A Novel Passive Bandwidth Reservation Algorithm based on Neural Networks Path Prediction in Wireless Environments

Peppino Fazio, Floriano De Rango, Ivano Selvaggi
Dept. of Electronic, Information and Systems (DEIS)
University of Calabria
Arcavacata di Rende (CS), Italy
{pfazio, derango}@deis.unical.it, ivoselva@gmail.com

Abstract—**In this paper, providing Quality of Service (QoS) to mobile hosts belonging to a real-time service class is the main issue and a new path prediction scheme based on Neural Networks (NNs) is proposed, in order to obtain a good management of system resources and service continuity, by making advanced bandwidth reservation on the Access Points (APs) where a generic user will move into. The prediction algorithm is based on two types of neural networks: the first one predicts the next cell by considering the current position and the direction of the host when the service request starts and the second one is recursively applied on previous predictions to obtain the sequence of predicted cells based on the past history. The choice of a "next cell" during the prediction phase is made by considering the probability that the user moves toward a specific adjacent cell. We hypothesize that users move according to the Smooth Random Mobility Model (SRMM) and the performance of the prediction system are analyzed in terms of prediction error for different hand-off events.**

Keywords – Prediction, Resource Reservation, Neural Networks, Wireless Networks

I. INTRODUCTION

In wireless environments, the congestion level that a Mobile Host (MH) can suffer is often different from a coverage area to another one, because the bandwidth management of each Access Point (AP) of the considered system is always independent from the neighbouring AP conditions. For example, the problem to maintain the continuity of multimedia services during a hand-off event is hard to solve: in wireless networks, one of the most important QoS factor is the hand-off Call Dropping Probability (CDP).

When a MH hands-off from the old AP to a new one, the available resource in the new coverage system may be inadequate to provide to the MH the expected service (the congestion level may vary, while the perceived service quality may fall below requested lower bounds, in the worst case, the connection may be dropped). Thus, when a MH executes a hand-off procedure, it may find scarce resource availability in the new location and the active connection can be seriously degraded. One of the most important issue is the guaranteeing of adaptive QoS to MHs in an appropriate architecture, capable to reserve bandwidth levels and to offer guaranteed services: it is the Integrated Services Packet Networks (ISPNs) with MHs, while Mobile ReSerVation Protocol (MRSVP) is applied for exchanging state information of wireless networks [1]. Owing to the problems above, deterministic service guarantees, commonly used in wired networks, become unsuitable in a wireless scenario and a flexible service model, which allows variable QoS, is needed [2,3].

The only way to ensure a certain grade of QoS and service continuity to mobile users is the employment of the passive reservation policy: reserving bandwidth for a MH not only in the cell when the call has originated, but making it over all the cells that the MH will probably visit during the active connection. So, the QoS guarantee can be achieved if enough resources are reserved before a new cell is visited, that is to say an "exact" prediction of next crossing cell that a MH is going to enter is made. The proposed algorithm does not make a step-by-step prediction (as the most of the works in literature), but it predicts the whole set of probably visited cells, before the flow between source and destination is activated.

The MRSVP is based on active and passive reservations and it is capable to pre-reserve a certain amount of bandwidth for Mobile Independent Predictive (MIP) (for tolerant real-time applications, that can allow some bounded data packet delay variations) flows in the current cell and in the future ones, guaranteeing the desired QoS during hand-off events [2]. The main contribution of this paper is the enhancement of the pre-reservation phase through the introduction of neural networks, using them in a recursive way in order to make the needed advanced reservations; in this way, the wastage of resources is heavily avoided and system utilization is seriously improved, making MIP users having an adequate service continuity.

The rest of the paper is organized as follows: next section presents a summary of the existing related work that concerns the topic of this paper; brief overviews of the SRMM and Mobile RSVP are given in section III; the prediction algorithm is presented in section IV and the obtained results with conclusions are respectively summarized in sections V and VI.

II. RELATED WORK

There are many works in literature regarding the issue of resource reservations prediction in ISPNs, where each flow can receive a different QoS, which must be negotiated at the beginning of sessions, between flows and network, by the RSVP protocol or the Mobile-RSVP/Dynamic-RSVP protocol in mobile scenarios [4].

It is very important to describe accurately the movement patterns of mobile users in wireless cells, analyze them and, then, make the appropriate predictions in order to obtain the desired guarantees. In [5] a next location prediction algorithm based on an MMP (Mobility Pattern Profile) is presented. A MMP has been designed to estimate individual mobility based on each MH's movement history. They make use of MMP to predict the probability of entering a certain next crossing cell. In [6] the Adaptive Bandwidth Reservation Scheme is adopted: it provides QoS guarantees and lower connection dropping probability by reserving bandwidth in the target cell and all neighboring cells simultaneously. The Shadow Cluster (SC) concept is introduced. Cells exchange MH Location and movement pattern information for nearby cells after every period of time to predict hand-off probability for MH and reduce unnecessary waste reservation of bandwidth.

In [7,8] a prediction technique based on the cell stay time (CST) evaluation of a mobile user under a Random Way Point mobility model is proposed. A formula that binds speed, cell radius and variation around the average speed is calculated and resource reservation techniques have been proposed.

In [11] the authors propose a prediction-based location management scheme for locating a MH, which depends on its history of movement pattern. A Multilayer Neural Network (MNN) model for mobile movement prediction is designed to predict the future movement of a MH. The performance of the method has been verified for prediction accuracy by considering different movement patterns of a MH and learning accuracy.

In [14] the authors propose another prediction algorithm for an efficient resource reservation with mobility prediction. Their work is based on an algorithm which stores and uses the history of user's positions within the current cell to predict the next cell. All the mentioned works present a good efficiency and precision when predicting one cell based on the knowledge of the current one, so they perform a step-by-step prediction; none of them is designed for in-advance predicting (although with a certain percentage of error) all the cells that user will probably visit. The work illustrated in this paper is based on a new idea concerning Neural Networks (NNs) theory [10], which helps to solve the main problem of low system utilization when heavy MIP traffic percentage is present in the system.

III. SMOOTH RANDOM MOBILITY MODEL (SRMM) AND SIGNALING PROTOCOL

The management of user mobility and its analysis is an important issue before the application of the prediction algorithm. The employed predictors can be trained and an efficient set of passive reservations can be determined, as illustrated in next section).

Many mobility models in the literature consider that the new choice for speed v and direction φ is not correlated to previous values (such as in the Random WayPoint Mobility Model). This may cause unrealistic movement behavior with sudden speed changes ($\frac{\partial v(t)}{\partial t} \rightarrow \infty$) or sharp turnings (high $\frac{\partial \varphi(t)}{\partial t}$ when v is high). The SRMM includes both autocorrelation features. This work employs the Smooth Random Mobility Model (SRMM) proposed in [9] for a two-dimensional set of cell clusters; this model makes users' movements smoother and more realistic than previous random models, because it relates speed and direction changes; in addition it leads to a general set of analytical expressions, that can be used for different wireless scenarios (urban, rural, etc.). The speed is changed incrementally by the current acceleration and also the direction change is smooth: once a station is intended to turn, the direction is (in general) changed in several time steps until the new target direction is achieved. For details please refer to [9].

The Internet best-effort service does not offer any guarantee about available bandwidth, network propagation delays, jitter and packet delivery. MIP services use a pre-reservation phase to reserve bandwidth for a MH in the current cell and in the cells that MH will probably visit (passive and active reservations). An active reservation is made by a user only on the current access point (for MDP class, as will be seen later), while passive reservations are made only on the remote cells that the user will visit during its connection (users belonging to MIP class request passive reservations). In a typical 2D wireless scenario each wireless cell is covered by an access point (or base station, depending on the context) that is wired connected to the sender node through a wired Routing Subnet; a MH makes a service request in a certain coverage area through a RESV message; the reservation request is then routed to the sender node by an appropriate set of MRSVP messages. The cell where the request originates is the active one, while the future next cell is the passive one. Before making the real reservation request, the MH must ensure itself that there is enough bandwidth availability on the current cell (and on the passive ones if it belongs to MIP class), so it sends the Pre_Reservation (Pre_RESV) message to the local access point (and to the remote ones through its local access point and the switching subnet if it belongs to MIP class); the involved access points answer with a positive acknowledgement if possible. If the MH does not receive all the positive acknowledgements, then it will try the connection later; on the contrary, it will perform its reservation request by sending the RESV message, that can be an active_RESV if the request is made only to the local access point or a SPEC if it must reach some remote access points for passive in-advance reservations. Finally, the access points answer with a positive confirmation if the request can be accepted. Details about the adopted Call Admission Control (CAC) scheme can be found in [8], while details on MRSVP can be found in [4].

IV. RECURSIVE PATH PREDICTION WITH BASED ON NNS

In the last years, Neural Networks (NNs) have been considered not only in the economics studies but also in telecommunications issues. NNs are well-developed for learning functional relationships and they arise from imitations of biological neural systems, providing a simple application of parallel computation and have been extended to solve system learning and optimization problems [10].

Any function can be approximated by a piece-wise linear function, which can be denoted as a "basis function". Let $\delta(k,\rho)$ be the general basis function used for approximation, where k is the index of the function and ρ is the state. The state can be considered in terms of function samples, where the function is evaluated at a finite set of "states" ρ_l, $l=1,2,\dots$. For simplicity the term ρ_l is exchanged with l directly, so a function $F(\rho)$ can be approximated by a basis:

$$\widetilde{F}(l,r)=\sum_{k=0}^{K} r_k \delta(l) \tag{1}$$

where $r=\{r_1, \dots, r_K\}$ are the weights associated with the basis set $\{\delta_0(l), \dots, \delta_K(l)\}$. Eq. (1) represents a single-layer NN and for more complex functions, multiple layers may be used [10].

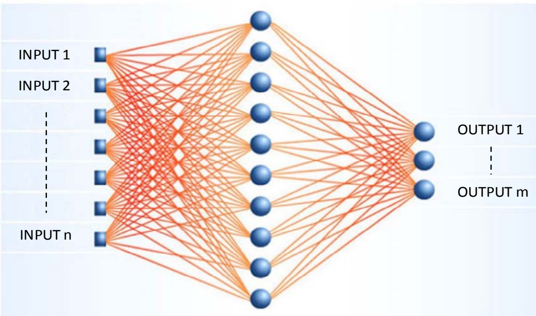

Figure 1. The generic structure of a NN.

In this work the, so-called, multi-layer NNs are used: they contain two or more single-layer networks, such as:

$$\widetilde{F}(l,r)=\sum_{k=1}^{K} r(k)\theta\left(\sum_{t=1}^{T} r(k,t)x_t(l)\right) \tag{2}$$

where the base function $\theta(s)$ is a smooth monotonic function taking values in (0,1) or in (-1,1) (e.g. $(1+e^{-s})^{-1}$ or $tanh(s)$); fig. 1 represents a generic structure of a NN.

The performance function of the neural network is normally chosen to be the Minimum Mean Squared Error (MMSE) for each pattern on the training set. The best approximation, in an MMSE sense, for a given set of basis functions is:

$$r=\arg\min_{r\in R^{K+1}}\sum_{l}\left|F(l)-\widetilde{F}(l,r)\right|^2 \tag{3}$$

Sometimes, there are functions of the state known to be important or useful in the prediction. In these cases, intermediate functions, called features, are introduced to capture the important aspects of the current state. Let the feature vector associated with state l be denoted as $f(l)$, the single layer network now can be written as $\widetilde{F}(l,r)=\sum_k r_k\delta_k(f(l))$. Features can be obtained by prior-knowledge of the network or heuristic policies.

As known in literature, a NN is composed of three main parts: the input vector (or input layer), the hidden layer and output vector (or output layer). In order to choice an appropriate structure for the NN, it must be noticed that, for our purposes, when a wireless connection starts and the service request is made, no past information about user's movements is available, so for the first hand-off event the prediction must be different from the next ones; for this reason, two kinds of reservations are necessary before the initiation of an active flow in our scenario and two different NNs are considered:

- *NN1*: the prediction of the first adjacent cell after the active one is mandatory, but no information about history movements are available;
- *NN2*: from the second prediction till the end of the predicted path, a different structure is required, because historical information are available (like previous predicted cells).

The number of probably visited cells can be obtained with the same approach of [7], taking into account the Call Holding Time (CHT) distribution and the Cell Stay Time distribution (CST), so the number of hand-off events can be a-priori determined (so, the number of iterations of the prediction algorithm is also known). A Wireless-LAN (WLAN) coverage area, generally with a circular shape, can be approximated with a m-edge regular polygon as depicted in fig. 2 (m can be considered as an input control parameter). A set S_{ho} of m possible movement directions can be then obtained: if movement directions are indicated with $d_1 \dots d_m$, where $d_j=\underline{\theta}\cdot(2\cdot j-1)/2$ rad., $\underline{\theta}=2\pi/m$ rad. and $j=1..m$, then $S_{ho}=\{d_1, .., d_n\}$ and $|S_{ho}|=m$. In our work the value of m has been fixed to 6, because this ensures a good level of accuracy and makes the polygonal approximation also suitable for cellular networks.

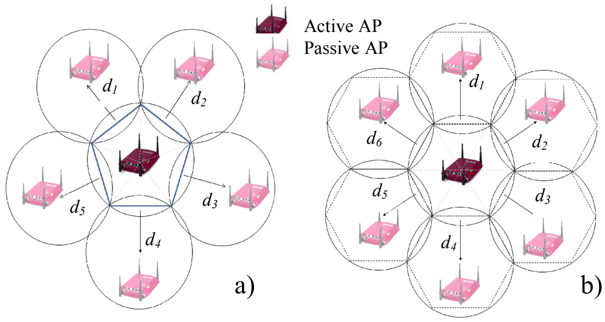

Figure 2. AP coverage area approximation with a) m=5 and b) m=6.

If C is the set of the considered cells (the cells that cover the simulated geographical region), a path P can be considered as a set of n movements from a source point to a destination and can be indicated as:

$$P=\left\{(c,d)\mid c\in C, d\in S_{ho}{}^{n}\right\} \tag{4}$$

The main role of NNs in our scheme is the employment of their ability to capture the relations between the past and the future positions of MHs, if a mobility model is fixed.

In order to use a NN, it is necessary to choose its topology, the neural model, the number of neurons and the training algorithm. Different topologies are proposed for NNs and each one is dedicated to a specific purpose; after a deep analysis of the literature [10],[11],[12],[15],[16] it has been observed that a feed-forward NN with Back-Propagation (BP) [12] training algorithm is adequate for our purposes (we tried to use also Resilient BP R-Prop or Broyden-Fletcher-Goldfarb-Shanno method BFGS, but no differences in the final results have been observed).

Input and output values belong to the range [0, 1], so each neuron of the output and hidden layers can use a sigmoidal function [10]: as explained later, the output of the considered neural networks consists of the set of m probability values, where the *i-th* one represents the probability of handing-off the current cell on direction i.

The movement pattern of *NN1*, indicated with X^{NN1}, represents the current position and the movement direction of the considered user, so:

$$X^{NN1} = (bs_i, ap_i, d_i), \tag{5}$$

where bs_i indicates the cluster identifier, ap_i indicates the cell identifier and $d_i \in S_{ho}$ indicates the movement direction of the sample i. So the number of input neurons is 3 (since we have 3 input values), the number of output neurons is m (as the number of fixed directions) and, as demonstrated in [12], a number of neurons in the hidden layer equals to 20 is adequate, because makes the network accurate and not so complicated in terms of computations, during the training and the prediction phases (many campaigns of simulations with higher numbers of nodes in the hidden layer have been carried out and no enhancements have been observed).

For *NN2*, the movement pattern indicated with X^{NN2} represents the history of movements of the considered user, where v is its dimension, so:

$$X_v^{NN2} = \{d_1, d_2, ..., dv\}, \tag{6}$$

where d_i is the direction at *i-th* step. After different simulations, we decided to fix v=5 (it leads to a negligible error while considering history pattern). The output set is the same of *NN1*. The structures of *NN1* and *NN2* are illustrated in fig. 3.

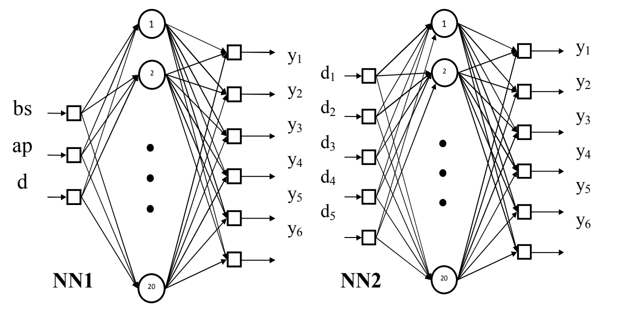

Figure 3. Network structures for NN1 and NN2.

In [12] the classic NN prediction performance is analyzed: the error goes from 7% to 70% (if considering regular or casual paths); so, the main aim of this paper is the enhancement of pattern prediction. The basic idea is the estimation of next-cells probabilities, instead of the prediction of a single next cell. For this reason, the number of output neurons is equal to m. The output of *NN1* and *NN2* is a vector Y of probabilities of handing-off to the m directions for the given movement pattern X_i. To carry out a single movement prediction, the considered input is the pattern X_i given to the NN and the obtained output will be the vector Y_{i+1}, which contains direction probabilities (it is used to determine the movement pattern X_{i+1}). A recursive method has been designed for multiple movement predictions: the input pattern X_{i+1} is obtained from predicted output Y_{i+1} and it is used to generate the input pattern for the next step X_{i+2}.

As discussed, the chosen training algorithm is the BP; the algorithm that is executed for user x can be resumed with the following pseudo-code:

Init) *Determine the number H of probably visited cells;*
Determine bs_0, ap_0 and d_0 for user x; set h=0;
Step 0) *Create an input pattern $X_0^{NN1}=(bs_0, ap_0, d_0)$;*
Step 1) *Activate NN1 with X_0^{NN1} and store the output Y_i;*
Step 2) *While (h<H)*
Step 3) Evaluate from Y_i the probabilities of each near cell for the i-th step of the prediction; h++;
Step 4) For each probability value that is higher than a fixed threshold T, do steps 5-10;
Step 5) *Take the direction d_{ij} associated with $Y_i[j]$;*
Step 6) *Ask the network to obtain the identifiers bs_i, ap_i of the adjacent cell on direction d_{ij};*
Step 7) *Make the passive reservation on bs_i, ap_i;*
Step 8) *Create an input pattern $X_i^{NN2} = (d_{i-l}, d_{i-l+1}, ..., d_{i-1}, d_i)$ starting from the history of user's direction;*
Step 9) *Activate NN2 with X_i^{NN2} and store the output Y_{i+1};*
Step 10) *Evaluate from Y_{i+1} the probabilities of handing-out on the m different directions;*
End;
End.

Concerning the Time Complexity *TC* of the algorithm, from [15], for a feed-forward neural network having N_I input nodes, N_H hidden nodes and N_O output nodes, the time required to propagate an input pattern to the output nodes is $O(N_I N_H + N_H N_O)$. So, from the pseudo-code above, it is clear that *TC* is $O(N_{I1}N_{H1}+N_{H1}N_{O1}) + +O[H*m*(N_{I2}N_{H2}+N_{H2}N_{O2})]$; setting N_{MAX} as the maximum number of nodes in a layer, then $TC= =O[H*m*N^2_{MAX})]$. Two main goodness metrics are considered:

a) Prediciton error e: it can be defined as the percentage of cells on which the MH does not find a passive reservation:

$$e = \frac{err}{len}, \tag{7}$$

where *err* is the observed number of errors in the prediction and *len* is the length of the pattern (in terms of number of cells that the MH will probably visit);

b) Prediction wastage w: it can be defined as the percentage of cells that have not been visited during the active session of the MH:

$$w = \frac{P - len}{size}, \tag{8}$$

where P is the number of predicted cells, len is the length of the pattern and $size$ is the dimension of the simulated network (in terms of cells). At this point, the considered metric can be defined in order to evaluate the goodness of neural predictions through an index:

$$index_{goodness} = \alpha \cdot e + (1 - \alpha) \cdot w\,, \tag{9}$$

where α is a weight and indicates the importance that is given to the wastage and to the error.

V. Performance Evaluation

Simulated network consists of 7 clusters of wireless cells with a radius of 250 meters; users move toroidally, according to the SRMM, with the same mobility parameters of [8]. An exponentially distributed CHT with mean λ=180s has been considered.

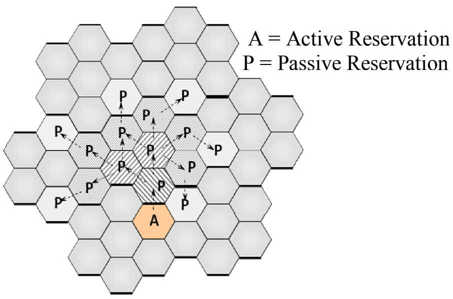

Figure 4. The simulated network and an example of passive reservations.

Many campaigns of simulations have been carried out, in order to appreciate the correctness of the proposed prediction scheme. The main parameters of the implemented simulator, based on the NNFW [13], are the threshold T (if the relation y_i>T is verified then the adjacent cell on direction d_i is considered for passive reservation) and the maximum number of cells (MAX_c) on which the resources should be reserved in a passive way for the considered hand-off event; α has been fixed to 0.5 in order to give the same level of priority to e and w. Fig. 5 shows the trend of the goodness index for *NN1*, varying the values of T and MAX_c: some combinations of the values of T and MAX_c must be disregarded because they lead to an unacceptable prediction error. The minimum value of i is obtained for T =0.2 and MAX_c=2.

Fig. 6 illustrates the trend of the goodness index for *NN2*: the same considerations of the previous case can be made; in fact the optimum (minimum) value is obtained for T =0.2 and MAX_c=2. Fig. 7 shows the trend of the number of tried corrections, during flows lifetimes: when a user reaches a "non-predicted" cell, the algorithm for NN2 is repeated one more time, taking as input vector the sequence of visited cells and, if possible, a new amount of bandwidth is reserved on newer cells; for the optimum values of T and MAX_c the number of needed corrections is 1.

Fig. 8 illustrates the trend of the prediction error for *NN1* and *NN2*, with and without the correction action: it is evident that if the correction action is carried out, the prediction error decreases and the difference is more evident for higher T values.

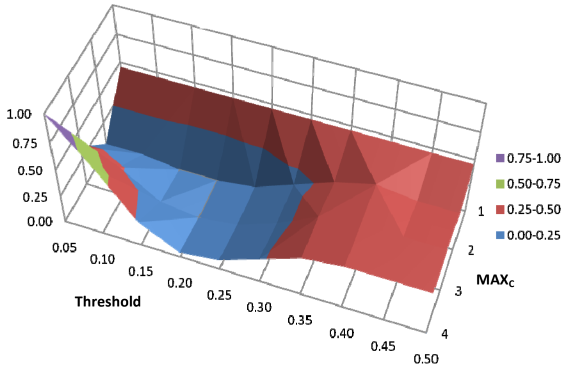

Figure 5. Goodness trend of NN1.

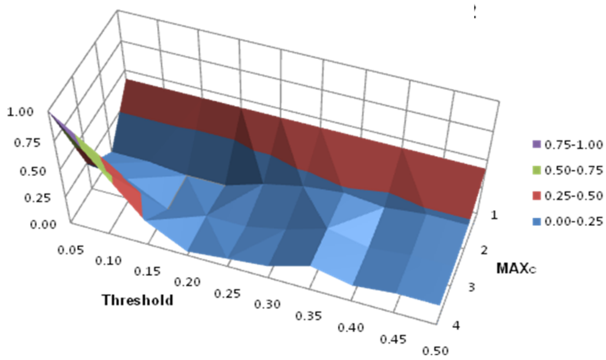

Figure 6. Goodness trend of NN2.

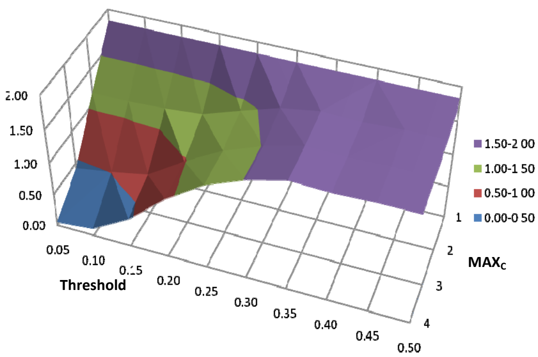

Figure 7. Average number of tried trajectory corrections.

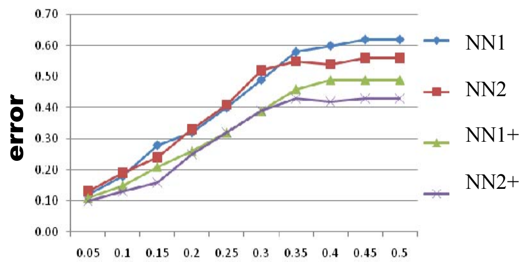

Figure 8. Prediction error for NN1 and NN2.

VI. CONCLUSIONS

In this paper a new pattern prediction scheme for wireless networks and based on NNs has been proposed. The proposed scheme faces the problem of pre-reserving passive bandwidth for MIP flows over the cells that compose the system, trying to minimize the wastage of passive resources. The main aim is the prediction of the most probable cells that MHs will probably visit; it is based on the study of mobility models. Simulations results have revealed that the prediction scheme performs well in terms of error and resource wastage, in fact the proposed idea reaches a precision of 69%-84% with a system utilization of 75%-90%. In this way, proper QoS guarantees can be given.

REFERENCES

[1] Talukdar A. K., Badrinath B.R. and Acharya A., "On Accomodating Mobile Hosts in an Integrated Services Packet Network", in the *Proceedings of the IEEE INFOCOM '97*, April 1997.

[2] A.Alwan, R.Bagrodia, N.Bambos, M.Gerla, L.Kleinrock, J.Short, and J.Villasenor, "Adaptive Mobile Multimedia Networks", *IEEE Personal Comm. Magazine*, Apr.1996, pp.34-51.

[3] Gomez, J. and A.T. Campbell, "Havana: Supporting Application and Channel Dependent QOS in Wireless Networks", *ACM Journal on Wireless Networks (WINET)*, Vol. 9, Issue 1, pp. 21-35, January 2003.

[4] A.K.Talukdar, B.R.Badrinath, A.Acharya, "MRSVP: A Resource Reservation Protocol for an Integrated Services Network with Mobile Hosts," *Wireless Networks*, Kluwer Journal, pp.5-19, 2001.

[5] C. Choi, M. I. Kim, T. J. Kim, S. J. Kim, "Adaptive Bandwidth Reservation Mechanism Using Mobility Probability in Mobile Multimedia Computing Environment," IEEE Local Computer Networks, Nov. 2000, pp. 76-85.

[6] D. A. Levine, I. F. Akyildiz, and M. Naghshmeh, "The Shadow Cluster Concept for Resource Allocation and Call Admission in ATM-Based Wireless Networks," IEEEIACM Trans. Networking, vo1.6, Feb. 1997, pp. 1-10.

[7] F. De Rango, P.Fazio, S.Marano, "Mobility Prediction and Resource Reservation in WLAN Networks under a 2D Mobility Models," *63rd Vehicular Technology Conference* (VTC Fall 2006), Montreal, Canada, Sept..25-28, 2006.

[8] F. De Rango, P.Fazio, S.Marano, "Utility-Based Predictive Services for Adaptive Wireless Networks with Mobile Hosts," I*EEE Transaction on Vehicular Technology* (March.2009).

[9] C.Betstetter, "Mobility Modeling in Wireless Networks: Categorization, Smooth Movement, and Border Effects," *ACM SIGMOBILE Mobile Computing and Communications Review*, Vol. 5, Issue 3, pp. 55–66, July 2001.

[10] L. Fausett, "Fundamentals of Neural Networks: Architectures, Algorithms and Applications", Prentice Hall 1994.

[11] B. P. Vijay Kumar, P. Venkataram, "Prediction-based location management using multilayer neural networks", 2002.

[12] J. Biesterfeld, E Ennigrou, K. Jobmann, "Location Prediction in Mobile Networks with Neural Networks", Jun. 1997, pp. 207-214.

[13] Neural Network FrameWork, http://www.nnfw.org.

[14] S. Kwon, H. Park, K. Lee, "A Novel Mobility Prediction Algorithm Based on User Movement History in Wireless Networks " , lecture notes in computer science in Systems Modeling and Simulation, SpringerLink 2005.

[15] K.J. Batenburg, W.A. Kosters, "A Neural Network Approach to Real-Time Discrete Tomography", IWCIA LNCS 4040, pp. 389–403, Springer-Verlag Berlin Heidelberg, 2006.

[16] S.-C. Liou, Y.-M. Huang, "Trajectory Predictions in Mobile Networks", International Journal of Information Technology, Vol. 11, 2005.