Traffic Forecasting for Mobile Networks with Multiplicative Seasonal ARIMA Models

Jia Guo [1] Yu Peng [1] Xiyuan Peng [1] Qiang Chen [2] Jiang Yu [2] Yufeng Dai [2]
1.Automatic Test and Control Institute, Harbin Institute of Technology
Harbin 150080, China
2.China Mobile Communications Corporation Heilongjiang Co.Ltd
Harbin 150028, China
[1]Email: *guojia.gm@gmail.com*

Abstract *–Traffic forecasting is an important task which is required by overload warning and capacity planning for mobile networks. Based on analysis of real data collected by China Mobile Communications Corporation (CMCC) Heilongjiang Co.Ltd, this paper proposes to use the multiplicative seasonal ARIMA models for mobile communication traffic forecasting. Experiments and test results show that the whole solution presented in this paper is feasible and effective to fulfill the requirements in traffic forecasting application for mobile networks.*

Keywords *– mobile network management, traffic forecasting, multiplicative seasonal ARIMA models.*

I. INTRODUCTION

Rapid development of mobile communication industry brings the overload warning and capacity planning of mobile networks to play increasingly important roles. Traffic forecasting is an important task which is required by overload warning and capacity planning [1]. Traffic forecasting for mobile networks is difficult to carry out manually due to efficiency and accuracy.

Recently, researches on traffic forecasting and traffic model for mobile networks have been proposed [1]-[6]. Holt-Winter's exponential smoothing is used as a forecast method for traffic in [1]. In [2] authors analyze traffic characterization which can be used to model the traffic in cellular networks. Based on mobility prediction, authors in [3] propose a traffic forecasting method by estimating the number of users that have active sessions at each location. A model for cellular traffic forecasting is proposed in [4] by integrating a vehicular traffic model. In [5] authors propose a method of forecasting traffic systematically based on the user's properties and the environment information. However, the information is not easy to obtain. An analytical model for channelized cellular mobile circuit-switched systems that support general arbitrary distributed handoff traffic is proposed in [6]. Moreover, most of previous efforts at mobile communication traffic forecasting focus on a small quantity of large-scale network (e.g., city). Little research has been conducted in applying traffic forecasting to huge amounts of cells in mobile networks.

Based on analysis of real data collected by NMS (Network Management Systems) of CMCC Heilongjiang Co.Ltd, this paper proposes to use the multiplicative seasonal ARIMA models for traffic forecasting. We focus on designing a practical scheme to process historical data of cells in mobile networks and generate reliable forecasting results for overload warning and capacity planning.

The remainder of the paper is structured as follows. Section II presents analysis of real traffic data. In Section III, we describe the scheme design, along with introduction of traffic forecasting method. The test results and discussion are presented in Section IV. We conclude in Section V.

II. TRAFFIC DATA ANALYSIS

The real data are collected by cells hourly including record time, cell name, TCH traffic, etc. NMS generates a daily file with about 500,000 records. Traffic data is calculated by the unit Erlang. Traffic characteristics of cells can be obtained from traffic data. We can regard traffic data as nonstationary random signals. Traffic intensity always changes by day and week. Cells usually have higher traffic during daytime than nighttimes in a day. When it comes to a week, the changes are more complicated. Some cells have higher traffic during weekdays than weekends or vice versa and some cells have no significant changes in weekdays and weekends. Moreover, several special factors (e.g., festivals, emergency) always cause traffic unusual changes which cannot be forecasted by statistical analysis limited to the deficient information.

Figure 1 shows an example of traffic intensity changes during Dec.2, 2007 (Sunday) to Dec.8, 2007 (Saturday) with 24 measurements per one day. This cell has higher traffic during weekdays than weekends, and has higher traffic during daytime than nighttimes.

Figure 2 shows an example of traffic intensity changes during Dec.2, 2007 (Sunday) to Dec.8, 2007 (Saturday) with 24 measurements per one day. This cell has no significant changes in weekdays and weekends, and has higher traffic during daytime than nighttimes.

978-1-4244-3864-8/09/$25.00 ©2009 IEEE

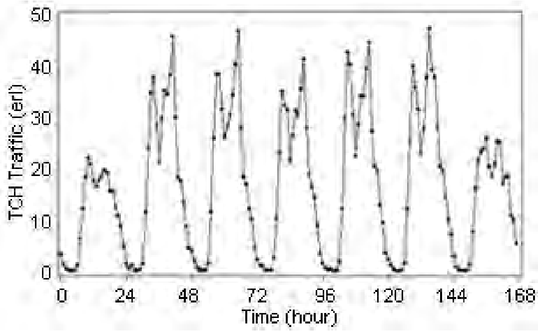

Fig. 1. Traffic data with significant weekly changes.

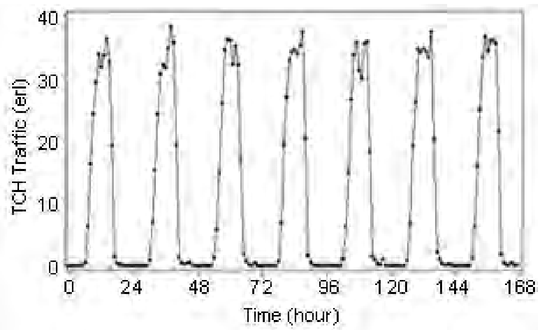

Fig. 2. Traffic data with no significant weekly changes.

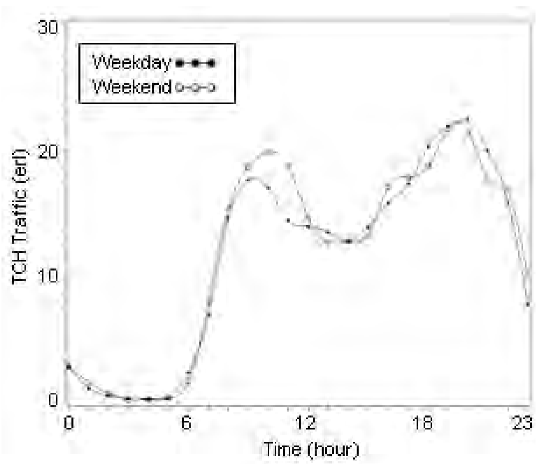

Fig. 3. Traffic intensity versus 24 hours of one day (a cell from residential district category).

Based on the above analysis, we tentatively classify all the cells into four categories: residential district, commercial district, street and university campus, according to cell location, cell size, number of channels, traffic intensity, etc. Different categories have different traffic characteristics. For example, Figure 3 and Figure 4 show the average traffic intensity over the hours of the weekdays and weekends of a cell from residential district category and a cell from street category respectively.

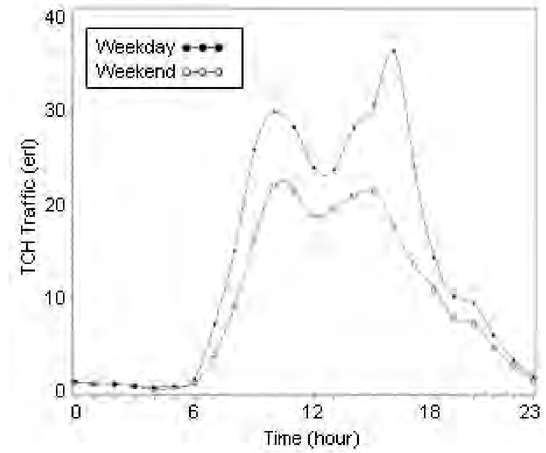

Fig. 4. Traffic intensity versus 24 hours of one day (a cell from street category).

III. FORECASTING METHOD AND SCHEME DESIGN

A. Forecasting method

Data analysis shows that the mobile communication traffic data of CMCC Heilongjiang Co.Ltd have seasonality and trend generally. Therefore, this paper proposes to use the multiplicative seasonal ARIMA models for traffic forecasting. The multiplicative seasonal ARIMA models are also named as Box-Jenkins methodology. The modeling and forecasting procedure involves the following steps: i) model identification (i.e., determine fitted model structure of the time series), ii) parameter estimation and checking (i.e., determine the values of parameters), iii) forecasting (i.e., forecasting future values of the time series based on the fitted model). Due to space constraints, we will not go into details of the method, and refer the reader to [7, 8].

B. Scheme design

In consideration of data update frequency and application requirements, we designed the scheme of traffic forecasting for mobile networks with multiplicative seasonal ARIMA models. The scheme design involves the following steps:

1)Data import. The original data (e.g., record time, cell name, TCH traffic) are imported to local database, and the data should be transformed to database formats.

2)Data pre-processing. Due to system faults or other factors, real data always have missing values or erroneous values. In this step, missing values and erroneous values are pre-processed.

3)Modelling and forecasting. The modelling and forecasting of traffic data are completed in this step including identification, estimation and forecasting. First of all, the identification part loads data from database and differences the series data if desired, then computes statistics (e.g. autocorrelation function) to help identify models. Secondly, the estimation part specifies a model, produces estimates of its parameters and prints diagnostic information by which to check the model. Parameter estimation is done using Maximum Likelihood Estimation. The best model is chosen as the one that provides the smallest AIC (Akaike Information Criterion) with orders which do not exceed 2. Finally, the forecasting part generates forecast values using the parameter estimates produced by the previous estimation part.

4)Result output. Report generation, data and figure outputs are completed in this step. All the forecast results output to database.

IV. RESULTS AND DISCUSSIONS

Base on the real data collected hourly during Nov.4, 2007 to Dec.8, 2007 by NMS of CMCC Heilongjiang Co.Ltd, we did modelling and forecasting tests. The data were divided into two parts: first part contains data during first 28days (672 hours) and considered as the modelling data; second part consist of rest 7 days (168 hours) and considered as the evaluation data. Due to the characteristics of ARIMA models, the forecasting precision decreases along with the increase of forecasting-step [7]. Consequently, in consideration of data update frequency and application requirements, we choose 168 hours (i.e., one week) as the forecasting-step.

We did tests for 269 cells from four different categories presented in Table 1. The results were evaluated by Normalized Root Mean Square Error (NRMSE) specified by equation (1).

$$\mathrm{NRMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\frac{y_t - \hat{y}_t}{y_t}\right)^2} \tag{1}$$

where the y_t, $\hat{y}_t$ are the real value of data and its corresponding forecast value at time t, respectively. And n is total number of forecast values.

Table 1. Number of cells from four categories.

Category	Number of cells
residential district	21
commercial district	76
street	139
university campus	33
total	269

We calculated the average NRMSE for each category and all the cells. The results are presented in Table 2. For all the cells we tested on, the average NRMSE is 0.2061; the best result is 0.0810 and the worst result is 0.3618. In the four categories, we get the best result of average NRMSE for university campus, and the worst one for residential district.

Table 2. Average NRMSE.

Category	Average NRMSE
residential district	0.2448
commercial district	0.2023
street	0.2054
university campus	0.1932
total	0.2061

The frequency of NRMSE of all the 269 cells is presented in Figure 5. The NRMSE values distribute in the range of [0.0810, 0.3618]. And there are 246 cells whose NRMSE values distribute in the range of [0.1097, 0.3019]. That is 91.45% of all the 269 cells.

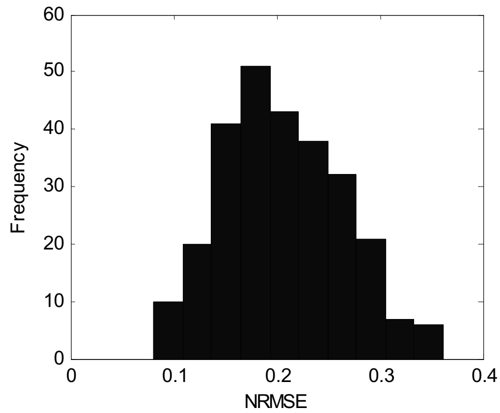

Fig. 5. The frequency of cells with NRMSE(total).

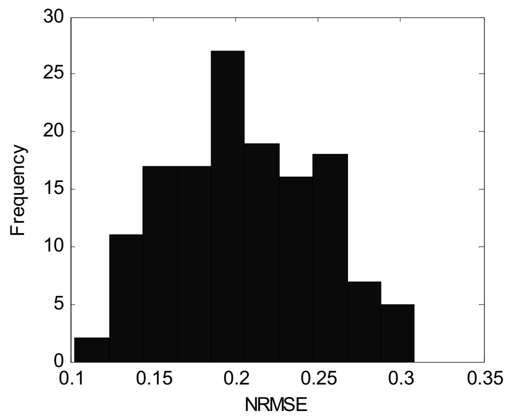

Fig. 6. The frequency of cells with NRMSE(street category).

Due to space constraints, we will not show results of all the four categories. Figure 6 presents the frequency of NRMSE of cells from street category. The NRMSE values distribute in the range of [0.1029, 0.3087]. And there are 125 cells whose NRMSE values distribute in the range of [0.1283, 0.2672]. That is 89.93% of all the 139 cells from street category.

The test results show that forecast errors are acceptable. Moreover, Figure 5 and Figure 6 denote that proposed technique also has strongly robust property.

Figure 7 presents the forecasting results of the cell which has the best NRMSE result in our tests. We obtain the fitted model as follows: ARIMA$(2,0,1)\times(2,1,0)_{168}$, with ϕ_1 =1.1005, ϕ_2 =-0.1302, Φ_1 =-0.5836, Φ_2 =-0.5060, Θ_1 =0.9228. Figure 7 shows that proposed technique gets accurate results for this cell.

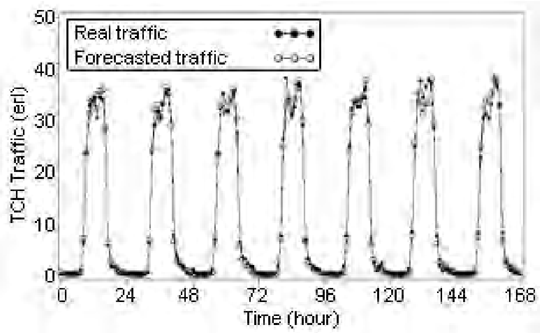

Fig. 7. Forecasting results with the best NRMSE.

Figure 8 presents the forecasting results of the cell which has the worst NRMSE result in our tests. We obtain the fitted model as follows: ARIMA$(2,0,1)\times(0,1,1)_{168}$, with ϕ_1=1.4033, ϕ_2=-0.4092, θ_1 =0.9054, Θ_1 =0.8227. Figure 8 shows that the forecasted traffic is always higher than real traffic.

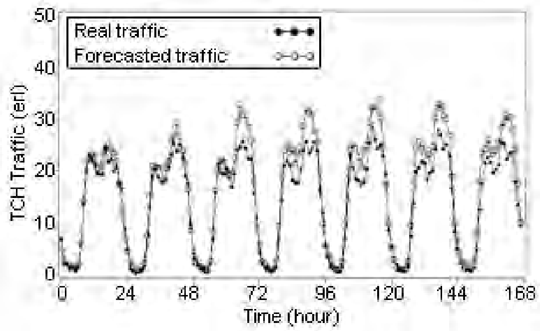

Fig. 8. Forecasting results with the worst NRMSE.

Based on analysis and test results above, proposed technique with multiplicative seasonal ARIMA models can get satisfactory and reliable results for modeling and forecasting traffic for mobile networks.

We design a traffic analysis and forecasting system for mobile networks based on proposed technique. Figure 9 shows the framework of the system. Oracle database is adopted for data storage and management. Modeling and forecasting module obtains data from database and outputs results to database. Users can make parameter setting and visual inspection through data visualization and user interface module which is developed in Java.

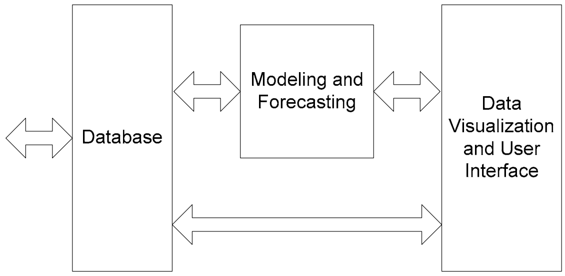

Fig. 9. Block diagram of Traffic Analysis and Forecasting System.

V. CONCLUSION

This paper presented a technique for forecasting traffic of cells in mobile networks. We applied proposed technique to real data collected by NMS of CMCC Heilongjiang Co.Ltd. The whole solution proposed in this paper is feasible and effective for traffic forecasting for mobile networks, and it could be served as a support for overload warning and capacity planning.

ACKNOWLEDGMENT

This work was supported in part by China Mobile Communications Corporation Heilongjiang Co. Ltd under the Technology Development Projects under Grant [2008]1891.

REFERENCES

[1] Denis Tikunov and Toshikazu Nishimura, "Traffic Prediction for Mobile Network using Holt-Winter's Exponential Smoothing". *15th International Conference on Software, Telecommunication and Computer Networks*, pp. 310-314, Sep. 2007.

[2] Houda Khedher and Fabrice Valois, "Traffic Characterization for Mobile Networks". *IEEE 56th Vehicular Technology Conference*, Vol. 3, pp. 1485-1489, May. 2008.

[3] Haitham Abu-Ghazaleh and Attahiru Sule Alfa, "Mobility Prediction and Spatial-Temporal Traffic Estimation in Wireless Networks". *IEEE 67th Vehicular Technology Conference*, pp. 2203-2207, May. 2008.

[4] Raffaele Bolla and Matteo Repetto, "A new model for network traffic forecast based on user's mobility in cellular networks with highway stretches". *International Journal of Communication Systems*, Vol. 17, No. 10, pp. 911-934, Dec. 2004.

[5] Yoshikazu Akinaga, Shigeru Kaneda, Noriteru Shinagawa and Akira Miura, "A proposal for a mobile communication traffic forecasting method using time-series analysis for multi-variate data". *IEEE Global Telecommunications Conference 2005*, Vol. 2, pp. 1119-1124, Nov. 2005.

[6] Samya Bhattacharya, Hari Mohan Gupta and Subrat Kar, "Traffic Model and Performance Analysis of Cellular Mobile Systems for General Distributed Handoff Traffic and Dynamic Channel Allocation". *IEEE Transactions on Vehicular Technology*, Vol. 57, No. 6, pp. 3629-3640, Nov. 2008.

[7] George E P Box, Gwilym M Jenkins and Gregory C Reinsel, "Time Series Analysis: Forecasting and Control, 3rd ed". *Englewood Cliffs, NJ: Prentice-Hall*, pp.377-391, 1994.

[8] Peter J Brockwell and Richard A Davis, "Time series: Theory and Methods, 2nd ed". *New York: Springer-Verlag*, pp. 250-253, 1991.