. It is thus desirable that the training data needed by the prediction model to return reliable results is mimimized.

Contribution. We present an original forecasting model that satisfies the provisions imposed by anticipatory MEC above. Our proposed model exploits the functional characteristics of the per-service traffic demands in an unsupervised manner by considering the correlation exhibited among multiple datastreams. More specifically, the method automatically selects the historical demands observed at MEC facilities that best correlate with the historical time series recorded at the MEC facility targeted for prediction. All the correlated datastreams are then input to a multiple-input single-output (MISO) autoregressive moving average modeling scheme.

This approach yields several significant advantages: *(i)* it is fully unsupervised; *(ii)* it is flexible enough to adapt to the heterogeneous dynamics in the demand generated by different kinds of mobile service over base station clusters; *(iii)* it compensates for limited historical data by leveraging similarities in the measurements in different regions.

Extensive tests with large-scale real-world traffic demands demonstrate the effectiveness of our proposed model.

II. Forecasting model

In order to present our proposed forecasting model, we first provide a formal definition of the mobile service demand prediction problem (Sec. II-A). We then introduce a basic autoregressive moving-average approach (Sec. II-B), and explain how we extend it to a correlation-based solution that can meet our requirements, by exploiting the relationships observed in the offered mobile traffic load at different locations (Sec. II-C).

A. Problem formulation

Let us consider a mobile network where a set $\mathbb{M}$ of MEC facilities control the virtualized resources of users associated with a large number of base stations[1]. Specifically, each MEC facility is in charge of managing virtual machines or containers for mobile subscribers served by a specific subset of base stations. A MEC facility $i \in \mathbb{M}$ can thus monitor the load offered at its managed base stations, acquiring samples $x_i^t(s) \in \mathbb{R}_0^+$ that denote the volume of traffic observed at time t for a specific mobile service $s \in \mathbb{S}$. Over time, these samples form a datastream $\mathbb{x}_i^t(s) = \{x_i^1(s), \ldots, x_i^t(s)\}$ that represents the time series of measured traffic volumes for service s at MEC facility i, up to the current time t.

Our problem is predicting the future datastream samples $x_i^{t^*}(s)$, where $t^* > t$. A reliable forecast of upcoming samples allows foreseeing fluctuations in the demand for a specific service: in turn, this allows starting in advance the relocation of virtualized resources, *e.g.*, across network slices [9]. Indeed, resource relocation is critical to balance load among MEC facilities, but it involves migrating virtual machines or containers, which are fairly slow operations; anticipating them can lead to substantial improvements in network management and users' quality of experience [10].

Formally, at each MEC facility i, we aim at providing an accurate estimate

$$\hat{\mathbb{x}}_i^{t,h}(s) = \{\hat{x}_i^{t+1}(s), \hat{x}_i^{t+2}(s), \ldots, \hat{x}_i^{t+h}(s)\}$$

of the next h samples – referred to as the prediction *horizon* hereinafter – of the demand for a specific service s.

The following important remarks are in order. First, if not mentioned otherwise, we assume in the following $h = 1$, *i.e.*, an horizon that corresponds to the next sample $x_i^{t+1}(s)$; we will however also investigate the case where $h > 1$. Second, we consider that MEC facilities are loosely synchronized, so that the indices t in their recorded datastreams are consistent; in our case studies, a precision of the synchronization in the order of seconds is sufficient, which is a very reasonable supposition. Third, our problem formulation can accommodate any geographical distribution of MEC facilities.

B. Baseline autoregressive moving-average model

Our forecasting model stems from a basic autoregressive moving-average (ARMA) method. ARMA models leverage an input-output representation of the process under observation [11], which is described in terms of two polynomials accounting for the autoregressive (AR) and moving-average (MA) parts. Specifically, the AR compound performs regression based on lagged values of the input, while the MA part compensates for the error, which is approximated by a linear combination of discrepancies computed from past instances of the process. The AR and MA components are characterized

[1] We use the term base station to indicate all possible implementations of the radio access infrastructure, including cellular (*e.g.*, 3G, 4G, or beyond), Wi-Fi or other envisioned (*e.g.*,, mmWave, Li-Fi) technologies – see Fig. 1. Our problem formulation is independent of the access network realization.

by possibly different orders, denoted as p and q, respectively. The general form of the ARMA model, for $h = 1$, is

$$\hat{x}_i^{t+1}(s) = \kappa + \epsilon_i^{t+1}(s) + \sum_{j=0}^{p-1} \phi_j x_i^{t-j}(s) + \sum_{j=0}^{q-1} \theta_j \epsilon_i^{t-j}(s), \quad (1)$$

where κ is a constant, $\epsilon_i^{t+1}(s) \sim N(0, \sigma^2)$ is a zero-mean Gaussian noise with $\sigma^2 = \mathrm{Var}[\hat{\mathbb{x}}_i^{1,\tau}(s) - \mathbb{x}_i^{\tau}(s)]$ with $\tau \in [t - q + 1, t]$, *i.e.*, the variance of errors in the previous q estimated samples. The AR and MA parameters ϕ_i and θ_i weight the contributions of past samples and errors, respectively[2].

The expression is easily extended to the case of $h > 1$, iterating over t and using in the AR component estimated samples to represent the demand at time instants $t^* > t$ that have not yet occurred.

C. Correlation-based model with exogenous input

Our proposed approach builds on the ARMA model above, and extends it so as to exploit the relationships that characterize the traffic demand of a same mobile service across different geographical areas [12]. In a nutshell, the historical offered loads observed at all MEC facilities for a same service are inspected for affinity, and strongly correlated time series are used as exogenous inputs to reinforce the ARMA model.

Formally, given the datastreams $\mathbb{x}_i^t(s) = \{x_i^1(s), \dots, x_i^t(s)\}$ recorded at all MEC facilities $i \in \mathbb{M}$ for a service $s \in \mathbb{S}$ up to the current time t, the model computes the sample Pearson correlation coefficient for each MEC facility pair $a, b \in \mathbb{M}$ as

$$C_{a,b}^t(s) = \frac{\mathbb{v}_a^t(s)[\mathbb{v}_b^t(s)]^\intercal}{\sqrt{\mathbb{v}_a^t(s)[\mathbb{v}_a^t(s)]^\intercal}\sqrt{\mathbb{v}_b^t(s)[\mathbb{v}_b^t(s)]^\intercal}},$$

where

$$\mathbb{v}_i^t(s) = \mathbb{x}_i^t(s) - \overline{x}_i^t(s),$$

and $\overline{x}_i^t(s)$ is the mean of all samples in $\mathbb{x}_i^t(s)$, while $[\cdot]^\intercal$ is the transpose operator. The c datastreams that are the most correlated with that of each MEC facility i are collected in a set $\mathbb{C}_i^t$. Finally, $\mathbb{C}_i^t$ is fed to a multiple-input single-output (MISO) ARMA model with exogenous input (ARMAX)

$$\begin{aligned}\hat{x}_i^{t+1}(s) \quad = \quad & \epsilon_i^{t+1}(s) + \sum_{j=0}^{p-1} \phi_j x_i^{t-j}(s) + \\ & \sum_{j=0}^{q-1} \theta_j \epsilon_i^{t-j}(s) + \sum_{c \in \mathbb{C}_i^t} \sum_{j=0}^{r-1} \eta_j x_c^{t-j}(s), \quad (2)\end{aligned}$$

where η_j weights the order-j (out of total order r) contribution of samples of the c-th exogenous input, *i.e.*, $x_c^{t-j}(s)$.

As a concluding remark, from an engineering viewpoint, implementing the MISO ARMAX forecasting model above requires that MEC facilities exchange their datastreams, or communicate them to some controller located, *e.g.*, at the network operations center (as in Fig. 1) or at some intermediate location in the core network. However, these data transfers introduce a negligible overhead when compared to the actual traffic volumes generated by mobile services; moreover, they do not have strict quality-of-service requirements, as the prediction process occurs over timescales of minutes at MEC level, and datastream transfer delays in the order of seconds would be acceptable.

III. DATASET

In order to evaluate the accuracy of the forecasting model described in Sec. II, we use measurement data collected in an operational 3G/4G network. Specifically, passive probes tapping into the Gn interface at the Gateway GPRS Supporting Nodes (GGSN) in the 3G Packet Switched Core and into the S5/S8 interface at the Packet Data Network Gateway (P-GW) in the 4G Evolved Packet Core were employed to monitor the GPRS Tunneling Protocol user plane (GTP-U) and record all mobile data traffic in the network.

The mobile service associated to each IP session was identified by means of Deep Packet Inspection (DPI) on the collected data traffic, adopting dedicated fingerprinting techniques in presence of encrypted traffic[3]. In our study, we focus on four representative classes of mobile services.

- *Gaming* encompasses a large number of mobile videogaming applications. Popular examples are Pokemon Go, Clash of Clans, Clash Royale, or Candy Crush.
- *Messaging* includes the two dominant mobile chat services in the country, *i.e.*, Snapchat and WhatsApp.
- *Social media* covers the three top social networking services in terms of mobile data traffic generated in the target country, *i.e.*, Facebook, Twitter and Instagram.
- *Video streaming* consists of mobile applications that are dedicated to distributing video content, including YouTube, iTunes, NetFlix, DailyMotion, among others.

Our choice of mobile service classes is driven by two considerations: *(i)* these services are extremely popular, and generate a significant portion of the overall mobile data traffic; *(ii)* they are all good examples of applications that could benefit from the presence of MEC facilities, to reduce latency (gaming, video streaming), or to relieve the core network from part of its load via caching of favored contents (messaging, social media, video streaming).

The measurements were carried out in a large region of 760 km^2, which encompasses a major metropolitan area, its suburban surroundings, and the rural areas around those. We assume that MEC facilities are deployed in the target region so that each facility controls all access infrastructures that provide coverage to a specific geographical area. In order to define such areas, we leverage a fine-grained administrative division of the regions, which leads to a capillary deployment of MEC facilities that is fairly uniform in space – a highly desirable feature for latency reduction. Ultimately, 123 MEC facilities are present in the reference region, each covering a median area of around 5 km^2. For each MEC facility and every

[2] All parameters are independently tuned for each MEC facility i and service s. For the sake of clarity, we omit the corresponding indices in equation (1).

[3] Due to confidentiality agreements with the mobile network operator, we cannot disclose additional implementation details.

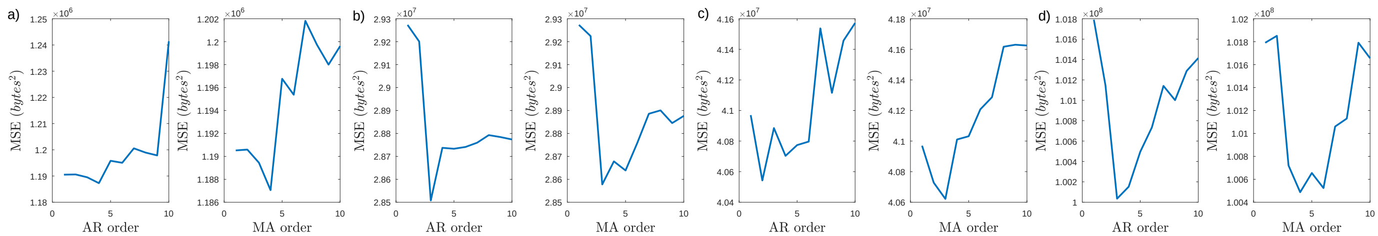


Fig. 3. MSE versus AR and MA orders of our forecasting model. Results refer to the best performing model for each order, as obtained from leave-one-day out experiments on historical data collected at one MEC facility. Plots refer to (a) gaming, (b) messaging, (c) social media, and (d) video streaming.

mobile service class, we aggregate the traffic generated by all relevant applications over 5-minute time intervals, which is a reasonable timescale for resource management in anticipatory MEC. Note that this implies that the time step t in Sec. II spans 5 minutes: for instance, a next-sample forecast ($h = 1$) predicts the total mobile traffic volume that will be observed in the following 5 minutes.

IV. Performance evaluation

We first discuss the parametrization of our proposed forecasting model (Sec. IV-A), and introduce the figures of merit we use to evaluate its performance (Sec. IV-B). We then present the evaluation results (Sec. IV-C).

A. Parametrization

All model parameters are selected by training the MISO ARMAX model presented in Sec. II-C on the historical datastreams recorded by each MEC facility for every mobile service class. To this end, we assume that each MEC facility is allowed to keep one week of data, due to storage limitations and personal data protection regulations, and that these data are leveraged as historical datastreams to tune the model.

Specifically, the parameters ϕ_j, θ_j, and η_j are computed using a standard iterative prediction-error method [13], and are typically assigned values in the range [-0.7,0.2]. The number of exogenous inputs c is set to 5, *i.e.*, around 2% of the external datastreams: this keeps the model simple while substantially improving performance, as detailed in Sec. IV-C.

The orders p, q and r of the model are instead selected by exhaustive search, using a leave-one-day out strategy over the available weekdays. Models with different p, q and r combinations are learned with data from all days except one, and their performance is assessed on the remaining day. The model achieving the lowest mean squared error (MSE) is finally selected as the reference one. We find that searching a space [1,10] is sufficient to identify clear minima in the MSE, for all orders: an example is provided in Fig. 3 for p and q, under different mobile service classes at one MEC facility.

B. Figures of merit

We recall that the actual demand generated at time t by the mobile service $s \in \mathbb{S}$ in the area under control of MEC facility $i \in \mathbb{M}$ is denoted as $x_i^t(s)$, whereas its estimate is $\hat{x}_i^t(s)$. We assess the prediction quality with the following metrics [14].

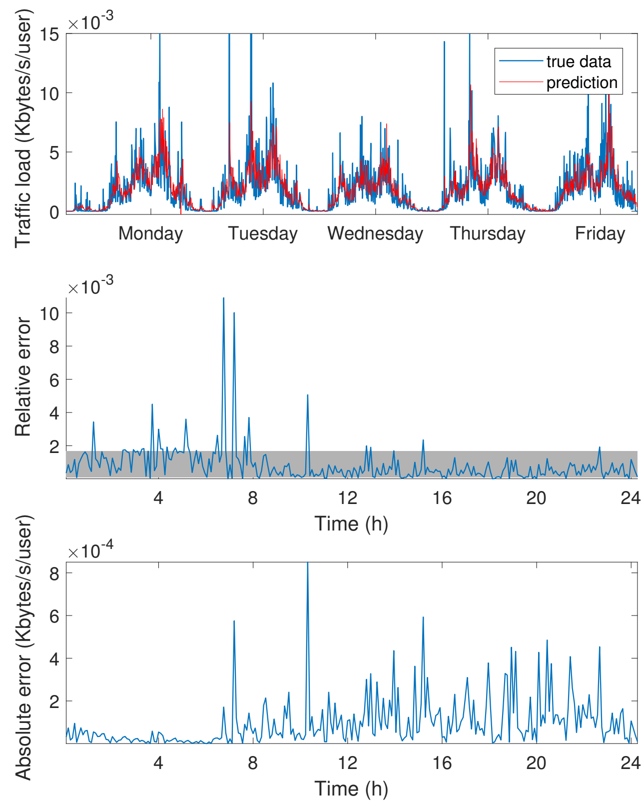
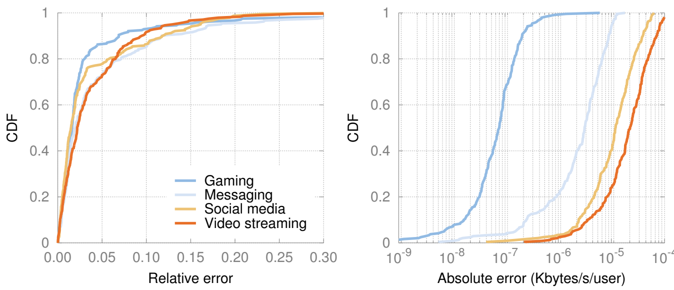


Fig. 4. Sample forecast, for the gaming mobile service class at one MEC facility. Actual and predicted traffic (top), relative error with 10$^{\text{th}}$-90$^{\text{th}}$ percentile range as the grey interval (middle), and absolute error (bottom).

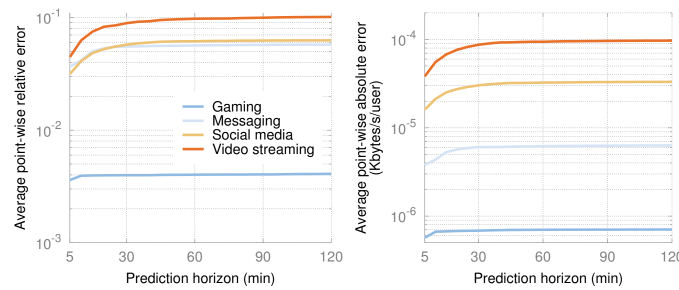
- The *relative error* measures the fractional error of the prediction with respect to the actual traffic. It is a dimensionless scalar $e^t_{rel,i}(s) = |\hat{x}_i^t(s)/x_i^t(s) - 1|$. This metric is used, *e.g.*, in [15].
- The *absolute error* measures the forecast error in terms of traffic volume. In order to make it comparable across MEC facilities that serve different user bases, we express it in kilobytes per second and per subscriber present in the MEC facility area[4]. Formally $e^t_{abs,i}(s) = |\hat{x}_i^t(s) - x_i^t(s)|$. This metric is used, *e.g.*, in [16].

[4]Note that we consider all mobile users in the area, including inactive users. This leads to low values of traffic volume per second and per subscriber. Ideally, including only users that generate traffic for a specific service would return more interpretable values, but it is an information we unfortunately do not have access to.



(a) (b)

Fig. 5. Cumulative distribution functions of (a) relative error and (b) absolute error. Results are aggregated for all MEC facilities, and curves refer to gaming, messaging, social media, and video streaming mobile services.



(a) (b)

Fig. 6. Average point-wise (a) relative error and (b) absolute error versus the prediction horizon. Results are means over all MEC facilities, and curves refer to gaming, messaging, social media, and video streaming mobile services.

C. Preliminary results

An illustrative example of the typical performance of the proposed MISO ARMAX model is provided in Fig. 4 for the gaming mobile service class at one randomly selected MEC facility. The top plot shows the demand for gaming services over five working days, Monday through Friday, and the prediction returned by our model. Apart from a few instants characterized by sudden traffic surges that are by their own nature hard to anticipate, the prediction overlaps well with the actual offered load.

The middle and bottom plots show the relative and absolute errors computed on the time series in the top plot. Both errors are shown over 24 hours, by averaging results over the original five days, so as to highlight possible daily dynamics. Indeed, we observe that the relative error tends to be higher overnight, when the traffic is lower and random fluctuations are more likely to occur. However, the small demand for services between midnight and 7 am translates into much higher absolute error during the daylight hours. Independently from these night-day oscillations, the accuracy of the forecast stays good in all cases, with a relative error below 0.2% over 90% of the time for the selected MEC facility and service.

Fig. 5 provides a more global view of the model performance. The plots show the cumulative distribution functions (CDF) of the relative and absolute error measured during five working days at all 123 MEC facilities in our reference region. In each plot, different curves maps to the four classes of mobile services we consider.

In the vast majority of the cases (85% to 95%, depending on the mobile service considered), the model provides forecasts with a relative error, in Fig. 5a, below 0.1: in other words, the estimate lies within 10% of the actual demand value. For all classes of mobile service, the relative error is below 0.02 (*i.e.*,, the estimate is within 2% of the true value) in half of cases at least. The mean relative errors are at 3.8%, 4.8%, 3.7%, and 3.9% for gaming, messaging, social media, and video streaming services, respectively. Interestingly, the performance of our forecasting model only differ slightly across services in Fig. 5a, despite the substantial differences in the dynamics of their respective demands shown in Fig. 2. This is a promising result towards the possibility of employing the proposed approach to predict traffic generated by very heterogeneous services. When looking at the absolute error, in Fig. 5a, the value ranges become different among services: despite close relative errors, the total volume of traffic generated by classes of mobile applications is not the same, and, *e.g.*, video streaming incurs an absolute error three orders of magnitude higher than that of gaming.

To conclude our preliminary experimental evaluation, we test the capability of the forecasting model to anticipate the demands for mobile services at predictions horizons farther than one future sample, *i.e.*, more than five minutes ahead in our settings. To the best of our knowledge, this is a first-of-its-kind investigation in the context of mobile traffic forecasting, as we could not find any equivalent analysis in the literature.

In the following, we consider so-called *point-wise* relative and absolute errors. These are computed by predicting the mobile traffic volume at different horizons $h \geq 1$, from each time step of the original demand time series and for every combination of MEC facility and mobile service class; then, the mean error of all predictions at a given horizon is calculated. The resulting average point-wise errors are plotted as functions of the prediction horizon in Fig. 6.

As one could expect, both the relative error, in Fig. 6a, and the absolute error, in Fig. 6b, increase as h grows from 5 minutes to two hours: capturing the mobile traffic dynamics father in the future is a harder task in general, as estimation errors cumulate in time. However, quite surprisingly, most of the additional error is accumulated during the first 30 minutes or so, and predictions at longer time intervals do not incur into substantial extra inaccuracy. Also, it is interesting to note that a constant offset between the curves outlines a similar response to increasing horizons when predicting diverse mobile services. Overall, the quality of the prediction remains acceptable for all horizons, with point-wise relative errors below 10% in all cases, and typically lower than 5%.

V. Related work

The body of works on predicting network traffic is very large, yet no previous study considered settings that are specific to anticipatory MEC as we do [17]. In the context of mobile networks, previous works have mainly addressed the problem forecasting traffic aggregate over all services: this is, in general, a simpler problem than anticipating per-service demands. Representative efforts in this direction include characterizations of the predictability of global radio access network (RAN) traffic via information theory [18], Markovian modeling of high, medium or low levels of mobile traffic volumes [19], and time series forecasting using autoregressive models [15], [20], [21] or machine learning approaches [16], [22], [23].

We would also like to mention that one of the cornerstones in the design of our proposed model, *i.e.*, the exploitation of similar datastreams as exogenous inputs, is not completely novel. A close principle was leveraged in a different forecasting contexts. In [24], the goal is predicting future locations of individuals. In that case, a cluster-based Bayesian framework exploited similarities in the mobility of a set of people to improve the movement prediction accuracy of each person. In [25], the objective is reconstructing potentially missing data from a sensor networks, and an ensemble model is leveraged to take advantage of time series correlations. Although the underlying principles in those works are comparable to ours, the methodology to implement them and the nature of the forecasting processes are completely different.

VI. Conclusions

We presented a novel forecasting model based on a MISO ARMAX approach, which is especially well suited for predicting per-mobile service traffic. Our model has the following key features: (i) it is a lean representation that does not need bulky training data, but takes advantage of strong easily-anticipated regular components [20], [26] that characterize mobile traffic demands; (ii) it employs spatially unconstrained exogenous inputs, leveraging the fact that mobile service usage is driven by human activities, which in turn depend on land use [27]. A preliminary performance evaluation with large-scale measurement data collected in an operational mobile network yields promising results: the proposed model achieves high accuracy, which makes it an ideal enabler for anticipatory MEC solutions.

Acknowledgments

This research was partly supported by the European Union Horizon 2020 Framework Programme under REA grant agreement no.778305 DAWN4IoE, and by the European Research Consortium for Informatics and Mathematics (ERCIM).

References

[1] EC H2020 5G Infrastructure PPP, "Pre-structuring model, version 2.0," http://5g- ppp.eu/, April 2014.

[2] NetWorld2020 ETP, "5g: Challenges, research priorities, and recommendations," https://www.networld2020.eu/, August 2014.

[3] R. W. Thomas, L. A. DaSilva, and A. B. MacKenzie, "Cognitive networks," in *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN 2005.*, Nov 2005.

[4] J. Jiang, V. Sekar, I. Stoica, and H. Zhang, *Unleashing the Potential of Data-Driven Networking*. Springer International Publishing, 2017.

[5] N. Bui, M. Cesana, S. A. Hosseini, Q. Liao, I. Malanchini, and J. Widmer, "A survey of anticipatory mobile networking: Context-based classification, prediction methodologies, and optimization techniques," *IEEE Communications Surveys Tutorials*, vol. 19, no. 3, 2017.

[6] H. Lee, S. Vahid, and K. Moessner, "A survey of radio resource management for spectrum aggregation in lte-advanced," *IEEE Communications Surveys Tutorials*, vol. 16, no. 2, 2014.

[7] A. Furno, D. Naboulsi, R. Stanica, and M. Fiore, "Mobile demand profiling for cellular cognitive networking," *IEEE Transactions on Mobile Computing*, vol. 16, no. 3, March 2017.

[8] "Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec," https://www.eugdpr.org/, April 2016.

[9] X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, "Network slicing in 5g: Survey and challenges," *IEEE Communications Magazine*, vol. 55, no. 5, pp. 94–100, May 2017.

[10] A. Strunk, "Costs of virtual machine live migration: A survey," in *Proceedings of the 2012 IEEE Eighth World Congress on Services*, ser. SERVICES '12. IEEE Computer Society, 2012.

[11] L. Ljung, Ed., *System identification (2nd ed.): theory for the user*. Prentice Hall PTR, 1999.

[12] C. Marquez, M. Gramaglia, M. Fiore, A. Banchs, C. Ziemlicki, and Z. Smoreda, "Not all apps are created equal: Analysis of spatiotemporal heterogeneity in nationwide mobile service usage," in *Proceedings of the 13th ACM International Conference on emerging Networking EXperiments and Technologies*, ser. CoNEXT '17, 2017.

[13] G. E. P. Box and G. Jenkins, *Time Series Analysis, Forecasting and Control*. Holden-Day, Incorporated, 1990.

[14] C. Ferri, J. Hernndez-Orallo, and R. Modroiu, "An experimental comparison of performance measures for classification," *Pattern Recognition Letters*, vol. 30, no. 1, 2009.

[15] F. Xu, Y. Lin, J. Huang, D. Wu, H. Shi, J. Song, and Y. Li, "Big data driven mobile traffic understanding and forecasting: A time series approach," *IEEE Trans. on Services Computing*, vol. 9, no. 5, 2016.

[16] J. Wang, J. Tang, Z. Xu, Y. Wang, G. Xue, X. Zhang, and D. Yang, "Spatiotemporal modeling and prediction in cellular networks: A big data enabled deep learning approach," in *Proceedings of the 36th Annual IEEE International Conference on Computer Communications*, ser. INFOCOM'17, May 2017.

[17] M. Joshi and T. H. Hadi, "A review of network traffic analysis and prediction techniques," *CoRR*, vol. abs/1507.05722, 2015.

[18] R. Li, Z. Zhao, X. Zhou, J. Palicot, and H. Zhang, "The prediction analysis of cellular radio access network traffic: From entropy theory to networking practice," *IEEE Communications Magazine*, vol. 52, no. 6, June 2014.

[19] M. Z. Shafiq, L. Ji, A. X. Liu, and J. Wang, "Characterizing and modeling internet traffic dynamics of cellular devices," in *Proceedings of the ACM SIGMETRICS Joint International Conference on Measurement and Modeling of Computer Systems*, ser. SIGMETRICS '11, 2011.

[20] M. Zhang, H. Fu, Y. Li, and S. Chen, "Understanding urban dynamics from massive mobile traffic data," *IEEE Transactions on Big Data*, 2017.

[21] S. Tom Au, G.-Q. Ma, and S.-N. Yeung, "Automatic forecasting of double seasonal time series with applications on mobility network traffic prediction," in *Joint Statistical Meetings*, 2011.

[22] A. Y. Nikravesh, S. A. Ajila, C.-H. Lung, and W. Ding, "An experimental investigation of mobile network traffic prediction accuracy," *International Journal of Big Data*, vol. 3, 2017.

[23] C. Zhang and P. Patras, "Long-term mobile traffic forecasting using deep spatio-temporal neural networks," in *Proceedings of the 19th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, ser. MobiHoc'18, June 2018.

[24] J. Jeong, M. Leconte, and A. Proutiere, "Cluster-aided mobility predictions," in *Proceedings of the 35th Annual IEEE International Conference on Computer Communications*, ser. INFOCOM'16, April 2016.

[25] C. Alippi, S. Ntalampiras, and M. Roveri, "Model ensemble for an effective on-line reconstruction of missing data in sensor networks," in *IEEE International Joint Conference on Neural Networks*. IEEE, 2013.

[26] B. Cici, M. Gjoka, A. Markopoulou, and C. T. Butts, "On the decomposition of cell phone activity patterns and their connection with urban ecology," in *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, ser. MobiHoc '15, 2015.

[27] A. Furno, M. Fiore, R. Stanica, C. Ziemlicki, and Z. Smoreda, "A tale of ten cities: Characterizing signatures of mobile traffic in urban areas," *IEEE Transactions on Mobile Computing*, vol. 16, no. 10, Oct 2017.