Wireless Traffic Modeling and Prediction Using Seasonal ARIMA Models

Article *in* IEICE Transactions on Communications · January 2003

DOI: 10.1093/ietcom/e88-b.10.3992 · Source: OAI

Yantai Shu
Tianjin University

Oliver W. W. Yang
University of Ottawa

Jiakun Liu
Chinese Academy of Sciences

Huifang Feng
The Northwest Normal University

IEICE TRANS. COMMUN., VOL.E88–B, NO.10 OCTOBER 2005

PAPER

Wireless Traffic Modeling and Prediction Using Seasonal ARIMA Models

Yantai SHU[†a)], ***Member***, **Minfang YU**[††], **Oliver YANG**[†††b)], **Jiakun LIU**[††††], ***and*** **Huifang FENG**[†], ***Nonmembers***

SUMMARY Seasonal ARIMA model is a good traffic model capable of capturing the behavior of a network traffic stream. In this paper, we give a general expression of seasonal ARIMA models with two periodicities and provide procedures to model and to predict traffic using seasonal ARIMA models. The experiments conducted in our feasibility study showed that seasonal ARIMA models can be used to model and predict actual wireless traffic such as GSM traffic in China.

key words: *traffic modeling, prediction, seasonal ARIMA models*

1. Introduction

According to the statistics of China Mobile of Beijing [1], the number of mobile phone users is increasing at an exponential rate. Therefore, there is an urgent need to study and understand the wireless traffic more deeply.

Traffic models have played a significant role in the design, engineering and performance evaluation of networks. In particular, time-series modeling holds a great promise as a tool for studying network traffic. The ARIMA (Auto Regressive Integrated Moving Average) model appears to be a good model to capture the behavior of the network traffic, and many variations of the ARIMA models have been broadly applied, e.g. the seasonal ARIMA model [2]. However, there is not much work and applications to wireless networking.

It is important to forecast wireless traffic workload in the planning, design, control and management of wireless networks. Traffic predictions have been used in various network research aspects, such as the long-range forecasting and planning of NSFNET [3], the linear prediction scheme used in the dynamic bandwidth allocation schemes for VBR video [4], and the predictive congestion control for broadband wide area networks [5]. Our work in this area includes the fractional ARIMA model in admission control [6], and the multiplicative seasonal ARIMA model for the prediction of traffic in the dial-up access network of Chinanet-Tianjin [7] in which only one periodicity is considered.

In this paper we provide a general expression for the multiplicative seasonal models with two periodicities. We propose a method to fit a seasonal ARIMA model to actual traffic traces and an adjusted traffic prediction method based on this mode. To the best of our knowledge, it is the first time to use a seasonal ARIMA model for modeling and predicting wireless traffic. Our many prediction experiments on the actual measured wireless traces demonstrated its feasibility.

This paper is organized as follows. Section 2 summarizes the basics mathematics of the ARIMA models. Section 3 studies the seasonal models used in our work. Section 4 shows how to build a seasonal ARIMA model to describe a traffic trace. Section 5 studies the traffic prediction using seasonal ARIMA models. Section 6 studies the feasibility of traffic modeling and prediction based on seasonal ARIMA models using the GSM (Global System for Mobile) traffic measured at the China Mobile-Tianjin Network. Section 7 is the applications. Section 8 is the concluding remarks.

For the remainder of this paper, the following notations pertain:

a_t a white noise with zero mean and variance σ^2
d the level of differencing
$e_t(h)$ a forecast error with a Normal distribution
h the lead time of forecast
p the order of autoregression
q the order of moving average
t the time; as a subscript, it also indicates the evolution of a series in time
u the desired probability that the observed value is less than the predicted value
B the backward-shift operator, i.e. $BX_t = X_{t-1}$
W_t a differenced seasonal ARIMA process
X_t a seasonal ARIMA process
$\hat{X}_t(h)$ the h-step forecast of X_t
$\hat{X}_t^u(h)$ the adjusted h-step forecast
ξ the bias in traffic prediction
$\hat{\sigma}_t^2(h)$ the minimum mean square error of the h-step forecasts
Ψ_j the coefficiences such that $\sum_{j=0}^{\infty} \Psi_j B^j = \Theta(B)\Phi^{-1}(B)\nabla^{-d}$
∇ the differencing operator, $\nabla = (1 - B)$
∇^d the dth differencing operator defined in the

Manuscript received October 8, 2004.
Manuscript revised April 3, 2005.
[†]The authors are with the Department of Computer Science, Tianjin University, Tianjin, 300072, China.
[††]The author is with China Academy of Space Technology, China.
[†††]The author is with the Faculty CCNR Lab, School of Information Technology and Engineering, University of Ottawa, Ottawa, Ontario, K1N 6N5, Canada.
[††††]The author is with the School of Science, Tianjin University, Tianjin 300072, China.
a) E-mail: ytshu@tju.edu.cn
b) E-mail: yang@site.uottawa.ca
DOI: 10.1093/ietcom/e88–b.10.3992

usual binomial expansion

$\Phi(B), \Theta(B)$ the polynomials of operator B

υ_j the coefficients such that $\sum_{j=0}^{\infty} \upsilon_j B^j = \Phi(B)\Theta^{-1}(B)\nabla^d$

Γ the Gamma function.

2. ARIMA Models

We summarize the mathematical properties of ARIMA processes here in order to introduce the notations used in the remainder of the paper. Let $\{X_t : t = \ldots, -1, 0, 1, \ldots\}$ be a time-series and B be the backward-shift operator, i.e. $BX_t = X_{t-1}$. Then, we let $\nabla = (1 - B)$ be the differencing operator and ∇^d is called the differencing operator defined in the usual binomial expansion, i.e.

$$\nabla^d = (1 - B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k \tag{1}$$

An ARIMA(p, d, q) process is a stochastic time-series process where d is the level of differencing, p is the autoregressive order, and q is the moving average order [2]. All three parameters are non-negative integers. Then the ARIMA(p, d, q) process can be described by the following relationship

$$\phi(B) \nabla^d X_t = \theta(B) a_t \tag{2}$$

where $\{a_t : t = \ldots, -1, 0, 1, \ldots\}$ is a Wiener (white noise) process with zero mean and variance σ^2, and

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p \tag{3}$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q \tag{4}$$

Both $\phi(B)$ and $\theta(B)$ are polynomials in complex variables with no common zeroes, and in addition $\phi(B)$ has no zeroes in the unit disk $\{B : |B| \le 1\}$.

3. Seasonal ARIMA Models

In general, we say that a series exhibits an s-periodic behavior when similarities in the series occur after s basic time intervals. For example, s-periodic series in Fig. 1 that we analyze later in this paper has an s of 7 days. After summarizing the seasonal ARIMA model, we shall give a general expression for the seasonal ARIMA model with two periodicities.

We first introduce the operator B^s such that $B^s X_t = X_{t-s}$ and the operator $\nabla_s = 1 - B^s$ such that $\nabla_s X_t = (1 - B^s) X_t = X_t - X_{t-s}$. The periodic effect implies that an observation for a particular time interval is related to the observations in same interval of earlier periods. Suppose that the tth observation X_t is for a particular interval. Mathematically, suppose X_t is the tth observation for a particular interval, then this observation can be linked to the observations in the same intervals of former periods by

$$\Phi(B^s) \nabla_s^D X_t = \Theta(B^s) a_t \tag{5}$$

where $\Phi(B^s), \Theta(B^s)$are polynomials in B^s of degrees P and Q respectively, in much similarity to (3) and (4) earlier and satisfying the stationarity and the invertibility conditions [2].

Now the error components $a_t, a_{t-1}, \ldots$ in these models are in general correlated. To take care of such relationship, we introduce an ARIMA(p, d, q) model as follows:

$$\phi(B) \nabla^d a_t = \theta(B) a_t \tag{6}$$

Combining (6) with (5), and following other details provided in [2], one arrives at

$$\phi_p(B) \Phi_P(B^s) \nabla^d \nabla_s^D X_t = \theta_q(B) \Theta_Q(B^s) a_t \tag{7}$$

In (7) the subscripts p, q, P, Q are added to distinguish the orders of the various operators.The resulting multiplicative process will be said to be of order $(p, d, q) \times (P, D, Q)_s$.

A similar argument can be used to obtain models with three or more periodic components to take care of multiple seasonalities. Therefore we can obtain models with two periodicities s_1 and s_2 as follows:

$$\begin{aligned} &\phi_p(B) \Phi_{P_1}(B^{s_1}) \Phi_{P_2}(B^{s_2}) \nabla^d \nabla_{s_1}^{D_1} \nabla_{s_2}^{D_2} X_t \\ &\quad = \theta_q(B) \Theta_{Q_1}(B^{s_1}) \Theta_{Q_2}(B^{s_2}) a_t \end{aligned} \tag{8}$$

In (8) the subscripts p, q, P_1, Q_1, P_2, Q_2 are added to distinguish the orders of the various operators and s_1, s_2 express the different periodicities existing in the time series. The resulting multiplicative process will be said to be of order $(p, d, q) \times (P_1, D_1, Q_1)_{s_1} \times (P_2, D_2, Q_2)_{s_2}$.

4. Building a Seasonal ARIMA Model to Describe a Trace

The basis of building our seasonal model is by spectrum analysis which is concerned with the exploration of cyclical patterns of data. Previous successful analysis [8] shows that this technique can uncover just a few recurring cycles of different lengths in the time series of interest. Therefore, we would like to use it to decompose our complex time series with cyclical components and obtain the few underlying sinusoidal functions of those particular periods.

Since there are several known ways of fitting ARMA models, we can take advantage of this by first transferring the ARIMA problem to an ARMA problem, and then identifying the necessary parameters. This can be done by rewriting (7) as

$$W_t = \nabla^d \nabla_s^D X_t \tag{9}$$

where

$$W_t = \phi_p^{-1}(B) \Phi_P^{-1}(B^s) \theta_q(B) \Theta_Q(B^s) a_t \tag{10}$$

Therefore, given a known time series X_t, the job is to estimate the parameter d and D from which we can obtain the seasonal ARMA model W_t from (9). The level of differencing d and D can be obtained from the basic statistical property of the time series. After differencing on X_t, we obtain W_t. The multiplicative seasonal ARMA model W_t may be viewed as a special form of ARMA model. Then, from W_t we can obtain the parameter $\phi_1, \phi_2, \ldots, \phi_p, \theta_1. \theta_2, \ldots, \theta_q, \sigma^2, \phi_{P1}, \phi_{P2}, \ldots, \phi_{PP}, \theta_{Q1}. \theta_{Q2}, \ldots, \theta_{QQ}$, according to the polynomial of $\phi(B)$, $\theta(B)$, $\Phi(B^s)$, $\Theta(B^s)$ from (10).

We propose the following procedure to fit a seasonal ARIMA model to the traffic trace. Explanation is provided where necessary.

Algorithm A: Procedure to fit a seasonal ARIMA model to traffic trace.

Step 1: Obtaining the periods such as s_1 and s_2 through spectrum analysis.

Step 2: Obtaining an estimate of d, D_1 and D_2 according to incremental analysis of the trace, determining d, D_1 and D_2 using ADF test [9].

Step 3: Performing differencing on X_t according to (9) to obtain a stationary series.

Step 4: Identifying the model by determining all the orders: Since our experience shows that p, P_1, P_2 and q, Q_1, Q_2, of fitted traffic ARIMA models are usually small, we propose to begin with candidate parameter sets that have small (p, q), (P_1, Q_1), (P_2, Q_2) values such as 0, 1, or 2, but where p, P_1, P_2 and q, Q_1, Q_2 should not be 0 simultaneously in one set. Then, we can select the best (p, q), (P_1, Q_1), (P_2, Q_2) combination according to the known model identification such as AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) [10].

Step 5: Estimating all the parameters. Use approximate maximum likelihood parameter estimation methods [11] to obtain all parameters: $\phi_1, \phi_2, \ldots, \phi_p, \theta_1. \theta_2, \ldots, \theta_q, \phi_{P_1 1}, \phi_{P_1 2}, \ldots, \phi_{P_1 P_1}, \theta_{Q_1 1}, \theta_{Q_1 2}, \ldots, \theta_{Q_1 Q_1}, \phi_{P_2 1}, \phi_{P_2 2}, \ldots, \phi_{P_2 P_2}, \theta_{Q_2 1}, \theta_{Q_2 2}, \ldots, \theta_{Q_2 Q_2}$

Step 6: Obtaining the fitted multiplicative seasonal ARIMA models from (8).

5. Prediction

One very useful application of our modeling research is to predict traffic. So in this section, we shall use the experience gained from the previous sections to provide a procedure to predict the future values of a time series from the current and past values.

5.1 Using Seasonal ARIMA Models to Forecast Time Series

Starting with the seasonal ARIMA process introduced in Sect. 3 and assuming causality and invertibility [12], we can re-write $\{X_t : t = \ldots, -1, 0, 1, \ldots\}$ as

$$X_t = \sum_{j=0}^{\infty} \Psi_j a_{t-j} \tag{11}$$

and

$$a_t = \sum_{j=0}^{\infty} \upsilon_j X_{t-j} \tag{12}$$

such that

$$\sum_{j=0}^{\infty} \Psi_j B^j = \theta_q(B)\Theta_Q(B^s)\phi_p^{-1}(B)\Phi_P^{-1}(B^s) \triangledown^{-d} \triangledown_s^{-D} \tag{13}$$

and

$$\sum_{j=0}^{\infty} \upsilon_j B^j = \phi_p(B)\Phi_P(B^s)\theta_q^{-1}(B)\Theta_Q^{-1}(B^s) \triangledown^{d} \triangledown_s^{D} \tag{14}$$

Equations (11) to (14) form the basis of our prediction algorithm. Since seasonal ARIMA models are linear models, we can use linear prediction to make forecasts based on the minimum mean square error (MMSE). Let $\hat{X}_t(h)$ be the MMSE-predictor of X_t. In another word, $\hat{X}_t(h)$ is the predicted value of X_{t+h} for a lead-time of h-steps. Simple derivation using the seasonal ARIMA process allows us to write the following,

$$\hat{X}_t(h) = -\sum_{j=1}^{\infty} \upsilon_j^{(h)} \hat{X}_{t+h-j} \tag{15}$$

where

$$\upsilon_j^{(1)} = \upsilon_j,$$

and

$$\upsilon_j^{(h)} = \upsilon_{j+h-1} - \sum_{i=1}^{h-1} \upsilon_i \upsilon_j^{(h-i)}, h > 1.$$

We can also write

$$\hat{X}_t(h) = \sum_{j=h}^{\infty} \Psi_j a_{t+h-j} \tag{16}$$

and obtain their MMSE $\hat{\sigma}_t^2(h)$ of the h-step forecasts given by

$$\hat{\sigma}_t^2(h) = E(X_{t+h} - \hat{X}_t(h))^2 = \sigma^2 \sum_{j=0}^{h-1} \Psi_j^2 \tag{17}$$

5.2 Traffic Prediction Based on Upper Probability Limit

When traffic control and management are based on prediction, new calls can be blocked if actual arrivals are continuously greater than predicted traffic value. Therefore, it would be useful to specify the probability limits of a given prediction algorithm [13]. But unlike the upper- and lower-limits used in normal prediction techniques, we only need to calculate the upper probability limit to specify the accuracy of traffic prediction in networks because the lower limit does not contribute to any calls blocking probability. Therefore, we need to adjust prediction technique from before.

To do it, let $\hat{X}_t^u(h)$ be the adjusted h-step forecast by adding a bias ξ_u to the minimum mean square error forecast, i.e.

$$\hat{X}_t^u(h) = \hat{X}_t(h) + \xi_u \tag{18}$$

Let u be the desired probability that an observed (actual) value is less than the predicted value; (1-u) is therefore the probability that the observed value is more than the predicted value. Assuming a forecast error $e_t(h)$ with a Normal distribution [14], then the relationship between ξ_u and u can be determined numerically beforehand via

$$P\left[e_t(h) \leq \xi_u\right] = u \quad 0.5 \leq u \leq 1 \tag{19}$$

For example, let $\sigma_t(1)$ denote the standard deviation of the one-step forecast error which we can obtain using (17). Then under the assumption of a Normal distribution for the forecast error $e_t(h)$, we can obtain the bias

$$\xi_u = z_u \times \sigma \tag{20}$$

where z_u is the uth quantile of a Normal distribution, and σ is the standard error of the $\{e_t(h)\}$, which is also equal to $\sigma_t(1)$. Let $\Phi(\cdot)$ be the probability distribution function with a Normal distribution. Then, we have $\Phi(0) = 0.5$ when $z_u = 0; \Phi(1) = 0.8412 \approx 0.84$ when $z_u = 1; \Phi(2) = 0.9772 \approx 0.98$ when $z_u = 2 \ldots$etc. Consequently, we can determine the relationship

$$\xi_u = \begin{cases} 0 & \text{when} \quad u = 0.5 \\ \sigma_t(1) & \text{when} \quad u = 0.84 \\ 2\sigma_t(1) & \text{when} \quad u = 0.98 \end{cases} \tag{21}$$

Finally, we summarize in the following our prediction algorithm for a network traffic stream.

Algorithm B: Procedure to predict traffic of a given upper-bound call blocking probability.

Step 1: From the QoS requirement of a particular network (e.g. call blocking probability), determine the value of u.
Step 2: From u,determine the value of ξ_u (e.g. from (21)).
Step 3: From the operating time scale of a particular network traffic-management mechanism which is based our prediction method, determine the value of the time granularity and the parameter h of the h-step traffic prediction.
Comment: Normally, we choose $h = 1$ from an engineering point of view.
Step 4: Use Algorithms A to construct a seasonal ARIMA models to fit the traffic trace.
Step 5: Predict the next value of the time series using h-step minimum mean square error forecast.
Step 6: Obtain the predicted traffic by adding a bias ξ_u from (18).

6. Feasibility Study

We have performed experiments on real-traffic traces to study the feasibility of the proposed algorithms (Algorithm B) on modeling and prediction. We first fit a seasonal ARIMA model for a trace. Then, we evaluate the performance of the prediction algorithms.

We used real traffic traces measured from the GSM Network of China Mobile of Tianjin. We have one original hourly traffic trace from 0:00 June 1 2001 (Friday) to 0:00 April 27 2002 (Saturday), a total of 330 days. Each value of the hourly traffic trace represents the sum of connection time in the past one hour. We accumulated the traffic in each day to obtain the daily traffic trace for the same 330 days. Each value of the daily traffic trace represents the sum of connection time in the past one day. Each sample of traffic is expressed in Erlang. We have used the previous 301 day data trace to do modeling and forecast next 28 day values. We compared the forecasted value with original value to evaluate the performance of the prediction algorithms.

6.1 Analyzing Actual GSM Traffic

Figure 1 and Fig. 2 show the original traces of the daily and hourly traffic respectively. While the daily traffic shows a periodicity of 7 (one week), the hourly traffic shows two periodicities of 24 (one day) and 168 (one week).

The periodogram based on the daily traffic trace is shown in Fig. 3. We can see that:

1. There is a peak when the frequency is about 0.14 which is called the main frequency. From this we can infer that the period of this network traffic is 7 (one week). This is in accordance with the actual situation.
2. There is a second peak when the frequency is about 0.28 which is called second harmonic. This second har-

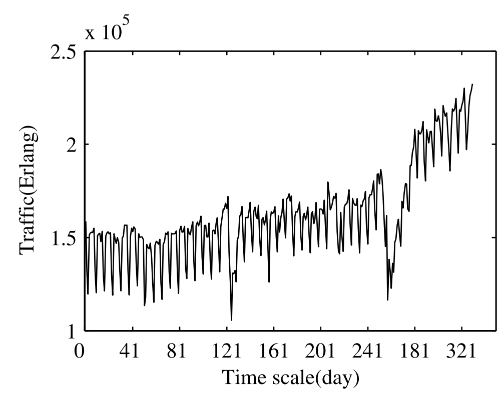

Fig. 1 Daily traffic trace.

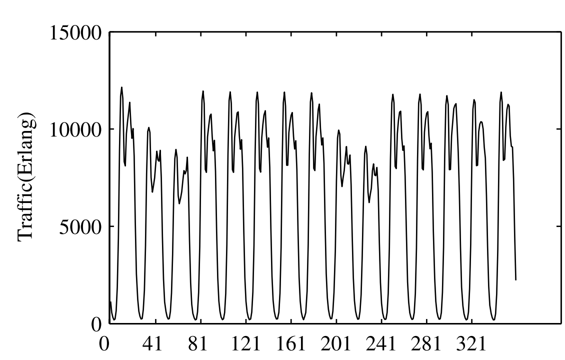

Fig. 2 Hourly traffic trace.

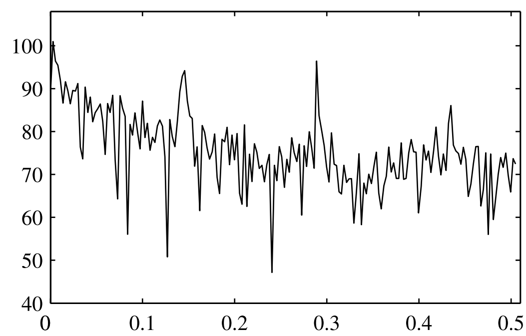

Fig. 3 Periodogram based on daily traffic trace.

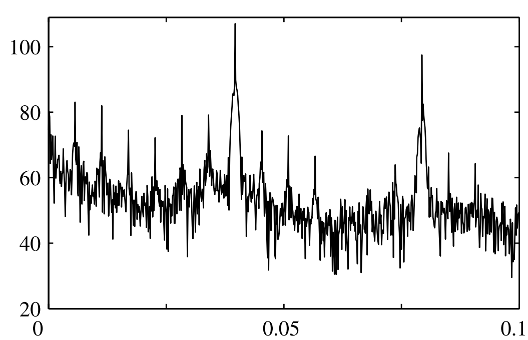

Fig. 4 Periodogram based on hourly traffic trace.

monic is formed because of the asymmetry of network traffic in the seven days period.

3. There is a third peak when the frequency is about 0.42 which is called third harmonic. This is due to the traffic on Saturday and Sunday that is far below the traffic in workdays. It forms sharp canyon in Fig. 1.

The periodogram based on the hourly traffic trace is shown in Fig. 4. From Fig. 4, we can also see that:

1. There is a peak when the frequency is about 0.042 which is called the main frequency. we can get the periodicity 1/0.042=24. That is to say that there is a periodicity of one day.
2. There is a second peak when the frequency is about 0.08 which is called second harmonic. This second harmonic is formed because of the asymmetry of network traffic in the one day period. This means that network traffic varies over hours in a single day.
3. There is another main frequency at 0.006 with second and third harmonics. This corresponds to the periodicity of 168 i.e. one week.

Thus, the hourly traffic shows two periodicities of 24 (one day) and 168 (one week).

6.2 Building Seasonal ARIMA Models for Actual GSM Traffic

After we have made the certain periods of the analyzed GSM traffic, we will then build the seasonal ARIMA model with periodicity of 7 for the daily trace, and the multiple seasonal ARIMA model with two periodicities of 24 and 168 for the hourly trace respectively.

First we preprocess the two traces: We notice over all, the GSM traffic increases linearly over time. However, during long holidays such as Chinese New Year and October 1st national day, we see a dramatic drop in traffic. These dips causes our predictions to report lower than the expected values. One solution for this problem is to "patch" these dips with correlating data from before and after the holiday. We do this by taking the average of corresponding date of the week and time of day during the period preceding and following the dip. We then input these values into the dip matching the correct corresponding time interval. Next we use step 1–6 of Algorithm A to process the two traces. The fitted seasonal ARIMA models are given in Table 1. The models of traffic traces show that these traffic traces exhibit periodicity.

Table 1 Fitted seasonal ARIMA models of 301 day GSM traffic.

Trace	Models
Daily traffic	$\theta_1 = -0.312, \phi_1 = 0.333$ $\Phi_{11} = -0.322$ $(1, 0, 1) \times (1, 1, 0)_7$
Hourly traffic	$\theta_1 = 0.0376, \Phi_{11} = -0.3728$ $\Phi_{21} = 0.708$ $(0, 1, 1) \times (1, 1, 0)_{24} \times (1, 0, 0)_{168}$

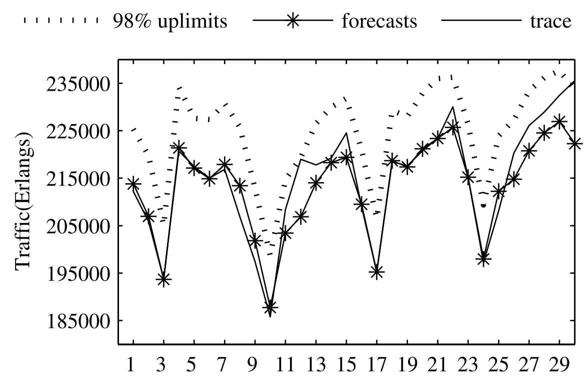

Fig. 5 Forecasts of 301 day daily traffic trace.

6.3 Traffic Prediction for Actual GSM Traffic

We have done the "minimum mean square error forecast" [2] experiments on various traces. In the following, we shall use the data from the daily and hourly traffic of 301 days to forecast the values of the next 28 days using the model we have built above. The result is shown in Fig. 5 and Fig. 6 We also show the upper probability 98% limit which correspond to a bias $\xi_u = 2\sigma_t(1)$ using adjusted traffic prediction of the daily and hourly traffic trace in Fig. 5 and Fig. 6 respectively.

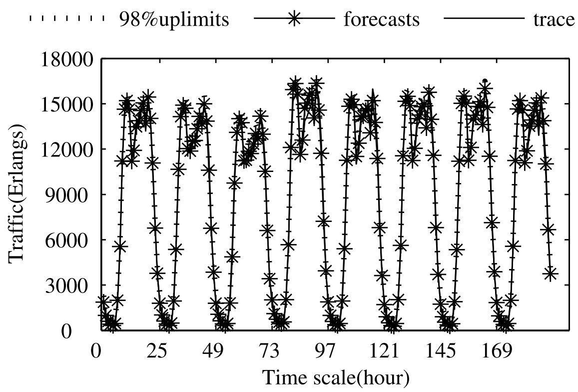

Fig. 6 Forecasts of 301 day hourly traffic trace.

Table 2 Numerical analysis (relative error) of 301 day daily traffic forecasts.

Lead time	Relative error
First week	0.00032
Second week	0.014611
Third week	0.003563
Fourth week	0.014605

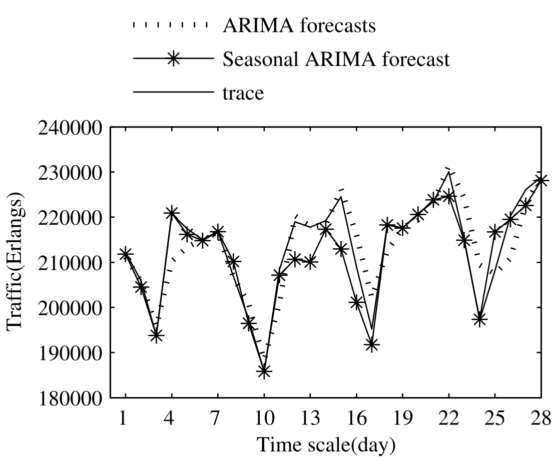

Fig. 7 Forecasts of 28 day daily traffic trace using seasonal ARIMA model comparing with using ARIMA model.

6.4 Comparing the Forecasts with the Actual Traffic Traces

We compare the forecasted values of the traffic with the actual traffic. These are shown in Fig. 5 and Fig. 6. The comparisons show that they are close. The relative error between forecasting values and actual values are all less than 0.02. The numerical analysis results of daily traffic forecasts are given in Table 2, and demonstrate the validity of our prediction method.

We have also used the previous 28 day data trace to do modeling and to forecast the values of next 28 days. Figure 7 compares the daily traffic forecast using the Seasonal ARIMA model with that using the ARIMA model. Visually, the prediction result of using seasonal ARIMA model is better than that using ARIMA model. To confirm our observation mathematically, we also present in Table 3 and Table 4, respectively, the computed relative error and the mean absolute error between the forecasting values and the actual values for the same two models. From these Tables, We can see that the precision of the seasonal ARIMA model is better than that of ARIMA model. Moreover, according to our analysis, the measured wireless traffic exhibits an obvious periodicity. Therefore, we can confidently use a seasonal ARIMA model for modeling and predicting wireless traffic. Lastly, notice that when the prediction error of ARIMA model is close to that of seasonal ARIMA model, our experiment results show that, when compared to the ARIMA model, our seasonal ARIMA model needs less parameters and samples to model and predict the same traffic. In conclusion, the seasonal ARIMA model based prediction has more adaptability to track the traffic load.

Table 3 Numerical analysis (relative error) of 28 day daily traffic forecasts using seasonal ARIMA model comparing with using ARIMA model.

Lead time	Relative error	
	Seasonal ARIMA	ARIMA
First week	0.002347	0.011943
Second week	0.015549	0.011221
Third week	0.015831	0.015985
Fourth week	0.012722	0.02443
Total four week	0.011612	0.015895

Table 4 Numerical analysis (mean absolute error) of 28 day daily traffic forecasts using seasonal ARIMA model comparing with using ARIMA model.

Lead time	Mean absolute error (Erlangs)	
	Seasonal ARIMA	ARIMA
First week	497.9458	2578.892
Second week	3345.817	2285.026
Third week	3337.631	3394.351
Fourth week	2777.018	5163.048
Total four week	2503.783	3341.1492

On the other hand, we have also used fractional ARIMA models to describe the GSM trace and forecast traffic, but we did not find any improvement which can probably be attributed to the weakness of the long-range dependency in the traffic characteristics.

7. Applications

Our seasonal ARIMA model based traffic prediction can be extended to network design, management, planning and optimization. For example, in order to determine the service performance or the number of channels for a predicted Erlang Load, one can use the GOS (Grade of Service) chart given by Erlang B Table and Erlang C Table [15], [16]. Also, the output of our traffic analysis can provide a number of performance measures that can give an idea of the system performance as follows.

We can define Grade of Service (GOS) to be the probability of any incoming calls being queued. Similarly, the Average Queue Depth is the average number of users waiting in the queue, and the Average Call Delay is the average time in queue, average across all incoming calls.

$$GOS = \frac{A^C}{A^C + C!\left(1 - \frac{A}{C}\right)\sum_{k=0}^{C-1}\frac{A^k}{k!}} \tag{22}$$

$$\text{Average Queue Depth} = \frac{GOS \times A}{C - A} \tag{23}$$

$$\text{Average Call Delay} = \frac{GOS \times H}{C - A} \tag{24}$$

where A (Erlangs) is the total number of hours of call traffic during the busy hour period, C (Number of Channels) is the number of working channels. H (Busy Hour) is the hour of the day with the heaviest call traffic load.

8. Conclusions

In this paper, we have studied a method of fitting multiplicative seasonal ARIMA models to measured traffic traces. We gave a general expression of the multiplicative ARIMA models with two periodicities and we proposed a practical algorithm for building seasonal ARIMA models. We proposed an adjusted traffic prediction method using seasonal ARIMA models. We have repeated the comparison with many prediction experiments on the GSM traces actually measured in the network of China Mobile of Tianjin. We found that the relative error between our forecasting values and the actual values are all less than 0.02. This lends a strong support to our prediction method. Our experiments showed that the seasonal ARIMA model is a good traffic model capable of capturing the properties of real traffic.

Future work includes applying seasonal ARIMA models for various traffic predictions and comparing performance of seasonal ARIMA and ARIMA models in detail. Furthermore, The application with details of predict-based network design, management, planning and optimization will be studied in our future work.

Acknowledgement

This research was supported in part by the National Natural Science Foundation of China (NSFC) under grant No. 60472078, by a grant from the Cisco University Research Program Fund at Community Foundation Silicon Valley, by the Natural Sciences and Engineering Research Council of Canada (NSERC) under Grant No. RGPIN-2003-42878, by Central of Liuhui Applied Mathed.

The authors express gratitude to Prof. Zhigang Jin and Xu Wang for their many helpful suggestions. They would also like to thank Jingang Bo, Jianzhong Bai and their colleagues at China Mobile of Tianjin who provided us with their measured trace.

References

[1] http://www.bmcc.com.cn
[2] G. Box, G. Jenkins, and C. Reinsel, Time Series Analysis: Forecasting and Control, 3rd ed., Prentice-Hall, Englewood Cliffs, NJ, 1994.
[3] N. Groschwitz and G. Ployzos, "A time series model of long-term NSFNET backbone traffic," Proc. IEEE ICC'94, pp.1400–1404, New Orleans, USA, May 1994.
[4] A. Adas, "Supporting real time VBR video using dynamic reservation based on linear prediction," Proc. IEEE INFOCOM'94, pp.1476–1483, Toronto, Canada, June 1994.
[5] G. Ramamurthy and B. Sengupta, "A predictive congestion control policy for broadband integrated wide area networks," Comput. Netw. ISDN Syst., no.28, pp.811–834, 1996.
[6] Y. Shu, Z. Jin, J. Wang, and O.W.W. Yang, "Prediction-based admission control using FARIMA models," Proc. IEEE ICC'00, pp.1325–1329, New Orleans, Louisiana, USA, June 2000.
[7] Y. Shu, Z. Jin, and D. Tian, "Dial-up user models and traffic prediction," Proc. Future Telecommunications Conference 2001, pp.90–93, Beijing, China, Nov. 2001.
[8] I.F. Gihman and A.V. Skorohod, The Theroy of Stochastic Processes, vol.1, Springer, Berlin, Heidelberg, NewYork, 1975.
[9] D.A. Dickey and W.A. Fuller, "Likelihood ratio statistics for autoregressive time serries with a unit root," Econometrica, vol.49, pp.1057–1072, 1981.
[10] H. Akaike, Time Series Analysis and Control Through Parametric Models, Applied Time Series Analysis, Academic Press, New York, 1978.
[11] J. Haslett and A.E. Raftery, "Space-time modeling with long-memory dependence: Assessing Ireland's wind power resource," Appl. Stat., vol.38, no.1, pp.1–50, 1989.
[12] P. Brockwell and R. Davis, Time Series: Theory and Methods, 2nd ed., Springer Verlag, New York, 1991.
[13] S. Basu and A. Mukherjee, "Time series models for Internet traffic," Proc. IEEE INFOCOM'96, pp.611–620, San Francisco, USA, March 1996.
[14] R.G. Addie, "Traffic will be more Gaussian in future," Proc. ATNAC'96, Melbourne, Dec. 1996.
[15] K. Tutschku and P. Tran-Gia, "Spatial traffic estimation and characterization for mobile communicationnetwork design," IEEE J. Sel. Areas Commun., vol.16, no.5, pp.804–811, 1998.
[16] R.M. Whitaker and S. Hurley. "Evolution of planning for wireless communication systems," Proc. HICSS'03, pp.296–305, 2003.

Yantai Shu is a professor of Computer Science at Tianjin University, China, serving as vice president of the university from 1993–1997. His current interests are focused on computer communication networks, wireless networks, real-time systems, modeling and simulation. He received Chinese equivalent to B.S., M.S., and Ph.D. in electronics engineering from Tianjin University. From 1974 until 1991 he was employed by the Institute of Plasma Physics, Academia Sinica in research positions. He is a member of the IEEE and the ACM. He has published more than 120 papers and contributed to one book.

Minfang Yu is an engineer in China Academy of Space Technology. She received her M.S. in Computer Science, Tianjin University, China, in 2004. Her current interests are focused on computer networks, modeling and simulation, Wireless Multimedia network.

Oliver Yang is a professor in the School of Information Technology and Engineering at University of Ottawa, Ontario, Canada. Dr. Yang received his Ph.D. degree in Electrical Engineering from the University of Waterloo, Ont., Canada in 1988. His research interests are in the modeling, analysis and performance evaluation of computer communication networks, their protocols, services and interconnection architectures, queueing theory, simulations, computational algorithms and their applications. Dr. Yang has published more than 200 technical papers in wireline, wireless and photonic networks. He is currently an editor of the IEEE Communication Magazine.

Jiakun Liu is a professor in School of Science at Tianjin University, China. His research interests are in random process, performance evaluation of computer networks.

Huifang Feng is currently a Ph.D. student in Computer Science Department, Tianjin University. She received her M.S. degree in computer science from Lanzhou University, Lanzhou, China, in 1996. Her research interests are in computer communication networks, wireless networks, modeling and simulation.