Traffic Prediction for Mobile Network using Holt-Winter's Exponential Smoothing

Denis Tikunov
Ritsumeikan University
1-1-1 Noji-higashi, Kusatsu, Shiga, Japan
E-mail: denis@gin.ics.ritsumei.ac.jp

Toshikazu Nishimura
Ritsumeikan University
1-1-1 Noji-higashi, Kusatsu, Shiga, Japan
E-mail: tnt@is.ritsumei.ac.jp

Abstract: **This paper presents a technique for evaluation of future radio traffic of circuit switched services in Erlang for near-term outlook using Holt-Winter's exponential smoothing. The proposed traffic prediction technique relies on analysis of traffic data on cells. We propose how to classify traffic data for prediction. The result from a trial in history data of commercial GSM/GPRS network is presented. It shows that predicted traffic is good fitted with original one within our classification. The technique should be used in automated overload warning and express estimation of future traffic for capacity planning.**

1. INTRODUCTION

Mobile networks are faced rapid changes, because of new technologies, new services, and increasing number of subscribers. The capacity planning and overload warning became crucial important issue. Reducing operational costs should be achieved by using overload warning for mobile networks.

The capacity planning is the process of adjusting the capacity of a cell to do work in response to changing or predicted demands. The capacity planning is a part of network planning. The aim of capacity planning is to define the maximum number of transceivers (TRXs) at each base station (BTS) and required capacity of the cell and the base station site at each physical location [6].

Areas with high traffic (e.g. downtown) have a priority for the radio capacity planning and overload warning, due to fast traffic growth, high number of subscribers and high profit from area. One of the common ways for evaluating current state of the radio network is comparing the traffic values on cells with a threshold defined by Erlang B formula in order to evaluate a blocking rate [1].

In our paper, we propose to use the Holt-Winter's exponential smoothing for short-term traffic prediction of circuit switched services as an express method for overload warning and capacity planning. The intuition behind proposed technique is to use trend and seasonal character of traffic for prediction future values. The future traffic should be compared with thresholds (which define a particular blocking rate by Erlang-B formula) in order to warn about future overload. Moreover, the proposed method can also be applied to traffic prediction of circuit switched services in the 3G or multi-systems networks since Erlang model is suitable for circuit switched traffic. The main objectives for designing technique are good prediction accuracy and easiness of implementation.

This paper is organized as follows. In Section 2, we present previous efforts at traffic prediction. Section 3 presents classification of traffic's patterns based on analysis of real traffic statistics. In Section 4, we introduce concepts of Holt-Winter's exponential smoothing method to traffic prediction. The experiment's results and discussion are presented in Section 5. We conclude in Section 6.

2. RELATED WORK

Network models for traffic prediction can be found in References [10, 11, and 12]. An explicit mobile prediction is proposed in Reference [10], where the authors present a formula for the transition probability of a mobile station in adjacent cells in future periods. The computation of blocking probabilities is presented in Reference [11], where authors propose multi-rate multi-class Erlang loss formulas for each class of traffic. In Reference [12], authors present probabilistic handover estimation for a mobile station in a neighbor cell using recent handover history. Many proposed approaches predict handovers or blocking probabilities by using statistical data. Some of the methods predict the time and next cell of handover based on a per-user basis, which requires huge computational resource. On the other hand, all history-based schemes are difficult to store and to update traffic histories for huge amount of cell. Moreover, statistics along short time can effectively reflect changes because of traffic variations. Hence, long collected data can generate more stable predictions, but the prediction might not reflect fast variations which are results from accidents and slowing downs. Moreover, simulating of a large-scale network (radio access network) is typically not feasible.

In the Reference [3], Holt-Winter's exponential smoothing method is used for aberrant behavior detection in IP network. The proposed method captures the history of the network traffic and predicts the future traffic rate in the form of a confidence band using additive Holt-Winter's method. When the variance of the network traffic continues to fall outside of the confidence band, an alarm is raised. The reference [4] introduces low-order AutoRegressive Integrated Moving Average (ARIMA) model for weekly approximation of IP network. The main disadvantage of ARIMA modeling approach is that it is a complex technique, which requires a great deal of experience. Although it often produces satisfactory results, those results depend on the researcher's level of skill because ARIMA model building is an empirical treatment of systematically identifying, evaluating, diagnosing, and predicting time series. This cycle of model building continuously requires the control of expert analysts, which can be very expensive. The major advantages of Holt-Winter's methods are their simplicity and low cost, since the methods do not require maintenance of high skilled expert analyst. Moreover, the Holt-Winter's methods can be applied almost automatically.

3. PATTERNS IN TRAFFIC DATA

Analysis of traffic datasets shows different behavior of traffic in cells. For classification of traffic data, we selected main factors of traffic behavior:

- Day variations. For example, during daytime cell has busiest hours, but in nighttime it has small value of traffic.
- Week variations. For example, cell has high traffic during weekdays and small traffic during weekends or vise versa.
- Accidental variations. For example, national holidays, accidents etc.

An accidental variation of traffic has random or rare characteristics and can not be forecasted by statistical analysis, because of lack of information in dataset. Day and week variations have systematic cycle behavior and may be forecasted by statistical analysis. Real traffic consists of all three variations, but for prediction, we should select datasets with one variation, otherwise algorithm for prediction will be complex and forecasted results will have large errors. We believe that accidental variations should be excluded from traffic prediction in overload warning or cell planning etc.

According to traffic values, datasets were classified into three types:

- High intensive traffic cell. The average traffic has large values - more than 50% of CS traffic channels - TCH load at 2% blocking rate during whole day or whole week (e.g. airport area, railway station area). This type of cell has not exactly specified week variations in behavior, traffic values are large during day variations.
- Medium intensive traffic cell. The average traffic has large values during some hours in a day or several days in a week – more than 50% of CS TCH load at 2% blocking rate (e.g. urban and suburban area). For such cells, traffic patterns have easily defined day and week variations.
- Low intensive traffic cell. The average traffic has small values – less than 50% of CS TCH load at 2% blocking rate (rural area). The cell may have day or week traffic variations.

For overload warning and capacity planning, cells with high and medium intensive traffic should be considered as the most interesting and profitable. So, cells with high and medium traffic intensity have to be used for traffic prediction. Figure 1 shows an example of the high intensive traffic cell, statistics given during 16 days with period 24 measurements per one day. During all observed time, statistics have large values without week deviation. In the shown example, curve has "V" shape i.e. traffic have two busiest hours per day.

Figure 2 presents an example of the medium intensive traffic cell during 14 days with period 24 measurements per one day. The moving average line shows that traffic depends on weekday i.e. for first five days (workday), traffic value has large values and during next two days (weekend), traffic value became small. Then during next week, traffic value large like on previous workdays and during second weekend, traffic also became small as weekend before. The traffic on the cell presented on the figure has two components in short term: daily and weekly variations.

The analysis of statistics shows that the cells' traffic behavior can be divided into three types described above. For traffic prediction, we selected cells with high and medium intensive traffic. Prediction was not performed for cells with low intensive traffic because the cells do not need capacity planning (the cells do not need extension of number of TRX).

The traffic behavior on a cell consists of trend and season components. Season component may have two parts: day and week portion of deviation. The traffic patterns during the

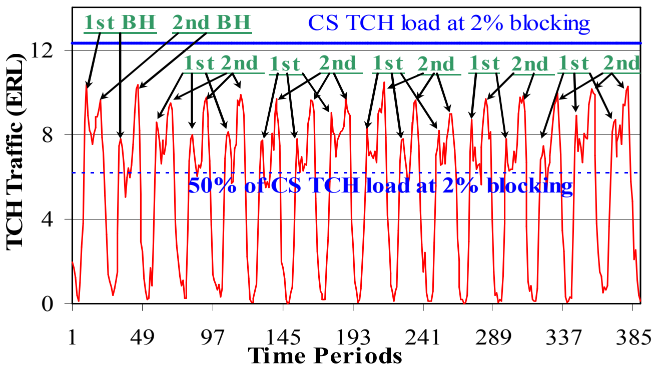

Figure 1. Cell with high intensive traffic.

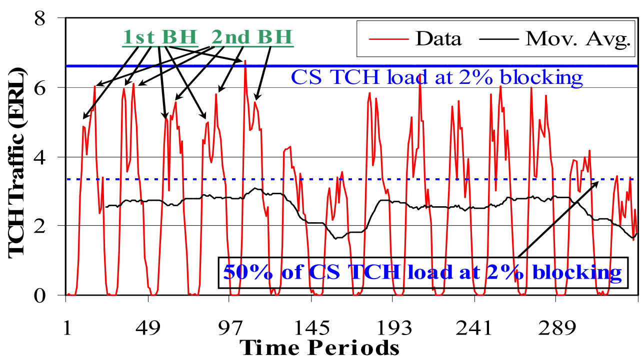

Figure 2. Cell with medium intensive traffic.

same period among different seasons have the same characteristics and may have the trend component.

In case of cells with high intensive traffic, all available statistics were selected for prediction. This is because such type of cells has not weekly variation in the statistics. For daily variation, we selected 24 statistics variables as a period for one season, since we suppose that cell behavior is always the same for the same periods.

In case of cells with medium intensive traffic, the statistics of days with high traffic (e.g. workdays) were selected for prediction. The traffic statistics of days with small traffic (e.g. weekend) were discarded for prediction. During statistics analysis, we found that the period for week deviation did not always same with formula 5+2 (five workdays plus two days of weekend), sometimes the week deviation were 6+1 and others. It can be considered as result of subscribers' behavior and cell's location. Suitable statistics for prediction should be selected for each cell separately based on comparison with average level. Figure 2 shows that traffic of 6th and 7th days are different from previous five days. For the first five days, we can easily define "V" shape (two Busiest Hours), but for 6th and 7th days, we cannot define such shapes. The moving average line also shows that the average level of traffic for 6th and 7th days deceased. For such case, statistics during 6th, 7th, 13th and 14th days were discarded as they are not suitable data for prediction. All other statistics data were used for prediction.

4. PREDICTION METHOD

The Holt-Winter's forecasting technique for seasonal time series is basically a quantitative forecasting method that uses mathematical recursive functions to predict the trend behaviour [5]. It uses a time series model to make predictions assuming that the future will follow the same pattern as the past.

The technique has two types: one for additive seasonality (additive Holt-Winter's method) and one for multiplicative seasonality (multiplicative Holt-Winter's method) [7]. The seasonality of the data is multiplicative if the effect of the season increases with an increase in the level of the time series. It is additive if the seasonal effect does not depend on the current level of the time series and if the level can simply be added or subtracted from a forecast.

The equations of additive Holt-Winter's method are:

$$S_t = \alpha\left(\frac{Y_t}{I_{t-L}}\right) + (1-\alpha)(S_{t-1} - b_{t-1}), \quad I_t = \beta\left(\frac{Y_t}{S_t}\right) + (1-\beta)I_{t-L},$$
$$b_t = \gamma(S_t - S_{t-1}) + (1-\gamma)b_{t-1}, \quad F_{t+m} = (S_t + mb_t)I_{t-L+m}. \tag{1}$$

Where S_t is a smoothed series at the time t, Y_t is an observed value at time t, I_t is a seasonal component at the time t, b_t is a trend smoothing component at the time t, L is number of periods in one season, F_{t+m} is a forecast for m periods ahead, m is number of predicted periods, α (alpha) is an overall smoothing parameter, β (beta) is a seasonal smoothing parameter, γ (gamma) is a trend smoothing parameter.

The equations of multiplicative Holt-Winter's method are:

$$S_t = \alpha\left(\frac{Y_t}{I_{t-L}}\right) + (1-\alpha)(S_{t-1} - b_{t-1}), \ I_t = \beta\left(\frac{Y_t}{S_t}\right) + (1-\beta)I_{t-L},$$
$$b_t = \gamma(S_t - S_{t-1}) + (1-\gamma)b_{t-1}, \ F_{t+m} = (S_t + mb_t)I_{t-L+m}. \tag{2}$$

Parameters α, β and γ must be any numbers between 0 and 1 not include [7, 8]. The parameters values of α, β, γ are selected to minimize Root Mean Squared Error (RMSE) for either additive Holt-Winter's method or multiplicative one. The parameters are calculated for each cell separately.

The initial conditions for additive Holt-Winter's method:

$$S_t = \sum_{t=1}^{L} \frac{Y_t}{L}, \ b_t = 0, \ I_t = Y_t - S_t \ \textit{for } t = 1 \textit{ to } L. \tag{3}$$

The initial conditions for multiplicative Holt-Winter's method are given by:

$$S_t = \sum_{t=1}^{L} \frac{Y_t}{L}, \ b_t = 0, \ I_t = \frac{Y_t}{S_t} \ \textit{for } t = 1 \textit{ to } L. \tag{4}$$

RMSE defines the error between forecast results and real history data. We select particular method with set of parameters for traffic prediction between additive Holt-Winter's exponential smoothing and multiplicative one. A main criterion for selecting the methods is a smallest value of RMSE.

5. RESULTS AND DISCUSSIONS

In our experiment, the proposed traffic prediction technique was applied to history statistics of commercial GSM/GPRS network. The statistics consist of traffic values of each cell in mobile network collected by NMS (Network Management Systems) hourly, during 25 days. We divided statistics into 2 parts: first part contains data during first 18 days and considered as the input data for prediction; second part consist of rest 7 days and considered as the data for checking result of prediction. Totally, one cell has dataset consist of 432 values and forecast consists of 168 values of traffic. For experiment, we selected cells with high intensity of traffic patterns and medium ones as discussed in Section 3. In our experiment, we predicted traffic for 77 cells. The thresholds for classification traffic patterns (50% of CS TCH load at 2% blocking rate) and configurations of cells are presented in Table 1. The CS TCH load at 2% blocking rate calculated with the Erlang B formula. Note that in our experiment, cells also have the fixed packet data traffic channels dedicated only for packed switched services - GPRS service. We processed statistics using Holt-Winter's exponential smoothing described in Section 4.

We estimated accuracy of the proposed technique by Normalized Root Mean Squared Error (NRMSE) which calculated by following equation.

$$NRMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\frac{Y_t - \hat{Y}_t}{Y_t}\right)^2}. \qquad (5)$$

Where Y_t is the real value of data at time t, $\hat{Y}_t$ is the predicted value at time t, n is total number of fitted forecast values.

Table 1. The classification thresholds relative cell's configurations.

TRX, TCH per cell	CS TCH load at 2% blocking, Erlang	Classification threshold, Erlang
1, 5	1.66	0.83
2, 12	6.62	3.31
3, 19	12.33	6.17
4, 26	18.38	9.19
5, 32	23.73	11.87
6, 40	31	15.5

We calculated the average NRMSE for period from 9 a.m. until 9 p.m. In our experiment, the average NRMSE is 0.1585; the best result of NRMSE is 0.0807 and the worst one is 0.3033. Table 2 presents the relation between range of average real traffic on a cell and NRMSE. The cells with large value of average traffic have better predicted result (indicated by small NRMSE) than cells with small traffic.

Table 2. Prediction Error based on Range of Real Traffic

Range of Average Real Traffic (Erlang)	NRMSE
< 5	≥ 0.28
5 - 10	0.1607
10 - 15	0.1259
15 - 20	0.1228

The analysis of NRMSE, predicted traffic value, real traffic value for each hour of daytime periods separately, gave us the following conclusions.

- The predicted traffic is good fitted with original one.
- During periods with value of traffic less than 5 Erlang, the average NRMSE is always more than approximately 0.28.

The result of our experiment shows that in our settings, we could not predict traffic less than 5 Erlang by Holt-Winter's exponential smoothing precisely.

Figure 3 presents the frequency of the cells, which have specific average level of NRMSE. The NRMSE has distributed between values 0.12 and 0.18 and upper level 0.26 for the most cells. The distribution of NRMSE relative to average traffic on cells is shown in Figure 4.

We analyzed the result of the experiment for the "best" cell (cell with NRMSE=0.0807) and the "worst" (cell with NRMSE=0.3033) one in detail.

The "best" cell has α=0.136, β= 0.001 and γ=0.230. Figure 5 presents the real traffic, predicted traffic, and two bounds (upper and lower) for confidence interval 95% on the "best" cell. From visual inspection of presented plot, one can conclude that proposed technique behaves well for this particular cell. In order to be able to quantify the quality of the prediction for the "best" cell, we calculated the relative prediction error. Figure 6 presents the relative error between predicted traffic and real traffic for 6 day from 9 a.m. until 9 p.m. on the cell. Negative error assumes that the predicted traffic was higher than the real one. As it can be seen from the Figure 6, the relative error fluctuates with time, but is centered on zero. More specifically, the average relative prediction error was 3.29% for 6 days on the "best" cell. This means the proposed technique is neither to underestimate nor to overestimate the future traffic.

The "worst" cell has α=0.651, β= 0.001 and γ=0.999. Figure 7 presents the real traffic, predicted traffic, and two bounds (upper and lower) for confidence interval 95% on the "worst" cell. Figure 8 presents the relative error between predicted traffic and real traffic for 5 day from 9 a.m. until 9 p.m. on the cell. As it can be seen from Figure 7 and Figure 8, the predicted traffic is usually higher than real one.

The average relative prediction error was -24.42% for 5 days on the "worst" cell.

In both cases, predicted traffic has the same "V" shape of curve (two busiest hours during one day) like real data.

The prediction error (i.e. the difference between predicted value and real one) has a crucial importance for capacity planning because values of predicted traffic are intended for estimation of future network capacity. If predicted values are larger than real, it may lead the wrong decision that some particular cell need optimization and planning. Estimation of the reasonable prediction error (i.e. the error which enough

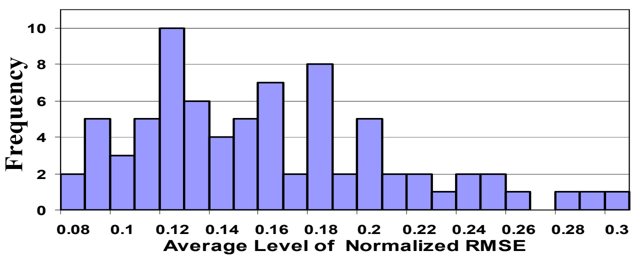

Figure 3. The frequency of cells with specific average levels of Normalized RMSE.

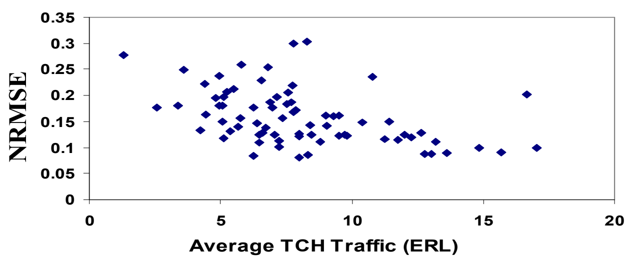

Figure 4. The distribution of NRMSE relative to average traffic.

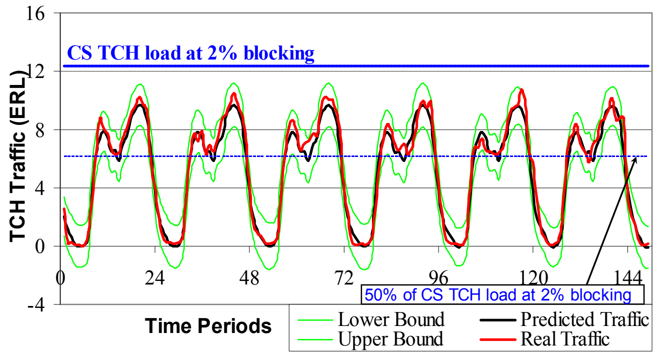

Figure 5. The six days prediction for the "best" cell.

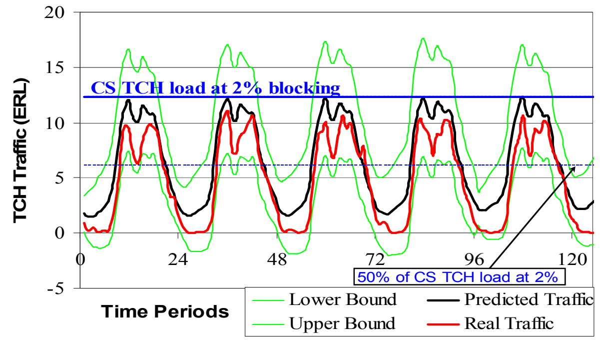

Figure 7. The five days prediction for the "worst" cell.

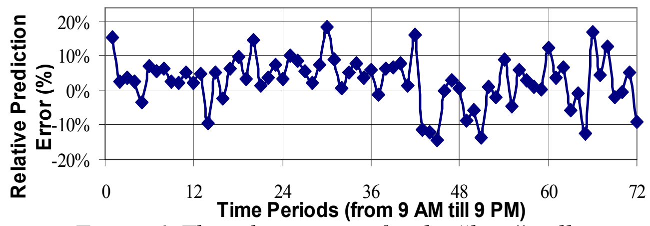

Figure 6. The relative error for the "best" cell.

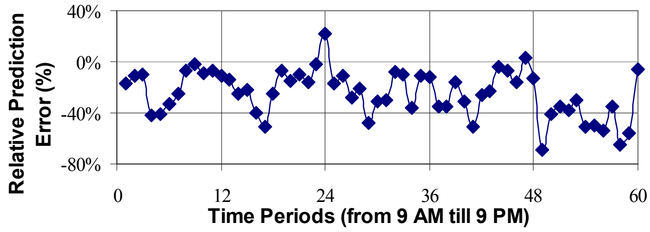

Figure 8. The relative error for the "worst" cell.

for taking right decision) should be executed in capacity planning process. Traffic prediction for long term (e.g. one month) may not give desirable preciseness for capacity planning because dynamic network developing and introducing new services may change traffic behaviour dramatically. Therefore, in proposed traffic prediction technique, we focus on a near-term outlook.

Note that Holt-Winter's exponential smoothing can be used to predict the demand in traffic on a particular cell for one week in the future. Prediction should be used for overload warning and capacity planning of mobile network to identify the time when the predicted traffic will exceed the operational thresholds. The prediction technique should be used with caution. The further ahead in the future we tries to predict, the larger error margin should be allowed.

Although the proposed technique was evaluated on GSM/GPRS network, the proposed technique can be applied even for 3G mobile network for estimation of future value traffic for circuit switched and Grade of Service because Erlang model is suitable for it [6].

6. CONCLUSION

This paper presented a technique for predicting traffic in Erlang on a cell of radio access network. We applied the proposed technique to real GSM/GPRS statistics and got estimation of the technique - the average NRMSE is 15.85%. The analysis of our experiment shows that the higher traffic on a cell has, the less NRMSE is.

Proposed technique is easy to implement, and it should be realized in fully automated tool for overload warning and express traffic estimation.

REFERENCES

[1] T. Halonen, J. Melero, J. Romero: "GSM, GPRS & EDGE Performance: Evolution towards 3G UMTS," John Wiley, 2002.
[2] H. Leijion: "Basic forecasting theories: a brief introduction," ITU, Tech. Rep., November 1998.
[3] J. D. Brutlag: "Aberrant Behavior Detection in Time series for Network Monitoring," Proceedings of the 14th USENIX conference on System administration, pp. 139 – 146, 2000.
[4] K. Papagiannaki, N. Taft, Z.-L. Zhang and C. Diot: "Long-Term Forecasting of Internet Backbone Traffic: Observations and Initial Models," 22nd Annual Joint Conference of the IEEE Computer and Communications Societies, vol. 2, pp. 1178-1188, 2003.
[5] NIST/SEMATECH e-Handbook of Statistical Methods [Online] Available: http://www.itl.nist.gov/div898/handbook/
[6] J. Lempiain and M. Manninen: "Radio interface system planning for GSM/GPRS/UMTS," Kluwer Academic, 2001.
[7] P.R. Winters: "Forecasting Sales by Exponentially Weighted Moving Averages," Management Science, vol. 6, pp. 324-342, 1960.
[8] C. Chatfield and M. Yar: "Holt-Winters Forecasting: Some Practical Issues," The Statistician, vol. 37, pp. 129-140, 1988.
[9] G. Gomez and R. Sanchez (editors): "End-to-End Quality of Service over Cellular Networks: Data Services Performance Optimization in 2G/3G," John Wiley & Sons, 2005.
[10] D.A. Levine, I.F. Akyildiz and M. Naghshineh: "A resource estimation and call admission algorithm for wireless multimedia networks using the shadow cluster concept," In IEEE/ACM Transactions on Networking; vol. 5, issue 1, pp. 1–12, 1997.
[11] K. Mitchell and K. Sohraby: "An analysis of the effects of mobility on bandwidth allocation strategies in multi-class cellular wireless networks," In Proceedings of IEEE Infocom '01, vol. 2, pp. 1005 – 1011, April 2001.
[12] S. Choi and K.G. Shin: "Adaptive bandwidth reservation and admission control in QoS-sensitive networks," In IEEE Transactions on Parallel and Distributed Systems, vol. 13, issue 9, pp. 882 – 897, 2002.