Cost-Function-based Network Selection Strategy in Integrated Wireless and Mobile Networks

Wei Shen and Qing-An Zeng
Wireless and Mobile Networking Laboratory
Department of Computer Science, University of Cincinnati
Cincinnati, OH 45221-0030, USA
Email:{shenwi, qzeng}@ececs.uc.edu

Abstract

Any single type of existing wireless and mobile network cannot provide all types of services, e.g., wide coverage and high bandwidth. An integrated wireless and mobile network is introduced by combining different types of networks to provide more comprehensive services. In an integrated wireless and mobile network, a mobile terminal equipped with heterogeneous network interfaces can connect to different types of networks. Therefore, how to select a desired network is an important issue in the integrated wireless and mobile network. Although some network selection strategies have been proposed, most of them are designed to meet users' needs, such as bandwidth, money cost, or power consumption. Also, the system performance has not been touched. In this paper, we propose a cost-function-based network selection (CFNS) strategy in an integrated wireless and mobile network from system's perspective. We also analyze the system performance of the proposed strategy using theoretical model and simulations. The results show that the proposed network selection strategy affects multiple system parameters, which needs to be handled carefully.

1 Introduction

Wireless and mobile networks have experienced a great success in the past few years. However, any single type of existing wireless and mobile network cannot provide all types of services, e.g., wide coverage and high bandwidth. In order to provide more comprehensive services, an integrated wireless and mobile network is introduced by combing different types of networks. In an integrated wireless and mobile network, a mobile terminal is equipped with heterogeneous network interfaces, i.e., multi-mode terminal. When a mobile user having a multi-mode terminal generates a new call or originating call in an integrated wireless and mobile network, it can select connections among different types of networks based on a network selection strategy. An active multi-mode terminal also can change its connections between different types of networks. Such process of changing the connections between different types of networks is called *vertical handoff*. Obviously, the *network selection* and *vertical handoff decision* are two important processes in an integrated wireless and mobile network. In [1], a mobile user always selects the network with the highest bandwidth from all the available networks during its communication. Therefore, the only concern of the network selection and vertical handoff decision for the mobile user is bandwidth. This is good for the service quality from the user's point of view. In [2], the authors proposed that a data call is kept in the higher bandwidth network as long as possible to increase the throughput and a voice vertical handoff call makes a handoff as soon as possible to avoid handoff delay. Their strategy can achieve higher throughput for data calls and low handoff delay for voice calls. In [3], a network selection strategy that only considers mobile users' power consumption was introduced. In order to maximize the battery life, the mobile user selects the uplink or downlink that has the lowest power consumption from all the available networks. In [4], the authors proposed a policy-enabled network selection strategy, where a mobile user can select the "best" network based on the user's preferences in bandwidth, price, or power consumption.

Although the above network selection strategies have their own advantages, they are all designed to meet individual mobile user's needs. They did not put much attention on the system performance, such as the blocking probability of originating calls and the forced termination probabilities of horizontal and vertical handoff calls. In this paper, we propose a cost-function-based network selection (CFNS) strategy in an integrated wireless and mobile network from system's perspective. The proposed network selection strategy is also analyzed using theoretical model and simulations.

21st International Conference on
Advanced Information Networking and Applications Workshops (AINAW'07)
0-7695-2847-3/07 $20.00

This paper is organized as follows. In the next section, we introduce our proposed system model for an integrated wireless and mobile network. In Section 3, a cost-function-based network selection (CFNS) strategy is presented. Section 4 analyzes the performance of the proposed CFNS strategy using theoretical method. Section 5 provides the performance evaluation of the system through theoretical analysis and simulations. Finally, we conclude the paper in Section 6.

2 System Model

We consider an integrated wireless and mobile system having K different types of networks. We assume that the entire service area of the system is covered by network N_1 that consists of many homogeneous cells and provides a low bandwidth service. Assume that network N_i $(2 \leq i \leq K)$ is randomly distributed in the service area covered by network N_1 and provides a higher bandwidth service than network N_1. Network N_i has limited coverage, which only covers some portion of the entire service area. For example, a cellular network (N_1) covers several WLANs $(N_2, N_3, ..., N_K)$.

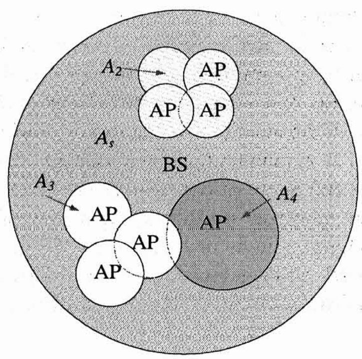

Figure 1. Coverages of different types of networks

For the purpose of simplicity, we focus on one cell of network N_1, where some area is covered by high bandwidth networks as shown in Figure 1. We assume that each cell of network N_i $(1 \leq i \leq K)$ has a circular cell shape. We define the area covered by network N_i $(1 \leq i \leq K)$ as area A_i. The area that is only covered by network N_1 is defined as area A_s, which indicates that it is only covered by a single network. In area A_i $(2 \leq i \leq K)$, mobile users may have more than one connection option. We assume that each cell of network N_i $(1 \leq i \leq K)$ has B_i bandwidth units. It is necessary to clarify that each bandwidth unit is a logic channel.

We assume that mobile users are uniformly distributed in the service area. They move in all the directions with equal probability. The moving speed of mobile user follows an arbitrary distribution with a mean value of $E[V]$. When an active mobile user changes its connection from its current serving network N_i to network N_j (for all i, j), a handoff call (request) is generated in network N_j. If $i = j$, the handoff call is named a horizontal handoff call. If $i \neq j$, it is called a vertical handoff call .

3 Network Selection Strategy

As we introduced in Section 1, existing network selection strategies focused on the user's needs. Our motivation is to design a network selection strategy from system's perspective, which can also meet certain user's needs. In the following, we discuss how our proposed network selection strategy works.

When an originating call is generated, the proposed network selection strategy works as follows: *1)* if there is no free bandwidth unit, the originating call is blocked; *2)* if only one available network has free bandwidth units, the originating call is accepted by that network; *3)* if there are more than one available network having free bandwidth units, all these candidate networks are compared based on network selection strategy and the originating call is accepted by the most desired network.

Since we focus on network selection strategy in this paper, a horizontal handoff call is handled in a traditional way like [5] and a vertical handoff call is handled in the following ways: when an active mobile user moves from area A_i into adjacent area A_j $(i \neq j)$, it changes its connection from network N_i to N_j if network N_j has a higher bandwidth than N_i and there are free bandwidth units in N_j. If area A_j is not covered by network N_i, the mobile user has to change its connection to other available networks. If there is no free bandwidth unit in other available networks, the vertical handoff call is forcedly terminated.

Our proposed network selection strategy prefers an originating call to be accepted by a network with a low traffic load and stronger received signal strength, which can achieve better traffic balance among different types of networks and a good service quality for mobile users. Consequently, we define a cost function to combine traffic load and received signal strength. The cost to use network N_i for an originating call is defined as

$$C_i = w_g \cdot G_i + w_s \cdot S_i, \text{ for } i = 1, 2, ..., K, \qquad (1)$$

where G_i is the complementary of normalized utilization of network N_i. S_i is the relative received signal strength from network N_i. w_g $(0 \leq w_g \leq 1)$ and w_s $(0 \leq w_s \leq 1)$ are the weights that provide preferences to G_i and S_i, respectively.

21st International Conference on
Advanced Information Networking and Applications Workshops (AINAW'07)
0-7695-2847-3/07 $20.00 © 2007 IEEE

The constraint between w_g and w_s is given by $w_g + w_s = 1$. G_i in Equation 1 is defined as B_{if}/B_i, where B_{if} is the available bandwidth units of network N_i and B_i is the total bandwidth units of network N_i. In general, stronger received signal strength indicates better signal quality. Therefore, an originating call prefers to be accepted by a network that has a higher received signal strength. However, it is difficult to compare received signal strengths among different types of networks because they have different maximum transmission powers and receiver threshold. As a result, we propose to use a relative received signal strength S_i to compare different types of networks. S_i is defined by

$$S_i = \frac{P_i^c - P_i^{th}}{P_i^{max} - P_i^{th}}, \text{ for } i = 1, 2, ..., K, \tag{2}$$

where P_i^c is the current received signal strength from network N_i. P_i^{th} is the receiver threshold from N_i. P_i^{max} is the maximum transmitted signal strength from N_i. Note that we only consider path loss in the propagation model. Consequently, the received signal strength P_i^c (in decibel) from N_i is given by $P_i^{max} - 10\gamma log(r_i)$, where r_i is the distance between the mobile user and the base station (or access point) of network N_i. γ is the fading factor that is generally in the range of $[2, 6]$. The receiver threshold P_i^{th} from network N_i is given by $P_i^{max} - 10\gamma log(R_i)$.

If an originating call has more than one connection option, the costs for all the candidate networks are calculated using cost function of Equation (1). The originating call is accepted by a network that has the largest cost, which indicates the "best" network.

In the following, we discuss two special cases of the proposed CFNS strategy, i.e., $w_g = 1$ and $w_g = 0$. When $w_g = 1$, the cost function of Equation (1) only considers G_i. It gives rise to another network selection strategy and we call it traffic balanced-based network selection (TBNS) strategy. This strategy tries to achieve the best traffic balance among different types of networks. In this case, an originating call is accepted by a network which has more free bandwidth units. When $w_g = 0$, the proposed CFNS strategy gives rise to another network selection strategy, i.e., received signal strength-based network selection (RSNS) strategy. It is obvious that the only concern of selecting a desired network is based on the received signal quality.

4 Theoretical Analysis

In this section, we analyze the performance of the system that uses the proposed CFNS strategy. As network N_1 is assumed to have homogeneous cells, we only focus on a single cell in network N_1 as shown in Figure 1, here we call it marked cell. We only consider one higher bandwidth network (N_2). In order to simplify the analysis, we assume that network N_2 only has a cell in the marked cell and it does not cross the boundary of the marked cell. In the following, we first introduce a traffic model and then analyze the system performance using Markov model.

4.1 Traffic Model

A two-dimensional fluid flow model assumes that the mobile users are uniformly distributed throughout the area and each mobile user moves in any directions with equal probability. Using fluid flow model, the average outgoing rate μ_{A_s} of area A_s is given by

$$\mu_{A_s} = \frac{E[V]L_{A_s}}{\pi A_{A_s}}, \tag{3}$$

where A_{A_s} is the area of A_s, L_{A_s} is the length of perimeter of A_s, and $E[V]$ is the average of the moving speed V (random variable) of mobile users. A_{A_s} is given by $\pi R_1^2 - \pi R_2^2$, where R_1 and R_2 is the cell radius of network N_1 and N_2. L_{A_s} is equal to $2\pi R_1 + 2\pi R_2$. Since the A_s outgoing mobile users consist of the mobile users out of the marked cell and mobile users moving into area A_2, we decompose μ_{A_s} as the average cell outgoing rate $\mu_{A_{so}}$ and the average A_{s2} outgoing rate $\mu_{A_{s2}}$. The length $L_{A_{so}}$ of the boundary of the marked cell is equal to $2\pi R_1$ and the length $L_{A_{s2}}$ of the boundary between A_s and A_2 is equal to $2\pi R_2$. Consequently, the average cell outgoing rate $\mu_{A_{so}}$ is given by

$$\mu_{A_{so}} = \mu_{A_s} \cdot \frac{L_{A_{so}}}{L_{A_{so}} + L_{A_{s2}}} = \frac{E[V]L_{A_{so}}}{\pi A_{A_s}}. \tag{4}$$

The average outgoing rate $\mu_{A_{s2}}$ from area A_s to A_2 is given by

$$\mu_{A_{s2}} = \mu_{A_s} \cdot \frac{L_{A_{s2}}}{L_{A_{so}} + L_{A_{s2}}} = \frac{E[V]L_{A_{s2}}}{\pi A_{A_s}}. \tag{5}$$

Thus, the dwell time $T_{A_{so}}$ (random variable) that a mobile user stays in area A_s before moving out the marked cell has a mean value of $1/\mu_{A_{so}}$. Similarly, the dwell time $T_{A_{s2}}$ (random variable) that a mobile user stays in area A_s before moving into area A_2 has a mean value of $1/\mu_{A_{s2}}$.

Using fluid flow model, the average outgoing rate μ_{A_2} of mobile users in area A_2 is given by

$$\mu_{A_2} = \frac{E[V]L_{A_{s2}}}{\pi A_{A_2}}, \tag{6}$$

where A_{A_2} is the area of A_2 that is equal to πR_2^2. The dwell time T_{A_2} (random variable) that a mobile user stays in area A_2 before moving out of area A_2 has a mean value of $1/\mu_{A_2}$. The call holding time T_c (random variable) is also assumed to have a mean value of $1/\mu_c$.

We assume that the average arrival rate of originating calls in the marked cell is λ_O. Because mobile users are

21st International Conference on
Advanced Information Networking and Applications Workshops (AINAW'07)
0-7695-2847-3/07 $20.00 © 2007 IEEE

uniformly distributed in the whole cell, the average arrival rate λ_{OA_s} of originating calls in area A_s is given by $\lambda_O \cdot A_{A_s}/A_{A_1}$. The average arrival rate λ_{OA_2} of originating calls in area A_2 is given by $\lambda_O \cdot A_{A_2}/A_{A_1}$.

If the number of active mobile users in area A_s is U_{A_s}, the average number of active mobile users in area A_s is $E[U_{A_s}]$. Thus, the rate of mobile users moving out of the marked cell without completing their communications can be given by $E[U_{A_s}]\mu_{A_{so}}$. Since network N_1 is assumed to have homogeneous cells, the arrival rate λ_{HH} of horizontal handoff calls to the marked cell is equal to the rate of mobile users moving out of the marked cell without completing communications, which is $E[U_{A_s}]\mu_{A_{so}}$.

4.2 Performance Analysis

We assume that the arrival processes of originating calls and handoff calls are Poisson processes. We also assume the dwell times and call holding time follow exponential distributions. From the above assumptions, we can use a three-dimensional Markov chain to model the marked cell. The state of the marked cell can be defined using three-tuple nonnegative integers (i, j, k), where i is the sum of active mobile users in area A_s, j is the sum of active mobile users of network N_1 in area A_2, and k is the sum of active mobile users of network N_2. In the following, we analyze state transitions from the current state (i, j, k) event by event.

Event (a): When an originating call is generated in area A_s or a horizontal handoff call arrives from a neighboring cell of network N_1, it is accepted by N_1 if there are free bandwidth units in the marked cell of network N_1. Consequently, there is a state transition only when $i + j < B_1$. The transition rate is the total arrival rate of originating calls in A_s and horizontal handoff calls, which is given by $\lambda_{OA_s} + \lambda_{HH}$ if $i + j < B_1$. If there is no free bandwidth unit in the marked cell of network N_1, the state does not change which indicates that the call is blocked or forcedly terminated.

Event (b): When an active mobile user completes its call in area A_s or moves out of the marked cell of network N_1 before the call completion, the current state (i, j, k) transits to state $(i-1, j, k)$ with a transition rate $i \cdot (\mu_c + \mu_{A_{so}})$ if $i > 0$.

Event (c): When an originating call is generated in area A_2, there are two possible state transitions that correspond to the call acceptances to networks N_1 or N_2, respectively. The call acceptance is based on the proposed CFNS strategy. Consequently, the transition rate from state (i, j, k) to $(i, j, k+1)$ is given by

$$\begin{cases} \lambda_{OA_2}, & \text{if } k < B_2 \text{ and } i + j = B_1. \\ [1 - P(C_1 > C_2)] \cdot \lambda_{OA_2}, & \text{if } k < B_2 \text{ and } i + j < B_1. \end{cases} \tag{7}$$

The transition rate from state (i, j, k) to $(i, j+1, k)$ is given by

$$\begin{cases} \lambda_{OA_2}, & \text{if } k = B_2 \text{ and } i + j < B_1. \\ P(C_1 > C_2) \cdot \lambda_{OA_2}, & \text{if } k < B_2 \text{ and } i + j < B_1. \end{cases} \tag{8}$$

$P(C_1 > C_2)$ is the probability that the cost of network N_1 is larger than the cost of N_2.

Event (d): When an active mobile user in network N_1 completes its call in area A_2, the current state (i, j, k) transits to state $(i, j-1, k)$ with a transition rate $j\mu_c$ if $j > 0$. Similarly, when an active mobile user of network N_2 completes its call, the current state (i, j, k) transits to state $(i, j, k-1)$ with a transition rate $k\mu_c$ if $k > 0$.

Event (e): When an active mobile user moves from area A_s to A_2, the mobile user makes a vertical handoff to network N_2 if there are free bandwidth units in N_2. Otherwise, the mobile user still keeps its connection with network N_1. Both state transitions have the same transition rate as $i\mu_{A_{s2}}$ if $i > 0$.

Event (f): An active mobile user in area A_2 may connect to network N_1 or N_2. If the mobile user connects to N_1, it still keeps its connection with N_1 when it moves out of area A_2. The transition rate is given by $j\mu_{A_2}$ if $j > 0$. If the active mobile user connects to network N_2, it has to make a vertical handoff from N_2 to N_1. If there is no free bandwidth unit in N_1, the vertical handoff call is forcedly terminated. Therefore, there are two possible state transitions for the arrival of a vertical handoff call, which implies successful vertical handoff and fail one. Both state transitions have the same transition rate as $k\mu_{A_2}$ if $k > 0$.

All the above state transitions are out of the current state (i, j, k). The state transitions from other states to the current state (i, j, k) can be easily obtained using the similar way. The equilibrium state probabilities $P(i, j, k)$s are related to each other by global balanced equations that can be obtained from the above state transitions. Beside the state balanced equations, the summation of all the equilibrium state probabilities is equal to 1. Therefore, all the equilibrium state probabilities can be obtained using SOR (Successive Overrelaxation) algorithm by solving the above equations. In the following, we present the major performances like the blocking probability of originating calls, the forced termination probability of horizontal and vertical handoff calls using equilibrium state probabilities $P(i, j, k)$s.

Blocking probability of originating calls: When an originating call is generated in area A_s, it is blocked if there is no free bandwidth unit in the marked cell of network N_1. Therefore, the blocking probability of originating calls in area A_s is given by

$$P_{BA_s} = \sum_{i=0}^{B_1} \sum_{k=0}^{B_2} P(i, B_1 - i, k). \tag{9}$$

21st International Conference on
Advanced Information Networking and Applications Workshops (AINAW'07)
0-7695-2847-3/07 $20.00 © 2007 IEEE

When an originating call is generated in area A_2, it is blocked only when there is no free bandwidth unit in both networks N_1 and N_2. As a result, the blocking probability of originating calls in area A_2 is given by

$$P_{BA_2} = \sum_{i=0}^{B_1} P(i, B_1 - i, B_2). \qquad (10)$$

Since the mobile users are uniformly distributed in the whole cell, the average blocking probability of the whole cell is given by

$$P_B = \frac{\lambda_{OA_s} P_{BA_s} + \lambda_{OA_2} P_{BA_2}}{\lambda_O}. \qquad (11)$$

Forced termination probability of horizontal or vertical handoff calls: When an active mobile user moves into the marked cell from a neighboring cell of network N_1, it is terminated if there is no free bandwidth unit in the marked cell. When an active mobile user of network N_2 moves out area A_2 and makes a vertical handoff to network N_1, the vertical handoff call is also terminated if there is no free bandwidth unit in the marked cell of network N_1. Therefore, the forced termination probability P_{HH} of horizontal handoff calls or the forced termination probability P_{VH} of vertical handoff calls is given by Equation 9.

5 Numerical results and discussions

In our analysis model, we consider an integrated wireless and mobile network that has two types of networks N_1 and N_2. We assume that the radius of each cell in N_1 and N_2 is 600 m and 200 m. The total bandwidth units in each cell is set to $B_1 = 20$ and $B_2 = 5$. The average of moving speed is set to $E[V] = 10$ m/s, and w_g is set to 0.3 in CFNS strategy. We also use simulations to verify our analysis. In the simulations, we assume that network N_1 has 10×10 cells that fill the entire service area. When an mobile user in the edge cell moves out of the service area, we assume that the same mobile user moves into the other side of the service area through a horizontal handoff between two corresponding cells. By doing this, we can simulate a very large coverage of wireless and mobile networks using limited number of cells.

First of all, we examine the blocking probabilities of originating calls in different areas of the marked cell. Figures. 2 and 3 show the blocking probabilities of originating calls in different areas in TBNS, RSNS, and CFNS strategies. The offered traffic load is defined as λ_O/μ_c. We can observe that the blocking probability of originating calls in area A_2 in RSNS strategy is slightly lower than that in CFNS strategy. However, this benefit sacrifices the blocking probability of originating calls in area A_s. Without considering the traffic load, more bandwidth units in network N_1 are allocated to the originating calls that are generated in area A_2. Therefore, the originating calls that are generated in area A_s have less chance to obtain the bandwidth unit from network N_1. As a result, the blocking probability of originating calls of area A_s in RSNS strategy is the highest. The simulation results also verify our analysis.

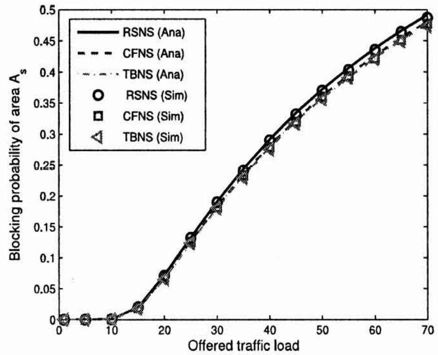

Figure 2. Blocking probability of originating calls in area A_s

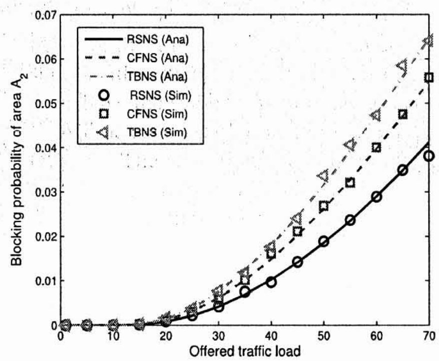

Figure 3. Blocking probability of originating calls in area A_2

We then examine the blocking probability of originating calls of the whole cell. As shown in Figure 4, the blocking probabilities of originating calls in TBNS, RSNS, and CFNS strategies have a small difference. The RSNS strategy achieves the worst system performance because it does not consider the traffic load between different types of networks, while received signal strength is the only decision factor.

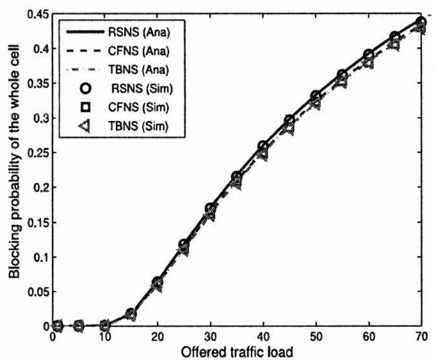

Figure 4. Blocking probability of originating calls in the whole cell

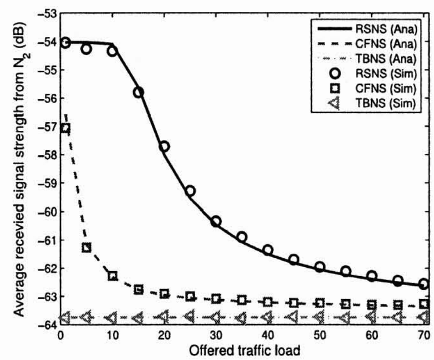

Figure 5. Average received signal strength (in dB)

In RSNS strategy, the originating calls in area A_2 is accepted by network N_1 if both networks (N_1 and N_2) have free bandwidth units and the cost of network N_1 is larger than that of network N_2. The originating call that has a stronger received signal strength from network N_2 has a larger cost of network N_2, which has a higher probability to be accepted by network N_2. Therefore, the average received signal strength of the originating calls accepted by network N_2 becomes larger in RSNS strategy. In this paper, we compare the average received signal strength given the transmitted signal strength is equal to 0 dB. Figure 5 shows the average signal strength in TBNS, RSNS, and CFNS strategies. As a special case of CFNS strategy, RSNS strategy ($w_g = 0$) achieves the strongest average received signal strength. On the other hand, as another special case of CFNS strategy, TBNS strategy ($w_g = 1$) achieves the worst received signal strength because TBNS strategy does not consider the received signal strength. When the offered traffic load increases, the probability that networks N_1 has free bandwidth units decreases. As a result, the call generated in area A_2 has to be accepted by network N_2 that has free bandwidth units without considering the cost function. Therefore, the average received signal strength becomes worse when the offered traffic load increases as shown in Figure 5.

6 Conclusions

In this paper, we proposed a system model for an integrated wireless and mobile network. We also proposed a cost-function-based network selection strategy to combine the traffic balance and received signal strength among different types of networks. The system performance was analyzed using theoretical model and simulations. The results have shown that the proposed cost-function-based network selection strategy can achieve a tradeoff between the blocking probability of originating calls and the average received signal strength, which are very important for both systems and users.

References

[1] M. Stemm and R. H. Katz. "Vertical handoffs in overlay wireless networks," *ACM Mobile Networking (MONET), Special Issue on Mobile Networking in the Internet*, pp. 335-350, 1997.

[2] Q. Zhang, C. Guo, Z. Guo, and W. Zhu, "Efficient mobility management for vertical handoff between WWAN AND WLAN," *IEEE Communications Magazine*, vol.41, no.11, pp. 102-108, November, 2003.

[3] M. Nam, N. Choi, Y. Seok and Y. Choi, "WISE: Energy-efficient interface selection on vertical handoff between 3G Networks and WLANs," *Proceedings of the IEEE PIMRC 2004*, pp. 692-698, 2004.

[4] H. Wang, R. Katz, and J. Giese, "Policy-enabled handoffs across heterogeneous wireless networks," *WMCSA 99*, pp. 51-60, 1999.

[5] Q-A. Zeng and D. P. Agrawal, "Modeling and Efficient Handling of Handoffs in Integrated Wireless Mobile Networking," *IEEE Transactions on Vehicular Technology*, Vol. 51, No.6, pp. 1469-1478, November, 2002.

21st International Conference on
Advanced Information Networking and Applications Workshops (AINAW'07)
0-7695-2847-3/07 $20.00 © 2007 IEEE