Learning Wi-Fi Performance

Julien Herzen
EPFL
Switzerland
julien.herzen@epfl.ch

Henrik Lundgren
Technicolor
France
henrik.lundgren@technicolor.com

Nidhi Hegde
Alcatel-Lucent Bell Labs
France
nidhi.hegde@alcatel-lucent.com

***Abstract*—Accurate prediction of wireless network performance is important when performing link adaptation or resource allocation. However, the complexity of interference interactions at MAC and PHY layers, as well as the vast variety of possible wireless configurations make it notoriously hard to design explicit performance models.**

In this paper, we advocate an approach of "learning by observation" that can remove the need for designing explicit and complex performance models. We use machine learning techniques to learn implicit performance models, from a limited number of real-world measurements. These models do not require to know the internal mechanics of interfering Wi-Fi links. Yet, our results show that they improve accuracy by at least 49% compared to measurement-seeded models based on SINR. To demonstrate that learned models can be useful in practice, we build a new algorithm that uses such a model as an oracle to jointly allocate spectrum and transmit power. Our algorithm is utility-optimal, distributed, and it produces efficient allocations that significantly improve performance and fairness.

I. Introduction

We have witnessed a rapid adoption of Wi-Fi for home, enterprise and hotspot wireless networks. The result is often dense deployments of interfering Wi-Fi links that contend for a limited amount of spectrum. At the same time, these networks are under an ever increasing pressure to deliver a higher performance. Recent and ongoing IEEE amendments, such as 802.11n and 802.11ac, address this demand by including techniques such as wider channel bandwidths and faster modulation schemes. However, these enhancements put even more stress on the scarce spectrum resource and are sensitive to the operating conditions to deliver the effective performance improvements. Wider channels increase spectrum usage and can create harmful interference. Higher modulation schemes require a higher SNR and less interference to correctly decode transmissions. It is therefore increasingly important to carefully allocate resources such as spectrum and transmit power.

Efficient resource allocation requires realistic models. However, 802.11 networks – and especially those using newer amendments with variable bandwidth – are notoriously hard to model. They exhibit several performance intricacies due to complex interactions between the MAC and PHY layers, which manifest themselves in frequency, spatial and time domains. Existing performance models for 802.11 networks, such as the Bianchi model [1], usually adopt *explicit* and *bottom-up* approaches; they model the actual mechanics of the protocol (for example, the CSMA/CA procedure of the MAC layer in [1]) in order to compute throughput figures. These explicit models often give precious insights on the internal or asymptotic properties of 802.11. However, to remain tractable, these models have to rely on a set of simplifying assumptions (e.g., homogeneous PHY parameters in [1]), which prevent their use to predict the impact of different PHY layer configurations, such as variable channel widths. In contrast, textbook models based on the SINR (signal to interference-plus-noise ratio) can be used to capture some of the phenomena occurring at the PHY layer. However, in turn, these models do not take the MAC layer into account and, as we will observe, they do not capture the actual performance of interfering links when CSMA/CA is employed.

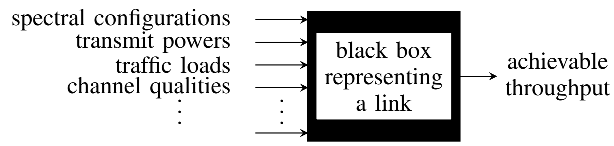

Figure 1: Black box representation of a link. It takes various configuration and topological features related to a given link and its neighbors as inputs, and it outputs a throughput.

In this paper we argue that, as far as quantitative performance predictions are concerned, it can be more efficient to learn *implicit* and *top-down* models directly from a set of observed measurements. We treat Wi-Fi links as black boxes with potentially unknown internal mechanics (see Figure 1). Such a black box takes some parameters as inputs (such as the spectral configurations of a link and its neighbors, as well as topological features such as current measurements of channel qualities), and it outputs a throughput value. Our goal is to find any function providing an accurate mapping between (potentially never-observed) inputs and outputs. In particular, we do not attempt to seed a pre-existing model (such as SINR-based or Markov-based) with measurements. Rather, we show that in some cases it can be more efficient to *learn the model itself* from a limited set of measurements.

Constructing useful black boxes is difficult for two main reasons. First, they must capture a fair level of complexity; the cross-layer relationships between the various input parameters and the obtained throughput are usually complex, multi-modal, nonlinear and noisy. Second, it is infeasible to simply measure the link performance for each possible combination of inputs.

Instead of conducting exhaustive measurements, we observe

The work presented in this paper was supported (in part) by the Swiss National Science Foundation under grant number 200021-146423. Part of this work was carried out while the authors were working at Technicolor.

that a statistical representation of these black boxes can be learned by observing a limited number of input/output combinations. Using supervised machine learning techniques, it is possible to generalize the observations made on this limited subset of measurements, while still capturing the complex relationships between the inputs. We build such implicit models using real-world measurements and we test them systematically, by asking them to predict the throughput for links and configurations that have never been observed during the initial measurement phase. We observe that our learned black boxes improve prediction accuracy over models based on the SINR, which is usually the preferred metric for allocating resources such as spectrum or transmit power.

Finally, we demonstrate the usefulness of this "learning by observation" approach, by using one such black box as an oracle for allocating spectrum and transmit power in a dynamic fashion. In particular, we design and implement a complete, utility-optimal algorithm for the joint allocation of spectrum and transmit power. Our algorithm does not rely on a central controller, and requires only local collaboration between neighboring access points (APs). Yet, it converges to a global, network-wide solution of the utility maximization problem. We observe on a testbed deployment that it reacts well to various optimization objectives, such as maximizing throughput and/or fairness. In this context, our black box oracle is instrumental for capturing the intricate interference patterns and finding efficient configurations. To the best of our knowledge, it is the first implementation of a utility-optimal algorithm for spectrum allocation.

The paper is organized as follows. In Section II, we motivate our approach. In Section III, we present our method to learn black box performance models. We evaluate the accuracy and generalization of our models in Section IV. We then present an application for spectrum and transmit power allocation in Section V and discuss the limitations of our models in Section VI. Finally, we present related work in Section VII and give some concluding remarks in Section VIII.

II. Motivation

Although existing models for 802.11 often provide precious insights into various tradeoffs of Wi-Fi performance (see e.g., [1], [9]), obtaining precise and quantitative performance predictions in general cases – i.e., with varying PHY layer conditions – remains a notoriously difficult problem.

To see this, consider for instance networks where nodes can use variable-width channels. Careful allocation of such spectrum chunks is necessary to properly configure recent IEEE 802.11n and 802.11ac amendments with channel bonding [8]. As noted in [26], modeling performance is difficult when several interfering links use channels of variable widths that are possibly overlapping. In fact, a model explicitly designed for this task would have to take into account several complex effects occurring in time, space and frequency domains. For example, using a wide channel bandwidth creates interference in frequency domain, but using a narrow bandwidth increases packet transmission times, which can create more interference in time domain (due to the rate anomaly problem of MAC layers based on CSMA/CA [11]). In addition, for a fixed transmit power, a narrow bandwidth packs more Watts per Hertz, which improves the transmission range [5], but also increases interference in spatial domain.

In general, performance depends in a highly complex way on the actual topology, channel qualities, spectral configurations, etc. This complexity is further exacerbated if the nodes have arbitrary traffic loads, or if they can adapt their transmit powers; although transmit power adaptation can potentially improve spectral re-use [4], it is rarely used in practice as the impact on performance is difficult to predict [22].

For these reasons, there is to our knowledge no model that captures all of the above-mentioned phenomena, and most of the works proposing models or optimizations for the PHY layer (e.g., [22], [23], [26]) are constrained to use SINR-based models. Although SINR models can provide a characterization of the capacity at the PHY layer, they are not meant to capture 802.11 performance and, as we will see now, they can fail to capture important CSMA/CA performance patterns.

A. An Example where SINR Models are Inappropriate

We now consider a real example from our testbed, with two interfering links l and k operating with 802.11n on the same channel of width 20 MHz. We give more details on our testbed and experimental setup in Section IV-A. Both links send saturated UDP traffic. Link l has a fixed transmit power set to 12 dBm, and k varies its transmit power from 3 dBm to 21 dBm. We measure the throughput obtained by l, for two different pairs of links (l, k). For comparison, we also compute the information-theoretic capacity c_l of link l as

$$c_l = \text{constant} \cdot \log_2 (1 + \text{SINR}_l), \qquad (1)$$

where the constant factor accounts for the bandwidth and MIMO configuration, and SINR_l denotes the SINR of link l. On such a two-link setup, the SINR is given by

$$\text{SINR}_l = \frac{P_{l \leftarrow l}}{N_0 + P_{l \leftarrow k}}, \qquad (2)$$

where $P_{l \leftarrow l}$ (resp. $P_{l \leftarrow k}$) denotes the received power at the receiver of l (as measured by our NICs) from the transmitter of l (resp. from the transmitter of k), and N_0 is the background noise (also reported by our NICs).

We show both the measured throughput and the theoretic capacity for the two link pairs on Figure 2. The (schematized) topologies are shown at the top of the figure. For the first link pair, the throughput obtained by l decreases by about 50% when k increases its transmit power. This is due to an increased likelihood of collision at l's receiver and carrier-sensing activation at l's transmitter, as k increases its effective interference range. This qualitative trend is captured by the theoretical capacity, which decreases when $P_{l \leftarrow k}$ increases. However, in this case, the magnitude of the theoretical capacity is much higher than the actual throughput of the link.

The situation is different (and more surprising at first sight) for the second link pair. Here, we can decompose the measured

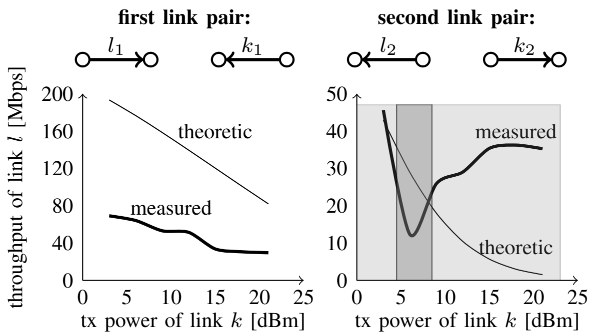

Figure 2: Measured throughput and theoretical capacity of l, when k varies its transmit power. The results are shown for two different pairs of links (l_1, k_1) and (l_2, k_2) from our testbed.

performance in three distinct regimes (represented by three shaded regions in the figure). When k's transmit power is low, the links are nearly independent and l suffers little interference from k. When k's transmit power grows to intermediate values, k starts interfering with l. In this case, l carrier-senses k, and interference mitigation is done in time-domain via CSMA/CA. However, a closer inspection of packets reveals that k *itself* does not have a good channel quality (as it uses only an intermediate transmit power), which forces it to use relatively robust (and slow) modulations. As a result, in this intermediate regime, k consumes a significant portion of the time to transmit its packets, which reduces l's throughput (due to the rate-anomaly). Finally, when k uses a large transmit power, it also uses faster modulations, which has the apparently paradoxical effect of increasing l's throughput.

In this second example, the information-theoretic formulation for the capacity does not capture all these "802.11-specific" cross-layer and multi-modal effects. Instead, it shows a monotonic dependency on transmit power, because it treats the case of Gaussian channels subject to constant and white noise interference.In fact, in the cases where a time-sharing scheme such as CSMA/CA is employed, links often have the opportunity to transmit alone on the channel, thus without observing any interference at all during their transmission[1].

It is thus clear that even in such a simple setting, a resource allocation algorithm relying on monotonic expressions of the SINR is likely to take bad decisions. Despite these problems – and despite the fact that SINR models are usually not considered strong predictors of wireless performance – these models are still the models of choice for allocating resources at the PHY layer, due to their generality: By adapting judiciously the power values in the SINR Equation (2), it is possible to use variable transmit powers (as we just did), but also partially overlapping channels [23] and variable bandwidths [26] as inputs of SINR models. In addition, a large body of literature on *optimal* resource allocation also relies on SINR models in various contexts [3], [10], [15], [19], [22]. By contrast, MAC layer models such as Bianchi's are often accurate with homogeneous PHY configurations, but cannot be used to capture such heterogeneous PHY configurations.

[1]This is also the reason the actual throughput might be largely above the predicted capacity.

III. Learning Performance Models

A. Approach

A natural step to improve the accuracy of SINR-based models is to seed (or fit) some parameters (for instance, a factor controlling the magnitude of the prediction) to the observations of actual measurements. The approach of seeding a model with measurements can be appropriate for networks with collaborative APs, such as enterprise networks, and it has been taken in [21], [26], [27] and others (see Section VII for a discussion). We now show that, if one has the possibility of conducting an initial measurement phase, then it is possible to directly learn the model itself from the data, instead of fitting or seeding a previously existing model. Our overall approach consists of the three following steps.

1) **Measurement phase:** This phase consists in performing N short-duration controlled experiments. Considering again the black box representation of Figure 1 (although generalized for several links), each experiment consists in measuring the throughput of a given link l, for one particular combination of inputs (which we call *features*). This phase is relatively short; we observe in Section IV-D that it is possible to "learn" our entire indoor testbed with reasonable accuracy in less than 6 hours.

2) **Learning phase:** Once the measurements are obtained, this phase consists in finding a mathematical function that maps the features to observed throughputs. This function should approximate the throughput well on the measured data points. However, to be useful, it must not overfit existing measurements, which are intrinsically noisy. Instead, it should generalize to unseen combinations of input features (which can potentially relate to unseen nodes and links). Supervised machine learning provides us with precisely the tools to handle this challenge.

3) **Black box representation:** Once a good function has been found, we can discard the measurements and use the function itself to cheaply compute throughput predictions. Such black boxes can then be used by the APs themselves for selecting efficient configurations (e.g., with predicted throughputs satisfying traffic demands) without probing.

Importantly, we observe in Section IV-C that learned models continue to be useful in new or unseen environments, and that the training procedure does not need to be repeated when new wireless links come and go. We detail our procedure in the remainder of this section.

B. Feature Selection

Consider a link l, for which we want to predict saturated throughput (i.e., under saturated traffic load[2]) for arbitrary spectrum and transmit power configurations, given a set $\mathcal{N}_l$ of K neighboring links with arbitrary conditions, configurations

[2]We target saturated throughput because it is the maximum achievable throughput in a given configuration. In particular, we assume that if throughput t is achievable, then any throughput $t' < t$ is also achievable.

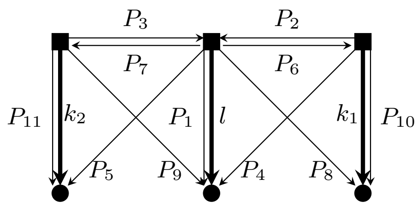

Figure 3: Throughput prediction setting for a link l and two neighboring links $\mathcal{N}_l = \{k_1, k_2\}$. We wish to predict the throughput that l could potentially obtain, given the various received powers $P_1, \ldots, P_{11}$, as well as the physical rates, channel widths, center frequencies, and traffic loads of k_1 and k_2.

and traffic demands. Such a scenario is shown in Figure 3 for $K = 2$. The features must include factors that impact the performance and are measurable by the transmitter of l and its immediate neighbors. We selected the following list of features, because it is known that they all have an immediate impact on performance [4], [5], [11], [23]:

- The power received by each node of l from every transmitting node, and the power received by every other node, from the transmitter of l. These quantities are denoted $P_1, \ldots, P_{11}$ in Figure 3 (assuming downlink traffic, from the APs to their clients). They depend on the transmit powers and the various channel gains, and they can be easily measured online by commodity hardware using RSSI (received signal strength indicator). There are $5K + 1$ such power quantities in general.
- The channel widths used by l and by the links in $\mathcal{N}_l$. There are $K + 1$ such values.
- The spectral separations between the center frequency used by l, and the center frequencies used by all the other links in $\mathcal{N}_l$. There are K such values.
- The K average traffic loads of the links in $\mathcal{N}_l$.
- The physical rates (determined by the 802.11n MCS index) used on each link in $\mathcal{N}_l$. There are again K such values.

Adding up the above-mentioned features, we have access to $d := 9K + 2$ quantities to estimate the throughput that l can obtain in the presence of K interferers. Note that this list of features is not an exhaustive list of the factors affecting performance that can be known or measured by the APs. For instance, we could make it more complete by including the packet sizes, higher order statistics to describe the traffic loads of interferers (instead of the mean only), or more detailed PHY layer information (e.g., capturing multipath effects or frequency-selective fading). Including more features could further increase the predictive power and generality of the learned models. However, the features selected here already allow us to build useful models, while having the advantage of being simple and easy to acquire with commodity hardware.

C. Measurement Phase

The initial measurement phase consists of N measurements with different combinations of features. Some of the features can be directly controlled (namely, the channel widths, spectral separations and traffic loads) and others cannot (the received powers depend both on the transmit powers and channel gains, and the physical rates depend on the auto-rate mechanism used by the APs). Each of the N measurements consists of two sub-experiments. We first perform an experiment during which l is silent, in order to obtain a corresponding vector $\mathbf{x} \in \mathbb{R}^d$ of features (some of which are controlled, others are measured). We then repeat the experiment with l sending saturated traffic, and measure its throughput t_l. Our goal is to expose the learning procedure to as wide a variety of situations as possible. To this end, we apply the following sampling procedure for each of the N data points.

We start by selecting a link l uniformly at random among all the links formed by all the nodes of the network. We then sample K random interfering links, where K itself is randomly drawn between 0 and max_K, and max_K denotes a fixed upper bound on K. For l and the K links in $\mathcal{N}_l$, we sample transmit powers and spectral configurations uniformly at random from the set of configurations that do produce some interference (i.e., such that each link in $\mathcal{N}_l$ uses a band at least adjacent or partially overlapping with l). Finally, for each link k in $\mathcal{N}_l$, we sample a traffic load in the interval $(0, h(w_k)/K]$, where $h(w_k)$ is a value representing the maximum throughput achievable on an isolated link using bandwidth w_k. We take $h(20 \text{ MHz}) = 80$ Mbps and $h(40 \text{ MHz}) = 130$ Mbps in our training procedure, in line with the maximum achievable throughput of our 802.11n cards. Our goal is to predict performance for *arbitrary* interfering loads, and sampling the loads in this way allows us to expose the learning procedure to different environments with both light and heavy contention. In particular, we measured that the offered loads of the nodes in $\mathcal{N}_l$ was above capacity (i.e., saturated) in about 54% of the experiments (mainly due to inter-neighbors interference). The remaining experiments consist of non-saturated conditions.

Once the configurations have been chosen, we perform the first experiment with only the K interfering links active. During this experiment, we measure the average physical rates used by each of the K links in $\mathcal{N}_l$, and we group all the above-mentioned features in a vector $\mathbf{x}_i$. In order to vary K between 0 and max_K but keep features vectors of fixed dimension d, we append $9 \cdot (max\_K - K)$ default "flag" values to $\mathbf{x}_i$, using -110 dBm for all the power values, and setting all the remaining features to zero[3]. We then perform the second experiment in the same conditions, but with link l sending saturated traffic, and we measure its achieved throughput. Each of the two sub-experiments constituting each of the N data points needs only to last a few seconds (in order to measure physical rates and throughput), and the whole procedure is easily automated.

D. Learning

Let us write $\{(\mathbf{x}_1, t_1), \ldots, (\mathbf{x}_N, t_N)\} \subset \mathbb{R}^d \times \mathbb{R}$ for our set of measurements. Our goal is now to find a function $f : \mathbb{R}^d \to \mathbb{R}$ that maps $\mathbf{x}_i$ to a value close to t_i for each measurement i. Learning the function f from the observed data is a regression

[3]The current number of interfering links K is thus an implicit feature, encoded by the presence/absence of flag values.

problem, and we consider the following techniques (see [2], [14] for more details).

Regression tree: This technique fits a binary tree to the data. Each feature vector corresponds to a path in the tree (from the root to a leaf), and each leaf corresponds to a (discretized) throughput value. The resulting model is elegant, because it yields predictions that can be evaluated by a sequence of "if-else" clauses on the features[4]. However, the obtained trees are usually sub-optimal and the hard decision thresholds can affect generalization and accuracy.

Gradient Boosted Regression Trees (GBRT): This technique combines the predictions of M regression trees. Given a feature vector $\mathbf{x}$, the throughput is predicted as $\hat{t} = f(\mathbf{x}) = \sum_{m=1}^{M} \pi_m h_m(\mathbf{x})$. In this expression, $h_m(\mathbf{x})$ denotes the prediction of the m-th tree, and the π_m's are the weighting coefficients (learned with gradient boosting [14]). We obtain the number of trees M as well as their depth by cross-validation. Using several trees has the potential to largely improve the predictive power compared to a single tree, however as we will see, it might still be subject to overfitting.

Support Vector Regression (SVR): For a feature vector $\mathbf{x}$, this method outputs a predicted throughput given by $\hat{t} = f(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i k(\mathbf{x}_i, \mathbf{x}) + b$, where the α_i's and b are the fitted parameters. The function $k(\cdot, \cdot)$ is the kernel function, and we use a kernel specified by $k(\mathbf{x}_i, \mathbf{x}) = \exp(-\gamma \|\mathbf{x} - \mathbf{x}_i\|^2)$, where γ is a parameter obtained by cross-validation. This technique has a high descriptive power, and it can efficiently prevent overfitting.

SINR-based model: As a comparison to pure machine-learning techniques, we also fit SINR-based models to our measurements. In particular, we compute the theoretical capacity c_l of link l as $c_l = \Gamma \cdot w_l \cdot \log(1 + \mathrm{SINR}_l)$, where Γ is a constant that is fitted to measurements (using a least square fit), in order to correct for the magnitude problem mentioned in Section II. In addition, we also use the approach proposed in [23] to account for partially overlapping channels; namely, we scale each power value appearing in the SINR Equation (2) by an appropriate value that accounts for the spectral overlap (assuming perfect bandpass filters). To the best of our knowledge, such models are the only existing models that can capture arbitrary spectral configurations with variable widths and transmit powers.

IV. Evaluation of Performance Predictions

In this section, we evaluate the accuracy and generalization of the different learning strategies in various conditions.

A. Experimental Setup and Methodology

1) Experimental Setup: We use a testbed of 22 nodes spread over an entire floor of an office building (see Figure 4). The nodes are Alix 2D2 boards, equipped with Atheros AR9220 wireless adapters. They run the OpenWrt 10.03 Linux distribution with the open source ath9k wireless drivers, and they use the default Minstrel autorate algorithm. We employ 20 and 40 MHz channel widths with 802.11n, 2×2 MIMO, and 10 different transmit power values in the set $\{3\text{dBm}, 5\text{dBm}, \ldots, 21\text{dBm}\}$. We use the 5.735-5.835 GHz band.

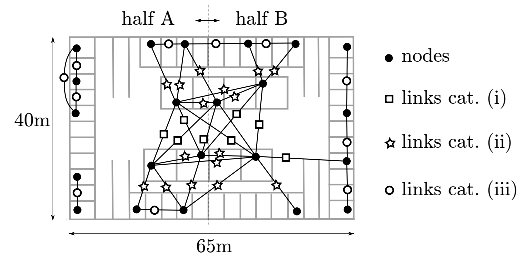

Figure 4: Layout of our 22-nodes wireless testbed. We also show the different link categories and the two halves of the testbed used in the experiments of Section IV-C2.

2) Methodology: We want to test predictions for unknown combinations of features. As such, we only predict throughputs for data points that do not appear in the N measurements used for learning (or training). To this end, we always split our total set of measurements into a *training set* and a *test set*. The training set consists in the actual N measurements used for learning the models and their parameters, whereas the test set is used only once, for measuring the final accuracy.

We gathered a trace of about 8900 measurements[5], with $max\_K = 3$. This set is voluntarily larger than what is actually needed, in order to allow us to test the effect of the number of measurements N on the models quality.

To evaluate the accuracy of predictions, we use the *coefficient of determination* R^2. If we have a test set with n throughput measurements $t_1, \ldots, t_n$ and a given model predicts the throughputs $\hat{t}_1, \ldots, \hat{t}_n$, then it is given by $R^2 := 1 - \left(\sum_i (t_i - \hat{t}_i)^2\right) / \left(\sum_i (t_i - \bar{t})^2\right)$, where $\bar{t}$ is the average throughput, given by $\bar{t} = \frac{1}{n}\sum_i t_i$. Concretely, the R^2-score quantifies how well a predictor does, compared to the simplest baseline strategy, which always predicts the mean throughput. It is equal to 1 for perfect predictions. We also compute the RMS error (RMSE), defined as $\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_i (t_i - \hat{t}_i)^2}$. We used the Python machine learning package `scikit-learn` to learn the various models.

B. Prediction Accuracy

In order to compare the accuracy of the different classes of models, we perform 50 consecutive splits of our measurements in training and test sets (50-fold cross-validation). For each split, we evaluate the R^2-score and RMSE, and we show the average and standard deviations in Figure 5(a) for each class of model. In addition, we also show the detailed distribution of prediction errors in Figure 5(b) for models based on SVR and GBRT.

It appears clearly that the learned models, in particular the ones based on SVR and GBRT, perform significantly better than the SINR-based models. In terms of R^2-score, learned SVR and GBRT models improve the prediction accuracy by 54% and 71%, respectively, compared to SINR models (which,

[4]For instance, on a simplistic tree of depth 2, a regression path could look like: "if received power $\leq X$ and frequency offset $> Y$, then predict Z".

[5]Our dataset is publicly available: http://www.hrzn.ch/data/lw-data.zip

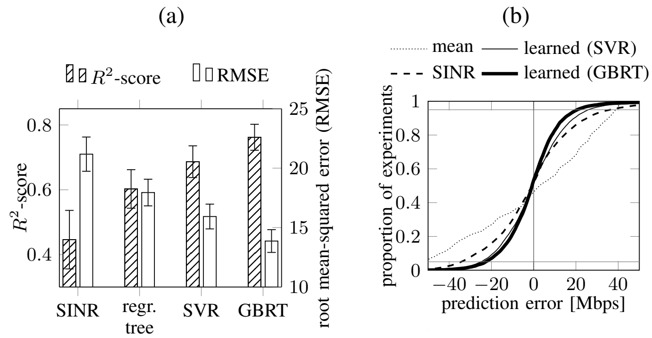

Figure 5: Summary of prediction performance for various models (a) and empirical CDF of prediction errors (b). The "mean" model in plot (b) represents the errors obtained by a baseline predictor that always predicts the mean throughput of the training set.

we recall, are the only known class of models capturing phenomena such as overlapping channels). In terms of error distribution, 90% of the errors made by learned models are between -25 Mbps and 25 Mbps, whereas 90% of the errors made by SINR-based models are between -35 Mbps and 36 Mbps. The fact that learned models are more accurate is remarkable; it demonstrates that, as far as performance prediction is concerned, learning abstract models coming from the machine learning domain can be much more efficient than trying to fit (or seed) pre-existing specialized models.

In order to visualize the actual predictions in detail, we also show a scatter plot of the throughputs predicted by SINR models and learned SVR models, against the actual measured throughputs, in Figure 6. Clearly, SVR models perform much better and produce fewer outlying predictions than SINR models. Note that obtaining *perfect* predictions is impossible here, considering the fact that both the measured features and the throughput are highly noisy variables, measured with commodity hardware. To illustrate this, we examine in more detail the features corresponding to the worst prediction obtained by both models (shown by an arrow on the plots – incidentally, this is the same point for both models). This point corresponds to a link l subject to no (controlled) interference (i.e., $K = 0$), with an apparently good channel quality (the measured RSSI is -59 dBm), and using a bandwidth of 40 MHz, supposedly yielding the largest capacity. Yet, despite these features, the measured throughput was low. We can only speculate about the causes for this discrepancy (it may be due to especially high noise or software factors). In any case, this example illustrates the limits of throughput predictability with imperfect information.

C. Generalization

Due to the evaluation on test set, the previous results address cases where predictions are produced for unseen combinations of *features*. We now attempt to push our models further, by predicting throughputs for unknown *links*, potentially belonging to different environments.

1) Predictions for Unknown Links: For each possible link l, we remove both l and its reverse link (obtained by inverting the transmitter and the receiver of l) from the training set. We then predict throughput for each data point that contains l (or its reverse link), and show the results in Figure 7. Compared with Figure 5(a), some models (especially the ones based on regression trees) see their accuracy slightly decreased. However, the models learned with SVR still perform remarkably well; in terms of R^2-score, their accuracy is reduced by less than 4%, and they still improve the accuracy by 49% compared to SINR-based models.

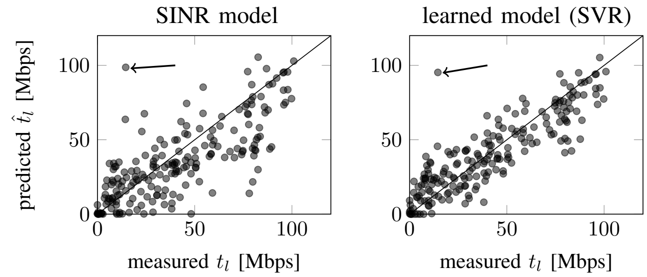

Figure 6: Predicted versus measured throughput, for SINR and a learned model, on a test set of 200 points.

2) Different Environments: We now manually divide the links present in our trace in three distinct categories, depending on the type of attenuation that they experience. The categories are shown in Figure 4, and they correspond to the following link division: (i) links that traverse mostly empty space, (ii) links that traverse sparsely spaced walls and (iii) links that traverse densely spaced walls.

For each category, we remove all the links (and their reverse) belonging to this category from the training set. We then build the test set so as to predict throughput for links belonging *only* to this category. The goal of this experiment is to test prediction accuracy in the worst possible conditions: each model is learned on links that operate in conditions radically different than the conditions prevailing during the actual predictions. In addition to the three link categories (i)-(iii), we also split our testbed in two halves A and B (also shown in Figure 4). The resulting accuracies are shown in Figure 8. Even in these difficult cases, the learned models based on SVR show a graceful degradation and keep a relatively high accuracy (with R^2-scores always larger than 0.54). When predicting on half B with models learned on half A, models based on SVR even obtain similar accuracies as when learning using the full testbed. This allows us to draw some conclusions on the extent to which our method generalizes. Even when learning models on a different part of the testbed, or using radically different links, abstract models based on machine learning still have far more predictive power than measurement-seeded models based on SINR.

D. How Much Learning is Needed?

Finally, we measure the accuracy as a function of the training set size N. For different N, we learn models using N experiments sampled at random from our experiment trace. We then predict the throughput for all the other experiments, and measure the R^2-score. The results are shown in Figure 9. Using $N = 100$ training experiments is enough to obtain better accuracy than SINR models, and $N = 1000$ experiments

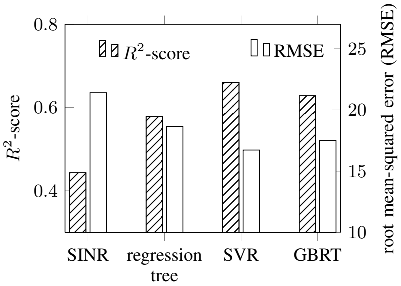

Figure 7: Prediction accuracy on links that have never been observed during the learning phase.

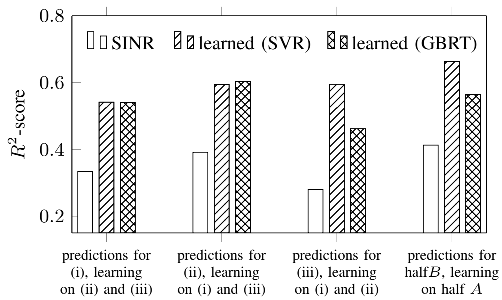

Figure 8: Prediction accuracy for never-observed groups of links. Even in the difficult cases, learned models largely outperform SINR models.

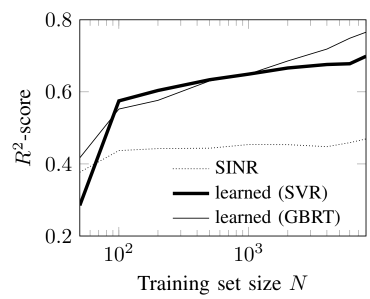

Figure 9: Prediction accuracy as a function of number of measurements N (note the logarithmic scale of the x-axis).

already yield good predictive accuracies. If each experiment lasts 10 seconds (which is the duration that we employed), an efficient performance model for an entire building-scale network such as ours can be learned in less than 6 hours.

V. Spectrum Allocation

To show that learned models can be useful in practice, we designed and implemented a complete decentralized algorithm for the joint allocation of spectrum (channel center-frequency and bandwidth) and transmit power. Due to space constraints, we only give a brief overview of our algorithm, and refer the reader to [16] for more details.

A. Algorithm for Spectrum and Transmit Power Allocation

Overall Description. We consider a utility maximization setting, where each link l is attached a utility function $U_l : \mathbb{R} \to \mathbb{R}$. The utility functions are arbitrary and need not be concave. The APs running our algorithm wake up after random time intervals (with typical mean durations of a few minutes). When an AP A wakes up, it contacts its direct neighboring APs, and requests information about the current configurations, channel measurements and traffic loads (i.e., the current features). Using this information, A predicts the achievable throughput (and the corresponding utility) on each of its attached links, for each possible configuration of spectrum and transmit power. Note that, even though these predictions must account for complex spectrum and channel settings, they are cheap and easy to obtain using one of our learned model. A then samples a new configuration, according to the Gibbs distribution, which gives more weight to configurations with large achievable utilities. It can be shown that, even though the algorithm acts distributively, it converges towards configurations such that the link throughputs x_l's maximize $\sum_l U_l(x_l)$, where the sum runs over all the links in the network. Our algorithm has similar convergence properties as [3] and [15], and we give its full specification in [16].

Implementation. We implemented the complete distributed algorithm in about 3000 lines of C++ code, using *Click* [20] in user space. Our implementation comprises a distributed neighbor discovery mechanism, in the form of a rendez-vous protocol. The APs periodically switch to a pre-determined 20 MHz channel (to have the largest possible communication range), and send a broadcast frame that contains their public (wired) IP address. The neighboring APs that overhear this address then use their wired connection for the actual collaboration. In particular, the APs communicate various up-to-date features that they measure from neighboring APs, as well as from their own and neighboring clients. They also inform their neighbors about their own current traffic loads (which they can easily measure themselves). The performance predictions are all obtained with a black box model learned with SVR. In our evaluation, the algorithm deals with random configurations, channel gains and AP-clients combinations that have in general never been observed during the learning procedure.

B. Algorithm Evaluation

Experimental Methodology. Unless otherwise stated, we use the following experimental methodology. We randomly select between 8 and 10 AP-client pairs among the 22 nodes of our testbed. Each pair starts in a random configuration of channel, width and transmit power. The APs send saturated UDP traffic generated by `iperf` to their clients. The mean wake-up time is set to 600 seconds (meaning that each AP "reevaluates" its spectral configuration every 10 minutes on average). In addition, we use the following three utility functions $U_l(x_l)$ in our study:

- $U_l(x_l) = x_l$. When all links use this utility function, the optimization target consists in maximizing the sum of throughputs, irrespective of other considerations such as fairness. We denote this utility function U^{thr}.
- $U_l(x_l) = \log(1 + x_l)$. Using this function is equivalent to maximizing proportional fairness; we denote it U^{prop}.
- $U_l(x_l; \alpha) = (1 - \alpha)^{-1} x_l^{1-\alpha}$. This is the α-fairness utility function defined in [24], with $\alpha > 1$. Taking $\alpha \to \infty$ yields allocations that are max-min fair, and $1 < \alpha < \infty$ represents a compromise between proportional fairness and max-min fairness. We denote this function U^{α}.

Finally, we benchmark our algorithm against the one proposed in [19]. This algorithm finds configurations of channel center frequencies that minimize the overall interference. We augment it to sample bandwidths and transmit powers as follows, for a fair comparison. We modulate the power received by a node a from a node b by (i) the transmit power used by b and (ii) the overlap between a's receive spectrum mask and b's transmit spectrum mask (see [23]), assuming perfect band-pass filters. We run the algorithm [19] (with our augmented metric) offline, using the whole testbed channel gains matrix in input, for 1000 iterations. The resulting allocations are denoted K+, and are run for 1000 seconds in our testbed. This is repeated 10 times to obtain confidence intervals.

Performance. We conduct experiments where all the links use U^{thr}, U^{prop}, or U^{α} with $\alpha = 4$. Figure 10 shows the steady state throughput and Jain's fairness index, obtained by computing $\left(\sum_l x_l\right)^2 / (L \cdot \sum_l x_l^2)$, where x_l is the throughput obtained by link l. Quite remarkably, the practical results obtained on the testbed reflect well the objectives of the various utility functions: U^{thr} provides the greatest throughput, while both U^{prop} and U^{α} improve fairness. Furthermore, in line with theoretical expectations, U^{α} provides slightly better fairness and lower throughput than U^{prop}. To the best of our knowledge, this is the first observation that the framework of utility maximization can be used with spectrum assignment to achieve various optimization objectives in a real testbed.

The key element that allows our algorithm to jointly optimize the network over all three parameters is our learned black box model. To see this, we also plot the throughput and fairness for U^{thr} obtained when our algorithm uses a measurement-seeded SINR model (instead of a black box) in Figure 10 (labeled SINR). It appears clearly that both throughput and fairness are improved (and less variable) when using a learned model.

Selected Configurations. Figure 11 shows the distribution of transmit powers selected by the algorithm (over all nodes and all experiments), for the three utility functions. Fairer policies use lower transmit power for a higher fraction of time (see U^{α} vs U^{thr}). This means that the aggressiveness of the configurations (here in spatial domain), can be directly controlled by the utility functions.

We now study the impact of traffic loads, as it is taken into account by our models. We perform experiments where each link l has a traffic load $load_l$, which is randomly chosen between 10 and 80 Mb/s. We use the utility functions $U_l(x_l) = \min\{x_l/load_l, 1\}$. This function is equivalent to U^{thr} when all links are saturated, and it is maximized as long as all links obtain a throughput that satisfy their demand. Figure 12 shows the proportion of time that the algorithm selects a 40 MHz channel width (on the left y-axis), and the average transmit power in dBm (right y-axis), as a function of traffic load at the AP that makes these choices.

We observe an elegant load-balancing pattern, as APs with heavier loads use more spectrum and larger transmit powers. This is a desirable feature (see e.g., [12]), and it directly relies on the ability of our learned models to suitably capture the interference effects of variable traffic loads. Interestingly, note that, when *all* APs generate 100 Mb/s of load (labeled "all 100" in Figure 12), the APs lower their resource consumption compared to cases with heterogeneous loads. This is because, in these cases, heavily-loaded APs compete with other heavily-loaded APs, and they naturally collaborate to share spectrum equitably. Overall, our black box oracle allows the resource allocation to finely load balance spectrum usage as a function of the utilities and fairness objectives.

VI. Limitations and Discussion

We have evaluated our learned models in static conditions, a setting for which throughput prediction is somewhat easier (compared to say, high mobility with fast fading, short channel coherence times, etc). This is because, in this paper we deliberately restrict ourselves to using features easily accessible on commodity hardware (e.g., RSSI measurements). Such features are only meaningful on relatively coarse timescales (typically seconds) and cannot capture such fast-changing phenomena. In this sense, our black boxes suffer the same timescale limitations as any model (including SINR) using similar measurements. Whether a similar learning framework could be applied to shorter timescales is left for future work.

Importantly, using features operating at relatively coarse timescales already allows our learned models to be useful in practice. In Section V, we considered a setting where the global spectrum consumption are re-evaluated every few minutes by the APs[6]. Such global, relatively slow-varying spectrum allocation complements well (and provides more spectrum to) existing PHY techniques operating at fast timescales, such as interference cancellation and alignment.

VII. Related Work

Performance Models. Several papers propose measurement-based approaches to model performance and interference in 802.11 networks. In particular, [18], [21], [25], [27] use initial measurement campaigns, where the number of measurements is typically a function of the number of nodes present in the network. [27] fits a model based on the SINR in order to estimate the packet loss probability, whereas [18], [25] and [21] use measurements-based Markov chain models to predict the capacity and/or interference of 802.11 networks. All of the above models are agnostic to the spectral configurations of the nodes, and they are designed to work when the links operate with a fixed channel width. In this paper, we also use an initial measurement phase. However, we are not constrained to using any particular model, but rather employ machine learning to learn *any* suitable model that captures both PHY and MAC layer complexities together.

[13] observes that measurements at the OFDM subcarrier level largely improves the accuracy of performance prediction. Unfortunately, the method does not take interference into account, and it cannot be used to make performance predictions when several links operate at the same time.

Finally, a few papers propose to use machine learning techniques in the context of wireless networks. [7] discusses the use of k-NN for link adaptation and [6] proposes an architecture for cognitive radios with learning abilities. However, these works do not attempt to predict performance. To the best of our knowledge, ours is the first work using machine learning to predict actual Wi-Fi performance.

Resource Allocation. Some recent works consider simultaneous channel center frequency and width allocation for

[6] At faster timescales, the overhead of switching to different spectrum bands on commodity hardware would exceed the benefits of employing efficient spectrum allocations.

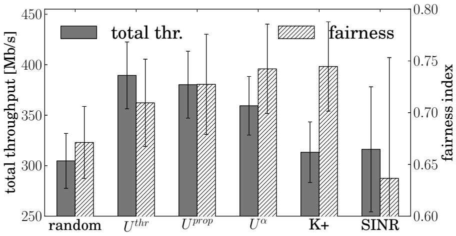

Figure 10: Throughput and fairness for different utility functions. We also show the results for random configurations, the "K+" algorithm, and our algorithm with U^{thr}, but using the measurement-seeded SINR model instead of a learned model.

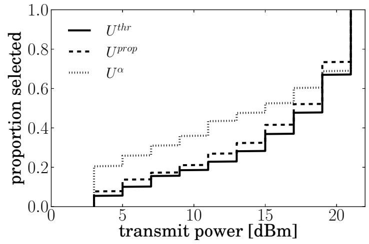

Figure 11: CDF of transmit powers selected with different utilities and sampling policies. Fair policies select lower transmit powers.

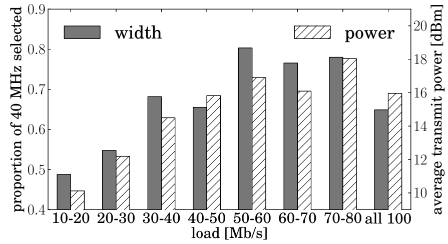

Figure 12: Proportion of time a bandwidth of 40 MHz is selected, and selected transmit power, as a function of traffic load. Efficient load balancing is achieved, as heavily loaded APs sample wider bandwidths and larger transmit powers.

802.11 networks. [26] runs the spectrum allocation jointly with scheduling decisions at a central controller, and [17] proposes a distributed algorithm for the joint allocation of center frequencies and bandwidths. None of these algorithms considers the transmit power, and they do not adapt to various utility functions. Our learned models predict achievable throughputs, which can be directly plugged into the utility maximization framework. This removes the need to use indirect optimization objectives (such as minimization of interference, which often does not coincide with performance maximization [17]).

The theoretical works that are the closest to ours are [3], [15]. These papers propose optimal algorithms for channel and/or power selection, but do not consider channel width. Further, they have not been implemented in real networks.

VIII. Conclusions

We investigated and validated a new approach for predicting the performance of Wi-Fi networks. Rather than manually fitting complex models to capture complex dependencies, we showed that it is possible to directly learn the models themselves, from a limited set of observed measurements. This approach bypasses the usual modeling process, which requires both deep knowledge and tedious analysis, and yet often yields models that are either too restricted or too inaccurate. We observed that abstract black box models built using supervised machine learning techniques – without any deep knowledge of the complex interference dynamics of 802.11 networks – can largely outperform the dominant class of SINR-based models. Further, we have shown that these models still work when they have to predict performance for links that have never been observed during the learning phase.

We have used one such model as an oracle in a new distributed utility-optimal resource allocation algorithm. We observed that our algorithm adapts well to various optimization criteria, and that our learned model is instrumental for achieving good performance in these tangled settings.

References

[1] G. Bianchi. Performance analysis of the ieee 802.11 distributed coordination function. *IEEE JSAC*, 2000.

[2] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag New York, 2006.

[3] S. Borst, M. Markakis, and I. Saniee. Distributed power allocation and user assignment in OFDMA cellular networks. In *Allerton*, 2011.

[4] I. Broustis, J. Eriksson, S. V. Krishnamurthy, and M. Faloutsos. Implications of power control in wireless networks: a quantitative study. In *PAM*, 2007.

[5] R. Chandra, R. Mahajan, T. Moscibroda, R. Raghavendra, and P. Bahl. A case for adapting channel width in wireless networks. In *ACM SIGCOMM*, 2008.

[6] C. Clancy, J. Hecker, E. Stuntebeck, and T. O'Shea. Applications of machine learning to cognitive radio networks. *IEEE Wireless Communications*, 2007.

[7] R. C. Daniels and R. Heath. An online learning framework for link adaptation in wireless networks. In *IEEE ITA, 2009*.

[8] L. Deek, E. Garcia-Villegas, E. Belding, S.-J. Lee, and K. Almeroth. The impact of channel bonding on 802.11n network management. In *ACM CoNEXT '11*.

[9] M. Durvy, O. Dousse, and P. Thiran. Modeling the 802.11 protocol under different capture and sensing capabilities. In *INFOCOM*, 2007.

[10] R. Etkin, A. Parekh, and D. Tse. Spectrum sharing for unlicensed bands. *Selected Areas in Communications, IEEE Journal on, 2007*.

[11] M. H. Franck, F. Rousseau, G. Berger-sabbatel, and A. Duda. Performance anomaly of 802.11b. In *IEEE INFOCOM*, 2003.

[12] R. Gummadi and H. Balakrishnan. Wireless networks should spread spectrum based on demands. In *ACM HotNets*, 2008.

[13] D. Halperin, W. Hu, A. Sheth, and D. Wetherall. Predictable 802.11 packet delivery from wireless channel measurements. In *SIGCOMM'10*.

[14] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning: data mining, inference and prediction*. Springer, 2008.

[15] N. Hegde and A. Proutiere. Optimal decentralized spectrum sharing: A simulation approach. In *Proc. CISS*, Mar. 2012.

[16] J. Herzen, N. Hegde, and H. Lundgren. Learning wi-fi performance (technical report). http://www.hrzn.ch/publications/learning-tr.pdf.

[17] J. Herzen, R. Merz, and P. Thiran. Distributed spectrum assignment for home wlans. In *IEEE INFOCOM*, 2013.

[18] A. Kashyap, S. Ganguly, and S. R. Das. A measurement-based approach to modeling link capacity in 802.11-based wireless networks. In *ACM MobiCom*, 2007.

[19] B. Kauffmann, F. Baccelli, A. Chaintreau, V. Mhatre, K. Papagiannaki, and C. Diot. Measurement-based self organization of interfering 802.11 wireless access networks. In *IEEE INFOCOM*, 2007.

[20] E. Kohler, R. Morris, B. Chen, J. Jannotti, and M. F. Kaashoek. The click modular router. *ACM Transactions on Computer Systems, 2000*.

[21] Y. Li, L. Qiu, Y. Zhang, R. Mahajan, and E. Rozner. Predictable performance optimization for wireless networks. In *SIGCOMM*, 2008.

[22] V. Mhatre, K. Papagiannaki, and F. Baccelli. Interference mitigation through power control in high density 802.11 WLANs. In *IEEE INFOCOM*, 2007.

[23] A. Mishra, V. Shrivastava, S. Banerjee, and W. Arbaugh. Partially overlapped channels not considered harmful. In *ACM Sigmetrics*, 2006.

[24] J. Mo and J. Walrand. Fair end-to-end window-based congestion control. *IEEE/ACM Transactions on Networking (ToN)*, 8(5):556–567, 2000.

[25] L. Qiu, Y. Zhang, F. Wang, M. K. Han, and R. Mahajan. A general model of wireless interference. In *ACM MobiCom*, 2007.

[26] S. Rayanchu, V. Shrivastava, S. Banerjee, and R. Chandra. FLUID: Improving throughputs in enterprise wireless lans through flexible channelization. In *ACM MobiCom*, 2011.

[27] C. Reis, R. Mahajan, M. Rodrig, D. Wetherall, and J. Zahorjan. Measurement-based models of delivery and interference in static wireless networks. In *ACM SIGCOMM*, 2006.