Adaptive Linear Prediction Augmented Autoregressive Integrated Moving Average Based Prediction for VBR Video Traffic

T. Raghuveera, P.V.S. Kumar and K.S. Easwarakumar
Department of Computer Science and Engineering
Faculty of Information and Communication Engineering
Anna University, Chennai, 600 025, India

Abstract: Problem statement: Network traffic prediction plays a vital role in the optimal resource allocation and management in computer networks. This study introduces an ARIMA based model augmented by Adaptive Linear Prediction (ALP) for the real time prediction of VBR video traffic. The synergy of the two can successfully address the challenges in traffic prediction such as accuracy in prediction, resource management and utilization. **Approach:** ARIMA application on a VBR video trace results in a component wise representation of the trace which is then used for prediction. The ALP applied afterwards, ensures consistency in error fluctuation and better accuracy in turn. **Results:** Performance evaluation of the proposed method is carried out using RMSE. The prediction accuracy is improved by 23% and the error variance is reduced by 23%. **Conclusion:** The performance of the proposed method is thoroughly investigated, by applying it on video traces of different qualities and characteristics.

Key words: Traffic prediction, Autoregressive Integrated Moving Average (ARIMA), VBR video, Adaptive Linear Prediction (ALP), packet loss, traffic characteristics, video traffic, prediction accuracy, prediction process, univariate time, linear combination, error variance

INTRODUCTION

Traffic Prediction is the process of predicting future network traffic based on the characteristics of the past traffic. It has wide applications in the realm of networking, as it enhances the overall network performance through better resource utilization. Rapid advancements in the hardware, storage and video compression technologies enable the user to access video contents in a ubiquitous fashion (Kang *et al*., 2010). If the resources are allocated according to the peak rate of the video traffic, no packet loss occurs, but a substantial amount of the resources are wasted during transmission. On the other hand, if the bandwidth is not allocated close to the peak rate, large delays and excessive packet loss may be experienced. There exists high degree of correlation in video trace generated from a video encoder; this phenomenon can be used for traffic prediction. The prediction, when combined with dynamic resource allocation, can provision both network efficiency and QoS guarantees.

Traffic prediction can be done in two different ways. If the characteristics of the entire traffic are known in advance before the outset of the prediction process, it is known as offline traffic prediction. Here the constraint is that the algorithm must be equipped with full characteristics of the traffic to enable it for the prediction process. On the other hand if the prediction is done on the fly, it is known as online traffic prediction. Achieving better prediction accuracy in the online scheme is more challenging due to the uncertainty in the future traffic patterns. Capturing the traffic characteristics to the entirety is a near impossible task. Thus a good statistical model is expected to encapsulate the significant features of the traffic.

Generally video traffic has more bandwidth requirements than any other genre of traffic. Moreover it is sensitive to violation of QoS compliance. This makes video traffic as the ideal candidate for traffic prediction related experiments as well as network performance evaluation. So to develop a prediction scheme that performs well for VBR video inputs is treated as the ultimate challenge in this area.

There are three types of frames in a standard MPEG based video encoding scheme: I (Intracoded), P (Predictive) and B (Bi-directionally Predictive). I Frame type is self-contained. P Frame type carries the information difference from the preceding I or P type

Corresponding Author: Raghuveera Tripuraribhatla, Department of Computer Science and Engineering,
Faculty of Information and Communication Engineering Anna University, Chennai, 600 025, India

frame. B Frame type contains the interpolated information between consecutive I or P frame type pairs. The GOP structure specifies the number and temporal order of P and B frames between two successive I frames. GOP structure is represented by GOP(S, s), where S is the distance between successive I frames and s is the distance between consecutive P frames or the distance between I frame and following P frame. For example, GOP (16, 2) denotes the frame sequence "IBPBPBPBPBPBPBPB".

ARIMA (Autoregressive Integrated Moving Average) is a statistical methodology in time series analysis which is chiefly used in the forecast or prediction of future terms based on the characteristics of the past terms. It is a combination of three components namely the Auto Regression (AR), Integration (1) and Moving Average (MA). The model is generally referred to as an ARIMA (p,d,q) model where p, d, and q are integers greater than or equal to zero and refer to the order of the autoregressive, integrated, and moving average parts of the model respectively. When one of the terms is zero, it's usual to drop AR, I or MA. For example, an I(1) model is ARIMA(0,1,0), and a MA(1) model is ARIMA(0,0,1). This study makes good use of the ARIMA for the prediction of future frame size from the past, given a frame size process.

The sections to follow deal with Related Works, ARIMA, Proposed Work, Experimental Setup, Results and Discussion and conclusion, in order.

Related works: A great deal of effort has been put in to device a network traffic prediction scheme subject to the constraints of prediction accuracy, relative easiness of the method and implementation, the varieties of input that can be successfully processed by the scheme etc. All these constraints are complementary to one another and hence researches are oriented towards addressing at least any of these. Sharmila Chabba *et al.* (2009) made a thorough research over the use of neuro-fuzzy systems for the transmission of data on IP networks. A novel algorithm was developed by (Mohammad *et al.*, 2010) for the measurement and prediction of the traffic in B-ISDN. Thavasi and Natarajan (2011) had researched to find out a new way to manage the queue size in communication networks and thereby helping traffic prediction. A prediction methodology for the traffic prediction in LEO was developed by (Zihe *et al.*, 2011). A time series forecasting methodology was proposed by (Shabri *et al.*, 2009) using artificial neural network and statistical model. A comparison of univariate time series methods was suggested by (Assis *et al.*, 2010) in which they established that the combined ARIMA/ GARCH model

outperformed ARIMA and GARCH. Cui *et al.* (2011) developed an algorithm for traffic management in heterogeneous networks. Bashier and Talal (2007) introduced a univariate ARIMA model for forecasting and prediction related purposes. Suhartono (2011) used seasonal ARIMA model for forecasting.

ARIMA: It is a well established fact that ARIMA based prediction performs the best for traffics with no seasonality or trend present in it (Kang *et al.*, 2010). So first of all the input trace is preprocessed for the removal of the seasonality components and trend, if any, for making it fit for ARIMA. Then the trace is decomposed into several component processes and represented as linear combination of its own past values along with the past values of a newly generated ARIMA process obtained from the original trace as done in (Kang *et al.*, 2010). The prediction is done over the ARIMA model, to yield the predicted values of the future frame sizes. A comparison between predicted and actual values is done to evaluate the performance of the model.

Let X_t be the input frame size sequence with a regular fixed GOP pattern for a VBR compressed video, denoted by GOP (s,S) where s and S being the difference between successive P to (I or P) and consecutive 1-1 frame size. The sample process X_t is decomposed as:

$$X_t = X_t^s + X_t^S + \varepsilon_t$$

Where:

X_t^s, X_t^S = Denote the seasonal components which appear in every

s^{th}, S^{th} = Sample, respectively

ε_t = Error term. Then the differencing operation is performed multiple times for each lag. The difference orders D and d are set prior to performing the differencing operation. Assuming a GOP pattern GOP (16, 2) the differenced process Y_t c an be formulated as:

$$Y_t = \left(1 - B^2\right)^d \left(1 - B^{16}\right)^D X_t$$

where, B is the backward operator, which is widely used in statistics to make the time series expression more compact and is given by:

$$B^k X_t = X_{t-k}$$

Thus $(1-B)X_t$ denotes the differenced time series, $X_t - X_{t-1}$.

Let ARIMA model under consideration be $(1,1,1)_{2\times}(1,1,1)_{16}$, making $d = D = 1$, the above equation can be rewritten as:

$$Y_t = (1 - B^2)(1 - B^{16})X_t$$
$$= (1 - B^{16} + B^2 - B^{18})$$

Expanding using the above convention:

$$Y_t = X_t - X_{t-16} - X_{t-2} + X_{t-18} \qquad (1)$$

In deriving the state space model for ARIMA, we need to represent X_t as a linear combination of its past values. From Eq. 1 it is verifiable that:

$$X_t = X_{t-2} + X_{t-16} - X_{t-18} + Y_t$$

The differenced process Y_t is a multiplicative ARMA process ARMA $(1,1)_{2\times}(1,1)_{16}$. This can be represented as:

$$(1 - \varphi B^2)(1 - \Phi B^{16}) Y_t = (1 + \theta B^2)(1 + \Theta B^{16}) \varepsilon_t$$

where, ϕ, Φ, θ and Θ are coefficients of moving average and autoregressive expression.

To make the above equation more manageable a new AR process Z_t is introduced as:

$$Z_t = (1 - \varphi B^2)(1 - \Phi B^{16}) Z_t = \varepsilon_t$$

From Z_t we rewrite Y_t as:

$$Y_t = Z_t + \theta Z_{t-2} + \Theta Z_{t-16} + \theta\Theta Z_{t=18} \qquad (2)$$

Equation 1 can be rewritten by substituting for Y_t using Eq. 2 as:

$$X_t = \bar{X}_t + \bar{Z}_t \qquad (3)$$

Where:

$$\bar{X}_t = X_{t-2} + X_{t-16} - X_{t-18}$$

$$\bar{Z}_t = Z_t + \theta Z_{t-2} + \Theta Z_{t-16} + \theta\Theta Z_{t-18}$$

Thus the terms:

X_{t-2}, X_{t-16}, Z_{t-2}, Z_{t-16}, ,Z_{t-18} are enough to express X_t. But since these terms are certain lag apart, all the terms

X_{t-1} through X_{t-18} and Z_t through Z_{t-18} are mandatory for properly representing X_t. The Eq. 3 above is the governing equation of the ARIMA and hence the entire prediction process. It is used to generate the predicted frame size sequence. The error is measured by taking the difference between actual value and predicted value in each step.

Proposed work: This segment of the work chiefly comprises of the application of ALP over the result obtained in the previous segment. Two major modifications to the traditional ALP have been proposed here. Initially a threshold is chosen for the prediction error based on the characteristics of the input trace. This can ensure that the error values obtained would not vary beyond the threshold in either direction, giving out a lesser variance in error. This flexibility imparts a great deal of freedom in network management while allocating resources, as the fluctuation of error is curtailed by the threshold value. This has been shown schematically in Fig. 1.

Secondly the policy of using variable step sizes in ALP has been adopted. Adaptive Linear Prediction is a rudimentary and renowned technique in the fields of signal processing and forecasting.

ALP is a linear prediction mechanism where the predicted value can be expressed as a linear combination of a certain number of past values of the same process. The fundamental equation that governs a pth order linear predictor is expressed as:

$$\hat{X}_{t+N} = \sum_{l=0}^{N} w_t(1) X_{t-1} = w_n^T X_t$$

where, $w_t = (w_t(0),...,w_t(p-1))^T$ denotes the prediction filter coefficient vector which minimizes the mean square error. Parameter p indicates the number of past values used for prediction.

Let $e_t = X_{t-N} - \hat{X}_{t-N}$ be the prediction error at the tth instant.

Starting with an initial estimate of filter coefficient w, and for each new data point, the ALP method updates w_t using the recursive equation by:

$$w_{t+1} = w_t + \frac{\mu e_t X_t}{\|X_t^2\|}$$

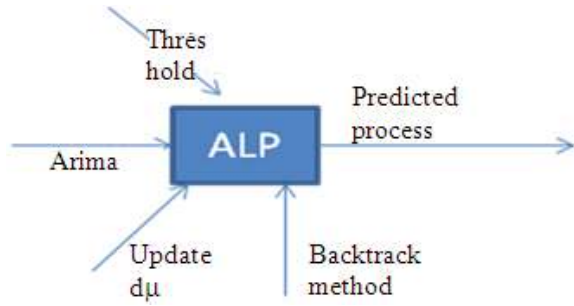where, μ is the step size and is usually a fixed value. Also $\mu \in (0,2]$ (Haykin, 1991).

Fig. 1: Schematic diagram of the proposed work

During the process of updating of weight vector the generic ALP, above suggests a parameter called μ, which indicates the step size and would be a fixed value. But it has been verified in (Zhao *et al*., 2002) that the usage of a variable step size for the prediction of each value could result in better prediction accuracy. Here we present a variable step size prediction technique along with the incorporation of a threshold value to reduce the error variance.

We have altered the ALP mechanism to include a variable step size value at every prediction step. To derive the value of step size to be used for predicting the next frame size, a modified version of the variable step size prediction equation in (Zhao *et al*., 2002) is used. The equation is as follows:

$$\mu_{k+1} = \alpha\mu_k + \gamma\left(q_1 e_k^2 + q_2 e_{k-1}^2\right)$$

where, α, γ, q_1, q_2 are constants whose values were determined in (Zhao *et al*., 2002) as 0.98, 0.015, 0.7 and 0.3 respectively. For our experimentation we have altered the above eq. as follows:

$$\mu_{k+1} = \alpha\mu_k + \frac{\gamma\left(q_1 e_k^2 + q_2 e_{k-1}^2\right)}{e_k^2 + e_{k-1}^2} \qquad (4)$$

μ_i and e_i represent step size and prediction error respectively for the ith value in the trace.

Algorithm 1:

For (I = k; i> 0; i--)
$\mu_{k+1} = \mu_I \ni e_i < T$

This change of equation can be explained by the observation that the original equation is largely unfit for huge frame sizes, generally present in high quality video frame size sequences. The method here predicts each frame size of the trace using ALP with a step size value measured by Eq. 4. It then checks for the error. If the error doesn't fall below the threshold limits the method gets a fresh value for step size as demonstrated in the following procedure.

The procedure to find out the step size value to be used in $(k+1)^{th}$ prediction:

$$\mu_{k+1} = \begin{cases} \mu & \text{obtained using Equation (4), if } e_{k+1} < T \\ \mu & \text{obtained using Algorithm (1) otherwise} \end{cases}$$

Where:

$e_{k+1} =$ Prediction error in the $(k+1)^{th}$ step
$T =$ Threshold

MATERIALS AND METHODS

A variety of analyses has been performed here for video traffic with varying qualities for 10 minute traces taken from the movie Star Wars and from NBC News (http:// www.traces.eas.asu.edu/). The two traces analyzed here are taken from videos of different ilk; the movie Star Wars IV has rapid scene changes and insignificant correlations. On contrary, the NBC news possesses relatively high degree of correlation. This establishes that the procedure explained here is capable of processing and yielding good results regardless of the broad type of video being used as input.

The general attributes of the input video traces are as follows. All the traces being experimented here are encoded in H.264 standard. Each of them is of duration 10 minutes and contains a total of 18000 frames. There will be 30 frames every second. The GOP pattern is G16B1 and CIF resolution is 352×288. The frame statistics of the individual input video traces used for analysis is summarized in Table 1.

The implementation procedure is summarized as follows:

Step-1: Convert the original sample trace sequence to an ARIMA sequence using Equation 3.
Step-2: Perform predictions over ARIMA sequence using step-size adjusted Adaptive Linear Predictor.
Step-2.1: During every prediction step, use the procedure as given in Algorithm 1, to update the step-size variable, and also compute the prediction error.
Step-2.2: Generate the complete sequence recursively using the above step.

Table 1 Frame size statistics of input video traces

Name	Quality	Max frame size	Min frame size	Avg. frame size	Burstiness
Star wars IV	Low	13344	168	699.58	19.06
	Medium	82152	168	6332.60	12.97
	High	476912	19600	111352.00	4.28
NBC- News	Low	19024	168	1553.70	12.25
	Medium	140040	168	19227.00	7.28
	High	706032	61056	228174.00	11.56

Table 2: Variance comparison

		Variance			Percentage of performance improved		
Input trace	Quality	ALP	ARIMA	ARIMA with ALP	ARIMA over ALP	ARIMA with ALP over ALP	ARIMA with ALP over ARIMA
Star wars IV	Low	14517.414	12026.137	10023.760	17.16	30.95	16.65
	Medium	56156.529	50700.641	45510.812	9.72	18.96	10.24
	High	181345.601	162231.823	137691.160	10.54	24.07	15.13
	Low	11610.331	9467.550	7935.801	18.46	31.65	16.18
NBC-News	Medium	13007.780	12752.034	10401.548	1.97	20.04	18.43
	High	186534.925	156091.600	120225.710	16.32	35.55	22.98

Table 3: RMSE Comparison

|		RMSE %			Percentage of performance improved		
Input trace	Quality	ALP	ARIMA	ARIMA augmented with ALP	ARIMA over ALP	ARIMA augmented with ALP over ALP	ARIMA augmented with ALP over ARIMA
Star Wars IV	Low	4.64	3.87	3.13	16.59	32.54	23.64
	Medium	9.21	8.29	6.47	9.98	18.89	10.97
	High	14.29	11.76	10.40	17.70	27.22	13.07
NBC- News	Low	4.83	3.75	3.24	22.36	32.91	15.74
	Medium	7.87	6.72	5.12	14.61	29.86	21.73
	High	13.98	12.96	10.12	7.29	20.45	16.54
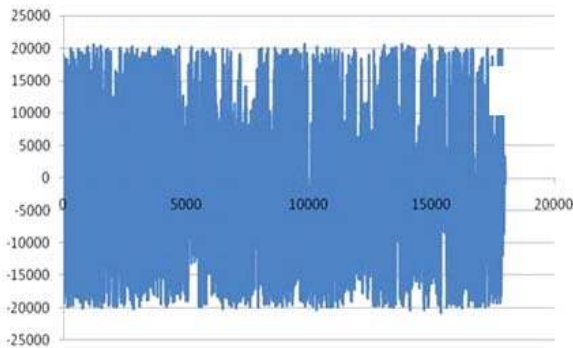
Fig. 2: Fluctuation of error (high quality-star wars iv)

RESULT AND DISCUSSION

A detailed analysis is performed to verify how striking the feature of having a provision to specify the threshold to make the prediction error consistent. Here the parameter is the error variance. Right through the analysis the error threshold is kept as 10 percent of the average frame size. It has been observed that our method gives the minimum variance and outperforms ARIMA and traditional ALP. This has been quantitatively portrayed in Table 2. It can be observed from the Table 2 that combining ARIMA with variable step size ALP is performing the best for all of the input samples. Also ARIMA stands superior to the traditional ALP.

So generically the method presented here, i.e., ARIMA, incorporated with ALP of variable step size, can be learned to perform well for input traces of substantial differences in quality. When traces with extremely opposing characteristics (Star Wars IV and NBC News) were tested with the method, the results were found to be quite satisfactory, based on which we justify that the proposed method performs relatively better and can handle traffic of any genre efficiently.

A comparison of performance for the different prediction schemes of traditional ALP, ARIMA and ARIMA augmented by variable step size ALP, are carried out, using RMSE as the parameter. The results obtained are summarized in Table 3. Figure 2 is a depiction of the error file of a high quality video trace of the movie Star Wars II, obtained after the application of the proposed method. The threshold value fixed was 20000 and it can be easily observed that the fluctuation in error is limited by the threshold.

CONCLUSION

An ARIMA based mechanism augmented by ALP for the prediction of VBR video traffic was devised and implemented. Upon evaluating the performance using RMSE an improvement of 30% was observed. Also the error variance was reduced by 19%. The methodology was justified through a detailed analysis of different qualities of VBR video traces taken from two different video files. The provision of specifying a threshold value to limit the error fluctuation effectively controls the variation in prediction errors. Here the threshold value was fixed empirically based on the characteristics of the input trace. The work can be extended by conceiving a mathematical model to derive the optimal threshold value from the attributes of the input trace.

REFERENCES

Assis, K., A. Amran, Y. Remali and H. Affendy, 2010. A comparison of univariate time series methods for forecasting cocoa bean prices. Trends Agric. Econ., 3: 207-215. DOI: 10.3923/TAE.2010.207.215

Bashier, A.A. and B. Talal, 2007. Forecasting foreign direct investment inflow in Jordan: Univariate arima model. Edn. J. Soc. Sci., 3: 1-6. DOI: 10.3844/jssp.2007.1-6

Cui, Y.. Y. Xu, R. Xu and X. Sha, 2011. A multi-radio packet scheduling algorithm for real-time traffic in a heterogeneous wireless network environment. Inform. Technol. J., 10: 182-188. DOI: 10.3923/ITJ.2011.182.188

Haykin, S., 1991. Adaptive Filter Theory. 2nd Edn., Englewood Cliffs, N J Prentice Hall, United State or United Kinddom, ISBN: 0130132365 9780130132369, pp: 817-844

Kang. S., S. Lee, Y. Won and B. Seong , 2010. On-Line prediction of non-stationary variable-bit-rate video traffic. IEEE Transactions on Signal Processing, 06 Nov, IEEE Signal Processing Society, Hangyang University Seoul, South Korea, pp: 1219-1237. DOI: 10.1109/TSP.2009.2035983

Mohammad, A., R. I. Mansi and M. Hussien, 2010. Algorithm development for measurement and prediction the traffic in broad band integrated service networks. J. Comp. Sci., 6: 1164-1169. DOI: 10.3844/jcssp.2010.1164-1169

Shabri, A., R. Samsudin and Z. Ismail, 2009. Forecasting of the rice yields time series forecasting using artificial neural network and statistical model. J. Applied Sci., 9: 4168-4173. DOI: 10.3923/JAS.2009.4168.4173

Suhartono, 2011. Time series forecasting by using seasonal autoregressive integrated moving average. subset, multiplicative or additive model. J. Math. Stat., 7: 20-27. DOI: 10.3844/jmssp.2011.20-27

Thavasi, S. and N. Natarajan, 2011. An effective queue management scheme for data communication. J. Comput. Sci., 7: 348-351. DOI: 10.3844/jcssp.2011.348-351

Zhao, H., N. Ansari, Y.Q. Shi, 2002. A fast non-linear adaptive algorithm for video traffic prediction. Proceedings of the International Conference on Information Technology: Coding and Computing, Apr. 08-10, Las Vegas, Nevada, pp: 0054. http://doi.ieeecomputersociety.org/10.1109/ITCC.2002.1000359

Zihe, G., G. Qing and N. Zhenyu, 2011. A distributed routing algorithm with traffic prediction in LEO satellite networks. Inform. Technol. J., 10: 285-292. DOI: 10.3923/itj.2011.285.292