Dial-up User Models and Traffic Prediction

Yantai Shu, Zhigang Jin, Dongxu Tian
Department of Computer Science, Tianjin University,
Tianjin 300072, China
Tel.: +86-22-27402274, Fax: +86-22-27404544
ytshu@tju.edu.cn

ABSTRACT

One important aspect to offer QoS in networking is to be able to analyze and predict network traffic. In this paper, we have measured the 2163 dial-up network traffic, built an ARIMA model for it, and on this basis, predicted the future network t raffic. All the work we have done give a reference for future network capacity plan.

Key words: *Network measurement, Network prediction, ARIMA model*

1. INTRODUCTION

The advent of high-speed networking has introduced opportunities for new applications such as video conferencing, scientific visualization and medical imaging. These applications have stringent performance QoS (Quality of Service) requirements [1], and it is a very important research. One key aspect to meet this requirement is to be able to analyze and predict network traffic. By analyzing the network traffic, making accurate prediction based on the traffic modeling, doing the network capacity plan and making corresponding adjust in advance, we can assure with certain confidence the whole operation quality.

The Computer Network Performance Group of Tianjin University measured the dial-up network traffic of 2163 and others at the Chinanet-Tianjin Network, analyzed the measurements, built traffic models and made prediction on the future network traffic. These measurements and analysis offer useful information to the ISP (Internet Service Provider).

The main tasks we have accomplished are as follows:

(1) Measuring the dial-up network traffic such as 2163 and so on. We mainly measured the daily network users' overall quantity in year 2000, overall time length about the dial-up users' online time and average time length (unit: minute)

(2) Doing the spectrum analysis about the 2163 traffic and its progressive traffic. By our analysis, we can find that the traffic's frequency is 0.14, this is to say that it has a period of 7 days which is in accordance with the people's work schedule. Under this foundation, we can make analysis and statistical check and build appropriate network traffic models. We did some comparison and experiments, and at last we chose the ARIMA model for our traffic prediction

(3) With the network models, making prediction for the network traffic one month ahead, and comparing it with the actual traffic.

(4) Considering the actual operation of Chinanet-Tianjin Network, we make some valuable recommendations.

2. TRAFFIC MEASUREMENT

Measurement is the basis of modern science. It offers fundamental parameters and basis for the analysis and prediction for the computer and network system and other researches. It also offers a method for verifying the results of analysis and prediction [2,3]. Because of the restrictions of network measurement technology, for a long time, network researchers can't get adequate amount of actual network traffic data for their research. In recent years, it is possible for the network researchers to measure and analyze vast traffic data as the development of the network measurement technology increases.

We will introduce some basic measurement principles and technologies here.

As other physical object's measurement, network measurement also needs some measuring method. At first we must define the characteristics of the measured object,

and then set down measuring method and at lastly analyze and verify the measured results.

To carry out network traffic measurement, we need insert a probe into the measured system. The probe may be hardware or software, or it could be the combination of both hardware and software. In the process of measurement, some measured parameters can be directly gotten, for example: CPU's utility, how many data being transferred at some moment in Ethernet. There are some measured parameters can't be directly measured, but they can be gotten by indirect method. For example, A=B+C, A and C can be directly measured, B=A-C, then we can get B.

In the process of measurement, some measured parameters' magnitude is so small that it exceeds the precision bound of the measurement system. For these parameters we can use the method that we can magnify the measured parameter. For example, when we need precisely measure a transferred time transferring a data packet between two given hosts, we can properly control the measurement circumstance, and get the time for one thousand data packets transferred.

With the above methods, we did a long-term measurement for the 2163 traffic of Chinanet-Tianjin. Figure 1 is the measurement of 2163 traffic varying with the day. These data are from January 21st of 2000 to December 20th of 2000.

In figure 1, the abscissa denotes time, and its unit is date, the y-axis denotes the total time users use, and its unit is minute. From this figure, we can see that although the 2163 traffic has some fluctuation, but the 2163 traffic increased much quickly, and it continues the trend. This is because that the ISP can be accessed conveniently.

3. BUILD NETWORK TRAFFIC MODEL

To further analyze the traffic's characteristics and predict it to instruct the network plan, we should build the network model [3]. Time series model is a kind of statistical model broadly used in network traffic analysis. Time series are a series of observation value gotten in the term of the sequence of time. Many data are in the term of time series: A packet's observation according to second in Ethernet, MPEG's bit number per second, error times' observation per second of wireless network packets, and so on. It can be seen as a time series we measured above according to times, so we can apply some time series model to the traffic model and we can build the ARIMA network traffic model.

Time series include linear model of stationary time series: AR (Auto Regression), MA (Moving Average), ARMA model and linear model of unstable time series — ARIMA model. For the periodicity of the data we measured, we build the network model using ARMIA model with periodicity. The raw periodogram figure of 2163 traffic is shown as figure 2.

In figure 2 the abscissa represents frequency and y-axis represents energy. We can see from this figure that:

(1) There is a peak when the frequency is about 0.14,which is called main frequency. From this we can infer that there is some periodicity in this network traffic and this period can be gotten from formula: 1/0.14=7, that is to say that the period of this network traffic is one week. This is in accordance with the actual situation.

(2) There is another peak when the frequency is about 0.28,which is called second harmonic. From this we can get another period which is 1/0.28=3.5. This second harmonic is formed because of the asymmetry of network traffic in the seven days period. This is owing to that the traffic of Monday and Sunday is not be symmetry and so is the traffic of Friday and Saturday and so on. The traffic in Saturday and Sunday is far below the traffic in workday.

From above we have made certain the period of the analyzed network traffic, then we will build some proper network model using ARIMA model with periodicity in Splus. Figure 3 is the parameter diagnostic figure of 2163 traffic.

4. TRAFFIC PREDICTION

It is very important for the resource distribution and control to predict the network traffic 's future varying trend. It has been proved that under some condition, the average value of the time series in statistical balance is equal to the statistical average value. Prediction means that we predict the statistical parameters of the whole system from the history values of the time series in mathematics. The prediction of time series plays a very important role in the production and people living, and it is broadly applied in the dynamic system research domain and the time series research domain [4]. Via the effectively observed value of some time series in time t, we can predict its

future value in time $t+t$, and thus build the optimized foundation for the resource distribution and control. In the following, we will predict the future values of the 2163 traffic and compare it with the actual traffic using the model we have build above. These are shown in figures 4 and figure 5.

Figure 4 started with zero and figure 5 started with non-zero, and we can see the latter one as the magnification of the former figure. In both figures, we will use three hundreds days' traffic and predict the future thirty days' value and we will give you the actual curve and the predictive curve separately.

From figure 4 and figure 5 we can see that our prediction on the whole tallies with the actual traffic's varying trend. So we can inspect the actual traffic according to the prediction to better instruct the distribution and control of resources and then we can make a plan for the network capacity.

For the whole year's prediction of 2163, we predict three hundreds' values ahead according to the three hundreds' history values (from January 21st 2000 to November 15th 2000). At first we will give the whole year's prediction and then we will compare the actual traffic with the predicted traffic value and give you the comparative curve, as figure 6 and figure 7 show.

5. CONCLUSION

Now we have predicted the 2163's whole year traffic and compared our predicted values with the actual traffic values from January 2001 to February 2001. From the above figure we can see that our predicted values are comparatively in accordance with the actual values. In the next year, the network traffic quickly increased compared with the corresponding period of the last year. As it has the advantage such as high speed, cheap and so on, the 2163 traffic has the distinct increase trend compared with the corresponding period of the last year, about double of it. For the sake of satisfying the development of the network traffic, we should adjust the corresponding capacity and extend it.

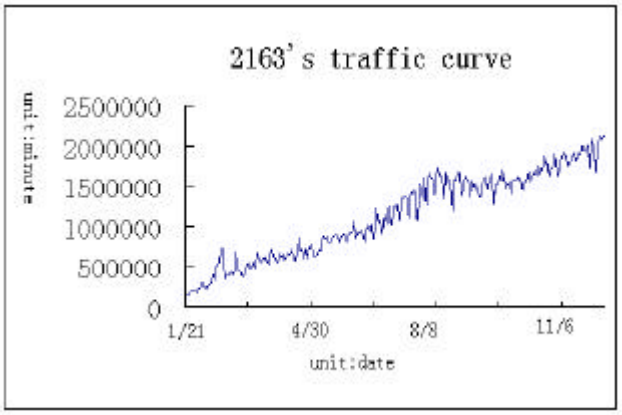

Figure1: 2163's actual traffic curve

6. ACKNOWLEDGMENT

This research was supported in part by the National Natural Science Foundation of China (NSFC) under grant No. 69872025, the Natural Science Foundation of Tianjin under grant No. 993800211 and a Cisco university grant.

REFERENCES

[1] Huizhang et al. " Service Discipline for Guaranteed Performance Service in Packet-switching Networks" Proceedings of the IEEE, 83(10), Oct 1995.

[2] R. L. A. Cottrell and J. Halperin et al. "Effects of Internet Performance on Web Response Times" http://www.slac.stanford.edu/comp/net/wan-mon/ping/correlation.html" , Dec. 1996

[3] Thomas M. Chen, " Network Traffic Measurements and Experiments" *IEEE Communication Magazine*, 2000,5

[4] " Object-oriented network performance prediction Jerry et al. Proceedings of the Conference on Winter Simulation", 1992

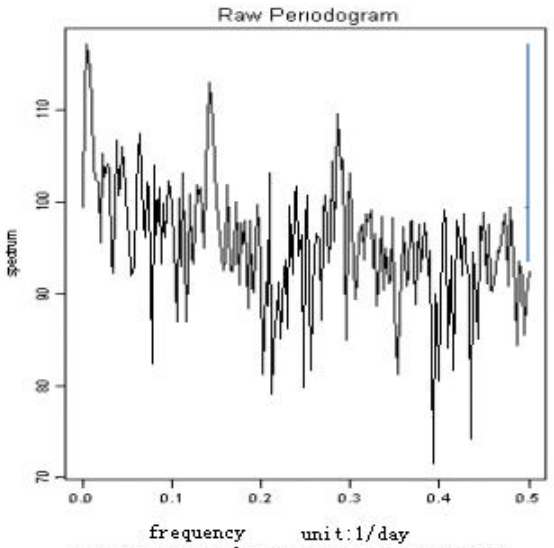

Figure 2:2163's traffic periodogram

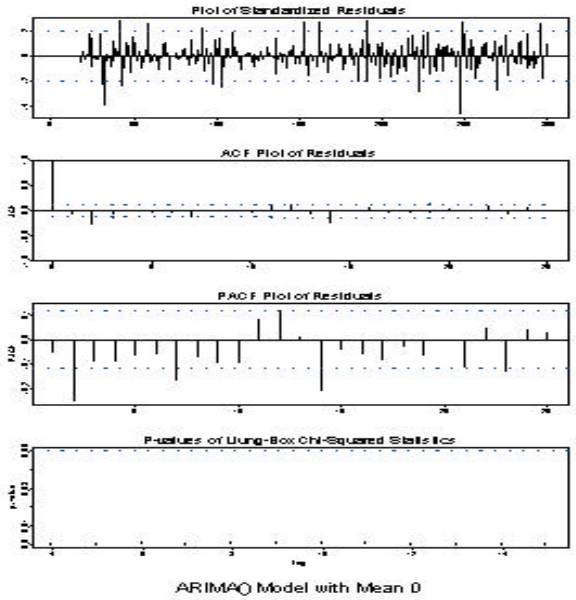

Figure 3: 2163 traffic's parameters diagnostic figure

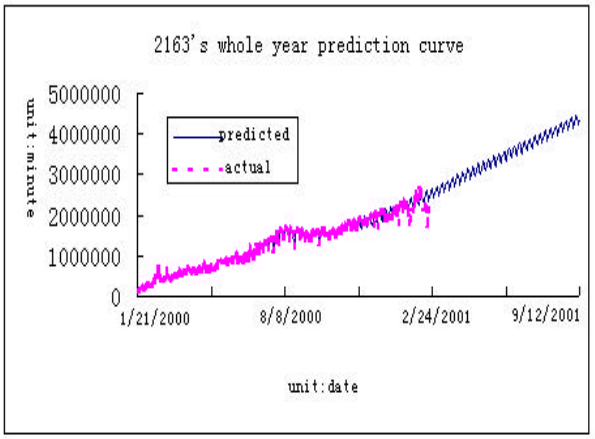

Figure 6:2163's whole year prediction curve

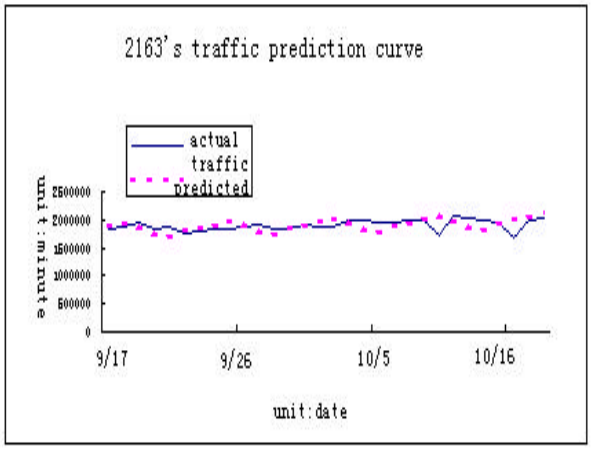

Figure 4:2163's traffic prediction curve

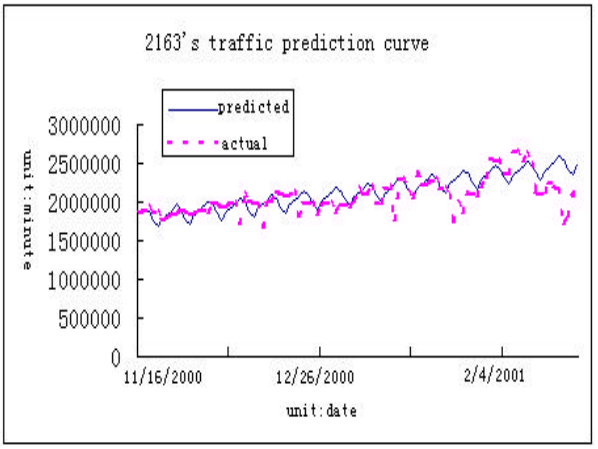

Figure 7:2163's traffic prediction curve

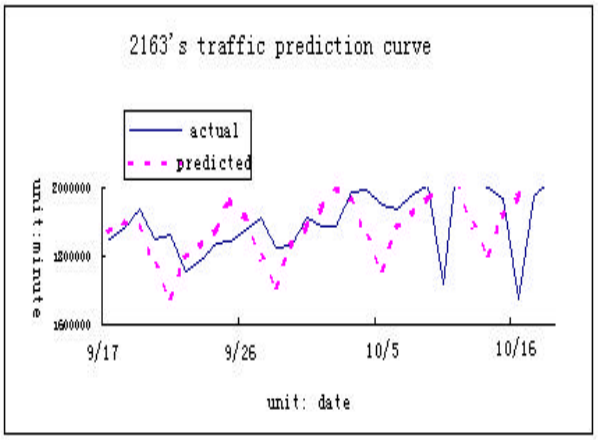

Figure 5:2163's traffic prediction curve