Modelling Throughput Prediction Errors as Gaussian Random Walks

Nicola Bui[12], Joerg Widmer[1]

[1]IMDEA Networks Institute, Leganes (Madrid), Spain
[2]UC3M, Leganes (Madrid), Spain

***Abstract*—One of the most critical aspects of anticipatory networking is assuming that future system conditions can be estimated. In this paper we address how accurate the current state of the art predictors are in providing a forecast of short term throughput. We propose a simple model for the short term prediction error based on Gaussian Random Walks that allows for mathematical analysis of the impact of imperfect future knowledge on network optimization.**

I. Introduction

Recent interest in anticipatory networking is motivated by the increasing volume of mobile data traffic and the need for energy efficiency in mobile networks [1]. In particular, the work in [2]–[4] discusses how future network state information can be obtained and exploited to add a temporal dimension to network optimization problems.

In addition, the dynamics of cellular data networks have been thoroughly studied in many works such as those by Paul [5] or Shafiq [6], which show the link between network performance and user mobility as well as other intrinsic correlated aspects that advocate prediction feasibility.

In our previous contribution [7], we analyzed the state of the art in mobile throughput and user mobility prediction in order to derive a composite model for prediction error. Short term prediction, shown in Fig. 1 on the left, is most often based on time series filtering techniques [8], [9], while medium and long term prediction, shown on the right, is usually derived from mobility aspects and networks dynamics.

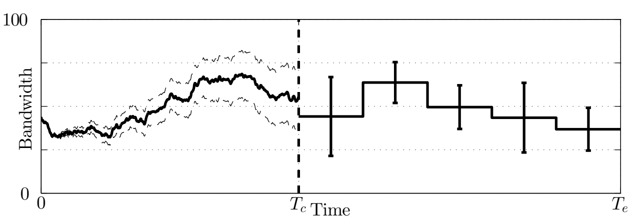

Fig. 1. Capacity availability prediction uncertainties: short term predictors are useful until time T_c, while the medium term model is used until slot T_e.

In this paper, we will focus on predictors represented in the left part of the figure, where the solid line represent a possible trace of throughput evolution. The dashed lines show the boundary of the region the prediction is likely to fall in. The short term predictors start to be useless at time T_c when the prediction error is as big as that obtained by randomly drawing the next samples from the statistic distribution of the original phenomenon. Statistical distribution can be used from T_c until T_e, the time when no statistical considerations can be derived from mobility prediction.

In particular, we model the short term prediction error of a Gaussian random walk, which provides a close fit to the original random process and allows for a simple mathematical analysis of the impact of imperfect prediction on network optimization.

In Section II we discuss the model use to derive mobile networks characteristics and the filtering technique we used for prediction. Section III gives details on prediction error and present our Gaussian random walk model. Section IV concludes the paper.

II. System model

This section focuses on the prediction of downlink rate between base station (eNodeB) and user equipment (EU). User throughput in mobile networks depends on several aspects and it is most always modeled as a function of the signal to interference plus noise ratio (SINR) γ and the number of active users in the cell K. The SINR is usually modeled as a function of the distance d between eNodeB and UE, the K active users and the scheduler type.

In what follows we specifically address LTE technology and the related throughput model proposed by Østerbø [10] for the case of proportional fair scheduler and the K users uniformly distributed in the cell coverage area. According to this model, the throughput g can be expressed as:

$$g = \eta(\gamma_0 d^{-\alpha} r(K))/K, \tag{1}$$

where $\eta(x)$ is a piece-wise constant function associating throughput to SINR ranges, γ_0 is a constant scaling factor related to environmental and system parameters (e.g., transmit power, antenna gains, etc.), $\alpha \in [2, 4]$ is the exponent of the pathloss law and $r(K)$ is the fast fading gain and depends on K to model opportunistic gain achieved by the scheduler.

A throughput value obtained from Eq. (1) has a coherence time T_f which is inversely proportional to the user movement speed s [11]. Thus, we average $\lceil T_s/T_f \rceil$ throughput values to filter fast fading variations.

In order to obtain user movement traces we let the user move with constant speed and direction in an area where cells are randomly placed. In particular, the position of each eNodeB along the user path is chosen so that the maximum distance between the UE and the closest eNodeB is never larger than a given communication range. Every T_s seconds the UE-eNodeB

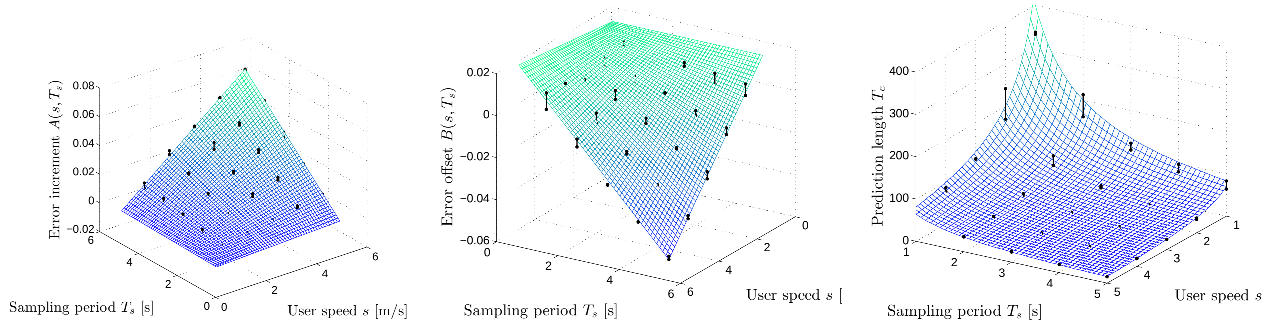

Fig. 2. Comparison between the collected data and the fitted model varying s and T_s. Starting from the left, the figures show the approximation of the A (left), B (center) and T_c (right) parameters as surfaces and the distance from the surface to the actaul data as lines.

distance is measured as the distance between the UE and the closest eNodeB in the area. The eNodeB random placement is equivalent to assume random variation in the user speed and constant distance between eNodeBs. In order to contain the dimensionality of the problem, in this paper we only study movement sequences characterized by constant speed and direction.

For any given tuple of parameters (s, T_s) we can generate any number of sequences $D(s, T_s) = d_i, i \in [1, T_l]$ of any length T_l. Subsequently, we can generate throughput sequences $G(s, T_s, K) = g_i, i \in [1, T_l]$, where g_i is obtained by averaging $\lceil T_s/T_f \rceil$ values obtained from d_i through Eq. (1).

For what concerns prediction itself, we limited our focus on autoregressive and moving average (ARMA) filters. We choose this technique, because it is well studied and it is simple to implement in mobile phones. The basic ARMA model is as follows:

$$X_i = c + \varepsilon_t + \sum_{j=1}^{p} \varphi_j X_{i-j} + \sum_{k=1}^{q} \theta_k \varepsilon_{i-k}, \qquad (2)$$

where c is a constant, ε_i are white noise error terms, φ_j, θ_k, p and q are the autoregressive and the moving average coefficients and their respective orders and X_i is the reference signal. This model is referred to as an ARMA(p, q) with reference to the order of the two parts of the filter. To determine the order to be used, we followed the Box-Jenkins method [12] using automatic inspection of autocorrelation and sampled partial autocorrelation functions.

To generate error sequences and their statistics we operate as follows: first we obtain the optimal order of the ARMA filters to be used and, for each tuple of speed and sampling period (s, T_s), we generate a single very long training sequence $G_T(s, T_s, K)$ from which we tune the filter coefficients; subsequently, we generate shorter throughput sequences $G_i(s, T_s, K), i \in [1, 100]$ to test the filter on. In particular, we obtain filters $F(s, T_s, K)$ from $G_T(s, T_s, K)$ and we use the filters to predict the sequences $\tilde{G}_{ij}(s, T_s, K) = \tilde{g}_{ijk}, i \in [1, 100], j \in [1, 100], k \in [\max\{p, q\} + j, \max\{p, q\} + T_p + j]$ or, in other words, from each of the 100 sequences we generate 100 predicted sequences starting at different points and long T_p values. Finally, we compute errors $e_{ijk} = \tilde{g}_{ijk} - g_{ik}$ and the error sequences $E_{ij}(s, T_s, K)$ from which we further obtain the sequences $\sigma_k^2(s, T_s, K) = \mathrm{E}[(e_{ijk} - \mu)^2]/\sigma_G^2$, which represent the variance of the k-th prediction error normalized to the variance σ_G^2 of the original training signal $G_T(s, T_s, K)$.

III. Prediction error model

This section proposes to use a Gaussian random walk to approximate the sequences $E_{ij}(s, T_s, K)$. Gaussian random walks are interesting, because their total variance at time t is proportional to the interval duration and they can be expressed as a sum of i.i.d Gaussian random variables.

Before approaching the fitting of the model itself, we verified that assuming the error sequences to be drawn from zero mean normal distribution was a valid hypothesis. To do so, we perform the Kolmogorov-Smirnov [13] test between the generated error sequences and theoretical normal distributions with zero mean and the same variance as the error sequences. All the tests performed rejected the null hypothesis according to which the error and the normal distributions are not equal.

Subsequently, by visual inspection of the $\sigma_k^2(s, T_s, K)$ we noticed that: *i*) it increases with the prediction distance k, *ii*) the steepness is increasing with both s and T_s; *iii*) the minimum error is decreasing with both s and T_s; *iv*) T_c can be obtained as the minimum k so that $\sigma_k^2(s, T_s, K) = 1$; *v*) the number of active users K has a negligible impact on the prediction error.

Thus we are looking for a family of linear equations that approximates the variance sequence:

$$\sigma_k^2(s, T_s) = \begin{cases} A(s, T_s)k + B(s, T_s) & k \le T_c/T_s \\ 1 & \text{otherwise} \end{cases}, \qquad (3)$$

where $A(s, T_s)$ represent the steepness and $B(s, T_s)$ the offset of the process or, in other words, how fast the prediction reliability decreases and how large is the intrinsic randomness of the process respectively. According to *ii*) and *iii*) we fit two linear functions on the sT_s product to approximate A and B respectively and we obtain:

$$\begin{aligned} A(s, T_s) &= A_1 s T_s + A_2 \\ B(s, T_s) &= B_1 s T_s + B_2. \end{aligned} \qquad (4)$$

Fig. 2 shows how close the model fits the data. The model coefficients have been obtained from the original sequences by imposing a perfect match for $s = 1$ and $T_s = 1$ and minimizing the least square error in the other points. In particular, Fig. 2(a) and 2(b) show the surfaces obtained from Eq. (4) and the distance from the actual data and the surfaces. Also, Fig. 2(c) show the prediction validity length derived from the model $T_c = T_s(1 - B(s, T_s))/A(s, T_s)$ (surface) compared to the same obtained from the data.

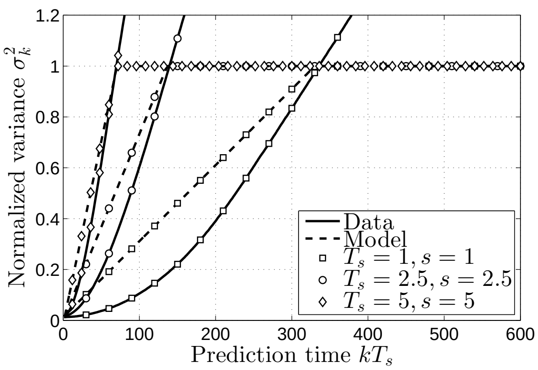

Fig. 3. Comparison between model and data for different s, T_s couples.

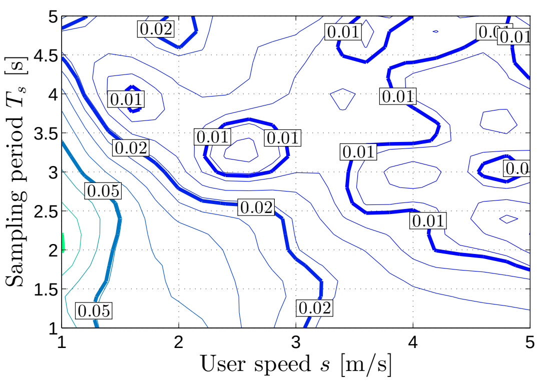

Fig. 4. Contour plots of the average distance between the model and the data.

Fig. 3 visualizes how the model fits the data for a few speed-sampling time couples: solid and dashed lines represent the normalized variance σ_k^2 obtained from the data and from the model respectively; square, diamond and circle markers identify the (s, T_s) couple as $(1, 1)$, $(2.5, 2.5)$ and $(5, 5)$ respectively. In all the three cases the model fits the curves reasonably well and is always providing a conservative approximation: the predicted error is always larger than obtained from actual data.

Fig. 4 shows contour plots of the average approximation error. Bold lines are marked with the actual error value, which is most always smaller than 2 %, but for $1 \leq T_s \leq 3$ and $s < 1.5$ where it is slightly larger than 5 %. This is mainly due to two effects: the randomness of the original signal is higher and the linear fitting is less appropriate for small s and T_s as a consequence of the stronger impact of fast fading and a slower prediction reliability degradation respectively.

Finally, we conclude that Gaussian random walks can be used as a valid model for short term prediction errors since they can reproduce the main characteristics of the original random process. Also, random walks allow for an easier analysis of prediction based optimization problems: in fact, it is possible to approximate the distribution of the prediction error as a sum zero mean Gaussian variables: one of variance $B(s, T_s)$ accounting for the sequence inherent randomness and k with variance $A(s, T_s)$ each to account for decreasing reliability of the prediction after k steps.

Hence, since the model is conservative with respect to the uncertainty introduced by imperfect prediction, optimization algorithms' performance obtained through this approximation are conservative as well. Thus it will be possible to derive optimization algorithms leveraging on the prediction reliability in order to mitigate the effects of uncertainties.

IV. Conclusion

This paper analyzed the short term prediction error for throughput sequences and proposed an approximated model based on Gaussian random walks. The model provides a compact description of prediction error based on user mobility parameters and prediction time.

Due to its simplicity, the model allows for closed form mathematical analysis of resource allocation optimization problem exploiting throughput prediction. Our current investigation topics include the model validation real data and its exploitation to derive performance bounds for resource allocation algorithm working with imperfect knowledge.

Acknowledgment

The research leading to these results was partly funded by the European Union under the project eCOUSIN (EU-FP7-318398).

References

[1] S. Wang, Y. Xin, S. Chen, W. Zhang, and C. Wang, "Enhancing spectral efficiency for lte-advanced and beyond cellular networks [guest editorial]," *IEEE Wireless Communications*, vol. 21, no. 2, pp. 8–9, April 2014.

[2] Z. Lu and G. de Veciana, "Optimizing stored video delivery for mobile networks: The value of knowing the future," in *IEEE INFOCOM*, Turin, Italy, April 2013, pp. 2706–2714.

[3] Q. Xu, S. Mehrotra, Z. Mao, and J. Li, "PROTEUS: network performance forecast for real-time, interactive mobile applications," in *ACM MobiSys*, Taipei, Taiwan, June 2013, pp. 347–360.

[4] H. Abou-zeid, H. Hassanein, and S. Valentin, "Optimal predictive resource allocation: Exploiting mobility patterns and radio maps," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, 2013, pp. 4714–4719.

[5] U. Paul, A. P. Subramanian, M. M. Buddhikot, and S. R. Das, "Understanding traffic dynamics in cellular data networks," in *INFOCOM, 2011 Proceedings IEEE*. IEEE, 2011, pp. 882–890.

[6] 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. New York, NY, USA: ACM, 2011, pp. 305–316. [Online]. Available: http://doi.acm.org/10.1145/1993744.1993776

[7] N. Bui, F. Michelinakis, and J. Widmer, "A model for throughput prediction for mobile users," in *European Wireless*, Barcelona, Spain, May 2014.

[8] N. Sadek and A. Khotanzad, "Multi-scale high-speed network traffic prediction using k-factor gegenbauer arma model," in *IEEE ICC*, vol. 4, Paris, France, June 2004, pp. 2148–2152.

[9] Y. Qiao, J. Skicewicz, and P. Dinda, "An empirical study of the multiscale predictability of network traffic," in *IEEE HDPC*, Honolulu, Hawaii USA, June 2004, pp. 66–76.

[10] O. Østerbø, "Scheduling and capacity estimation in lte," in *Teletraffic Congress (ITC), 2011 23rd International*, 2011, pp. 63–70.

[11] T. S. Rappaport *et al.*, *Wireless communications: principles and practice*. Prentice Hall PTR New Jersey, 1996, vol. 2.

[12] S. Makridakis and M. Hibon, "Arma models and the boxjenkins methodology," *Journal of Forecasting*, vol. 16, no. 3, pp. 147–163, 1997.

[13] H. W. Lilliefors, "On the kolmogorov-smirnov test for normality with mean and variance unknown," *Journal of the American Statistical Association*, vol. 62, no. 318, pp. 399–402, 1967.