This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the WCNC 2009 proceedings.

Markov Decision Process-based Adaptive Vertical Handoff with RSS Prediction in Heterogeneous Wireless Networks*

Ben-Jye Chang[1], Jun-Fu Chen, Cheng-Hsiung Hsieh and Ying-Hsin Liang

Abstract — In a **heterogeneous wireless network (HWN) that consists of various wireless networks (e.g., WiMAX and WiFi) and cellular communications (e.g., B3G and 4G), vertical handoff acts as an important mechanism for achieving continuous seamless transmissions and improving grade of service. This work thus proposes an adaptive vertical handoff algorithm with predictive RSS to reduce unnecessary handoff while increasing utilization and decreasing connection dropping significantly. The proposed approach determines the optimal target network in two phases: polynomial regression RSS prediction and Markov decision process analysis. Numerical results indicate that the proposed adaptive approach outperforms other approaches in the number of vertical handoffs and SWGoS while yielding competitive utilization.**

***Keywords*—Markov decision process, adaptive prediction, vertical handoff, heterogeneous wireless networks, RSS**

I. Introduction

Important wireless networks, including WiMAX [1] and WiFi [2], and mobile communications, such as WCDMA [3] and HSDPA/HSUPA [4], are being rapidly developed to achieve high transmission rate and better service of quality (QoS). These heterogeneous wireless networks obviously have differences in data rates, transmission ranges, traffic classes and access costs. Vertical handoff is a significant mechanism for fulfilling seamless data transfer when mobile nodes cross the overlay area between adjacent heterogeneous wireless networks. Several challenges exist in vertical handoff, including receiving RSSs from different networks, various service classes, determining the optimal target network for handoff and avoiding vibrating TCP congestion window [6][7]. The related works of handoff can be classified into three types: 1) RSS-based related approaches [8]-[12], 2) Cost-based related approaches [13]-[15], and 3) Others [16][17]. However, some critical issues still have not yet solved in these three-type related works, e.g., In the RSS-based approaches the RSS is dependent on the distance, they still has the same problem in heterogeneous wireless networks. In the cost-based approaches, using a single threshold will cause the ping-pong effect. In the other approaches, the computation complexity is too high while using a periodical fast Fourier transform-based decay detection to determine the decay of received signal.

To solve these problems in vertical handoff, a MDP cost-based RSS-prediction approach is proposed herein to perform vertical handoff in HWNs. The main contributions include 1) Adaptive RSS-prediction vertical handoff approach, and 2) Determining the optimal target network to handoff into.

The remainder of the paper is organized as follows. Section II describes the network model and performance metrics for evaluations. Section III presents the adaptive vertical handoff with the predictive RSS approach. The MDP cost-based approach is detailed in Section IV. Section V evaluates the performance of vertical handoff in several important metrics, and compares the results to those of other approaches. Section VI draws some important conclusions.

II. Network Model

This section first defines the model of a HWN, and then evaluates the proposed approach. We model a HWN covering an area with length L and width W consists of a single UMTS/B3G network, M WMANs and N WLANs. A WMAN and a WLAN are denoted as $WMAN_m$ and $WLAN_n$, respectively, where m (i.e., $0 \le m < M$) and n (i.e., $0 \le n < N$) represent the network indexes. We also assume that the B3G UMTS network covers the entire area. Each WLAN is with the same transmission radius of R_{11} and that of each WMAN is R_{16}. M WMANs and N WLANs are randomly deployed over the area. The signal strength received from $WLAN_n$ and $WMAN_m$ at any mobile node are denoted as RSS_n^{11} and RSS_m^{16}, respectively. Each mobile node supports three interfaces to access WMAN, WLAN and B3G. A mobile node chooses only one interface for communication at any moment. Due to node mobility, a mobile node moves from a network i to another network j. If their network-types are different, the mobile node should perform vertical handoff for continuous transmission. Conversely, if the network-types of these two networks are the same, the mobile node performs horizontal handoff.

Then, in evaluations the proposed approach is compared with important related approaches in different performance metrics: number of vertical handoffs, GoS and network utilization. First, the ping-pong effect occurs when a mobile node moves around the overlay area between two networks, causing unnecessary handoffs and increasing the handoff overhead. Fewer handoffs indicate a better handoff algorithm. Second, a handoff connection will be dropped

* This research was supported in part by the National Science Council of Taiwan, ROC, under contracts NSC-96-2221-E-324-019 and NSC-97-2221-E-324-006-MY3.

Ben-Jye Chang is with the Institute of CSIE, National Yunlin University of Science and Technology, Taiwan, ROC. (email: changb@cyut.edu.tw)
Jun-Fu Chen and Cheng-Hsiung Hsieh are with the Department of CSIE, Chaoyang University of Technology, Taiwan, ROC.
Ying-Hsin Liang is with the Department of CSIE, Nankai University of Technology, Taiwan, ROC. (e-mail: t136@nkut.edu.tw)
1. The corresponding author of this paper.

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the WCNC 2009 proceedings.

when the available bandwidth of the target network is not enough. The revenue loss of dropping connections is much higher than that of blocking of new connections. To maximize the revenue to the network provider, the Sum of Weighted Grade of Service (SWGoS) is adopted as an important metric, which is formulated by

$$SWGoS = W_B \frac{\sum_{k=1}^{K} B_k}{K} + W_D \frac{\sum_{k=1}^{K} D_k}{K},$$

where B_k and D_k represent the blocking and dropping probabilities of class k traffic, and W_B and W_D represent the weights of blocking and dropping connections, respectively. In this paper, W_B and W_D are set to one and ten, respectively [19]. Third, network utilization is adopted as another performance metric. Higher network utilization indicates better performance.

III. Adaptive Vertical Handoff with Predictive RSS

The proposed approach for vertical handoff in heterogeneous wireless networks is detailed in two aspects. First aspect is the adaptive vertical handoff with predictive RSS mechanism that is described in this section. Then, the time complexity of the proposed approach is analyzed.

The adaptive vertical handoff with predictive RSS mechanism consists of two primary phases: the phase of predictive RSS with a hysteresis, and the phase of determining optimal wireless network to handoff in. In the first phase, a scheme of polynomial regression-based prediction of RSS is proposed to cooperate with a hysteresis in order to predict whether a mobile node moves closer to or away from the monitored network. In the second phase, the handoff cost is determined based on the Markov decision process (MDP) method [22]. The network with the least handoff cost is then selected as the optimal handoff network. The second phase achieves load balancing and non-dropping. The following three subsections describe the adaptive vertical handoff approach.

3.1. The Importance of Hysteresis and Predictive RSS Handoff

The RSS-based scheme with a hysteresis is adopted to avoid unnecessary handoff due to the ping-pong effect. For instance, H_1 shown in Fig. 1 indicates the hysteresis between B3G and WMAN, where the margins of H_1 are bounded by the lower threshold, $S_{WMAN,1}$, and the higher threshold, $S_{WMAN,2}$, and the margin interval is denoted as $\Gamma_{H_1} = |S_{WMAN,2} - S_{WMAN,1}|$. Although mobile nodes transmit/receive well between these two thresholds and thus avoid unnecessary handoff, they suffer from low data rate and weak RSS when the received RSS of the serving network is close to the low threshold. Such a low data rate and weak RSS causes serious low utilization and high dropping probability. The main reason is explained below and as shown in Fig. 2.

Figure 2 shows a mobile node *MN* is moving from network 1 to network 2. In the RSS-based handoff approach, the mobile node performs handoff when it is in the overlap area; therefore, the handoff point may occur at point 2, 3, or 4. This causes a serious ping-pong effect if the mobile node moves around the overlap area. In the RSS with a hysteresis approach, the mobile node performs handoff at point 4. Therefore, the mobile node receives too weak RSS from *network 1*. This results in a low data rate and high dropping probability. Thus, this work proposes a predictive RSS-based handoff approach to perform handoff at point 2 when the mobile node predicts that is moving toward *network 2*. The predictive RSS mechanism has two advantages. First, the handoff process can be performed before RSS becomes weak, and thus obtains better QoS and higher data rate. Second, it not only avoids unnecessary handoff, but also minimizes dropping probability obviously.

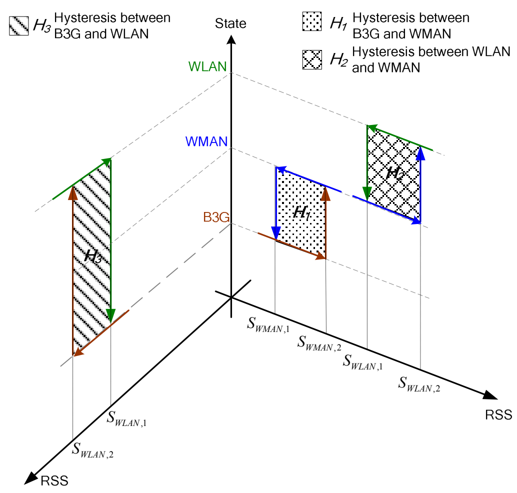

Fig. 1. The thresholds of three hysteresises

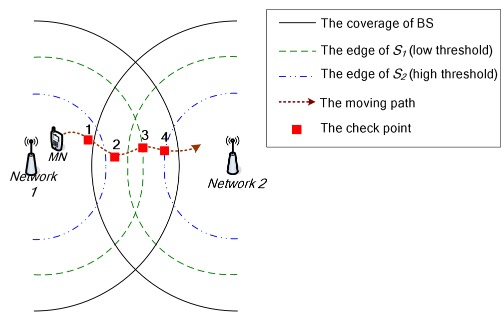

Fig. 2. Handoff points in different handoff approaches

3.2. The Polynomial Regression-Based Predictive RSS Handoff approach

Based on the hysteresis mechanism, this work proposes a polynomial regression predictive RSS handoff approach that consists of two steps: the preprocess step and the RSS prediction step.

Step 1. The preprocess step

In prediction, some previous RSSs are important for determining the next predictive RSS. To intensify the polynomial regression-based curve fitting, the preprocess of accumulated generating operation, which is based on [20], is adopted for achieving the accuracy of the prediction results. The preprocess generates a new data sequence by summation of previous n data, which is formulated by

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the WCNC 2009 proceedings.

$$S'(n) = \sum_{p=1}^{n} S(p), \tag{1}$$

where $S(p)$ means the original data sequence, and $S'(n)$ is the new sequence after executing the preprocess. The purpose of the preprocess is to smooth the fitting curve.

Step 2. The RSS prediction step

After the preprocess step, the new data sequence is used as the input data for curve fitting in the RSS prediction step. The predictive RSS in the new sequence is denoted as $RSS'_{prediction}$, which is computed by

$$RSS'_{predication} = F_{pre}(t+1), \tag{2}$$

where $F_{pre}(\cdot)$ is a polynomial function with $Z+1$ unknown coefficients (i.e., a_i, $0 \le i \le Z$). That is

$$F_{pre}(t) = a_0 t^0 + a_1 t^1 + a_2 t^2 + \cdots + a_Z t^Z. \tag{3}$$

The preprocess predictive RSS of the previous hth process is

$$F_{pre}(t_h) = a_0 t_h{}^0 + a_1 t_h{}^1 + a_2 t_h{}^2 + \cdots + a_Z t_h{}^Z.$$

Let R_h be the RSS of the previous hth process. Then, the sum of the square of difference between actual R and predictive $F_{pre}(t\,)$ of previous n processes is defined as D, i.e.,

$$D = \sum_{h=1}^{H} [R_h - (a_0 t_h{}^0 + a_1 t_h{}^1 + a_2 t_h{}^2 + \cdots + a_Z t_h{}^Z)]^2. \tag{4}$$

For determining each coefficient a_i, we view each a_i in each polynomial as a variable and take partial differential of each a_i in (4), and then set each partial differential equation to zero. After that, we can obtain the following equivalent polynomials,

$$\begin{cases} a_0\left(\sum_{h=1}^{H} t_h{}^0\right) + a_1\left(\sum_{h=1}^{H} t_h{}^1\right) + \cdots + a_Z\left(\sum_{h=1}^{H} t_h{}^Z\right) = \sum_{h=1}^{H} t_h{}^0 R_h \\ a_0\left(\sum_{h=1}^{H} t_h{}^1\right) + a_1\left(\sum_{h=1}^{H} t_h{}^2\right) + \cdots + a_Z\left(\sum_{h=1}^{H} t_h{}^{Z+1}\right) = \sum_{h=1}^{H} t_h{}^1 R_h \\ \quad\vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ a_0\left(\sum_{h=1}^{H} t_h{}^Z\right) + a_1\left(\sum_{h=1}^{H} t_h{}^{Z+1}\right) + \cdots + a_Z\left(\sum_{h=1}^{H} t_h{}^{2Z}\right) = \sum_{h=1}^{H} t_h{}^Z R_h. \end{cases} \tag{5}$$

After determining all the summations of t in (5), we can simplify this matrix of polynomials in (5) to an upper triangle matrix of polynomials. Each coefficient a_i can be determined by the Native Gauss Elimination method [21] as

$$a_i = \frac{1}{C_{i,i}^{(i)}}\left(b_i^{(i)} - \sum_{j=i+1, j\le Z}^{Z} C_{i,j}^{(i)} a_j\right), \tag{6}$$

where $i = Z, Z-1, Z-2, \cdots, 0$.

Consequently, the polynomial function of $F_{pre}(t+1)$ in (2) is obtained, i.e., the prediction of the preprocessed RSS, $RSS'_{prediction}$, is determined. Since $RSS'_{prediction}$, i.e., $S'(n+1)$, is computed by (2), the predictive RSS of the original sequence, $RSS_{prediction}$, i.e., $S(n+1)$, can be determined by the reverse transformation of (1) as

$$S(n+1) = S'(n+1) - S'(n). \tag{7}$$

We give an experiment of the polynomial regression-based predictive RSS. Fig. 3 demonstrates the actual and predictive RSSs of a mobile node moves in a heterogeneous wireless network, in which only three previous RSSs (i.e., $H = 3$) are adopted for each prediction. In Fig. 3, the predictive RSS always predicts accurately even though the actual RSS is at the next point of each peak, which is the most difficult prediction point. The proposed prediction achieves an excellent Mean Relative Error (MRE) of 0.7256% in this experiment.

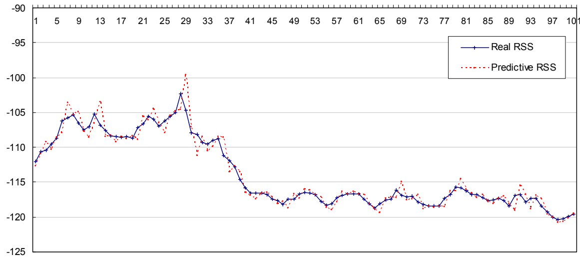

Fig. 3. Predictive results by using polynomial regression

After obtaining the predictive RSS of each neighbor BS, mobile nodes use the predictive RSS to determine the mode node moves closer to or away from the monitored target network. If current RSS of the serving network is less than the low threshold, the mobile node checks whether the predictive RSS of each target network is stronger enough or not by comparing the predictive RSS with the high threshold, i.e., $RSS_{prediction}^{WLAN,n} \ge S_{WLAN,2}$ or $RSS_{prediction}^{WMAN,m} \ge S_{WMAN,2}$. The target network, whose predictive RSS satisfies above condition, will be added to the corresponding candidate list. Meanwhile, the mobile node checks whether the condition of vertical or horizontal handoff is satisfied or not. If it is true, the mobile node performs vertical handoff based on the network-type selection order of WLAN>WMAN>B3G. The benefit of using predictive RSS is to help the mobile node to know whether it is toward the target AP/BS or not. If yes and current RSS of the serving AP is lower than the low threshold, the mobile node performs handoff early. Consequently, predictive RSS brings two advantages: 1) dropping of handoff connections is avoided, and 2) higher data rate and transmission quality are achieved.

IV. MDP Cost-based Approach

In most RSS-based handoff algorithms, the target network with the strongest signal is always selected as the network to handoff into. Since mobile nodes select the same strongest network as the handoff-in network, it causes the available bandwidth of the strongest network is not enough for these mobile nodes and thus yields unbalance traffic loading among networks. This section thus proposes the Markov decision process (MDP) cost-based approach to define the handoff cost for all candidate wireless networks. Then, the candidate wireless network with the least handoff cost will be chosen as the target network to handoff into, and thus achieves the goal of load balancing.

In this paper, the MDP approach formulates the bandwidth allocation problem as a Markov decision process and obtains the handoff cost for carrying a handoff connection by the network according to Markov decision theory [22]. Although MDP may cause huge state space for exact modeling into a Markov decision process, all previous researches made two assumptions, namely link independence assumption and path cost separability assumption [23,24,25,26]. A simplified link model is proposed in [25] to further reduce the state space and computational complexity for multi-service loss networks. Moreover, MDP analysis has been adopted to determine the optimal code for single and multiple codes of resource management in B3G/UMTS successfully [27].

Based on the Markov decision process, the state of a wireless network is described by the bandwidth occupancy, and the network is assumed to support K classes of services, each with different QoS requirements. Now, we model each state of a wireless network w of $WLAN$ or $WMAN$ with total capacity C_w as a Markov process with a birth rate of λ_k^i and death rate of μ_k^i, where k is the traffic class and i represents the state of bandwidth occupancy. Let $\pi(i)$, $i = 0,\ 1,\ 2, ...,\ C_w$, be the equilibrium probability of being in state i, which satisfies

$$\mu_k^i \pi(i) = \lambda_k^i \pi(i-1), \quad i = 1, ..., C_w, \tag{8}$$

and

$$\sum_{i=0}^{C_w} \pi(i) = 1. \tag{9}$$

By forming a Markov decision process on the birth-death process, the long-term average reward loss rate, g, and a set of relative values, $v(i)$, can be obtained, using the following set of simple expressions:

$$v^w(C_w) - v^w(C_w - 1) = \frac{\sum_{j=1}^{K} r_j^w \lambda_j^w}{\overline{\lambda}_{C_w-1}^w} \frac{E(\overline{\lambda}^w, C_w)}{E(\overline{\lambda}^w, C_w - 1)}, \tag{10}$$

$$v^w(i) - v^w(i-1) = \frac{g}{\overline{\lambda}_{i-1}^w E(\overline{\lambda}^w, i-1)}, \quad 1 \le i < C_w \tag{11}$$

where

$$E(\overline{\lambda}^w, i) = \frac{\frac{1}{i!}\prod_{j=0}^{i-1} \overline{\lambda}_j^w}{\sum_{n=0}^{i} \frac{1}{n!}\prod_{j=0}^{n-1} \overline{\lambda}_j^w}, \tag{12}$$

and

$$g = \sum_{j=1}^{K} r_j^w \lambda_j^w - C_w(v^w(C_w) - v^w(C_w - 1)) \cdot \tag{13}$$

The r_j^w and λ_j^w are the network reward and the arrival rate of class j traffic on network w, respectively.

Equations (10)-(13) provide a simple way of obtaining the cost of carrying a class k connection, where the network reward is set to $r_k^w = r_k^{\varpi} \times \frac{b_k^w}{Z}$, and the arrival rate λ_k^w of class k connection on network w is computed on-line based on the Exponentially Weighted Moving Average (EWMA) model of $\lambda_{k,new}^w = (1-\alpha) \cdot \lambda_{k,old}^w + \alpha \cdot \frac{\tilde{\lambda}_k^w}{1 - B_k^w}$.

Based on the difference of relative values obtained from (10)-(11), the handoff cost of carrying a class k connection on the network w with occupancy i is denoted as $\Gamma_k^w(i)$ and defined as

$$\Gamma_k^w(i) = \begin{cases} \frac{v^w(i + b_k) - v^w(i)}{\mu_k} & \textit{if} \quad i + b_k \le C_w, \\ \infty & \textit{otherwise}, \end{cases} \tag{14}$$

where b_k^w is the bandwidth required of the class k handoff connection at network w. If there are several available neighbor networks in the candidate list, the network with the least cost will be chosen as the optimal target network, NET_{opt}, for the mobile node to handoff into. As a result, the MDP cost-based achieves several advantages: selecting the optimal target network, increasing transmission quality and achieving load balancing.

Finally, we analyze the time complexity of the MDP-based adaptive predictive RSS approach. The analysis includes two parts: 1) the polynomial regression-based predictive RSS approach, and 2) the MDP cost-based approach for determining the optimal target network. First, the handoff algorithms compute the predictive RSS based on solving Z Linear Equations by the Gauss Elimination, where each factor of any linear equation is determined by H previous RSSs. The time complexity of Gauss Elimination is $O(Z^3)$, where Z is the dimension of the $Z \times Z$ square matrix. In the worst case, there are at most N WLANs and M WMANs within the transmission range of a mobile node. The total time complexity of handoff detection procedure is thus

$$O(NZ^3) + O(MZ^3). \tag{15}$$

Since Z is a constant when the order of polynomial function is decided, it needs a constant time for each predictive RSS. Consequently, if $N \ge M$, the time complexity becomes $O(NZ^3)$; otherwise, it is $O(MZ^3)$.

Second, the MDP cost-based approach is proposed to select the network with the least handoff cost from the candidate list as the optimal target network to handoff into. The time complexity for determining the minimum MDP-based cost network is $O(N)$ or $O(M)$ for the WLAN or WMAN, respectively. Finally, the time complexity of the proposed MDP cost-based predictive RSS approach is $O(NZ^3) + O(N)$, if $N \ge M$. Otherwise, it becomes $O(MZ^3) + O(M)$. Since Z is a constant value, the time complexity is $O(N)$ if $N \ge M$; otherwise, it is $O(M)$, where N and M are the numbers of WLANs and WMANs, respectively.

V. NUMERICAL RESULTS

This section evaluates the proposed approach (i.e., denoted by *PreRSS-Poly+MDP*) by comparing various performance metrics, including number of vertical handoffs, SWGoS and network throughput. Several compared approaches include 1) the RSS-based approach with a single threshold (denoted by *RSS-T*) [8] and 2) the RSS-based approach with two thresholds as an interval of a hysteresis (denoted by *RSS-HT*) [10], and 3) the adaptive cost-based approach with a predictive RSS hysteresis (denoted by *PreRSS-Cost*) [28].

In simulations, the evaluated heterogeneous wireless networks consist of a single B3G/UMTS network, 100 WLANs and 10 WMANs, where WLANs and WMANs are randomly deployed. The topology covers an area with 3000(m) of length and 3000(m) of width. The number of mobile nodes ranges from 10 to 70 and the mobility of random-way point is adopted for each mobile node with random direction and random velocity from 1 to 25 m/s. The transmission range of B3G covers the whole area, that of each WLAN is 100(m), and that of each WMAN is 1000(m). The bandwidth of B3G/UMTS, WLAN and WMAN are 384 Kbps, 54Mbps and 15Mbps, respectively.

The simulation considers two classes of traffic: the constant bit rate (CBR) and variable bit rate (VBR) classes. The CBR traffic is assumed to arrive at the heterogeneous network to a Poisson distribution with arrival rate λ. The average holding time of the CBR traffic is exponentially distributed (μ) and its mean is normalized to unity. On the other hand, the VBR traffic is assumed to arrive based on a Pareto distribution with the parameters of $\alpha\_on = 1.1$, $\alpha\_off = 1.5$ and $\beta = 1$, where α is the shape parameter and β is the scale parameter.

The RSS received by a mobile node is different when it uses different wireless networks. In WLAN, the RSS is computed based on

$$RSS = 10 \cdot \log_{10}\left(\frac{100}{(distance \cdot 39.37)^{RSS\_factor}}\right) \text{ (dbm)}, \quad (16)$$

where RSS_factor is the signal decay factor and we set 2.8 to RSS_factor [29]. The RSS is affected not only by distance but also by environment factors of transmissions. In this work, $S_{WLAN,1}$ and $S_{WLAN,2}$ are set to -80 and -75 (dbm), respectively. In WMAN, the RSS received by mobile node is computed based on the propagation model in [29]. In the propagation model, when the distance between a mobile node and a BS is d (m), the $RSS(d)$ is formulated by

$$RSS(d) = P_{trans} - PL(d)_{\sigma}, \quad (17)$$

where P_{trans} is the transmit power and $PL(d)$ is the path loss at distance d. The path loss at distance d is defined as

$$PL(d) = S + 10 \cdot n \cdot \log(d), \quad (18)$$

where S denotes the path loss constant that depends on the propagation environment and n denotes the path loss exponent which range is from 2 to 4. Meanwhile, $S_{WMAN,1}$ and $S_{WMAN,2}$ are set to -95 and -92 (dbm), respectively.

First, we evaluate the performance metrics under different arrival rates ranging from 6 to 16, as demonstrated in Figs. 4-6. In Fig. 4, the proposed *PreRSS-Poly+MDP* approach yields the least number of vertical handoffs, but *RSS-T* results in the worst results. Meanwhile, the numbers of vertical handoffs of all approaches increase when the arrival rate increases. The number of vertical handoffs of *PreRSS-Poly+MDP* gently increases while the arrival rate increasing, but that of *RSS-T*, *RSS-HT* and *PreRSS-Cost* increase obviously. In Fig. 5, since *PreRSS-Poly+MDP* determines the optimal network based on the analysis of Markov decision process, it avoids performing handoff to a target network, which has strong RSS but lacks enough bandwidth for the handoff connection. As a result, the dropping probability of *PreRSS-Poly+MDP* is zero and thus yields the least SWGoS. However, SWGoS of *RSS-T* and *RSS-HT* are almost above 7.7 under different arrival rate, as shown in Fig. 5. *PreRSS-Cost* yields competitive SWGoS to *PreRSS-Poly+MDP*, because it considers the network that has the most available bandwidth as the target network and thus avoids dropping. Then, the network utilization is evaluated as demonstrated in Fig. 6. Since the data rate of WLAN and WMAN is larger than that of cellular communications, e.g., B3G/UMTS, more the period of staying at WLAN or WMAN results in higher utilization. Therefore, more the number of vertical handoffs to WLAN or WMAN a mobile node yields, more the utilization the network has. From Fig. 6, we can observe that *RSS-T* yields the highest network utilization. *PreRSS-Poly+MDP* yields competitive utilization to that of *PreRSS-Cost* and *RSS-HT*. Unfortunately, in Fig. 4 *RSS-T* and *RSS-HT* yield much higher number of vertical handoffs than *PreRSS-Poly+MDP*.

In summary, the proposed *PreRSS-Poly+MDP* approach results in the least number of vertical handoffs, the best SWGoS and competitive utilization than *RSS-T*, *RSS-HT* and *PreRSS-Cost*. The SWGoS of *PreRSS-Poly+MDP* is very low, but that of *RSS-T* and *RSS-HT* obviously increase as the arrival rate, mobility, or NDS increases. The primary reason is that the *PreRSS-Poly+MDP* selects the WLAN/WMAN with the least cost as the optimal target network for handoff, where the cost is determined by the Markov decision process analysis. Specifically, the proposed approach achieves load balancing and avoids unnecessary handoff, while yielding competitive utilization. Consequently, *PreRSS-Poly+MDP* outperforms other approaches in all performance metrics of vertical handoff in heterogeneous wireless networks.

VI. CONCLUSIONS

This work proposes a polynomial regression predictive RSS approach with the MDP-based optimal network selection for handoff in heterogeneous wireless networks. The proposed approach consists of a two-phase procedure. In the first phase, a predictive RSS based on the polynomial regression with a hysteresis algorithm is proposed to predict whether a mobile node moves closer to or away from the monitored wireless network. In the second phase, the handoff cost is determined based on the Markov decision process (MDP) analysis, and the candidate network with the least handoff cost is selected as the optimal handoff network. The time complexity of the proposed approach is analyzed as $O(N)$ or $O(M)$, where N and M are the number of WLANs and WMANs, respectively. Numerical results indicated that the proposed approach outperforms other approaches in number of vertical handoffs and SWGoS, while yielding competitive utilization. Meanwhile, the proposed approach yields none dropping probability, which meets the requirement for optimal vertical handoff in heterogeneous wireless networks.

VII. REFERENCES

[1] "IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems," *IEEE 802.16e Standard*, Feb. 2006.

[2] "IEEE Standard for Local and Metropolitan Area Networks Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," *IEEE 802.11e Standard*, Nov. 2005.

[3] H. Holma and A. Toskala, "WCDMA for UMTS: Radio Access for Third Generation Mobile Communications," *John Wiley & Sons Inc.*, 2004.

[4] H. Holma and A. Toskala, "HSDPA/HSUPA for UMTS: High Speed Radio Access for Mobile Communications," *John Wiley & Sons Inc.*, 2006.

[5] D. Johnson, C. Perkins, and J. Arkko, "Mobility Support in IPv6," *IETF RFC 3775*, June 2004.

[6] I.F. Akyildiz, J. McNair, J.S.M. Ho, H. Uzunalioglu, and W. Wang, "Mobility Management in Next-generation Wireless Systems," *Proceedings of the IEEE*, Vol. 87, Issue 8, pp. 1347-1384, Aug. 1999.

[7] K.R. Santhi, V.K. Srivastava, G. SenthilKumaran, and A. Butare, "Goals of True Broad Band's Wireless Next Wave (4G-5G)," *2003-Fall IEEE Vehicular Technology Conference*, Vol. 4, pp. 2317-2321, Oct. 2003.

[8] G.P. Pollini, "Trends in handover design," *IEEE Communications Magazine*, Vol. 34, Issue 3, pp.82-90, Mar. 1996.

[9] N. Zhang, and J.M. Holtzman, "Analysis of Handoff Algorithms Using Both Absolute and Relative Measurements," *IEEE Transactions on Vehicular Technology*, Vol. 45, Issue 1, pp.174-179, Feb. 1996.

[10] P. Marichamy, S. Chakrabarti, and S.L. Maskara, "Performance Evaluation of Handoff Detection Schemes," *IEEE Conference on Convergent Technologies For The Asia-Pacific 2003*, Vol. 2, pp. 643 - 646, Oct. 2003.

[11] K. Pahlavan, P. Krishnamurthy, A. Hatami, M. Ylianttila, J.P. Makela, R. Pichna, and J. Vallstron, "Handoff in Hybrid Mobile Data Networks," *IEEE Personal Communications*, Vol. 7, Issue 2, pp. 34-47, Apr. 2000.

[12] H. Bing, C. He, and L. Jiang, "Performance Analysis of Vertical Handover in a UMTS-WLAN Integrated Network," *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications 2003*, Vol. 1, pp.187-191, Sep. 2003.

[13] J. McNair and Z. Fang, "Vertical Handoffs in Fourth-generation Multinetwork Environments," *IEEE Wireless Communications*, Vol. 11, Issue 3, pp. 8-15, June 2004.

[14] H.J. Wang, R.H. Katz, and J. Giese, "Policy-enabled Handoffs Across Heterogeneous Wireless Networks," *IEEE Mobile Computing Systems and Applications Conference*, pp. 51-60, Feb. 1999.

[15] Z. Fang and J. McNair, "Optimizations for Vertical Handoff Decision Algorithms," *IEEE WCNC2004,* Vol. 2, pp. 867-872, Mar. 2004.

[16] Q. Zhang, C. Guo, Z. Guo, and W. Zhu, "Efficient Mobility Management for Vertical Handoff between WWAN and WLAN," *IEEE Communications Magazine*, Vol. 41, Issue 11, pp.102-108, Nov. 2003.

[17] C. Guo, Z. Guo, Q. Zhang, and W. Zhu, "A Seamless and Proactive End-to-end Mobility Solution for Roaming across Heterogeneous Wireless Networks," *IEEE JSAIC*, Vol. 22, Issue 5, pp. 834-848, June 2004.

[18] S. Pack, J. Choi, T. Kwon, and Y. Choi, "Fast-Handoff Support in IEEE 802.11 Wireless Networks," *IEEE Communications Surveys & Tutorials*, Vol. 9, Issue 1, pp. 2-12, 2007.

[19] S. Boumerdassi and A.-L. Beylot, "Adaptive Channel Allocation for Wireless PCN," *Mobile Networks and Applications*, Vol. 4, No. 2, pp. 111–116, June 1999.

[20] T.-C. Chang, K.-L. Wen, and M.-L. You, "The Study of Regression Based on Grey System Theory," *IEEE International Conference on Systems, Man, and Cybernetics*, Vol. 5, pp. 4307-4311, Oct. 1998.

[21] S.C. Chapra and R.P. Canale, "Numerical Methods for Engineers," *McGraw-Hill*, 2002.

[22] R.A. Howard, "Dynamic Programming and Markov Processes," *John Wiley & Sons Inc.*, 1960.

[23] T.J. Ott and K.R. Krishnan, "State Dependent Routing of Telephone Traffic and the Use of Separable Routing Schemes," *International Teletraffic Congress,* Vol. 11, pp. 867-872, Sep. 1985.

[24] S.-P. Chung and K.W. Ross, "Reduced Load Approximations for Multirate Loss Networks," *IEEE Transactions on Communications*, Vol 41, No. 8, Aug. 1993.

[25] R.-H. Hwang, J.F. Kurose, and D. Towsley, "State Dependent Routing for Multirate Loss Network," *IEEE Globecom'92*, pp. 565-570, Dec. 1992.

[26] R.-H. Hwang, May 1993, "Routing in High-Speed Networks", Ph. D. Thesis, also available from Computer Science Dept. Technical Report 93-43, Univ. of Massachusetts Amhert.

[27] B.-J. Chang and P.-S. Chang, "Multicode-based WCDMA for Reducing Waste Rate and Reassignments in Mobile Cellular Communications," *Computer Communications*, Vol. 29, Issue 11, pp.1948-1958, July 2006.

[28] B.-J. Chang and S.-Y. Lin, "Mobile IPv6-based Efficient Vertical Handoff Approach for Heterogeneous Wireless Networks," *Wiley Wireless Communications and Mobile Computing*, Vol. 6, Issue 5, pp. 691-709, Aug. 2006.

[29] K. Pahlavan and P. Krishnamurthy, "Principles of Wireless Networks: A Unified Approach," *Prentice Hall PTR*, 2001.

[30] The network simulator ns-2, http://www.isi.edu/nsnam/ns/

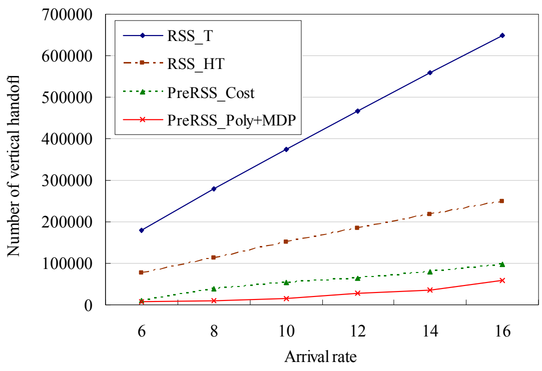

Fig. 4. Number of vertical handoffs under various arrival rates

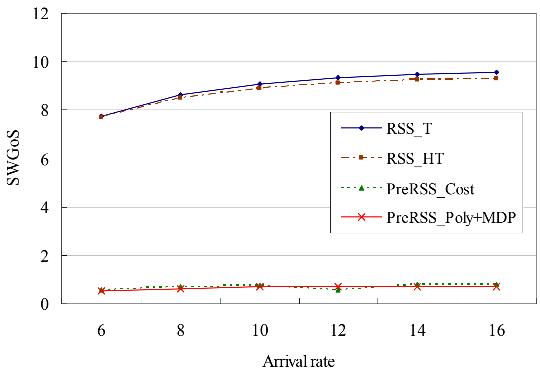

Fig. 5. SWGoS under various arrival rates

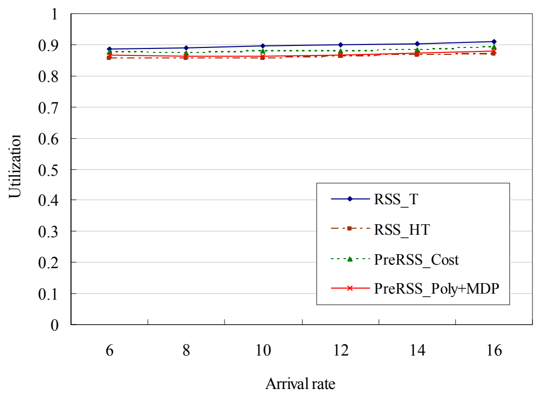

Fig. 6. Network utilization under various arrival rates