A Predictive Call Admission Control Algorithm for Wireless/Mobile Networks

Renyong Wu, Guangxi Zhu, Xiaofeng Lu, Guoqin Ning
Department of Electronics & Information Engineering, Huazhong University of Science & Technology
Wuhan, Hubei province, P.R.China, 430074

Abstract*-In this paper, a novel method is presented to guarantee the predefined quality-of-service (QoS) requests such as handoff-dropping probability by forecasting the next possible number of handoff dropping calls, using auto regressive moving average (ARMA) model, to update threshold through a close-loop control, while increasing the statistical multiplexing gain (the number of users provided service in a wireless/mobile system at a time) in providing acceptable QoS grades for every connections. We have tested our algorithm through simulations on a wide variety of traffic input. Our simulation results suggest that the forecasting-based approach enables us to achieve a higher level of statistical multiplexing gain while still reliably meeting the target handoff-dropping probability.

I. INTRODUCTION

In a cellular system, user mobility results in handoff calls. To guarantee the handoff-dropping probability, call admission control (CAC) has to be implemented to determine a right amount of reserved resources for handoff calls. There have many research productions on call admission control for wireless/mobile networks.

It had been proved by Ramjee [1] that the permanent guard-channel scheme was optimal to minimize a linear objective function of the handoff-dropping probability and new-call-blocking probability. And several distributed CAC schemes have been proposed to dynamically calculate the possible handoff-dropping probability in order to determine whether to admit a new call request [2], [3]. The two distributed policies outperform a trunk reservation system; however it requires status information exchange among different cells. In [4], the concept of shadow cluster was developed farther, which assumes that detailed knowledge of the users' routes in the system has been pre-declared as a basis for predicting future load requirement, so as to reduce the call-dropping probability by predictive resource reservation. Such predictive information is passed to the base stations of the cells that the mobile resides as well as the neighboring cells, thus the drawback of this scheme is the requirement of the detailed track information and the signaling involved for each call setup. Further more, a tentative shadow cluster needs to be implemented for every handoff call as well as every new call when a mobile is handed off to other cells. To simplify these distributed algorithms, Epstein and Schwartz have developed two schemes, multimedia one-step prediction [5] and independent multi-class one-step prediction [6] respectively, which require only basic knowledge of the users' mobility parameters and the state information of adjacent cells. The two methods use one-step prediction of the overload probability in the home and neighbor cells in deciding whether to admit a new user into a multi-class cellular system, while the latter additionally ensures the predefined call blocking probability profile. In both algorithms, the base station controller in the home cell and each of the adjacent cells only need to compute the predicted capacity using a single prediction time instant, however the signaling exchange among cells will be still involved for each call.

As what Zhang have summarized in [7], most of the existing CAC methods try to model the factors that impact the resource demands and then derive the demands from the model of the impacting factors, while not model the demands of handoff calls directly. Trying to cope with the complexity of modeling the impacting factors forces existing methods to impose stringent assumptions on how these factors behave, interact with each other, and impact the amount of resources required by handoff calls. The proposed methods [7] use only local information to model the instantaneous amount of resource demands directly, based on Wiener processes and time series analysis. Its major drawback is that the amount of reserved resource should be set to the upper limit of the confidence interval, and the limited simulation for non-stationary handoff call arrivals assumes that the arrival rate of handoff in "training period" is much more larger than the value in future period, which certainly leads to bandwidth overestimation. In [8], two time series-based models for predicting handoff load: adaptive exponential smoothing technique and autoregressive integrated moving average are proposed, and the predicted value is used directly. Both schemes try to model the amount of resource demands and forecast the future value based on the assumption of stationary time series. In fact, the practical patterns of time series may be time variable.

In this paper, a new resource prediction and reservation

1-4244-0063-5/06/$20.00 ©2006 IEEE

methods is presented to overcome some of the limitations of existing methods by forecasting the next possible instantaneous amount of handoff dropping calls directly, while meeting the target handoff-dropping probability (denoted as $P_{D\text{-}tar}$) and minimizing the blocking probability. The scheme monitors the state of cellular system at each interval, i.e., arrive rate, the number of handoff dropping calls and the handoff-dropping probability, and forecast the next possible number of handoff dropping calls. The proposed methods use local information only.

The rest of this paper is organized as follows: The forecasting-based dynamic guard channel scheme is proposed in Section Ⅱ. Simulation results are provided in Section Ⅲ. Finally, concluding remarks are given in Section Ⅳ.

II. FORECASTING-BASED DYNAMIC GUARD CHANNEL SCHEME

Based on distinct QoS requirements, the services can be classified into three categories [9]: real-time streaming, real-time block transfer, and non-real-time services.

Handoff-guaranteed service is adopted to represent real-time streaming application in this paper; it is usually necessary for service provider to maintain a minimum bandwidth during its lifetime, the target handoff-dropping probability should be zero. For real-time block transfer service, it can tolerate some extent delays and is categorized as Handoff-prioritized service; the target handoff-dropping probability is fulfilled in statistical meaning. Non-real-time applications do not need a minimum bandwidth to set up a connection. No call admission control or resource reservation is required here, thus it is beyond the discussion scope of the paper.

We suppose that the guaranteed resources are reserved one to two intervals in advance to simplify the simulation model for handoff-guaranteed service. The assumption can be implemented for soft handoff deployed for next-generation multimedia wireless networks. For the handoff-prioritized service class, aggregate resources are reserved to maintain a reasonably low target handoff-dropping probability.

In our scheme, there are three types of real-time traffic listed in the decreasing order of priorities: the handoff-guaranteed handoff (HGH) call, the handoff-prioritized handoff (HPH) call, and the new call. The resource reservation for HGH calls is guaranteed by admitting HGH calls unconditionally. For HPH calls, a pre-reservation is achieved by the proposed dynamic guard-channel scheme.

In this section, a measurement-based dynamic guard-channel scheme is presented for the handoff-prioritized services to take into account the correlation of future handoff dropping number with present and past handoff dropping number, and it is different from traditional schemes in that the Markov chain is not adopted. Formally, an ARMA (p, q) process is a classical approach for modeling stationary time series that for each t satisfies

$$X_t = c + \phi_p X_{t-p} + \theta_q \varepsilon_{t-q} \tag{1}$$

where c, ϕ_i and θ_i are unknown parameters, $\{\varepsilon_t\}$ are white noise with mean zero and variance σ^2. When q = 0, X_t is correlated with many previous values, denoted by an auto regressive model AR (p). In practice, small order AR models (p=1,2, or 3) are often sufficient.

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + \varepsilon_t \tag{2}$$

In this paper, the cellular system is measured at regular intervals (every second in this paper), and the proposed scheme works as follows. At each measurement instant, the current and two latest handoff dropping number of HPH calls are recorded. From the three records, we use a simple AR (3) to forecast the next possible handoff dropping number as following:

$$\widetilde{H}_D(n+1) = \phi_1 H_D(n) + \phi_2 H_D(n-1) + \phi_3 H_D(n-2) \tag{3}$$

Where $H_D\,(n\text{-}i)(i=0,1,2)$ represents the current and two latest dropping number of HPH calls, $\widetilde{H}_D(n+1)$ denotes the forecasted value of dropping number in next interval.

Only a single reservation threshold (denoted as Th in this paper) is deployed to control resources reservation for HGH and HPH calls, which all include three traffic classes in the simulation. There are two types of approaches used to obtain the handoff-dropping probabilities: modeling based and measurement based. The modeling-based approaches use theoretical models to deduct the dropping probability via mathematical analysis by assuming some parameters of the system model. In general, a simple theoretical model cannot approximate a real network system perfectly without making some unrealistic assumptions. Furthermore, if the CAC problem is modeled as a queuing system, usually the results that can be obtained are those that hold in the average case and transient analysis becomes cumbersome and sometimes impossible. The measurement- based approaches use observed network state parameters, obtained by some measurements, to achieve the object such that the handoff-dropping probability is ensured to be less than a target value in statistical meaning, thus usually have light computational complexity and can adapt themselves to the changing conditions of practical networks.

Let $H_N\,(n\text{-}i)(i=0,1,2)$ denote the corresponding three numbers of HPH calls, so the dropping probability (denoted as P_D in following parts of this paper) is defined as

$$P_D(n) = \sum_{i=0}^{2} H_D(n-i) \Big/ \sum_{i=0}^{2} H_N(n-i) \tag{4}$$

Our dynamic scheme seeks to maximize Th to maximize the statistical multiplexing gain in providing acceptable P_D. Note that our design moving of dynamic guard-channel scheme is based on the simple idea: Given the traffic input being stable (e.g., following Possion distribution) and CAC system has entered 'steady' state, the reservation threshold Th will usually be an appropriate value and P_D will be no larger than $P_{D\text{-}tar}$ ($P_D \leqslant P_{D\text{-}tar}$). Thus the basic idea is: Whenever handoff dropping happens at the sample instant, it is obvious that we should decrease the value of Th so as to maintain the handoff-dropping probability a reasonable small value. The feedback value to update Th will be determined in the light of the difference between P_D and $P_{D\text{-}tar}$. The scheme is described in Fig.1, where *a*, *b*, *c* and *d* are four factors to map the forecasted dropping number of next instant into bandwidth resources value and are functions of $H_D(n-i)(i=0,1,2)$, capacity and average arrival rate of HPH calls.

update $\widetilde{H}_D(n+1)$ and P_D

if handoff dropping happens

if $P_D > P_{D\text{-}tar}$

$Th = Th - a * \widetilde{H}_D(n+1)$

else

$Th = Th - b * \widetilde{H}_D(n+1)$

else

if $P_D < P_{D\text{-}tar}$ & $\sum_{i=0}^{2} H_D(n-i) > 0$

$Th = Th + c * \widetilde{H}_D(n+1)$

if $P_D < P_{D\text{-}tar}$ and $\sum_{i=0}^{2} H_D(n-i) = 0$

$Th = Th + d$

Fig.1. Forecasting-based guard-channel algorithm

III. NUMERICAL RESULTS

In this section, we use the stochastic self-driven discrete-event model, i.e., a network model of a single cell with bandwidth capacity C_T, to evaluate the performance of the proposed algorithm and compare to the measurement-based Timer (capital letter used to show difference) CAC scheme [10]. In the Timer CAC scheme, the time counter Timer is adopted to avoid frequent updates of Th. If and only if $P_D < P_{D\text{-}tar}$ till the Timer expire, the Th is increased to improve channel utilization. Otherwise, we renew the Timer. Here Timer is set to be 120. The selection of new call to be blocked from the new calls or the handoff call to be rejected from the handoff calls is random.

In next-generation multimedia wireless networks like Wideband-CDMA (W-CDMA), each call request may declare a certain service class and call priority as defined by the 3rd Generation Partnership Project (3GPP) [11], and each class corresponds to a feasible series of quality of service (QoS) levels ranging from the best QoS level to the worst QoS level. Here, the best QoS level represents *Maximum bitrate* and the worst one represents *Guaranteed bitrate* [11]. We assume there are three types of real-time multimedia traffic, for example, video, audio, and voice, and each type includes three service classes, if there have enough bandwidth resources, each of them requires four, two or one unit bandwidth(s) from the network, respectively. Even under the worst condition, we have to allocate 1.4, 0.7 or 0.48 unit bandwidths to them. Here, decimals imply that connections can be allocated less than one resources unit. It is practical especially in packet switching networks. From the best QoS level to the worst QoS level there have 25 uniform middle grades. Here, a Hopfield neural network is used in simulation to allocate bandwidth resources adaptively and fairly among the admitted calls (it is beyond the scope of this paper). Among the generated new calls, handoff-guaranteed and handoff-prioritized calls, we randomly select 50% as voice, 25% as audio, and the remaining 25% as video applications. To simplify the simulation model, we assume that both the cell residence time and call holding time are exponentially distributed with the same mean η.

In this paper, some system parameters were set as follows: The total bandwidth resource capacity of the cell is $C_T = 100$. For the new call service, $\lambda_N = 0.5$ call/s, $\eta_N = 100$ s. For the handoff-guaranteed service, the average arrival rate $\lambda_G = 0.05$ call/s, $\eta_G = 25$s. For the handoff-prioritized service, $\lambda_P = 0.5$call/s, $\eta_p = 50$s. The target handoff-dropping probability $P_{D\text{-}tar}$ of the handoff-prioritized service was 0.01 through the simulation. At the beginning, C_T is selected as the initial value of Th for a given cell. ϕ_i (i =1,2,3) in (3) are fixedly set to 0.9, 0.05, and 0.05 respectively(or other values to give a large enough weight to current dropping number of HPH calls and ensure the sum to be one).

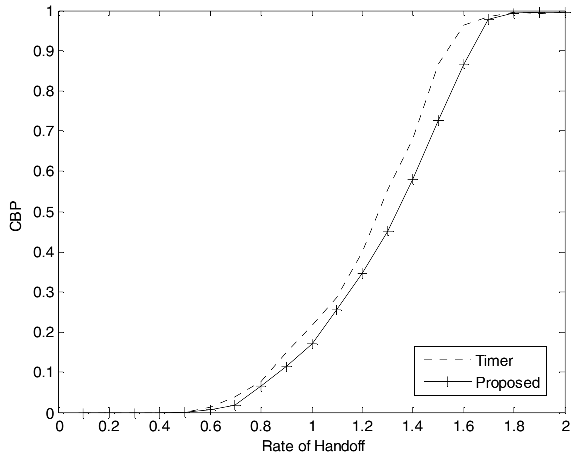
As we described in Fig.1, after we have forecasted the possible dropping number of the next instant (next time unit), the four factors, i.e. *a*, *b*, *c* and *d,* will map the dropping number into bandwidth resources value to update call admission threshold. In this simulation, *a* (=0.765) is the average minimum bandwidth value for a call, and the other three factors are calculated as follows:

$$b = \alpha \cdot \lambda_P \qquad \alpha < 1 \tag{5}$$

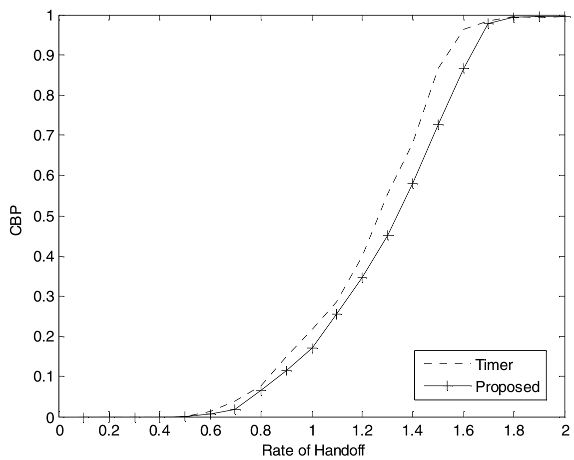
$$c = \beta \cdot \frac{1}{\lambda_P} \qquad \beta < 1 \tag{6}$$

$$d = \gamma + \rho \frac{1}{\lambda_P} \qquad \gamma < 1 \tag{7}$$

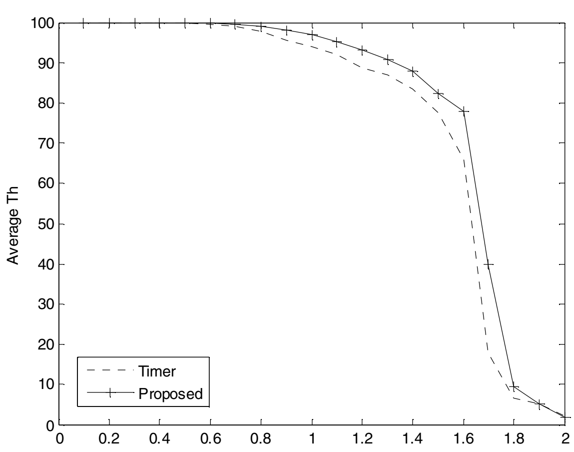
Here, given the traffic input following stable, simulations have shown that α, β, γ,and ρ in (5)-(7) should always be set values between 0.01 and 0.1 to obtain good enough performance, here they are set as 0.01, 0.01, 0.01, and 0.01 respectively.



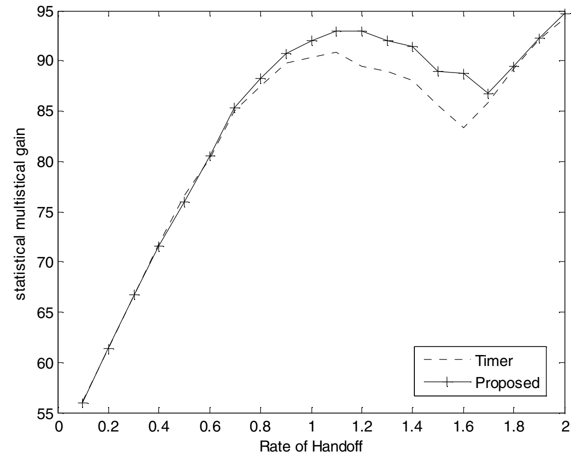
(a) Handoff-dropping probabilities



(b) New-call-blocking probabilities



(c) Average value of Th



(d) Statistical multiplexing gain

Fig.2. Performance under different arrival rate (call/s) of handoff-prioritized service.

Now, we present how our CAC can guarantee QoS to users and compare to the scheme in [10] (denoted as Timer scheme below) in terms of the handoff-dropping probability P_D, the new-call-blocking probability P_B, Th, and statistical multiplexing gain (the number of users provided service in a wireless/mobile system at a time).

To investigate the system performance of the two algorithms under different traffic loads, we conducted a set of simulations with different arrival rates of handoff-prioritized service, while maintaining all other settings as those in the above scenario. Fig. 2(a) shows that the two handoff-dropping probabilities increase slowly, and remain at the target with increasing arrival rate before the rate equals 1.7, and then the two P_Ds increase linearly because Th is close to zero from then on. Fig. 2(b) shows that the new-call-blocking probability increases with increasing arrival rate of handoff-prioritized service. It is noted that P_B of our CAC scheme is notable lower than that of the Timer scheme before the arrival rate equals 1.7. Fig. 2(c) shows that the average threshold (Th) of the Timer algorithm is notable lower than that of our CAC scheme, which can explain in part why the new-call-blocking probability of the Timer algorithm is notable higher than that of our scheme. It means that more resources are reserved in the Timer algorithm to maintain a lower handoff-dropping probability than our scheme, while increasing new-call-blocking probability. Fig. 2(d) shows that the statistical multiplexing gain of the network in the Timer scheme is remarkably less than that of our CAC scheme under moderate and high load conditions. It is worth noting that both of the average statistical multiplexing gains increase till the arrival rate λ_P reaches 1.1, where the Erlang load of the cell system is 106.75, which is nearly 81% of the system capacity (in this simulation, the system capacity equals C_T / a).

When λ_P equals 1.6 (or 1.7 for proposed method), the average statistical multiplexing gains reach their minimums. The drop of statistical multiplexing gain in Fig. 2(d) reveals that the admission control algorithm has to "waste" more resources to guarantee dropping probability. From then on, both of the two gain values increase because the two threshold values start being close to zero. It implies that most of resources are used to satisfy HPH calls from then on.

IV. CONCLUSION

In this paper, we proposed a forecasting-based call admission control for wireless/mobile networks. Our scheme monitors the state of cellular system at regular intervals, i.e., the number of handoff dropping calls, and forecast the next possible number of handoff dropping calls. Based on the forecasted value and the handoff-dropping probability at each interval, the reservation threshold Th is updated and then the call admission decisions can be made according to the value of Th at the next instant.

One limitation is that we have not studied the relations between the four parameters (*a*, *b*, *c* and *d) and system* capacity C_T and average arrival rate of HPH calls. It will be considered in the future work.

REFERENCES

[1] R.Ramjee, D.Towsely,and R.Nagarajan, "On optimal call admission control in cellular networks,"wireless networks, vol.3,no.1, pp.29-41,1997.

[2] M. Naghshineh and S. Schwartz, "distributed call admission control in mobile/wireless networks," IEEE J. Select. Areas Commun., vol. 14, pp.711-717, May 1996.

[3] S. wu, K. Y. M. Wong, and B. Li, "A new distributed and dynamical call admission policy for mobile wireless networks with QoS guarantee," in Proc. 9th IEEE Int. Symp. Personal, Indoor and Mobile Radio Communications, vol. 1, Sept. 1998, pp. 260-264.

[4] D. Levine, I. Akyildiz, and M. Naghshineh, "A resource estimation and call admission algorithm for wireless multimedia networks using the shadow cluster concept," IEEE/ACM Trans. Networking, vol. 5, pp. 1-12, Feb. 1997.

[5] B. Epstein and M. Schwartz, "QoS-based predictive admission control for multi-media traffic," in Broadband Wireless Communications, M. Luise and S. Pupolin, Eds. Berlin, Germany: Springer-Verlag, 1998, pp. 213-224.

[6] B. Epstein and M. Schwartz, "Predictive QoS-based Admission Control for multiclass traffic in cellular wireless networks," IEEE J. Select. Areas Commun., vol. 18, pp.523-534, March 2000.

[7] T. Zhang et al. "Local predictive resource reservation for handoff in multimedia wireless IP networks," IEEE J. Select. Areas Commun., vol. 19, pp.1932-1941, Oct. 2001.

[8] Kelvin L. Dias et al. "Predictive call admission control for all-ip wireless and mobile networks," ACM LANC'03 OCT. 3-5, 2003.

[9] T.C. Kuok, "Residential Broadband Internet Services and Applications Requirements," IEEE Communications Magazine, vol.35, pp.76-83, June 1997.

[10] L. Huang, S. Kumar, and C.-C.J.Kuo. "Adaptive resource allocation for multimedia QoS management in wireless networks," IEEE Trans. Veh. Technol., vol.53, pp. 547-558, March 2004.

[11] 3GPP, QoS concept and architecture, Technical specification TS 23.107 (march 2002), http://www.3gpp.org.