Generalized Cost Function Based Forecasting for Periodically Measured Nonstationary Traffic

Balaji Krithikaivasan, Yong Zeng, Deep Medhi

***Abstract*— In this paper, we address the issue of forecasting for periodically measured nonstationary traffic based on statistical time series modeling. Often with time series based applications, minimum mean square error (MMSE) based forecasting is sought that minimizes the square of the positive as well as the negative deviations of the forecast from the unknown true value. However, such a forecasting function is not directly applicable for applications such as predictive bandwidth provisioning in which the negative deviations (under-forecast) have more impact on the system performance than the positive deviation (over-forecast). For instance, an under-forecast may potentially result in insufficient allocation of bandwidth leading to short term data loss. To facilitate a differential treatment between the under and the over-forecasts, we introduce a generalized forecast cost function that is defined by allowing different penalty associated with the under and the over-forecasts. We invoke mild assumptions on the first order characteristics of such penalty functions to ensure the existence and uniqueness of the optimal forecast value in the domain of interest. The sufficient condition on the forecast distribution is that all the orders of the moments are well-defined. We provide several possible classes of penalty functions to illustrate the generic nature of the cost function and its applicability from a dynamic bandwidth provisioning perspective. A real network traffic example using several classes of penalty functions is presented to demonstrate the effectiveness of our approach.**

I. Introduction

One of the challenging issues that arises in a nonstationary network traffic environment is to ensure that the resource requirements, in particular, bandwidth, of the underlying traffic is met most of the time while avoiding excessive allocation of resources as well. However, nonstationary network traffic do exhibit cyclic pattern at various seasonal cycles, i.e., on a daily (24-hour) basis, a weekly basis, and so on. Consequently, a viable option is to assess the expected peak requirement over a 24-hour window period and allocate the bandwidth accordingly. Such an option, however, while resulting in negligible data loss, leads to excessive allocation of resources during the non-peak periods. Thus, there arises a need to develop a traffic forecasting mechanism that can predict the resource requirements ahead of time on a finer scale (say, for every fifteen minutes) while taking into account the relative importance of the under-forecast and the over-forecast. Such an effective forecast can then be translated into bandwidth requirement and be provisioned in a pro-active manner.

The first step toward facilitating traffic prediction, is to characterize the underlying traffic dynamics through an appropriate stochastic model based on the available measurements. In this context, statistical time series based models have been proposed in the literature to model and predict the short-term as well as long-term traffic behavior in Internet backbone networks; for example, see [2], [7], [9], [10], [11], [12]. The advantage of using time series models lies in its ability to adapt the prediction (derived from the model) based on a finite moving window of traffic trace history. Consequently, from the provisioning perspective, an effective bandwidth envelope closely matching the traffic dynamics can be obtained. The question arises then is that "what is an effective forecasting strategy, assuming a best-fit model is in place?"

Conventionally, a forecasting function that achieves the minimum mean-square error (MMSE) among all the possible forecast functions is sought in most applications. The reasons are two-fold: 1) the MMSE forecast minimizes the square of the deviation of predicted value in either directions from the unknown true value, 2) for a larger class of distributions, MMSE forecast equals the conditional expectation of the distribution which is trivial to compute. From the bandwidth provisioning perspective, however, under-forecast (negative deviation), which may be translated into data loss, should be treated differently from the over-forecast (positive deviation), which may decrease the utilization. In other words, a weighted deviation function becomes more desirable for the provisioning application. We introduce such a generic forecast cost function framework that can accommo-

Manuscript received December 9, 2005; revised February 26, 2007.

B. Krithikaivasan and D. Medhi, Computer Science & Electrical Engineering Department, University of Missouri-Kansas City, MO 64110 USA (e-mail: dmedhi@umkc.edu).

Y. Zeng, Department of Mathematics and Statistics, University of Missouri-Kansas City, MO 64110 USA (e-mail: zengy@umkc.edu).

date a weighted approach toward the forecast deviations. Precisely, we can potentially associate different penalty functions with the under and the over-forecasts in the forecast cost function which is then minimized to obtain the desirable optimal forecast subjected to the corresponding penalties. Under the general setup of our forecast cost function framework, we show that MMSE forecast can always be obtained as a special but less desirable case (see Section III-A.2).

Our major contribution in this work is that we introduce a generalized penalty function approach that accounts for both under and over-forecasts differently and enumerate the sufficient conditions on these penalty functions such that the existence and the uniqueness of forecast value that minimizes the forecast cost function, is guaranteed. For the derivation of this important result, we assume that the knowledge of conditional forecast distribution is available from the identification of a time series model. In order for our result to hold, the forecast distribution must belong to the class of distributions for which all the moments are well-defined (less than ∞). In general, as long as a distribution has a moment generating function, then all moments exist; this is a fairly common assumption that includes distributions such as normal, exponential, or log-normal distributions.

We illustrate the generality of our result by considering several classes of penalty functions that are applicable in a real network. In particular, using a real network data collected on the Internet link connecting the University of Missouri–Kansas City to MOREnet, we present quantitative results for the first and the second order polynomial functions and piecewise linear function. Our intent behind using the real network data is to show the effectiveness of generic cost function based forecast against the conventional MMSE forecast in a real network. Furthermore, the quantitative results bring out the impact of various parameters involved in deriving the desirable optimal forecast.

The rest of the paper is organized as follows: We present the generic forecast cost function in Section II and prove the existence of a unique minimizer of such a function under mild assumptions. In Section III, we present various penalty functions that obey the assumptions from Section II to demonstrate the generality of our result. In Section IV, we present a real network traffic example to illustrate the applicability of our result to real networks. Finally, we present a summary of the work.

II. Generalized Forecast Cost Function

Suppose, the traffic on a network link is measured periodically (say, every fifteen minutes). A total of $t-1$ such measurements yield a equally spaced time series $z_1, z_2, \ldots, z_{t-1}$ where the subscripts represent the time indices. Assume, the best-fit model is identified for the time series through time series modeling methodologies. Let $\mathcal{F}_{t-1}$ be the σ-algebra generated by $z_1, z_2, \ldots, z_{t-1}$ and $F_{t-1}(\cdot)$ represent the one-step ahead conditional distribution, given $\mathcal{F}_{t-1}$ with the conditional density function $f_{t-1}(\cdot)$. It may be noted that $F_{t-1}(\cdot)$ is known from the best-fit time series model and the observations up to time $t-1$. Let z_t be the traffic at time t and given $\mathcal{F}_{t-1}$, z_t follows $F_{t-1}(\cdot)$. Let $E_{t-1}[z_t]$ be the conditional expectation with respect to z_t given $\mathcal{F}_{t-1}$ and $\tilde{z}_t$ be the forecast at time t. Further, let $U(\tilde{z}_t, z_t)$ and $L(\tilde{z}_t, z_t)$ refer to the generic penalty functions for over and under-forecasts, respectively. Then, we define the total cost of obtaining forecast $\tilde{z}_t$ as follows:

$$\mathcal{C}_t(\tilde{z}_t) = E_{t-1}[\ U(\tilde{z}_t, z_t)\ I_{\{z_t<\tilde{z}_t\}} + L(\tilde{z}_t, z_t)\ I_{\{z_t>\tilde{z}_t\}}\ ], \quad (1)$$

where the indicator function is defined as

$$I_{\{a>b\}} = \begin{cases} 1 & \text{if } a > b \\ 0 & \text{otherwise.} \end{cases}$$

Such a cost function (1) can be used for forecasting for time-series based on models such as AutoRegressive Integrated Moving Average (ARIMA) [5] and AutoRegressive Conditional Heteroskedasticity (ARCH)-type models [6], [8] with i.i.d normal innovations. This aspect will be discussed later in Section IV-A.

For a given forecast $\tilde{z}_t$, the first term in (1) characterizes the penalty due to over-forecast while the second term characterizes the penalty due to under-forecast. Furthermore, (1) can be written based on the linearity of expectation as below:

$$\begin{aligned} \mathcal{C}_t(\tilde{z}_t) &= E_{t-1}[\ U(\tilde{z}_t, z_t)\ I_{\{z_t<\tilde{z}_t\}}\ ] \\ &\quad + E_{t-1}[\ L(\tilde{z}_t, z_t)\ I_{\{z_t>\tilde{z}_t\}}\ ] \\ &= \int_0^{\tilde{z}_t} U(\tilde{z}_t, z_t)\ f_{t-1}(z_t)\ dz_t \\ &\quad + \int_{\tilde{z}_t}^{\infty} L(\tilde{z}_t, z_t)\ f_{t-1}(z_t)\ dz_t. \end{aligned} \quad (2)$$

For notational convenience, we set

$$J_U(a, b; \tilde{z}_t) = \int_a^b U(\tilde{z}_t, z_t)\ f_{t-1}(z_t)\ dz_t$$

$$J_L(a, b; \tilde{z}_t) = \int_a^b L(\tilde{z}_t, z_t)\ f_{t-1}(z_t)\ dz_t$$

$$I_U(a, b; \tilde{z}_t) = \int_a^b \frac{\partial U(\tilde{z}_t, z_t)}{\partial \tilde{z}_t}\ f_{t-1}(z_t)\ dz_t$$

and

$$I_L(a,b;\tilde{z}_t) = \int_a^b \frac{\partial L(\tilde{z}_t, z_t)}{\partial \tilde{z}_t}\ f_{t-1}(z_t)\ dz_t.$$

Thus, (2) can be rewritten as

$$\mathcal{C}_t(\tilde{z}_t) = J_U(0,\tilde{z}_t;\tilde{z}_t) + J_L(\tilde{z}_t,\infty;\tilde{z}_t). \tag{3}$$

We are interested in determining z_t^* such that z_t^* achieves the minimum cost among all positive $\tilde{z}_t$'s, i.e., $\mathcal{C}_t(z_t^*) \leq \mathcal{C}_t(\tilde{z}_t)$ for $\tilde{z}_t \in [0,\infty]$. It may be noted that since the measurements (of traffic trace) are always positive, we restrict the interval of definition for the conditional distribution $F_{t-1}(\cdot)$ of forecast to $[0,\infty]$.

Below, we enumerate a set of assumptions on the penalty functions $U(\tilde{z}_t, z_t)$ and $L(\tilde{z}_t, z_t)$ in order to ensure that a unique minimizer z_t^* exists for the forecast cost function given in (2).

Assumption 1:

(i) For a given nonnegative $\tilde{z}_t$, $U(\tilde{z}_t, z_t)$ is a non-increasing continuous function of z_t for $z_t \in [0,\tilde{z}_t]$, and $U(m,m) = 0$ for $m \in [0,\infty]$ (see Remark 2).

(ii) The partial derivative of $U(\tilde{z}_t, z_t)$ with respect to $\tilde{z}_t$ exists almost everywhere and it is positive and has at most a countable number of discontinuities for z_t in the interval $[0,\tilde{z}_t]$ for $\tilde{z}_t \in [0,\infty]$, i.e., if $\mathcal{P}_U$ is the set containing points of discontinuity then,

$$\frac{\partial U(\tilde{z}_t, z_t)}{\partial \tilde{z}_t} > 0 \text{ for } z_t \in [0,\tilde{z}_t)\backslash\mathcal{P}_U.$$

(iii) For $c_1, c_2 \in [0,\infty]$ with $c_1 < c_2$,

$$\frac{\partial U(\tilde{z}_t, a)}{\partial \tilde{z}_t}\Big|_{\tilde{z}_t=c_1} \leq \frac{\partial U(\tilde{z}_t, a)}{\partial \tilde{z}_t}\Big|_{\tilde{z}_t=c_2},$$

where $0 < a < c_1$ and the partial derivative exists at $z_t = a$ in the chosen interval, i.e., $\dfrac{\partial U(\tilde{z}_t, a)}{\partial \tilde{z}_t}$ is non-decreasing in $\tilde{z}_t$.

Remark 1: In simple terms, the above means that $U(\tilde{z}_t, z_t)$, (i) is nonincreasing and continuous, (2) has the partial derivative that exists almost everywhere, positive and is at most countable discontinuities, and (iii) is convex.

Remark 2: Condition (i) of Assumption 1 implies that the closer the under-realization z_t to the given forecast $\tilde{z}_t$, the lesser the penalty due to over-forecast and the penalty is *zero* if realization of z_t equals the forecast $\tilde{z}_t$ (an ideal case).

Remark 3: Condition (i) of Assumption 1 further implies that $U(\tilde{z}_t, z_t)$ is bounded from above for $z_t \in [0,\tilde{z}_t)$ which implies $U(\tilde{z}_t, z_t)$ is *Riemann integrable*. Furthermore, $f_{t-1}(z_t)$ being a probability density function, is Riemann integrable as well. Thus, $J_U(0,\tilde{z}_t;\tilde{z}_t)$ is *Riemann integrable*, i.e., it is well-defined.

Remark 4: Since $U(\tilde{z}_t, z_t)$ is bounded and non-increasing continuous for $z_t \in [0,\tilde{z}_t)$ it follows that the partial derivative of $U(\tilde{z}_t, z_t)$, if it exists, is bounded almost everywhere for $z_t \in [0,\tilde{z}_t)$. Then, following condition (ii) of Assumption 1, $I_U(0,\tilde{z}_t;\tilde{z}_t)$ is *Riemann integrable*.

Assumption 2:

(i) For a given nonnegative $\tilde{z}_t$, $L(\tilde{z}_t, z_t)$ is a non-decreasing continuous function of z_t for $z_t \in [\tilde{z}_t,\infty]$, and $L(m,m) = 0$ for $m \in [0,\infty]$ (see Remark 6).

(ii) The partial derivative of $L(\tilde{z}_t, z_t)$ with respect to $\tilde{z}_t$ exists almost everywhere and it is negative and has at most a countable number of discontinuities for z_t in the interval $(\tilde{z}_t,\infty]$ for $\tilde{z}_t \in [0,\infty]$, i.e., if $\mathcal{P}_L$ is the set containing points of discontinuity then,

$$\frac{\partial L(\tilde{z}_t, z_t)}{\partial \tilde{z}_t} < 0 \text{ for } z_t \in (\tilde{z}_t,\infty]\backslash\mathcal{P}_L.$$

(iii) For $c_1, c_2 \in [0,\infty]$ with $c_1 < c_2$,

$$\frac{\partial L(\tilde{z}_t, b)}{\partial \tilde{z}_t}\Big|_{\tilde{z}_t=c_1} \leq \frac{\partial L(\tilde{z}_t, b)}{\partial \tilde{z}_t}\Big|_{\tilde{z}_t=c_2},$$

where $c_2 < b < \infty$ and the partial derivative exists at $z_t = b$ in the chosen interval, i.e., $\dfrac{\partial L(\tilde{z}_t, b)}{\partial \tilde{z}_t}$ is non-decreasing in $\tilde{z}_t$.

(iv) For a given nonnegative $\tilde{z}_t$, $J_L(\tilde{z}_t,\infty;\tilde{z}_t)$ is Riemann integrable, i.e., it is well-defined (see Remark 7).

(v) For a given nonnegative $\tilde{z}_t$, $I_L(\tilde{z}_t,\infty;\tilde{z}_t)$ is Riemann integrable (see Remark 8).

Remark 5: In simple terms, the above means that $L(\tilde{z}_t, z_t)$ (i) is nonincreasing and continuous, (2) has the partial derivative that exists almost everywhere, is negative and is at most countable discontinuities, and (iii) is convex. The other two conditions, (iv) and (v), are technical integrability assumptions.

Remark 6: Condition (i) of Assumption 2 implies that the farther the over-realization z_t to the given forecast $\tilde{z}_t$, the higher the penalty due to under-forecast and the penalty is *zero* if realization of z_t equals the forecast $\tilde{z}_t$ (an ideal case).

Remark 7: Condition (i) of Assumption 2 does not imply that $L(\tilde{z}_t, z_t)$ is bounded from above for $z_t \in (\tilde{z}_t,\infty]$. Thus, condition (iv) is necessary for $\mathcal{C}_t(\tilde{z}_t)$ to be well-defined.

Remark 8: Since the partial derivative of $L(\tilde{z}_t, z_t)$ may not be bounded from above for $z_t \in (\tilde{z}_t,\infty]$, condition (v) is necessary for the derivative of $\mathcal{C}_t(\tilde{z}_t)$ with respect to $\tilde{z}_t$ to exist.

It may be noted that the sets $\mathcal{P}_U$ and $\mathcal{P}_L$ may be empty. In other words, $U(\tilde{z}_t, z_t)$ and $L(\tilde{z}_t, z_t)$ may be defined in such a way that their partial derivatives with respect to $\tilde{z}_t$ is continuous for z_t in the respective intervals.

Lemma 1: Let $U(\tilde{z}_t, z_t)$ and $L(\tilde{z}_t, z_t)$ be defined in such a way that both Assumptions 1 and 2 hold, then the first derivative of cost function $\mathcal{C}_t(\tilde{z}_t)$ is a strictly increasing continuous function of $\tilde{z}_t$ in the interval $[0, \infty]$.

Proof: From (2), using Leibniz rule, we have

$$\begin{aligned}\frac{d\mathcal{C}_t(\tilde{z}_t)}{d\tilde{z}_t} &= [U(\tilde{z}_t, \tilde{z}_t) - L(\tilde{z}_t, \tilde{z}_t)] f_{t-1}(\tilde{z}_t) \\ &\quad + \int_0^{\tilde{z}_t} \frac{\partial U(\tilde{z}_t, z_t)}{\partial \tilde{z}_t}\ f_{t-1}(z_t)\ dz_t \\ &\quad + \int_{\tilde{z}_t}^{\infty} \frac{\partial L(\tilde{z}_t, z_t)}{\partial \tilde{z}_t}\ f_{t-1}(z_t)\ dz_t. \end{aligned} \tag{4}$$

Following condition (i) of Assumptions 1 and 2, the first term in (4) vanishes. Then, (4) is simplified as

$$\frac{d\mathcal{C}_t(\tilde{z}_t)}{d\tilde{z}_t} = I_U(0, \tilde{z}_t; \tilde{z}_t) + I_L(\tilde{z}_t, \infty; \tilde{z}_t). \tag{5}$$

Following Remark 3 and condition (v) of Assumption 2, both $I_U(0, \tilde{z}_t; \tilde{z}_t)$ and $I_L(\tilde{z}_t, \infty; \tilde{z}_t)$ are Riemann integrable. Therefore, it is clearly evident that the first derivative of $\mathcal{C}(\tilde{z}_t)$ is a continuous function of $\tilde{z}_t$.

Let $c_1, c_2 \in [0, \infty]$ with $c_1 < c_2$ and define

$$D = \frac{d\mathcal{C}_t(\tilde{z}_t)}{d\tilde{z}_t}\Big|_{\tilde{z}_t = c_2} - \frac{d\mathcal{C}_t(\tilde{z}_t)}{d\tilde{z}_t}\Big|_{\tilde{z}_t = c_1}.$$

Then, it follows from (5) that

$$\begin{aligned} D &= I_U(0, c_2; c_2) + I_L(c_2, \infty; c_2) - I_U(0, c_1; c_1) \\ &\quad - I_L(c_1, \infty; c_1) \\ &= I_U(0, c_1; c_2) + I_U(c_1, c_2; c_2) + I_L(c_2, \infty; c_2) \\ &\quad - I_U(0, c_1; c_1) - I_L(c_1, c_2; c_1) - I_L(c_2, \infty; c_1) \\ &= \{I_U(0, c_1; c_2) - I_U(0, c_1; c_1)\} \\ &\quad + \{I_U(c_1, c_2; c_2) - I_L(c_1, c_2; c_1)\} \\ &\quad + \{I_L(c_2, \infty; c_2) - I_L(c_2, \infty; c_1)\}. \end{aligned} \tag{6}$$

It may be noted that both $I_U(a, b; m)$ and $I_L(a, b; m)$ where $\tilde{z}_t = m$ for some $m > 0$ are evaluated as follows:

$$I_U(a, b; m) = \int_a^b \frac{\partial U(\tilde{z}_t, z_t)}{\partial \tilde{z}_t}\Big|_{\tilde{z}_t = m}\ f_{t-1}(z_t)\ dz_t$$

and

$$I_L(a, b; m) = \int_a^b \frac{\partial L(\tilde{z}_t, z_t)}{\partial \tilde{z}_t}\Big|_{\tilde{z}_t = m}\ f_{t-1}(z_t)\ dz_t.$$

From condition (iii) of Assumption 1, we have

$$I_U(0, c_1; c_2) - I_U(0, c_1; c_1) \geq 0 \tag{7}$$

and from condition (iii) of Assumption 2, we have

$$I_L(c_2, \infty; c_2) - I_L(c_2, \infty; c_1) \geq 0. \tag{8}$$

Then, we also infer from condition (ii) of Assumption 1 that

$$I_U(c_1, c_2; c_2) > 0$$

and from condition (ii) of Assumption 2 that

$$I_L(c_1, c_2; c_1) < 0$$

implying

$$I_U(c_1, c_2; c_2) - I_L(c_1, c_2; c_1) > 0. \tag{9}$$

Following (7), (8) and (9), we have $D > 0$. Since, c_1 and c_2 are chosen arbitrarily, the lemma holds true. ■

Theorem 1: Under the Assumptions 1 and 2, there exists a unique z_t^* such that $\tilde{z}_t = z_t^*$ minimizes the cost function $\mathcal{C}(\tilde{z}_t)$ among all $\tilde{z}_t \in [0, \infty]$.

Proof: Since the first derivative of the cost function $\mathcal{C}(\tilde{z}_t)$ is a strictly increasing continuous function in the interval $[0, \infty]$ (from Lemma 1), it should attain its minimum at $\tilde{z}_t = 0$ and its maximum at $\tilde{z}_t = \infty$ in the interval $[0, \infty]$. Below, we investigate these boundary values in order to prove the desired result.

Recall from (5) that

$$\frac{d\mathcal{C}(\tilde{z}_t)}{d\tilde{z}_t} = I_U(0, \tilde{z}_t; \tilde{z}_t) + I_L(\tilde{z}_t, \infty; \tilde{z}_t).$$

Then, we have

$$\frac{d\mathcal{C}(\tilde{z}_t)}{d\tilde{z}_t}\Big|_{\tilde{z}_t = 0} = I_U(0, 0; 0) + I_L(0, \infty; 0). \tag{10}$$

The first term in (10) vanishes, while the second term

$$I_L(0, \infty; 0) < 0 \tag{11}$$

following condition (ii) of Assumption 2 implying that the *minimum* value of the first derivative of cost function is *negative*.

Similarly,

$$\frac{d\mathcal{C}(\tilde{z}_t)}{d\tilde{z}_t}\Big|_{\tilde{z}_t = \infty} = \lim_{\tilde{z}_t \to \infty} I_U(0, \tilde{z}_t; \tilde{z}_t) + \lim_{\tilde{z}_t \to \infty} I_L(\tilde{z}_t, \infty; \tilde{z}_t). \tag{12}$$

The second term in (12) vanishes, whereas the first term

$$\lim_{\tilde{z}_t \to \infty} I_U(0, \tilde{z}_t; \tilde{z}_t) > 0 \tag{13}$$

as a result of condition (ii) of Assumption 1 implying thereby that the *maximum* value of the first derivative of cost function is *positive*.

Then, from Weierstrass intermediate value theorem [13], there exists at least one $z_t^* \in (0, \infty)$ such that the derivative vanishes. However, from Lemma 1, it follows that z_t^* should be unique thereby, guaranteeing a unique minimizer for the cost function $\mathcal{C}(\tilde{z}_t)$. ■

III. Penalty Function - Examples

In this section, we exemplify the generality of class of penalty functions from section II. In particular, we consider penalty functions in the form of polynomial functions and piecewise linear functions. For these examples, we also show that all the assumptions stated above are satisfied.

A. Polynomial functions

Consider the following:

$$U(\tilde{z}_t, z_t) = \kappa_1(\tilde{z}_t - z_t)^n$$
$$L(\tilde{z}_t, z_t) = \kappa_2(z_t - \tilde{z}_t)^n,$$

where constants $\kappa_1, \kappa_2 > 0$ and n is a positive integer ($n \geq 1$). Note that these functions form a natural generalization of conventional MMSE forecast function; for example, if we set $n = 2$ and $\kappa_1 = \kappa_2 = 1$, then it reduces to the MMSE forecast function. Parameters κ_1 and κ_2 allow a user to assign different weights to under and over forecast.

Clearly, the above polynomial function satisfies condition (i) of Assumption 1 and also condition (i) of Assumption 2. Below, we show that the rest of conditions are satisfied as well.

(a)
$$\frac{\partial U(\tilde{z}_t, z_t)}{\partial \tilde{z}_t} = \kappa_1\ n\ (\tilde{z}_t - z_t)^{n-1},$$
which is greater than zero for $z_t \in [0, \tilde{z}_t)$ and equals to zero if $z_t = \tilde{z}_t$. It is evident that $\mathcal{P}_U$ is a null set (no discontinuities). Similarly,
$$\frac{\partial L(\tilde{z}_t, z_t)}{\partial \tilde{z}_t} = -\kappa_2\ n\ (z_t - \tilde{z}_t)^{n-1},$$
which is less than zero for $z_t \in (\tilde{z}_t, \infty]$ and equals to zero if $z_t = \tilde{z}_t$. Again, $\mathcal{P}_L$ is also a null set (no discontinuities). Thus, condition (ii) of Assumptions 1 and 2 are satisfied.

(b) For $c_1, c_2 \in [0, \infty]$ with $c_1 < c_2$ and $n > 1$, we have
$$\frac{\partial U(\tilde{z}_t, a)}{\partial \tilde{z}_t}\Big|_{\tilde{z}_t = c_1} < \frac{\partial U(\tilde{z}_t, a)}{\partial \tilde{z}_t}\Big|_{\tilde{z}_t = c_2},$$
where $0 < a < c_1$. Similarly, we have
$$\frac{\partial L(\tilde{z}_t, b)}{\partial \tilde{z}_t}\Big|_{\tilde{z}_t = c_1} < \frac{\partial L(\tilde{z}_t, b)}{\partial \tilde{z}_t}\Big|_{\tilde{z}_t = c_2},$$
where $c_2 < b < \infty$. For $n = 1$, we have an equality relationship in both the cases. Thus, condition (iii) of Assumptions 1 and 2 are satisfied.

(c) For the given $L(\tilde{z}_t, z_t)$, $J_L(\tilde{z}_t, \infty; \tilde{z}_t)$ can be expressed in terms of the first n moments of the forecast distribution. Since, all the moments of the forecast distribution are assumed to be well-defined and n is finite, condition (iv) of Assumption 2 is satisfied. Furthermore, $I_L(\tilde{z}_t, \infty; \tilde{z}_t)$ can also be expressed in terms of the first $n-1$ moments of the forecast distribution following (a). Thus, condition (v) of Assumption 2 is also satisfied.

Thus, if the penalty functions $U(\tilde{z}_t, z_t)$ and $L(\tilde{z}_t, z_t)$ are polynomial functions of the form given above, following Lemma 1 and Theorem 1, it can be established that the forecast cost function given in (2) has a unique minimum z_t^*. Observe that when $n \geq 1$ for any real number, parts of (a), (b) and (c) in the beginning of this section remain true. Thus, Theorem 1 implies the existence of the unique minimizer. When $n \in (0, 1)$, for example, $n = 1/2$, then the Assumptions 1(iii) and 2(iii) fail and Theorem 1 does not apply.

Below, we briefly illustrate how to compute z_t^* for the first order and the second order polynomials ($n = 1$ and $n = 2$) based on the results from [11]. The derivations of these results are also presented in detail in [11].

1) Computation of z_t^* for case $n = 1$: For $n = 1$, after replacing the factors κ_1 and κ_2 using a ratio r, i.e., $r = \dfrac{\kappa_1}{\kappa_2}$ with $r < 1$ (i.e., more penalty for under-forecasting), we obtain the penalty functions as follows:

$$U(\tilde{z}_t, z_t) = r(\tilde{z}_t - z_t)$$
$$L(\tilde{z}_t, z_t) = (z_t - \tilde{z}_t).$$

Upon substituting these penalty functions in (1) and minimizing the resulting cost function, we obtain the following result:

$$F_{t-1}(\tilde{z}_t) - \frac{1}{1+r} = 0, \tag{14}$$

where $F_{t-1}(\tilde{z}_t) = P[z_t \leq \tilde{z}_t \,|\, \mathcal{F}_{t-1}]$. The required optimal forecast z_t^* is the value of $\tilde{z}_t$ that satisfies (14). In other words, z_t^* is the $(\frac{1}{1+r})^{\text{th}}$ quantile of F_{t-1}. When $r = 1$, z_t^* is the median of F_{t-1}. In fact, for $r > 1$, we still have z_t^* satisfying equation of the form (14) by replacing r with $\tilde{r} = \frac{1}{r}$ in (14).

Since, $F_{t-1}(\tilde{z}_t)$ is a *strictly increasing* function of $\tilde{z}_t$ for any given distribution, (14) can be solved numerically using a binary search algorithm. For $r < 1$ (the desirable range), the lower and upper limits for the search algorithm can be taken as the value of z_t corresponding to the median and the maximum capacity of the provisioning system respectively. The convergence rate of the search algorithm depends on the conditional distribution obtained from the fitted time series model.

2) Computation of z_t^* for case $n = 2$: Similar to the case $n = 1$, we obtain the penalty functions for $n = 2$ as follows:

$$U(\tilde{z}_t, z_t) = r(\tilde{z}_t - z_t)^2$$
$$L(\tilde{z}_t, z_t) = (z_t - \tilde{z}_t)^2.$$

Upon substitution in (1) followed by minimization, we obtain the optimal forecast z_t^* to be the value of $\tilde{z}_t$ that satisfies

$$\tilde{z}_t - G(\tilde{z}_t) = 0, \tag{15}$$

(a fixed-point equation) where

$$G(\tilde{z}_t) = \frac{E_{t-1}[z_t] - (1-r)\ Q_{t-1}(\tilde{z}_t)}{1 - (1-r)\ F_{t-1}(\tilde{z}_t)} \tag{16}$$

with $0 < r < 1$ and

$$Q_{t-1}(\tilde{z}_t) = \int_0^{\tilde{z}_t} z_t\ f_{t-1}(z_t)\ dz_t.$$

For $r > 1$, we obtain

$$G(\tilde{z}_t) = \frac{\tilde{r}\ E_{t-1}[z_t] + (1-\tilde{r})\ Q_{t-1}(\tilde{z}_t)}{\tilde{r} + (1-\tilde{r})\ F_{t-1}(\tilde{z}_t)}. \tag{17}$$

The uniqueness of z_t^* shown above for any integral values of n guarantee the uniqueness of the fixed-point satisfying (15). Note here that for $r(\tilde{r}) = 1$, (17) becomes $G(\tilde{z}_t) = E_{t-1}[z_t]$ and the optimal forecast z_t^* is the conditional mean of the distribution $F_{t-1}(\cdot)$ identified from the time series model. Thus, the conditional mean $E_{t-1}[z_t]$ can be taken as the initial value of $\tilde{z}_t$ for the fixed-point iteration algorithm for any given $r < 1$ or $\tilde{r} < 1$. The following steps briefly illustrates the fixed-point iteration.

(i) Assume the initial value of $\tilde{z}_t = E_{t-1}[z_t]$.
(ii) Compute $G(\tilde{z}_t)$ using (16).
(iii) If the current value of $\tilde{z}_t$ equals $G(\tilde{z}_t)$ obtained from step (ii), stop the iteration and $z_t^* = \tilde{z}_t$. Otherwise, proceed to the next step.
(iv) Assign $\tilde{z}_t = G(\tilde{z}_t)$ and go to step (ii).

Since (15) has a unique fixed-point, the convergence of the iterative algorithm is guaranteed. Furthermore, it may be noted that the conditional mean of the distribution is also the MMSE forecast for any linear time series system. Thus, it is evident that we obtain MMSE forecast as a special case of second order polynomial cost function with equal costs for both under and over-forecasts. In other words, MMSE forecast matches with a particular instance of cost function less useful in reality. For further details, see [11].

3) Remarks on non-integral values of n: A more general case could be considering n to be rational. In particular, n can be chosen from the subset $\{\frac{1}{2}, 1, \frac{3}{2}, 2, \ldots.\}$ of set of rational numbers. It can be shown that if $n = \frac{1}{2}$ (i.e., square root function), not all of the conditions will be satisfied and hence z_t^* is not well defined in that case. However, these conditions are satisfied for polynomial functions for $n \geq 1$ in that subset. In fact, for all rational $n \geq 1$, it can be shown that a unique minimizer z_t^* exists. Furthermore, the result can be generalized to any real $n \geq 1$.

B. Piecewise Linear Functions

We now consider penalty functions in the form of piecewise linear functions. These functions form a special case of first order polynomial functions. Essentially, these functions enable us to impose penalty in a non-uniform manner depending upon the interval on the number axis in which the magnitude of the deviation error falls into. For example, in a real network, it may be desirable to consider a smaller penalty factor for the deviation error up to 20% of unknown z_t while, a relatively greater penalty factor if deviation error exceeds 20%. In fact, such a notion might be further extended considering more intervals. Here, we consider representations for $U(\tilde{z}_t, z_t)$ and $L(\tilde{z}_t, z_t)$ with n sub-intervals, i.e., the intervals $[0, \tilde{z}_t]$ and $[\tilde{z}_t, \infty]$ are divided into n sub-intervals in which the penalty functions $U(\tilde{z}_t, z_t)$ and $L(\tilde{z}_t, z_t)$ assume different linear forms, respectively.

Fig. 1 shows an example of $U(\tilde{z}_t, z_t)$ with 4 sub-intervals while Fig. 2, shows an example for $L(\tilde{z}_t, z_t)$ with 4 sub-intervals.

Now, we present the general representation as follows:

$$U(\tilde{z}_t, z_t) = \begin{cases} b_1\tilde{z}_t - m_1 z_t & a_0\tilde{z}_t \leq z_t \leq a_1\tilde{z}_t \\ b_2\tilde{z}_t - m_2 z_t & a_1\tilde{z}_t \leq z_t \leq a_2\tilde{z}_t \\ \vdots & \\ b_n\tilde{z}_t - m_n z_t & a_{n-1}\tilde{z}_t \leq z_t \leq a_n\tilde{z}_t \end{cases} \tag{18}$$

and

$$L(\tilde{z}_t, z_t) = \begin{cases} m_1' z_t - b_1'\tilde{z}_t & a_0'\tilde{z}_t \leq z_t \leq a_1'\tilde{z}_t \\ m_2' z_t - b_2'\tilde{z}_t & a_1'\tilde{z}_t \leq z_t \leq a_2'\tilde{z}_t \\ \vdots & \\ m_n' z_t - b_n'\tilde{z}_t & a_{n-1}'\tilde{z}_t \leq z_t \leq a_n'\tilde{z}_t, \end{cases} \tag{19}$$

where $a_0 = 0 < a_1 < a_2 < \ldots < a_n = 1$ and $a_0' = 1 < a_1' < a_2' < \ldots < a_n' < \infty$. In order to ensure zero penalty when z_t equals $\tilde{z}_t$ (following condition (i) of Assumptions 1 and 2), we consider $b_n = m_n$ and $b_1' = m_1'$. Choose $m_1 > m_2 > \ldots > m_{n-1} > m_n$ (to be consistent with the notion of imposing higher penalty for

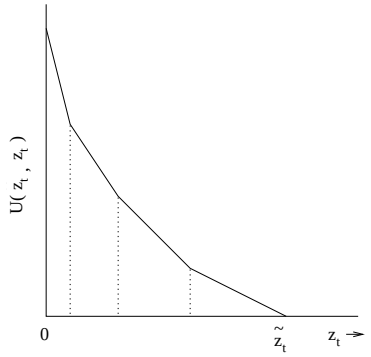

Fig. 1. Utilization Penalty Function - An Example

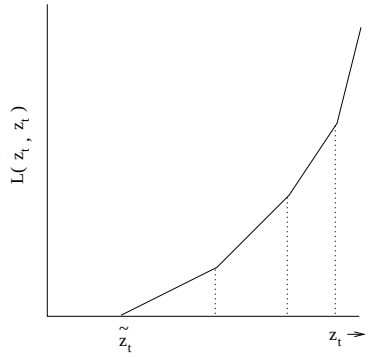

Fig. 2. Loss Penalty Function - An Example

higher deviation) and then, the continuity of $U(\tilde{z}_t, z_t)$ is ensured by computing b_n's as follows:

$$b_k = b_{k+1} + a_k(m_k - m_{k+1}) \quad k = 1, 2, \ldots, n-1. \tag{20}$$

Similarly, choose $m'_1 < m'_2 < m'_3 < \ldots < m'_n$ and then, the continuity of $L(\tilde{z}_t, z_t)$ is guaranteed by choosing b'_n's as follows:

$$b'_{k+1} = b'_k + a'_k(m'_{k+1} - m'_k) \quad k = 1, 2, \ldots, n-1. \tag{21}$$

It can be easily verified that condition (i) of Assumptions 1 and 2 is satisfied here. Below, we investigate all the other conditions:

(a)
$$\frac{\partial U(\tilde{z}_t, z_t)}{\partial \tilde{z}_t} = \begin{cases} b_1 & 0 \le z_t < a_1\tilde{z}_t \\ b_2 & a_1\tilde{z}_t < z_t < a_2\tilde{z}_t \\ \vdots & \\ b_n & a_{n-1}\tilde{z}_t < z_t \le a_n\tilde{z}_t, \end{cases}$$
which is clearly positive since $\{b_k\}_{k=1}^n > 0$. However, at points $a_k\tilde{z}_t, k = 1, 2, \ldots, n-1$, the partial derivative does not exist ($n-1$ discontinuities). Similarly,
$$\frac{\partial L(\tilde{z}_t, z_t)}{\partial \tilde{z}_t} = \begin{cases} -b'_1 & \tilde{z}_t \le z_t < a'_1\tilde{z}_t \\ -b'_2 & a'_3\tilde{z}_t < z_t < a'_2\tilde{z}_t \\ \vdots & \\ -b'_n & a'_{n-1}\tilde{z}_t < z_t \le a'_n\tilde{z}_t, \end{cases}$$
which is negative since $\{b'_k\}_{k=1}^n > 0$. We also have $n-1$ points of discontinuity here.

(b) For $c_1, c_2 \in [0, \infty]$ with $c_1 < c_2$, it readily follows from (a) that
$$\frac{\partial U(\tilde{z}_t, p)}{\partial \tilde{z}_t}\Big|_{\tilde{z}_t = c_1} \le \frac{\partial U(\tilde{z}_t, p)}{\partial \tilde{z}_t}\Big|_{\tilde{z}_t = c_2}$$
$$\frac{\partial L(\tilde{z}_t, q)}{\partial \tilde{z}_t}\Big|_{\tilde{z}_t = c_1} \le \frac{\partial L(\tilde{z}_t, q)}{\partial \tilde{z}_t}\Big|_{\tilde{z}_t = c_2},$$
where $0 < p < c_1$ and $c_2 < q < \infty$.

(c) For a given $L(\tilde{z}_t, z_t)$, $J_L(\tilde{z}_t, \infty; \tilde{z}_t)$ can be expressed in terms of the first moment of the forecast distribution. Thus, condition (iv) of Assumption 2 is satisfied. Since the partial derivatives of $L(\tilde{z}_t, z_t)$ evaluates to a constant value in each of the n subintervals, it follows that condition (v) of Assumption 2 is also satisfied.

Thus, the piecewise linear functions of the form given above with at most countable number of regions (intervals), satisfy all the imposed conditions. It then follows that we have a unique z_t^* that minimizes the cost function.

For the piecewise linear functions given above, the cost function is given as follows:

$$\mathcal{C}(z_t) = \sum_{i=1}^{n} \left\{ \int_{a_{i-1}\tilde{z}_t}^{a_i\tilde{z}_t} (b_i\tilde{z}_t - m_i z_t)\ f_{t-1}(z_t)\ dz_t \right\} + \sum_{j=1}^{n} \left\{ \int_{a'_{j-1}\tilde{z}_t}^{a'_j\tilde{z}_t} (m'_j z_t - b'_j\tilde{z}_t)\ f_{t-1}(z_t)\ dz_t \right\}. \tag{22}$$

where $a_0 = 0, a_n = a'_0 = 1, a'_n = \infty$. Upon differentiating (22) with respect to $\tilde{z}_t$ and equating to zero, we obtain (after simplification):

$$\sum_{i=1}^{n-1}(b_i - b_{i+1})F_{t-1}(a_i\tilde{z}_t) + (b_n + b'_1)F_{t-1}(\tilde{z}_t) + \sum_{j=1}^{n-1}(b'_{j+1} - b'_j)F_{t-1}(a'_j\tilde{z}_t) = b'_n. \tag{23}$$

Below, we briefly illustrate the computational aspects of step functions (used as penalty functions) for the cases $n = 1$ and $n = 2$, i.e., functions with one and two subintervals.

1) Computation of z_t^* for $n = 1$: It can be observed that for $n = 1$, (23) can be simplified as follows:

$$F_{t-1}(\tilde{z}_t) = \frac{b_1'}{b_1 + b_1'}. \tag{24}$$

In particular, if $b_1 = m_1 = \kappa_1$ and $b_1' = m_1' = \kappa_2$, then (24) becomes equivalent to (14). Thus, the computation of optimal forecast z_t^* is similar to that of first order polynomial case described in Section III-A.1.

2) Computation of z_t^* for $n = 2$: When we have step functions with two sub-intervals ($n = 2$) as penalty functions, simplifying (23) yields the following linear relation:

$$\begin{aligned} b_1 \int_0^{a_1\tilde{z}_t} f_{t-1}(z_t)dz_t + b_2 \int_{a_1\tilde{z}_t}^{\tilde{z}_t} f_{t-1}(z_t)dz_t \\ - b_1' \int_{\tilde{z}_t}^{a_1'\tilde{z}_t} f_{t-1}(z_t)dz_t - b_2' \int_{a_1'\tilde{z}_t}^{\infty} f_{t-1}(z_t)dz_t = 0. \end{aligned} \tag{25}$$

Without loss of generality, consider $b_2 = m_2 = r$ and $b_1' = m_1' = 1.0$. Using the relations from (20) and (21), the linear relation (25) can be shown to be equivalent to

$$\begin{aligned} a_1(m_1 - r)F_{t-1}(a_1\tilde{z}_t) + (1 + r)F_{t-1}(\tilde{z}_t) \\ a_1'(m_2' - 1)F_{t-1}(a_1'\tilde{z}_t) = 1 + a_1'(m_2' - 1). \end{aligned} \tag{26}$$

Since $F(\cdot)$ is a strictly increasing function, (26) can be solved for the optimal z_t^* using a binary search algorithm similar to that of $n = 1$ case. The lower and upper limits for the search algorithm can be appropriately chosen based on the known constants a_1, a_1', m_1 and m_2'. The convergence rate of the search algorithm depends upon the conditional distribution of the forecast model.

IV. A Real Network Traffic Example

In this section, we first present a summary of the time series model obtained for a real network example; a detailed discussion can be found in [10]. We then demonstrate the quantitative results for the first and the second order polynomial penalty functions, followed by results for a piecewise linear penalty function with two sub-intervals.

A. Data sets and Time series modeling

The measurements are collected on the Internet link connecting the University of Missouri–Kansas City (UMKC) to MOREnet (a Internet Service Provider) using a Multi Router Traffic Grapher (MRTG) traffic monitoring system; these measurements represent the nonstationary data rate of the traffic from outside world to UMKC averaged over every 5-minute interval (granularity). We have grouped the measurements into three data sets referred to as data set I, data set II and data set III. Each data set spans measurements over 24-hours for 15 days excluding weekends and any holidays as our interest is to model for the weekday behavior. The measurements are collected in the months of September, November and December of year 2003. For further details on data sets, see [10], [11]. The first ten days' data from the data sets were used in determining the parameters' estimates and then, tested on our model over the next five days.

First, we have averaged three successive sample points (of 5 minutes each) of the collected measurements to obtain the data rate over every 15-minute interval. Our intent behind this aggregation is under the assumption that a service provider may not want to do bandwidth updates (based on forecasts) more frequently then every fifteen minutes; this also gives us one-step prediction to look out for the next fifteen minutes. Note that this is done for the purpose of our study and the time series model presented is quite general. Over ten days, we thus consider 960 samples for model identification in each data set with 96 samples per day. In general, we denote the sample size per day using T; for this study, T happens to be 96. Fig. 3 shows a less-detailed view of 15-minute aggregate data rate observed in data set I while Fig. 4 shows a detailed view of the day 1 traffic data rate from data set I. A similar coarse-level behavior is observed in the other two data sets as well.

The first obvious observation from Fig. 3 is the evidence of the time-of-the-day effect. In a wider sense, the data rate behavior over any given weekday is similar to that of any other weekday in the sample. However, we have observed a few noticeable peaks on certain days in the data sets. For example, relatively higher peaks can be observed on day 1, day 6 and day 8 than the rest in Fig 3. These peaks result in outliers (data points that are extreme relative to the rest of the sample) that can distort the innovation distribution as well as the statistical estimates of the parameters of the resulting model. Nevertheless, these outliers are genuine and essential in characterizing the traffic dynamics since they might arise due to the inherent variability of the underlying traffic. Thus, we do not eliminate them in our analysis; rather, we transform the aggregated time series using the natural logarithmic transformation. Such a transformation is shown to be quite effective in stabilizing the data points (free of outliers). Following this transformation, based on the auto-correlation function, we identified the need for first differencing as well as seasonal differencing (based on duration T) in order to obtain a stationary

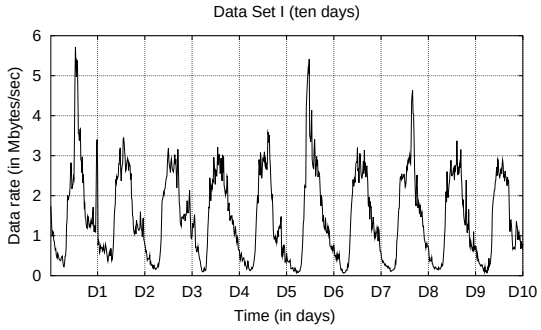

Fig. 3. Data set I: Less detailed view

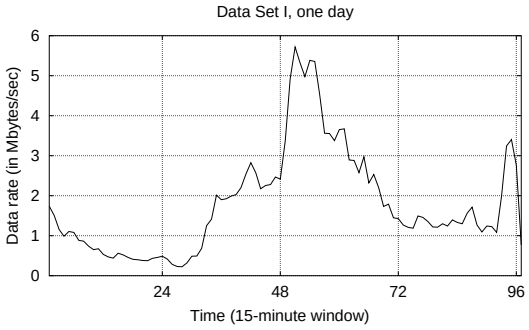

Fig. 4. Data set I: Detailed view of Day-1 (D1)

series of first order. In the resulting stationary series, we observed the presence of the clustering of large values and small values. Such a clustering effect arises primarily from the strong dependence of the innovation variance at any given sample point over the innovation variance of past sample points, i.e., innovation variance is correlated. In order to model this dependency, we include a conditionally heteroskedastic component as a part of the model. Thus, the time series model so obtained for the transformed series is a seasonal ARCH-type model where we restrict the process model for innovation (disturbances) to Gaussian distribution.

Let z_t represent the observed time series and w_t represent the log transformed series i.e., $w_t = \ln z_t$. Then, the stationary series y_t of first order obtained after the differencing operation at two levels is given as follows:

$$y_t = (w_t - w_{t-1}) - (w_{t-T} - w_{t-T-1}). \tag{27}$$

Following the discussion above, the proposed seasonal ARCH model for the transformed series y_t is given below:

$$y_t = \sum_{i=1}^{4} \varphi_i y_{t-i} + \sum_{j=1}^{2} \phi_j y_{t-jT} + \varepsilon_t \tag{28a}$$

$$\varepsilon_t = \eta_t \sqrt{h_t} \tag{28b}$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2. \tag{28c}$$

The unknown parameters of the model (28) are: $\{\varphi_i\}_{i=1}^4$ (autoregressive coefficients), $\{\phi_j\}_{j=1}^2$ (seasonal autoregressive coefficients), α_0 (ARCH intercept) and α_1 (first order ARCH co-efficient). Here, η_t is an independent and identically distributed normal random variable with zero mean and unit variance. However, it should be noted that ε_t are merely uncorrelated and not independent (higher moments may be correlated), i.e., $E[\varepsilon_t \varepsilon_{t-1}] = 0$ and $E[\varepsilon_t^2 \varepsilon_{t-1}^2] \neq 0$. It follows then, the conditional distribution of ε_t given the information up to and including time $t-1$. is normal with mean 0 and variance h_t. The proposed model for series y_t is chosen from various candidate models based on the smallest Bayesian Information Criterion (BIC) [5] (a commonly adopted conservative model selection criterion). For this model identification and estimation procedure, we used SAS, a well-known software package [14]. The parameters of the model are estimated based on the sample maximum likelihood function conditioned on the first $10T$ observations; see [5], [8] for more details on the statistical inference of the time series models. We present the estimates of all the parameters along with their t-values in Appendix I for data set I. For the other data sets, the estimates of these parameters are indeed different; however, the order of the model remains the same as (28). It may be noted that the estimate of α_1 is *highly statistically significant* from zero (i.e., at the 99% significant level to reject the null hypothesis that it is zero). Such a high statistical significance for the estimate of α_1 is observed in other two data sets as well. Thus, the ARCH-effect is justified.

It is evident from (28a) that the distribution of one-step ahead forecast of y_t is normally distributed with mean $E_{t-1}[y_t]$ and variance h_t. The reason that we restrict ourselves to one-step ahead forecasts is to make use of the new observation (z_t) as soon as it is available. Using relation (27) and $w_t = \ln z_t$, we can express (28a) in terms of z_t as follows:

$$\ln z_t = \sum_{i=1}^{5} \chi_i \ln z_{t-i} + \sum_{j=T}^{T+5} \chi_j \ln z_{t-j} + \sum_{k=2T}^{2T+1} \chi_k \ln z_{t-k} + \sum_{s=3T}^{3T+1} \chi_s \ln z_{t-s} + \varepsilon_t, \tag{29}$$

where χ's are computed from the linear and the seasonal autoregressive coefficients, φ and ϕ, respectively, from (28a). The one-step ahead minimum mean square error (MMSE) forecast $\tilde{z}_t$ of z_t (the desired quantity) at time $t-1$ is given by $E_{t-1}[e^{\ln z_t}]$, which can be obtained

from (29) as shown below. Let

$$C_{t-1} = \sum_{i=1}^{5} \chi_i \ln z_{t-i} + \sum_{j=T}^{T+5} \chi_j \ln z_{t-j} + \sum_{k=2T}^{2T+1} \chi_k \ln z_{t-k} + \sum_{s=3T}^{3T+1} \chi_s \ln z_{t-s}. \quad (30)$$

It may be noted that C_{t-1} is completely known at time t. Thus,

$$\begin{aligned} E_{t-1}[e^{\ln z_t}] &= e^{C_{t-1}}\ E_{t-1}[e^{\varepsilon_t}] \\ &= e^{C_{t-1}}\ e^{\frac{h_t}{2}}. \end{aligned} \quad (31)$$

Since, the conditional distribution of ε_t follows a normal distribution, we have e^{ε_t} to follow a "lognormal" distribution. Then, it follows from (31) that given the information up to and including time $t-1$, z_t is conditionally lognormally distributed with mean μ_t and variance σ_t^2 where

$$\mu_t = e^{C_{t-1}} e^{\frac{h_t}{2}} \quad (32)$$

$$\sigma_t^2 = e^{2C_{t-1}}\ e^{h_t}(e^{h_t} - 1). \quad (33)$$

Note that all the moments of a lognormally distributed random variable are well-defined and hence the random variable z_t has well-defined moments. Consequently, an optimal forecast z_t^* can always be obtained for penalty functions satisfying Assumptions 1 and 2 listed in Section II based on (1). The optimality of the forecast is based on the nature and the relative magnitude of the penalty components associated with the under and over-forecasts.

In the following sections, we evaluate the forecasting results based on polynomial penalty functions (first and second order) and piecewise linear penalty functions (with two sub-intervals) against the MMSE forecast as applied to Day-11 of data sets. Toward this end, we consider three metrics: (i) Fractional Increase in error due to Over-Forecast (FIOF), (ii) Fractional Decrease in error due to Forecast Misses (FDFM), and (iii) Difference in the number of Forecast Misses (DFM). Let the cumulative error observed due to the over-forecast and the under-forecast for the forecasting scheme S be denoted by CE_S^{of} and CE_S^{uf}, respectively. Furthermore, let NFM_S be the number of forecast misses observed for the forecasting scheme S. Then, the metrics are defined as follows:

$$\text{FIOF}_S = \frac{(CE_S^{of} - CE_{MMSE}^{of})}{CE_{MMSE}^{of}} \quad (34)$$

$$\text{FDFM}_S = \frac{(CE_S^{uf} - CE_{MMSE}^{uf})}{CE_{MMSE}^{uf}} \quad (35)$$

$$\text{DFM}_S = NFM_S - NFM_{MMSE}. \quad (36)$$

Note that the scheme S refers to one among the first order polynomial case, the second order polynomial case and the piecewise linear function case.

For space considerations, we present our results primarily for data set I. Nevertheless, we present plots in several figures, which also include results for data sets II and III for comparison. We first present the cumulative under and over-forecast and the number of forecast misses associated with the MMSE forecast in Table I. These values help us to comprehend the results on the three metrics in a better manner.

TABLE I
MMSE FORECAST DETAILS FOR DATA SET I

CE_{MMSE}^{of}	CE_{MMSE}^{uf}	NFM_{MMSE}
10065445.29432951	7921418.801668145	41

B. Results for first and second order polynomial functions

For ease of discussion, henceforth, we refer to the first order and the second order case as the "linear" and the "quadratic" scheme, respectively. Most of our discussions here will be for results based on data set I, while in some of the figures we also report results for data sets II and III for the purpose of comparison. In Fig. 5, we have shown the one-step ahead forecast envelope for the linear and quadratic penalty functions along with the MMSE forecast envelope for $r = 0.1$. For clarity, the samples are shown from near the end of Day-10 to the end of Day-11, split over three parts.

Notice from Fig. 5 that there is a distinct gap between the MMSE forecast envelope and the other two forecast envelopes. When $r = 0.1$, the penalty factor associated with the under-forecast is ten times that of over-forecast due to which, the forecast envelope associated with the linear and the quadratic scheme is much above that of the MMSE forecast envelope. As r increases, the gap decreases and eventually for a particular r value, the modified forecast envelope coincides with the MMSE forecast envelope. In particular, in the case of quadratic forecast scheme, it can be recalled from Section III-A.2 that the modified forecast envelope coincides with the MMSE forecast envelope when $r = 1.0$. On the other hand, for the linear scheme, the modified forecast envelope coincides with the MMSE forecast envelope for some $r < 1.0$. In other words, for $r = 1.0$, the modified forecast envelope falls below the MMSE forecast envelope. This is due to the fact that the modified forecast coincides with the "median" of the conditional forecast distribution for $r = 1.0$ which is *always smaller*

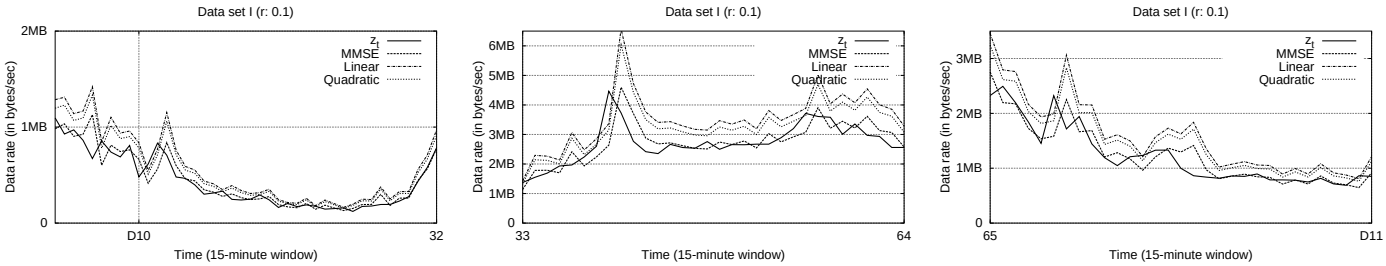

Fig. 5. MMSE forecast curve along with linear and quadratic scheme based forecast, shown from near the end of Day-10 (D10) to the end of Day-11 (D11); split over three parts to show finer details

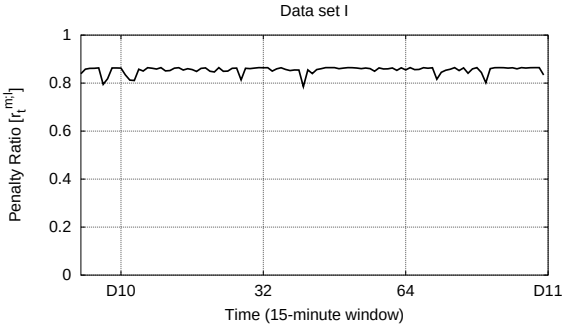

Fig. 6. Data set I: r associated with MMSE for linear scheme ($r_t^{m;l}$)

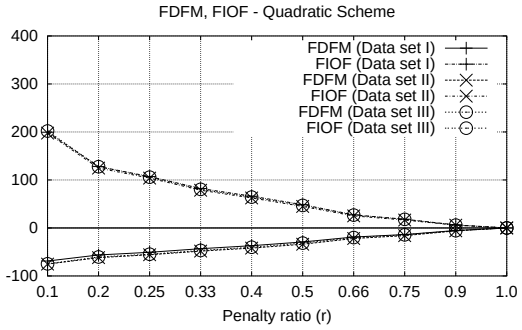

Fig. 7. FIOF and FDFM for linear scheme

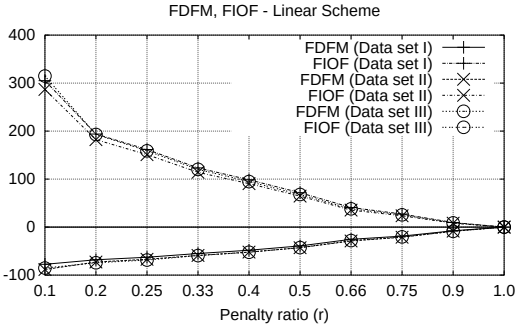

Fig. 8. FIOF and FDFM for quadratic scheme

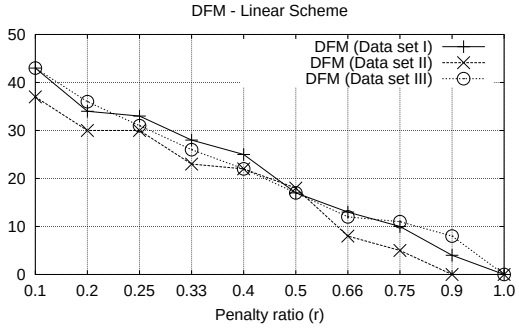

Fig. 9. DFM for linear scheme

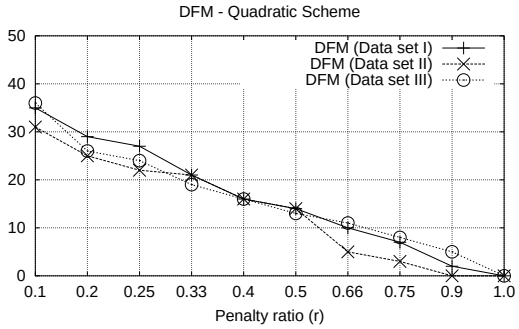

Fig. 10. DFM for quadratic scheme

than the mean (MMSE forecast) in the case of lognormal distribution. In fact, we computed the value of r required at each time instant in order to obtain the MMSE forecast from the forecast cost function associated with the linear scheme. For convenience, we denote this ratio by $r_t^{m;l}$. Then, $r_t^{m;l}$ can be computed at any instant t using (37) from (14).

$$r_t^{m;l} = \frac{1}{F(z_t^m)} - 1, \tag{37}$$

where z_t^m is the MMSE forecast at time t. Fig. 6 illustrates $r_t^{m;l}$ for all time instants from near the end of Day-10 to the end of Day-11 for data set I. As can be noticed, $r_t^{m;l}$ stays close to 0.85 (less than 1.0) with occasional little fluctuations.

In Figs. 7 and 8, we illustrate the results of FIOF and FDFM for various r values. One of the major observation is that we always achieve a higher FIOF and FDFM with the linear scheme than with the quadratic

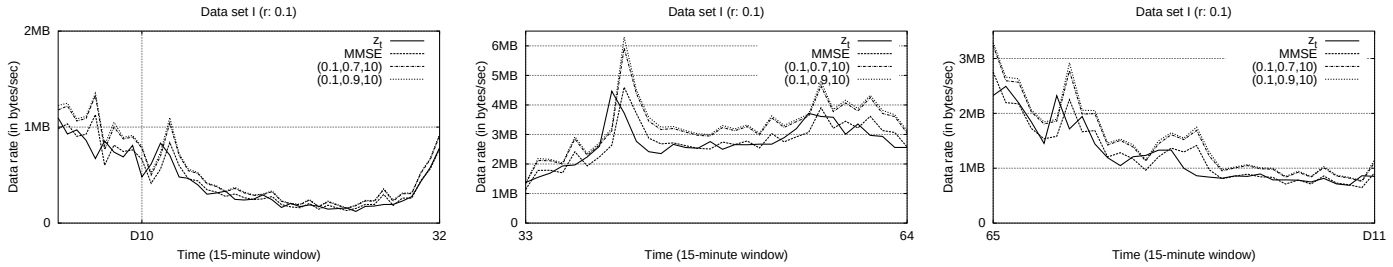

Fig. 11. MMSE forecast curve along with piecewise linear function based forecast, shown from near the end of Day-10 (D10) to the end of Day-11 (D11); split over three parts to show finer details

scheme. For r closer to zero, we notice approximately an exponential increase in FIOF for both the schemes due to an extensive bias toward avoiding under-forecast. For example, when $r = 0.1$, we observe around a 80% and 70% decrease in error due to under-forecast with an approximate 300% and 200% increase in error due to over-forecast respectively, for the linear and the quadratic scheme. As r increases, we witness a gradually decaying gain in FDFM while a gradual decrease in FIOF, eventually converging to zero gain/loss from that of MMSE forecast. Note that for $r = 0.9$ onward, FDFM becomes positive while FIOF becomes negative in the case of linear scheme implying an inferior performance with respect to the MMSE forecast scheme. Such a behavior is attributed to the fact that the forecast from the linear scheme coincides with the MMSE forecast for r in the neighborhood of 0.85 (see Fig. 6). Furthermore, we present the decrease in forecast misses associated with the linear and the quadratic schemes with respect to the MMSE forecast in Figs. 9 and 10. A significant decrease is observed in the number of forecast misses for values of r closer to zero, supporting the observed behavior of FDFM in Figs. 7 and 8. Upon increasing r, DFM converges toward zero. In particular, DFM reaches zero when $r = 1.0$ with the quadratic scheme while it becomes positive with the linear scheme. This behavior is consistent with the observation made in Figs. 9 and 10.

In summary, for the quadratic scheme, the optimal forecast z_t^* at time t is always above that of MMSE forecast z_t^m at time t, as long as r lies in the open interval $(0, 1)$. On the other hand, for the linear scheme, the optimal forecast z_t^* at time t will be above that of MMSE forecast for r taking values in the open interval $(0, a)$ and will fall below the MMSE forecast for r taking values in the closed interval $[a, 1]$ for some $a < 1.0$. The value of a, however, depends on the parameters of the conditional forecast distribution, which is lognormal in our case.

C. Results for Piecewise Linear Function $(n = 2)$

It is evident from (26) that the parameters involved in computing the optimal z_t^* are: r, a_1, m_1, a_1', m_2'. In our evaluation exercises, we consider $m_1 = rm_2'$ and $a_1' = \frac{1}{a_1}$. This means that the parameters involved are: r, a_1 and m_2'. The implication of choosing $m_1 = rm_2'$ here is to reduce the slope of the linear functions associated with the over-forecast ($U(\tilde{z}_t, z_t)$) in each interval uniformly by a factor of $1 - r$ relative to the respective linear functions associated with the under-forecast ($L(\tilde{z}_t, z_t)$).

Fig. 11 shows the one-step ahead forecast envelope for two sets of values for the parameters in the order (r, a_1, m_2') along with MMSE forecast envelope. Similar to the case of polynomial penalty functions, for clarity, the samples are shown from near the end of Day-10 to the end of Day-11 split over three parts. It is evident from Fig. 11 that the forecast envelopes corresponding to $(0.1, 0.7, 10)$ and $(0.1, 0.9, 10)$ are well above the MMSE forecast envelope. In other words, values of r closer to zero yields a more conservative forecast envelope than the MMSE forecast envelope. As we increase r while keeping both a_1 and m_2' constant, the slope of the linear functions associated with $U(\tilde{z}_t, z_t)$ increases which, in turn, brings the forecast envelope closer to the MMSE envelope. The choices of a_1 and m_2' primarily control the value of r that brings down this gap to a negligible level. For convenience, we use the notation $r_t^{m;pl}$ to refer to the value of r that yields the MMSE forecast at time t for some given a_1 and m_2'. To compute $r_t^{m;pl}$ at any instant t, we solve (38) for r obtained from (26) by substituting $a_1' = \frac{1}{a_1}$ and $m_1 = rm_2'$ to obtain:

$$r_t^{m;pl} = \frac{\frac{1}{a_1}(m_2' - 1)\{1 - F\left(\frac{z_t^m}{a_1}\right)\} + \{1 - F(z_t^m)\}}{\{F(z_t^m) + a_1(m_2' - 1)F(a_1 z_t^m)\}}, \tag{38}$$

where z_t^m is the MMSE forecast at time t.

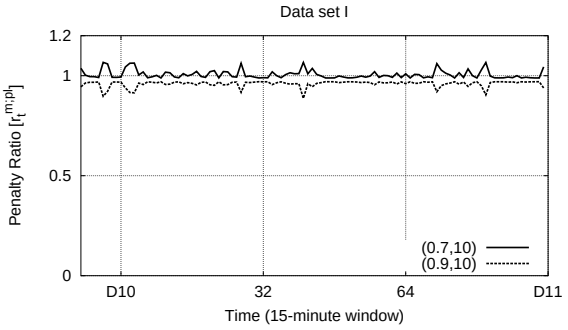

Fig. 12. Data set I: r corresponding to MMSE (r_t^m)

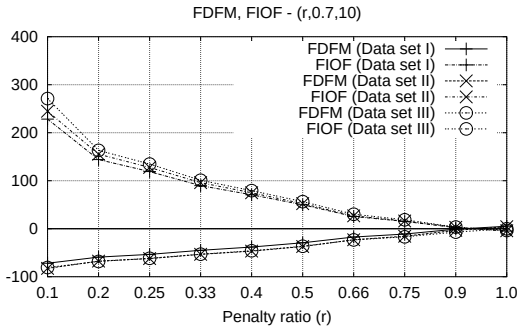

Fig. 13. FIOF and FDFM for Piecewise Linear Case $(r, 0.7, 10)$

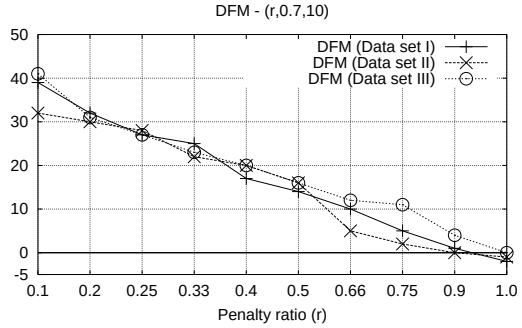

Fig. 14. DFM for Piecewise Linear Case $(r, 0.7, 10)$

Fig. 12 illustrates $r_t^{m;pl}$ for the two cases discussed above. The average values of $r_t^{m;pl}$ for the cases $(a_1 = 0.7, m_2' = 10)$ and $(a_1 = 0.9, m_2' = 10)$ are 1.01 and 0.96, respectively. As can be observed, the variation around the average is very minimal. Furthermore, we notice that the path of $r_t^{m;pl}$ over time for the case $(0.7, 10)$ resembles an approximate mirror image of the case $(0.9, 10)$. This near symmetrical behavior is possibly due to the inverse proportionality relationship between a_1 and a_1'.

In Fig. 13, we illustrate the results of FIOF and FDFM for various r values where the zero axis is associated with the MMSE forecast. For instance, when $r = 0.1$, i.e., for the case $(0.1, 0.7, 10)$, we observe around 70% decrease in error due to under-forecast (forecast misses) with an approximate 200% increase in error due to over-forecast. These values substantiate the observation made in Fig. 11 regarding the significant gap between the modified forecast envelope and the MMSE forecast envelope. Note that on increasing r, gain in FDFM decays gradually along with a gradual decrease in FIOF. Since the average value of $r_t^{m;pl}$ for the case $(0.7, 10)$ is just above 1.0, FDFM and FIOF remain negative and positive, respectively, even with $r = 1.0$. For completeness, the difference in forecast misses (DFM) is presented in Fig. 14.

To summarize, if piecewise linear functions with $n = 2$ sub-intervals are used as penalty functions, the optimal forecast z_t^* at any instant t for a given r in the half-closed interval (0, 1], depends on a few parameters. In our study, the parameters that control the modified forecast are reduced to two; a_1 and m_2'. However, this need not be the case in practice. In other words, there might not exist any dependencies between a_1 and a_1' and m_1 and m_2' as considered here. In any case, the choice of the values for these four parameters influences the optimal forecast z_t^* at any instant t.

In the case of data set II and data set III, a similar overall trend is observed for the FDFM, the FIOF and the DFM metrics with respect to increasing r in the half-closed interval $(0, 1]$ though the absolute values may vary; this trend can be seen in Figs. 13 and 14.

V. Summary

In this paper, we sought to differentiate the impact of under-forecast and over-forecast on the network performance by presenting a generic forecast cost function. This function, upon minimization, yields the desired forecast subjected to given penalties. Such an approach toward deriving forecast is especially more desirable in the context of adaptive bandwidth provisioning where the data loss (arising from under-forecast) is a more stringent criterion than the under-utilization of bandwidth (due to over-forecast). Our results hold for forecast distributions for which moments are well defined; this is a fairly general assumption that holds for distributions such as normal, exponential or log-normal distributions. In some practical situations, heavy-tailed distributions might be encountered, which do not have all the orders of moments well-defined. In such cases, an approximate method can be used, as described in [10].

We have enumerated a few mild assumptions on the first-order characteristics of penalty functions, respectively for data loss and under-utilization, that guarantee the existence of a unique forecast under the generic cost function framework.

We considered various classes of functions that can be candidate penalty functions in a real network. In particular, when penalty functions belong to polynomial classes of functions, they yield equations that are computationally easy to solve. In fact, we have presented the forecast equations (to be solved to obtain the desirable forecast) for the first and the second order polynomial functions. We have shown that the MMSE forecast can be obtained as a special case of second order polynomial function.

To assess the effectiveness of our approach, we presented quantitative results for the polynomial class of functions and piece-wise linear functions as applied to real world traffic measurements. One of the inferences from these exercises is that MMSE forecast can always be obtained as a special case from the corresponding forecast cost function. Such special cases however, turns out be less desirable from the bandwidth provisioning perspective. Thus, our approach toward prediction can be more general with respect to provisioning of bandwidth in a real network. The forecasting models proposed hare can be automated in a monitoring environment.

It may be noted that there is no one-size-fits-all time series model for all types of network traffic patterns and conditions. In a related work [10], we have noted that the time-series model is relatively stable for a link, but the coefficients of the model may change from time to time and should be updated. Since the measurement data we collected on UMKC's Internet link is primarily TCP-based, the general model is expected to be useful for TCP-based traffic, however, only at the aggregate level. For streaming UDP-based traffic, it remains to be seen applicability of such models. In our approach, updates are typically envisioned to be done every 15 minutes; since our computation takes only a few seconds on a standard pentium-based computer, the method proposed can be implemented in a real-world environment. Finally, this paper focusses on the generalized cost-based forecasting approach, which is a step in determining bandwidth provisioning; detailed results on bandwidth provisioning schemes can be found in [10].

APPENDIX I

We present the estimates of parameters along with their t-statistics in parenthesis evaluated through the maximum likelihood function for the data set used in our evaluation in Table II.

TABLE II
DATA SET I - PARAMETER ESTIMATES

φ_1	-0.147359828 (-4.35)	ϕ_1	-0.586686018 (-19.12)
φ_2	-0.10856727 (-3.95)	ϕ_2	-0.2604787 (-8.36)
φ_3	-0.099462816 (-3.25)	α_0	0.0332740145 (17.99)
φ_4	-0.137403164 (-4.11)	α_1	0.196148845 (3.82)

REFERENCES

[1] Y. Afek, M. Cohen, E. Haalman, Y. Mansour, "Dynamic Bandwidth Allocation Policies," *Proc. of IEEE INFOCOM'96*, pp. 880-887, March 1996.

[2] T. Anjali, C. Bruni, D. Iacoviello, G. Koch, C. Scoglio, "Filtering and Forecasting Problems for Aggregate Traffic in Internet Links," *Performance Evaluation*, vol. 58, no. 1, pp. 25-42, 2004.

[3] A. K. Bera, M. L. Higgins, "ARCH Models: Properties, Estimation and Testing," *Journal of Economic Surveys*, vol. 7, no. 4, pp. 305-362, 1993.

[4] T. Bollerslev, R. F. Engle, D. B. Nelson, "ARCH Models," *Handbook of Econometrics*, vol.4, pp. 2961-3038, 1994.

[5] G. E. P. Box, G. Jenkins, G. C. Reinsel, *Time Series Analysis: Forecasting and Control*, Third Edition, Prentice Hall, 1994.

[6] R.F. Engle, "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, vol. 50, no. 4, pp. 987-1007, July 1982.

[7] N. K. Groschwitz, G. C. Polyzos, "A Time Series Model of Long-Term NSFNET Backbone Traffic," *Proceedings of ICC'94*, pp. 1400-1404, May 1994.

[8] J. D. Hamilton, *Time Series Analysis*, Princeton University Press, 1994.

[9] B. Krithikaivasan, K. Deka, D. Medhi, "Adaptive Bandwidth Provisioning Envelope based on Discrete Temporal Network Measurements," *Proceedings of IEEE INFOCOM '04*, pp. 1786-1796, March 2004.

[10] B. Krithikaivasan, Y. Zeng, K. Deka, D. Medhi, "ARCH-based Traffic Forecasting and Dynamic Bandwidth Provisioning for Periodically Measured Nonstationary Traffic," *IEEE/ACM Trans. on Networking*, vol. 15, August 2007 (to appear).

[11] B. Krithikaivasan, "Forecasting Models and Adaptive Quantized Bandwidth Provisioning for Nonstationary Network Traffic," *Ph.D. Dissertation*, University of Missouri–Kansas City, May 2006.

[12] K. Papagiannaki, N. Taft, Z. Zhang, C. Diot, "Long-Term Forecasting of Internet Backbone Traffic: Observations and Initial Models," *Proceedings of IEEE INFOCOM'03*, pp. 1178-1188, April 2003.

[13] A. E. Taylor, W. R. Mann, *Advanced Calculus*, John Wiley, 3rd Edition, 1983.

[14] *SAS Documentation*, `http://v9doc.sas.com/sasdoc`.