Call Admission Control for QoS Provisioning in Multimedia All-IP Wireless and Mobile Networks

Kelvin L. Dias, Stênio F. L. Fernades, and Djamel F. H. Sadok

Abstract— **This paper proposes a novel call admission control (CAC) mechanism for wireless and mobile networks. Our proposal avoids per-user reservation signaling overhead and takes into account the expected bandwidth to be used by calls handed off from neighboring cells within a prediction interval through the Trigg and Leach Method (an adaptive exponential smoothing technique) [13]. Our scheme is compared through simulations with the ACR (Adaptive Channel Reservation) [10] scheme, a dynamic reservation-based proposal that uses GPS systems to extrapolate users' movement and to trigger reservations in the next predicted cell. The simulation results show that our proposal provides the best performance in terms of handoff dropping probability and can achieve similar levels of call blocking probability as compared to ACR. In addition, our proposal can grant an upper bound on handoff dropping probability even under very high loads.**

Index Terms— **Wireless and mobile networks, QoS, call admission control.**

I. INTRODUCTION

All-IP wireless and mobile networks represent the convergence of two key technologies: Internet and wireless cellular systems. The combination of both technologies suggests that a coming trend will be an increasing demand for IP based wireless/mobile access to traditional and multimedia applications with varying quality of service (QoS) requirements. Fig. 1 illustrates an envisioned scenario with heterogeneous wireless technologies integrated through IP mobility aware protocols (Mobile IPv4/IPv6, Cellular IP, Hawaii, etc.) that will seamlessly interwork with the global Internet [1], [2].

The research effort is especially challenging when dealing with provisioning of quality of service (QoS) guarantees. Users applications may experience performance degradation due to the properties of wireless channels and due to user mobility from handoffs. Handoff in wireless/mobile networks is the mechanism that transfers an ongoing call from the current cell as the mobile station (MS) moves through the coverage area of the system. If the target cell does not have sufficient available bandwidth, the call will be dropped. From the user's point of view handoff dropping is less desirable than the blocking of a new call.

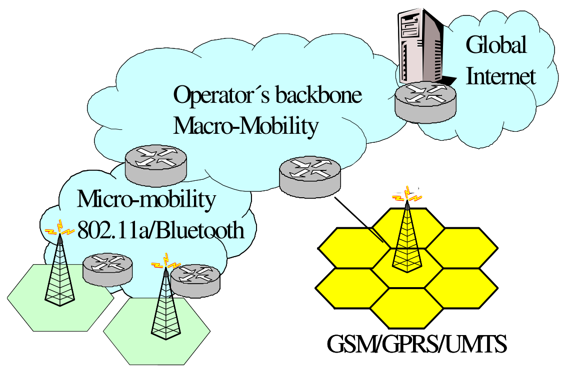

Fig. 1. A scenario for all-IP mobile/wireless networks.

An important component for mobile/wireless networks is the Call Admission Control (CAC) mechanism. It must be used to address the mobility effects, accepting or rejecting new users in the network. CAC schemes not only have to ensure that the network meets the QoS of newly arriving calls if accepted, but should also guarantee that QoS of existing calls does not deteriorate.

On the other hand, Internet frameworks for QoS provisioning rely, basically, on two architectures: Integrated Services (IntServ) [3] and Differentiated Services (DiffServ) [4]. While the IntServ architecture provides strict QoS guarantees through per-user explicit signaling for CAC and reservation using RSVP (Resource Reservation Protocol), it fails in providing the scalability objectives due its reservation based approach. The DiffServ proposal aims at providing less strict QoS guarantees through packet classification at network ingress and differentiation of the treatment according to a set of classes named PHB (Per Hop Behavior), hence offering better network scalability. The Bandwidth Broker (BB) is a network entity proposed for implementing resource management policies in the DiffServ architecture, including the CAC mechanism [5].

Kelvin. L. Dias, Stênio F. L. Fernades, and Djamel. F. H. Sadok are with Centro de Informática, Universidade Federal de Pernambuco, 50732-970; e-mail: {kld,sflf,jamel}@ cin.ufpe.br.

In wireless and mobile networks, reservation of resources is more challenging than in wired networks due the scarcity of bandwidth in wireless links. In our opinion, a scalable QoS architecture for wireless/mobile networks should provide CAC schemes that avoid excessive per-user signaling for wireless link reservation purposes.

This paper proposes a novel CAC mechanism for wireless and mobile networks. Our proposal avoids per-user reservation signaling and takes into account the expected bandwidth to be used by calls handed off from neighboring cells within a prediction interval through the Trigg and Leach Method (TL), an adaptive exponential smoothing technique [13]. Our scheme is compared through simulations with the ACR [10] (Adaptive Channel Reservation) scheme, a dynamic reservation-based proposal that uses GPS systems to extrapolate users' movement and to trigger reservations in the next predicted cell. The simulation results show that our proposal provides the best performance in terms of handoff dropping probability and can achieve similar levels of call blocking probability as compared to ACR. In addition, our proposal can grant an upper bound on handoff dropping probability even under very high loads.

The remainder of this paper is organized as follows. In section II, we describe the related research work. Section III gives an overview of the Trigg and Leach Technique for forecasting. We then present the novel CAC scheme in section IV. Performance results are presented in section V. Finally, we present our conclusions in section VI.

II. Related Work

Proposals for CAC in wireless/mobile networks present in the literature can be divided into two categories: fixed and dynamic strategies. Fixed strategies, such as the guard channel (GC) [6] scheme, give preferential treatment to handoff calls reserving a fixed number of channels exclusively for them. The advantage of this strategy is its simplicity because there is no need for the exchange of control information between base stations. However, this scheme is not flexible to handle varying traffic loads, since there is no information about current and neighboring cell's load.

Proposed dynamic reservation strategies [7], [8], [9], [10], [11], [12] extend the basic guard channel scheme according to the estimated handoff call rate derived from the number of calls in the neighboring cells and the mobility pattern of these calls to reserve bandwidth in advance in the next cell or in a group of cells. Resource reservation can be problematic, in general, due to the possibility of poor network utilization due to unnecessary blocking of new users and can get even worse if the reservation are made in several adjacent cells. Furthermore, these schemes imply a large amount of signaling overhead.

The scheme proposed in [7] uses the aggregate history of handoffs in each cell to predict the probability a call will be handed off to a certain neighboring cell. Based on handoff prediction, the number of channels are reserved in advance. Each base station records the number of handoff failures and adjusts the reservation by changing the estimation window size. One problem with history-based schemes is the overhead to develop, store and update traffic histories for the different cells. Furthermore, due to short-term changes (e.g., diversion of traffic due to accidents) and medium-term changes (e.g., traffic re-routing during road constructions), these estimates cannot be fully reliable.

The call admission control proposed in [8] takes into consideration the number of calls in adjacent cells, in addition to the number of calls in the admission cell. The authors developed a theoretical model to compute the requirements for handoff requests in order to maintain a target handoff dropping probability. The proposed model assumes that all bandwidth requests are identical, which is not valid if multimedia services with varying bandwidth requirements are to be supported by the network.

Next, we will describe some existing research that aims at optimizing bandwidth utilization (decreasing call blocking probability), but keeping low levels of dropping probability for handoffs.

In [9] a predictive channel reservation (PCR) scheme based on mobile positioning systems (GPS -Global Positioning System) is proposed. This scheme makes predictive channel reservation for each MS based on its current position and orientation. The reservation is triggered if the MS reaches a certain threshold distance from the next cell. A reservation may be deemed invalid (false reservation) if the MS changes its moving direction. In this case, the cancellation of the reservation must be sent to de-allocate the reserved channel. Furthermore, rather than strictly mapping each reserved bandwidth portion to the MS that made the reservation, all reserved bandwidth is used as a generic pool to serve handoff requests but not new calls. When a MS arrives from a neighboring cell after a handoff, it may use bandwidth from the reserved portion if there is any available. Otherwise, the handoff connection will compete in the free bandwidth portion with other new call attempts.

The ACR (Adaptive Channel Reservation) scheme was proposed in [10] and it is based on the PCR proposal, but it uses a threshold time instead of a threshold distance to trigger the bandwidth reservation in the next predicted cell. The authors argue that using a threshold time permits a better control of the different degrees of mobility to trigger the reservation in the next cell, avoiding waste of bandwidth due to unused reservations. For example, considering a MS located in the overlapping area of two adjacent cells with a very slow moving speed of this MS (close to 0) and requiring a channel for its call. If the PCR scheme is used, two channels (each cell has one channel occupied) will be occupied by this call, one channel is used for communication in the current cell and the other is reserved for this call in the adjacent cell because the threshold distance was reached. Since the MS of this call is almost stationary, the reserved channel, may not be used for the lifetime of this call. Consequently, PCR can lead to under-utilization of wireless channels.

The PCR as well as the ACR schemes introduce a lot of signaling messages for reservation and cancellation of false reservations. Moreover, the reservations can decrease the dropping probability at the expense of increasing the blocking probability, what may give rise to poor network utilization. The use of GPS for predicting user mobility is also advocated in proposals [11], [12].

In summary, while such dynamic reservation-based schemes have demonstrated significant performance advantages over well-engineered guard channels, the per-user dynamic reservation approach place computation and communication burdens on the network's infrastructure, which increases with the numbers of users and handoffs. Hence, the scalability and applicability of such solutions to future micro and pico-cellular networks is not well established.

III. FORECASTING PROCEDURES: EXPONENTIAL SMOOTHING AND VARIANTS

Exponential smoothing methods forecast time series by discounted past observations. They have become increasingly accepted because of their effortlessness compared to their superior overall performance. It is recommended for short-term prediction. Among the simplest methods is the ordinary (simple) exponential smoothing, which assumes no trend and no seasonality whereas Trigg and Leach procedure could be seen as its adaptive approach.

A. Simple Exponential Smoothing

Let Y_t denote a univariate time series. Simple exponential smoothing assumes that the forecast $\hat{Y}$ for period $t+h$ is given by a variable level $\hat{a}$ at period t

$$Y_{t+h} = \hat{a}_t . \tag{1}$$

which is recursively estimated by a weighted average of the observed and the predicted value for Y_t.

$$\hat{a}_t = \alpha Y_t + (1-\alpha)\hat{Y}_t . \tag{2}$$

$$\hat{a}_t = \alpha Y_t + (1-\alpha)\hat{a}_{t-1} . \tag{3}$$

where $0 < \alpha < 1$ is known as the smoothing parameter.

B. Adaptive Exponential Smoothing: Trigg and Leach

In order to assist the selection of α and to improve awareness capability of the predictor, a number of adaptive methods have been recommended in the literature. The most representative and widely used is the Trigg and Leach [13] technique. Its mainly advantage rely on the fact that there is no need to specify the smoothing parameter previously. Trigg and Leach procedure can regulate α whenever a change occurs in the time series basic structure. Let α_{t+1} be the one-step ahead smoothing parameter. So, the prediction in $t+1$ for the level is

$$\alpha_{t+1} = \frac{E_t}{M_t} . \tag{4}$$

where

$$E_t = \beta\varepsilon_t + (1-\beta)E_{t-1} . \tag{5}$$

$$M_t = \beta|\varepsilon_t| + (1-\beta)M_{t-1} . \tag{6}$$

and $\varepsilon_t = Y_t - \hat{Y}_t$ (prediction error at t).

One can easily show that α_{t+1} is between $[-1,1]$. Values close to zero point out a well-controlled prediction system (smaller prediction errors) whereas values near to the unity indicate an out of control prediction system (huge prediction errors). It is important to emphasize that α_{t+1} allow the system to reconcile by not being too reactive to changes. But most importantly, α_t will vary based on variations in the data pattern.

IV. THE PROPOSED CALL ADMISSION CONTROL SCHEME

Our novel CAC estimates the total amount of required bandwidth for future handoff calls using TL. The process for predicting the required bandwidth for handoff calls is local, that is, the base station uses only local information (collected bandwidth due to handoffs) that serves as the input for the prediction method without exchange of messages among neighboring cells to this end. Suppose that a base station knows the amount Ω of required resources for handoff calls at the current time t. The amount of resources required for handoffs $E(\Omega)$ at a future time $t+\Delta t$ can be predicted based on the current Ω and its predicted value from the previous time interval $t-\Delta t$. In order to offer statistical guarantees regarding the worst case handoff dropping probability (HDP) for the next time interval, we may use interval prediction based on confidence intervals and confidence levels. This way, we can determine a level L such that $Prob(\Omega \le L) = 1-HDP$. This level L is called $(1-HDP)*100\%$ upper confidence bound for Ω. This value is given by:

$$\psi = E(\Omega) + Z\gamma\sqrt{\left(\frac{\alpha}{2-\alpha}\sigma^2\right)}. \quad (7)$$

where $Z\gamma$ is the q-quartile of the standard Normal distribution of $N(0,1)$, α is the smoothing parameter, and σ^2 the sample variance.

The novel CAC should determines whether the admission cell has sufficient bandwidth to support the user requirements and takes into account the predicted handoff load for that cell. The following condition must be met:

$$\sum^{N} Bi + B + \psi \leq C . \quad (8)$$

Equation (8) verifies whether the admission cell has sufficient bandwidth to support the new request. *N* is the number of existing connections, *C* is the wireless link capacity and *Bi* is the bandwidth being used by the i^{th} connection in that cell. *B* is the bandwidth required by the newly requested connection. ψ is the upper confidence bound for the expected bandwidth due to handoff calls $E(\Omega)$ for the next prediction interval. For example, if the network operator has to guarantee a maximum target handoff dropping probability of 5%, ψ will be set to the 95% upper confidence bounds of the predicted bandwidth requirements for handoff calls $E(\Omega)$ for the next prediction interval. At the start of each interval, a new ψ is used to control the admission decision. Upon each handoff arrival in a cell, during a prediction interval, the current ψ is decreased by the MS's bandwidth that has arrived until it reaches 0 or a new prediction interval is initiated.

V. Simulation Results

The metrics of interest in this paper are: handoff dropping probability (HDP) and call blocking probability (CBP). The simulated model consists of a cellular network with 19 hexagonal cells in a toroidal way to avoid border effects. The unit of bandwidth is called bandwidth unit (BU), which is assumed to be the required bandwidth to support a voice connection as in [7], [12]. Each cell is assumed to have a fixed link capacity of 100 BUs. The traffic model used is similar to the one used in [7], [12]. Call requests are generated according to Poisson distribution with rate λ (call/cell/second) in each cell. The simulated traffic consists of users with bandwidth requirements of 1 BU (voice) and 4 BUs (video) with probabilities Rvo and 1-Rvo, respectively, where Rvo is also called the voice ratio as in [7]. In our simulations Rvo is set to 0.7, that is, 70% of voice traffic and 30% of video traffic. The lifetime of each call is exponentially distributed with mean 180s [9], [10], [12].

We studied two mobility scenarios: In scenario I, the MS chooses a direction and does not change its path while crossing cells during its call lifetime. The time that a call spends in a cell prior to handoff to another cell (residence time) is exponentially distributed with mean 60s, representing high mobility users that execute three handoffs on average considering the call lifetime of 180s. Upon each handoff a new residence time exponentially distributed is obtained.

The scenario II uses a probability of 70% to dictate changes in the MS's moving direction, that is, the MS can change its new target handoff cell at any moment in time while crossing a cell with a probability of 70%. If a MS's moving direction is changed, a new path is randomly selected among its six neighboring cells as well as a new residence time is chosen using the exponential distribution with the same mean residence time as that of scenario I. Upon each handoff, a new neighboring cell and residence time are chosen again. The offered load is expressed in terms of call requests generated according to a Poisson distribution in each cell (calls/cell/second).

The evaluation presented below refers to simulations using batch means technique with transient removal at the beginning of the simulation and adopting 10 batches, each for 10^4s. For each simulation run, the metrics were periodically collected and then averaged out. The graphs depict average values for the 10 runs. The bars are plotted for a 95% confidence level.

The prediction interval of our proposal (TL) is set to 60s and 10s in the simulated scenarios and the maximum target handoff dropping probability is set to 5%. The ACR's threshold time is set to 60s and 30s as indicated in the labels of the curves.

Fig. 2 and Fig. 3 show the scenario I (MS does not change its moving direction while crossing the cells). The best result for the HDP was achieved by our proposal (TL) over the entire range of the offered loads. However, ACR outperforms TL in terms of call blocking probability (CBP) for lower loads. It is important to note that ACR proposal aims at optimizing network utilization by postponing the reservation of bandwidth resources for handoff in the next predicted cell and, consequently, more users will be admitted into the network. On the other hand, the ACR scheme does not grant an upper bound for HDP in higher load situations as our proposal does. It is also important to note that our proposal achieves better levels of HDP to those of the ACR scheme, without the need for per-user signaling to make advanced reservations of bandwidth in the next predicted cell.

With regard to the ACR's HDP, the best result was obtained when the threshold time is set to 60s. Under higher loads, ACR performance can be degraded due to the postponed approach for triggering the reservation in the next predicted cell (as depicted, ACR-30 obtained the worst HDP). This way, it may be too late to request a reservation in the next cell only when the threshold time is reached.

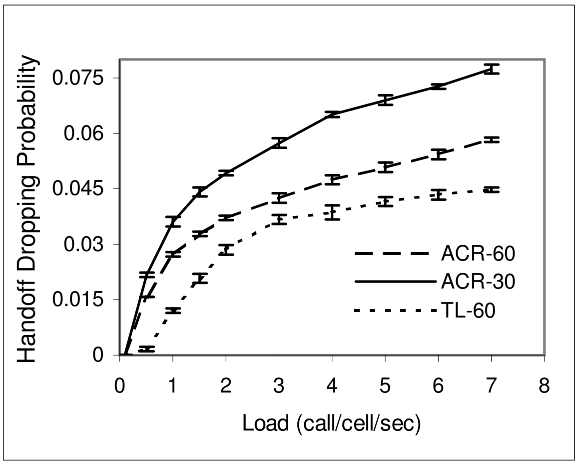

Fig. 2. Handoff dropping probability for scenario I (Deterministic mobility).

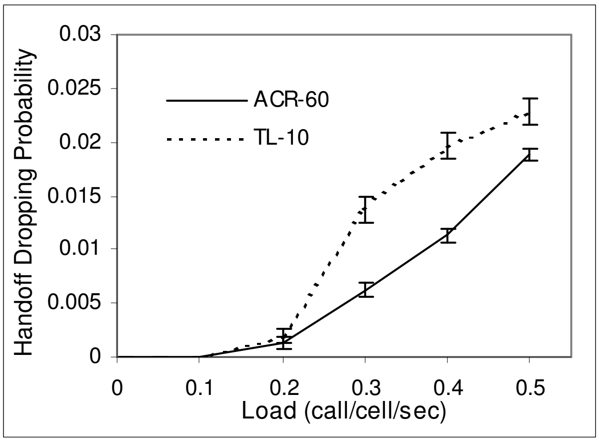

Fig. 4. Handoff dropping probability for scenario I (Deterministic mobility - Lower loads).

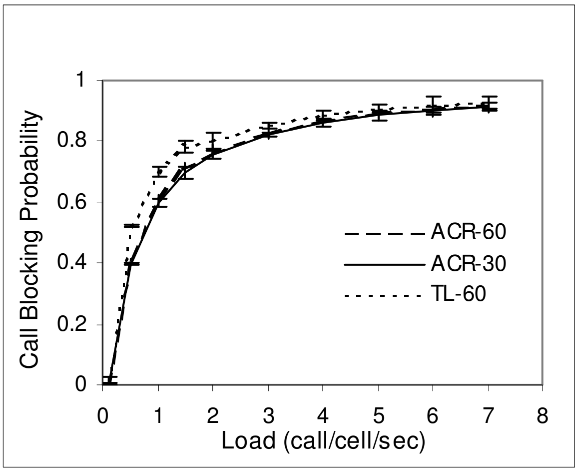

Fig. 3. Call blocking probability for scenario I (Deterministic mobility)

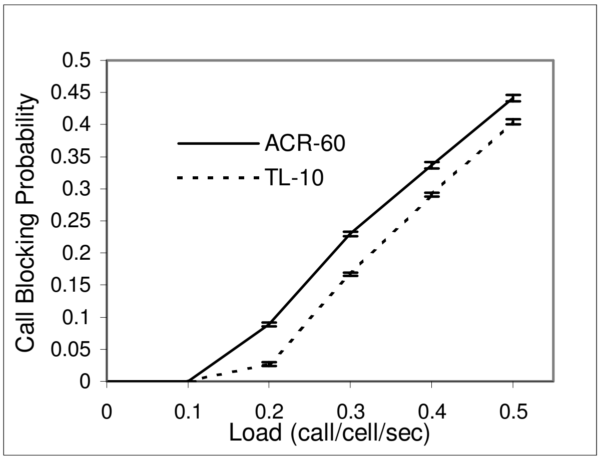

Fig. 5. Call blocking probability for scenario I (Deterministic mobility - Lower Loads)

In Fig. 3, our proposal obtained a slightly greater CBP than ACR´s for lower loads, but depending on the cellular operator's objectives, the TL's configuration parameters can be set to benefit the CBP and network utilization at the expense of a greater, but controlled HDP. To this end, Fig. 4 and Fig. 5 depict an experiment for lower loads using the same settings of the scenario I, except by the value of the TL's prediction interval that is set to 10s. The ACR's threshold time is set to 60s (the best setting in Fig. 2 and Fig. 3). As can be seen, for small values of the prediction interval, TL's CBP can be improved.

Fig. 6 and Fig. 7 depict the simulation results for the scenario II where a MS can change its moving direction while crossing a cell with probability equal to 70% (a new direction is randomly (uniformly distributed) selected among its six neighboring cells). The results are quite similar to scenario I (Fig. 2 and Fig. 3), but ACR obtained an even worst HDP whereas TL's kept below its 5% target. This random mobility scenario deteriorates ACR's HDP mainly due to the changes in MS moving direction that may be followed by the unavailability of sufficient bandwidth in the new cell MS is moving to. It is important to note that ACR signaling increases considerably due to request and cancellation of false reservations. TL scheme can accommodate the fluctuations due to this random mobility without the need for per-user tracking and signaling, providing a good solution without imposing any signaling overhead on the network.

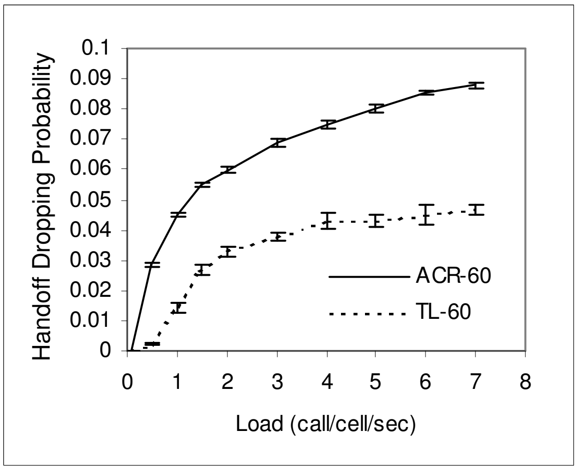

Fig. 6. Handoff dropping probability for scenario II (Random mobility).

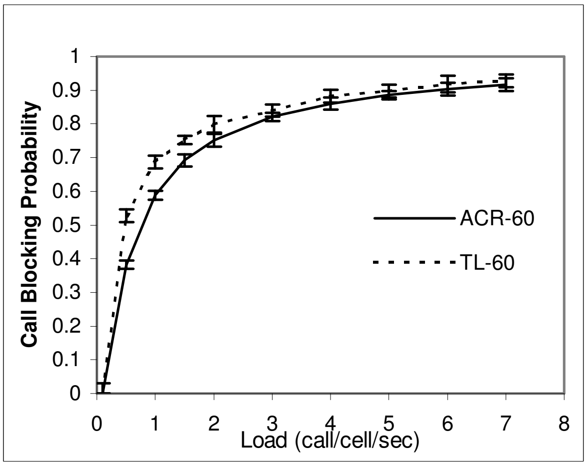

Fig. 7. Call blocking probability for scenario II (Random mobility)

VI. CONCLUSIONS

In this paper, a novel CAC scheme has been proposed for all-IP wireless and mobile networks. Our proposal avoids per-user reservation signaling through a predictive technique executed locally by each base-station. In order to predict the expected bandwidth of future handoffs we utilize an adaptive exponential smoothing method, called Trigg and Leach, which is effortless and does not impose computation overhead on the network elements. In addition, this method does not require a huge amount of saved data to perform forecasting. Our approach can also grant an upper bound on the handoff dropping probability even under higher loads. Our future work is concerned with the proposal of an adaptive algorithm to dynamically adjust the TL's prediction interval to optimize the bandwidth utilization depending on the current network load, mobility scenario and HDP objectives

REFERENCES

[1] L. Bow and W. Leroy, "Toward an all-IP based UMTS System Architecture," IEEE Network, 15(1), Jan-Feb 2001, pp.36-45.
[2] R. Berezdivin et al., "Next-Generation Wireless Communications Concepts and Technologies," IEEE Comm. Mag., 40(3), Mar 2002, pp. 108-116.
[3] R. Braden et al. " Integrated Services in the Internet Architecture: an Overview," IETF RFC 1633, Jun. 1994.
[4] S. Blake et al. "An architecture for Differentiated Services," IETF RFC 2475, Dec. 98.
[5] K. Nichols et al, "A Two-bit Differentiated Services Architecture for Internet. RFC 2638, Jul 1999.
[6] D. Hong, S. Rappaport, Traffic Model and Performance Analysis for Cellular Mobile Radio Telephone Systems with Prioritised and Nonprioritised Handoff Procedures IEEE Tran. Veh Tech., Aug. 1986
[7] S. Choi and K. G. Shin, "Predictive and Adaptive Bandwidth Reservation for Handoffs in QoS Sensitive Cellular Networks," In Proc. ACM SIGCOMM´98.
[8] M. Naghshineh and M. Schwartz, "Distributed Call Admission Control in Mobile/Wireless Networks," IEEE JSAC, 14(4), May 1996. pp. 711-717.
[9] M.H. Chiu and M. A. Bassiouni, "Predictive Schemes for Handoff Prioritization in Cellular Networks based on Mobile Positioning," IEEE JSAC, 18(3), Mar. 2000
[10] Z. Xu et al. " A New Adaptive Channel Reservation Scheme for Handoff Calls in Wireless Cellular Networks," Proc of IFIP Networking2002. pp 672-684.
[11] W.-S. Soh and H. S. Kim, "Dynamic Guard Bandwidth Scheme for Wireless Broadband Networks," Proc. IEEE Infocom'01, Anchorage, AK, Apr. 2001
[12] W.-S. Soh and H. S. Kim, "QoS Provisioning in Cellular Networks Based on Mobility Prediction Techniques," IEEE Comm. Mag., Jan. 2003, pp 86-92.
[13] D. W. Trigg, D. H. Leach, Exponential Smoothing with an Adaptive Response Rate, Operational Research Quarterly, vol. 18, 1967, pp. 53-59