

# End-to-End Bandwidth and Available Bandwidth Estimation in Multi-Hop IEEE 802.11b Ad Hoc Networks

Marco A. Alzate, José-Carlos Pagán, Néstor M. Peña, and Miguel A. Labrador

**Abstract**—In this paper we estimate the end-to-end total bandwidth (BW) and available bandwidth (ABW) of a path between a pair of nodes in an IEEE 802.11b ad hoc network, both as functions of packet length, using dispersion traces between probing packet pairs of different lengths. The pairs of packets that suffer the minimum delay are used to estimate the maximum achievable transmission rate (BW), while the variability of the dispersions is used to estimate the fraction of that BW that is available for data transmission (ABW). To model the dependence of the variability of the dispersion traces on the ABW, we use a neuro-fuzzy estimator. The system is shown to be timely, accurate and efficient.

**Index Terms**— Packet pair active probing, interference, neuro-fuzzy modeling.

## I. INTRODUCTION

ACCURATE and timely estimation of the end-to-end available bandwidth (ABW) is very important in such applications as resource allocation, data rate adjustment, traffic engineering, resource-constrained routing, capacity planning, peer-to-peer file sharing, etc. Although many useful tools to estimate this path parameter in wired networks exist (see, for example, [1] and references therein), they all consider the ABW as the unused capacity of the tight link. Unfortunately, this is not an applicable concept to wireless ad hoc networks due to the unreliable and shared nature of the medium [2].

Several researchers have tried to elucidate the throughput properties of wireless ad hoc networks considering mobility, radio transmission, and multiple access (see, for example, [3] and references there in). However, these important theoretical results have not led to practical estimation tools yet. Some estimators measure the fraction of time a node senses the channel idle, and multiply this fraction by the physical transmission capacity of the node (as in [4], among others). This measure is shared among the nodes of a path to estimate the ABW, which is calculated as the minimum measure

among the individual nodes (as in [5], for example). This per node estimation is not quite correct because it doesn't consider the occupation times of those links that cannot be used simultaneously, nor the additional overhead incurred when trying to use that idle capacity. Other authors use the self congestion principle to obtain the maximum achievable throughput (e.g. [6]) but, although this principle provides a more realistic measure of the available bandwidth, it poses important intrusiveness concerns in resource-scarce environments, such as wireless networks.

In this paper, we integrate an active probing technique that estimates the bandwidth (BW) of an end-to-end path between a source and a destination in an IEEE 802.11b ad hoc network [2] with a new neuro-fuzzy estimator that finds the end-to-end available bandwidth.

The gaps between those pairs of packets that suffer the minimum sum of one-way-delays are used to estimate the maximum achievable transmission rate (or bandwidth, BW) as a function of packet length, for any packet length. Then the variability of the dispersions is used to estimate the fraction of that bandwidth that is effectively available for data transmission, also as a function of packet length.

To consider implicitly all the phenomena that jointly affect the available bandwidth and the dispersion measures, a neuro-fuzzy identification system models their dependence. For example, even in the absence of competing flows, ABW can be less than BW due to self interference (when consecutive packets of the same flow compete among them on different links of the path). ABW can be further reduced by cross-traffic, which can (1) simply decrease the signal-to-noise ratio at some parts of the path, or (2) can interact through MAC arbitration at some other parts of the path, or (3) can even share some common queues along the path. Another largely ignored aspect that is indirectly captured by the neuro-fuzzy system is the fact that ABW can differ from the unused capacity because, once that unused capacity is to be occupied, the arrival of the new flow can re-accommodate the occupation pattern along the neighborhood of its path.

In order to consider all these interacting aspects, the neuro-fuzzy estimator is trained on data collected from a large set of simulated scenarios, for which we used Qualnet@[7]. We carefully selected the scenarios to have enough samples of each of the effects mentioned above, and combinations of them, over a wide range of configuration parameters, so as to get a representative set of data for training, testing, and validation purposes. With all these data, the estimator learns how to infer the available bandwidth from the variability of

M. A. Alzate is with Universidad Distrital, Bogotá, Colombia (e-mail: malzate@udistrital.edu.co).

J-C Pagán is with Universidad de Puerto Rico, Mayagüez, PR (e-mail: jcpagan@cse.usf.edu).

N. M. Peña is with Universidad de los Andes, Bogotá, Colombia (e-mail: npena@uniandes.edu.co).

M. A. Labrador is with University of South Florida, FL (e-mail: labrador@cse.usf.edu).

This work was partially supported by Team TACLAN under contract number TACLAN-USF-07-16.

the dispersion traces. The system is designed so as to have good generalization properties and to be computationally efficient. As a result, we get an accurate, efficient, and timely end-to-end available bandwidth estimator.

The paper is organized as follows. Section 2 discusses the concept of Bandwidth in an IEEE 802.11b multi-hop ad hoc network and describes our bandwidth estimation tool. Section 3 describes some scenarios that show how this bandwidth can be reduced due to self and cross-interference, leading to our concept of available bandwidth. Section 4 describes the design process of our neuro-fuzzy estimator. Section 5 presents some evaluation results and Section 6 contains the conclusions of the paper.

## II. END-TO-END BANDWIDTH

Consider a pair of neighbor nodes A and B for which all network resources are completely available. The average time it takes A to transmit an  $L$ -bit packet to B has the form  $T(L) = \alpha_1 L + \beta_1$ , where  $\alpha_1$  is the cost, in seconds per bit, of transmitting one bit, and  $\beta_1$  is the average overhead cost, in seconds per packet, corresponding to the control timers (including the random backoff) and the transmission of control information (packet headers, acknowledgements and, possibly, RTS/CTS frames). So, the maximum transmission rate achievable over that link with a flow of  $L$ -bit packets is  $BW^{link}(L) = L/(\alpha_1 L + \beta_1)$ . If the packets are to be sent under the same conditions over a multi-hop path where all the links belong to the same spatial channel (i.e. where the individual links of the path can only be used one at a time), collisions could occur due to self interference. If we assume perfect scheduling of packets, the maximum transmission rate achievable over that path with a flow of  $L$ -bit packets is  $BW^{channel}(L) = L/(L\sum\alpha_j + \sum\beta_j)$ , where  $\alpha_j$  and  $\beta_j$  are the corresponding parameters of the  $j^{th}$  link in the path. Finally, on a longer path, simultaneous transmissions over some links would be possible if they belong to different spatial channels. In this case, the spatial channels are given by the maximal cliques in the contention graph (see [2] for details) and the bandwidth of the end-to-end path will be given by the minimum of the bandwidths of the constituent channels. If the path goes through  $H$  spatial channels with  $n_i$  links in the  $i^{th}$  channel, the end-to-end bandwidth of the path will be

$$BW^{path}(L) = \min_{i=1\dots H} \frac{L}{L \sum_{j=1}^{n_i} \alpha_{i,j} + \sum_{j=1}^{n_i} \beta_{i,j}} \quad (1)$$

where  $\alpha_{i,j}$  and  $\beta_{i,j}$  are the parameters of the  $j^{th}$  link of the  $i^{th}$  channel. Notice that, in a wired network, each link is an independent spatial channel, so that  $n_i = 1$  and  $\beta_{i,1} = 0$ , leading to the widely accepted equivalence between the end-to-end maximum achievable transmission rate and the capacity of the narrow link in the path,  $C = \min_{i=1\dots H} C_i$  [1]. In general, whether we have a single-hop path, a multi-hop path within a single spatial channel, or a multi-hop multi-channel path, the end-to-end bandwidth will take the form

$$BW^{path}(L) = \frac{L}{\alpha L + \beta} \quad (2)$$

The estimation procedure developed in [2] sends pairs of back-to-back probing packets of lengths  $L_0$  and  $L_1$  bits every  $T$  seconds. If each packet of a pair of  $L_i$ -bits suffers no queueing, retransmission or scheduling delays, then their sum of one way delays (*sowd*) would be minimum and the dispersion between them (or *gap*) will reveal a sample of the transmission time  $\alpha L_i + \beta$ , for  $i=0,1$ . If the network is not heavily loaded, the probability of having such a pair will be greater than zero, so we will get at least one valid measure with high probability if we send enough pairs. Having  $BW^{path}(L_0)$  and  $BW^{path}(L_1)$ , we can solve for  $\alpha$  and  $\beta$  in Equation (2), in order to obtain the end-to-end bandwidth as a function of the packet length (see [2] for validating experiments).

## III. END-TO-END AVAILABLE BANDWIDTH

In wired networks, it is widely accepted the definition of ABW as the unused capacity of the tight link. But in wireless multi-hop ad hoc networks, the unused bandwidth can differ from the available bandwidth because, due to interference, the unused capacity may not be completely available. Indeed, once a new flow is established in the given path to occupy some of that unused capacity, the interfering cross-traffic can re-accommodate itself in response to the new flow, changing the perception of the new flow about its available bandwidth. So, it is tempting to define ABW as the throughput achieved by a saturated source. However, due to self interference, a saturated node could reduce its throughput far below of what a less impatient source might obtain. The ABW could also be defined as the maximum achievable transmission rate that does not disturb current flows, but this is a very elusive definition because, due to the interactions in the shared medium, even a very low rate new data flow could affect current flows.

Just as an illustration of some of the possible interactions between a pair of neighbor links, consider the simple scenario of Figure (1a). Two links of 2 Mbps are formed by pairs of nodes 320 m apart, transmitting 1400-byte packets. (In all our scenarios we use a two-ray path loss propagation model, 15 dBm of transmission power, and -89 dBm of receiver sensitivity.) The second link is slowly moving to the right, increasing the distance  $d$  between senders. Figure (1b) shows the availability,  $x(L) = ABW(L)/BW(L)$ , for  $L = 1400$  bytes, with and without RTS/CTS signaling at the MAC layer. Up to 500 m, we can see the effect of the MAC protocol arbitrating the medium along with some unfairness against link 1 due to the capture effect. From that point on, it is no longer possible for the senders to sense the other's carrier directly. When the RTS/CTS mechanism is not used, this leads to repeated collisions that reduce the throughputs of both sources to very low values. When the RTS/CTS mechanism is used, Tx<sub>2</sub> can sense Tx<sub>1</sub>'s carrier virtually, but Tx<sub>1</sub> remains blind to Tx<sub>2</sub>'s transmissions. As the distance grows farther, the second link becomes isolated from any interference from the first link, while Rx<sub>1</sub> remains immersed in the "noise" produced by Tx<sub>2</sub>.

This leads to a big region of more than 500 m in which link 2 owns the whole channel bandwidth while link 1 can only use the leftovers. At 1200 m, the interference of Tx<sub>2</sub> on Rx<sub>1</sub> has been reduced to the point that some packets from Tx<sub>1</sub> begin to go through, but the links do not become independent until the distance exceeds one kilometer and a half.

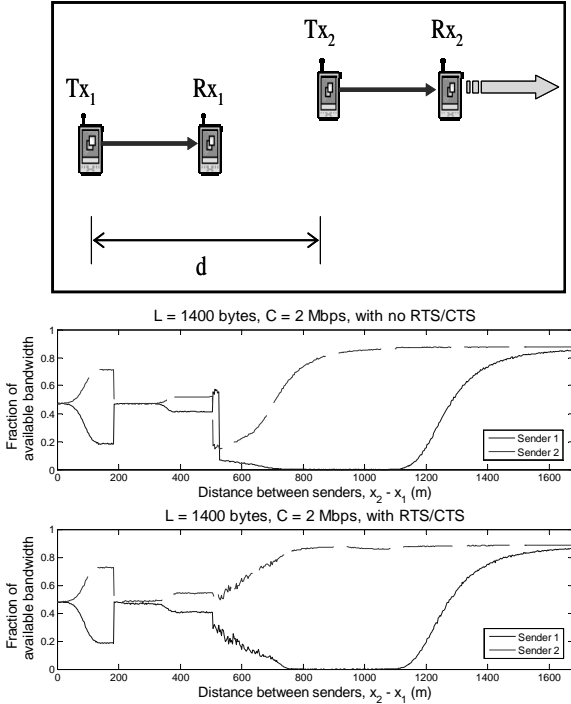


Fig. 1. (a) Scenario for studying the interaction between two competing links. (b) fraction of channel capacity available for each link.

Another possibility is to define ABW as the maximum throughput achievable by a CBR flow in the path, where the maximization is performed over the range of input data rates. Although, intuitively, this definition makes better sense, it is the most unfriendly for estimation purposes, because it requires the estimator to explore different transmission rates in order to find the one that maximizes the throughput. However, instead of doing this process on-line, we can collect accurate and representative data to feed a machine learning process that would relate the statistics of the packet pair dispersion measures with the true maximum achievable rate in the path. We first conduct an experiment to measure the probing packet dispersions and, then, we replicate the same experiment to measure the available bandwidth as the maximum achievable throughput. Then we compute the ratio between the maximum achievable throughput and the bandwidth (or “availability”,  $\chi = ABW/BW$ ) in order to relate it to the variability of the dispersion measures. The underlying hypothesis is that, since the probing packet pair dispersions are affected by the same phenomena that determines the current ABW, we could avoid the costly search of an optimal input rate if we can infer it from the statistics of the dispersion trace.

Consequently, our pragmatic definition of ABW would be given as follows:

$$ABW(L) = \max_{\lambda > 0} \left[ \lim_{t \rightarrow \infty} \frac{(n(t) - 1)L}{t - t_1} \right] \quad (3)$$

where  $L$  is the length, in bits, of the transmitted packets,  $\lambda$  is the transmission rate at the source,  $t_1$  is the reception time of the first of an infinite sequence of packets and  $n(t)$  is the number of packets received up to time  $t$ . To evaluate (3) experimentally, we send a large number (1000) of packets at the given rate  $\lambda$ . If the receiver gets less than 25% of the transmitted packets, we consider the loss probability is too high and set the available bandwidth to zero for that input rate. Otherwise, we compute the throughput for this rate as  $\Gamma(L; \lambda) = (n - 100)L / (t_n - t_{100})$ , where  $n$  is the last received packet, which arrived at  $t_n$ . We consider the first 100 received packets as part of the transient period. Then, through bracketing, we find the value of  $\lambda$  that maximizes  $\Gamma$ , which becomes  $ABW(L)$ . Since we can keep constant conditions in our experimental scenarios, this procedure gives a very accurate measure of ABW.

#### IV. NEURO-FUZZY END-TO-END AVAILABLE BANDWIDTH ESTIMATOR

Our work is aimed at designing a system capable of learning, from sample data, the intricate relations between ABW and dispersion measurements. So, we must collect a representative set of data to determine whether the dispersion measurements carry enough information for a significant estimation of ABW or not. If that is the case, we must use that data to train a neuro-fuzzy system. We review this process here.

##### A. Data Collection

With the procedure described in Section 3, we have the possibility to collect a large data set that relates the dispersion measurements of the active probing packet pairs with the corresponding ABW on different scenarios. The data set must reflect the most important features of the underlying characteristics of any IEEE 802.11b ad hoc network, which include the interaction between competing flows by buffer sharing, by MAC arbitrated channel sharing or, simply, by an increased noise.

All these aspects of the dynamic behavior of an IEEE 802.11b ad hoc network (and their combinations) can be captured using the network configuration shown in Figure 2, where we can change different parameters to explore a wide range of cross-traffic interference conditions. In particular, we changed the distance between nodes, the physical transmission rate, the use of RTS/CTS mechanism, the number of cross-traffic flows, the origin and destination of each cross-traffic flow, the transmission rate and packet length of each cross-traffic flow, and the buffer size at the IP layer. For each condition, we found the ABW between each of the 21 pairs of nodes of the second row, for four different packet lengths, averaged over 10 independent simulations. Then we replicated each experiment to take a dispersion trace of probing traffic for each measured ABW. This way we obtained 6000 samples,

where each sample consisted of a traffic dispersion trace, a corresponding  $BW(L)$  function, and four measured availabilities for four different packet lengths,  $\{x(L_i) = ABW(L_i)/BW(L_i), i=0..3\}$ .

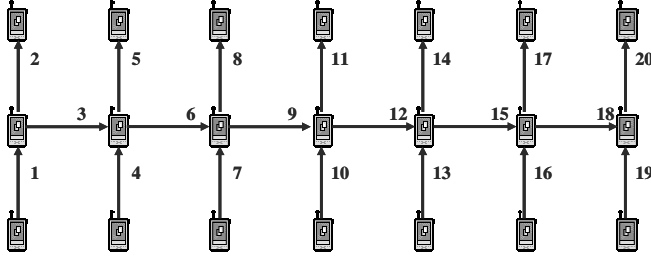


Fig. 2. Test Scenario for data collection.

We group the traces in analysis windows of 200 pairs overlapped every 4 pairs and, for each analysis window, we considered the following statistics of the dispersion trace for the two probing-packet lengths,  $L_0$  and  $L_1$ :

- $\theta_1(L_i)$  = mean of the gap between packets of a pair of  $L_i$ -bit packets
- $\theta_2(L_i)$  = standard deviation of the gap between packets of a pair of  $L_i$ -bit packets
- $\theta_3(L_i)$  = mean of the sowl (sum of one way delays) of a pair of  $L_i$ -bit packets
- $\theta_4(L_i)$  = standard deviation of the sowl of a single pair of  $L_i$ -bit packets

where, in each analysis window, the *gaps* and *sowds* are centered and normalized with respect to the gap between the packets that suffered the minimum *sowl*, in order to get comparable magnitudes over different network conditions. The vector of eight input parameters will be denoted as  $\theta$ , while the vector of four input parameters corresponding to a given packet length  $L$  will be denoted a  $\theta(L)$ .

Figure 3 shows the probability density functions (pdf) of each component of  $\theta(L)$  within the collected data for  $L_1 = 1400$  bytes, conditioned on a low (less than the median) or high (greater than the median) availability, where similar results hold for  $L_0=100$  bytes. A low availability tends to increase the values of the parameters and disperse them over a wider range, as compared to a high availability. It is this discrimination property what we want to exploit in our available bandwidth estimator.

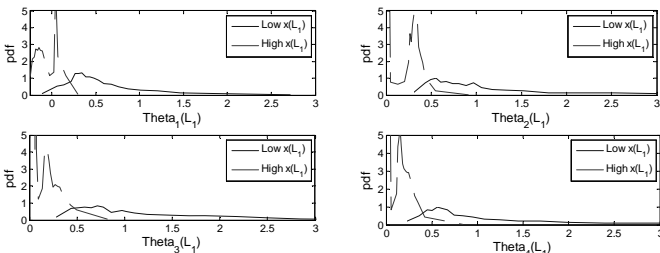


Fig. 3 Pdf of the availability (two upper plots) and the input parameters  $\{\theta\}$ , for short packets,  $L_0 = 100$  bytes (left group), and long packets,  $L_1 = 1400$  bytes (right group).

We first use a fuzzy clustering algorithm to identify regions in the input space that show strong characteristics or

predominant phenomena. Then, the clustered data is used to train simple neural networks, which can easily learn such phenomena. The local training data is selected through alphacuts of the corresponding fuzzy sets, and the antecedent membership functions are used to weight the outputs of the locally expert neural networks, according to the following simple rules

$$\text{IF } [\theta \text{ is in cluster } j] \text{ THEN } [x(L_i) = \text{neuralnet}(i,j)] \quad (4)$$

A large number of clusters can give a high accuracy at a cost of a high computational complexity. Since the efficient use of computational resources is an important requirement for our system (mobile devices are supposed to run on battery), we chose to locally train two networks on two different subsets of the input data. The local data was selected through a fuzzy c-means clustering algorithm on the whole set of input parameters. This choice leads to good regularity and generalization properties and a good compromise between bias and variance errors, while keeping a low computational complexity.

Figure 4 shows the two local regions projected over the availability space. Clearly, the circled light region ( $r_1$ ) exhibits some cross-traffic interaction while the dotted dark region ( $r_2$ ) corresponds to high availability and small interference conditions in the path, bringing a nice physical interpretation of the fuzzy rules of Equation (4). The global model takes the following form

$$\hat{x}(L_i) = f_i(\theta|r_1)\mu_{r_1}(\theta) + f_i(\theta|r_2)\mu_{r_2}(\theta) \quad (5)$$

where  $f_i(\theta|r_j)$  is the output of the locally expert network for  $L_i$ -bit packets in the  $j^{\text{th}}$  region, and  $\mu_{r_j}(\theta)$  is the membership function of the set of input parameters in the  $j^{\text{th}}$  cluster. The neuro-fuzzy estimator, shown in Figure 5, estimates the availability for four different packet lengths (100, 750, 1400 and 2000 bytes). Each neural network has eight neurons in the input layer, two neurons in the hidden layer, and one neuron in the output layer, for a total of 21 parameter per network. The cluster centers add 16 more parameters, so the global model has 184 parameters. The extra computational complexity resides in the computation of the statistics  $\theta$ , which is in the order of the analysis window size, 160 samples (320 packets). The total computational complexity is of  $O(n)$ , where  $n$  is the sum of the number of parameters of the model and the window length. In our case, this translates to 512 floating point operations per estimate.

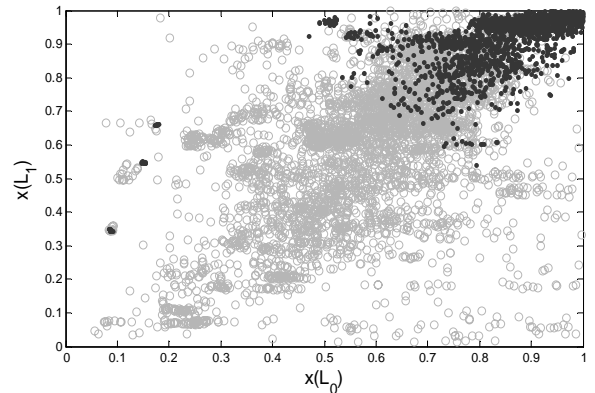


Fig. 4 Grouping input data in clusters helps in classifying the output data.

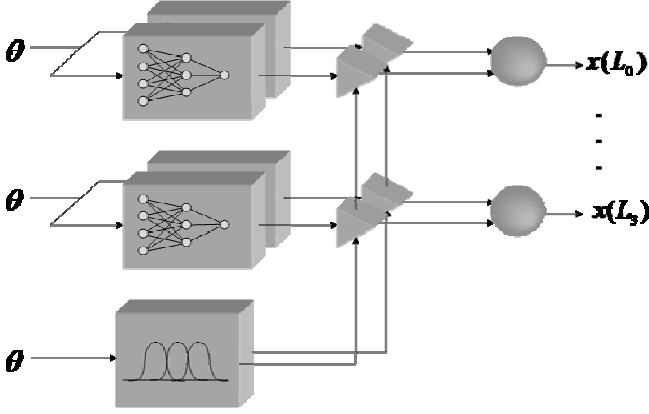


Fig. 5 Structure of the neuro-fuzzy estimator.

Locally trained submodels increased significantly the learning capacity of the whole system, as can be appreciated in Figure 6, which shows the final estimation results on the collected test data using the following settings: small packet size  $L_0 = 100$  bytes, large packet size  $L_1 = 1400$  bytes, time between pairs  $T = 0.25$ , analysis window size  $W = 320$  packets. We can notice that, unless the network is heavily loaded, the estimation is highly accurate. Indeed, the accuracy is within 15% for more than 80% of the test samples (which were never seen during training) and, within 10% for more than 90% of the samples for those samples with availability greater than 0.5.

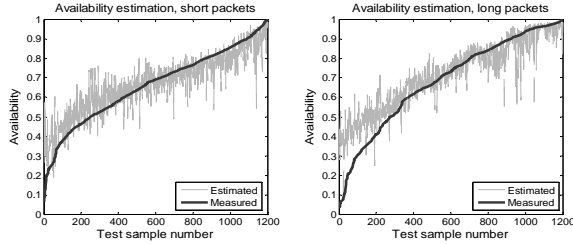


Fig. 6 Availability estimation on the test trace (data not seen by the system during training).

Having  $BW(L)$  and four samples of the availability  $x(L)$ , we can interpolate  $ABW(L)$  as a function of packet length by adjusting one of the following functional forms, according to the ratio of lost packets,

$$ABW_1(L) = \frac{L}{\mu L + \eta}, \quad ABW_2(L) = \gamma L \exp(-\lambda L) \quad (6)$$

## V. PERFORMANCE EVALUATION

We conducted many validation experiments with several scenarios using different network sizes, mobility conditions, data transmission rates, cross traffic intensities and configuration parameters. Here we present two of them. In the first one, we consider the five-node scenario shown in Figure 7 to find the available bandwidth between nodes 1 and 2. All nodes transmit at 2 Mbps and use the RTS/CTS mechanism. Node 1 moves along the dotted trajectory at a constant speed of 2 m/s. Figure 7 shows some positions requiring a path with one, two, three, and four hops. The link between nodes 3 and 4 carries a cross-traffic VBR flow that sends 2000-byte packets

at an average rate of 750 Kbps. The results of the probing packet dispersion analysis are shown in Figure 8, where they are compared with the true available bandwidth, obtained by looking for the maximum achievable throughput at different positions. Notice the high accuracy of the estimation and the detailed resolution of the ABW trace. This resolution is achieved by advancing 320-packet analysis windows every 8 packets. Under no losses, this is equivalent to obtaining, every second, the average ABW on the past 40 seconds.

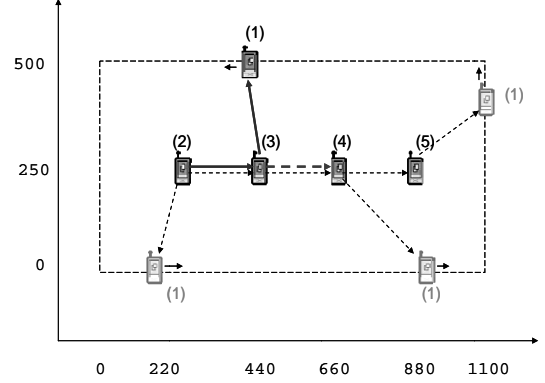


Fig. 7 First mobility scenario for testing our estimation method.

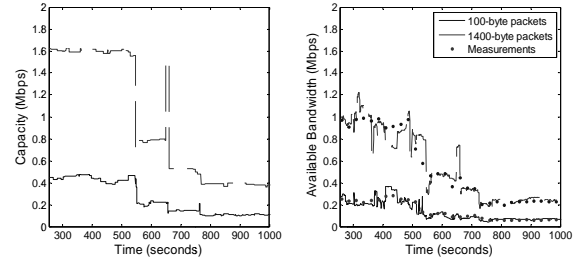


Fig. 8 Capacity and Available Bandwidth in the scenario of Figure 7.

Figure 9 shows another experiment, where a node moves at constant speed in a spiral trajectory within a grid of 25 nodes in a 1200m x 1200m area, as shown in the bottom left corner. A 200 kbps VBR cross traffic flows through the second row of nodes. The graph at the top of the figure shows  $BW(L)$  – continuous line– and  $ABW(L)$  –dashed line– for  $L=1400$  bytes, as a function of time, between the first node and the mobile node. The bottom right figure shows  $BW(L)$  –continuous line– and  $ABW(L)$  –dashed line– as a function of  $L$ , in a given instant of time. The stem lines show the true ABW for packets of 100, 750, 1400 and 2000 bytes and the BW for 100 and 1400 bytes. Notice the high accuracy of the estimation. Similar results have been obtained on scenarios with different mobility conditions, cross traffic conditions, physical data transmission rates, with and without using the RTS/CTS mechanism.

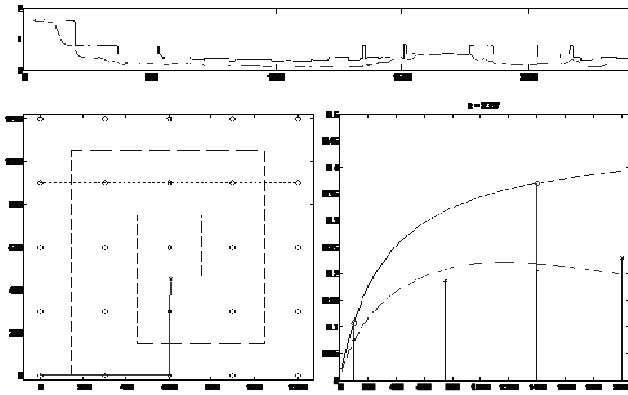


Fig. 9. Validation experiment under mobility and cross-traffic

## VI. CONCLUSIONS

We have presented a neuro-fuzzy estimator that use a dispersion trace of probing packet pairs to compute accurate functional forms of the end-to-end bandwidth and available bandwidth of a path between two nodes in an IEEE 802.11b ad hoc network. The estimator exploits the fact that the phenomena that determine the available bandwidth also determine the variability of the dispersion trace, as revealed by the distribution of the dispersion statistics conditioned on a high or low availability. This way the inference procedure, which models the dependence of the variability of the

dispersion traces on the true fraction of available bandwidth, takes into account most of the interactions involved in data transmission over such a network. Simulation experiments under very different conditions show the neuro-fuzzy estimator to be accurate and efficient in the use of communication and computational resources.

## REFERENCES

- [1] Prasad, R., Murray, M., Dovrolis, C. and Claffy, K. "Bandwidth Estimation: Metrics, Measurement Techniques, and Tools", IEEE Network, November/December 2003.
- [2] Alzate, M., Salamanca, M., Peña N. and Labrador, M. "End-to-End Mean Bandwidth Estimation as a Function of Packet Length in Mobile Ad Hoc Networks", Proceedings of the 12th IEEE Symposium on Computers and Communications, 2007.
- [3] Qiu, L., Zhang, Y., Wang, F., Han, M., Mahajan, R. "A General Model of Wireless Interference", Proceedings of ACM Mobicom, 2007
- [4] Sarr, C., Chaudet, C., Chelius, G. and Guérin, I. "A Node-Based Available Bandwidth Evaluation in IEEE 802.11 Ad Hoc Networks", Proceedings of First International Workshop on System and Networking for Smart Objects, 2005.
- [5] Sarr, C., Chaudet, C., Chelius, G. and Guérin, I. "Improving Accuracy in Available Bandwidth Estimation for 802.11-based Ad Hoc Networks", INRIA Research Report N° 5935, June 2006.
- [6] Johnsson, A., Melander, B., and Bjorkman, M. "Bandwidth Measurement in Wireless Network", Technical Report, Malardalen University, Sweden, March 2005.
- [7] Qualnet Simulator, version 4.0, available from <http://www.scalable-networks.com/>.