Modelling Network Traffic Using Time Series Analysis - A Review

Mbulelo Brenwen Ntlangu
Council for Scientific and Industrial Research
Pretoria, Gauteng
bntlangu@csir.co.za

Prof. Alireza Baghai-Wadji
University of Cape Town
Cape Town, Western Cape
alireza.baghai-wadji@uct.ac.za

ABSTRACT

With the advent of Internet of Things (IoT) technology, the need for tools which facillitate the development and management of network based services has become increasingly important. Issues such as network security and quality of service are no longer just the concern of ISPs and big corporations, but have now become the intimate concern of private users with implications that directly affect the personal lives and businesses of citizens. At the heart of addressing these concerns lies the problem of modelling the networks on which interconnected devices now operate. While the activity of provisioning the networks and responding to threats may still lie in the hands of networking specialists, the ability to at least know when something is amiss remains intrumental to establishing the confidence and peace of mind of IoT users.

In this paper we broadly review the historical development of network traffic modelling and trace a path that leads to the use of time series analysis for the said task. A basic introduction to time series analysis is provided in order to facillitate the theoretical discussion regarding the feasibility and suitability of time series analysis techniques for modelling network traffic. The intention is to provide an orientation, for the interested novice, to the domain of time series analysis for network traffic modelling; and to advocate the necessity and utility of further study in these domains for application to IoT concerns.

CCS CONCEPTS

• **Mathematics of computing** → **Time series analysis**; • **General and reference** → *Surveys and overviews*; • **Information systems** → *Traffic analysis*; • **Security and privacy** → Intrusion detection systems;

KEYWORDS

Internet of things, network traffic modelling, time series analysis, multiplicative ARIMA models.

1 INTRODUCTION

Modelling of network traffic is a notoriously difficult problem. This is primarily due to the ever-increasing complexity of network traffic and the different ways in which a network may be excited by user activity. The ongoing development of new network applications, protocols, and usage profiles further necessitate the need for models which are able to adapt to the specific networks in which they are deployed. These considerations have in large part driven the evolution of statistical profiles of network traffic from simple Poisson processes to non-Gaussian models that incorporate traffic burstiness, non-stationarity, self-similarity, long-range dependence (LRD) and multi-fractality. The need for ever more sophisticated network traffic models has led to the specification of a myriad of traffic models since. Many of these are listed in [1, 2]. Much of the complexity of network traffic can be attributed to the unpredictable ways in which humans influencing the network traffic behave. This becomes less of a factor in networks comprised of IoT devices as much of the traffic is generated by devices which function autonomously and in a more deterministic fashion. This then suggests that the activity of building models for IoT network traffic might prove to be substantially fruitful.

2 HISTORICAL DEVELOPMENT

Early work on network traffic modelling, based on principles from telephony, proposed the Poisson process as a model [1–3] for the packet arrival process. The inter-arrival times of network packets were considered to be exponentially distributed, with the density function of the model given by ($t > 0$):

$$f(t) = \lambda e^{-\lambda t} \tag{1}$$

The Poisson model is considered suitable when the arrivals are assumed to emanate from a large number of independent Poisson sources. The Poisson distribution is such that the superposition of these Poisson sources gives rise to another Poissson process whose rate is the sum of the rates of the independent Poisson processes. The mean and variance of the Poisson distribution are also given by the rate parameter, $\lambda > 0$ [2]. In time, the emergence of modern high speed networking technologies as well as the plethora of applications and services that came into use meant that the Poisson model was no longer able to fully capture the complexity of network traffic [1]. In the seminal work of Leland et al. (1994), the authors performed an empirical study of ethernet LAN traffic to provide evidence of the hypothesised self-similar, fractal, and scaling behaviour of network traffic. The work gives a mathematical presentation of the properties of self-similar processes as well as two approaches for modelling network traffic exhibiting these

BDIOT2017, December 20–22, 2017, London, United Kingdom

ACM ISBN 978-1-4503-5430-1/17/12…$15.00
https://doi.org/10.1145/3175684.3175725

behaviours. In their discussion regarding the significance and application of self-similarity to network traffic, Leland et al. posit that self similarity presents itself in ethernet traffic due to the absence of a “natural” burst length, where instead burstiness manifests on a wide range of time scales such that “traffic spikes ride on long term ripples, which ride on longer term swells, etc.” [4]. This is what is said to give ethernet traffic its self-similar or “fractal quality” [4].

In their subsequent work, Paxson and Floyd (1995)[3] used traces of wide area network traffic in an empirical study of the error that is produced when using the Poisson process to model TCP, FTP and Telnet sessions, connections and packet arrivals. These authors also concluded that “ethernet traffic is better modelled as a self similar process...”[3]. Willinger et al. (1995) responded to the findings of [4] and [3] by providing a physical explanantion for the self similarity that is observed in modern network traffic. In this work, the authors employed the ON/OFF source description of network traffic, also known as the "packet train model", to show that if each individual ON/OFF source is characterised by the "Noah Effect" (high variability or infinite variance i.e. very long ON or OFF periods occuring with non-negligible probability), then effectively these individual sources have characteristics which manifest on a wide range of time scales. The aggregate effect of these sources is then to produce network traffic that is self-similar and/or Long Range Dependent (LRD) - termed the "Joseph Effect" [5].

Traditional models, that is up until the discovery of self-similarity in network traffic, had always been assumed to be finite variance models; like the exponential and geometric distributions. The result of this assumption is that aggregated traffic was then modelled as having no significant correlations in the long term, which is contrary to what is observed in reality [5]. This self-similar and LRD model of network traffic is seen to have provided a significant advancement to our understanding of network traffic dynamics [6]. Infinite variance or long range dependent processes are typically modelled using the class of heavy tailed distributions. These distributions, which include some Pareto and stable distributions, are usually parameterised by a heaviness-of-the-tail parameter α. Alpha is also related to the Hurst parameter, which represents a measure of the degree of self-similarity in a signal [4, 5].

In the works of Simross-Wattenburg et al. [7, 8] the alpha-stable distribution was used to model the marginals of binned network packet counts. Based on these marginals a generalised likelihood ratio test (GLRT) was applied towards classifying the observed traffic into normal and anomalous traffic patterns and hence the detection of flash crowds and denial of services attacks. By modelling network traffic according to the alpha-stable family of distributions, which are statistical distributions with heavy tails, the authors posit that they are able to account for the high variability which manifests from the bursty nature of network traffic. The use of the alpha-stable family of distributions is further motivated by the fact that they are the limiting distribution in the Generalised Central Limit Theorem (GCLT) which states that the sum of random variables, which are distributed according to heavy tailed distributions i.e. distributions with tails that decay as: $|x|^{-\alpha-1}$, where $0 < \alpha < 2$, tend to be distributed according to a stable distribution as the number of random variables becomes large. The Gaussian distribution is a special case where $\alpha = 2$.

Following the discovery of the fractal nature of network traffic, a popular approach became to use fractional Brownian motion (fBm) or rather its increment process, fractional Gaussian noise (fGn) as a means of modelling network traffic[1, 9, 10]. In addition to the long term characteristics of network traffic, Riedi and Willinger [6] studied the small time scaling behaviour of network traffic. This lead to the proposal of multifractality, which was attributed to the network protocol mechanisms. See [11–13] for some of the multifractal models applied to network traffic. The case of multifractality in network traffic was however disputed in [14], where the authors cite the multifractal behaviour that had been observed as “...a misinterpretation due to a lack of power in the statistical methodology[14]”.

In the work of Scherrer et al. (2007) [15], network traffic is modelled using a Gamma distribution to fit the marginals and an ARFIMA process to fit the temporal characteristics of the data. The Poisson distribution is said to represent one extreme where the data is not very aggregated, while the Gaussian distribution emerges for highly aggregated data, as suggested by the Central Limit Theorem (CLT). Gamma distributions are chosen in [15] because they are able to describe data that has a distribution which lies within the transitionary area between the Poisson and Gaussian distributions. This, in turn, allows them to provide a description of network traffic that is applicable on a wider range of scales. The ARFIMA process introduced to the model is particularly well suited to describing the covariance structure of the data, including its short and long range dependence characteristics. The ARFIMA process, in fact, represents a whole family of fractionally integrated processes, which [15] chooses to restrict to the subclass of ARFIMA $(1, d, 1)$ processes, where d is the fractional order of differencing determined from the data. In the next section the basic precepts of time series modelling are introduced, which will lead us into a discussion of existing implementations of time series analysis based models for network traffic.

3 THEORY

3.1 Stochastic Processes

A stochastic process may be considered to be a phenomenon or system which generates values that are non-deterministic. Typically these give rise to time series, which are sequential observations of the generating process taken at equidistant points in time. Denoting the N successive time instances when observations of the process are made $t = 1, 2, ..., N$, we obtain the time series $\{z(t)\} = \{z(1), z(2), ..., z(N)\} = \{z_1, z_2, ..., z_N\}$. It is noted that this set of observations indexed by time (a time series) is but a single realisation of the infinitely many that could have been realised from the generating stochastic process [16]. Each observation in the time series is a specific manifestation of a single random variable drawn from a population of possible values that are distributed according to some probability density function $f(z_t)$. In the aforementioned time series, z_t, we thus have N random variables distributed according to N probability density functions $f_1(z_1), f_2(z_2), ..., f_N(z_N)$. The set of N observations $z_1, z_2, ..., z_N$ are what is referred to as a sample realisation.

3.2 Deriving the Moments of the Joint Distribution

Each pair of random variables z_k, z_{k+j} are subject to a joint distribution of the two random variables $f_{k,k+j}(z_k, z_{k+j})$. By extension, the entire set of N random variables is subject to a joint distribution $f_{1,2,...,N}(z_1, z_2, ..., z_N)$. The first step in attempting to model an unknown stochastic process is to study the characteristics of a realisation of that process. The most basic way of characterising a time series is by virtue of its mean values. Considering the random variables z_t to be distributed according to a multivariate probability density function, $f_{1,2,...,N}(z_1, z_2, ..., z_N)$, we obtain the following expressions for the moments of the distribution about the origin:

$$\mu_{n_1,n_2,...,n_N} = E[Z_1^{n_1} Z_2^{n_2} ... Z_N^{n_N}] \tag{2}$$

where $n_1, n_2, ..., n_N$ denote the n^{th} order moment for the random variables $Z_1, Z_2, ..., Z_N$ respectively, and can take on any integer values $1, 2, 3, ...$ The discrete case, amounts to

$$\sum_{i_1} \sum_{i_2} ... \sum_{i_N} z_1^{n_1} z_2^{n_2} ... z_N^{n_N} f_{1,2,...,N}(z_1, z_2, ..., z_N) \tag{3}$$

If we set $n_1 = 1$, which corresponds to the first moment of the random variable Z_1, and set the rest of the moment indices to zero, $\mu_{1,0,0,...}$ we obtain the mean of Z_1, $E[Z_1] = \mu_{Z_1}$ Likewise for $\mu_{0,1,0,...}$, we obtain the mean of Z_2, $E[Z_2] = \mu_{Z_2}$, for $\mu_{0,0,1,...}$, we obtain the mean of Z_3, $E[Z_3] = \mu_{Z_3}$ and so on [17]. This gives us the equivalent of determining the first moment of the individual marginal distribution of each random variable independently. The familiar integral form of this operation is given by:

$$E[x] = \int_{-\infty}^{\infty} x f(x) dx \tag{4}$$

where $f(x)$ is the probability function of the random variable x.

3.3 Covariance and Autocorrelation

A characteristic of a time series that is readily computed and forms the basis of subsequent analysis is the covariance function. The covariance measures the linear association between variates X and Y. When the association between variates from the same stochastic process is being considered, this covariance is referred to as the autocovariance. To simplify the discussion, we consider the linear association between only two variates Z_t and $Z_{t+\tau}$, where t is the time index of the variate and τ is any number of time lags $1, 2, ...$ Consequently, we need only deal with the $(a, b)^{th}$ moments of the bivariate distribution $f_{t,t+\tau}(Z_t, Z_{t+\tau})$. The autocovariance is then obtained by setting $a = b = 1$ where a and b are now the indices of the moments of Z_t and $Z_{t+\tau}$ respectively. The autocovariance is then given by

$$\mu_{1,1} = E[(Z_t - \mu_{Z_t})(Z_{t+\tau} - \mu_{Z_{t+\tau}})] \tag{5}$$

which, in the discrete case, amounts to

$$\mu_{1,1} = \sum_i \sum_j (z_i - \mu_t)(z_j - \mu_{t+\tau}) f_{t,t+\tau}(Z_t, Z_{t+\tau}) = \gamma_{t,t+\tau} \tag{6}$$

The (auto)correlation between the variates Z_t and $Z_{t+\tau}$ is obtained by normalising the (auto)covariance using the product of the variances $\sigma^2_{Z_t}$ and $\sigma^2_{Z_{t+\tau}}$ as follows:

$$\rho_{t,t+\tau} = \frac{\gamma_{t,t+\tau}}{\sqrt{\sigma^2_{Z_t} \sigma^2_{Z_{t+\tau}}}} \tag{7}$$

If the autocovariance between all variates which are separated by the same number of lags is the same, regardless of when they occur in time, then the stochastic process is said to be covariance stationary. This is to say that the autocovariances $\gamma_{t,t+\tau}$ are not dependent on the time index t, but rather on the number of lags, τ, separating the random variables. Consequently, we denote the autocovariance between all variates of a stationary stochastic process separated by τ lags γ_τ and the corresponding autocorrelations ρ_τ.

3.4 The Linear Filter Transfer Function

Time series modelling may be thought of as defining operations for the purpose of transforming a highly dependent, and possibly non-stationary process $\{Z_t\}$, into an uncorrelated white noise process. Conversely, time series modelling may be equivalently considered as the act of determining the linear filter which, when operating on independent random shocks, produces the process $\{Z_t\}$. The random shocks $a_t, a_{t-1}, a_{t-2}, ...$ are typically considered to be independent random variables that are normally distributed with mean equal to zero and some fixed variance σ^2_a. Usually one does not observe these shocks directly, but rather observes the realisations of the process itself, $\{z_t\}$. The linear filter ψ simply produces a weighted sum of previous shocks as follows:

$$z_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2}, ... \tag{8}$$

where μ represents the level of the process. In practise the filter ψ must be determined from the observations of the process,$\{z_t\}$. The filter may also be considered to be constituted of the ratio of two polynomials $\phi(B)$ and $\theta(B)$

$$\psi(B) = \frac{\theta(B)}{\phi(B)} \tag{9}$$

$$\psi(B) = \frac{1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q}{1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p} \tag{10}$$

where B is the backshift operator. The determination of these two polynomials is in fact one of the primary concerns of time series analysis. $\psi(B)$ is called the transfer function of the linear filter relating z_t to a_t [16] as follows:

$$z_t = \psi(B) a_t \tag{11}$$

Denoting $\psi(B)^{-1}$ as $\pi(B)$, we obtain the transfer function of the linear filter relating a_t to z_t [16] as follows:

$$\begin{aligned} a_t &= \psi(B)^{-1} z_t \\ &= \pi(B) z_t \end{aligned} \tag{12}$$

3.5 The Autoregressive AR(p) Model

An autoregressive (AR) process is one in which the current value of a process, z_t, is represented as a weighted sum of previous values of the process and the current random shock. This is essentially a regression of the process against itself:

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \ldots + \phi_p z_{t-p} + a_t \tag{13}$$

If we collect all the terms involving the values of the process on the left hand side and factor out the polynomial in B, $\phi(B)$, we obtain

$$(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p) z_t = a_t \tag{14}$$

which can be written economically as $\phi(B)z_t = a_t$. The equivalent representation in terms of $\psi(B)$ is then

$$z_t = \frac{a_t}{(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)} = \psi(B) a_t \tag{15}$$

An AR process is defined by a set of p weights, which are the coefficients of the polynomial in B, $\phi(B)$. The AR process of order p is denoted $AR(p)$. We note in particular that the polynomial $\phi(B)$ is finite in order and hence in the extent of its weights, as opposed to the corresponding transfer function $\psi(B) = \phi(B)^{-1} = \sum_{j=0}^{\infty} \psi_j B^j$, which is infinite in its extent. In order for the AR process to be stationary, it is required that the weights ψ_j of the transfer function $\psi(B)$ form a convergent series. This imposes the restriction on AR polynomial $\phi(B)$ - that the roots of the characteristic equation $(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p) = 0$ should lie outside the unit circle.

3.6 The Moving-Average $MA(q)$ Model

A moving-average (MA) process is one in which the current value of a process, z_t, is represented as a weighted sum of previous innovations or random shocks a_t that are input to the process. This amounts essentially to an averaging of the most recent shocks.

$$z_t = \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q} + a_t \tag{16}$$

If we factor out the polynomial in B, $\theta(B)$, we obtain

$$z_t = (1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q) a_t \tag{17}$$

which can be written economically as $z_t = \theta(B)a_t$. The equivalent representation in terms of $\psi(B)$ becomes

$$z_t = (1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q) a_t = \psi(B) a_t \tag{18}$$

As seen above, a MA process is defined by a set of q weights, which are the coefficients of the polynomial in B, $\theta(B)$. The MA process of order q is denoted MA(q). We note in particular that the polynomial $\theta(B)$ is finite in order and hence in the extent of its weights. Naturally then, since $\psi(B) = \theta(B)$, the MA's corresponding transfer function, $\psi(B)$, is also finite in its extent. The corresponding inverse transfer function $\pi(B) = \psi(B)^{-1}$ by contrast is infinite in its extent. We see that in order for the MA process to be stationary, there is no requirement on the transfer function $\psi(B)$ as it naturally forms a convergent series for MA(q) processes. If we however wish to express the MA process as an AR(∞) process by simply inverting the MA operator $\theta(B)$, and hence have the MA process be invertible [18],

$$\begin{aligned} a_t &= \frac{z_t}{(1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q)} \\ &= (1 + \pi_1 B + \pi_2 B^2 + \pi_3 B^3 + \ldots) z_t \\ &= \pi(B) z_t \end{aligned} \tag{19}$$

it is required that the weights of the inverse transfer function $\pi(B)$ be convergent on or within the unit circle. This imposes a restriction on the MA polynomial $\theta(B)$ - that the roots of the characteristic equation $(1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q) = 0$ lie outside the unit circle.

3.7 The Auto-Regressive Moving-Average ARMA(p, q) Model

Due to the fact that in practice a given process may not be parsimoniously represented as a purely MA or AR process, it may be necessary to incorporate both AR and MA components into the expansion of the process as follows:

$$z_t = \phi_1 z_{t-1} + \ldots + \phi_p z_{t-p} - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q} + a_t \tag{20}$$

which, if we collect all the terms involving the process values on the left hand side, may be written as

$$\begin{aligned} -\phi_1 z_{t-1} - \ldots - \phi_p z_{t-p} + z_t &= -\theta_1 a_{t-1} - \ldots - \theta_q a_{t-q} + a_t \\ (1 - \phi_1 B - \ldots - \phi_p B^p) z_t &= (1 - \theta_1 B - \ldots - \theta_q B^q) a_t \\ \phi(B) z_t &= \theta(B) a_t \\ z_t &= \phi(B)^{-1} \theta(B) a_t \end{aligned} \tag{21}$$

The mixed autoregressive moving-average (ARMA) process with order p AR polynomial $\phi(B)$ and order q MA polynomial $\theta(B)$ is denoted ARMA(p, q). The transfer function of the ARMA(p, q) process $\psi(B)$ relates the process values z_t to the random shocks a_t according to $z_t = \psi(B)a_t$, thus we obtain

$$\begin{aligned} \psi(B) &= \phi(B)^{-1} \theta(B) \\ &= \frac{1 - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q}}{1 - \phi_1 z_{t-1} - \ldots - \phi_p z_{t-p}} \end{aligned} \tag{22}$$

3.8 Modelling Non-stationary Time Series

The theoretical framework established thus far provides us with a simple, yet powerful, set of tools to model stationary processes of many varieties. Unfortunately, the assumption of stationarity cannot be made for most of the processes encountered in real world applications, particularly network traffic. The first task, therefore, is to perform some form of preparation of the data in order to make the data ammenable to modelling via our tools. Concretely, our goal is to manipulate the data into a form such that it can be sufficiently described by an ARMA(p, q) model of the form in equation 22. We define a non-stationarity process as one whose mean and covariance change over time. Non-stationarity can arise in numerous different ways. In this work, however, we will restrict the discussion to three of the most common sources of non-stationarity, namely the presence of trend, seasonality and unit roots.

3.8.1 *Trend.* Trend may informally be considered as a long-term change in the mean, where the mean level of the process is a function of time. It may turn out that an observed trend is in fact a cyclic variation that only becomes visible when the process is observed over a very long time period, nevertheless, it remains worthwhile to consider such cyclic components as trends when the wavelength is far greater than the observed time series [19]. Assuming an observed process $\{x_t\}$ is comprised of a linear trend added to white gaussian noise ϵ_t, $\{x_t\}$ may be formulated as

$$\begin{aligned} x_t &= m_t + \epsilon_t \\ &= \alpha + \beta t + \epsilon_t \end{aligned} \tag{23}$$

where mean level of the process is then given by $m_t = \alpha + \beta t$. This may be referred to as the "trend term". Ideally the trend term would be a deterministic function of time describing the global trend of the process. However, in reality, the scope of the trend term may be reduced to describing only local trends; and consequently the parameters α and β may be required to vary over time. Trend may be modelled with the view to either use the model to remove the trend from the data or to use the trend model for forcasting. Alternatively, a transformation may be applied to the data to simply eliminate the trend. A general method of modelling trend is to fit polynomials to the data. For example, one might fit a parametric family of curves to the data of the form

$$m_t = a_0 + a_1 t + a_2 t^2 \tag{24}$$

Coefficients a_0, a_1, and a_2 can then be obtained by a least squares estimation procedure where $\sum_t (x_t - m_t)^2$ is minimised [20]. Another way of handling trend is to consider it as the task of defining a linear filter whose output has all the long term variation removed i.e. a high pass filter [19]. This typically takes the form of a smoothing operation on the data

$$S(x_t) = \sum_{\tau=-m}^{m} a_\tau x_{t+\tau} \tag{25}$$

where $\{a_\tau\}$ is a set of weights (or filter coefficients), and $S(x_t)$ represents the smoothing operation applied to x_t. The smoothed time series can then be used for forecasting or subtracted from the original time series to obtain residuals, where the long term trend is now no longer present.

$$Res(x_t) = x_t - S(x_t) \tag{26}$$

A non-stationary time series that can be made stationary via the removal of trend is called stationary about a trend or trend-stationary.

3.8.2 Unit Roots. In the case of units roots, a simple differencing, equivalent to the application of the operator $\nabla(= 1 - B)$ to the data, is able to bring the process to stationarity [16], where B is the backshift operator. We define a general unit root, also called a difference stationary process, in terms of the transfer function $\psi(B)$ as

$$(1 - B)^d z_t = \delta + \psi(B) a_t \tag{27}$$

where δ is the mean of the stationary process $(1 - B)^d z_t$. An initial plot of the autocorrelation of the data may assist in identifying unit root processes. Unit root processes are typically characterised by an (partial)autocorrelation function that decays slowly towards zero.

3.8.3 Seasonality. Seasonality is a phenomenon that arises when the observations of a process typically manifest patterns which are cyclic. This can often be seen in data that is collected periodically (e.g. daily, weekly etc.). When dealing with non-stationary data that is seasonal, it is important to note that the analysis of the data must be considered at two (or more) timescales; namely from observation to observation and from season to season. The observation to observation timescale may be defined as the "one lag" timescale. The season to season timescale may be defined as the "s lag" timescale, where s is the suspected period of the seasonal component in the data. The definition of the differencing operator defined above can be extended to define the seasonal differencing operator as $\nabla_s = (1 - B^s)$ where ∇_s denotes the fact that the difference is now computed between observations separated by s time lags. This operation can in turn be interpreted as dealing with the non-stationarity at the s lag timescale. Consequently, the AR and MA polynomials used to model the seasonal component are defined in terms of $\tilde{B} = B^s$ as opposed to B.

$$\begin{gathered} \Phi(\tilde{B}) z_t = \Theta(\tilde{B}) u_t \\ (1 - \Phi_1 B^s - \Phi_2 B^{2s} - \ldots - \Phi_P B^{Ps}) z_t = \\ (1 - \Theta_1 B^s - \Theta_2 B^{2s} - \ldots - \Theta_Q B^{Qs}) u_t \end{gathered} \tag{28}$$

where $\{u_t\}$ is an input process which may not necessarily be uncorrelated. The effect of defining the ARMA polynomials in terms of B^s is seen in equation 28, where the coefficients of $\Phi(\tilde{B})$ and $\Theta(\tilde{B})$ are not applied to adjacent observations as in $(1 + B + B^2 + \ldots) z_t$, but rather to observations separated by s lags as in $(1 + B^s + B^{2s} + \ldots) z_t$, where B is the backshift operator. In this way, the ARMA(P, Q) model specified by $\Phi(\tilde{B})$ and $\Theta(\tilde{B})$ captures the correlation between every s^{th} observation of the process.

3.9 Model Building

In attempting to build models that describe the data that we observe from network traffic, the following basic steps, as advocated by Box and Jenkins, are usually followed:

- Prepare the data such that it is ammenable to analysis under the assumption of covariance stationarity [18].
- Obtain estimates of the parameters p and q, which are the orders of the autoregressive(AR) and moving-average(MA) polynomials respectively for an ARMA(p, q) model [16, 18].
- Estimate the parameters (that is the $p + q$ coefficients) of the autoregressive and moving-average polynomials $\phi(B)$ and $\theta(B)$ respectively that form part of a general ARMA model.
- Perform diagnostic analysis to assess the model's ability to reproduce the characteristics observed in the original data.

4 CURRENT IMPLEMENTATIONS OF TIME SERIES ANALYSIS MODELS

In this section we look at some of the successful applications of time series analysis to network traffic modelling observing the areas where further work is warranted.

In the gamma-ARFIMA model proposed in [15], the choice of the ARFIMA$(1, d, 1)$ subclass of processes is somewhat arbitrary. In general, we cannot be certain upfront about what the order parameters, p and q of an ARIMA(p, d, q) process that is modelling a given network traffic stream, will be. Traditionally, the approach proposed by [16] is to examine plotted graphs of the data along with its autocorrelations and partial autocorrelations in order to get an intuition of what these parameters might be. While this is always a valuable thing to do, as preliminary analysis, it is ad hoc and subjective for the purpose of selecting a model [21].

In [22] a seasonal ARIMA model is used to model wireless GSM traffic. The authors proceed by performing a spectral analysis of the data in order to identify cyclical patterns in the data. This is used to compose a traditional multiplicative seasonal ARIMA model. The authors extend this framework by providing an expression which allows for two periodicities to be incorporated into the model. The proposed procedure then proceeds along the same lines as method 2 of [23] with the additional assumption that the order of the fitted models can be limited to the range $[0, 2]$.

Procedures for automatically determining the parameters p and q are proposed in [23]. Therein the authors propose, firstly, exploring every possible autoregressive, AR(p), representation of the data choosing the combination of p, d and s that results in the lowest value of the Aikake Information Criterion (AIC). Secondly, a grid search of all possible combinations of p and q may be conducted, where once again, the minimisation of the AIC is used as the objective criterion. This can be extended to include candidate values of s and d. This is however somewhat of a brute force approach, which may be expensive in terms of memory and computation.

In [21], the instability of model selection is discussed and a number of metrics to quantify the stability of model selection are provided. In the event of having multiple candidate models for a prediction task, instead of attempting to find the "true model", the Aggregated Forecast Through Exponential Re-weighting(AFTER) algorithm is proposed. This entails running multiple competing models at the same time and weighting their output according to their performance in previous time instances. In this way the stability and accuracy of the resulting forecasts is improved. The success of this scheme is of course still dependent on having at least identified suitable candidate models to combine.

Laner et al. in [24, 25] motivate an approach for modelling and simulating network traffic based on three sequential tranformations of a zero mean, unit variance, i.i.d, Gaussian noise process. These transformations act to modulate the cumulative distribution, autocorrelation and cross-correlation functions of I independent Gaussian noise processes in order generate output processes that resemble real network traffic. The first tranformation forms a weighted sum of the I input processes in order introduce cross-correlation between them, thus simulating the cross-correlation that occurs between the traffic streams of separate applications running concurrently on a network. The second tranformation performs a linear time invariant (LTI) filtering of each process, in the form of an ARMA transfer function, in order to introduce the desired autocorrelation structure to the samples. The third tranformation is a memoryless polynomial transformation, which shapes the distribution of the output processes. Laner et al. advocate that treating the modelling and simulation problem according to these three separate concerns results in analytically tractable, parsimonious, efficient and low complexity means of generating synthetic network traffic samples. This however remains to be proven.

In [26], Iqbal et al. study the power and performance of online, one-step ahead, traffic predictors. The authors posit that by accurately predicting traffic and identifying periods of idling or low traffic, a given system can be placed into a low-power state thereby using processing resources more efficiently. To this end, predictors from three categories are studied, namely time series analysis based predictors, artificial neural networks(ANNs), and wavelet transform based predictors. The results of the study show that the Double Exponentially Weighted Smoothing (DES) and ARMA models incurred some of the lowest computational overhead among the compared techniques, while also performing the best in terms of prediction accuracy over the Caida(DES was the best), Auckland and Bellcore (ARMA was the best) datasets.

5 DISCUSSION AND FUTURE WORK

There are still some difficulties in objectively identifying appropriate time series analysis models. Much of the model identification in literature still depends on the discretion of the analyst who has to study various plots of the signal and its features. Ideally the entire model building exercise should be completely data-driven and automated. Furthermore the models developed in the networking domain need to be adaptive in order to maintain accuracy under realistic operational conditions.

In their paper Stadnytska et al. (2008) evaluated the Minimum Information Criterion (MINIC), smallest canonical correlation method (SCAN), and the extended sample autocorrelation function(ESACF) procedures for automatic ARMA model identification. The details of these procedures can be found in [27] and the references therein. While the study exposed the ineptitude of these techniques, they still retain utility as evidenced by their inclusion in SAS. This shows that the development of automated model identification algorithms, however crude, provides significant value.

In their paper, Tran and Reed (2004) investigate the application of time series analysis to the prediction of temporal I/O arrival paterns. Of particular value to our discussion are their contributions towards automatic detection of non-stationarity and seasonality for model identification. The authors devise an approach that makes use of the fact that after sufficient differencing, the significant spikes that remain in the autocorrelation and partial autocorrelation functions, if positioned at regular intervals, are indicative of seasonality. The distances (in lags) between successive spikes is computed, taking the distance that occurs most frequently as the period of the suspected seasonal component [28].

In the paper by Huang and Shih (2003) [29], ARMA models are proposed for use in in making short term forecasts of load in power grid systems. The proposed method incorporates the use of higher order cumulants to determine the order of AR and MA processes that model a given time series. The premise of the procedure is that the higher order cumulants (> 2) of a Gaussian process are zero. The non-Gaussian process can thus be distinguished from the Gaussian component [29]. The authors go on to prescribe the use of the bispectrum, which is a 2D Fourier transform of the third order cumulant, to test for Gaussianity. The subsequent model identification then follows two paths. If the stationary time series is found to be Gaussian, then model identification proceeds in the usual fashion via analysis of the ACF and PACF of the series. Alternatively, the model identification is performed via a cumulant-based order determination procedure, the details of which can be found in [30].

6 CONCLUSION

Network traffic has been found to be fractal, displaying high variability at certain scales while also exhibiting correlations over long time scales. Time series analysis specifically tries to describe a process as a linear combination of its previous realisations and random innovations. This has the benefit of capturing explicitly the covariation between sample observations. For network traffic modelling, this means that we are no longer beholden to the assumption of identical and independently distributed samples. We can capture both the short and long range dependence of network traffic data using time series models, potentially avoiding the need to deal with the intricacies of multifractality. Time series analysis also gives us the tools to deal with some of the difficulties encountered in network traffic data such as non-stationarity and seasonality. A survey of the literature in a wide array of fields has found that there are numerous techniques which hold promise for addressing the particular needs of traffic modelling in the IoT domain, and that an effort to consolidate and improve these techniques may provide powerful tools for the management and securing IoT networks.

REFERENCES

[1] Walter Willinger, Murad S Taqqu, and Ashok Erramilli. A bibliographical guide to self-similar traffic and performance modeling for modern high-speed networks. *Stochastic networks: Theory and applications*, pages 339–366, 1996.

[2] Balakrishnan Chandrasekaran. Survey of network traffic models. *Waschington University in St. Louis CSE*, 567, 2009.

[3] Vern Paxson and Sally Floyd. Wide area traffic: the failure of poisson modeling. *IEEE/ACM Transactions on Networking (ToN)*, 3(3):226–244, 1995.

[4] W.E. Leland, M.S. Taqqu, W. Willinger, and D.V. Wilson. On the self-similar nature of ethernet traffic (extended version). *IEEE_J_NET*, 2(1):1–15, February 1994.

[5] Walter Willinger, Murad S. Taqqu, Will E. Leland, and Daniel V. Wilson. Self-similarity in high-speed packet traffic: Analysis and modeling of ethernet traffic measurements. *Stat. Sci.*, 10(1):67–85, February 1995.

[6] Rudolf H Riedi and Walter Willinger. Toward an improved understanding of network traffic dynamics. *Self-similar Network Traffic and Performance Evaluation, eds Park and Willinger, Wiley*, pages 507–530, 2000.

[7] Federico Simmross-Wattenberg, Antonio Tristan-Vega, Pablo Casaseca-de-la Higuera, Juan Ignacio Asensio-Perez, Marcos Martın-Fernandez, Yannis A Dimitriadis, and Carlos Alberola-Lopez. Modelling network traffic as $\alpha-$stable stochastic processes. an approach towards anomaly detection. *Proc. VII Jornadas de Ingenierıa Telematica (JITEL)*, pages 25–32, 2008.

[8] Federico Simmross-Wattenberg, Juan Ignacio Asensio-Perez, Pablo Casaseca de-la Higuera, Marcos Martin-Fernandez, Ioannis A. Dimitriadis, and Carlos Alberola-Lopez. Anomaly detection in network traffic based on statistical inference and α-stable modeling. *IEEE_J_DSC*, 8(4):494–509, July 2011.

[9] Ilkka Norros. On the use of fractional brownian motion in the theory of connectionless networks. *IEEE Journal on selected areas in communications*, 13(6):953–962, 1995.

[10] Jacques Vehel and Rudolf H Riedi. Fractional brownian motion and data traffic modeling: The other end of the spectrum. *Fractals in Engineering*, 1997.

[11] Rudolf Riedi and Jacques Lévy Véhel. *Multifractal properties of TCP traffic: a numerical study*. PhD thesis, INRIA, 1997.

[12] Rudolf H Riedi, Matthew S Crouse, Vinay J Ribeiro, and Richard G Baraniuk. A multifractal wavelet model with application to network traffic. *IEEE Transactions on Information Theory*, 45(3):992–1018, 1999.

[13] Flávio Henrique Teles Vieira, Gabriel Rocon Bianchi, and Luan Ling Lee. A network traffic prediction approach based on multifractal modeling. *Journal of High Speed Networks*, 17(2):83–96, 2010.

[14] Darryl Veitch, Nicolas Hohn, and Patrice Abry. Multifractality in tcp/ip traffic: the case against. *Computer Networks*, 48(3):293–313, 2005.

[15] A. Scherrer, N. Larrieu, P. Owezarski, P. Borgnat, and P. Abry. Non-gaussian and long memory statistical characterizations for internet traffic with anomalies. *IEEE_J_DSC*, 4(1):56–70, January 2007.

[16] George E. P. Box, Gwilym M. Jenkins, and Gregory C. Reinsel. *Time Series Analysis*. Wiley-Blackwell, June 2008.

[17] Maurice Bertram Priestley. *Spectral analysis and time series*. Elsevier, 2004.

[18] James D. Hamilton. *Time Series Analysis*. Princeton University Press, 1994.

[19] Douglas C Montgomery, Cheryl L Jennings, and Murat Kulahci. *Introduction to time series analysis and forecasting*. John Wiley & Sons, 2015.

[20] Peter J. Brockwell and Richard A. Davis. *Introduction to Time Series and Forecasting*. Springer New York, 2002.

[21] Hui Zou and Yuhong Yang. Combining time series models for forecasting. *International journal of Forecasting*, 20(1):69–84, 2004.

[22] Yantai Shu, Minfang Yu, Jiakun Liu, and Oliver WW Yang. Wireless traffic modeling and prediction using seasonal arima models. In *Communications, 2003. ICC'03. IEEE International Conference on*, volume 3, pages 1675–1679. IEEE, 2003.

[23] Asrul H Yaacob, Ian KT Tan, Su Fong Chien, and Hon Khi Tan. Arima based network anomaly detection. In *Communication Software and Networks, 2010. ICCSN'10. Second International Conference on*, pages 205–209. IEEE, 2010.

[24] Markus Laner, Philipp Svoboda, and Markus Rupp. Modeling randomness in network traffic. In *ACM SIGMETRICS Performance Evaluation Review*, volume 40, pages 393–394. ACM, 2012.

[25] Markus Laner, Philipp Svoboda, and Markus Rupp. Parsimonious network traffic modeling by transformed arma models. *IEEE Access*, 2:40–55, 2014.

[26] Muhammad Faisal Iqbal and Lizy K John. Power and performance analysis of network traffic prediction techniques. In *Performance Analysis of Systems and Software (ISPASS), 2012 IEEE International Symposium on*, pages 112–113. IEEE, 2012.

[27] Tetiana Stadnytska, Simone Braun, and Joachim Werner. Comparison of automated procedures for arma model identification. *Behavior research methods*, 40(1):250–262, 2008.

[28] Nancy Tran and Daniel A Reed. Automatic arima time series modeling for adaptive i/o prefetching. *IEEE Transactions on parallel and distributed systems*, 15(4):362–377, 2004.

[29] Shyh-Jier Huang and Kuang-Rong Shih. Short-term load forecasting via arma model identification including non-gaussian process considerations. *IEEE Transactions on power systems*, 18(2):673–679, 2003.

[30] Jerry M Mendel. Tutorial on higher-order statistics (spectra) in signal processing and system theory: Theoretical results and some applications. *Proceedings of the IEEE*, 79(3):278–305, 1991.