Traffic Modeling, Prediction, and Congestion Control for High-Speed Networks: A Fuzzy AR Approach

Bor-Sen Chen, *Senior Member, IEEE*, Sen-Chueh Peng, *Member, IEEE*, and Ku-Chen Wang

***Abstract*—In general, high-speed network traffic is a complex, nonlinear, nonstationary process and is significantly affected by immeasurable parameters and variables. Thus, a precise model of this process becomes increasingly difficult as the complexity of the process increases. Recently, fuzzy modeling has been found to be a powerful method to effectively describe a real, complex, and unknown process with nonlinear and time-varying properties. In this study, a fuzzy autoregressive (fuzzy-AR) model is proposed to describe the traffic characteristics of high-speed networks. The fuzzy-AR model approximates a nonlinear time-variant process with a combination of several linear local AR processes using a fuzzy clustering method. We propose that the use of this fuzzy-AR model has greater potential for congestion control of packet network traffic. The parameter estimation problem in fuzzy-AR modeling is treated by a clustering algorithm developed from actual traffic data in high-speed networks. Based on adaptive AR-prediction model and queueing theory, a simple congestion control scheme is proposed to provide an efficient traffic management for high-speed networks. Finally, using the actual ethernet-LAN packet traffic data, several examples are given to demonstrate the validity of this proposed method for high-speed network traffic control.**

***Index Terms*—Cell loss rate, fuzzy-AR approach, quality of service (QoS), traffic prediction.**

I. INTRODUCTION

THE asynchronous transfer model (ATM)-based broad-band integrated service digital network (B-ISDN) is a high-speed transport network designed to support all variable service with different requirements for quality of service (QoS) (e.g., cell loss rate, delay, and delay jitter) and a broad range of statistical characteristics [1]–[10]. However, the additional flexibility needed to accommodate different traffic sources may cause serious congestion problems, resulting in severe buffer overflow, cell loss, and degradation in the required QoS. In order to guarantee the QoS, effective congestion control schemes are needed. Because congestion control schemes depend on modeling for their traffic characteristics and available network resources [22], [38], [39], an accurate estimation of traffic model is crucial for successful congestion control of ATM networks. These traffic models need to be accurate and able to represent the statistical characteristics of the actual traffic. If the traffic models do not accurately represent actual traffic, they may overestimate or underestimate network performance.

Traffic models are analyzed based on goodness-of-fit, number of parameters needed to describe the model, parameter estimation, and analytical tractability. Recently, traffic models have been described as stationary or nonstationary [1], [2], [13], [18], [21], [23], [24], [37]. Stationary traffic models can be classified in general into two classes: short-range dependent models and long-range dependent models. Short-range dependent models include Markov models and regression models (i.e., AR, MA, ARMA) [10], [12], [15], [16]. These traffic models have a correlation structure that is significant for relatively small lags. Long-range dependent traffic models, such as the fractional autoregressive integrated moving average (F-ARIMA) and fractional brownian motion have significant correlation even for large lags [3], [4], [7], [8], [11], [20]. In most cases, actual traffic does not fulfill the stationary assumption, whereas it does satisfy nonstationary, uncertain, and even nonlinear assumptions.

Recently, fuzzy modeling has been developed to very successfully represent real linear and nonlinear (time invariant or time variant) uncertain systems and it has had excellent application in system control designs. To describe stationary, nonstationary, or nonlinear high-speed traffic in networks, we have proposed a fuzzy-AR model to capture these characteristics of actual traffic. This model is easier to implement in a digital computer and is more persuasive than the conventional models such as AR, ARMA, ARIMA, TES, and DAR regression models [1], [2].

In this fuzzy-AR model, the actual traffic data is divided into several clusters by fuzzy partitions, with each cluster described by an AR model. In this situation, the actual traffic process is described by combining these AR processes via fuzzy clustering rules [27]–[32], [36]. This model has been found to describe a real process very successfully with nonlinear and nonstationary properties. A clustering algorithm is proposed to provide a systematic procedure to partition this traffic data so that the number of fuzzy rules and the shapes of fuzzy sets in the fuzzy-AR model are determined to achieve good fitness of data.

The suggested clustering algorithm for parameter estimation of fuzzy-AR model is composed of two steps: coarse tuning and fine tuning [27]. The coarse tuning is a fuzzy c-regression modeling (FCRM) algorithm using hyperplane-shaped clustering technique. This is a fast but rough algorithm. The fine

Manuscript received April 5, 1999; revised March 1, 2000.

B.-S. Chen and K.-C. Wang are with the Department of Electrical Engineering, National Tsing-Hua University, Hsin-Chu 30043, Taiwan, R.O.C.

S.-C. Peng is with the Department of Electrical Engineering, National Huwei Institute of Technology, Huwei, Yunlin 632, Taiwan, R.O.C.

Publisher Item Identifier S 1063-6706(00)08462-9.

tuning is a gradient descent algorithm used to precisely adjust parameters of the fuzzy-AR model to achieve precise parameter estimation.

Inherently, the fuzzy-AR model is a prediction-typed modeling. After the parameter estimation of fuzzy-AR model has been assembled from actual traffic data, the prediction of packet traffic can then be solved. Precise traffic prediction is an important factor for the success of congestion control in high-speed networks, and the fuzzy-AR model can produce reliable and accurate forecasts under congested traffic condition. A good prediction method such as this can maximize the use of available capacities in high-speed networks, thus saving transmission time and minimizing congestion costs.

By utilizing real-time traffic information from network facilities, the traffic may be appropriately assigned to each candidate route. However, since traffic flow patterns change all the time, a control strategy based on previous traffic flow patterns may be irrelevant within a millisecond. For example, even if an alternate route selected for diversion is not congested at the time of selection, one part of the chosen route may be congested by the time when the packet data reaches that part of the network. Thus, forecasting future traffic flow variables for each link along diversion routes is a necessary process for selecting the most efficient alternate route. If a prediction model representing the traffic flow-fluctuation over time could be developed, that model would have the potential to provide reliable forecasts for solving this route guidance problem called the admission problem, with the guarantee of QoS.

Generally speaking, the fuzzy-AR model has the advantage of excellent description of stationary or nonstationary packet traffic in high-speed networks. Inherently, it is also suitable for prediction of the packet traffic in broad-band networks. Therefore, the fuzzy-AR model has an important potential application in traffic congestion control. In this study based on AR-prediction model and queueing theory, a simple feedback congestion control scheme is developed to control network traffic to provide efficient management of high-speed networks. Finally, in order to demonstrate these advantages, some comparisons of the proposed fuzzy-AR model with other conventional models are given to illustrate the superiority of the proposed model for the prediction and congestion control of the packet network traffic.

The structure of this paper is as follows. The problem description is given in Section II. Then, in Section III, the parameter estimation of the fuzzy-AR traffic model with the clustering algorithm is described. In Section IV, some comparisons of the fuzzy-AR model with other traffic models are given. Congestion control using traffic prediction is discussed in Section V. In Section VI, simulation examples are provided to demonstrate the validity of the fuzzy-AR model. Section VII gives the conclusions.

II. Problem Description

High-speed networks based on ATM cell transmission are expected to carry a variety of traffic types in an integrated fashion. This differs considerably from the cases of circuit-switched telephone networks, designed primarily to handle voice communication and, from the much more recent packet-switch networks, designed to handle data transmission. A good traffic model plays a significant role in the design and engineering of both of these network types. It appears that the integration of packetized voice, packetized video, packetized images, and computer-generated data traffic, each with its own multi-objective QoS, requires the development of a relatively sophisticated traffic model to carry out accurate design and performance evaluation.

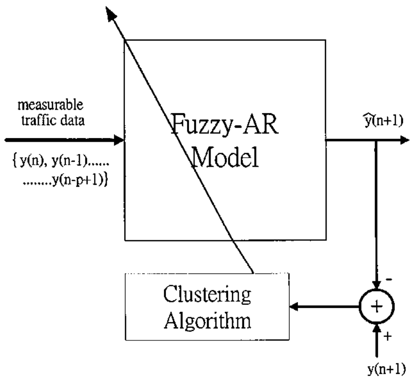

Fig. 1. Fuzzy-AR modeling for prediction of actual traffic data.

With the emergence of diverse communication services such as data, voice, and video, as well as increasing switching and multiplexing in networks [14], [19], the packet traffic of high-speed networks becomes a nonstationary or even nonlinear process. In this situation, conventional linear traffic models are not suitable to capture the traffic behavior of broad-band and high-speed networks. The proposed fuzzy-AR model is a nonlinear mapping of an input data vector to a scale output and it has excellent capability to describe these traffic characteristics.

In general, the fuzzy-AR traffic model contains four components: rules, fuzzifier, inference engine, and defuzzifier. The role of the fuzzifier is to map the crisp input data values to fuzzy sets defined by their membership function, and depending on the degree of "possibility" of the input data. The goal of the defuzzifier is to map the output fuzzy sets to crisp output values. The fuzzy inference engine defines how the system should make inferences through the fuzzy rules contained in the rules base in order to determine the output fuzzy sets.

The fuzzy-AR model can be represented by combining several AR models via fuzzy rules (see Fig. 1) as follows [28], [29]:

$$
\begin{aligned}
\text{Rule } R^l: &\text{ If } y(n) \text{ is } M_1^l\left(q_{11}^l, q_{12}^l\right) \\
&\text{ and } y(n-1) \text{ is } M_2^l\left(q_{21}^l, q_{22}^l\right), \ldots, \\
&\text{ and } y(n-p+1) \text{ is } M_p^l\left(q_{p1}^l, q_{p2}^l\right) \\
&\text{ then } y_l(n+1) \\
&\qquad = a_{l,0} + a_{l,1} y(n) + \cdots + a_{l,p} y(n-p+1)
\end{aligned} \tag{1}
$$

where $R^l (l = 1, 2, \ldots, c)$ denotes the lth fuzzy rule; $y(n-j)$ $(j = 0, 1, \ldots, p-1)$ is the measurable traffic data; $y_l(n+1)$ is the output of the fuzzy rule R^l. $a_{l,i}$, $i = 0, 1, \ldots, p$ can be considered as a residue of the model. Inherently, this is a prediction-typed model. M_j^ls are bell-typed membership functions

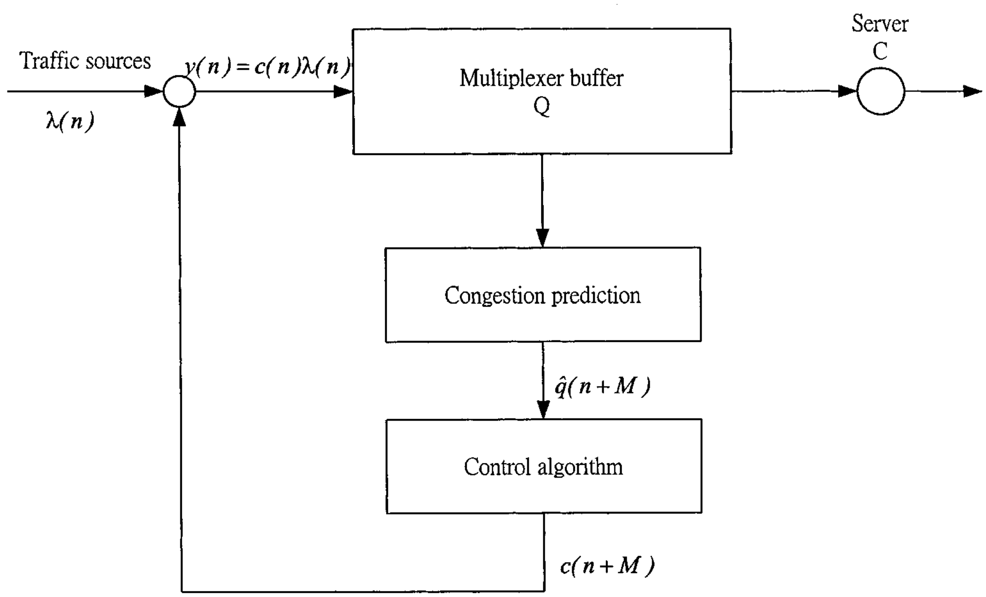

Fig. 2. Traffic control algorithm.

with only two variables, i.e., $q_{jk}^l, k=1,2$ in the jth fuzzy membership function of the lth fuzzy rule

$$M_j^l\left(q_{j1}^l, q_{j2}^l\right) = \exp\left\{-\left(\frac{y(n-j)-q_{j1}^l}{q_{j2}^l}\right)^2\right\}. \quad (2)$$

The fuzzy-AR model in (1) can be represented by a weighted average of the $y_l(n+1)$s, where the weight $w_l(n)$ of the $y_l(n+1)$ is determined by an operation performed on the membership functions associated with that output of the fuzzy-AR model [27]–[29]

$$\hat{y}(n+1) = \frac{\sum_{l=1}^{c} w_l y_l(n+1)}{\sum_{l=1}^{c} w_l}$$
$$w_l = \prod_{j=1}^{p} M_j^l\left(q_{j1}^l, q_{j2}^l\right) \quad (3)$$

i.e., the fuzzy-AR model in (1) is equivalent to the weighted AR model in (3), i.e., a nonlinear and time-variant weighting of local AR models. Due to its simplicity, the equivalent fuzzy-AR model in (3) is employed in this study for modeling and prediction of high-speed packet network traffic. The fuzzy-AR model in (3) is determined by the number of input variables and the number of fuzzy variables or membership functions associated with each input variable. In general, increasing the number of the fuzzy rules or the number of coefficients at the AR model results in a more accurate model, but at the expense of increased computational complexity. Choosing the appropriate number of parameters is a very important problem because the number of fuzzy rules and coefficients in the fuzzy rule will affect the performance of the fuzzy-AR model and the computational complexity. To describe traffic characteristics of high-speed networks, the problem of estimating the parameters q_{jk}^l ($l=1,2,\ldots,c, j=1,2,\ldots,p$ and $k=1,2$) and $a_{l,i}$ ($l=1,2,\ldots,c$ and $i=0,1,\ldots,p$) of the fuzzy-AR model will be discussed in the next section. In brief, the estimation and prediction problems involve how to find a suitable number of rules, a proper partition of the feature space, and the accuracy of parameter estimation.

Traffic conditions in networks vary considerably over time. In order to look ahead, a prediction model that describes the time-varying traffic conditions throughout the high-speed network is needed. Here, three variables (i.e., current traffic information, past traffic information, and prediction traffic information) are investigated to assess the variability in traffic conditions. Current traffic information is one of essential means of estimating traffic trends and it is a component of the prediction model. Past traffic information is used to smooth out abrupt changes in the current traffic flow to avoid extreme forecasts. Traffic prediction information is introduced to represent the dynamic nature of traffic flow. The quality of the fuzzy-based AR prediction model depends on the way by which those three variables are combined.

If all the link data throughout the network are available, it will be possible to more efficiently control the traffic of the high-speed network. In order to design such a control, an advanced and precise prediction model is required. In other words, a high-speed network scheme is to keep the promise of ubiquity, convenience, affordability, and reliability. Thus, an accurate system model must be constructed to provide acceptably precise performance prediction in a reasonable amount of time. The fuzzy-AR model provides a powerful method for predicting the traffic in high-speed networks. In particular, effective traffic estimation is required when there are corresponding demands for network resources. Needless to say, poor prediction invariably leads to inappropriate design and management decisions, which, in addition to impacting users in tangible adverse ways, can also bring about a sense of disappointment and a perception that the new technology is being “oversold” to the public.

In order to demonstrate the excellence of the fuzzy-AR model for high-speed network traffic, the convergence analysis of the proposed model is evaluated with actual traffic data. Generally speaking, the fuzzy-AR model can be tuned to approximate any linear or nonlinear dynamic of packet network traffic and, thus, represent traffic behaviors. The first purpose of this current work is to present a new traffic model based on fuzzy clustering algorithm with coarse tuning and fine tuning techniques. This model captures the characteristics of the packet network traffic and at the same time solves the prediction problem for packet network traffic.

High-speed networks offer great flexibility and efficiency by statistically multiplexing different types of multimedia traffic with different QoS requirements (e.g., cell loss rate, delay, and delay jitter) and a broad range of statistical characteristics. However, the additional flexibility to accommodate different traffic sources may cause serious congestion problems, which will result in severe buffer overflow, cell loss and degradation in the required QoS. Thus, effective congestion control and traffic management schemes have to be developed. The main function of traffic control is to regulate the flow of the traffic into the network such that it approximately matches the capacity of the network's limited resources (see Fig. 2). One of the problems that makes congestion control in high-speed networks difficult is the uncertainty and highly time-varying nature of the diverse mix of traffic sources. Furthermore, due to the very small cell transmission time and small buffer sizes, it is imperative that any effective congestion control algorithm must be simple with minimal reaction time. In this work, we propose a feedback traffic control algorithm which is based on the proposed AR-prediction. The main control action in our algorithm is to reduce the peak rate of traffic sources when the AR-prediction scheme predicts possible congestion in the multiplexer. The motivation behind our use of this AR-fuzzy prediction model is that it is very effective in learning and predicting nonlinear complex systems, these making it an ideal tool to handle high-speed networks.

In the next section, we introduce a useful approach to identify the proposed fuzzy-AR model using real traffic data. This identification method includes a parameter estimation for the membership functions and a parameter estimation for local AR models. After the parameter estimations, the prediction of actual traffic data can then be successfully solved. Based on the fuzzy AR-prediction of possible congestion, a simple feedback traffic control algorithm is developed to efficiently mitigate the congestion of network traffic.

III. Parameter Estimation of Fuzzy-AR Traffic Model Using a Clustering Algorithm

In this section, we introduce a fuzzy clustering algorithm, which both combines the excellent system description capability of network traffic by the fuzzy-AR model and also simplifies the identification method of fuzzy-AR model. In this study, the clustering algorithm is presented to estimate the parameters of the fuzzy-AR model based on the prediction of the packet network traffic. Details are given below.

First, we propose an identification method to estimate the parameters of the fuzzy-AR model (3) from real traffic data. The proposed identification method consists of a cascade of a coarse tuning algorithm (i.e., FCRM algorithm) [27]–[32] and a fine-tuning algorithm (i.e., gradient descent algorithm). The traffic data is clustered by the proposed method to provide the structure of the fuzzy-AR model. In order to obtain the fuzzy-AR model, the identification algorithm should include:

- the choice of the number of fuzzy rules;
- the parameter learning of the membership functions;
- the parameter estimation of the local AR models.

In brief, the suggested fuzzy-AR modeling algorithm is composed of two steps: the first one is a coarse tuning, which determines consequent parameters roughly using FCRM clustering, and the second one is a fine tuning, which adjusts the premise and consequent parameters more precisely by gradient descent algorithm. Both tuning procedures are repeated to find an appropriate number c of clusters.

In the following paragraphs, we show how to determine the optimum consequence and premise parameters are determined in order to minimize the performance index. The performance index has been defined as a root-mean-square (rms) of the output errors, i.e., the differences between the traffic data of the actual network and those of fuzzy-AR traffic model.

A. Rough Tuning by FCRM Algorithm

The fuzzy-AR traffic model in (1)–(3) can be represented by the following:

$$y_l(n+1) = a_{l,0} + a_{l,1}y(n) + \cdots + a_{l,p}y(n-p+1) \quad (4)$$

$$\hat{y}(n+1) = \frac{\sum_{l=1}^{c} w_l y_l(n+1)}{\sum_{l=1}^{c} w_l}. \quad (5)$$

The linear local AR model in the lth cluster can be rewritten by the following vector equation:

$$y_l(n+1) = \mathbf{Z}^T(n)\mathbf{A}_l \quad (6)$$

where $\mathbf{Z}(n) = [1, y(n), y(n-1), \ldots, y(n-p+1)]^T$ and $\mathbf{A}_l = [a_{l,0}, a_{l,1}, \ldots, a_{l,p}]^T$. When a set of traffic data $(\mathbf{Z}(n), y(n+1))$, $n = 0, 1, 2, \ldots$ is given, we can obtain the AR model parameters $\mathbf{A}_l$ $(l = 1, 2, \ldots, c)$ roughly by the following steps.

Step 1) Set $M = 0$ and assume data $S = (\mathbf{Z}(0), y(1)), \ldots, (\mathbf{Z}(n-1), y(n))$. Define a $c \times n$ weighting matrix U as in [36]

$$U = \begin{bmatrix} u_{1,1} & u_{1,2} & \cdots & u_{1,n} \\ u_{2,1} & u_{2,2} & \cdots & u_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ u_{c,1} & u_{c,2} & \cdots & u_{c,n} \end{bmatrix}$$

$$0 \le u_{l,j} \le 1, \quad \text{where } 1 \le l \le c, \quad 1 \le j \le n \quad (7)$$

and

$$\sum_{l=1}^{c} u_{l,j} = 1, \quad \forall j = 1, \ldots, n. \quad (8)$$

Initialize the weighting matrix U with random values between zero and one such that the constraints in (8) are satisfied and they initialize the parameters $\mathbf{A}_l = 0$.

Step 2) Define the cost function for FCRM at the Mth iteration as the following equation as in [27], [36]:

$$J=\sum_{l=1}^{c} J_l=\sum_{l=1}^{c}\sum_{j=1}^{n} u_{l,j}^{m} d_{l,j}^{2} \tag{9}$$

where $u_{l,j}$ is between zero and one. The equation $d_{l,j} = \|y(j) - \mathbf{Z}^T(j-1)\mathbf{A}_l\|$ denotes the Euclidean norm between the jth actual traffic data and the output of the lth fuzzy-AR model and $J_l = \sum_{j=1}^{n} u_{l,j}^{m} d_{l,j}^{2}$.

To find the optimum solution under the weighting constraints in (8), the necessary conditions for (9) to reach a minimum can be found by forming a new cost function Φ, which rewrites (8) and (9) as follows:

$$\begin{aligned}\Phi &= J+\sum_{j=1}^{n}\lambda_j\left(\sum_{l=1}^{c}u_{l,j}-1\right)\\ &=\sum_{l=1}^{c}\sum_{j=1}^{n}u_{l,j}^{m}d_{l,j}^{2}+\sum_{j=1}^{n}\lambda_j\left(\sum_{l=1}^{c}u_{l,j}-1\right)\end{aligned} \tag{10}$$

where $\lambda_j,\ (j=1,2,\ldots,n)$ are the Lagrange multipliers for the n constraints in (8). By differentiating Φ with $u_{l,j}$ ($l=1,2,\ldots,c$ and $j=1,2,\ldots,n$), the necessary condition for (9) to reach its minimum is

$$u_{l,j}=\frac{1}{\sum_{k=1}^{c}\left(\frac{d_{l,j}}{d_{k,j}}\right)^{2/(m-1)}}. \tag{11}$$

At the Mth iteration, compute the cost function according (9) and compute a new U using (11).

Step 3) If either the cost function is below a certain tolerance value or its improvement over previous iteration is below a certain threshold then stop; otherwise, go to Step 4).

Step 4) Calculate new cluster representatives at the $(M+1)$th iteration, $y_l(n+1) = \mathbf{Z}^T(n)\mathbf{A}_l$ using $u_{l,k}$ obtained in Step 2) and then a weight recursive least square (WRLS) algorithm is given as follows:

$$\mathbf{A}_l(k+1)=\mathbf{A}_l(k)+\mathbf{H}(k)[y(k+1)-\mathbf{Z}^T(k)\mathbf{A}_l(k)] \tag{12}$$

$$\begin{aligned}\mathbf{H}(k)&=\mathbf{S}(k+1)\mathbf{Z}(k)\\ &=\frac{\mathbf{S}(k)\mathbf{Z}(k)}{\frac{1}{u_{l,k}}+\mathbf{Z}^T(k)\mathbf{S}(k)\mathbf{Z}(k)}\end{aligned} \tag{13}$$

$$\mathbf{S}(k+1)=[I-\mathbf{H}(k)\cdot\mathbf{Z}^T(k)]\cdot\mathbf{S}(k) \tag{14}$$

where $k=1,2,\ldots,n$ and $l=1,2,\ldots,c$. The initial value of the algorithm is $\mathbf{S}(0)=\alpha I$

Step 5) Go to Step 2) and $M=M+1$.

For bell-typed membership functions (2), the parameter estimations of $q_{j,1}^{l}$ and $q_{j,2}^{l}$ can be roughly obtained from the weighting matrix U in above steps [27].

$$q_{j,1}^{l}=\frac{\sum_{k=1}^{n}u_{l,k}y(k-j)}{\sum_{k=1}^{n}u_{l,k}} \tag{15}$$

$$q_{j,2}^{l}=\sqrt{2\cdot\frac{\sum_{k=1}^{n}u_{l,k}\left(y(k-j)-q_{j,1}^{l}\right)^2}{\sum_{k=1}^{n}u_{l,k}}}. \tag{16}$$

The FCRM clustering algorithm can adjust the certainty grade of each fuzzy-AR model by its classification performance; that is, when the traffic data is misclassified by the algorithm, the certainty grade of the fuzzy-AR model is decreased. On the contrary, when the traffic data is correctly classified, the certainty grade is increased.

After the rough tuning based on the FCRM algorithm, a fine tuning based on gradient descent algorithm is performed.

B. Fine Tuning by Gradient Descent Algorithm

From (4) and (5), we can write the prediction traffic data $\hat{y}(n+1)$ in the fuzzy-AR traffic model as follows:

$$\begin{aligned}\hat{y}(n+1)&=\frac{\sum_{l=1}^{c}w_l\left(n,q_{j,k}^{l}\right)y_l(n+1)}{\sum_{j=1}^{c}w_j\left(n,q_{j,k}^{l}\right)}\\ &=\sum_{l=1}^{c}\left(\frac{w_l\left(n,q_{j,k}^{l}\right)}{\sum_{j=1}^{c}w_j\left(n,q_{j,k}^{l}\right)}\right)y_l(n+1)\\ &=\sum_{l=1}^{c}\beta_l\left(n,q_{j,k}^{l}\right)y_l(n+1)\end{aligned} \tag{17}$$

where $\beta_l(n,q_{j,k}^{l}) = (w_l(n,q_{j,k}^{l})/\sum_{j=1}^{c}w_j(n,q_{j,k}^{l}))$, and $q_{j,k}^{l}$s are parameters of the membership function M_j^l and

$$y_l(n+1)=a_{l,0}+a_{l,1}y(n)+\cdots+a_{l,p}y(n-p+1) \tag{18}$$

$$\begin{aligned}\hat{y}(n+1)&=\sum_{l=1}^{c}\beta_l\left(n,q_{j,k}^{l}\right)[a_{l,0}+a_{l,1}y(n)\\ &\quad+\cdots+a_{l,p}y(n-p+1)]\\ &=\mathbf{V}^T\left(n,q_{j,k}^{l}\right)\mathbf{A}\end{aligned} \tag{19}$$

where $\mathbf{V}(n,q_{j,k}^{l}) = [\beta_1(n,q_{j,k}^{l}),\beta_2(n,q_{j,k}^{l}),\ldots,\beta_1(n,q_{j,k}^{l})y(n),\ldots,\beta_c(n,q_{j,k}^{l})y(n),\ldots,\beta_c(n,q_{j,k}^{l})y(n-p+1)]^T \in R^{M\times 1}$, the consequent parameter $\mathbf{A} = [a_{1,0},a_{2,0},\ldots,a_{c,0},a_{1,1},\ldots,a_{c,1},\ldots,a_{c,p}]^T \in R^{M\times 1}$, and M is the number of terms in (19). The task of parameter estimation in the fuzzy-AR clustering algorithm is to determine $q_{j,k}^{l}$ and $\mathbf{A}$, by minimizing the following criterion:

$$\begin{aligned}J\left(q_{j,k}^{l},A\right)&=\|\mathbf{Y}-\hat{\mathbf{Y}}\|^2\\ &=\left\|\mathbf{Y}-\mathbf{\Pi}\left(q_{j,k}^{l}\right)\mathbf{A}\right\|^2\end{aligned} \tag{20}$$

where $\mathbf{Y} = [y(1),y(2),\ldots,y(N)]^T \in R^{N\times 1}$, $\hat{\mathbf{Y}} = [\hat{y}(1),\hat{y}(2),\ldots,\hat{y}(N)]^T \in R^{N\times 1}$, and (see (21) at the bottom of the next page).

Remark 1: In [32], it was previously shown that the parameter estimation of the fuzzy-AR model by the gradient descent algorithm can be divided in two steps. The optimal consequent parameter $\mathbf{A}^*$ can be obtained by the least squares method, which is a linear estimation problem. However, determining the optimization $q_{j,k}^{l^*}$ is a nonlinear estimation problem.

The iterative method of the fuzzy-AR traffic model based on the above minimum square error can be improved by a fine fuzzy partition. This fine tuning method consists of two procedures, i.e., the premise parameters procedure and the consequent parameters procedure [27], [31], [36].

a) *Premise parameters procedure:* According to the fuzzy-AR model in (4) and (5), the premise parameter q_{jk}^ls of the fuzzy-AR model can be precisely adjusted by the following learning rule:

$$\begin{aligned}\triangle q_{jk}^l(n+1) &= q_{jk}^l(n+1) - q_{jk}^l(n)\\ &= -\eta \frac{\partial}{\partial q_{jk}^l}\left(\frac{e^2(n)}{2}\right)\\ &= \eta(y(n) - \hat{y}(n))\\ &\quad \cdot (y_l(n) - \hat{y}(n)) \frac{1}{\sum_{l=1}^{c} w_l} \frac{\partial w_l}{\partial q_{jk}^l}\end{aligned}$$

or

$$\begin{aligned}q_{jk}^l(n+1) &= q_{jk}^l(n) + \eta(y(n) - \hat{y}(n))\\ &\quad \cdot (y_l(n) - \hat{y}(n)) \frac{1}{\sum_{l=1}^{c} w_l} \frac{\partial w_l}{\partial q_{jk}^l}\end{aligned} \tag{22}$$

where η is a premise learning rate, $y(n)$ is the actual traffic data, $\hat{y}(n)$ is the output of the fuzzy-AR model, and $e(n) = y(n) - \hat{y}(n)$.

b) *Consequent parameters procedure:* The consequent parameters $\mathbf{A}$ of the fuzzy-AR traffic model in (4) and (5) can be finely adjusted by the following learning method:

$$\begin{aligned}\triangle a_{l,j}(n+1) &= a_{l,j}(n+1) - a_{l,j}(n)\\ &= -\gamma \frac{\partial}{\partial a_{l,j}}\left(\frac{e^2(n)}{2}\right)\\ &= \gamma(y(n) - \hat{y}(n)) \frac{w_l y(n-j)}{\sum_{l=1}^{c} w_l}\end{aligned}$$

or

$$a_{l,j}(n+1) = a_{l,j}(n) + \gamma(y(n) - \hat{y}(n)) \frac{w_l y(n-j)}{\sum_{l=1}^{c} w_l} \tag{23}$$

where γ is a consequent learning rate.

After the parameters of $q_{j,k}^l$ and $a_{l,j}$ have been estimated from real traffic data via the proposed rough and fine tuning algorithms in the above subsections, the fuzzy-AR modeling in (4) and (5) is finished. Then, based on the constructed fuzzy-AR model in (4) and (5), a one-step-ahead prediction $\hat{y}(n+1)$ can be obtained from the previous data $y(n), y(n-1), \ldots, y(n-p+1)$.

Remark 2: The main computations of rough tuning are in the recursive estimation of $\mathbf{A}_l(k)$ in (12) and in the estimation of $q_{j,1}^l$ in (15) and $q_{j,2}^l$ in (16). The main computations of fine tuning are in the recursive estimations of $q_{j,1}^l$ and $q_{j,2}^l$ in (22) and $a_{l,j}(n)$ in (23). All of these recursive parameter estimations can be calculated in real time if the computer computation capability is fast enough.

Remark 3: Initially, rough tuning plays the most important role of parameter estimation. As the parameter estimation converges, only the fine algorithm is necessary. In this situation, the rough tuning algorithm can be cut off and the computing load is reduced.

Remark 4: Because the fuzzy-AR model has nonlinear behavior, accurate analysis is difficult. However, in [31] it was shown that the fuzzy-AR model in (4) and (5) has universal approximation capabilities. That is, for any given packet traffic data $y(n+1)$ and arbitrary $\epsilon > 0$, such that $\sup_n |y(n+1) - \hat{y}(n+1)| < \epsilon$, if the number and the order of the fuzzy-AR model are sufficiently large.

Remark 5: For the initial condition at $k = 0$, we have $\hat{A}_l(0) = 0$ and $S(0) = \alpha I$, where α is a small positive constant and I represents an identity matrix. For the AR model in (4), the M-step ahead traffic prediction is recursively obtained by

$$\begin{aligned}\hat{y}_l(n+1) &= Z^T(n)\hat{A}_l(n)\\ \hat{y}_l(n+2) &= Z_1^T(n)\hat{A}_l(n)\\ &\vdots\\ \hat{y}_l(n+m) &= Z_{m-1}^T(n)\hat{A}_l(n)\\ &\vdots\\ \hat{y}_l(n+M) &= Z_{M-1}^T(n)\hat{A}_l(n)\end{aligned} \tag{24}$$

where

$$\begin{aligned}Z_{m-1}^T(n) = [1 \quad &\hat{y}_l(n+m-1) \quad \hat{y}_l(n+m-2) \quad \cdots\\ &\hat{y}_l(n+1) \quad y_l(n) \quad \cdots \quad y_l(n+m-p+1)]\end{aligned}$$

and $\hat{A}_l(n)$ is obtained recursively from the proposed rough and fine tuning algorithm in the above subsections.

Summing up the above discussion, the fuzzy-AR modeling has excellent convergence characteristics with actual traffic data in a high-speed network. In other words, the fuzzy-AR model could accurately capture real network traffic behavior. Based on the clustering algorithm, the fuzzy-AR model is proposed to predict the packet traffic in high-speed network. Using the fuzzy-AR model to predict the packet traffic in high-speed network, we do not need to assume any statistical characteristic of the real traffic data. However, this procedure relies on the knowledge of a set of the past packet traffic values and current traffic information. The prediction with the fuzzy-AR model allows

$$\Pi\left(q_{j,k}^l\right) = \begin{bmatrix}\mathbf{V}^T\left(1, q_{j,k}^l\right)\\ \mathbf{V}^T\left(2, q_{j,k}^l\right)\\ \vdots\\ \mathbf{V}^T\left(N, q_{j,k}^l\right)\end{bmatrix} = \begin{bmatrix} v_0\left(1, q_{j,k}^l\right) & v_1\left(1, q_{j,k}^l\right) & \cdots & v_M\left(1, q_{j,k}^l\right)\\ v_0\left(2, q_{j,k}^l\right) & v_1\left(2, q_{j,k}^l\right) & \cdots & v_M\left(2, q_{j,k}^l\right)\\ \vdots & \vdots & \ddots & \vdots\\ v_0\left(N, q_{j,k}^l\right) & v_1\left(N, q_{j,k}^l\right) & \cdots & v_M\left(N, q_{j,k}^l\right)\end{bmatrix} \in R^{N\times M} \tag{21}$$

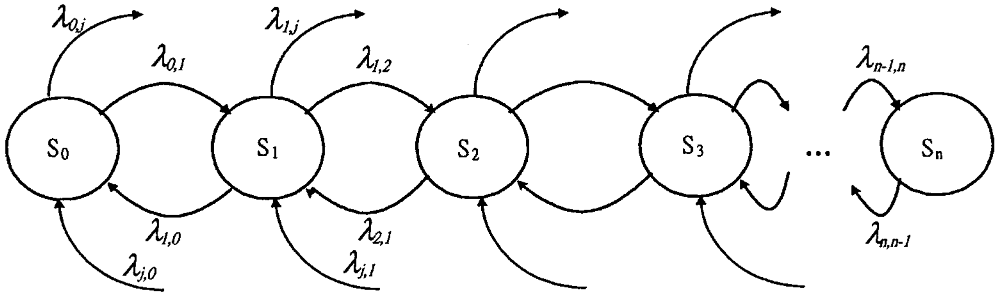

Fig. 3. MMPP process.

us to obtain a good approximation of the actual traffic for the development of the congestion control, and the multiplexer service capacity. Recently, due to reduction in price and increase in power of hardware, the computation cost has been reduced substantially. In this situation, the recursive parameter estimation algorithms in rough tuning and fine tuning of fuzzy-AR model for high-speed network can be performed in real time. In the next section, we will demonstrate the advantage of the fuzzy-AR traffic model by comparing it with other traffic models.

IV. Comparisons with Other Traffic Models

We begin by discussing a number of traffic models described in the literature, which represent network traffic, then pointing out some defects of these models. Furthermore, we will illustrate that the fuzzy-AR model could overcome the shortcomings in conventional traffic models and that this fuzzy-AR model provides excellent performance in capturing the traffic properties of high-speed networks.

A. Poisson Model

Poisson models are the oldest traffic models. A Poisson process can be characterized as a renewal process whose interarrival times X_n $(n = 1, 2, \ldots)$ are exponential distributed with rate $\lambda : p(X_n \leq t) = 1 - \exp(-\lambda t)$, and where the number of arrivals in disjoint intervals is statistically independent [10]. Poisson processes have some elegant analytical properties [1], [2], [16], [17]. First, the superposition of independent Poisson processes results in a new Poisson process whose rate is the sum the component rates. Second, the independent increment property renders Poisson a memoryless process. Third, Poisson processes are fairly common in traffic applications that physically comprise a large number of independent traffic streams, each of which may be quite general. However, the Poisson arrival process, used extensively in network performance analysis and the prediction of packet traffic, does not work, particularly when dealing with the stream-type continuous bit-rate traffic formed in high-speed networks. Recently, measurements of both LAN and wide-area TCP (Internet) traffic indicate that the Poisson model breaks down in a number of real-life application areas [25], [26]. The reason for this is that conventional Poisson processes cannot capture the long-range dependence of high-speed network traffic. Since the fuzzy-AR model via clustering algorithm has the capability of classifying different types of traffic data into different clusters, it could be applied to analyze and predict the mixed traffic data in broad-band networks. In other words, we could classify the data packet, voice packet and video packet into the different clusters by fuzzy clustering algorithm. Therefore, this model can capture both the short-range dependence and long-range dependence of packet traffic of high-speed networks. After clustering the mixed traffic data, we can analyze and successfully predict the packet traffic.

B. Markov-Modulated Poisson Process Model

The most commonly used Markov-modulated model is the Markov-modulated Poisson process (MMPP) model, which combines the simplicity of the modulating (Markov) process with that of the modulated (Poisson) processes (see Fig. 3). In this model, while in state S_k, arrivals occur according to a Poisson process with rate λ_k. The MMPP process allows the modeling of time-varying source, while keeping tractable the analytical solution of related queueing performance. The problems of MMPP are that the parameters estimation of MMPP (i.e., the arrival rate λ_k, the transition rate, and the suitable number of states S_k) are very difficult. To simplify the complexity of the MMPP model, the MMPP model could not have an infinite number of states and it must quantize the model parameters into finite number of the model parameters. Therefore, the MMPP model will produce quantization errors and serious modeling errors. Furthermore, the MMPP model is not suitable for packet traffic prediction. The fuzzy-AR model has an advantage over MMPP model in that it determines the model parameters easily and accurately. The parameters estimation of the fuzzy-AR model has self-learning capability by training via input traffic data. And the quantization of parameters is not needed in the fuzzy-AR model. Therefore, the fuzzy-AR model is more powerful and accurate than the MMPP model.

C. Autoregressive Traffic Model

The autoregressive model of order p has the following form:

$$y(n+1) = a_0 + \sum_{k=1}^{p} a_k y(n-k+1) + e(n) \quad n > 0 \quad (25)$$

where $y(n-k+1)$s are the real traffic data, $e(n)$ is zero-mean, independent identically distributed (i.i.d.) random variable, and $a_k, k = 0, 1, \ldots, p$ are real constants. Such models are particularly suitable for modeling VBR-coded video—a projected major consumer of bandwidth in emerging communication networks. However, the AR model is only suitable to model the case with short-range dependence structures in packet network traffic. Recently, measurements of packet traffic in the internet indicate that the traffic is self-similar or fractal-like (long-range dependence) in nature [35]. Furthermore, the traffic may be time-varying and nonlinear. The term self-similar, means that the statistical characterization of the traffic is essentially invariant with the time scale. Therefore, the AR model fails to capture the behavior of a high-speed network traffic. The clustering, time-varying and nonlinear characteristics of a fuzzy-AR model can take over the phenomenon of self-similarity or fractal property of packet network traffic. From the simulation results in next section, we can see that the fuzzy-AR model has excellent ability to describe the burstiness of packet traffic. Actually, the degree of self-similarity is related to the burstiness of packet traffic. In other words, the fuzzy-AR model can be used to efficiently characterize the self-similarity and long-range dependence structures in high-speed networks.

In the above discussion, based on the clustered, time-varying and nonlinear nature of fuzzy combination of AR clusterings, the fuzzy-AR model is found to be applicable to different types of actual packet traffics in high-speed networks. Based on traffic prediction using a fuzzy AR-model, a simple but efficient congestion control algorithm is proposed in the following section.

V. Congestion Control Using Traffic Prediction

The predicted traffic congestion with current queue information in the buffer can be used as a measure of congestion. Although ATM provides additional flexibility for accommodating different traffic sources, however, it may cause serious congestion problems that will result in severe buffer overflow, cell loss, and degradation in the required QoS. To overcome this, effective congestion control and traffic management are developed in this section.

Consider a multiplexer buffer queue with size of Q, where $y(n)$ is the number of cells that arrive between the time $n-1$ and n. The queue has a constant service rate of C cells per unit time. Denoting by $q(n)$ the queue length at time n, then we have the following Lindley's equation [34], [40] (see Fig. 2) indicating

$$q(n+1) = \min(\max(q(n) + y(n+1) - C, 0), Q) \quad (26)$$

cells are lost if the buffer overflows.

In this study, the occupancy of the queue is considered as an indication of traffic congestion in the multiplexer buffer. In this situation, the congestion prediction of the network is obtained sequentially as follows:

$$\begin{aligned} \hat{q}(n+1) &= \min(\max(q(n) + \hat{y}(n+1) - C, 0), Q) \\ \hat{q}(n+2) &= \min(\max(\hat{q}(n+1) + \hat{y}(n+2) - C, 0), Q) \\ &\vdots \\ \hat{q}(n+M) &= \min(\max(\hat{q}(n+M-1) \\ &\quad + \hat{y}(n+M) - C, 0), Q) \end{aligned} \quad (27)$$

where $\hat{q}(n+i)$ in the right-hand side of (27) is obtained from the previously stage, and traffic predictions $\hat{y}(n+i), i = 1, 2, \ldots, M$ are obtained from the fuzzy-AR prediction algorithm in the previous section.

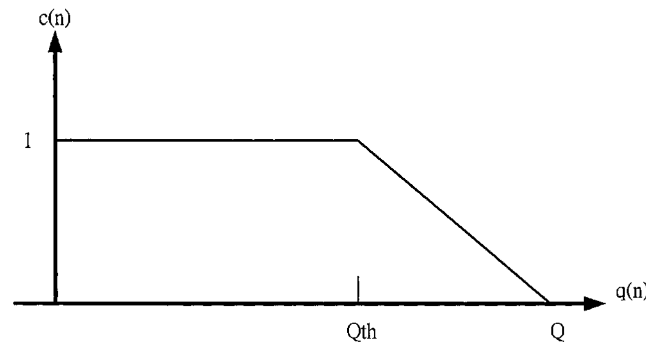

Fig. 4. Control signal $c(n)$.

Due to the very small cell transmission time and small buffer size, any effective congestion control algorithm must be simple with minimal reaction time. In this paper, we propose a feedback traffic control algorithm, which is based on the congestion prediction of the queue length of (27). The main control action of our scheme is to reduce the peak rate of traffic sources when the adaptive AR-prediction algorithm predicts possible congestion in the multiplexer buffer. Consider a multiplexer buffer with size Q and denote a threshold limit of queue length as Q_{th}, for example, $Q_{\mathrm{th}} = 0.5Q$. Our control signal $c(n)$ is generated by the following congestion control algorithm (see Fig. 4)

$$c(n+M) = \begin{cases} 1, & \text{if } \hat{q}(n+M) \le Q_{\mathrm{th}} \\ \frac{Q-\hat{q}(n+M)}{Q-Q_{\mathrm{th}}}, & \text{if } Q_{\mathrm{th}} \le \hat{q}(n+M) \le Q \\ 0, & \text{if } Q \le \hat{q}(n+M) \end{cases} \quad (28)$$

where the congestion prediction $\hat{q}(n+M)$ is obtained from (27) with $M = 1, 2, \ldots$.

The control signal is inversely proportional to the occupancy (congestion) of the queue when $\hat{q}(n+M) > Q_{\mathrm{th}}$. If the $\hat{q}(n+M)$ reaches the threshold Q_{th}, a control signal $c(n+M)$ is sent back and each source reduces its rate. Let us denote the current source rate as $\lambda(n)$. Then the current input rate $y(n)$ to the multiplexer is as follows:

$$y(n) = c(n)\lambda(n). \quad (29)$$

Obviously, this approach is a closed-loop control system. There are several advantages associated with the proposed control algorithm based on feedback throttling. First, it is not of the reactive control type since it is applied at the input access node to the network and its speed is not limited by the propagation delay in the multiplexer. Due to the use of prediction-type control algorithm, any control action taken will be in time to alleviate the potential congestion. Second, it can be widely applied, regardless of the type of encoding used. Third, through decreasing the peak rate, it provides the means for the maximum possible shaping of the input arrival process. Consequently, the bandwidth allocated to the input cell can be reduced, while the same required QoS is still achieved. Also, the statistical multiplexing gain is enhanced since more sources can be supported for each multiplexer. Therefore, the proposed control algorithm is more suitable for traffic smoothing.

In the next section, some practical examples with computer simulations are given with comparisons to demonstrate the

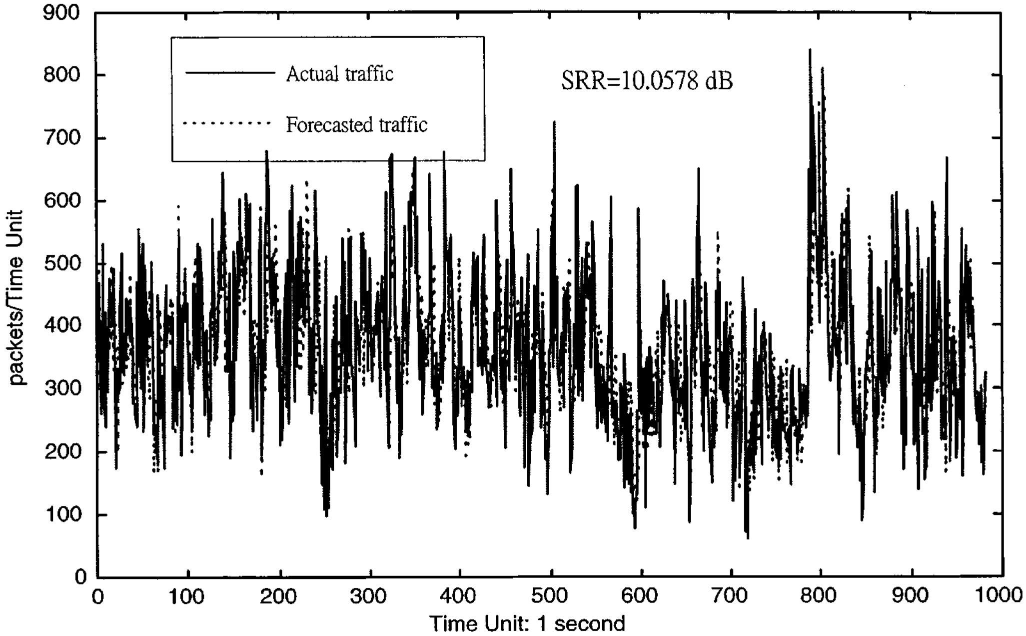

Fig. 5. Actual and forecasted packet traffic using fuzzy-AR model in Case 1 of Example 1.

validity of the prediction and congestion control based on the fuzzy-AR model by the high-speed network traffic data.

VI. Simulation

Example 1: In this example, we use the LAN traffic data collected by Leland and Wilson [3], [32] for simulations. The data were collected between August 1989 and February 1992 on several ethernet LAN's at the Bellcore Morristown Research and Engineering Center, NJ. Leland and Wilson have presented a preliminary statistical analysis of this traffic data, i.e., the self-similar behavior [3], [33]. This self-similar or fractal-like behavior of aggregate ethernet LAN traffic is very different both from conventional telephone traffic and from currently considered formal models for packet traffic. We use four cases to demonstrate the performance of the proposed fuzzy-AR model via the clustering algorithm from actual ethernet traffic data. That is, the proposed traffic model could capture the self-similar characteristics of the ethernet traffic through the following simulation results. In these cases, the actual traffic data were measured from the ethernet LAN's with different time scales. Four cases are discussed as follows.

Case 1: The actual LAN packet traffic data were collected by Leland *et al.* [3] over a 1000 s measurement period. The data measurement in Bellcore, started at 11:25 A.M., 29 August 1989, and has 1 000 000 packet arrivals, although we selected only a sample of the first 1000 time units (time unit $= 1$ s) for parameter estimation and prediction in our simulation.

In this case, we use three fuzzy rules and five consequent parameters in each rule in the fuzzy-AR model to predict the actual traffic data. Applying the clustering algorithm in Section III, we have obtained the simulation results shown in Fig. 5, in which the solid line presents the actual traffic packet $y(n+1)$ and the dotted line represents the prediction result $\hat{y}(n+1)$ by the fuzzy-AR model with the clustering algorithm.

In order to illustrate the performance of our proposed method, the prediction signal-to-error ratio (SRR) is defined as a performance index

$$\mathrm{SRR} = 10 \cdot \log_{10} \frac{E[y^2(n+1)]}{E[e^2(n+1)]} \quad (\mathrm{dB})$$

where $e(n+1) = y(n+1) - \hat{y}(n+1)$. In this case, the SRR is 10.0578 dB.

Case 2: The actual LAN packet traffic data over a 1000 s measurement period collected by Leland *et al.* [3] is used in this case. The traffic data is different from the packet data in Case 1. The traffic data measurement, started at 11:46 P.M., 30 October 1989, has 1 000 000 packet arrivals, although for convenience only a sample of the first 1000 time units (time unit $= 1$ s) is used for parameter estimation and prediction of fuzzy-AR modeling.

In this case, three fuzzy rules and five consequent parameters in each rule are used in the fuzzy-AR model. Fig. 6 shows the actual traffic data $y(n+1)$ and the forecasted packet traffic $\hat{y}(n+1)$ by the proposed fuzzy-AR model. In this case, the prediction performance SRR is 7.8443 dB.

Case 3: The actual LAN packet traffic data over a 10 000 s measurement period collected by Leland *et al.* were used with 10 s as one time unit. In this case, the measured traffic data is identical with Case 2 but different in measurement period (10 000 s) and time unit (10 s). Here, we use three fuzzy rules and five consequent parameters in each fuzzy rule of the fuzzy-AR model. The simulation results are shown in Fig. 7. In this case, the prediction performance SRR is 7.2657 dB. Since the prediction of 10 s ahead is more difficult than the prediction of 1 s ahead, the performance in this case is worse than that of Case 2.

Case 4: The actual LAN packet data over a 100 000 s measurement period collected by Leland *et al.* is used in this case.

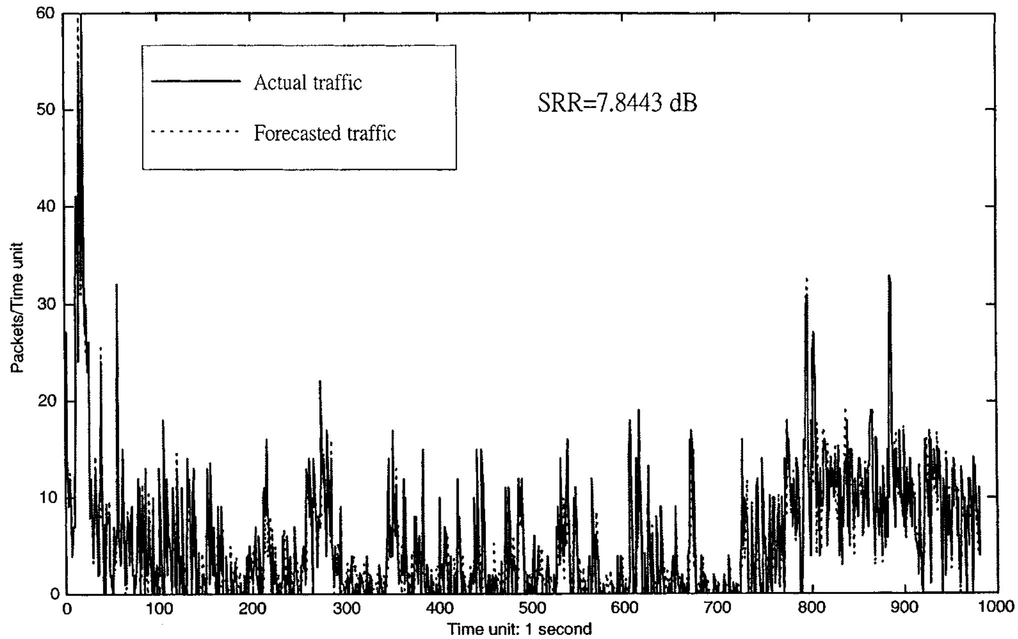

Fig. 6. Actual and forecasted packet traffic using fuzzy-AR model in Case 2 of Example 1.

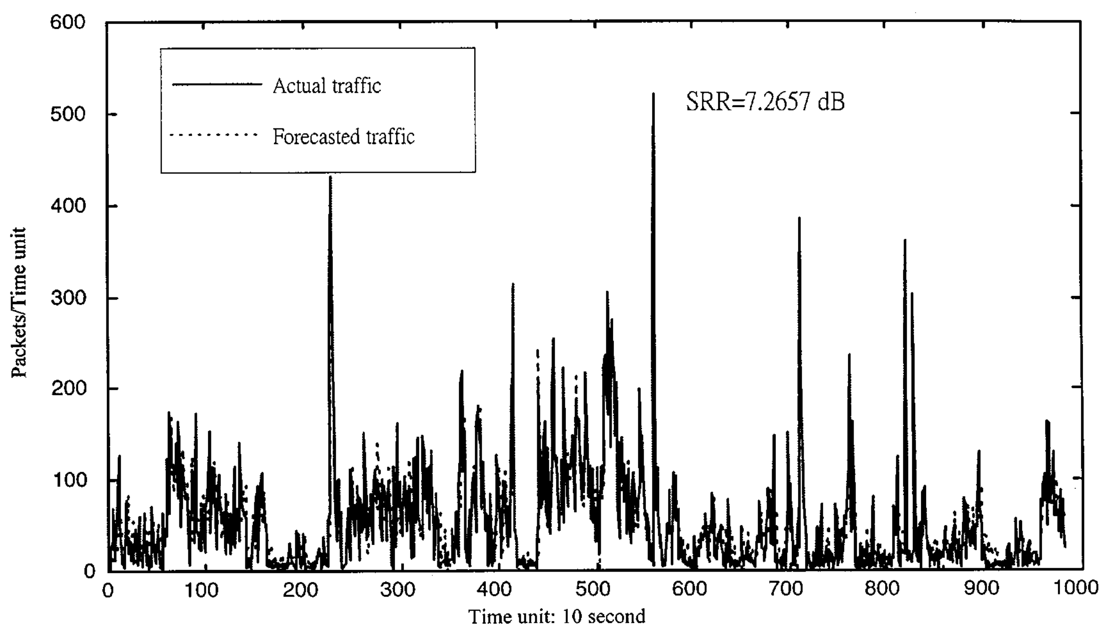

Fig. 7. Actual and forecasted packet traffic using fuzzy-AR model in Case 3 of Example 1.

The measured traffic data is also identical to Case 3 but different in measurement period (100 000 s) and time unit (100 s). The fuzzy-AR model is the same as Case 3. The simulation results are shown in Fig. 8, indicating the prediction SRR is 6.8875 dB. The performance is worse than the previous cases because the prediction of 100 seconds ahead is more difficult than the prediction of either 1 or 10 s ahead. However, the performance can still be considered very good for the prediction of packet network traffic.

Remark 6: The computing times for calculating one iterative tuning to update the parameters in the above four cases are all within 0.0035 s by workstation ultra 2, which is much smaller than the time units in the above four cases. Therefore, the proposed prediction algorithm can be performed in real time and has potential for traffic congestion control.

From Figs. 5–8, it is seen that there is very sharp variation of the packet traffic, hence, we cannot reasonably hope to model the arrivals using a simple traffic model or the conventional traffic model. Modeling ethernet traffic using other conventional models (i.e., Poisson, AR, MMPP models, and so on) cannot accurately reflect the long-range dependence in actual traffic and will significantly underestimate performance measures such as average packet delay or maximum queue size. In comparison, prediction results from the conventional AR model of order 120 in above cases are given in Figs. 9–12. Obviously, this model cannot capture the nonlinear and nonstationary properties of

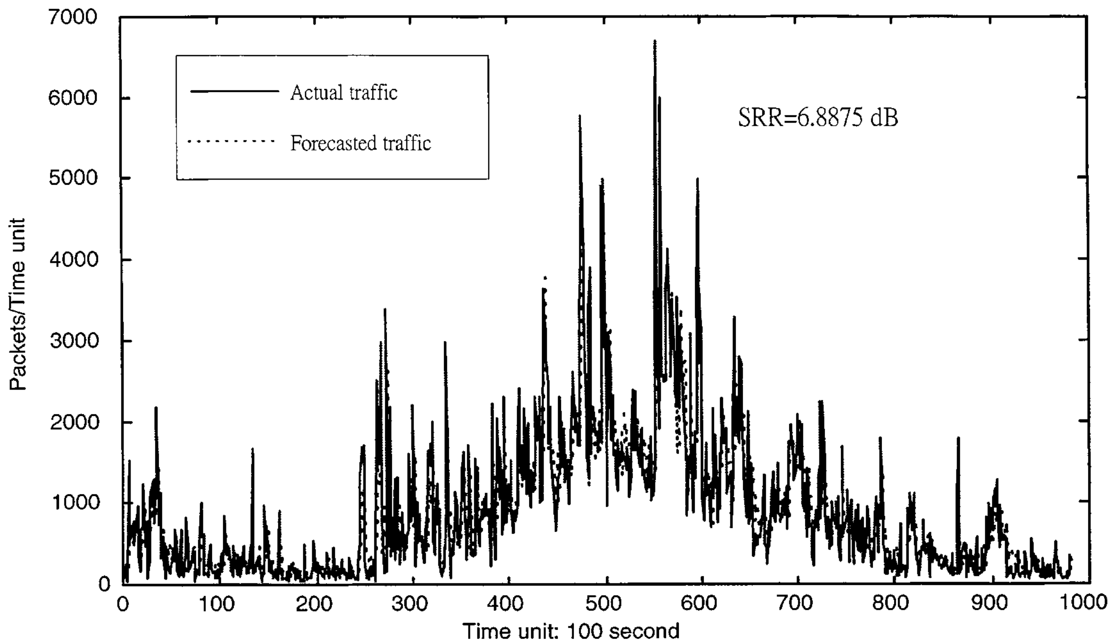

Fig. 8. Actual and forecasted packet traffic using fuzzy-AR model in Case 4 of Example 1.

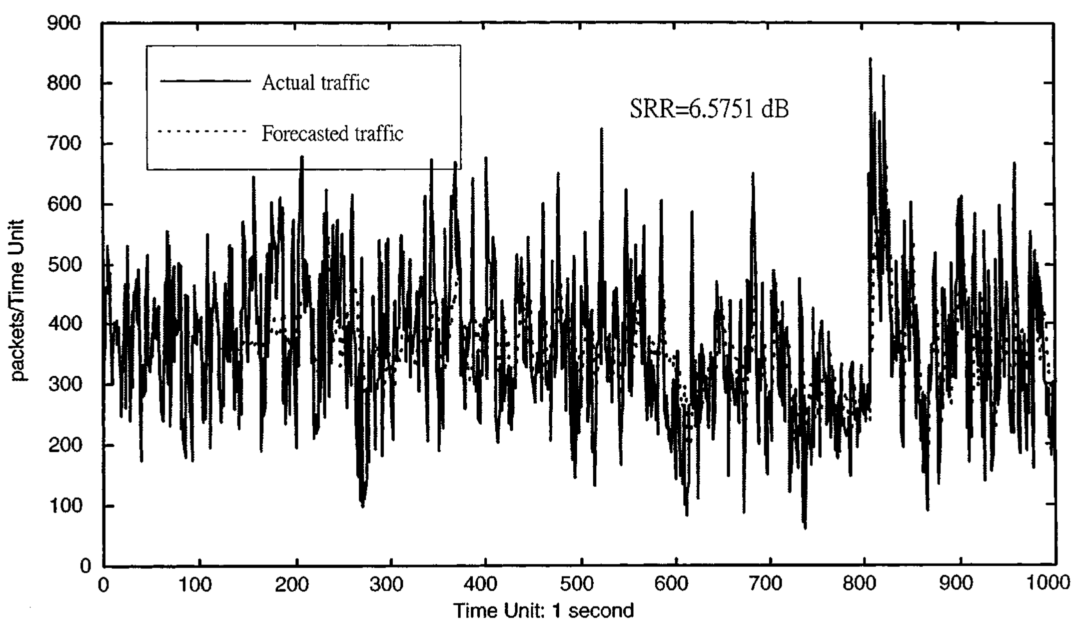

Fig. 9. Actual and forecasted packet traffic using AR model in Case 1 of Example 1.

high-speed networks traffic even with very high order. Through the simulation results, we find that the fuzzy-AR modeling via clustering algorithm has good capability to predict the actual packet traffic in high-speed networks.

Furthermore, in order to exploit the significance of traffic prediction on the congestion control, a congestion control design is given as the following.

Based on the results from the previous section, our congestion control algorithm is proposed as follows: if $0 \leq \hat{q}(k+M) \leq 0.5Q$, then each source does not reduce its rate. If $0.5Q < \hat{q}(k+M) \leq Q$, then a control signal is sent back and each source reduces to $(Q - \hat{q}(n+M)/Q - Q_{\rm th})$ of current rate. In these cases, the cell loss rate with different buffer sizes are shown in Figs. 13–16. Obviously, the congestion control based on the proposed traffic prediction method has superior performance than the other control methods without traffic prediction.

Example 2: In this example, multimedia network traffic with video and voice sources is used to test our proposed method. The video source is the actual VBR video traffic for a movie called *Star Wars*, which contains quite a diverse mixture of material ranging from low-complexity motion scenes to those with very high action. This source use an intraframe compression code similar to JPEG. The traffic is generated from the output of a simple

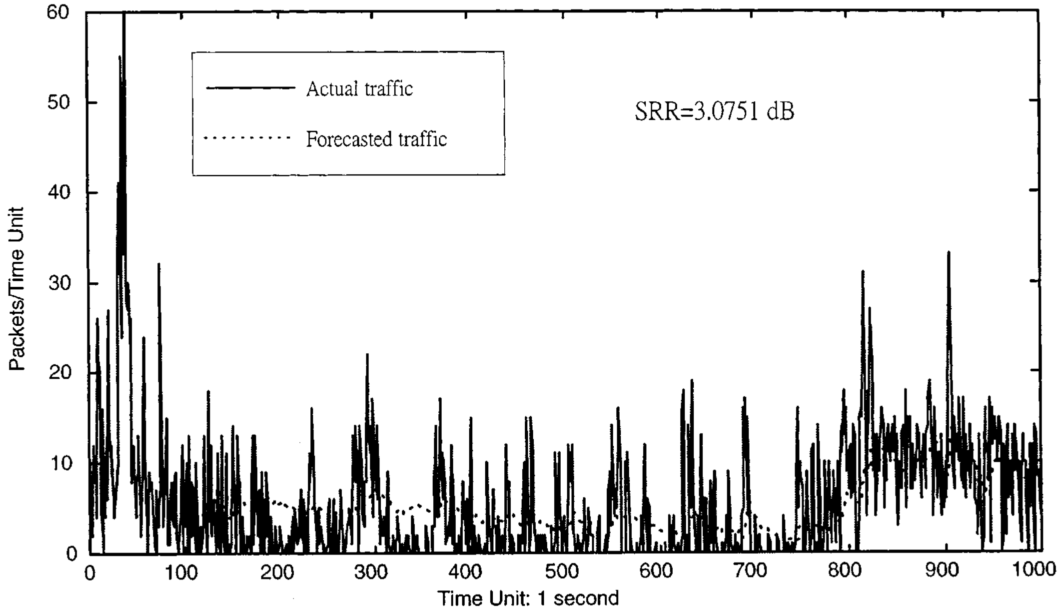

Fig. 10. Actual and forecasted packet traffic using AR model in Case 2 of Example 1.

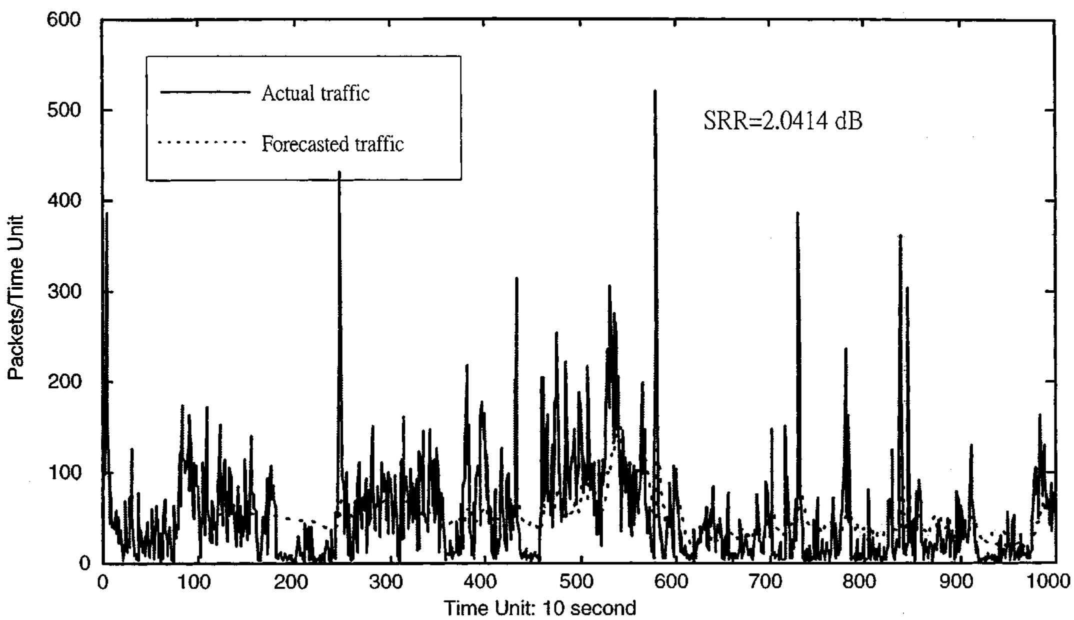

Fig. 11. Actual and forecasted packet traffic using AR model in Case 3 of Example 1.

variable bit-rate coder. Note that only the original film frames are coded (i.e., 24/s). The data was collected in 1993 by Garrett and Vetterli and is being made available to promote research on traffic behavior and control in ATM networks. It is available over the internet via anonymous ftp from ftp.bellcore.com in directory pub/vbr.video.trace. This video traffic has been shown to exhibit long-range dependence and self similarity.

The voice source is produced by an interrupted Poisson process (IPP) [40], [41]. A voice source alternates between talk spurts (active) and silent periods. In an IPP model, each voice source is characterized by ON (corresponding to talk spurts) and OFF (corresponding to silence duration) periods, which appear in turn. During the ON period, the interarrival time of packets are exponentially distributed and no packets are generated during the OFF period. The time spent in ON and OFF states is exponentially distributed with mean $1/\alpha$ and $1/\beta$, respectively. To specify this model completely, we assume that the peak bit-generation rate during the active period is 32 kb/s, the mean bit-generation rate is 11.2 kb/s, the mean talk spurt is $1/\beta = 600$ ms, and the mean silence period is $1/\alpha = 200$ ms. The arrival process is sampled at every sampling period T_s. Here, $T_s = 50$ ms. In the following, we perform several cases to test our proposed prediction method and the feedback congestion control algorithm. Note that in all cases the mean

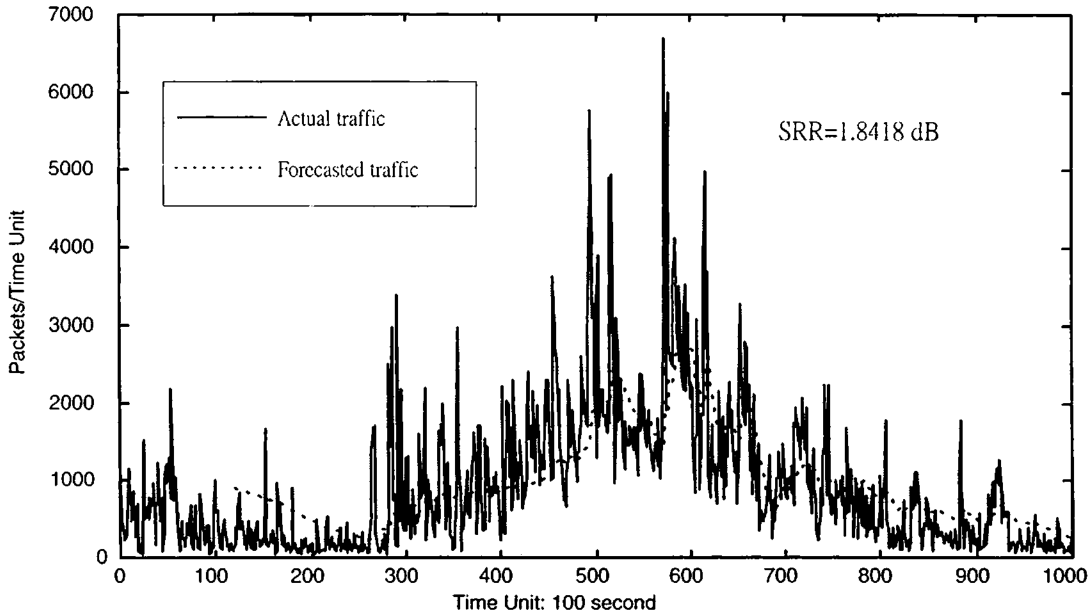

Fig. 12. Actual and forecasted packet traffic using AR model in Case 4 of Example 1.

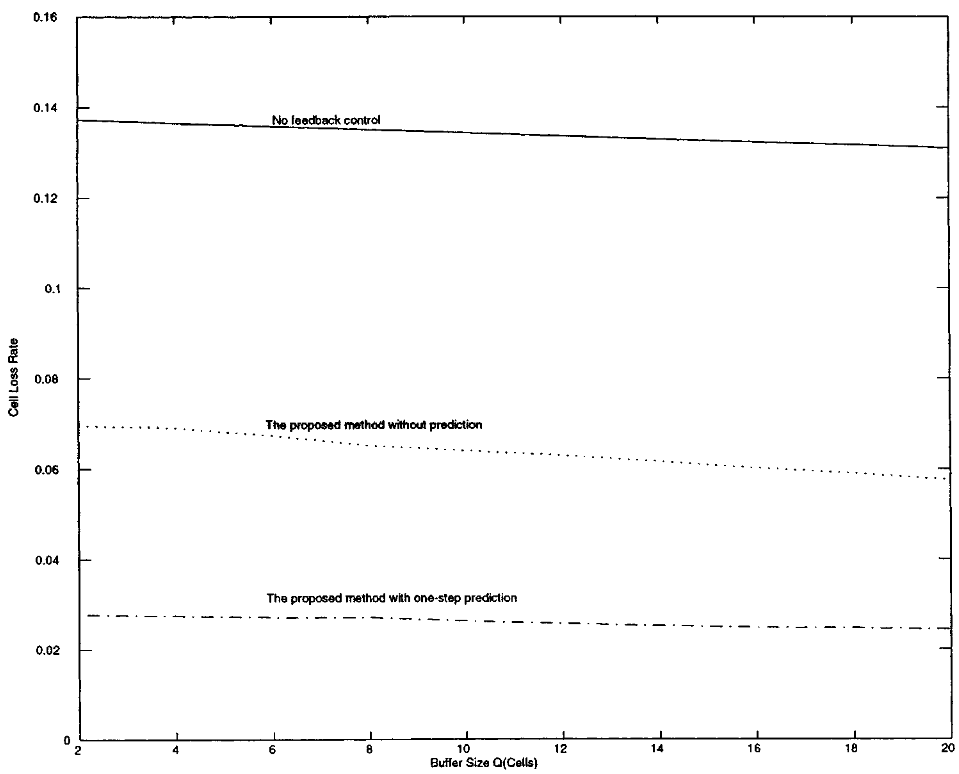

Fig. 13. Cell loss rate against buffer size in Case 1 of Example 1.

arrival rate and the service rate yield a utilization (i.e., the ratio between the mean arrival rate and the service rate) of 0.8.

Case 1: In the first case, we use a heterogeneous superposition arrival process of one video source and two voice sources to model multimedia traffic. Fig. 17 shows the original and predicted traffic data in which every sample unit is 0.25 s. The mean arrival rate is 5.08 Mb/s and service rate is 6.35 Mb/s. In this case, the proposed fuzzy-AR model is employed for traffic prediction, the SRR is 22.64 dB. After traffic prediction, the proposed control method in example 1 is used to treat the congestion problem. Fig. 18 shows the cell loss rate against the buffer size of the proposed control method. It is seen that the proposed control mechanism provides the best performance.

Case 2: In this case we use a heterogeneous superposition arrival process of two video sources and two voice sources to model multimedia traffic. Fig. 19 shows the original and pre-

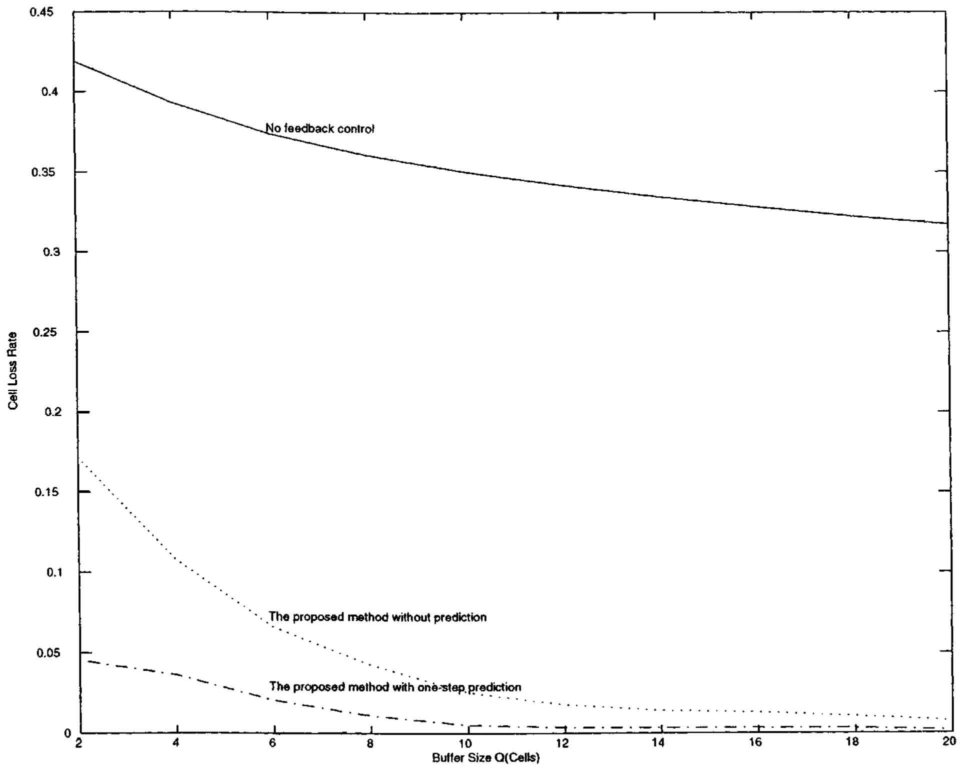

Fig. 14. Cell loss rate against buffer size in Case 2 of Example 1.

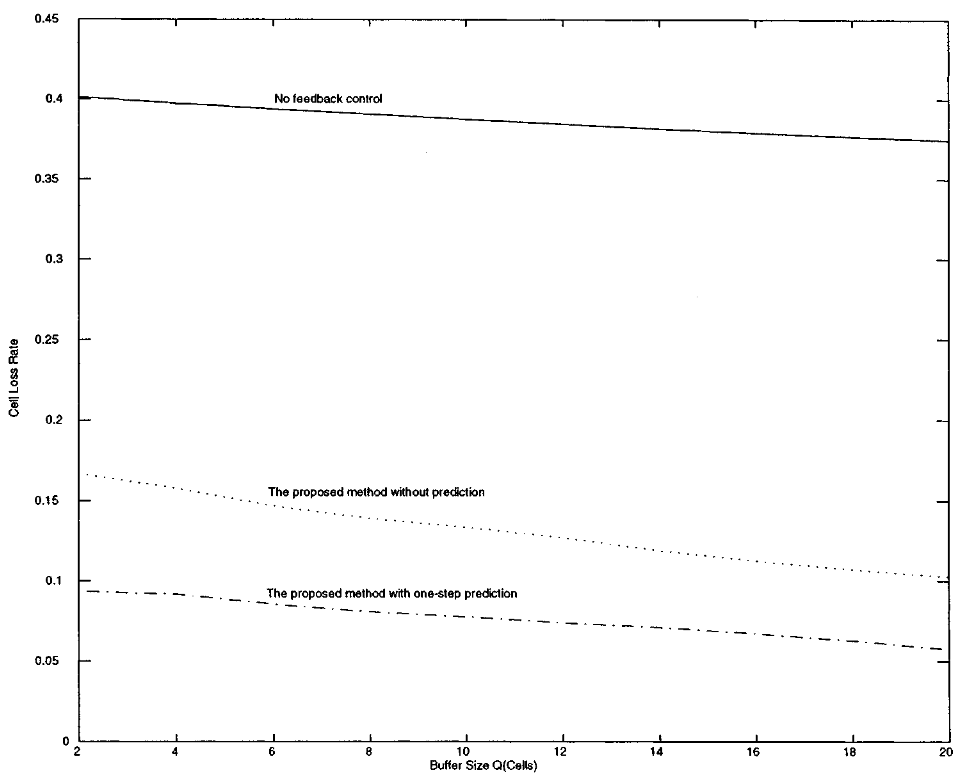

Fig. 15. Cell loss rate against buffer size in Case 3 of Example 1.

dicted traffic data and every sample unit is 0.25 s. The mean arrival rate is approximately 10.37 Mb/s. The service rate is 12.96 Mb/s. In this case, the proposed fuzzy-AR model is employed to predict the network traffic and the SRR is 27.0874 dB. After the traffic prediction, the proposed method in Example 1 is used for congestion control. Fig. 20 shows the cell loss rate versus buffer size. The results show the proposed mechanism has a superior performance for multimedia traffic.

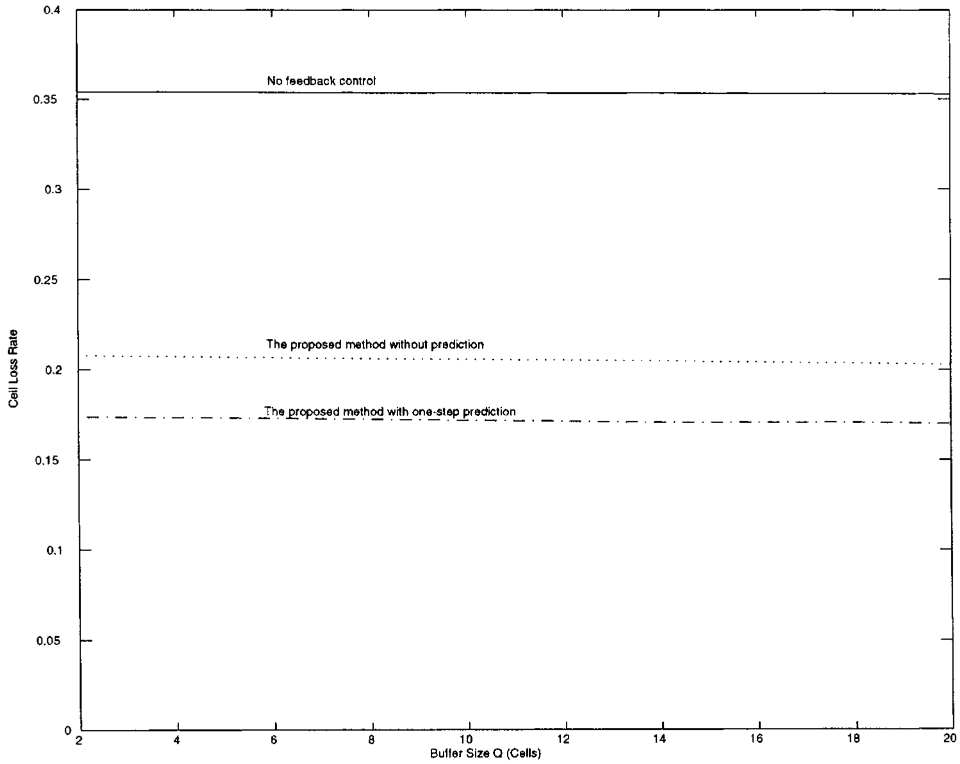

Fig. 16. Cell loss rate against buffer size in Case 4 of Example 1.

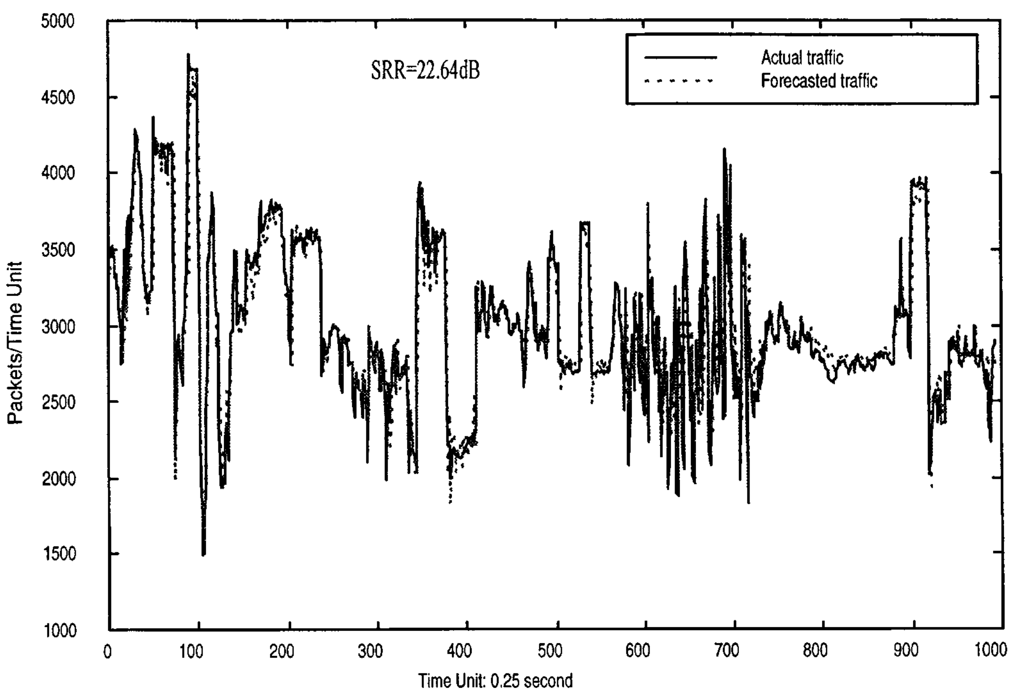

Fig. 17. Actual and forecasted packet traffic using fuzzy-AR model in Case 1 of Example 2.

VII. Conclusion

In this work, an efficient and accurate model is proposed for predicting actual traffic data in high-speed networks. The major problem in the modeling and prediction of traffic packets in high-speed networks is the complex interaction among the traffic flows. The fuzzy-AR model can circumvent this problem by using the fuzzy clustering algorithm. Based on a rough tuning algorithm and a fine tuning algorithm, a recursive parameter estimation algorithm of fuzzy-AR model is developed to achieve

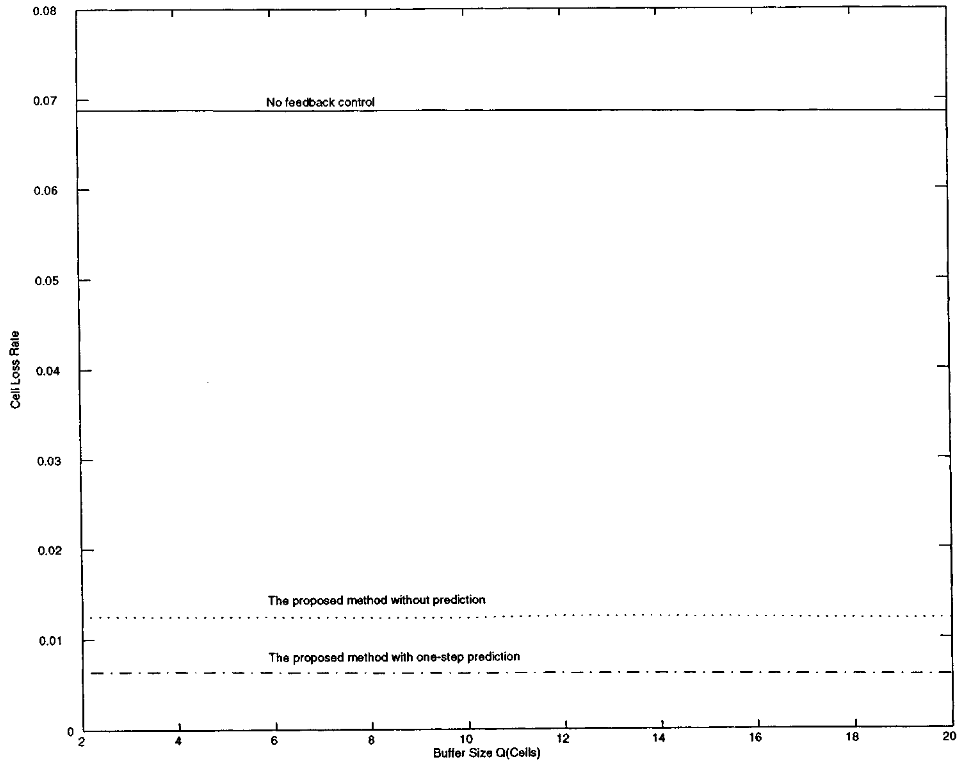

Fig. 18. Cell loss rate against buffer size in Case 1 of Example 2.

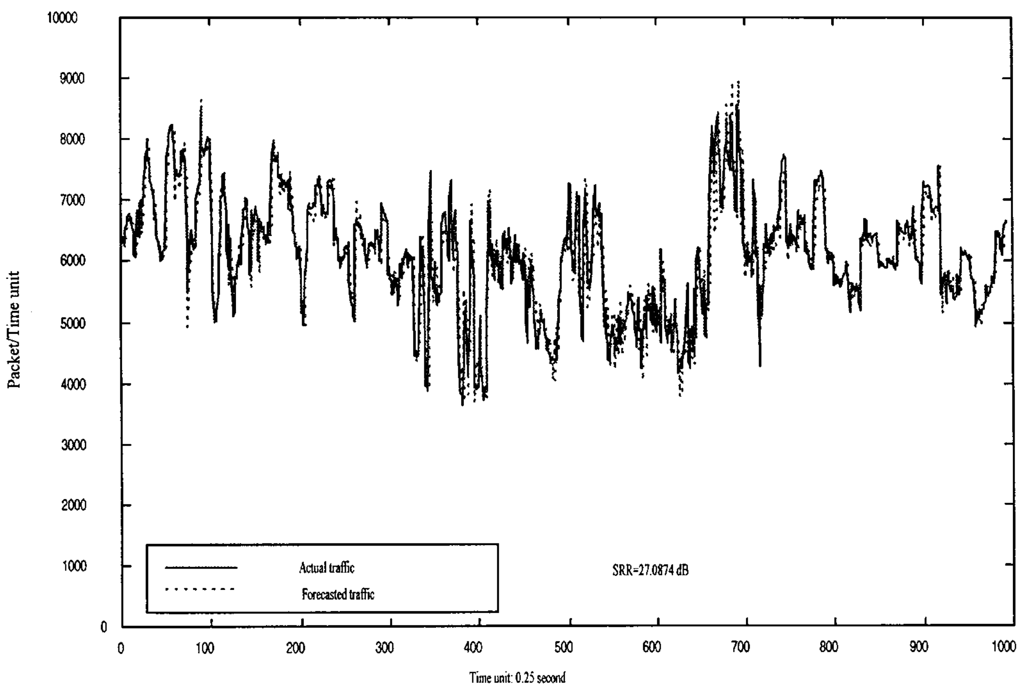

Fig. 19. Actual and forecasted packet traffic using fuzzy-AR model in Case 2 of Example 2.

precise prediction of actual traffic in high-speed networks. The proposed fuzzy-AR model has the potential for high utilization and congestion control in high-speed networks. Based on the prediction of traffic congestion by this proposed fuzzy-AR model, a simple but efficient congestion control algorithm is developed to smooth the input arrival process through decreasing the peak bit rate.

In numerical simulations, the results of this proposed method have exhibited excellent performance when used with actual traffic data. Several experimental simulations based on practical traffic data indicate that the fuzzy-AR model based on clustering algorithm can give very good prediction without knowledge of the statistical characteristics of the actual traffic data, therefore providing satisfactory traffic congestion control.

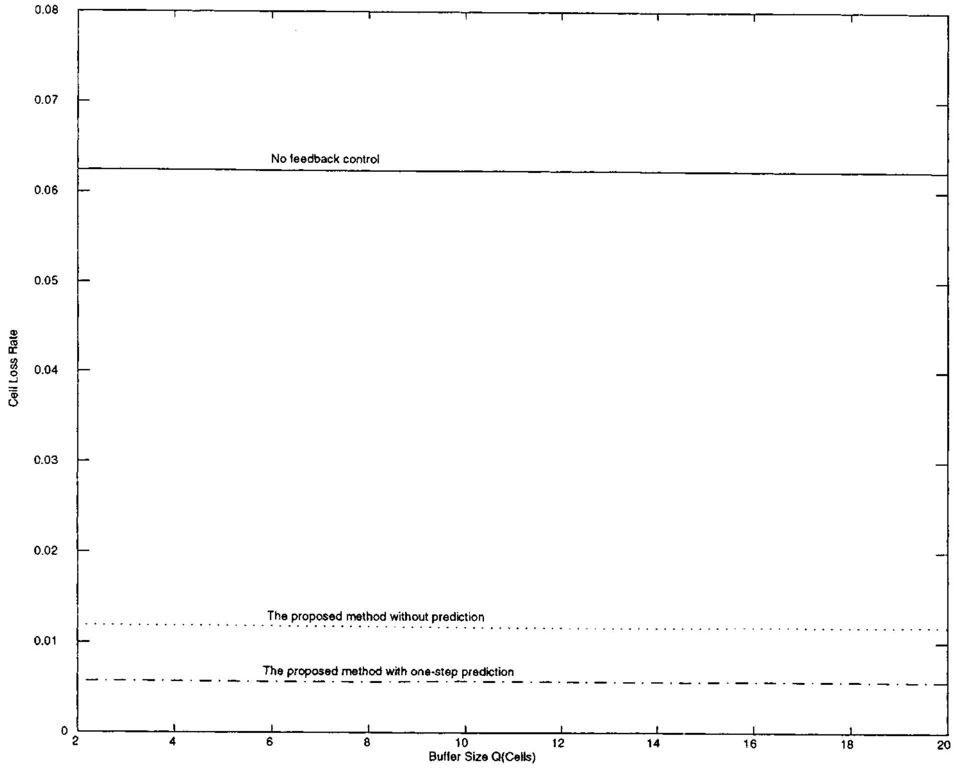

Fig. 20. Cell loss rate against buffer size in Case 2 of Example 2.

In brief, we have proposed an effective traffic model using fuzzy-AR modeling, which has excellent capability to describe the characteristics of broad-band networks and provide precise prediction for efficient congestion control. Furthermore, it is easy to calculate in real time and simple to implement in workstation computers. Therefore, this method has significant potential for practical network traffic control design. Finally, we have also verified the validity of the fuzzy-AR model using actual traffic data in Bellcore LAN network and from multimedia broad-band networks.

REFERENCES

[1] V. Frost and B. Melamed, "Traffic modeling for telecommunications networks," *IEEE Commun. Mag.*, pp. 70–81, Mar. 1994.
[2] A. Adas, "Traffic models in broadband networks," *IEEE Commun. Mag.*, pp. 82–89, July 1997.
[3] 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. Networking*, vol. 2, pp. 1–15, Feb. 1994.
[4] J. Beran, R. Sherman, M. S. Taqqu, and W. Willinger, "Long-range dependence in variable-bit-rate video traffic," *IEEE Trans. Commun.*, vol. 43, pp. 1566–1579, July 1995.
[5] H. Heffes and D. Lucantoni, "A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance," *IEEE J. Select. Areas Commun.*, vol. SAC-4, pp. 856–868, Sept. 1986.
[6] C. Shim, I. Ryoo, J. Lee, and S. Lee, "Modeling and call admission control algorithm of variable bit rate video in ATM networks," *IEEE J. Select. Areas Commun.*, vol. 12, pp. 332–344, Feb. 1994.
[7] J. Beran, *Statistics for Long Memory Processes*. London, U.K.: Chapman Hall, 1994.
[8] J. Hosking, "Fractional differencing," *Biometrica*, pp. 165–176, 1981.
[9] B. Bensaou, S. T. C. Lam, H.-W. Chu, and D. H. K. Tsang, "Estimation of the cell loss ratio in ATM networks with a fuzzy system and application to measurement-based call admission control," *IEEE/ACM Trans. Networking*, vol. 5, pp. 572–584, Aug. 1997.
[10] V. Paxson and S. Floyd, "Wide area traffic: The failure of Poisson modeling," *IEEE/ACM Trans. Networking*, vol. 3, pp. 226–244, June 1995.
[11] W. M. Lam and G. W. Wornell, "Multiscale representation and estimation of fractal point processes," *IEEE Trans. Signal Processing*, vol. 43, pp. 2606–2617, Nov. 1995.
[12] L. Ott, *An Introduction of Statistical Methods and Data Analysis*, 3rd ed. Boston, MA: PWS-Kent, 1990.
[13] D. P. Heyman and T. V. Lakshman, "Source models for VBR broadcast-video traffic," *IEEE/ACM Trans. Networking*, vol. 4, pp. 40–48, Feb. 1996.
[14] S. Dravida and V. R. Saksena, "Analysis and engineering of a voice/data packet multiplexer," *IEEE Trans. Commun.*, vol. 41, pp. 1656–1667, Nov. 1993.
[15] N. Akar and E. Arikan, "Markov modulated periodic arrival process offered to an ATM multiplexer," *Perform. Eval.*, vol. 22, pp. 175–190, 1995.
[16] V. N. Bhat, "Renewal approximations of the switched Poisson processes and their applications to queueing systems," *J. Operational Res. Soc.*, vol. 45, pp. 345–353, 1994.
[17] K. Sriram and W. Whitt, "Characterizing superposition arrival processes in packet multiplexers for voice and data," *IEEE J. Select. Areas Commun.*, vol. SAC-4, pp. 833–846, Sept. 1986.
[18] K. K. Leung, W. A. Massey, and W. Whitt, "Traffic models for wireless communication networks," *IEEE J. Select. Areas Commun.*, vol. 12, pp. 1353–1364, Oct. 1994.
[19] G. F. Williams and A. Leon-Garcia, "Performance analysis of integrated voice and data hybrid-switched links," *IEEE Trans. Commun.*, vol. COM-27, pp. 695–705, June 1984.
[20] J. Beran, "A goodness-of-fit test for time series with long range dependence," *J. Roy. Statist. Soc. B*, vol. 54, no. 3, pp. 749–760, 1992.
[21] D. Cox and V. Isham, *Point Processes*. London, U.K.: Chapman Hall, 1980.
[22] H. Fowler and W. Leland, "Local area network traffic characteristics, with implications for broadband network congestion management," *IEEE J. Select. Areas Commun.*, vol. 9, pp. 1139–1149, Sept. 1991.
[23] E. Fuchs and P. E. Jackson, "Estimates of distributions of random variables for certain computer communications traffic models," *Commun. ACM*, vol. 13, no. 12, pp. 752–757, Dec. 1970.
[24] R. Jain and S. Routhier, "Packet trains-measurements and a new model for computer network traffic," *IEEE J. Select. Areas Commun.*, vol. SAC-4, pp. 986–995, Sept. 1986.

[25] V. Paxson, "Empirically-derived analytic models of wide-area TCP connections," *IEEE/ACM Trans. Networking*, vol. 2, pp. 316–336, Aug. 1994.
[26] ——, "Growth trends in wide-area TCP connections," *IEEE Networks*, vol. 8, pp. 8–17, July/Aug. 1994.
[27] E. Kim, M. Park, S. Ji, and M. Park, "A new approach to fuzzy modeling," *IEEE Trans. Fuzzy Syst.*, vol. 5, pp. 328–337, Aug. 1997.
[28] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-15, pp. 116–132, Jan./Feb. 1985.
[29] M. Sugeno and T. Yasukawa, "A fuzzy-logic-based approach to qualitative modeling," *IEEE Trans. Fuzzy Syst.*, vol. 1, pp. 7–31, Feb. 1993.
[30] R. Hathaway and J. C. Bezdek, "Switching regression model and fuzzy clustering," *IEEE Trans. Fuzzy Syst.*, vol. 1, pp. 195–204, Aug. 1993.
[31] L. X. Wang, *A Course in Fuzzy Systems and Control*. Englewood Cliffs, NJ: Prentice-Hall, 1997.
[32] S. G. Cao and N. W. Rees, "Identification of dynamic fuzzy models," *Fuzzy Sets Syst.*, pp. 307–320, 1995.
[33] W. E. Leland and D. V. Wilson, "High time-resolution measurement and analysis of LAN traffic: Implications for LAN interconnection," in *Proc. IEEE INFOCOM'91*, Bal Harbour, FL, 1991, pp. 1360–1366.
[34] M. Schwartz, *Broadband Integrated Networks*. Englewood Cliffs, NJ: Prentice-Hall, 1996.
[35] B. Tsybakov and N. D. Georganas, "On self-similar traffic in ATM queues: Definitions, overflow probability bound, and cell delay distribution," *IEEE/ACM Trans. Networking*, vol. 5, pp. 397–409, June 1997.
[36] J. S. R. Jang, C. T. Sun, and E. Mizutani, *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*. Englewood Cliffs, NJ: Prentice-Hall, 1997.
[37] K. Meier-Hellsterm, P. E. Wirth, Y. L. Yan, and D. A. Hoeflin, "Traffic models for ISDN data users: Office automation application," in *Teletraffic and Datatraffic in a Period of Change (Proc. 13th ITC, Chpenhagen, 1991)*, A. Jensen and V. B. Iversen, Eds. Amsterdam, The Netherlands: North-Holland, 1991, pp. 167–172.
[38] D. M. Chiu and R. Jain, "Analysis of increase and decrease algorithms for congestion avoidance in computer networks," *Comput. Networks ISDN Syst.*, vol. 17, no. 1, pp. 1–14, June 1989.
[39] R. Dighe, C. J. May, and G. Ramamurthy, "Congestion avoidance strategies in broadband packet networks," in *Proc. INFOCOM'91*, Bal Harbour, FL, Apr. 1991, pp. 295–303.
[40] Z. Fan and P. Mars, "Access flow control scheme for ATM networks using neural-network-based traffic prediction," *Inst. Elect. Eng. Proc. Commun.*, vol. 144, no. 5, pp. 295–300, 1997.
[41] A. A. Tarraf, I. W. Habib, and T. N. Saadawi, "Statistical analysis of CCSN/SS7 traffic data from working subnetworks," *IEEE J. Select. Areas Commun.*, vol. 12, pp. 1088–1095, Aug. 1994.

Bor-Sen Chen (M'82–SM'89) received the B.S. degree from Tatung Institute of Technology, Taipei, Taiwan, R.O.C., in 1970, the M.S. degree from National Central University, Taiwan, in 1973, and the Ph.D. degree from the University of Southern California, Los Angeles, in 1982.

From 1973 to 1987, he was a Lecturer, Associate Professor, and Professor at Tatung Institute of Technology. He is now a Professor at National Tsing Hua University, Hsin-Chu, Taiwan. His current research interests are in control and signal processing.

Dr. Chen has received the Distinguished Research Award from National Science Council of Taiwan four times. He is a Research Fellow of the National Science Council and the Chair of the Outstanding Scholarship Foundation.

Sen-Chueh Peng (M'92) was born in Kaohsiung, Taiwan, R.O.C. He received the B.S. degree from Kaohsiung Institute of Technology, Kaohsiung, Taiwan, and the M.S. and Ph.D. degrees in electrical engineering from National Tsing-Hua University, Hsinchu, Taiwan, in 1986 and 1992, respectively.

From 1992 to 1998, he was an Associate Professor at National Yun-Lin Polytechnic Institute, Yunlin, Taiwan. He is now a Professor at National Huwei Institute of Technology, Yunlin, Taiwan. His areas of research interest are signal processing and robust control.

Ku-Chen Wang received the B.S. and M.S. degrees from Tsing Hua University, Hsinchu, Taiwan, R.O.C., in 1996 and 1998, respectively.