An ARIMA based Approach for Traffic Prediction

Sunil Kumar P V
Dept. of CSE, MES College of Engg.,
Kuttippuram, Kerala

ABSTRACT
Network traffic prediction plays a vital role in the optimal resource allocation and management in computer networks. This paper introduces an ARIMA based model for the real time prediction of VBR video traffic. The methodology presented here can successfully addresses the challenges in traffic prediction such as accuracy in prediction, resource management and utilization. ARIMA application on a VBR video trace results in a component wise representation of the trace which in turn used for prediction. A brief introduction of the classic prediction scheme of ALP along with a quantitative comparison of the ARIMA with ALP is also presented. 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 18%.

Keywords-Traffic prediction, ARIMA, ALP, VBR Video**.**

1. 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 [1]. 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 [2].

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. Here the algorithm takes traffic characteristics as input and simultaneously does the 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 in entirety is a near impossible task. Thus a good statistical model is expected to encapsulate the significant features of the traffic.

Traffic prediction can be very well applied to any form of network traffic; nevertheless the researches in this area extortionately prefer video traffic. This can be reasoned on the basis of the high bandwidth and QoS requirements demanded by video traffic as well as its sensitiveness to violation of QoS compliance. Generally video traffic has more bandwidth requirements than any other genre of traffic. This makes video traffic as the ideal candidate for traffic prediction related experiments as well as network performance evaluation. It is arguable at this juncture that any prediction scheme that works well for video traffic can yield better results for other traffic types. So to develop a prediction scheme that performs well for VBR video inputs is treated as the ultimate challenge in this area.

The ensuing discussion unravels some basic terminologies needed further in the literature. The terms that are frequently being referred to, in any literature on traffic prediction are the types of frames as well as the GOP. Also this work, in particular uses the statistical prediction method namely ARIMA. A brief description follows.

There are three types of frames in a standard MPEG based video encoding scheme: I (Intracoded), P (Predictive) and B (Biderectionally Predictive). I Frame type is self-contained. P Frame type carries the information difference from the preceding I or P type 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" [1].

The rest of the paper is organized as follows. Section II presents a survey of the related works in the domain. Section III gives an insight to the Adaptive Linear Prediction Method. Section IV carries a detailed description of the proposed ARIMA model for traffic prediction. Experimental results are presented in Section V followed by the conclusion in Section VI.

2. LITERATURE REVIEW
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. Doulamis et al. [3] used the AR(1) process to represent the relationship between each type of frame in consecutive GOP's and added another layer to represent inter-GOP behavior. Several works explicitly model the time dependent behavior, i.e., regular GOP pattern of VBR traffic: ARIMA, gamma based autoregression (GBAR) model [4], GOP GBAR model [5], and GOP ARIMA model [6]. However, the GBAR and GOP GBAR models yield exponentially decreasing autocorrelations while empirical VBR traffic has much more slowly decreasing autocorrelations. The GOP ARIMA model successfully represents the slowly decaying sample autocorrelations of the VBR sequence. Grossglauser et al. [7] proposed the use of Renegotiated

Constant Bit Rate (RCBR) service to support VBR video. They based their approach on the AR(1) model and used heuristics to predict the future bandwidth. Adaptive linear prediction (ALP) has been popular for real-time prediction of the VBR bandwidth process [7]. While linear prediction is very simple and fast, it does not properly capture the inter and intra-GOP correlations. In addition, it cannot quickly react to structural changes in the underlying sequence, e.g., scene change, and it is vulnerable to noisy input, e.g., intracoded B frame. Yoo [8] extended the work of Adas et al. [9] by incorporating a threshold-based scene detection scheme. Zaho H et al. [2] carried out experiments with smoothed traffic. This was done by taking I frames alone into account. They again normalized the traffic and the resulting sequence was represented as a time series. Then the future frames are predicted from past frames of this sequence. The analyses were done for fixed and various step sizes. Sang A et al. [10] tried to find a measure on the maximum number of future frames to be looked into, so that the predicton error is minimal. They also derived that the prediction interval size and prediction acuracy vary inversely. W. Xu and A. G. Qureshi [11] formed a sequence of GOP lengths used this for smoothing the traffic. Yoo S J [8] analysed traffic with rapid scene changes. It has been found that I and P frames prform poorly while predicting traffic with scene changes. So it was proposed to separate out the B frames and different methods were employed for B frames and for I and P frames.

3. ALP

Adaptive Linear Prediction is a rudimentary and renowned technique in the domain of signal processing and forecasting. A plethora of experiments had happened where the researchers center on ALP and pave the way for novel prediction methodologies that vary in effectiveness, easiness, accuracy and so on. It 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:

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

Where $\mathbf{w_t}$ denotes the prediction filter coefficient vector which minimizes the mean square error. $\mathbf{X_t}$ is a set of current and previous values of X_t. 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 t^{th} instant.

Starting with an initial estimate of filter coefficient ***w***, and for each new data point, the ALP method updates $\boldsymbol{w_t}$ using the recursive equation by

$$\mathbf{w}_{t+1} = \mathbf{w}_t + \frac{\mu e_t X_t}{\left\| X_t^2 \right\|}$$

Where μ is the step size and is usually a fixed value and $0 < \mu < 2$ [12]. The denominator in the right most term is the dot product used in the process of updating the weight vector. The value of step size determines how fast the prediction algorithm gets converged. The choice of step size is basically the process of resolving a conflict between the minimization of prediction error and speed in which the algorithm converges. For lager values of step size the convergence will be quick but the prediction error is more. For smaller μ values the error can be reduced at the expense of a slow convergence.

The ALP mechanism, as the name suggests, is self adaptive. That is the method finds out the prediction error in every step and the weight vector to be used in the prediction of next frame gets updated based on the current error value. This provision imparts a great deal of accuracy to the method since the weights used for the past terms in the next step of prediction will be lesser than current values for more prediction errors and vice versa.

The value of *p* denotes the total number of previous frames to be taken in to account for the prediction of the current frame. In other words It is an indication of how far into the future can network traffic be predicted with confidence. If a large value is chosen for the prediction interval it would provide sufficient time for control actions. But this would result in large prediction error. So it is highly important to optimize the value chosen for prediction interval. In fact it is a tradeoff between the prediction error magnitude and time for control actions[13].

4. PROPOSED WORK - ARIMA

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 (I) 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 work makes good use of the ARIMA for the prediction of future frame size from the past, given a frame size process.

The input to the ARIMA model is the trace consisting of a frame size sequence. Here an ARIMA based mechanism is used to predict the future frame size based on the past and current frame sizes. It is a well established fact that ARIMA based prediction performs the best for traffics with no seasonality or trend present in it [1]. 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 [1]. 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.

Initially the input trace is prepared. For this the seasonality components must be removed. Another added advantage of this process is that the input can be decomposed and expressed in terms of additive components so that a separate model can be used for each subsequence.

Let $\boldsymbol{X_t}$ 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 I to I frame size. The sample process $\boldsymbol{X_t}$ is decomposed as:

$$\mathbf{X_t} = x_t^s + x_t^S + \varepsilon_t$$

Where X_t^s and X_t^S denote the seasonal components which appear in every s^{th} and S^{th} sample, respectively and ε_t is the 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 $\mathbf{Y_t}$ can be formulated as:

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

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 $d = D = 1$, "(1)" can be rewritten: $Y_t = \left(1 - B^2\right)^d \left(1 - B^{16}\right)^D X_t = \left(1 - B^{16} - B^2 + B^{18}\right)X_t$

Expanding using the above convention

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

In deriving the state space model for ARIMA, we need to represent $\mathbf{X_t}$ as a linear combination of its past values. From "(2)" it is verifiable that:

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

The differenced process $\mathbf{Y_t}$ is a multiplicative ARMA process ARMA $(1,1)_2 \times (1,1)_{16}$. This can be represented as:

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

Where □, *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: $Z_t = \left(1 - \phi B^2\right)\left(1 - \theta B^{16}\right)Z_t = \varepsilon_t$

From $\mathbf{Z_t}$ we rewrite $\mathbf{Y_t}$ as:

$$Y_t = Z_t + \theta Z_{t-2} + \Theta Z_{t-16} + \theta\,\Theta Z_{t-18} \qquad (3)$$

Equation (3) can be rewritten by substituting for Y_t using "(2)" as

$$X_t = X_{t-2} + X_{t-16} - X_{t-18} + Z_t + \theta Z_{t-2} + \Theta Z_{t-16} + \theta\,\Theta Z_{t-18} \quad (4)$$

Thus the terms $X_{t-2}, X_{t-16}, X_{t-18}, Z_t, Z_{t-2}, Z_{t-16}, Z_{t-18}$ 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 Xt**.** Equation (4) 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.

The ARIMA model proposed here performs better than the classic prediction technique of Adaptive Liner Prediction (ALP). A detailed analysis is carried out based on RMSE and error variance as the measures. In both the cases the proposed method was found to outperform the classic ALP.

5. EXPERIMENTAL RESULTS

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 [14]. 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 using H.264 or MPEG4. 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 I

TABLE I. FRAME STATISTICS OF INPUT VIDEOS

Name	**Quality**	**Frame Size**			**Burstiness**
		Maximum	*Minimum*	*Average*	
Star Wars IV	Low	13344	168	699.58	19.06
	Medium	82152	168	6332.64	12.97
	High	476912	19600	111352	4.28
NBC News	Low	19024	168	1553.07	12.25
	Medium	140040	168	19227.5	7.28
	High	706032	61056	228174	11.56

A comparison of performance for the different prediction schemes of traditional ALP, ARIMA and using RMSE as the parameter. Three different qualities of each of the video inputs are included in the analysis to verify the nuances in the robustness of the proposed method. The results obtained are summarized in TABLE II.

It can be observed from the table that the ARIMA is performing the best for all of the input samples compared to ALP. So generically the method presented here, 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 it is justified that the proposed method performs relatively better and can handle traffic of any genre efficiently.

A fresh set of analysis is performed to verify how much leverage the application of ARIMA methodology imparts to the error variance. Error variance is selected as the performance parameter as it gives a clear picture of how effective the method is. So here the parameter is the *error variance*. It has been observed that the method present here gives the minimum variance and outperforms traditional ALP. This has been quantitatively portrayed in TABLE III. ARIMA methodology outperforms the traditional ALP by 18% in terms of error variance.

TABLE II. ROOT MEAN SQUARE ERROR COMPARISON

Input File	Quality	Root Mean Square Error		Percentage of Performance Improved
		ALP	*ARIMA*	
Star Wars IV	Low	464	387	16.59
	Medium	921	829	9.98
	High	1429	1176	17.7
NBC News	Low	483	375	22.36
	Medium	787	672	14.61
	High	1398	1296	7.29

TABLE III. ERROR VARIANCE COMPARISON

Input File	Quality	Error Variance		Percentage of Performance Improved
		ALP	*ARIMA*	
Star Wars IV	Low	14517.414	12026.137	17.16
	Medium	56156.529	50700.641	9.72
	High	181345.601	162231.823	10.54
NBC News	Low	11610.331	9467.55	18.46
	Medium	13007.78	12752.034	1.97
	High	186534.925	156091.6	16.32

6. CONCLUSION

An Autoregressive Integrated Moving Average (ARIMA) based mechanism for the real prediction of VBR video traffic was devised and implemented. A study to find out the performance improvement compared to the classic prediction technique, Adaptive Liner Prediction (ALP). Right through the work two performance measures were taken into account viz. the Root Mean Square Error (RMSE) and the error variance. It was found that the proposed ARIMA method stands superior to the traditional ALP when compared based on both the performance measures used here. Video traces with substantial variations in the type to which they belong to as well as quality were analyzed. An improvement of 23% was observed for RMSE and a reduction of error variance amounting to 18% was found for error variance to justify the proposed method.

7. REFERENCES

[1] S. Kang, S. Lee, Y. Won, B. Seong, "On-Line Prediction of Nonstationary Variable-Bit-Rate Video Traffic", IEEE Transacions on Signal Processing, vol. 58, No. 3, March 2010, pp 1219 - 1237.

[2] H. Zaho, N. Ansari, Y.Q Shi, "A Fast Non-linear Adaptive Algorithm for Video Traffic Prediction", Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC.02), 2002.

[3] N. Doulamis, A. Doulamis, and S. Kollias, "Modeling and adaptive prediction of VBR MPEG video sources," in Proc. IEEE ThirdWorkshop on Multimedia Signal Process., Sep. 1999, pp. 27 - 32.

[4] N. Rozic and M. Vojnovic, "Source modeling of MPEG video," in Proc.IEEE GLOBECOM '97, 1997, pp. 1429 - 1433.

[5] D. Heyman, "The GBAR source model for VBR videoconferences' 'IEEE/ACM Trans. Netw., vol. 5, pp. 554 - 560, 1997.

[6] Y. Won and S. Ahn, "GOP ARIMA: Modeling the non-stationarity of VBR process," ACM/Springer Multimedia Syst. J., vol. 10, no. 5, pp.359 - 378, Aug. 2005.

[7] M. Grossglauser, S. Keshav, and D. N. C. Tse, "RCBR: A simple and efficientservice for multiple time-scale traffic," IEEE/ACM Trans. Netw., vol. 5, no. 6, pp. 741 - 755, Dec. 1997

[8] S.-J. Yoo, "Efficient traffic prediction scheme for real-time VBR MPEG video transmission over high-speed networks," IEEE Trans.Broadcasting, vol. 48, no. 1, pp. 10 - 18, 2002.

[9] A. Adas, "Using adaptive linear prediction to support real-time VBR video under RCBR network service model," IEEE/ACM Trans. Netw., vol. 6, pp. 635 - 644,Oct. 1998.

[10] A. Sang, S. Li, "A Predictability Analysis of Network Traffic", IEEE/ACM Trans. Netw., vol. 10, no. 8, pp. 71 - 75, Dec. 1999.

[11] W. Xu and A. G. Qureshi, "Adaptive Linear Prediction of MPEG Video Traffic", Fifth International Symposium on Signal Processing and its Applications, ISSP '99, Brisbane, Australia, 22-25 August, 1999

[12] S Haykin, "Adaptive Filter Theory"; Englewood Cliffs, N J Prentice Hall,1991.

[13] A. Sang and S. Li, "A Predictability Analysis of Network Traffic", in Proc. INFOCOM, 2000, pp.342-351.