

Change Detection in Teletraffic Models

Rittwik Jana and Subhrakanti Dey

Abstract—In this paper, we propose a likelihood-based ratio test to detect distributional changes in common teletraffic models. These include traditional models like the Markov modulated Poisson process and processes exhibiting long range dependency, in particular, Gaussian fractional ARIMA processes. A practical approach is also developed for the case where the parameter after the change is unknown. It is noticed that the algorithm is robust enough to detect slight perturbations of the parameter value after the change. A comprehensive set of numerical results including results for the mean detection delay is provided.

Index Terms—Autoregressive integrated moving average, change detection, long memory processes, Markov modulated Poisson process.

I. INTRODUCTION

CHANGE detection algorithms have been studied extensively for the past 50 years [1], [2]. Adaptive identification algorithms can track only slow fluctuations of the characteristic parameters and are not suited for detection of abrupt changes in general. Detection of abrupt changes, however, is necessary in many applications like fault detection in navigational systems, onset detection in seismic signal processing, segmentation of speech signals and congestion control of wide area computer networks, etc. We are interested in the problem of online detection of an abrupt change, with the minimum delay in detection with a constraint on the mean time between false alarms. Most of the signals treated in [2] are time-series models (linear or nonlinear). Following techniques similar to [3], an online change detection algorithm was developed in [4] for Markov-modulated time-series models based on the cumulative sum (CUSUM)-like test derived from Page's test and the sequential probability ratio test (SPRT) [5]. In this paper, we develop online change detection algorithms based on a CUSUM-like procedure for teletraffic models, ranging from traditional models like Markov-modulated Poisson process (MMPP) to a class of current self-similar models proposed for internet traffic, specifically long-memory time series models.

It is known that the CUSUM test is optimal in the sense that it optimizes the worst mean delay in detection when the mean time between false alarms goes to infinity [2]. The CUSUM algorithm also gives the infimum of the worst mean delay for a class of stopping times with preassigned rate of false alarms

[2]. Simulations in [6] illustrated via a comparative study of five algorithms (for a change in the mean of a Gaussian distribution) that the CUSUM algorithm is more robust and efficient than the others. Optimal properties for the CUSUM algorithm are discussed in detail in [2]. Since the idea of the test is based on calculating the logarithm of the ratio of the conditional likelihood functions before and after the change, we can extend the CUSUM algorithm to the case of dependent observations (although initially it was designed for independent observations), where calculations of such likelihood functions are possible (e.g., hidden Markov models [3] and Markov-modulated time series [4]).

In this paper, we develop on-line change detection algorithms based on the CUSUM-like test for teletraffic models. The first model we treat is a traditional model for traffic data in communication networks known as the Markov modulated Poisson process (MMPP). It has been extensively used for modeling processes like overflow from a finite trunk group and superposition of packetized voice processes and packet data (see [7] and references therein). Although this model can account for time-varying arrival rates and captures some of the important correlations between interarrival times, it cannot model long-range dependence (or long memory with hyperbolically decaying autocorrelation functions arising as a striking feature of the so-called "self-similarity"). Starting with the seminal paper [8], it was shown in many other works that self-similarity existed in ATM traffic, compressed digital video streams, and web traffic between browsers and servers (e.g., see [9]). Although self-similarity is measured by the so-called Hurst parameter (taking values between 0.5 and 1.0), it has been shown that high degrees of self-similarity (Hurst parameter >0.7) have a detrimental effect on network performance, including packet loss and queueing delay (see [9] and references therein). In addition, increasing load on the Ethernet increases the degree of self-similarity. In the context of multimedia traffic such as video and voice with their diverse quality of service (QoS) requirements, it is important therefore that the effect of self-similarity on interconnected issues like queueing delay, packet loss, and buffer sizes is well understood. Hence, the issue of detecting changes in the degree of self-similarity is quite important in such applications. In this paper, we address this issue by developing online change detection algorithms for a class of long-range dependent processes, namely, Gaussian fractional ARIMA (FARIMA) processes.

It is to be remembered, however, that these algorithms are optimal when the parameters before and after the change are exactly known. In most realistic situations, that is not the case (e.g, network traffic). Although we can extend such methods to generalized likelihood ratio (GLR) tests (assuming that the parameters before the change are known but not so after the

Manuscript received May 5, 1999; revised August 31, 1999. The associate editor coordinating the review of this paper and approving it for publication was Dr. Alex B. Gershman.

R. Jana is with the Telecommunications Engineering Group, Research School of Information Systems and Engineering, The Australian National University, Canberra, Australia.

S. Dey is with the Department of Systems Engineering, The Australian National University, Canberra, Australia.

Publisher Item Identifier S 1053-587X(00)01545-2.

change), we can also develop more suboptimal methods by substituting the parameters before the change by their estimated values and running a bank of change detection algorithms with different assumed values for the parameters after the change. Although the closest value will result in a better performance, such algorithms are obviously computationally quite expensive, particularly when the number of parameters is large. Note also that such detection algorithms can potentially be made adaptive. For a comparative study of three such algorithms in the context of time series segmentation, see [10]. There are also other designs of detection algorithms, based on the efficient score (gradient of the log-likelihood function with respect to the parameter vector) or, more generally, on alternative estimating functions (see [11]–[14] for some related robustness issues). Analytical or even approximate computations of the mean delay in detection and mean time between false alarms are quite difficult in the case when the sequence of the test statistic is not independent and identically distributed and will not be considered in this paper for our algorithms except for some simulation results.

II. SIGNAL MODELS

A. Markov-Modulated Poisson Process (MMPP)

An MMPP is a doubly stochastic Poisson process where the rate of the Poisson process is modulated by the state of a Markov chain. Let $s_t \in S = \{1, 2, \dots, N\}$, $t \in \mathbb{N}^+$ denote a finite state, discrete-time, homogeneous Markov chain with a transition probability matrix $A = (a_{ij})$, where $a_{ij} = P(s_{t+1} = i | s_t = j)$ and initial probability distribution given by π such that $P(s_0 = i) = \pi(i)$. The number of arrivals (e.g., of data packets) in the t th time slot is modeled by a Poisson process and is denoted by n_t . The rate of arrival during the t th time slot is given by μ_t , where essentially, μ_t is modulated by s_t . Let $\mu_t \in \mu \triangleq \{\mu(1), \mu(2), \dots, \mu(N)\}$, where $\mu_t = \mu(i)$ if $s_t = i$. Define $B(n_t) = \text{diag}\{b_1(n_t), b_2(n_t), \dots, b_N(n_t)\}$, where $b_i(n) = P(n_t = n | s_t = i)$. Due to the Poisson nature of the process n_t , we obviously have $P(n_t = n | s_t = i) = (\mu(i)^n / n!) \exp(-\mu(i))$. The complete parameter space for the MMPP is then characterized by $\lambda = (A, \pi, \mu)$. The problem at hand is to detect a change from one parameter space $\lambda_H \triangleq (A^H, \pi^H, \mu^H)$ to another $\lambda_K \triangleq (A^K, \pi^K, \mu^K)$. Note that the superscript l used later with the notations above, e.g., in $B(n_t)^l$, will usually identify the parameter space before or after the change for $l = H, K$, respectively.

B. Long-Memory Processes

By generalizing the well-known ARIMA(p, d, q) models of Box-Jenkins [15], it is possible to relax the degree of differencing d to any real value to model long-term persistence. In what follows, we consider stationary fractional ARIMA processes with $0 < d < 1/2$ and Gaussian innovations.

1) *Fractional ARIMA*: We formally define a FARIMA($0, d, 0$) process to be discrete-time stochastic process $\{x_t\}$ represented as

$$\Delta^d x_t = \epsilon_t \quad (1)$$

where the operator Δ^d is defined by

$$\Delta^d = (1 - B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k B^k \quad (2)$$

where $\binom{d}{k} = (\Gamma(d+1)/\Gamma(k+1)\Gamma(d-k+1))$ is defined for any real d , and B is the backward shift operator defined by $Bx_t = x_{t-1}$. $\{\epsilon_t\}$ is a sequence of independent identically distributed (i.i.d.) random variables. In this paper, we assume ϵ_t is Gaussian distributed with mean 0 and variance unity.

When $d = 0$, the FARIMA($0,0,0$) process is a white noise with a constant spectral density. For $0 < d < 1/2$, the FARIMA($0, d, 0$) process is stationary with long memory (for a proof, see [16]). The correlations are all positive and decay monotonically and hyperbolically to zero as the lag increases. For the purpose of this work, we assume $0 < d < 1/2$, since for $d \geq 1/2$, the process is not stationary, at least not in the usual sense, since in this case, the usual definition of the spectral density results in a nonintegrable function (see [16]). When $-1/2 < d < 0$, the FARIMA($0, d, 0$) process has short memory since it can be shown [16] that the sum of all correlations is zero. Note that the Hurst parameter is related to d by $d = \text{Hurst parameter} - 1/2$. For $0 < d < 1/2$, the covariance function of $\{x_t\}$ is

$$\gamma_k^x = E\{x_t x_{t-k}\} = \frac{(-1)^k \Gamma(1-2d)}{\Gamma(k-d+1)\Gamma(1-k-d)} \quad (3)$$

and the correlation function

$$\rho_k^x = \frac{\Gamma(1-d)\Gamma(k+d)}{\Gamma(d)\Gamma(k+1-d)} \quad (4)$$

A more general model of the ARIMA family, namely FARIMA(p, d, q), can be defined by a stochastic process $\{y_t\}$ represented as

$$\phi(B)\Delta^d y_t = \theta(B)\epsilon_t \quad (5)$$

where Δ^d is the fractional differencing operator, $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$, $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$, and ϵ_t is a white noise process. The effect of d on distant observations decays hyperbolically as the lag increases, whereas the effects of ϕ and θ parameters decay exponentially. Thus, d may be chosen to describe correlation in a time series between distant observations, whereas ϕ and θ describe the short-term correlation.

In practice, it is expected that the FARIMA(p, d, q) processes are likely to be of most interest when p and q are small [17]. We will consider the simplest of such processes, namely, FARIMA($1, d, 0$) and FARIMA($0, d, 1$). These two time series are given by

$$(1 - \phi B)\Delta^d y_t = \epsilon_t \quad (6)$$

$$\Delta^d y_t = (1 - \theta B)\epsilon_t \quad (7)$$

To ensure stationarity and invertibility of $\{y_t\}$, we assume $0 < d < 1/2$, $|\phi| < 1$, $|\theta| < 1$. The covariance functions of the FARIMA($1, d, 0$) and FARIMA($0, d, 1$) processes can be found in [17] and are given in the Appendix.

III. ONLINE CHANGE DETECTION

Consider a sequence of i.i.d. random variables y_t with a probability density $f_\lambda(\cdot)$ depending on a parameter space character-

ized by λ . Before the unknown change time t_{opt} , the parameter space is given by $\lambda = \lambda_H$, and after the change, it is $\lambda = \lambda_K$. Assuming that the parameter space λ_H and λ_K are completely known *a priori*, the problem is then to detect and estimate this change in parameter. A CUSUM algorithm that uses the logarithm of likelihood ratios to detect such a change was devised by Page [1] in 1954. The key statistical properties of this ratio can be summarized as follows. A change in the parameter space λ is reflected as a change in the sign of the mean of the log-likelihood ratio. Let $S_t = \sum_{i=1}^t s_i$, $s_t = \log(f_{\lambda_K}(y_t)/f_{\lambda_H}(y_t))$. The typical behavior of S_t shows a negative drift before change and a positive drift after change. Although this test was originally designed for an independent sequence $\{y_t\}$, a similar CUSUM like test can be designed for dependent sequences of $\{y_t\}$ as well [2]. This is based on measuring the ratio of the conditional likelihood functions before and after the change. Note from [2] that such a sequential CUSUM-like procedure in a manner similar to Page's recursive test can be written as a recursion in the test statistic S_t in the following manner:

$$S_t = \max\{0, S_{t-1} + g(t)\} \quad (8)$$

where

$$g(t) = \log\left(\frac{f_{\lambda_K}(y_t|y_{t-1}, \dots, y_0)}{f_{\lambda_H}(y_t|y_{t-1}, \dots, y_0)}\right). \quad (9)$$

As far as the change is concerned, the relevant information lies in $g(t)$: the difference between the log-likelihood functions conditioned on the parameter spaces λ_H and λ_K , respectively. The key property that allows detectability in a CUSUM-like procedure is $E(g(t)|\lambda_H) < 0$, $E(g(t)|\lambda_K) > 0$. To detect a change from λ_H to λ_K , usually, a threshold h is set such that a change is detected when $S_t > h$. There are two quantities associated with the detection: the mean time between false alarms and the mean delay in detection. The general nature of these quantities are that the mean time between false alarms increases approximately exponentially with increasing values of the threshold, and mean delay in detection increases approximately linearly with increasing values of the threshold. These two properties make the CUSUM test quite useful. Computation of the exact conditional log-likelihood functions is possible (as shown in the next section) for the MMPP process but not for long-memory processes. Therefore, we resort to approximate computations for long-memory processes.

A. MMPP

Consider an MMPP defined in Section II-A. Note that here, we replace $\{y_t\}$ by the observation sequence $\{n_t\}$. Define the following forward variable $\alpha_t^l \triangleq (\alpha_t^l(1), \dots, \alpha_t^l(N))'$, where obviously, $\alpha_t^l \in \mathbb{R}^N$, such that the following recursion in α_t^l holds:

$$\begin{aligned} \alpha_t^l &= B(n_t)^l A^l \frac{\alpha_{t-1}^l}{\sum_i \alpha_{t-1}^l(i)} \\ \alpha_0^l &= B(n_0)^l \pi^l, \quad l = H, K. \end{aligned} \quad (10)$$

Remark 1: Note that in the right-hand side of the first equation in (10), α_{t-1}^l is normalized to avoid numerical problems.

It is easy to show that according to the above recursion (see the Appendix for a proof), $\alpha_t^l(j)$ is equal to the quantity $f_{\lambda_l}(n_t, n_{t-1}, \dots, n_0, s_t = j)/f_{\lambda_l}(n_{t-1}, \dots, n_0)$. Then, it easily follows that $f_{\lambda_l}(n_t|n_{t-1}, \dots, n_0)$, $l = H, K$ is given by $f_{\lambda_l}(n_t|n_{t-1}, \dots, n_0) = \sum_j \alpha_t^l(j)$. It is clear that we can now perform computations of the test statistic as given by (8) and (9).

B. Gaussian FARIMA

In this section, we show how we can extend the online change detection algorithm for long-memory time series, specifically, Gaussian FARIMA(0, d , 0), FARIMA(1, d , 0) and FARIMA(0, d , 1).

1) *Gaussian FARIMA(0, d , 0)*: Let $X = \{x_0, \dots, x_{T-1}\}$ be a T -length sequence of observations from a Gaussian FARIMA(0, d , 0) process with $0 < d < \frac{1}{2}$ satisfying (1). For the FARIMA(0, d , 0), the parameter subject to change is d . For FARIMA(1, d , 0), the parameters concerned are d, ϕ and for FARIMA(0, d , 1), they are d, θ . We assume that the parameters before and after the change are completely known. In the next section, where we present simulation results, we consider a practical suboptimal scheme for the case where the parameter (d in that particular case) is unknown after change. Next, we describe how we can go about performing a CUSUM-like test for long memory time series like those given by (1) and (7).

It is well known [16] that an exact likelihood computation of a long-memory time series, e.g., FARIMA(0, d , 0), is computationally prohibitive due to the covariance matrix being high dimensional (for long time series) and often numerically unstable for certain values of d such that inverting the matrix might be a problem. There are several ways to compute an approximate likelihood function like ("Whittle's approximate MLE;" see [18]). We take an alternative approach as given in [16]. Consider [1]. Assuming that the long-memory time series has a causal linear representation, we could write x_t as

$$x_t = \sum_{l=1}^{\infty} b(l)x_{t-l} + \tilde{\epsilon}_t \quad (11)$$

where $\{\tilde{\epsilon}_t\}$ is a sequence of i.i.d. innovations, and asymptotic properties of the AR coefficients $b(\cdot)$ can be found in [16]. If we knew the infinite past of x_t given by x_s, s , we could reconstruct the sequence of i.i.d. innovations $\tilde{\epsilon}_s, s \leq t$. Instead of the infinite past, if only a finite number of past values is observed, the innovations can be estimated by $\tilde{\epsilon}_t = x_t - \sum_{l=1}^{t-1} b(l)x_{t-l}$. Here, we would further truncate the memory such that we only consider M past samples. This is to prevent growing computational needs with increasing length of the time series. In that case, we represent x_t by

$$x_t = \sum_{l=1}^M \beta_{MI} x_{t-l} + e_t(d) \quad (12)$$

where $\sum_{l=1}^M \beta_{MI} x_{t-l}$ is the best linear prediction of x_t , given the past M samples. For a FARIMA(0, d , 0) series, the predictor taps β_{MI} are given by

$$\beta_{MI} = - \binom{M}{l} \frac{\Gamma(l-d)\Gamma(M-d-l+1)}{\Gamma(-d)\Gamma(M-d+1)}. \quad (13)$$

$e_t(d)$ denotes the prediction error at time t , where the dependence on d is explicitly shown. An approximate log-likelihood function $L_t(X_t, d)$ can be calculated by [where $X_t = (x_t, x_{t-1}, \dots, x_1)$] [16], [19]

$$L_t(X_t, d) = -\frac{t}{2} \log(2\pi) - \frac{t}{2} \log \left(\frac{\sum_{i=1}^t \sigma_i^2(d)}{2t} \right) - \frac{1}{2} \sum_{i=1}^t \frac{e_i^2(d)}{\sigma_i^2(d)} \tag{14}$$

where the mean squared prediction error $\sigma_t^2(d) \triangleq E[e_t^2(d)]$ is given by [19]

$$\sigma_t^2(d) = \Gamma(t)\Gamma(t - 2d)/(\Gamma(t - d)^2). \tag{15}$$

An approximate CUSUM-like test can now be devised (we call it approximate because of the approximate log-likelihood function) by computing $g(t)$ in (9) by the logarithm of the ratio of $f_{d_K}(x_t|x_{t-1}, \dots, x_1)$ and $f_{d_H}(x_t|x_{t-1}, \dots, x_1)$, where d_H, d_K are the respective d values before and after the change. It is not hard to see that this is given by (due to Bayes' Theorem)

$$g(t) = (L_t(X_t, d_K) - L_{t-1}(X_{t-1}, d_K)) - (L_t(X_t, d_H) - L_{t-1}(X_{t-1}, d_H)). \tag{16}$$

2) *Gaussian FARIMA(1, d, 0) and FARIMA(0, d, 1)*: The online change detection algorithm is essentially based on the computation of the approximate log-likelihood function for both FARIMA(1, d , 0) and FARIMA(0, d , 1). These approximations are based on the best linear prediction of the time-series given finite number of past samples. Hence, we can repeat the same procedure [as done in the previous subsection for Gaussian FARIMA(0, d , 0)] to obtain the predictor taps, the prediction error covariance, and finally, equations similar to (14) and (16). We will need to use the covariance formulae for FARIMA(1, d , 0) and FARIMA(0, d , 1) (see the Appendix), and to compute the predictor taps recursively, we can use the Levinson-Durbin algorithm [20].

We can potentially consider a case where the long-memory processes like Gaussian FARIMA(p, d, q) for nonzero p or q can have Markov modulated AR or MA parameters. This is an analytically hard problem and will be considered elsewhere.

IV. SIMULATION

In our simulations, the MMPP has an underlying Markov chain that takes values in a 4-D state space. We assume that the process changes from a parameter space λ_H to λ_K after the first 1000 points and then changes back to λ_H after another 1000 points. Different transition probability matrices A_H and A_K were used, along with two different sets of rate of arrivals of packets μ_H and μ_K (details are given in the Appendix). All simulation results presented in this section are averaged over 50 realizations.

Fig. 1 shows the plot of the test statistic clearly showing the changes at $t = 1000$ and $t = 2000$. Fig. 2 shows the corre-

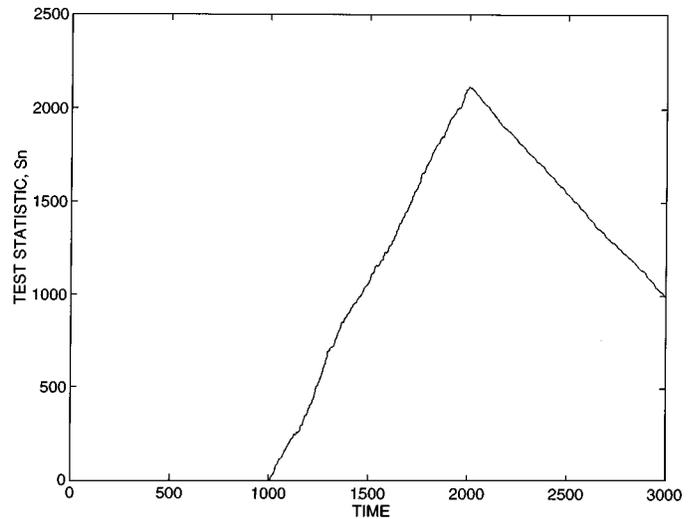


Fig. 1. Change detection for a MMPP process.

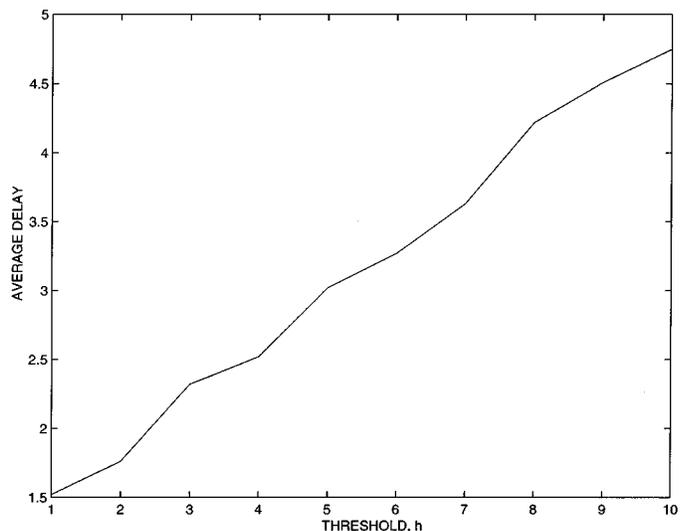


Fig. 2. Delay in detection for a MMPP process. Plot of average delay-threshold.

sponding mean delay in detection (averaged over 50 realizations) versus the detection threshold h . As the threshold is increased, the time required to detect a discernible change increases approximately linearly.

Figs. 3–5 show the change detection for long-memory processes. The data for the FARIMA(0, d , 0) process was generated using a statistical software package SPW, and FARIMA(1, d , 0) and FARIMA(0, d , 1) were generated by passing a FARIMA(0, d , 0) process through appropriate filters. Each process was subdivided into three equal segments of length 1000 such that the process changes from λ_H to λ_K at $t = 1000$ and back to λ_H from λ_K at $t = 2000$. In these simulations, the memory M of the approximate AR representation for the long-memory processes (see (12)) was chosen to be 60. It was noticed that increasing it further did not result in a substantial improvement in the performance of the detection algorithms. As seen in Figs. 3–5, the detection schemes clearly pick up the changes. Fig. 6 shows the delay in detection (averaged over 50 realizations) for the FARIMA(0, d , 0) process of

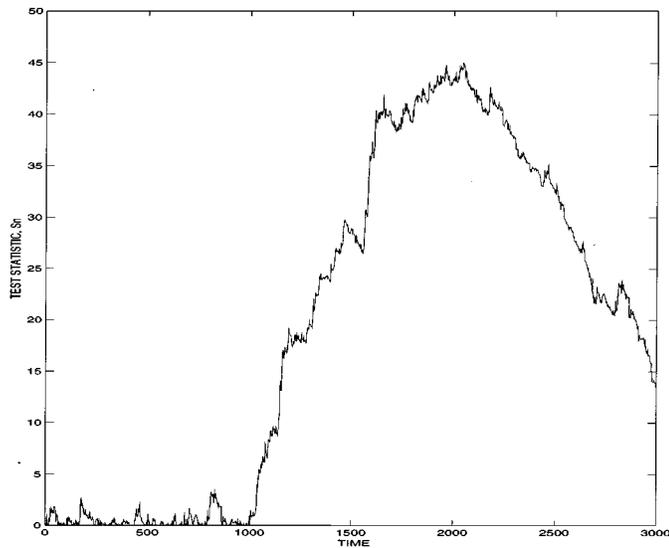


Fig. 3. Change detection for FARIMA(0, d , 0) process, $d_H = 0.1$ ($t = 0 \dots 999$), $d_K = 0.3$ ($t = 1000 \dots 1999$) and $d_H = 0.1$ ($t = 2000 \dots 2999$).

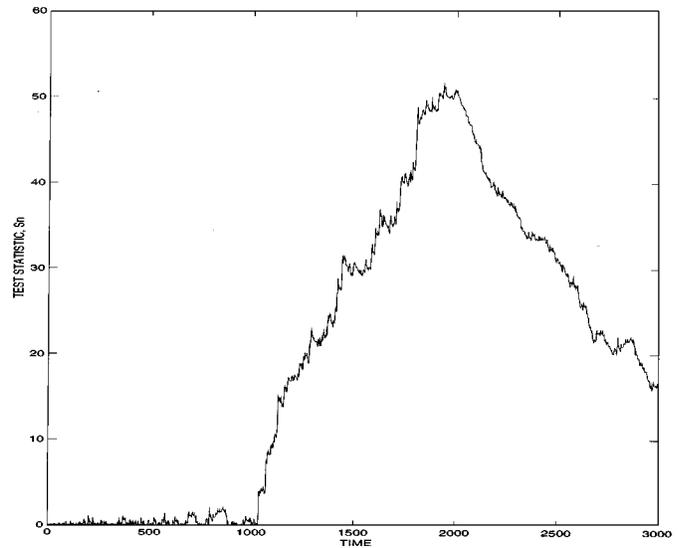


Fig. 5. Change detection for FARIMA(0, d , 1) process, $d_H = 0.1$, $\theta = 0.2$ ($t = 0 \dots 999$); $d_K = 0.3$, $\theta = 0.4$ ($t = 1000 \dots 1999$); $d_H = 0.1$, $\theta = 0.2$ ($t = 2000 \dots 2999$).

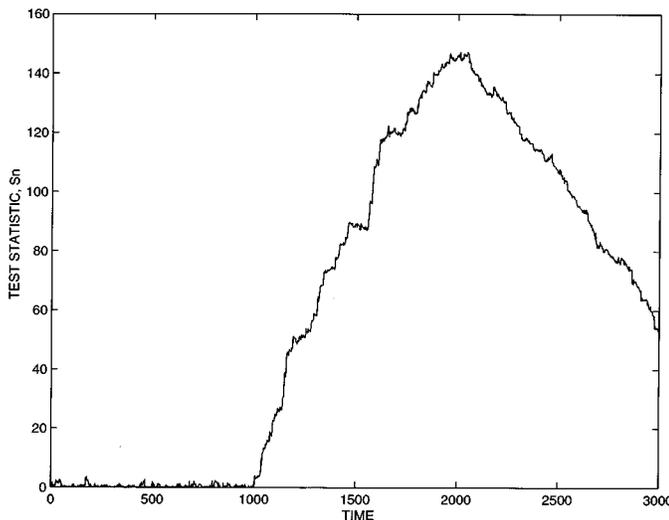


Fig. 4. Change detection for FARIMA(1, d , 0) process, $d_H = 0.1$, $\phi = 0.2$ ($t = 0 \dots 999$); $d_K = 0.3$, $\phi = 0.4$ ($t = 1000 \dots 1999$); $d_H = 0.1$, $\phi = 0.2$ ($t = 2000 \dots 2999$).

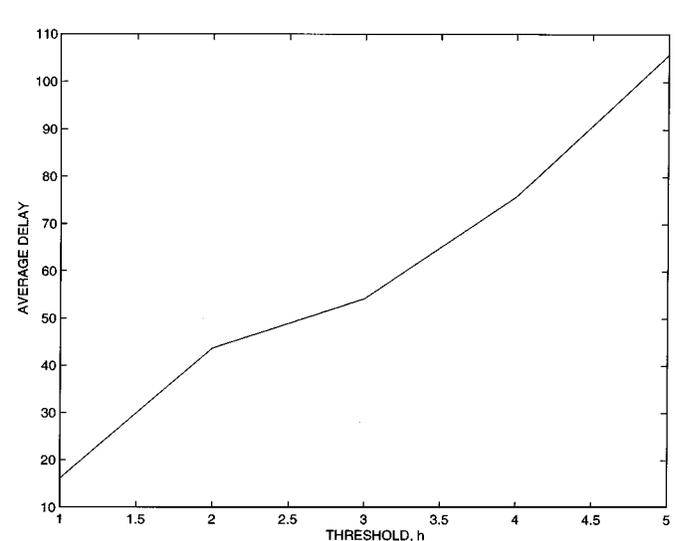


Fig. 6. Delay in detection for FARIMA(0, d , 0) process, $d_H = 0.1$, ($t = 0 \dots 999$); $d_K = 0.3$, $\theta = 0.4$ ($t = 1000 \dots 1999$); $d_H = 0.1$, $\theta = 0.2$ ($t = 2000 \dots 2999$).

Fig. 3, and as expected, the delay is seen to be approximately linear in the detection threshold.

Next, we show some results where the parameter after the change d_K is assumed to be unknown. A practical online approach would be to run several change detection algorithms in parallel with guessed values for d_K . If d_H , which is the parameter before the change, is not known, we can substitute d_H by some estimated value. In this case, the detection algorithm is expected to perform well when the estimation error is small. It is also well known that estimation of parameters of long-memory processes is computationally quite expensive. However, there are estimation schemes using approximate likelihood functions that achieve almost sure convergence to the true value of the parameter (e.g., d_H) at a rate of $n^{-1/2}$ (n being the length of the long memory time series) under certain regularity conditions holding for the spectral density for stationary Gaussian

processes. For a discussion on this and associated robustness issues, see [16]. In the case of unknown parameters before the change, we therefore assume that the estimated values are sufficiently “close” to the true parameter values before they are substituted for λ_H in the proposed change detection algorithm.

In what follows, we only concentrate on the case where d_K is unknown. The following figures (Figs. 7–11) show the plots for $t \geq 1000$ of the test statistic for various “guesses” for d_K for such an algorithm. We investigate the proposed detection algorithm using a filter bank with the FARIMA(0, d , 0) process. The 3000-sample long data sequence was generated using $d_H = 0.1$ for the first 1000 samples, $d_K = 0.3$ for the next 1000, and back to $d_H = 0.1$ in the final section. It is readily seen that for slight changes in assumed d_K , the test statistic is more “jittery” and has no clean transition boundaries at the time of change. More

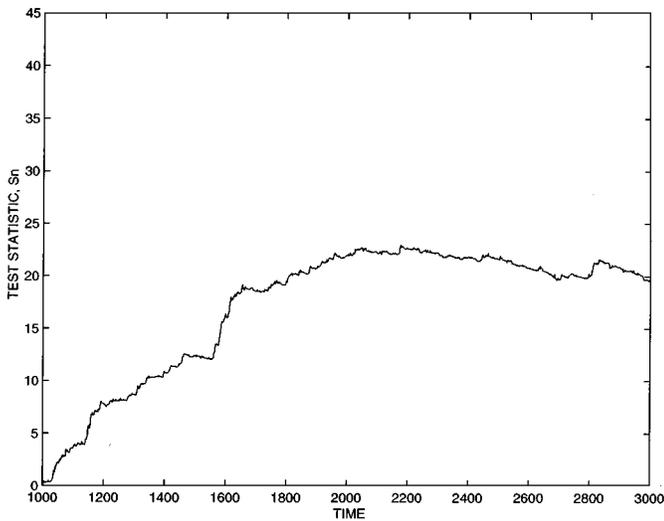


Fig. 7. Change detection for FARIMA(0, d , 0) process. Tested $d_K = 0.20$. Actual $d_K = 0.3$.

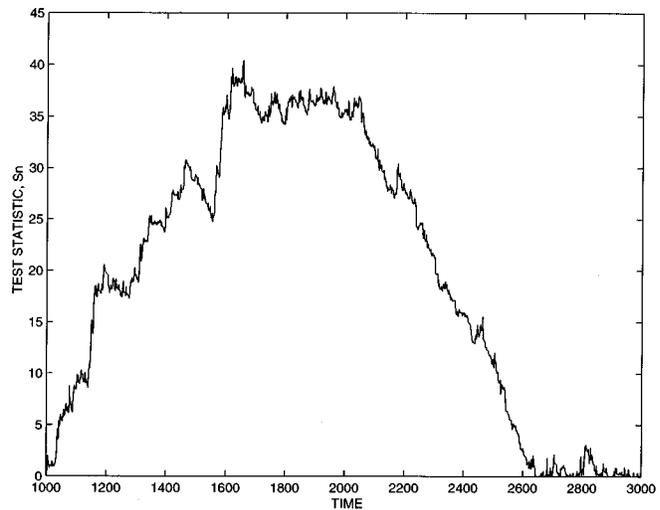


Fig. 10. Change detection for FARIMA(0, d , 0) process, Tested $d_K = 0.35$, Actual $d_K = 0.3$.

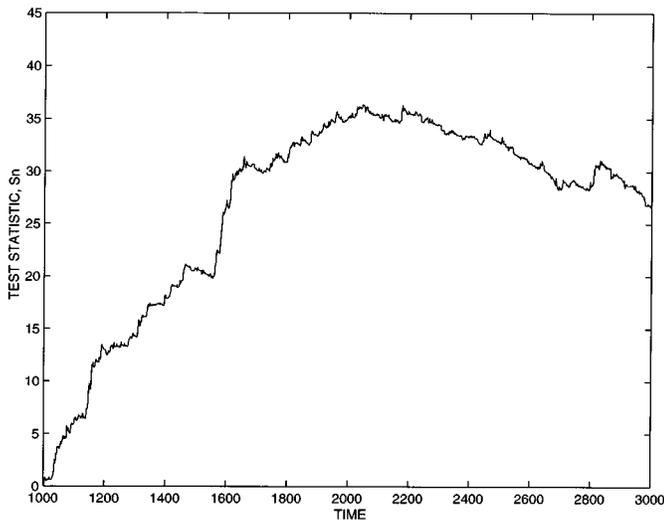


Fig. 8. Change detection for FARIMA(0, d , 0) process. Tested $d_K = 0.25$. Actual $d_K = 0.3$.

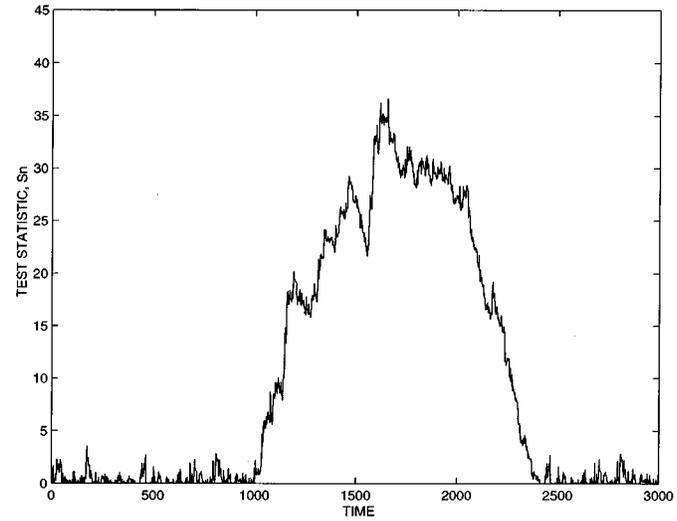


Fig. 11. Change detection for FARIMA(0, d , 0) process. Tested $d_K = 0.40$. Actual $d_K = 0.3$.

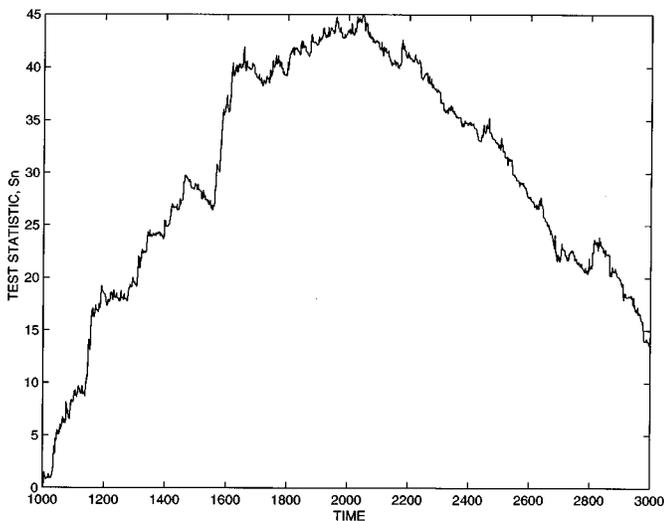


Fig. 9. Change detection for FARIMA(0, d , 0) process, Tested $d_K = 0.3$, Actual $d_K = 0.3$.

importantly, the average rate of change in the test statistic decreases for a branch assuming an incorrect d_K that is farther away from the true d_K . Fig. 9 shows the change for the correctly assumed d_K value (i.e., $d_K = 0.3$). Note that for the other branches, S_n plateaus at a lesser value. Therefore, a higher slope of increase in the values of the test-statistic and cleaner transitions with less jitter are indicative of a better guess. However, these are only empirical guidelines, and the complexity of this algorithm obviously increases exponentially with the number of parameters.

V. CONCLUSIONS

In this paper, we investigated the performance of a CUSUM-like algorithm for detecting changes in traditional teletraffic models such as MMPP's and models proposed more recently, such as long-memory processes, in particular, Gaussian FARIMA processes. Simulations illustrate that the detection algorithms perform well when the parameters before

and after the change are known. In the (realistic) case of unknown parameters after the change, our proposed scheme with a bank of parallel change detection algorithms shows high sensitivity to small variations in the parameters (after the change), thus indicating the potential of such a scheme when the number of parameters involved is small. Simulation results are also presented for mean delays in detection for these schemes. Research is currently underway for developing change detection algorithms for long-memory time series models with heavy-tailed non-Gaussian innovations, which are believed to be responsible for the self-similar nature in ATM and web traffic.

APPENDIX

Proof of (10): Define $\alpha_t^l(j) = (f_{\lambda_t}(n_t, n_{t-1}, \dots, n_0, s_t = j) / f_{\lambda_t}(n_{t-1}, \dots, n_0))$, $l = H, K$, $\forall t \geq 1$. In addition, define $\alpha_0^l = B(n_0)^l \pi_t$. It follows that for $t = 1$

$$\alpha_1^l(j) = \frac{f_{\lambda_t}(n_1, n_0, s_1 = j)}{f_{\lambda_t}(n_0)}. \quad (17)$$

Now

$$\begin{aligned} f_{\lambda_t}(n_1, n_0, s_1 = j) &= f_{\lambda_t}(n_1 | s_1 = j) f_{\lambda_t}(s_1 = j, n_0) \\ &= b_j^l(n_1) \sum_{i=1}^N f_{\lambda_t}(s_1 = j, s_0 = i, n_0) \\ &= b_j^l(n_1) \sum_{i=1}^N a_{ji}^l f_{\lambda_t}(n_0 | s_0 = i) P(s_0 = i) \\ &= b_j^l(n_1) \sum_{i=1}^N a_{ji}^l b_i^l(n_0) \pi_i \\ &= b_j^l(n_1) \sum_{i=1}^N a_{ji}^l \alpha_0^l(i). \end{aligned} \quad (18)$$

In addition, $f_{\lambda_t}(n_0) = \sum_i f_{\lambda_t}(n_0, s_0 = i) = \sum_i \alpha_0^l(i)$. Substituting this and (18) back in (17) and writing in matrix notation, we get

$$\alpha_1^l = B(n_1)^l A^l \frac{\alpha_0^l}{\sum_i \alpha_0^l(i)}. \quad (19)$$

Proceeding in a similar fashion, we can show that for any $t > 1$

$$\begin{aligned} f_{\lambda_t}(n_t, n_{t-1}, \dots, n_0, s_t = j) &= b_j^l(n_t) f_{\lambda_t}(s_t = j, n_{t-1}, \dots, n_0) \\ &= b_j^l(n_t) \sum_{i=1}^N a_{ji}^l f_{\lambda_t}(n_{t-1}, \dots, n_0, s_{t-1} = i). \end{aligned} \quad (20)$$

Here, $b_j^l(n_t) = f_{\lambda_t}(n_t | s_t = j)$. Now, since by definition $\alpha_{t-1}^l(i) = (f_{\lambda_t}(n_{t-1}, \dots, n_0, s_{t-1} = i) / f_{\lambda_t}(n_{t-2}, \dots, n_0))$, we can use (20) to obtain

$$\alpha_t^l(j) = b_j^l(n_t) \sum_i a_{ji}^l \frac{\alpha_{t-1}^l(i)}{\sum_i \alpha_{t-1}^l(i)}. \quad (21)$$

Writing it in the matrix notation, we get (10) for $t > 1$.

1) *Simulation Details:* Here, we provide the details of A^H, A^K, μ^H, μ^K for the simulations done to detect changes in a MMPP.

$$\begin{aligned} A^H &= \begin{bmatrix} 0.8 & 0.07 & 0.05 & 0.001 \\ 0.15 & 0.75 & 0.14 & 0.089 \\ 0.04 & 0.12 & 0.8 & 0.11 \\ 0.01 & 0.06 & 0.1 & 0.8 \end{bmatrix} \\ A^K &= \begin{bmatrix} 0.4 & 0.27 & 0.35 & 0.111 \\ 0.25 & 0.45 & 0.14 & 0.119 \\ 0.15 & 0.22 & 0.4 & 0.23 \\ 0.2 & 0.06 & 0.11 & 0.54 \end{bmatrix} \\ \mu^H &= [1 \ 2 \ 5 \ 8], \mu^K = [3 \ 6 \ 12 \ 15]. \end{aligned}$$

2) *Covariance Formulae for FARIMA(1, d, 0) and FARIMA(0, d, 1):* The covariance γ_k^y of the FARIMA(1, d, 0) process $\{y_t\}$ given by (7) is shown in (22) at the bottom of the page, where γ_k^x is given by (3), and $F(a, b; c; z)$ is the hypergeometric function given by [17]

$$F(a, b; c; z) = 1 + \frac{ab}{c \cdot 1} z + \frac{a(a+1)b(b+1)}{c(c+1) \cdot 1 \cdot 2} z^2 + \dots \quad (23)$$

Similarly, the covariance γ_k^y of the FARIMA(0, d, 1) process $\{y_t\}$ given by (7) is

$$\gamma_k^y = \gamma_k^x \frac{(1 - \theta^2)k^2 - (1 - d)\{(1 - d)(1 + \theta^2) - 2\theta d\}}{k^2 - (1 - d)^2}. \quad (24)$$

REFERENCES

- [1] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, pp. 100–115, 1954.
- [2] M. Basseville and L. V. Nikiforov, *Detection of Abrupt Changes*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [3] B. Chen and P. Willett, "Quickest detection of hidden Markov models, in," in *Proc. 36th IEEE CDC*, San Diego, CA, 1997, pp. 3984–3989.
- [4] S. Dey and S. I. Marcus, "Change detection in Markov-modulated time series, in," in *Proc. Inform. Decision Contr.*, Adelaide, Australia, Feb. 1999, pp. 21–24.
- [5] D. Siegmund, *Sequential Analysis—Tests and Confidence Intervals*. New York: Springer-Verlag, 1995.
- [6] S. W. Roberts, "A comparison of some control chart procedures," *Technomet.*, vol. 8, pp. 411–430, 1966.
- [7] W. Fischer and K. Meier-Hellstern, "The Markov-modulated poisson process," *Perf. Eval.*, vol. 18, pp. 149–171, Feb. 1992.

$$\gamma_k^y = \gamma_k^x \frac{\{F(1, d+k; 1-d+k; \phi) + F(1, d-k; 1-d-k; \phi) - 1\}}{(1 - \phi^2)} \quad (22)$$

- [8] W. E. Leland *et al.*, "On the self-similar nature of ethernet traffic," *IEEE/ACM Trans. Networking*, pp. 1–15, Feb. 1994.
- [9] Z. Sahinoglu and S. Tekinay, "On multimedia networks: Self-similar traffic and network performance," *IEEE Commun. Mag.*, vol. 37, pp. 48–52, Jan. 1999.
- [10] U. Appel and A. V. Brandt, "A comparative study of three sequential time series segmentation algorithms," *Signal Process.*, vol. 6, pp. 45–60, 1984.
- [11] I. V. Basawa, "Neyman-Le Cam tests based on estimating functions," in *Proc. Berkeley Conf. Honor Neyman and Keifer*, vol. 2, 1985, pp. 811–825.
- [12] —, *Generalized Score Tests Composite Hypothesis*, pp. 201–225, 1991.
- [13] M. Basseville, "On-board component fault detection and isolation using the statistical local approach," *Automatica*, vol. 34, no. 11, pp. 1391–1416, 1998.
- [14] Q. Zhang, M. Basseville, and A. Benveniste, "Early warning of slight changes in systems and plants with application to condition based maintenance," *Automatica*, vol. 30, no. 1, pp. 95–114, 1994.
- [15] G. E. P. Box and G. M. Jenkins, *Time Series Analysis Forecasting and Control*, 2nd ed. San Francisco, CA: Holden-Day, 1970.
- [16] J. Beran, *Statistics for Long-Memory Processes*. New York: Chapman & Hall, 1994.
- [17] J. R. M. Hosking, "Fractional differencing," *Biometrika*, vol. 68, pp. 165–176, 1981.
- [18] P. Whittle, "Estimation and information in stationary time series," *Ark. Mat.*, vol. 2, pp. 423–434, 1953.
- [19] Y. Yajima, "On estimation of long memory time series models," *Austral. J. Stat.*, vol. 27, pp. 303–320, 1985.
- [20] J. Durbin, "The fitting of time-series model," *Int. Statist. Rev.*, vol. 28, pp. 233–344, 1961.
- [21] V. V. Godambe, Ed., *Estimating Functions*. Oxford, U.K.: Clarendon, 1991.



Rittwik Jana received the B.E. degree in electrical and electronic engineering from the University of Adelaide, Adelaide, Australia, in 1994. He is currently pursuing the Ph.D. degree at the Australian National University, Canberra.

From 1996 to 1999, he held several positions at the Defense Science and Technology Organization, Canberra. His current research interests are in multiuser detection for cellular CDMA, adaptive signal processing, and protocol design for wireless voice and high-speed data systems.



Subhrakanti Dey was born in Calcutta, India, in 1968. He received the B.Tech. and M.Tech. degrees from the Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology, Kharagpur, in 1991 and 1993, respectively. He received the Ph.D. degree from the Department of Systems Engineering, Research School of Information Sciences and Engineering, Australian National University, Canberra, in 1996.

He is currently a Research Fellow with the Australian National University within the Department of Systems Engineering, where he has held a similar appointment from September 1995 to September 1997. From September 1997 to September 1998, he was a Postdoctoral Research Associate with the Institute for Systems Research, University of Maryland, College Park. His current research interests include hybrid systems, stochastic discrete event systems, performance analysis of communication networks, pattern classification, stochastic and adaptive estimation, and control and statistical and adaptive signal processing.