

# An autoregressive estimator for overhead reduction in Substitution Networks

Karen Miranda, Nathalie Mitton, Victor Ramos

► **To cite this version:**

Karen Miranda, Nathalie Mitton, Victor Ramos. An autoregressive estimator for overhead reduction in Substitution Networks. International Conference on Next Generation Mobile Applications, Services and Technologies (NGMAST), Sep 2015, Cambridge, United Kingdom. hal-01174852

**HAL Id: hal-01174852**

**<https://hal.inria.fr/hal-01174852>**

Submitted on 24 Sep 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# An autoregressive estimator for overhead reduction in Substitution Networks

Karen Miranda

Universidad Autónoma Metropolitana (UAM)  
Dept. of Applied Mathematics and Systems  
Cuajimalpa, Mexico City  
kmiranda@correo.cua.uam.mx

Nathalie Mitton

Inria Lille – Nord Europe  
40, avenue Halley  
Villeneuve d’Ascq, France  
nathalie.mitton@inria.fr

Victor Ramos *IEEE Senior Member*

Universidad Autónoma Metropolitana (UAM)  
WiNetSys Research Team  
Iztapalapa, Mexico City  
vicman@xanum.uam.mx

**Abstract**—A substitution network is a temporary network that self-deploys to dynamically replace a portion of a damaged infrastructure by means of a fleet of mobile routers. Some efficient solutions deploy robots based on active measurements. A robot/node in the network may use active link monitoring to assess the link quality towards its neighbors through the use of probe packets. Such probe packets are sent periodically at a given rate, and so, the accuracy of the measurements depends on the number and the frequency of exchanged packets. However, exchanging probe packets is energy and bandwidth consuming, thus active monitoring is considered as a costly mechanism. Even so, active link monitoring is a technique widely used on many network protocols. In this paper, we focus on an adaptive positioning algorithm (APOLO) to self-deploy a network. APOLO is based on active monitoring to gather essential information from nodes. Therefore, we show how autoregressive estimation may be used to reduce the overhead caused by the active measuring technique. Moreover, it is possible to use surrogate data rather than real data to feed APOLO without impacting its performance.

**Keywords**— Surrogate data; autoregressive processes; overhead; substitution network.

## I. INTRODUCTION

A substitution network is a temporary network that self-deploys to dynamically replace a portion of a damaged infrastructure by means of a fleet of mobile routers [1]. We focus in this paper on substitution networks made of autonomous wireless robots that automatically self-deploy to surrogate the damaged networks. Their movements are based on some active measurements. Active measurement is a widely used technique that provides insights about the network performance. Nevertheless, its main disadvantage is that it introduces additional packets, causing *communication overhead* when collecting such measurements. Such additional packets are energy and bandwidth consuming; however, the accuracy of the measurements depends on the number and the frequency of the exchanged packets. If the probe transmission rate increases, the accuracy of the insights will increase as well; nevertheless, the corresponding overhead increases proportionally consuming network resources. Obviously, there exists a compromise between the information accuracy and the transmission rate of probe packets [2]. Such trade-off is specially important when a decision making process relies on the information gathered by using the probe packets such as in the case of the *adaptive positioning algorithm*.

In a previous work, we have proposed an *adaptive positioning algorithm* called (APOLO) to self-spread mobile routers in substitution networks [3]. Such algorithm allows the routers to self-deploy by using controlled mobility based on the link quality. The link quality is measured in terms of delay, received signal strength (RSS), or signal to noise ratio by means of active measurement. To this end, our proposed algorithm sends out probe packets to all the nodes in the one-hop communication range; each node receiving the probe packets answers with reply packets; thus, the router computes the next movement and its new placement based on the replies received.

Reducing the packet transmission rate allows to scale down the corresponding overhead. However, regarding APOLO the trade off between the consumed resources and the data accuracy is reflected on the time it takes to routers to arrive at their final location since it takes more time to acquire the necessary data to make a decision.

We may find some similar problems in the literature. For instance, the Hello protocol, also known as the neighbor discovery protocol, allows neighbor nodes to establish and maintain the neighborhood relationships in wireless ad hoc networks [4]. For that purpose, each node sends Hello messages at regular intervals to claim/notify its existence. Thus, the main issue is to determine the optimal transmission rate for messages. Likewise, in wireless sensor networks (WSN) the communication task consumes most of the energy [5]. One method to reduce energy consumption is to reduce the number of messages exchanged between nodes. The goal of data reduction techniques is, precisely, to reduce the data exchanged between the sink and the sensor nodes. In particular, data prediction techniques reduce the amount of information sent by building a model of the data evolution. In both cases, as presented in [6], it is possible to use autoregressive modeling to solve the trade-off between accuracy and resources consumed.

In this paper, we propose to use surrogate data obtained by means of an autoregressive model to avoid data starvation periods and reduce the overhead in the network. The remaining of the paper is structured as follows. Section II describes the importance of the overhead-accuracy trade off for network characterization. Then, Section III reviews the concepts of time series used in this paper. In Section IV, we describe our proposal to use surrogate data in order to reduce the exchange of probe packets. Section V describes the simulation settings and present our results. Finally, we sketch out some concluding

remarks in Section VI.

## II. OVERHEAD AND ACCURACY COMPROMISE

Active measurement is a well known approach to characterize the conditions and performance of a given network. To that end, active measurement techniques introduce extra packets into the network, which are known as *probe packets* [7]. Such probe packets provide insights about the network behavior like packet loss, delay, and jitter [8].

Nevertheless, adding packets to the network traffic poses several challenges, particularly on wireless networks. For example, since the IEEE 802.11b [9] amendment may use several data rates namely 1 Mbps, 2 Mbps, 5.5 Mbps, and 11 Mbps, the amount of resources consumed by a packet relies on the rate at which the packet is transmitted; i.e., a given packet transmitted at 11 Mbps consumes less resources than an identical packet transmitted at 1 Mbps. Furthermore, it is necessary to consider the additional resources in terms of storage and analysis of the aggregated traffic as well as the cost induced by latency. Therefore, there exists a trade-off between the amount of probe packets transmitted and the accuracy of the obtained values [10]. This is specially important when a decision-making process depends on the data collected during the sampling intervals.

Apart from error as an important metric to evaluate different mechanisms, the ultimate goal is to make the same decision with less information. Particularly, our deployment algorithm needs to gather data about the link quality to decide the direction on which the router must move. Such a task depends on the transmission rate of probe packets to collect data for link quality. In other words, if the rate is too high for the decision making process, valuable resources such as energy or bandwidth would be wasted; conversely, if the rate is too low, it would cause data starvation periods where the router needs to make a decision but it does not have enough information to do so.

To illustrate the overhead and accuracy compromise in substitution networks, we consider a simple scenario composed of one source, one destination, and one router. First, we execute the adaptive positioning algorithm without modifications; then, we execute the same algorithm and we reduce the packets frequency by increasing the time between probe periods, that is, just sending one-fifth of the total of probe packets in normal conditions. Finally, we increase once again the time between probe periods, sending one-tenth of the probe packets. Figure 1 shows the corresponding behavior, where we may observe that by choosing the normal rate, the router reaches the middle point after 600 s of simulation. However, when we reduce the probe frequency, the time the router takes to reach the same point drastically increases. Thus, the overhead-accuracy trade-off has an evident impact on the performance of APOLO.

## III. SOME BACKGROUND ON TIME SERIES MODELS

Before describing our proposal to reduce overhead, we present a brief introduction to time series models for a good understanding of forecasting-based techniques. Time series forecasting methods are commonly used to predict future values of a process as a function of previous samples from a given series. In particular, the autoregressive (AR) model is

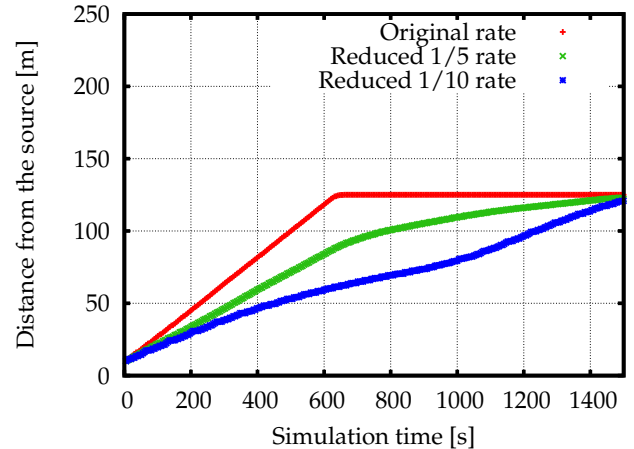


Fig. 1. Impact of the overhead-accuracy trade-off on the performance of the adaptive positioning algorithm.

widely used because of its simplicity and low complexity. This model predicts the value of  $X_{i+1}$ , which is denoted by  $\hat{X}_{i+1}$ , as a weighted sum of the previous  $M$  values of the process  $X_i$ . The AR( $M$ ) model is defined as follows:

$$\hat{X}_{i+1} = \sum_{m=1}^M \phi_m X_{i-m} + \epsilon_i, \quad (1)$$

where  $\phi_m$  represents the model coefficients,  $M$  is the model's order, and  $\epsilon_i$  is white noise. There are several approaches to estimate the values of  $\phi_m$  for  $m \in [1, \dots, M]$ , such as, Yule-Walker equations, ordinary least squares, maximum entropy estimates, geometric lattice method, and forward-backward method [11]. In practice, the selection of the order's model  $M$  for a given data is calculated by balancing the error generated by the model against the number of parameters. It is easy to reduce the error by increasing the order, however, this requires a greater number of unknown parameters, consequently, increasing the resource requirements for computation.

The autoregressive-moving average model (ARMA) and its generalization on the autoregressive integrated moving average model (ARIMA) are a combination of the AR and the moving average (MA) models. In both, the AR branch represents the dependency between the current value and the  $M$  previous values, while the MA branch represents the influence of current and past errors due to white noise on the current and future values [11].

## IV. DATA REPLACEMENT

In this section, we describe how to apply autoregressive estimation to reduce the overhead produced by the use of probe packets. First, we sketch our proposal to reduce the overhead by means of autoregressive estimation and next, we evaluate the usefulness of our proposal. As we report in Section II, it is important to reduce the number of artificial packets injected into the network preserving at the same time the accuracy of the measurements.

In [3], we have presented the adaptive positioning algorithm (APOLO) for self-deploying mobile routers in a

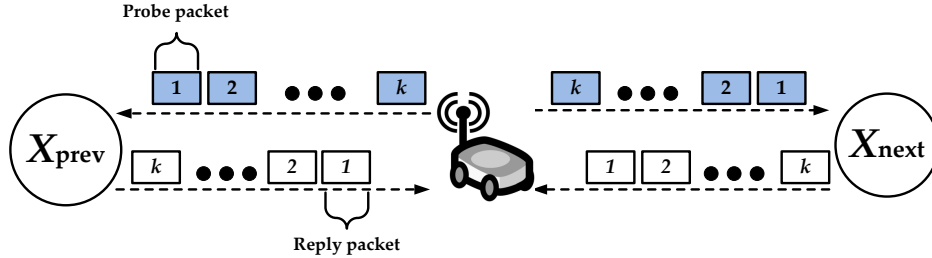


Fig. 2. Example of a probe period for two neighbors.

substitution network. During the substitution network lifetime, APOLO is executed in each mobile router to determine whether it has to move by using the feedback of the link quality coming from one-hop neighbors. APOLO consists of three major stages. *i)* APOLO measures the link quality by means of one link parameter, e.g., RSS or SNR. *ii)* APOLO computes the gathered data and makes the movement decision, i.e., if the router needs to move or not to improve the link quality; and *iii)*, APOLO determines the direction of the movement and the router moves accordingly.

Figure 2 depicts a probe period for two neighbors,  $X_{prev}$  (previous hop) and  $X_{next}$  (next hop), and one mobile router. The router broadcasts probe packets to all neighbors in one-hop communication range and receives reply packets in a unicast fashion from such neighbors. At every probe period, the router sends  $k$  probe packets to its neighbors, for example, if the number of neighbors is two and  $k = 10$ , the router sends 20 probe packets and receives 20 reply packets. Of course, this number is proportional to the number of nodes within the range and according to the probe period frequency. Thus, the number of exchanged packets between the router and its neighbors has a rapid growth when the number of nodes in the network increases. It is important to recall that the probe period and decision making period are independent from each other, even though the decision making process relies on data collected by the probe packets.

As we mentioned in Section II, we may reduce the number of probe packets by reducing their transmission rate. However, this causes an increment in the deployment time. Therefore, we propose to avoid the data starvation periods by substituting the missing information with estimates generated by an estimator.

#### A. Autoregressive Estimator

We now describe our AutoRegressive EStimator (AREs) for overhead reduction. At the beginning of the simulation, every router  $r$  executes the *adaptive positioning algorithm* and keeps track of the averages obtained for each link sampled. Those average samples  $X_i, i = 0, 1, 2, 3, \dots$  form the time series of interest feeding our algorithm. Thus, each router builds one model per neighbor in the one-hop communication range to predict the link quality of its neighbors. We find that if autoregressive mechanisms are well tuned, important improvements may be achieved on the estimation of real data [12]. Therefore, we choose the autoregressive model as

detailed in Section III by using the Yule-Walker method to calculate the model coefficients. The model's order  $M$  is the value that minimizes the mean squared error, MSE, given by

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (\hat{X}_i - X_i)^2.$$

At decision period  $j$ , router  $r$  estimates the value of the link quality  $\hat{X}$  fed with its own estimates until a new probe period brings new data. Then, the algorithm corrects the divergence between real and estimated data by feeding the estimator with real data at each probe period. In other words, we use the estimated values to replace the actual values when the latter are not available due to the overhead reduction. Thereby, the decision making period is not delayed by decreasing the probe transmission rate. Once the actual data arrive, our estimator uses them to update the AR model.

We describe the major steps of our overhead reduction proposal as follows. At each router  $r$ , the average values of the link quality  $X_i, i = 0, 1, 2, 3, \dots$  constitute a time series. From it, each link in the routers neighborhood is associated to a time series, separately, with individual model's order and individual coefficient values. The estimator collects a set 10 samples, which is equivalent to 10 probe periods. Thus, at probe period  $j > 10$ :

- Each mobile router builds one model for each neighbor in the one-hop communication range.
- The estimator is fed with its own samples until the next probe period.
- The estimator retrieves the computed values.
- Based on the estimated values, the router moves accordingly.

## V. EVALUATION

We consider a wireless network composed of mobile routers that may move on the two-dimensional Euclidean space. We use "node" as a generic term for any device into the simulation neighborhood, for instance, mobile or classic routers. For the sake of simplicity, we assume that the transmission range  $R$  of a node  $u$  is the area in which another node  $v$  can receive/send messages from/to  $u$ ; i.e.,  $d(u, v) < R(u)$ ,

where  $d(u, v)$  represents the Euclidean distance between  $u$  and  $v$ , and therefore, it exists a link  $X$  between  $u$  and  $v$ . We assume that two nodes are “neighbors” when they are within the communication range of each other. We use  $X$  to refer to the links with the neighbors of a mobile router. Likewise, we assume that some of the devices are fixed, that traffic needs to be transferred between two fixed devices, and that the wireless routers dynamically move in the scenario and act as relays, regardless of the routing protocol.

We define the quality of a communication link, or just “link quality”, as the probability that a message transmitted on the link is successfully received, that is, the reliability of the link [13]. We use the received signal strength (RSS) as a value to measure the link quality since its value retrieves insight of the performance of a wireless network [14]. Nevertheless, other link quality measures may be used [3].

We use the term “broadcast” to refer to the message propagation in a router’s neighborhood in order to obtain the link measurements. Also, we refer to the control packets of routing protocols as “Hello” messages or beacons and to the packets used in active measurements as probe packets. Finally, we define the term *controlled mobility* as the ability of some nodes, specially routers, to move by themselves to a specific destination or with a specific goal, i.e., the opposite of a random moving [15].

In this paper, we use the Dynamic Source Routing (DSR) protocol for our set of simulations [16]. DSR is a self-maintaining routing protocol designed for multi-hop wireless networks composed of mobile nodes. DSR uses on demand routing allowing each source to determine the route used to transmit its packets to the corresponding destinations. DSR consists mainly of two mechanisms, *Route Discovery* and *Route Maintenance*. The Route Discovery cycle is used to find on demand a route between the source and the destinations. Route maintenance is used to ensure that the paths remain optimum and loop-free as network conditions change. DSR avoids additional traffic, for example, Hello packets by using source routing; i.e., by including the entire route into the packet header and by storing the routes in caches.

To evaluate our proposed estimator, we use the NS-2 network simulator [17]. We compare our proposal to the original *adaptive positioning algorithm*. First, APOLO is executed without any modifications. Then, we reduce the probe packets transmission rate, and replace the missing data by using estimates generated by means of AREs. We use the relative error to assess the accuracy of our estimator. The relative error is equal to:

$$RE = \frac{|\hat{X}_i - X_i|}{X_i}. \quad (2)$$

We use the received signal strength (RSS) as input for both algorithms, nevertheless, it is possible to consider other link metrics, such as signal-to-noise ratio (SNR) or round-trip time (RTT). We execute 50 simulations of 1500 s for each algorithm in order to obtain average results. In all the simulations, the IEEE 802.11 DCF standard and the two-ray ground reflection model are used.



Fig. 3. Scenario 1: One source, one destination, and one mobile router.

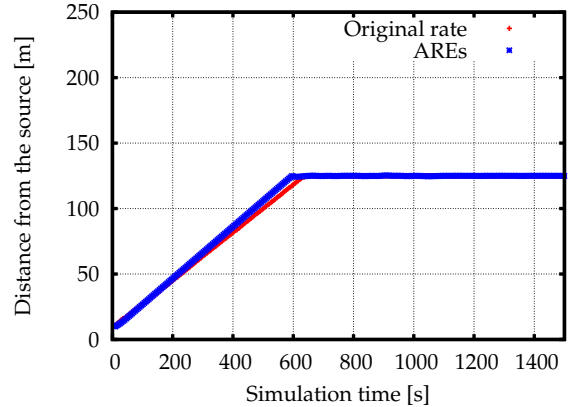


Fig. 4. Comparing position algorithm performance using actual data and surrogate data as a function of time.

TABLE I. PACKET EXCHANGE COMPARISON

	Probe Periods	Probe Pkts.	Reply Pkts.
Without change	428.7	8345.3	8344.14
Reduction to 20%	89.2	1730.86	1729.8
Reduction to 10%	45	863	862.3

We consider a scenario composed of two nodes,  $n_1$  and  $n_2$ , out of the range of each other, the distance between them is 250 m. In order to communicate  $n_1$  and  $n_2$ , a router  $r_1$  must be deployed. We consider that such router starts at 10 m close from  $n_1$  and it must move to allow the communication between  $n_1$  and  $n_2$  such as is shown in Figure 3. Specifically, our proposal constructs an AR model for each neighbor denoted by  $\hat{X}_{prev}$  (previous hop) and  $X_{next}$  (next hop), respectively.

In Figure 4, we compare the performance of our proposal, replacing the data with estimates, with the original positioning algorithm that uses the actual values. We show that by using autoregressive estimation, the router achieves the optimal position after  $\sim 600$  s as the original algorithm, i.e., the router makes the same decision for its movement decision based on the estimated values with the same delay. Furthermore, we present the relative error for every sample of the model built in Figure 5. In both plots, most of the values for the relative error are smaller than 0.2; i.e., less than 20% of the error after the sample number 200. RE achieves the smallest values during the simulation, this corresponds to the instant of time when the router achieves the optimal position.

Table I summarizes the gain in terms of the number of probe periods and the exchanged probe and reply packets. From such a table, we observe a considerable reduction of up to 80% on overhead during the simulation.

Moreover, we modify the scenario depicted in Figure 3 by adding a mobile router ( $r_2$ ) and increasing the distance between  $n_1$  and  $n_2$ . Then, we compare the performance of the positioning algorithm using actual data and surrogate data.



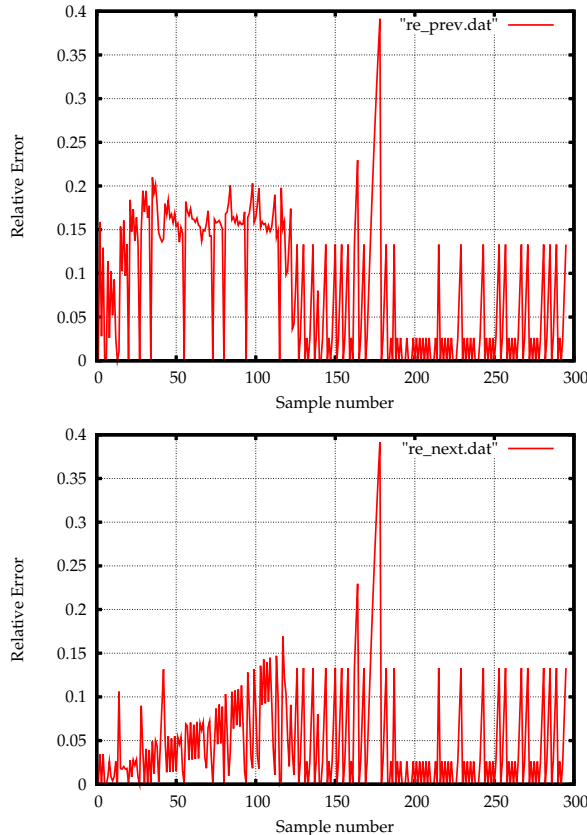


Fig. 5. The relative error of the estimated values for  $X_{\text{prev}}$  (previous hop) and  $X_{\text{next}}$  (next hop) by using the Yule-Walker method as a function of sample number.

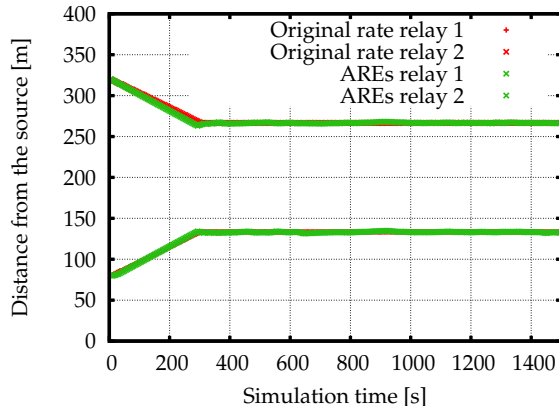


Fig. 6. Performance of the positioning algorithm using actual data and surrogate data as a function of time.

The results are presented in Figure 6 where both routers achieve their final position 300 s by executing both algorithm confirming the results above presented.

## VI. CONCLUSION

In this paper, we employed autoregressive estimation in wireless substitution networks for overhead reduction. We have reduced the number of probe packets exchanged by up to 80%, while at the same time maintaining the accuracy in the movement of routers. In both, the original positioning

algorithm and our proposal, the router achieves the same position by expending the same time. Clearly, replacing actual samples with accurate estimates is an effective approach to reduce the overhead caused by active measurement.

Finally, we have chosen a gathering information cycle of 10 probe periods to calculate the estimator coefficients. Nevertheless, these values may be not optimal, thus, our future work in this direction will focus on finding the optimal values for the percentage reduction and gathering period. We will evaluate different values and will observe the impact on our mechanism.

## REFERENCES

- [1] T. Razafindralambo, T. Begin, M. Dias de Amorim, I. Guérin Lassous, N. Mitton, and D. Simplot-Ryl, "Promoting Quality of Service in Substitution Networks with Controlled Mobility," in *Proceedings of 10th International Conference on Ad-hoc, Mobile, and Wireless Networks (ADHOC-NOW)*, Paderborn, Germany, Jul. 2011, pp. 248–261.
- [2] N. Baccour, A. Koubâa, L. Mottola, M. A. Zúñiga, H. Youssef, C. A. Boano, and M. Alves, "Radio Link Quality Estimation in Wireless Sensor Networks: A Survey," *ACM Transactions on Sensor Networks (TOSN)*, vol. 8, no. 4, pp. 34:1–34:33, Sep. 2012.
- [3] K. Miranda, E. Natalizio, and T. Razafindralambo, "Adaptive Deployment Scheme for Mobile Relays in Substitution Networks," *International Journal of Distributed Sensor Networks (IJDSN)*, vol. 2012, 2012.
- [4] V. C. Giruka and M. Singhal, "Hello Protocols for Ad-Hoc Networks: Overhead and Accuracy Tradeoffs," in *Proceedings of the IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WOWMOM)*, Taormina, Italy, Jun. 2005, pp. 354–361.
- [5] U. Raza, A. Camera, A. L. Murphy, T. Palpanas, and G. P. Picco, "What does model-driven data acquisition really achieve in wireless sensor networks?" in *Proceedings of the IEEE International Conference on Pervasive Computing and Communications (PerCom)*, Lugano, Switzerland, Mar. 2012, pp. 85–94.
- [6] X. Li, N. Mitton, and D. Simplot-Ryl, "Mobility Prediction Based Neighborhood Discovery in Mobile Ad Hoc Networks," in *Proceedings of the 10th International IFIP TC 6 Networking Conference (NET-WORKING)*, vol. 6640, Valencia, Spain, May 9-13 2011, pp. 241–253.
- [7] M. J. Luckie, A. J. McGregor, and H.-W. Braun, "Towards improving packet probing techniques," in *Proceedings of the 1st ACM SIGCOMM Workshop on Internet Measurement (IWM)*, San Francisco, California, USA, Nov. 1-2 2001, pp. 145–150.
- [8] M. Hasib and J. A. Schormans, "Limitations of Passive & Active Measurement Methods in Packet Networks," in *London Communications Symposium (LCS)*, London, UK, Sep. 2003.
- [9] IEEE, "IEEE 802.11, local and metropolitan area networks - specific requirements part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications, IEEE standard," 1999.
- [10] T. B. Reddy, B. S. Manoj, and R. R. Rao, "On the Accuracy of Sampling Schemes for Wireless Network Characterization," in *Proceedings of the IEEE Wireless Communications & Networking Conference (WCNC)*, Las Vegas, Nevada, USA, Mar. 2008, pp. 3314–3319.
- [11] J. G. Proakis and D. K. Manolakis, *Digital Signal Processing*. Prentice-Hall, 1996.
- [12] K. Miranda, V. Ramos, and T. Razafindralambo, "Using efficiently autoregressive estimation in Wireless Sensor Networks," in *Proceedings of the International Conference on Computer, Information, and Telecommunication Systems (CITS)*, Piraeus-Athens, Greece, May 7-8 2013, pp. 1–5.
- [13] M. R. Souryal, A. Wapf, and N. Moayeri, "Rapidly-Deployable Mesh Network Testbed," in *Proceedings of the Global Communications Conference (Globecom)*, Honolulu, Hawaii, USA, Nov. 2009, pp. 1–6.
- [14] E. Feo Flushing, J. Nagi, and G. A. Di Caro, "A mobility-assisted protocol for supervised learning of link quality estimates in wireless networks," in *International Conference on Computing, Networking and Communications (ICNC)*, Maui, Hawaii, USA, Jan. 2012, pp. 137–143.

- [15] E. Natalizio and V. Loscri, "Controlled mobility in mobile sensor networks: advantages, issues and challenges," *Telecommunication Systems*, vol. 52, no. 4, pp. 2411–2418, Apr. 2013.
- [16] D. B. Johnson, D. A. Maltz, and J. Broch, "Ad Hoc Networking." Addison-Wesley Longman Publishing Co., Inc., 2001, ch. DSR: The Dynamic Source Routing Protocol for Multihop Wireless Ad Hoc Networks, pp. 139–172.
- [17] NS, "Network Simulator v.2.29 (NS-2)," accessed on January 2013. [Online]. Available: <http://isi.edu/nsnam/ns/>