Computer Communications 31 (2008) 3741–3751

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

On traffic prediction for resource allocation: A Chebyshev bound based allocation scheme

R.G. Garroppo, S. Giordano *, M. Pagano, G. Procissi

Department of Information Engineering, University of Pisa, Via Caruso 16, I-56122 Pisa, Italy

ARTICLE INFO

Article history:
Available online 20 June 2008

Keywords:
FIR filters
Linear prediction
Chebyshev bound
Resource allocation

ABSTRACT

The paper presents a predictive approach to network resource allocation techniques. The rationale of this work is to use measurements to estimate future traffic behavior by prediction, and to use such an estimation to define the amount of future network resources that will be required by the considered traffic. In this framework, the paper presents the analysis and performance evaluation of classical and chaotic techniques for network traffic prediction. The performance parameters considered in the analysis are: the accuracy of predictors in capturing the actual behavior of traffic; the computational complexity for a realistic integration of such predictors into experimental testbeds; and the responsiveness with respect to traffic pattern variations. The analysis results show that the classical normalized linear mean square predictor achieves a satisfactory trade-off among the above mentioned metrics as it presents a medium level of complexity while achieving high performance in terms of prediction accuracy and responsiveness to network traffic changes. Then, using the normalized linear mean square predictor, we derive a bandwidth allocation strategy, named statistical delay bound (SDB), which guarantees a probabilistic bound on the delay experienced by packets traversing a network node. The paper presents the performance analysis of SDB showing that, in spite of the simplicity of the adopted predictive algorithm, the proposed measurement based technique allows to fulfill the project requirements and candidates for actual experimentation into prototypal routers which supports QoS mechanisms.

1. Introduction

The evolutionary trend of communication networks in which new services such as Voice over IP, audio/video streaming, videoconferencing and virtual private networks (VPNs) with quality of service are being offered to end users has led to the deployment of network architectures (e.g. DiffServ [1,2]) supporting QoS.

Moreover, over the last years, a considerable effort has been dedicated to the optimization of operational IP networks through the MPLS switching paradigm [3] and the so called traffic engineering [4].

The exigence of integration of such approaches arises the necessity of dynamic traffic control mechanisms, such as resource allocation techniques and admission control, to achieve optimal resource utilization and network performance [5].

Model-based approaches have demonstrated their weakness in that the long range dependence (LRD) property of network traffic [6] is not on-line distinguishable from non-stationarity and it does not permit an easy estimation of characteristic parameters (such as the Hurst parameter). This has impaired the development of easy, flexible and effective traffic models to be widely used for network dimensioning.

In addition to the above mentioned reasons, novel concepts like network self-sizing and reconfigurability have led to consider *measurement based* approaches as the most promising in order to address (even in approximate ways) common issues in Diffserv aware traffic engineering (DS-TE) [7] context such as constraint-based path computation, LSP dimensioning, scheduler parameters selection and so on.

In this scenario, the objective of this paper is to propose a novel measurement based strategy for resource allocation which integrates traffic measures and prediction. To this aim, the comparison of different prediction techniques and the definition of a resource allocation algorithm that uses the information produced by the selected traffic predictor are of paramount relevance.

The first part of the paper is devoted to the presentation of several prediction algorithms (classic, chaotic and geometric) and to the investigation of their performance in terms of the above mentioned criteria: accuracy, complexity and responsiveness. The prediction algorithm which will achieve the best trade-off among the three criteria will be integrated into the resource allocation strategy derived in the second part of the paper.

As above mentioned, the second part of the article is devoted to the derivation and performance evaluation of the statistical delay

* Corresponding author. Tel.: +39 050 2217539.
E-mail addresses: r.garroppo@iet.unipi.it (R.G. Garroppo), s.giordano@iet.unipi.it (S. Giordano), m.pagano@iet.unipi.it (M. Pagano), g.procissi@iet.unipi.it (G. Procissi).

0140-3664/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.comcom.2008.05.019

bound (SDB) resource allocation strategy. SDB aims at computing the service capacity to be allocated a connection in order to guarantee a probabilistic bound on the maximum queueing delay experienced by packets of the connection itself.

In summary, the paper is organized as follows. After the presentation of related works on traffic prediction for resource allocation, Section 3 gives the basic notions of time series prediction, while Section 4 reports the performance comparison of three different prediction algorithms. Section 5 describes the SDB allocation strategy whose performance is analyzed in Section 6. Finally, Section 7 concludes the paper with final remarks.

2. Related works

Although time series prediction is a classic topic in signal processing, its adoption into networking contexts is becoming quite frequent. Recently, different predictors have been presented and a new active queue management (AQM) scheme exploiting the predictions of future traffic intensity has been proposed in [8].

Focusing on resource allocation algorithms based on traffic prediction, different solutions have been proposed depending on network technology. In particular, in [9] the authors propose an access control protocol for high speed downlink packet access (HSDPA) system based on bandwidth allocation between real time and non-real time traffic. By applying a traffic model to the real time traffic in the current time slot, the system can predict the residual capacity in the next time slot, thus it can optimize the scheduling of non-real time data transmission. It is relevant to note that predictions are based on a model assumption for the considered traffic, and that the time quantum, T_p, is of the order of the transmission time interval length (usually tens of ms).

In [10], the author proposes a new dynamic satellite bandwidth allocation technique (named predictive resource reservation access, PRRA) which is based on accurate videoconference traffic prediction. The proposed model is then combined with results on data traffic modeling and prediction presented in [11] and is shown to provide very good throughput and delay results. In this case, the time quantum T_p is of the order of a round trip time for a GEO satellite system, i.e. around 500 ms.

In both these works, the predictions are based on traffic models; it is well known that the accuracy of model-based approaches depend on the reliability of the assumed source models. Often, to simplify the model of the system, Markov chains are used to model source traffic; this assumption ignores the LRD nature of traffic. In our previous work [12], the LRD nature of traffic has motivated the investigation of chaotic prediction to develop bandwidth allocation strategies.

Differently from these works, in [13] the dynamic bandwidth allocation strategy designed to support variable bit rate (VBR) video traffic does not assume any model for the traffic source. The strategy predicts the bandwidth requirements for future frames by using adaptive linear prediction that minimizes the mean square error. The adaptive technique does not require any prior knowledge of the traffic statistics nor assume stationarity. In this case, the prediction based bandwidth allocation algorithm has been defined for the renegotiated constant bit rate (RCBR) network service model of ATM.

Two traffic predictors on frequency domain have been compared in [14], in terms of complexity and ability to guarantee the project parameter of the resource allocation algorithm. The authors assume the server utilization as the project parameter (hence we refer to that approach as constant server utilization algorithm). The focus of [14] is on VBR video transport over ATM network.

As opposed to the previous mentioned works, our goal is to define a bandwidth allocation algorithm based on traffic prediction, without any assumptions on the statistical features of traffic. Furthermore, our proposal uses the prediction algorithm in the time domain and the allocation strategy has been designed by assuming as project constraint the maximum queueing delay. Moreover, the system model considered in our study is general and does not refer to any specific technology. It can be adapted for resource allocation in cellular, satellite or wired networks.

Finally, the prediction algorithm as well as the bandwidth allocation scheme have been selected taken into account the actual implementation of the proposed solutions. The purpose of predictors is to capture the traffic intensity and accurately estimate the future arrivals over prediction intervals [15]. Also, they should exhibit high responsiveness with respect to sudden traffic changes. The huge effort towards actual experimentation of novel network architectures through the set up of field-trials [16,17] adds further critical constraints: ease of implementation and low computational complexity. The proposed solutions have been selected by considering these aspects.

3. Basics of traffic prediction

This section briefly presents some of the basic concepts of time series prediction. Since the topic is well known and widely investigated in signal processing, in the following a very concise description will be given and more details are left to the bibliography.

Time series prediction can be addressed through statistical or geometric techniques. This paper investigates both classes of algorithms, with special focus on linear and chaotic prediction for the first class, and on polynomial extrapolators for the class of geometric algorithms.

None of the presented algorithms, though, require the knowledge of a specific model of the underlying time series to be predicted. Statistical predictors only require the underlying stochastic sequence to be wide sense stationary, while geometric predictors are derived by simple algebraic reasoning.

The general problem of prediction can be stated as follows: given a set of observations of a stochastic process $x(n)$, give an estimation $\hat{x}(n+k)$ of the value $x(n+k)$ that the process x will assume k steps ahead.

In other words, given a vector of p observations, $\underline{x} = [x(n), x(n-1), \ldots, x(n-p+1)]$, the predicted value $\hat{x}$ is obtained by

$$\hat{x} = \Upsilon(\underline{x}) \tag{1}$$

where the function Υ is called *predictor*.

The topic of next paragraphs will be the derivation of the analytic expression of Υ for the different classes of predictors investigated.

3.1. Linear prediction

Linear prediction occurs whenever the function $\Upsilon(\underline{x})$ is *linear*. In other words, the problem is to determine the impulse response $h(n)$ of the linear filter h such that

$$\hat{x}(n+k) = x(n) \otimes h(n) = \sum_{i=0}^{p-1} h(i)x(n-i) \tag{2}$$

The filters coefficients can be determined according to arbitrary *optimality* criteria. One of the most famous and widely adopted prediction algorithm is the so-called *linear minimum mean square error (LMMSE)* predictor in which, the values $h(n)$ are derived by minimizing the mean square error of prediction:

$$\mathbb{E}[\varepsilon^2(n)] = \mathbb{E}[(x(n+k) - \hat{x}(n+k))^2] \tag{3}$$

In that sense, LMMSE is an *optimal* algorithm in that, within the class of linear filters, it minimizes the mean square error of predic-