Predictive Estimation of Wireless Link Performance from Medium Physical Parameters

Guillaume Kremer, Philippe Owezarski, Pascal Berthou, German Capdehourat

Predictive Estimation of Wireless Link Performance from Medium Physical Parameters

Guillaume Kremer[1,2b], Phillippe Owezarski[1,2c], Pascal Berthou[1,2b] and German Capdehourat[3]

[1]CNRS, LAAS, 7 avenue du colonel Roche, F-31400 Toulouse, France
[2]Univ de Toulouse, [b]UPS, [a]INSA, [c]LAAS, F-31400 Toulouse, France
[3]Instituto de Ingeniería Eléctrica, Facultad de Ingeniería, Universidad de la República, Uruguay
Email: {kremer, owe, berthou}@laas.fr, gcapde@fing.edu.uy

Abstract. In wireless networks, the physical medium is the cause of most of the errors and performance drops. Thus, an efficient predictive estimation of wireless networks performance w.r.t. medium status by the communication peers would be a leap ahead in the improvement of wireless communication. For that purpose, we designed a measurement bench that allows us to accurately control the noise level on an unidirectional WIFI communication link in the protected environment of an anechoic room. This way, we generated different medium conditions and collected several measurements for various PHY layer parameters on that link. Using the collected data we analyzed the ability to predictively estimate the throughput performance of a noisy wireless link from measured physical medium parameters, using SVR (Support Vector Regression). Finally, we ranked the pertinence of the most common physical parameters for estimating or predicting the throughput that can be expected by users on top of the IP layer over a WIFI link.

1 Introduction

Wireless networks are of essential importance nowadays. Users are more and more mobile, and access the Internet thanks to mobile devices as laptops, smart phones or tablets. Even when staying at home, users want to get rid of wires. However, the wireless medium does not provide the same capabilities as wired networks on copper or fiber. In wireless networks, the physical medium is limited in terms of capacity, and the cause of most of the errors and performance drops. From a user or administrator point of view, the quality of wireless communication can appear as very versatile and unpredictable. This makes wireless networks very complex to manage, and users often experience communication quality drops that are completely unexpected.

Monitoring wireless networks is then very difficult. Monitoring such networks at the IP layer is very inefficient (whereas it is the way it is done in wired networks with extremely good results). Some previous work tried to include the MAC level in the monitoring of wireless networks [1], but none integrates the full

monitoring of the network from physical to network layers. We nevertheless argue that this is the direction to follow, and propose our preliminary study to estimate the relations between the physical signal parameters and the performance at the network level. Physicians are doing very strong studies on the signal level, but do not study the impact on upper layers [2]. In this paper, it is proposed to bridge the gap between the signal and the digital world in wireless communication networks.

This paper then presents a double contribution:

- First, we designed and built a platform for benchmarking wireless communications. It is built in an anechoic chamber to fully control the experimental environment, and avoid external signals to disturb the behavior of the communicating devices and the quality of the measurements. We used on this platform the common digital communications devices that are widely used (laptops, tablets, smart phones), as well as dedicated signal measurement tools specifically designed for physicians. Anyway, because of space limit, this paper concentrates on the study of a WIFI link.
- Second, the paper presents the analysis of the relations between the PHY parameters of the WIFI connection, and the performance parameters on top of the IP layer. It aims at demonstrating that, at the opposite of wired networks, the monitoring of wireless network has to be done at the physical level. It is shown that using a very limited number of signal parameters (one or two), it is possible to very accurately estimate communication performance and quality parameters as network level throughput, delay or loss ratio. It is even possible to predict performance drops at the scale of one second. For this purpose, we rely on the SVR algorithms. SVR are supervised learning algorithms that are known to have good prediction capabilities and that succeed in many domains as long as these domains can provide accurate time series [3, 4]. Some other techniques have been tested as decision trees, KNN (k-Nearest Neighbor), etc. and exhibited the same results. Again, because of space limit, the paper only presents the results with the most common physical signal parameters as SNR or RSS for estimating the throughput obtained on top of the IP layer.

Finally, section 4 concludes this paper.

2 Experimental platform and dataset

2.1 Experimental conditions and measurement equipments

The implementation of a dedicated wireless testbed is a major requirement for our work. First of all, experimentations must be reproducible, allowing comparison between different sets of measurements and algorithms. This point is not trivial when using wireless networks as the environment factors have a hight impact on the network performances. Secondly, part of the originality of this work comes from the combination of measurements made at multiple network

layers, using electronics instruments and software tools. This was also a strong requirement to be able to monitor the physical layer (the wireless transmission), and compare it to the higher layers, from the mac layer information given by the network cards to the end-to-end layers as transport throughput for instance. The hardware introspection requirement have an impact on the components choice as explained below. Thirdly, the synchronization of all of these datasets was a sticky point, but absolutely required to ensure a good behavior of the learning algorithms.

2.2 Reproducibility requirement

Our wireless testbed was designed inside an anechoic room. An anechoic room is a protected RF room which simulates free space conditions. Our model of chamber is 4,10 meters long for 2,50 meters wide. Inside, walls are covered of microwave absorbers materials that break and scatter any wireless signal that would come from an inside source. The chamber is then free of any multi-path propagation. There are different types of absorbers, each of them is defined for a specific frequency range that allows us to use the anechoic chamber for different purposes and frequencies. The absorbers protect also the inner environment of the room from outside perturbations. This protected context minimizes the uncontrolled parameters of our communication.

2.3 Introspection requirement and components choice

Inside the anechoic chamber we placed two WIFI nodes. The nodes are controlled through a wired network to avoid interference with the wireless communication. The nodes are Avila-GW2348-4 gateway platforms and run a Linux OpenWrt OS. The boxes have an Intel Xscale processor, 64 MB of SDRAM and 16MBytes of Flash memory. The WIFI network controllers are based on the AR5414 chipset from Atheros which uses the ath5k driver and are attached to an omni-directional antenna. The choice of the wifi chipset and its driver was crucial because they define the amount of metrics and the accuracy that it will be possible to obtain. The ath5k driver is open-source and well documented thanks to an active online community support. It has also a good integration within the OpenWrt OS. The OpenWrt OS is flexible enough to allow the implementation of new functionalities so that it accelerates the upgrade of the bench.

In addition and because we were unable to capture the noise strength of the received signal with the Atheros hardware, we used an oscilloscope connected to the receiver antenna. It records the amplitude of the received signal. The oscilloscope chosen was a fast Lecroy WaveRunner which allows us to capture a maximum number of frame signal with little loss and record them on internal memory. The precision of this instrument gives us the ground truth required by the training methods used. It also embeds a large library of filters, and operators which can be applied on the input signals. The oscilloscope is also synchronized by NTP.

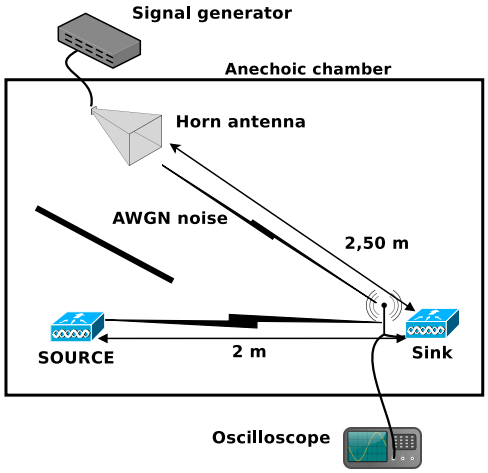

Fig. 1. Disposition of the different equipments in the wireless testbed. The cable from the receiver antenna is connected to a power splitter which enables the signal from the antenna to be dispensed similarly to the WIFI sink and to the Lecroy oscilloscope with neglictable signal alteration.

Synchronization requirement. As we used several equipments to get measurements, it is needed to have their clock very accurately synchronized. This was done with NTP by using a dedicated wired connection to a remote NTP server (accuracy with a shared network bus is not sufficient).

Software glues. The command and control of the nodes is done using distributed software deployed on the boxes and written in RUBY. The application allows the configuration of wireless networks settings and permits launching commands to control the measurement process. The system is based on Remote Method Invocation which gives access to a set of operational methods on each boxes and then permits the network to be controlled and configured transparently from a single node in the network.

The configuration of the network interfaces is done in promiscuous mode to capture any packets sensed by their antenna. The packets are captured at the MAC layer using the PCAP library and tools when they arrive at the kernel interface. The packets contain data from link to application layers, such as the 802.11 channel number, the type of frame at the MAC layer, or packet size at the network layer. Additionally, a packet also contains a RADIOTAP header which gives radio level information such as the received signal strength (RSS) reported by the ath5k driver.

We modified the ath5k drivers of the OpenWrt OS to permit, when possible, the propagation of packets with frame check sequence (FCS) errors to the upper layers, while on the original kernel they were discarded. The propagation is only possible if the error corrupted the data but not the header fields. Following this

modification the RADIOTAP header now contains a flag specifying whether a FCS error was detected when decoding the packet.

The Lecroy oscilloscope was set to capture and flush the data as soon as a frame is detected on the input cable. This happens when the amplitude of the sensed signal is above a specific threshold, set to be in between the current noise floor and the minimal amplitude value of a frame. This threshold has to be set in a way to prevent exceptional high noise values that could be incorrectly detected as a frame.

2.4 Experimental protocol

Noise generation. One of the objectives of our environment is to minimize the presence of these uncontrolled parameters on the communication. Another objective is to generate and control selected parameters that will impact our communications.

The noise and the interferences significantly impact the communication. In the set-up described in figure 1, we inject noise in the environment using a signal generator to perturb the communication. The signal generator is a device which emits RF signals. It can be configured to generate very realistic noise. Among the parameters of the generated noise, two important elements have a crucial impact: on a first hand the modulation used characterizes the main characteristics of the noise signal in the time and frequency domains (i.e. it characterizes the spectral occupancy of the generated signal, its fading or narrowness). On a second hand, the amplitude of the signal also affects the measured level of noise on the receiver side.

We found that the AWGN (Adaptive White Gaussian Noise) noise modulation was a good choice because it is a common model of noise. Moreover it can be used to impact the entire bandwidth of a 802.11g channel contrary to most other modulation schemes which produce narrow band noise. The noise level was determined empirically by testing the effects on the communication.

Finally, a major element that affects the noise generated in the anechoic chamber is the antenna. It characterizes the waveform, the direction and the amplitude of the noise wave. In order to perturb only one side of the communication we used a very directional antenna pointed to the receiving station.

We use IPERF to generate traffic between the two peers. The traffic is a TCP flow with a constant throughput of 24 Mb/s. The size of the packets is set to 1470 bytes.

Training and datasets. We generated different samples with different noise levels and different transmission powers. All the samples have the same duration of 5 minutes and will be used to constitute our training datasets. Table 1 sums up the characteristics of the different samples. The same experimental settings (transmission power and noise) are used for training and testing. Therefore a training dataset which contains all these samples will be considered as having full knowledge about the possible use cases met in the test dataset. Hence, to

test the generalization capacity of our algorithm, we built three different training datasets as described in table 1. These datasets differ by the quantities of samples they are made of, and consequently by the level of knowledge they represent.

Table 1. Constitutions and characteristics of our training sets. Each vector represents 1 second of measurements

Training set	**Dataset definition**
notation	{Tx Power (dBm); Noise Power (dBm)}; {sample 2};...
*Dataset*1 5323 vectors	{10;-20};{10;-17};{10;-15};{10;-13};{10;-10};{10;-7};{10;-5}; {20;-20};{20;-17};{20;-15};{20;-13};{20;-10};{20;-7};{20;-5}
*Dataset*2 2661 vectors	{10;-20};{10;-17};{10;-15};{10;-13};{20;-20};{20;-17};{20;-15};{20;-13}
*Dataset*3 1330 vectors	{10;-20};{10;-17};{10;-15};{10;-7};{10;-5};{20;-20};

2.5 SVM features definitions

Atheros Received Throughput. This is the performance metric of the communication that we are considering in this paper. It is computed from the PCAP captured at the receiver side of the transmission. It is defined by $BW_i = \sum_{k=1}^{k} L(P_i^k)$ with $k \in \mathbb{N}$. BW_i is the computed throughput at second i, $L(P_i^k)$ is the length of the payload at the network layer for the kth packet without FCS error captured during second i.

Atheros RSS. The Atheros RSS is extracted from the RSS field in the RADIOTAP headers of the packets included in the PCAP files. Given that $RSS(P_i^0)$ is the RSS of the kth packets without FCS error captured during second i, and R_i is the set of packets captured during second i, it is defined as $ATH\_RSS_i = \overline{R_i}$ with $R_i = \left\{RSS(P_i^0), RSS(P_i^1)), ..., RSS(P_i^k))\right\}$.

Lecroy RSS, SNR and noise. In addition to the Atheros values, we extract different metrics from the Lecroy datasets. These values are computed from the Root Mean Square (RMS) values of the raw data. These RMS values can be split into three parts, which are the data that are before, during and after the frame. The part of the data before and after the frame are the noise values and therefore can be used to extract the noise floor during the reception of that frame. We consider A and C, the sets of these points. Therefore we compute the average noise floor of the data during the reception of frame P with $N_P = \overline{A \cup C}$.

With M_i the set of noise levels extracted from the frames captured by the Lecroy oscilloscope during second i, we compute the feature for the noise floor at second i $LECR\_NOISE_i$ as $LECR\_NOISE_i = \overline{M_i}$ with $M_i = \left\{N_{P_i^0}, N_{P_i^1}, ..., N_{P_i^k}\right\}$.

The RSS of the received frame is computed on the first 8 symbols to comply with 802.11 standard. These points constitute the set D. Thus, similarly to previous equations, the RSS for a frame P is given by $R_P = \overline{D}$ and $LECR\_RSS_i = \left\{R_{P_i^0}, R_{P_i^1}, ..., R_{P_i^k}\right\}$, where $LECR\_RSS_i$ is the feature of the Lecroy RSS at second i.

Finally we compute the SNR S_P for frame P as the difference between the noise floor and the RSS of the frame P and therefore, similarly to previous formulas: $S_P = R_P - N_P$ and $LECR\_SNR_i = \overline{W_i}$ with $W_i = \left\{S_{P_i^0}, S_{P_i^1}, ..., S_{P_i^k}\right\}$.

3 Estimation of the relations between physical and performance parameters in WIFI communications

3.1 SVR based methodology

The 2nd contribution of this paper is the analysis of the relations linking the PHY layer parameters and the upper layers performance.

To that aim, we used the SVR algorithm which is commonly used in many patter recognition applications. It has been successfully applied before in various fields as a predictive method (see [5] and references therein). As noticed in the introduction, the choice of SVR can be seen as secondary and it could be replaced by any other machine learning algorithm like decision trees or KNN that perform similarly.

In a nutshell, SVR is a supervised learning algorithm able to find relationships in non-linear data. For that, SVR uses a trick which consists in mapping the input data into a higher dimensional feature space where linear relationship exist. It is made possible by using a non-linear function called a kernel. In our case, the selected kernel is a Gaussian Radial Basis Function (RBF). An important detail about the operational usage of SVR is that, in our configuration, three SVR parameters are required (C, γ and ϵ). They will impact the performance of the estimations as well as the generalization capability. Moreover, these settings will differ for each dataset, and therefore need to be carefully selected in each case. For our estimations, we used a grid search to select these SVM parameters. It is a common method which consists in an exhaustive test run of SVR training using generated settings combinations. We then select the best combination of C, γ and ϵ among the results.

To evaluate the estimations, two methods are used. First, we use the MSE. Given that $\hat{Y}_i$ are estimations and Y_i are the real values, the MSE is defined as $MSE = \frac{1}{N}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2$. Secondly we use the percentage of correct estimations noted $P(e < 1Mb)$. In this notation e represents the absolute error between the real value and the estimation. An estimation is judged correct if it differs from the real throughput value by less than 1Mbps.

3.2 Estimation results

Tables 2(a), 2(b), and 2(c) contain the results of the throughput estimation based on 6 different PHY or combinations of PHY parameters for respectively $Dataset1$, $Dataset2$, and $Dataset3$. The first column quotes the PHY parameters that have been used for the SVR estimation of the IP throughput. Columns 2 and 3 show the figures obtained respectively for the MSE and the probability $P(e < 1Mb)$. The two last columns give the ranking for the PHY parameters according to their ability to allow good estimations of the throughput. 1 corresponds to the best result among the 6 PHY parameters considered.

For $Dataset1$, i.e. the full one, the best result is obtained $LECR\_RSS + LECR\_NOISE$. Estimations are plotted on figure 2. This figure exhibits impressive matching between the real and estimated values of the throughput, with just very few outliers appearing (75% matchings). We got as impressive results for $Dataset2$, and $Dataset3$, but this time, the best results have been obtained with the $LECR\_SNR$ parameter (60% matchings). The difference of the results when using a full trace for the training compared to a sampled one exhibits the non empty intersection between PHY parameters as SNR, RSS and NOISE. These 3 parameters are closely related.

It nevertheless clearly appear with these figures that SNR, RSS and NOISE can help to perfectly estimate and predict (on a one second scale) the performance of the network at layers 3 and 4. Nevertheless, a deeper analysis on larger datasets that still need to be produced would allow a more accurate characterization of the link between PHY parameters and network performance. Actually, it appears that while the combined features metrics performance decrease, the overall performance of the RSS metrics 1 and 2 increases or stays more or less the same. This seems to suggest that the full training set was not adapted to these metrics. The difference between the full and the reduced sets is that the samples obtained with high noise are not present in the reduced datasets. This could be caused by incoherent values existing in $Dataset1$ because of the bad and noisy conditions. One possibility is that these values could deteriorate the model issued from the training process. This aspect needs to be considered for improving our platform and experiment protocol.

4 Conclusions and future work

The main contribution presented in this paper deals with the design of a generic platform for monitoring and analyzing wireless networks. This wireless testbed is set in the RF protected environment of an anechoic room, allowing us to control the perturbation on the physical medium by generating noise. It also has the originality to integrate pure physical signal measurement tools as Lecroy oscilloscopes for very accurate measurements serving as ground truth. Based on the collected data, the second contribution of the paper deals with exhibiting the importance of PHY parameters on network communication performance. The correlation between the physical environment and the communication performance is so strong that it is possible by only monitoring the SNR and the RSS

Table 2. Scores and pertinence of the estimations using physical layer metrics.

(a) Results with $Dataset1$ as training set.

n°	Physical layer parameter(s)	MSE (Mbps^2)	P($e < 1Mbps$) (% of estimations)	Pertinence ranking	
				MSE	$P(e < 1Mbps)$
1	ATH_RSS	11.24	35	6	6
2	$LECR\_RSS$	4.42	51	5	5
3	$LECR\_NOISE$	2.28	69	4	3
4	$LECR\_SNR$	1.69	64	3	4
5	$ATH\_RSS + LECR\_NOISE$	1.02	70	2	2
6	$LECR\_RSS + LECR\_NOISE$	0.88	75	1	1

(b) Results with $Dataset2$ as training set.

n°	Physical layer parameter(s)	MSE (Mbps^2)	P($e < 1Mbps$) (% of estimations)	Pertinence ranking	
				MSE	$P(e < 1Mbps)$
1	ATH_RSS	11	33	6	6
2	$LECR\_RSS$	3.9	59	4	2
3	$LECR\_NOISE$	5.4	55	5	4
4	$LECR\_SNR$	1.6	66	1	1
5	$ATH\_RSS + LECR\_NOISE$	2.3	49	3	5
6	$LECR\_RSS + LECR\_NOISE$	2.0	57	2	3

(c) Results with $Dataset3$ as training set.

n°	Physical layer parameter(s)	MSE (Mbps^2)	P($e < 1Mbps$) (% of estimations)	Pertinence ranking	
				MSE	$P(e < 1Mbps)$
1	ATH_RSS	10.17	34	6	5
2	$LECR\_RSS$	4.5	32	4	6
3	$LECR\_NOISE$	5.8	44	5	3
4	$LECR\_SNR$	1.6	62	1	1
5	$ATH\_RSS + LECR\_NOISE$	3.3	41	3	4
6	$LECR\_RSS + LECR\_NOISE$	2.53	49	2	2

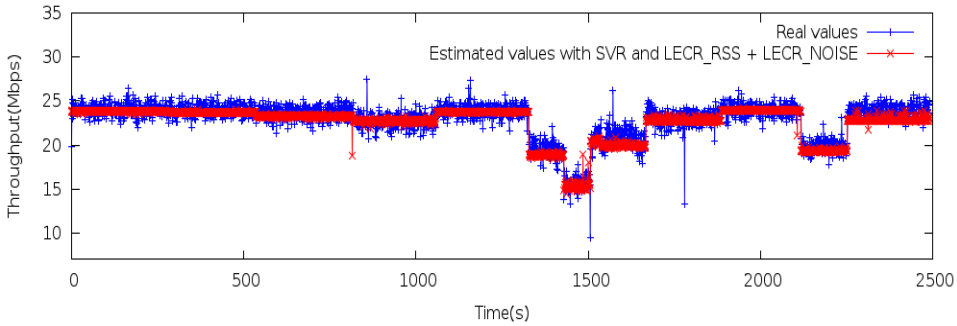

Fig. 2. Throughput estimation results obtained with the $LECR\_RSS + LECR\_NOISE$ metric compared to the real throughput.

of the signal to predict the performance level at the TCP/IP level. This result has been demonstrated using different kinds of models, in particular the SVR model.

Future work includes a large exploitation of our platform. We planned to generate large datasets for different kinds of wireless networks, including WIFI, UMTS, LTE, etc. and to open it to our research community, which is lacking such kinds of public datasets (to the best of our knowledge). We will also exploit this datasets by deeply analyzing them, understand how wireless networks behave, and then trying to improve the way we use and manage them.

5 Acknowledgments

This work is partially funded by the French National Research Agency (ANR) under two projects: the MAITRE project of the STIC AmSud program, and the RESCUE project of the VERSO program.

References

1. Claveirole, T., de Amorim, M.D.: Wipal and wscout, two hands-on tools for wireless packet traces manipulation and visualization. In: ACM Mobicom Workshop on Wireless Network Testbeds, Experimental Evaluation, and Characterization. (2008)
2. Lecointre, A., Dragomirescu, D., Plana, R.: New methodology to design advanced mb-iruwb communication system. IEE Electronics Letters **11** (2008)
3. Sapankevych, N., Sankar, R.: Time series prediction using support vector machines: A survey. Computational Intelligence Magazine, IEEE **4**(2) (2009) 24–38
4. Francois, J., Abdenur, H., State, R., Festor, O.: Machine learning techniques for passive network inventory. IEEE Transactions on Network and Service Management **7** (2010)
5. Basak, D., Pal, S., Patranabis, D.C.: Support Vector Regression. Neural Information Processing – Letters and Reviews **11** (October 2007)