Prediction-Based Admission Control for DiffServ Wireless Internet

Majid Ghaderi, Joe Capka and Raouf Boutaba
School of Computer Science, University of Waterloo, Waterloo, On, N2L 3G1
Email: {ghaderi, jcapka, rboutaba}@bbcr.uwaterloo.ca

***Abstract*— Future wireless Internet will consist of different wireless technologies that should operate together in a consistent way to provide seamless quality of service to wireless users. In this paper, a wireless cellular architecture overlaid with DiffServ domains is considered. We propose a flexible hierarchical framework for admission control based on this architecture which aims to keep the handoff dropping probability below a target level while maximizing the network utilization. The novelty of our proposal is that (1) our prediction-based admission control scheme considers not only intra-domain but also inter-domain handoffs, while (2) it is based on on-line bandwidth requirement prediction, and (3) benefits from different priorities among different service classes to improve the network utilization by accommodating high-priority handoffs at the expense of dropping low-priority calls. Simulation results show that our scheme outperforms the basic trunk reservation scheme with domain and cell-level reservations.**

I. INTRODUCTION

Future wireless Internet will support global roaming across multiple wireless and mobile networks, for example, from a cellular network to a satellite-based network to a high-bandwidth wireless LAN. Supporting quality of service (QoS) in such a heterogeneous wireless network is a challenging problem.

The IETF's differentiated services (DiffServ) framework [1] is an attempt to establish a global QoS architecture. This paper introduces an explicit call admission control (CAC) mechanism to complement the DiffServ framework in providing this global QoS architecture.

At call-level, two important parameters which specify the quality of service are the *call blocking probability* (P_b) and the *call dropping probability* (P_d). Typically, the goal of a CAC scheme is to maintain a prespecified target call dropping probability while minimizing the call blocking probability.

Fig. 1 shows the architecture that we consider for the wireless Internet. In this architecture a cellular network overlaid by DiffServ domains operates as the radio access network. Furthermore, each wireless access network, potentially has its own wireless technology and administrative policies. Without loss of generality, we assume that there is a one-to-one correspondence between administrative domains and DiffServ domains.

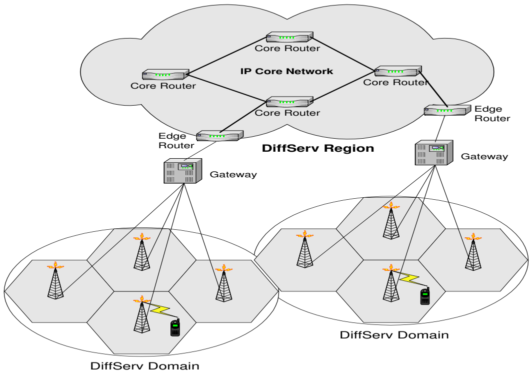

Fig. 1. Wireless DiffServ architecture for Internet.

In this architecture there are two different types of handoff calls:

1) *inter-domain:* between different domains; when there are some service level agreements (SLA) between neighboring domains and there are some service negotiation protocols, mobile users can move from one domain to another while keeping their connections alive.
2) *intra-domain:* in one domain; mobile users can move between neighboring cells inside each domain while receiving the same QoS.

Recently, Cheng and Zhuang [2] have considered DiffServ resource allocation in a domain-based cellular network. Their work is based on the *cell-cluster* concept proposed by Naghshineh and Acampora [3] where each cell-cluster corresponds to a DiffServ domain. In the cell-cluster approach, cells are grouped into clusters and each cluster is associated with a controller. A threshold is set for the whole cluster, then the cluster controller admits new calls as long as the number of occupied channels in the cluster is less than the threshold, and the cell where the new call is generated has a free channel to accept this new call. After admission to the cluster, no further communication is necessary with the cluster controller for handoffs between cells in that cluster. Cheng and Zhuang extend this basic scheme to include guard channels for each cell in the DiffServ domains, however their proposed scheme is still static in that it reserves a fixed number of guard channels for each cell and domain regardless of the traffic load. This can result in network underutilization.

In this paper, we propose a *prediction-based admission control* (PrBAC) for a DiffServ cellular Internet similar to the two-level scheme proposed in [4] and [5]. We extend their scheme to include DiffServ domains and benefit from relative priorities between different service classes. The proposed admission control may drop low-priority calls to accommo-

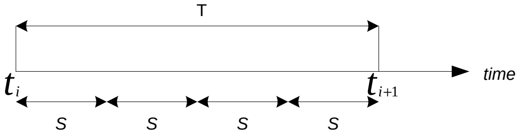

Fig. 2. Sampling mechanism in each control interval.

date high-priority handoffs. We consider both intra-domain and inter-domain handoffs when dynamically adjusting the reservation thresholds. Also, instead of using simple traffic patterns (Poisson arrivals and exponentially distributed call durations), PrBAC uses a *minimum mean square error* predictor (MMSE) [6] to predict the bandwidth requirements in each cell and in each domain. Because we directly predict bandwidth requirements independent of the underlying traffic characteristics, this scheme is very suitable for IP networks where traffic patterns are not Poisson [7].

Although using traffic prediction for admission control is not a new idea [8], [9], the novelty of our approach is that the MMSE predictor is on-line and does not rely on a specific traffic model. For example the FARIMA predictor used in [8] is very complex and can not be estimated using on-line traffic measurements. On the other hand, the ARIMA predictor of [9] is simpler but it is not suitable for *self-similar* Internet [10] traffic prediction as stated in [8] and similar papers. The key idea behind our approach is that we directly predict traffic from on-line measurements without involving any traffic modeling. To use a MMSE predictor we do not need to specify any traffic model.

The rest of the paper is organized as follows: Section II is dedicated to the proposed admission control scheme. In section III we analyze the performance of our scheme in terms of the call dropping probability. Simulation results are presented in section IV and finally, section V concludes the paper.

II. HIERARCHICAL ADMISSION CONTROL

In order to support inter-domain handoffs we use an admission control which is local to the domain, i.e, it does not need any information exchange with neighboring domains. The idea is that regardless of the complexity and overhead associated with distributed schemes, currently there is no standard protocol for exchanging information needed by distributed schemes between neighboring domains.

We extend this local algorithm to handle intra-domain handoffs as well, which leads to a simple and effective CAC scheme. In the gateway (GW), the bandwidth broker enforces DiffServ constraints while interacting with CAC component. In the base station (BS), the CAC component makes the admission decision based on the bandwidth requirements of new calls (can be extracted from their SLAs) and handoffs from both neighboring cells and neighboring domains by a prediction method based on the minimum mean square error predictor.

The proposed scheme, PrBAC, has a periodical control structure. At the beginning of each control interval of length T, each cell predicts the amount of bandwidth required to accept incoming handoffs during the current control interval. Then it reserves this amount of bandwidth to be used exclusively for handoffs until the end of this period. Fig. 2 shows the sampling mechanism used by the control algorithm where each control interval contains sampling points at distance s. The maximal sample taken in each interval is kept as the bandwidth usage for that interval. Below is the notation which will be used throughout this paper.

- B: the available bandwidth in the cell under consideration
- $B_T^u(t)$: the bandwidth allocated to all calls at time t
- $B_H^u(t)$: the bandwidth allocated to handoffs at time t
- B_T^i: the bandwidth usage during the control interval i
- $\hat{B}_T^i$: the predicted value of B_T^i
- B_H^i: the bandwidth required for handoffs that will arrive during the control interval i
- $\hat{B}_H^i$: the predicted value of B_H^i

The PrBAC scheme only takes care of handoffs belonging to expedited forwarding (EF) and assured forwarding (AF) classes. When it is necessary, PrBAC drops best effort (BE) calls in order to accommodate higher priority handoff calls. The only difference between EF and AF treatment is that for AF calls, PrBAC considers only their minimum bandwidth requirements.

A. Minimum Mean Square Error Predictor

To forecast the bandwidth usage for the current control interval, a MMSE predictor of order m is used. Let B denote the random variable to be predicted and $\hat{B}$ the predicted value of B. A MMSE predictor for B is given by

$$\hat{B} = \mathbf{W}\mathbf{B} + \varepsilon \tag{1}$$

where ε is the white noise error with mean 0 and variance σ_ε^2 and $\mathbf{B}$ is a vector of size m of the previous observations of B. In this equation, $\mathbf{W}$ is a weighting vector obtained as follows:

$$\mathbf{W} = \mathbf{\Gamma}\mathbf{G}^{-1} \tag{2}$$

where $\mathbf{G}$ is the autocovariance matrix and $\mathbf{\Gamma}$ is an autocovariance vector starting at lag m,

$$\mathbf{G} = \begin{bmatrix} \rho_0 & \rho_1 & \cdots & \rho_{m-1} \\ \rho_1 & \rho_0 & \cdots & \rho_{m-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{m-1} & \rho_{m-2} & \cdots & \rho_0 \end{bmatrix} \tag{3}$$

and

$$\mathbf{\Gamma} = [\ \rho_m \quad \ldots \quad \rho_1\ ]. \tag{4}$$

The autocovariance function ρ_k can be computed by

$$\rho_k = \frac{1}{m}\sum_{i=k+1}^{m} \mathbf{B}(i)\mathbf{B}(i-k) \tag{5}$$

where m is the order of the MMSE predictor. And finally, the mean squared error of the MMSE predictor is given by

$$\sigma_\varepsilon^2 = \sigma_B^2 - \mathbf{\Gamma}\mathbf{G}^{-1}\mathbf{\Gamma}'. \tag{6}$$

B. Admission Control at Base Station

Assume that a new call request arrives at time $t \in (0, T]$ during the control interval i. Let b denote the amount of bandwidth required by this call. Let $B^r_H(t) = \hat{B}^i_H - B^u_H(t)$ denote the residual amount of bandwidth that we have predicted to be used by the upcoming handoffs until the end of this interval, i.e., during interval $(t, T]$. Also, let $B^f_T(t) = B - B^u_T(t)$ denote the total amount of free bandwidth at time t. The admission control at a base station is described in algorithm 1.

```
if handoff call request then
    if H-DiffServ accepts then
        grant admission
    else
        reject
    end if
else
    if (N-DiffServ accepts) $\wedge$ $(B^f_T - B^r_H) > \alpha b$ then
        grant admission
    else
        reject
    end if
end if
```

Algorithm 1: Admission control at base station.

In this algorithm N-DiffServ is the standard DiffServ module which enforces DiffServ requirements at the ingress point to the domain. H-DiffServ is the same as N-DiffServ except that it may drop BE calls in order to accommodate EF and AF handoffs. Note that B^f_T and B^r_H include bandwidth usage of both EF and AF calls. The tuning parameter $\alpha \geq 1$ is an adaptable parameter that can be adjusted based on the difference between measured call dropping probability and the target P_d.

Algorithm 2 describes an additional admission condition which makes algorithm 1 more conservative. Algorithm 2 is used given that algorithm 1 accepts a new call.

```
if $(B^u_T(t) + b) > \hat{B}^i_T$ then
    if $\hat{B}^{i+1}_T < B$ then
        grant admission
    else
        reject
    end if
end if
```

Algorithm 2: Admission control: conservative condition.

As mentioned earlier, PrBAC uses the maximum amount of bandwidth usage sampled in each control interval to represent the amount of bandwidth required for that interval. If at time t it is found that the bandwidth requirement for the current interval is underestimated, then PrBAC looks ahead at the next control interval. If the predicted bandwidth requirements for the next interval, after accepting this new call, is greater than the total amount of available bandwidth then the new call request will be rejected.

C. Admission Control at Gateway

If a base station accepts a new call request then it will send this request to the domain gateway for second level admission. At this level, GW makes sure to take into account two considerations: (1) DiffServ constraints of domain, and (2) inter-domain handoffs.

The same algorithm we described for BSs can be applied in GWs considering a domain as a virtual cell. For this virtual cell, B_H is the bandwidth required for handoffs from neighboring domains (neighboring virtual cells) and B is the total bandwidth available in the domain. While it is possible to use the predicted values from the domain boundary cells at this level of the PrBAC, a direct prediction is preferred. This method has the advantage of less communication overhead and more accurate predictions due to aggregation. The more traffic is aggregated and smoothed, the more accurate prediction is possible.

Each BS will contact its corresponding GW only for new calls and handoffs from other domains. After admission, no more communication with GW is required for intra-domain handoffs. This reduced communication leads to fast handoff processing which is necessary to prevent QoS degradation at upper network layers (e.g., a delayed handoff process increases the packet loss and delay at network layer).

III. CALL DROPPING PROBABILITY

The accuracy of PrBAC is completely determined by the accuracy of the predictor. For example, if MMSE could predict the exact bandwidth requirements, then PrBAC could guarantee zero percent call dropping while achieving the optimal call blocking. This is not possible in practice.

During the life of a call, a mobile user may cross several cell boundaries and hence may require several successful handoffs. Failure to get a successful handoff at any cell in the path forces the network to drop the call. While the handoff failure probability, P_f, is an important parameter for network management, the probability of dropping a call, P_d, may be more relevant to the user and service provider. Nevertheless, call dropping probability is a system dependent parameter which is particularly affected by user mobility. Let H denote the number of handoffs during the life of a call, then $P_d = 1-(1-P_f)^H$ where H itself is a random variable that depends on several system parameters such as mobile velocity and cell size. In particular, the average probability of call dropping is given by $P_d = hP_f/(\mu + hP_f)$, where μ and h denote the average call completion and average handoff rate.

In the worst case, a handoff will fail when the predicted value $\hat{B}_H$ is less than the actual value B_H (it is possible to accept a handoff even in this situation due to the residual free bandwidth). Therefore, this is an upper bound for handoff failure:

$$\text{Pr(Handoff Failure)} \leq \Pr(\hat{B}_H < B_H) \tag{7}$$

equivalently, to satisfy the target handoff failure probability P_f, it is obtained that

$$\Pr(\hat{B}_H < B_H) \leq P_f \tag{8}$$

To guarantee that $\hat{B}_H > B_H$, we compute an upper confidence interval δ for the predicted value $\hat{B}_H$ as follows:

$$\Pr((B_H - \hat{B}_H) > \delta) \leq P_f \tag{9}$$

therefore,

$$\Pr(\varepsilon > \delta) \leq P_f \tag{10}$$

We know that ε is white noise with normal distribution $\mathcal{N}(0, \sigma_\varepsilon^2)$. Therefore $\delta = \sigma_\varepsilon \Phi(1 - P_f)$, where Φ is the inverse of the standard normal distribution.

IV. SIMULATION RESULTS

For the sake of simplicity there is only one traffic class with fixed bandwidth requirements in the simulated system. This basic implementation is enough to show the performance of the PrBAC scheme in comparison to the traditional trunk reservation scheme. We have also implemented the scheme proposed in [2] which we refer to as the *cell-domain admission control* (CDAC) scheme.

A. Simulation Parameters

Simulations were performed on a two-dimensional cellular system consisting of 19 hexagonal cells. Opposite sides wrap-around to eliminate the finite size effect. Each domain has 19 cells, each cell has 20 bandwidth units available and each new call requires one bandwidth unit. For the sake of simplicity, we have assumed that 12 cells out of 19 cells are bordering cells and 50% of their handoff traffic is due to the inter-domain handoffs. To predict handoffs, MMSE(10) which is a MMSE predictor with history of size 10 is used. Call durations and channel holding times are exponentially distributed with mean 20 and 5 units of time respectively. We also extended the basic CDAC scheme to support inter-domain handoffs. This extended version treats inter-domain handoffs similar to PrBAC. The target call dropping is set to $P_d = 0.01$ and the reservation thresholds for CDAC are 10% and 20% at cell-level and domain-level respectively.

B. MMSE Predictor Evaluation

To evaluate the accuracy of MMSE, the MMSE predictor is compared with several self-similar predictors including fGn [6], FARIMA [6] and GARMA [11] for IP traffic. In this experiment we used an Ethernet traffic trace (`pAug89.TL`) from Bellcore. Although we have used traffic from wired Ethernet, these results should remain valid for any traffic with the same degree of self-similarity (the so-called Hurst parameter for this traffic trace is 0.8). Considering that future cellular networks will be able to carry IP traffic (particularly in indoor environments such as wireless LANs), it seems reasonable to have the same traffic characteristics for wired and wireless IP traffic. For example, Jiang et al. [12] showed that cellular digital packet data traffic exhibits long-range dependencies.

TABLE I
THE ACCURACY OF PREDICTORS.

Predictor	SNR^{-1}
MMSE	0.27
fGn	0.32
FARIMA	0.22
GARMA	0.23

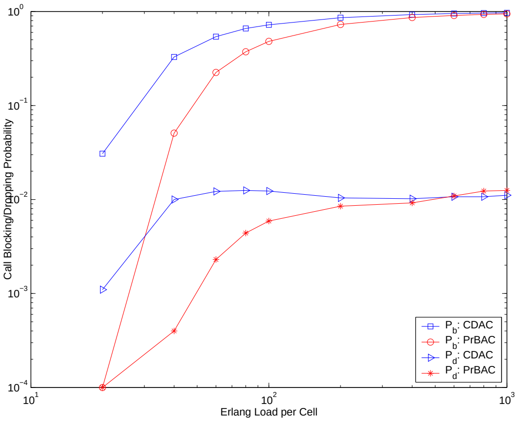

Fig. 3. Call blocking/dropping with Poisson traffic.

Finally, the reverse of *signal to noise ratio* defined as $\mathrm{SNR}^{-1} = \sum \varepsilon^2 / \sum B^2$ is used as the accuracy measure to compare these predictors. The smaller the SNR^{-1}, the more accurate the predictor. Table I summarizes the results of this comparison. In particular, it shows that the accuracy of MMSE is within 5% of the best predictor (FARIMA).

C. Results and Analysis

Simulations were done for a wide range of loads from 20 to 1000 Erlang load per cell. For each load, simulations were done by averaging over 4 samples, each for 10^5 new calls. In addition to call blocking/dropping probability as the QoS measures, call completion probability is also computed as the effective measure for network utilization. Call completion probability is given by $P_c = (1 - P_b)(1 - P_d)$.

Figures 3 and 4 show the QoS and utilization measures for Poisson generated traffic, where inter-arrival times for new and handoff calls are exponentially distributed. Furthermore, each cell of the system experiences the same rate of new arrivals and handoffs. Both schemes can provide a limit for call dropping probability while at the same time the call blocking probability of PrBAC is lower than CDAC.

As mentioned earlier, the performance of PrBAC is determined by the accuracy of MMSE. Although Poisson generated traffic is not a good test case for MMSE predictor, PrBAC performs better than static CDAC due to its dynamic nature. It is interesting to see the performance of both schemes under a different traffic pattern where traffic is more predictable than Poisson traffic.

Figures 5 and 6 show the call blocking/dropping and call completion probability for non-Poisson generated traffic,

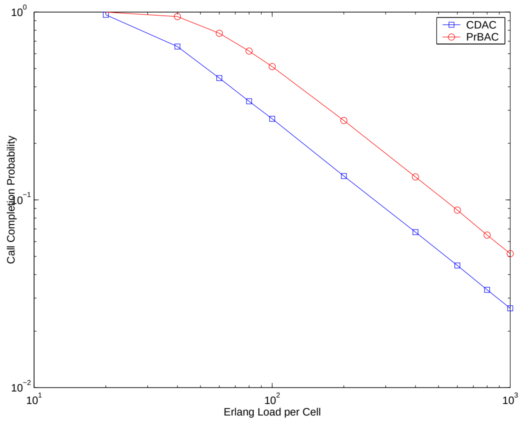

Fig. 4. Call completion with Poisson traffic.

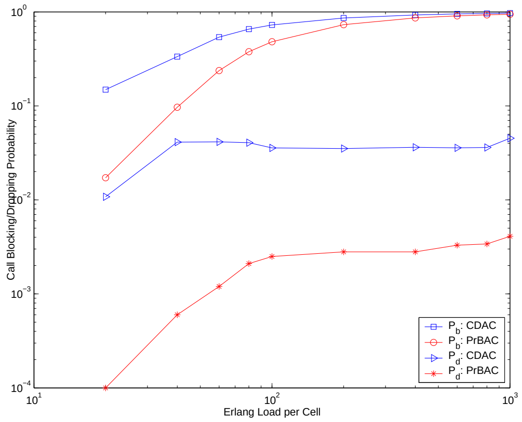

Fig. 5. Call blocking/dropping with non-Poisson traffic.

where the inter-arrival times for new and handoff calls are derived from an autoregressive model of order one, AR(1), with coefficients 0.5 and 0.8. An AR(1) model with coefficient θ is defined by $(X_t - \bar{X}) = \theta\,(X_{t-1} - \bar{X}) + Z_t$ where X is the stochastic process defined by the model, $\hat{X}$ is the mean of the process and Z is the deriving normal variable.

Although real traffic patterns are more complicated than any of the ones used, since the performance of PrBAC is better than CDAC in each tested case we can deduce that PrBAC will perform better than CDAC when presented with real traffic patterns. Of the tested cases the one using the simple AR(1) model shows the greatest performance difference between PrBAC and CDAC.

V. CONCLUSION

In this paper, we studied the application of traffic prediction to address the admission control problem in wireless mobile Internet. We proposed a hierarchical admission control scheme based on forecasting future handoff traffic. The key idea is to use online measurements to predict traffic directly without relying on any particular traffic model. Our scheme was validated using simulations for different types of traffic. The results of this paper can be used to build a simple and efficient admission control scheme for wireless IP networks, where traffic diversity prohibits conventional traffic modeling.

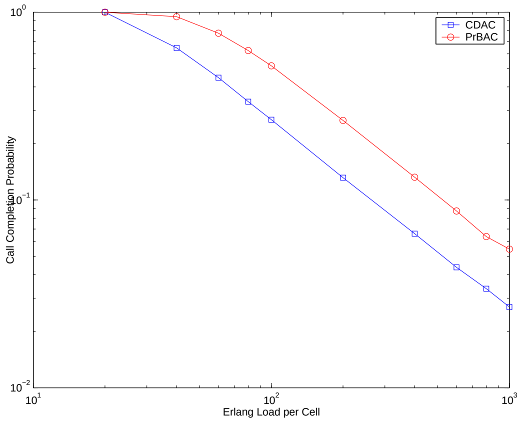

Fig. 6. Call completion with non-Poisson traffic.

This paper focused on constant bit-rate traffic. We are currently investigating the extension of PrBAC to variable bit-rate traffic. One possible approach is to predict the bandwidth usage based on the number of packets transmitted instead of the number of active calls.

REFERENCES

[1] S. Blake, D. L. Black, M. A. Carlson, E. Davies, Z. Wang, and W. Weiss, “An architecture for differentiated services,” IETF, RFC 2475, Dec. 1998.

[2] Y. Cheng and W. Zhuang, “Diffserv resource allocation for fast handoff in wireless mobile internet,” *IEEE Commun. Mag.*, vol. 40, no. 5, pp. 130–136, May 2002.

[3] M. Naghshineh and A. S. Acampora, “Design and control of micro-cellular networks with QoS provisioning for real-time traffic,” in *Proc. IEEE Universal Personal Communications*, San Diego, USA, Sept. 1994, pp. 376–381.

[4] J.-H. Lee, K. Lee, M.-R. Choi, and S.-H. Kim, “Two-level threshold for RCAC (region-based call admission control) in multimedia wireless networks,” in *Proc. IEEE International Conference on Networks*, Oct. 2001, pp. 208–213.

[5] H.-C. Lin and S.-S. Tzeng, “Double-threshold admission control in cluster-based micro/picocellular wireless networks,” in *Proc. IEEE VTC’00*, vol. 2, Tokyo, Japan, May 2000, pp. 1440–1444.

[6] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, 2nd ed. Springer-Verlag, 1991.

[7] V. Paxson and S. Floyd, “Wide-area traffic: The failure of poisson modeling,” *IEEE/ACM Trans. Networking*, vol. 3, no. 3, pp. 226–244, June 1995.

[8] Y. Shu, Z. Jin, J. Wang, and O. W. Yang, “Prediction-based admission control using FARIMA models,” in *Proc. IEEE ICC’00*, vol. 3, New Orleans, USA, June 2000, pp. 1325–1329.

[9] T. Zhang, E. Berg, J. Chennikara, P. Agrawal, J. C. Chen, and T. Kodama, “Local predictive resource reservation for handoff in multimedia wireless IP networks,” *IEEE J. Select. Areas Commun.*, vol. 19, no. 10, pp. 1931–1941, Oct. 2001.

[10] M. E. Crovella and A. Bestavros, “Self-similarity in world wide web traffic: Evidence and possible causes,” *IEEE/ACM Trans. Networking*, vol. 5, no. 6, pp. 835–846, Dec. 1997.

[11] R. Ramachandran and V. R. Bhethanabotla, “Generalized autoregressive moving average modeling of the Bellcore data,” in *Proc. IEEE LCN’00*, Tampa, USA, Nov. 2000, pp. 654–661.

[12] M. Jiang, M. Nikolic, S. Hardly, and L. Trajkovic, “Impact of self-similarity on wireless data network performance,” in *Proc. IEEE ICC’01*, Helsinki, Finland, June 2001, pp. 477–481.