ATM Dynamic Bandwidth Allocation Using F-ARIMA Prediction Model

Nayera Sadek, Alireza Khotanzad, and Thomas Chen
Department of Electrical Engineering
Southern Methodist University
Dallas, TX. 75275-0338 USA
nsadek@engr.smu.edu, kha@engr.smu.edu, tchen@engr.smu.edu

***Abstract*— Measurements of high-speed network traffic have shown that traffic data exhibits a high degree of self-similarity. Traditional traffic models such as AR and ARMA are not able to capture this long-range-dependence making them ineffective for the traffic prediction task. In this paper, we apply the fractional ARIMA (F-ARIMA) model to predict one-step-ahead traffic value at different time scales. F-ARIMA has the ability to capture both the short- and long-range dependent characteristics of the underlying data. We present a simplified adaptive prediction scheme to reduce the F-ARIMA computational complexity. The performance of the proposed F-ARIMA prediction model is tested on four different types of traffic data: MPEG and JPEG video, Ethernet and Internet. We also apply the F-ARIMA prediction model to a dynamic bandwidth allocation scheme. The results show that the performance of F-ARIMA outperforms the AR models. They also show that the prediction performance depends on the traffic nature and the time scale.**

Keywords- traffic prediction; self-similar; F-ARIMA; dynamic bandwidth allocation.

I. INTRODUCTION

High-speed network traffic measurements have shown that the traffic has a self-similar characteristic [1], [2]. The main feature of the self-similarity is that its autocorrelation function decays hyperbolically instead of exponentially as traditional traffic models, e.g. autoregressive (AR) and autoregressive moving average (ARMA). Thus, the traditional models are not able to capture the self-similar characteristic [1]. On the other hand, fractional autoregressive integrated moving average (F-ARIMA) is a self-similar model, and it has the ability to capture both the short-range dependent (SRD) and long-range dependent (LRD) characteristics. As such, this model would be useful as traffic predictor. In [6], F-ARIMA has been used in modeling Ethernet and video traffic. In [7], F-ARIMA parameters are estimated and used to predict the Ethernet traffic. A simplified scheme to estimate these parameters and build the model is also presented in [6], [7].

In ATM, traffic prediction is considered as the core of the preventive congestion control schemes such as the connection admission control (CAC) problems, and available bit rate (ABR) traffic rate control. CAC schemes are applied to decide whether to accept or reject new connection requests [10]. When the network accepts a new request, it allocates enough network resource to satisfy the required quality of service (QoS) without violating the QoS guaranteed to the existing connections. Bandwidth allocation approaches can be divided into two categories: static and dynamic [12], [13]. In static approaches, the allocated bandwidth remains constant during the connection lifetime. In dynamic allocation, the allocated bandwidth can be changed to increase the network utilization. Static allocation is effective for the constant bit rate (CBR) traffic service. Dynamic allocation is suitable for the variable bit rate (VBR) traffic service whose statistical characteristics may vary over time [12], [13]. It is difficult to allocate an optimal static amount of bandwidth for VBR traffic at the time when an admission decision is to be made.

F-ARIMA prediction model has been used to build a CAC scheme [8], and a dynamic bandwidth allocation scheme [9]. In this dynamic allocation scheme, the predicted traffic values are used directly regardless of the buffer length, and that may result in buffer overflow.

In this paper, we propose a simplified adaptive F-ARIMA based prediction scheme that avoids the computational complexity. The adaptation scheme gives the model the ability to keep track of the traffic changes and growth. We apply the F-ARIMA model to predict one-step-ahead traffic value at different time-scales. The performance of the proposed schemes is tested on four different types of traffic data, MPEG and JPEG video, Ethernet and Internet. We show also the application of the F-ARIMA model to the dynamic bandwidth allocation scheme proposed in [12]. This scheme limits the buffer length to a certain length, so it avoids the buffer overflow and cell loss. It can increase the network utilization and avoid congestion.

II. F-ARIMA PREDICTION MODEL

A. F-ARIMA Parameters Estimation

The fractional autoregressive integrated moving average model of order (p,d,q), denoted as F-ARIMA (p,d,q), is defined as:

$$\phi(B)\nabla^d x(n) = \theta(B)e(n) \qquad (1)$$

where: B is a lag operator, $x(n\text{-}1) = Bx(n)$.

$\phi(B)$ and $\theta(B)$ are two polynomial functions of degree p and q respectively, defined as:

$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$, and

$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$

All the zeros of ϕ(B), and θ(B) are outside the unit circle.

d is the differential factor, and it is calculated from the Hurst parameter H, where $d = H - 0.5$

∇^d is the fractional difference operator defined as:

$$\nabla^d = (1-B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k B^k \text{ , and}$$

$$\binom{d}{k} = \frac{d!}{k!(d-k)!} = \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)}$$

where Γ denotes the gamma function.

The F-ARIMA parameters are estimated from the historical traffic data. These parameters can be divided into two categories. The first category contains the d factor that presents the long-range dependent (LRD) property. The second category contains the ARMA parameters, p, q, $\phi(B)$, $\theta(B)$, and the variance of $e(n)$, that present the short-range dependent (SRD) property.

To compute the d factor, the Hurst parameter H is estimated using the Abry-Veitch (AV) wavelet method [3]. This estimation method provides an unbiased estimate of the H parameter with high robustness and low computational cost. For most traffic data, the estimated H parameter falls in the range of $H > 0.5$, so d will be in the range $0 < d < 0.5$ [1], [2], [3].

To estimate the ARMA parameters, we can convert the F-ARIMA process to an ARMA process through the following transformation [6], [7]: $w(n) = \nabla^d x(n)$.

The following steps are used to estimate the ARMA parameters [6], [7]:

1- Get a zero-mean traffic data $(x(n) - \mu)$, where $\mu = E[x(n)]$ with E denoting the expected value.
2- Apply the fractional difference operator: $w(n) = \nabla^d (x(n) - \mu)$, $w(n)$ is ARMA (p,q)
3- Determine p, and q by using the cross-validation scheme [14]. The values of p and q are small values, in the range of 0, 1, 2, and 3.
4- Estimate the parameters, $\phi(B)$, $\theta(B)$ and the variance of $e(n)$ [15]. Check to ensure that all the zeros of $\phi(B)$, and $\theta(B)$ are outside the unit circle.

In the cross-validation scheme [14], the training data is divided into modeling and validation sets. The modeling set is used to estimate the parameters of the ARMA models for different combinations of p and q. These parameters are used to predict the traffic in the validation set, and their performance is measured. That p and q combination that gives the best performance is selected. This scheme gives the model a better generalization ability.

B. F-ARIMA Prediction Scheme

For $0 < d < 0.5$, $x(n)$ is a stationary and invertible process, so the one-step ahead predicted value, denoted as $\hat{x}_{FARIMA}(n)$, can be estimated by [4], [7]:

$$\hat{x}_{FARIMA}(n) = \sum_{j=0}^{\infty} \psi_j e(n-j) \qquad (2)$$

where: $\sum_{j=0}^{\infty} \psi_j B^j = \theta(B)\phi^{-1}(B)\nabla^d$

In [7], they apply (2) to propose an adaptive prediction method to provide an upper probability limit. In this paper, we propose a simplified adaptive F-ARIMA prediction scheme. This schemes depends on the decomposition of the F-ARIMA(p,d,q) as F-ARIMA$(0,d,0)$ [5]. The basic idea is to predict the F-ARIMA$(0,d,0)$, and convert the predicted process to F-ARIMA(p,d,q). In this scheme, the order parameters, p, and q, remain fixed. We develop the following steps to predict the one-step-ahead traffic value and update the parameters easily:

1- Compute $y(n) = \{\theta(B)\}^{-1}\phi(B)x(n)$, $y(n)$ is F-ARIMA$(0,d,0)$.
2- Compute the best linear predictor of $y(n)$, by applying the following equation [4]:
$$\hat{y}(n) = \sum_{j=1}^{k} \beta_{kj} y(n-j),$$
where $\beta_{kj} = -\binom{k}{j} \cdot \frac{\Gamma(j-d)\Gamma(k-d-j)}{\Gamma(-d)\Gamma(k-d+1)}$
Theoretically, large values of k yield better results. In our implementation, k is adjusted according to the available data and to avoid any divergence.
3- Compute the predicted value $\hat{x}(n) = \theta(B)\{\phi(B)\}^{-1}\hat{y}(n)$.
4- Once the actual traffic value becomes available, update the historical data.
5- Estimate the parameters of $\phi(B)$, and $\theta(B)$.

6- Check for the convergence condition and check for all the zeros of $\phi(B)$, and $\theta(B)$ to be outside the unit circle. If all the conditions are satisfied, then go to step 1 and predict the next value.
If the parameters do not satisfy the conditions, then we estimate the H parameter using AV wavelet algorithm. Then, using the same p, and q, we estimate the parameters of $\phi(B)$, and $\theta(B)$.

To extend the horizon, the k-step-ahead predicted traffic value can be computed recursively [11]. For example, to obtain the two-step-ahead traffic value, the one-step-ahead predicted value is computed first. Then it is used with the other lagged traffic values to compute the two-step-ahead predicted value. This procedure is repeated to generate any further k predicted values. To update the F-ARIMA parameters, once the actual traffic values become available, the adaptive part of the proposed prediction scheme is applied (steps 4 to 6).

By increasing the value of k, we find that the recursive k-step-ahead traffic prediction may cause prediction error accumulation, especially when k becomes larger (k > 5). To avoid this error, we apply the F-ARIMA model to predict one-step-ahead traffic value at different time scales separately.

III. ATM DYNAMIC BANDWIDTH ALLOCATION

For the VBR traffic, static bandwidth allocation may result in low utilization. Dynamic bandwidth allocation can achieve better utilization, provided that the VBR can be predicted accurately. To evaluate the F-ARIMA prediction algorithm for this application, we model an ATM link as a first-in-first-out (FIFO) queue with finite length Q. The occupancy of the queue at any time n is given as $q(n)$. The lagged traffic values are used to predict the one-step-ahead traffic value, denoted as $\hat{x}(n)$. In [12], the actual allocated bandwidth is given by:

$$A(n) = \hat{x}(n)\left(1 + \frac{q(n-1) - \gamma.Q}{Q}\right) \tag{3}$$

where γ is the desired occupation ratio with $0 < \gamma < 1$.

The procedure for the dynamic bandwidth allocation is [12]:

1- Measure the current queue length $q(n\text{-}1)$.
2- Predict the traffic bandwidth $\hat{x}(n)$, compute the actual allocated bandwidth $A(n)$.
3- When $x(n)$ becomes available, update the F-ARIMA parameters.
4- Increment t, and go to step 1.

This algorithm is designed to keep the buffer occupancy around the value of γQ, so it avoids the buffer overflow and cell loss. It increases the network utilization and avoids congestion.

IV. EXPERIMENTAL RESULTS

The performance of the proposed F-ARIMA prediction model is tested on four different types of traffic data, MPEG and JPEG video, Ethernet and Internet. The used video traffic is the MPEG and JPEG version of the "Star Wars" movie that is available at [16]. The Telcordia (formerly Bellcore) Ethernet and Internet traffic are available at [17]. All the traffic data are processed to present number of packets (or cells) per unit time. Then they are aggregated at different time scales, e.g. 1, 5, 10 seconds. The traffic data is divided into two sets: training and testing. The training set is used to apply the cross-validation scheme and to estimate the models' parameters. Then these parameters are tested on the testing set ensuring blind test.

The utilized performance measure is the prediction signal to noise ratio (SNR) given as:

$$\mathrm{SNR} = 10\log_{10}\left(\frac{E(x(n)^2)}{E((x(n)-\hat{x}(n))^2)}\right) \quad \mathrm{dB} \tag{4}$$

where $E(.)$ is the expected value, $x(n)$ is the actual traffic value, and $\hat{x}(n)$ is the predicted traffic value. As the prediction accuracy increases, the prediction SNR becomes higher.

We compare the performance of the F-ARIMA model to the AR model. The order of AR (denoted as p) is estimated using the cross-validation scheme as mentioned in section II.A. The parameters of AR (denoted as $\phi(B)$) are estimated as in [15].

All the models use the lagged traffic values to predict one-step-ahead traffic value at different time scales. The performance results in terms of SNR are presented in Table I. The first column shows the time scale (TS) in sec. For each traffic data, we present the SNR of applying AR and F-ARIMA model. The best results are highlighted in bold.

Figs. 1 to 4 show samples of the results in graphical form. The actual traffic values of the MPEG, and JPEG video, Ethernet and Internet data at 1 sec are shown as solid line. The corresponding predicted values are superimposed as dotted line.

The results show that the performance of the F-ARIMA model outperforms the AR model. From Table I, we notice that the time-scale and the nature of the traffic play a significant role in the traffic prediction performance.

For small time-scale, the difference between F-ARIMA and AR model is small. For larger time-scales, the difference increases by about 1 to 2 dB depending on the nature of the traffic data.

TABLE I. PREDICTION SNR OF DIFFERENT TRAFFIC DATA

TS (sec)	MPEG		JPEG		Ethernet		Internet	
	AR	*F-ARIMA*	*AR*	*F-ARIMA*	*AR*	*F-ARIMA*	*AR*	*F-ARIMA*
1	13.2852	**13.4872**	23.076	**23.6735**	12.1476	**12.4199**	12.4135	**12.6893**
5	13.1059	**13.9029**	18.9845	**19.9448**	13.8275	**14.6888**	13.8813	**14.8311**
10	13.9234	**15.0434**	17.7972	**19.732**	13.4159	**14.605**	14.0734	**15.0198**

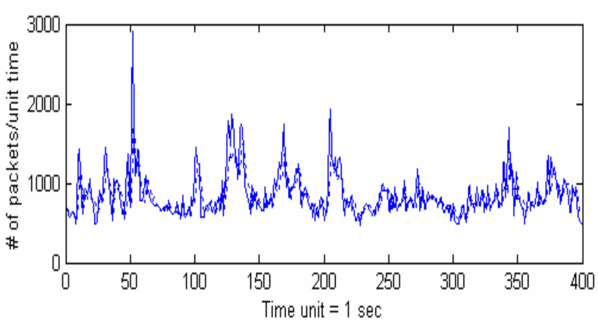

Figure 1. Actual and predicted MPEG video traffic

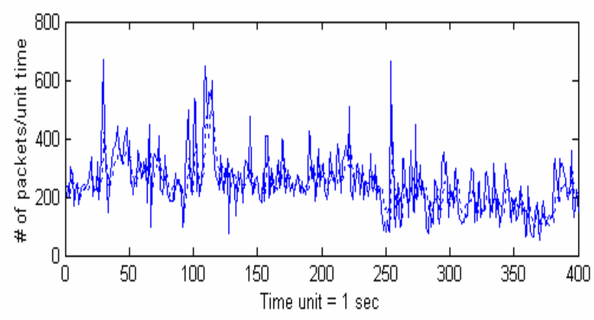

Figure 3. Actual and predicted Ethernet traffic

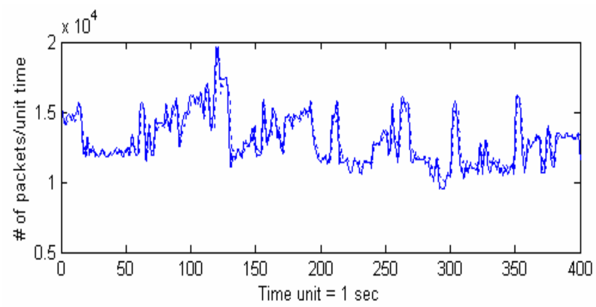

Figure 2. Actual and predicted JPEG video traffic

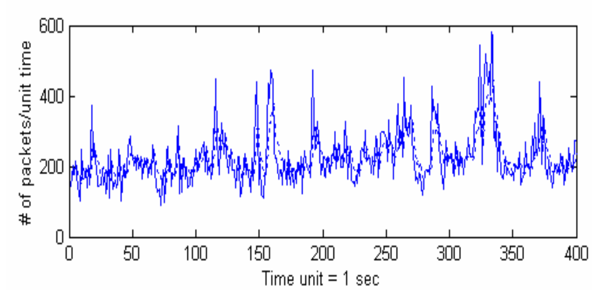

Figure 4. Actual and predicted Internet traffic

For the JPEG traffic data, as the time-scale increases, the SNR decreases. The used video traffic, “Star Wars”, is an action movie, and its scenes change very quickly. Thus, it is not easy to predict the amount of traffic over long time-scale. For the Internet traffic data, as the time-scale increases, the prediction performance increases. This result can be explained in term of the user’s habit of using the Internet. For example, during the day (longer time-scale) it is easy to predict when and how long the user will access the Internet, and how much traffic is expected over this period from user’s traffic history. From Table I, it is difficult to arrive at an explicit relation between the time scale and the prediction performance for the MPEG video, and Ethernet traffic.

The performance of the dynamic bandwidth allocation is tested on two different video traffic data, MPEG and JPEG. This traffic data is aggregated at time-scale 1 sec. In this experiment, γ is assigned to 0.5 implying that the buffer should be half-occupied all times.

Figs. 5 and 6 show the ratio of buffer occupation when we apply the F-ARIMA prediction model. From these Figs., we notice that the ratio of buffer occupation is around 0.5.

To compare the performances of the AR, and F-ARIMA models, Table II shows the mean and variance of the buffer length distribution when we apply these different models. From this table, we notice that the F-ARIMA model gives lower variance compared to the AR model. F-ARIMA has the ability to capture both the SRD and LRD traffic characteristics, while AR is able to capture the SRD characteristics only.

V. CONCLUSIONS

High-speed network traffic has self-similarity. The traditional traffic prediction models are not able to capture this characteristic. F-ARIMA is a self-similar model and it is able to capture both the SRD and LRD characteristics. The proposed F-ARIMA prediction scheme is adaptive, so the model is able to keep track with any traffic changes. We apply it to predict one-step-ahead traffic value at different time-scale. It is tested on four different traffic data, MPEG, and JPEG video, Ethernet and Internet. We compare the performance of F-ARIMA model to the AR model in terms of prediction SNR. The results show that the F-ARIMA model outperforms the other model. We also notice that the time scale and the nature of the traffic play important role in the prediction performance.

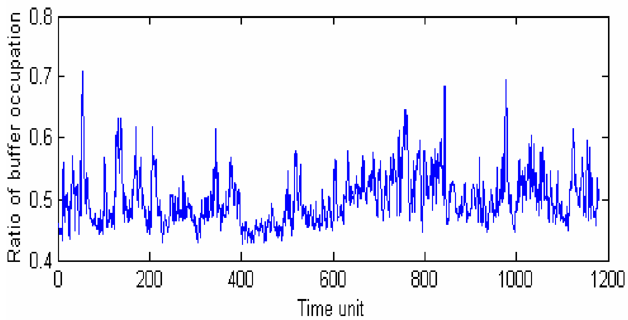

Figure 5. Ratio of buffer occupation using MPEG video traffic data

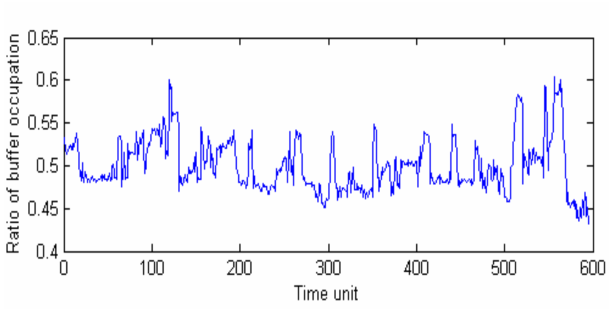

Figure 6. Ratio of buffer occupation using JPEG video traffic data

TABLE II. PARAMETERS OF BUFFER LENGTH DISTRIBUTION

Data	**MPEG**		**JPEG**	
	AR	***F-ARIMA***	***AR***	***F-ARIMA***
Mean	0.51	0.5	0.507	0.5
variance	0.07	0.038	0.019	0.0096

The proposed F-ARIMA model is also applied to a dynamic bandwidth allocation scheme. The used scheme forces the queue length to hover around any desired value, so it avoids buffer overflow and cell loss. It also can increase the network utilization and avoid congestion. We compare the performance of F-ARIMA to the AR models in terms of the variance around the desired queue length value. The F-ARIMA prediction model results in smaller variance.

REFERENCES

[1] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the self-similar nature of Ethernet traffic (extended Version)", *IEEE/ACM Trans. Net.*, vol. 2, no. 1, pp. 1-14, February 1994.

[2] J. Beran, R. Sherman, M. S. Taqqu, and W. Willinger, "Long-range dependence in variable-bit-rate video traffic", *IEEE Trans. Comm.*, vol. 34, no. 2/3/4, pp. 1566-1579, February/March/April 1995.

[3] P. Abry and D. Veitch, "Wavelet analysis of long-range-dependent traffic', *IEEE Trans. Info. Theory*, vol. 44, no. 1, pp. 2-15, January 1998.

[4] J. R. M. Hosking, "Fractional differencing", *Biometrika*, vol. 68, no. 1, pp. 165-176, 1981.

[5] J. R. M. Hosking, "Modeling Persistence in hydrological time series using fractional differencing", *Water Resources Research*, vol. 20, no. 12, pp. 1898-1908, December 1984.

[6] M. Corradi, R. G. Garroppo, S. Giordano and M. Pagano, "Analysis of f-ARIMA processes in the modeling of broadband traffic", *ICC'01*, vol. 3, pp. 964-968, 2001.

[7] Y. Shu, Z. Jin, L. Zhang, L. Wang and O. W. W. Yang, "Traffic prediction using FARIMA models", *ICC'99*, vol. 2, pp. 891-895, 1999.

[8] Y. Shu, Z. Jin, J. Wang and O. W. W. Yang, "Prediction-based admission control using FARIMA models", *ICC'00*, vol. 2, pp. 1325-1329, 2000.

[9] Z. Jin, Y. Shu, J. liu and O. W. W. Yang, "Prediction-based network bandwidth control", *Canadian Conference on Elec. And Comp. Eng.*, vol. 2, pp. 675-679, 2000.

[10] K. Shiomoto, N. Yamanaka, and T. Takahashi, "Overview of measurement-based connection admission control methods in ATM networks", *IEEE Comm. Surveys*, pp. 2-13, First Quarter 1999.

[11] A. Khotanzad, and H. Elragal, "Natural gas load forecasting with combination of adaptive neural networks", *IJCNN'99*, vol. 6, pp. 4096-4072, 1999.

[12] Z. Qing, W. Jingcheng, and X. Hongsheng, "Dynamic bandwidth allocation over ATM based on neural network prediction: Analysis and simulation", http://www4.ncsu.edu:8030/~qzhang4/Docs/Bandwidth.pdf

[13] T. Chan, W. Lau, and V. O. K. Li, "A measurement-based congestion alarm for self-similar traffic", *ICC'01*, vol. 5, pp. 1528-1533, 2001.

[14] A. Khotanzad, R. Afkhami-Rohani, and D. Maratukulam, "ANNSTLF-Artificial neural network short-term load forecaster- Generation three", *IEEE Trans. Power Sys.*, vol. 13, no. 4, pp. 1413-1422, November 1998.

[15] G. E. Box, G. M. Jenkins, *Time Series Analysis*, 3rd Edition, Prentice Hall, 1994.

[16] MPEG and JPEG traces are available at ftp://thumper.bellcore.com/pub/vbr.video.trace

[17] Internet traffic archive, available at http://ita.ee.lbl.gov/ .