Improved Radiometric Identification of Wireless Devices Using MIMO Transmission

Yan Shi, *Student Member, IEEE*, and Michael A. Jensen, *Fellow, IEEE*

***Abstract*—Authenticating wireless devices based on features of their transmitted waveform has become a topic of considerable interest. Recent work in this arena has shown that examination of features in a modulated waveform can lead to highly precise identification of 802.11 devices. This paper experimentally demonstrates improved identification accuracy and reduced sensitivity to identification parameters when such techniques are applied to the multiple transmitters within multiple-input multiple-output devices. It further studies an information theoretic technique for determining which radiometric features are most effective in discriminating devices and analyzes the identification performance as a function of the number of features used. Finally, because some radiometric features are sensitive to environmental conditions and, therefore, drift with time, the analysis explores the impact of radiometric feature drift both on feature prioritization and on identification performance.**

***Index Terms*—Computer network security, fingerprint identification, multiple-input multiple-output (MIMO) systems.**

I. Introduction

DEVICE identity management is a challenging issue in any network security solution. This is especially true for wireless networks where identity information is often relatively easy to forge, allowing an attacker to easily enter a network and possibly intercept sensitive information. As a result, most networks rely on cryptographic techniques based on establishment of secret encryption keys in an effort to ensure only trusted users are given access to the network [1], [2]. However, the challenges associated with implementation of key-based authentication, such as manual entry of public keys or the significant resource requirements of implementation [3], [4], and the potential security weaknesses associated with these approaches have led to research focusing on other authentication mechanisms that can replace or work in cooperation with cryptographic authentication techniques. Examples of such techniques include radio-frequency fingerprinting implemented via radiometric device identification [3], [4] or location identification [5], [6], the latter of which is appropriate only for stationary network nodes.

Because of its successful use in military and commercial network security, radiometric device identification has received considerable research attention in recent years. In this technology, unique imperfections inherent in the components used to construct the transmitter system hardware create a waveform signature (fingerprint or transceiverprint) described by a set of features that can uniquely determine the transmitter identification. Most commonly, the features considered are observed during signal transients [4]–[11]. However, recent work has shown that enhanced identification accuracy is enabled by focusing on characteristics of the modulated transmitted waveform [12]. This work further proposes an identification technique referred to as the PAssive RAdiometric Device Identification System (PARADIS) that when applied to a large set of network interface cards (NICs) achieves excellent identification accuracy.

Given this prior success of modulated-waveform-based radiometric identification, we experimentally evaluate its application to devices using multiple-input multiple-output (MIMO) technology. Specifically, by exploiting the enriched set of observable waveform features associated with the multiple transmitters of MIMO devices, we achieve improved identification accuracy and reduced sensitivity to algorithmic implementation parameters. Furthermore, because the increased set of possible features associated with a MIMO device can lead to an overly complex identification model, we implement a minimal-redundancy–maximal-relevance (mRMR) technique [13], [14] to order the observed features by their effectiveness in device discrimination. This ordering allows study of the identification performance as a function of the number of features used. Finally, while observing a set of features at one time and using this observed set to identify the device for an extended period thereafter is a commonly assumed scenario, we recognize that features that depend on analog hardware components will generally drift with temperature or other environmental parameters [15]. We therefore present an analysis of the impact of drift on feature selection and identification accuracy, with the results suggesting that identification accuracy for MIMO devices is less sensitive to feature drift than that for single-input single-output (SISO) devices.

II. Objective and Performance Quantification

A. Objective and Data Model

In the PARADIS technique that was previously proposed for SISO wireless devices [12], a number of features of the modulated waveform are measured for each frame of the transmitted data signal. Conceptually, the relevant features are measured for each device that is authorized to participate in network communication during a *training phase*, and the decision-making agents within the network are then given a database of these observed features. When one of these nodes later requests entrance

Manuscript received January 24, 2011; revised June 29, 2011; accepted July 09, 2011. Date of publication July 25, 2011; date of current version November 18, 2011. This work was supported in part by the U.S. Army Research Office under the Multi-University Research Initiative (MURI) Grant W911NF-07-1-0318. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Darko Kirovski.

The authors are with the Electrical and Computer Engineering Department, Brigham Young University, Provo, UT 84602 USA (e-mail: sigmanacl@gmail.com; jensen@ee.byu.edu).

Digital Object Identifier 10.1109/TIFS.2011.2162949

into the network, the decision-making agents compare their observations of these features with the characteristics recorded in the database during the *authentication phase*. These agents allow or deny network entrance based on the match between the observed features and those contained in the database. Prior work demonstrates that such an approach can uniquely identify different transmitters with relatively high accuracy. It is important to mention that this prior work suggests such radiometric identification should be viewed as an additional security layer to be used in combination with more traditional authentication techniques [12], an observation that is very reasonable given current technology.

More formally, consider N_d devices that are authorized to enter a network, with the nth device identified by a unique integer number h_n, $1 \leq n \leq N_d$. In this work, we sample the radiometric features at a temporal sample interval of T_0 over the time window $0 \leq t \leq T_s$. Let $g_{n,i}^{(k)}$ represent the kth temporal sample of the ith radiometric feature $(1 \leq i \leq N_f)$ for the nth device. For simplicity, we write one observation of the selected waveform features for the device as the $N_f \times 1$ vector $\mathbf{g}_n^{(k)} = [g_{n,1}^{(k)}, g_{n,2}^{(k)}, \cdots, g_{n,N_f}^{(k)}]^T$, where $\{\cdot\}^T$ indicates a transpose operation.

Given this notation, the goal of device identification is to find a function that uniquely maps the joint radiometric features to the device identity, or $\hat{h}_n = F(\mathbf{g}_n^{(k)})$, where $\hat{h}_n$ is the estimate of the true node identification h_n. The mapping function F can be trained via regression on a set of training data. For simplicity, we assume that F is trained using the first portion of the sampled data representing the temporal training duration $0 \leq t \leq T_t$, where $T_t = \gamma T_s$ and the scalar training ratio γ is bounded as $0 \leq \gamma < 1$. Over the time interval $T_t < t \leq T_s$, the trained model is applied to the observed data to determine its performance in properly identifying the radio devices.

Finally, observations from real radios show that the samples of many of the radiometric features have relatively large variance, which makes their use in device identification difficult. To reduce the variance observed in the measured features and to enhance reliability of the identification, the features are averaged over a sliding window of length $T_w = K_w T_0$, or

$$\overline{\mathbf{g}}_n^{(k)} = \frac{1}{K_w} \sum_{m=k-K_w+1}^{k} \mathbf{g}_n^{(m)}. \tag{1}$$

These averaged features are used to train the mapping function F. The impact of the choice of T_w is studied in the experimental results included.

B. Feature Selection

Before exploring the development of a classification algorithm that implements the mapping function from the radiometric features to the device identification, it is important to first consider the set of radiometric features that can effectively discriminate between different devices. Based upon the observations given in [12], we consider the value of parameters averaged over a transmission frame. Specifically, in this work we observe the error vector magnitude (EVM) of the data as well as the pilot tones of the orthogonal frequency-division-multiplexing (OFDM) signal, the error in the carrier center frequency (Freq Offset), the common phase error of the OFDM tones, and the error in the symbol clock of the transmitted signal. We also observe the origin offset (I/Q Offset), phase rotation (I/Q Rotation), and gain imbalance (I/Q Gain Imbalance) of the constellation defined by the in-phase (I) and quadrature (Q) components of the baseband signal for each frame, along with the correlation between the observed SYNC signal in the 802.11 frame and an ideal SYNC signal (SYNC Corr).

While it is intuitive that identification accuracy is maximized when using all available features for model training and application, naturally this leads to increased model and, therefore, computational complexity. Therefore, we wish to identify a subset of features whose behaviors are highly correlated with the device identification, which is a measure of their *relevance*. However, we also recognize that multiple features may have behaviors that are highly correlated, which is a measure of their *redundancy*. Therefore, in our search for a useful subset of features, we seek for maximum relevance and minimum redundancy.

A mathematical framework for determining this set of features is based on the mutual information. Relevance of a feature g_i is measured by the mutual information of that feature and the device identifier h [13], or

$$I(g_i, h) = \int \sum_{h \in \mathbf{h}} p(g_i, h) \log \frac{p(g_i, h)}{p(g_i)p(h)} dg_i \tag{2}$$

where $\mathbf{h}$ represents the set of N_d possible values of h and $p(\,\cdot\,)$ denotes a probability density function. The relevance of a set of selected features $\mathcal{S} = \{g_1, g_2, \cdots, g_{N_f}\}$ is defined as the average of the relevance of features in the set [13], or

$$V_{\mathcal{S}} = \frac{1}{N_f} \sum_{g_i \in \mathcal{S}} I(g_i, h). \tag{3}$$

Similarly, redundancy is measured by the mutual information of two different features. The redundancy of the feature set is defined as the average of the mutual information of each pair of features in the set, or

$$W_{\mathcal{S}} = \frac{1}{N_f^2} \sum_{g_i, g_j \in \mathcal{S}} I(g_i, g_j). \tag{4}$$

The goal is now to find the set that maximizes the relevance $V_{\mathcal{S}}$ while minimizing the redundancy $W_{\mathcal{S}}$. Prior work on this topic has shown that finding the feature set that maximizes the difference $V_{\mathcal{S}} - W_{\mathcal{S}}$ is one simple approach that achieves the objective [13], [14]. The mRMR feature selection algorithm resulting from this prior work is used in this paper to determine the optimal subset of features for device identification.

C. Identification Algorithm

Once a subset of features is selected from the available feature set, a classifier must be constructed based on feature observations. Generally speaking, the device identification depends nonlinearly on the radiometric features selected, and as a result, the regression can sometimes create a model whose order is too large (leading to excessive computational burden or other challenges [16], [17]) or that demonstrates unstable performance [18], [19]. To overcome this difficulty, for this study we select the modeling technique known as bootstrap aggregating or

bagging. For this technique, the training data is divided into M subsets, and each subset is used to train a nonlinear predictor known as the *kernel*. Each predictor is then used on the test data to determine the device identification, and the identification decision from each of the multiple kernels represents a *vote* for the final identification. In this way, if any kernel demonstrates unstable performance, its importance in the overall decision process is reduced [20]. Furthermore, the effect of averaging (through voting) reduces the variance in the final identification decision. Detailed analysis of this method along with a mathematical proof of its enhanced performance is provided in [21].

In our work, the kernel is the C4.5 decision tree (DT) algorithm [22] implemented in the data mining tool *Weka* [23] as algorithm J48. This kernel first constructs the discrete probability of observing each device—which in our case is uniform such that the probability of observing h_n is $p_n = 1/N_d$—and then computes the information required to identify the device as the entropy

$$\mathcal{E} = -\sum_{n=1}^{N_d} p_n \log_2 p_n. \tag{5}$$

For each of the N_f features, the kernel then divides the training data into L_i groups, where L_i is the number of values assumed by the ith feature (or the number of *bins* into which the range of the feature is divided), and computes the entropy $\hat{\mathcal{E}}_{\ell_i}$ for each group, $1 \leq \ell_i \leq L_i$. The information required to identify the device given the ith feature is then

$$\mathcal{E}_i = \sum_{\ell_i=1}^{L_i} \hat{p}_{\ell_i} \hat{\mathcal{E}}_{\ell_i} \tag{6}$$

where $\hat{p}_{\ell_i}$ is the discrete probability that an observation lies within the ℓ_ith group. The groups associated with the feature that achieves the largest difference $\mathcal{E} - \mathcal{E}_i$ are then used to construct the next branches in the decision tree. The procedure repeats with the remaining features to construct the tree until no more features are available or a branch ends with a group in which all elements have the same identification. The C4.5 algorithm actually has additional elements, such as normalization of the difference in entropy, that improve its performance. More details regarding this DT kernel are available in [22].

Prior work on radiometric device identification has instead used a support vector machine (SVM) for this implementation, demonstrating that it provides high accuracy in identification [12]. Our choice to depart from this is motivated simply by the fact that the bagging algorithm with a DT kernel results in reduced computational cost while achieving virtually identical accuracy (see Section III-F for a brief experimental comparison). Furthermore, prior analysis shows that the memory consumption of a DT in the *worst* case scales linearly with the problem dimensionality [24], while that of SVM in the *best* case scales linearly with the problem dimensionality [25], indicating that the memory requirements of the bagging algorithm implementation are less than or at most equal to those of SVM.

D. Performance Metrics

Effectively evaluating the overall performance of an identification model is not a trivial task, since application-dependent factors can dramatically impact the accuracy and average measures can obscure the impact of error events that are highly relevant for assessing the system reliability. In this work, we use two evaluation metrics that attempt to quantify the important performance behavior. The *classification precision* defined as

$$\eta_p = P(\hat{h}_n = h_n) \tag{7}$$

where $P(a = b)$ indicates the probability that $a = b$ is simply the probability that a device is correctly classified. The *worst true positive (TP) rate* defined as

$$\eta_w = \min_{a \in \mathbf{h}} P(\hat{h}_n = a \,|\, h_n = a) \tag{8}$$

measures the probability of correct device identification for the device that is most frequently incorrectly identified. We emphasize that both η_p and η_w depend on the choice of the averaging window size T_w, the training ratio γ, and most significantly the set of radiometric features. Therefore, the impact of these important parameters on identification accuracy are analyzed in detail in Section III.

E. MIMO Radios

Extension of this approach to radios with multiple transmitters is a relatively straightforward task. Specifically, the training and application of the model simply require that the device sends data from each of the transmitting radio-frequency chains sequentially. The radiometric features from each transmitter are then aggregated together to form a larger set of observed features. The entire framework described in this section is, therefore, applicable to MIMO radios without modification.

III. Experimental Results and Performance Analysis

A. Measured Data

The analysis uses data acquired from 21 different devices based on the 802.11n MIMO protocol, including three D-Link DIR-665 Extreme N routers and 18 D-Link DWA-556 PCI NICs. Each device is equipped with three antennas and is configured to transmit a stream of data using all antennas to a nearby MIMO access point. A portion of the signal from each antenna port of the device under test is coupled out using a coaxial tee, and the radiometric features of each coupled signal are measured using an Agilent N9010A EXA Vector Signal Analyzer (VSA) equipped with an 802.11n measurement personality. While such high-accuracy measurements have also been used in prior work [12], it should be pointed out that one area of important research involves the development of techniques for accurately identifying devices based on measurements with less certainty (such as measurements available with a deployed communication receiver).

To allow comparisons between the identification performance using SISO and MIMO signaling, one transmitter from each MIMO device is used in our analysis as a distinct SISO device. Note that we do not use all transmitters from each device as separate SISO devices, as in this case similarity of

features derived from common hardware would lead to unfairly poor performance for the SISO case. Finally, while the radiometric features are averaged over a frame of data, the values are recorded from the VSA at a sample interval of $T_0 = 0.2$ s over a total time of $T_s = 20$ min.

Ideally, the measurements from each transmitter would be conducted after the signal propagates over the wireless channel. This is particularly important for features such as the SYNC correlation or common phase error of the OFDM tones, since the multipath propagation inherent in many typical environments may tend to introduce error in the measurement of these features and therefore make them less reliable for accurate identification. Unfortunately, conducting such a measurement requires the ability to transmit from each radio separately.

Despite this limitation in the measurements, the resulting data remain useful for the purposes of our study. First, if a feature becomes less reliable for identification due to the measurement error created by the multipath propagation, the mRMR algorithm will demote its position in the feature list and possibly preclude its use for identification. The impact of this is studied in Section III-B. Second, a key objective of this study is to determine the benefit offered when identifying MIMO radios, and since the MIMO and SISO measurements are obtained in the same manner, this comparison remains valid. Finally, it is important to recognize that if multipath propagation impacts the reliability of a feature used for identification, then clearly the severity of the multipath becomes a parameter that must be considered as part of the analysis. The measurements performed here remove this as an independent parameter that might impact results. This observation motivates future studies where the impact of propagation on the identification performance is thoroughly studied.

The computations that follow are designed to evaluate the impact of different key parameters on the identification performance. Naturally, when sweeping one parameter, other parameters remain fixed. Therefore, unless otherwise specified, default parameters assume the values $T_w = 20$ s (smoothing window size), $\gamma = 30\%$ (training ratio), and $N_f = 5$ (number of selected features). Except when explicitly stated, performance for MIMO devices uses all three transmitters on the device ($N_{\mathrm{Tx}} = 3$) during training and identification.

B. Feature Selection

The first step in the process is to apply the mRMR feature selection approach to the observed data. This application orders the available features in decreasing significance for identification. Table I lists the top seven features selected by the algorithm in descending order of significance as a function of the number of transmitters N_{Tx} used on each device, where the subscript on a feature indicates the transmitter with which it is associated. We note that the first three features are independent of the number of transmitters and match exactly the observations provided in the original paper on the PARADIS algorithm [12]. This is encouraging, since the original work appears to report this ordering based on empirical observations while our work uses analysis to prove and confirm these observations. We further note that even when multiple transmitters are used, the algorithm selects only a single instance of the frequency offset, likely because in a single device all transmitters use a common local oscillator and, therefore, all streams have identical frequency offset values (meaning that additional frequency offset values are highly redundant with the first). A similar argument appears applicable to the SYNC correlation.

TABLE I
ORDERED SIGNIFICANCE OF THE FEATURES AS A FUNCTION OF THE NUMBER OF TRANSMITTERS N_{Tx} (SUBSCRIPTS INDICATE TRANSMITTER NUMBER)

Rank	Number of Transmitters (N_{Tx})		
	1	2	3
1	Freq Offset	Freq Offset$_1$	Freq Offset$_1$
2	SYNC Corr	SYNC Corr$_1$	SYNC Corr$_1$
3	I/Q Offset	I/Q Offset$_1$	I/Q Offset$_1$
4	EVM	I/Q Offset$_2$	I/Q Offset$_2$
5	Symbol Clock Error	EVM$_1$	I/Q Offset$_3$
6	I/Q Gain Imbalance	SYNC Corr$_2$	EVM$_1$
7	I/Q Rotation	EVM$_2$	EVM$_3$

This prioritization of the available features, however, does not indicate the number of features that should be used. While the prior work indicates that using all available features enhances identification accuracy [12], it is unclear how much incremental advantage each additional feature offers. Therefore, consider a simulation where the number of features used varies over the range $1 \leq N_f \leq 6$, selected in the order of significance as detailed in Table I. For each value of N_f, the identification model is created and applied. Fig. 1 shows the classification precision and the worst TP rate achieved as a function of N_f for SISO and MIMO ($N_{\mathrm{Tx}} = 3$) devices. We first observe in this plot that the performance for MIMO devices matches that achieved for SISO devices for $N_f \leq 3$ since, as detailed in Table I, the features used in the two cases are the same. However, the performance for MIMO devices becomes dramatically superior for $N_f > 3$, particularly when comparing the worst TP rate, meaning that the enriched feature set associated with MIMO transmission enables identification of devices that are more difficult to identify with SISO signaling. With reference to Table I, it is clear that the improved accuracy for MIMO devices stems from the fact that the identification algorithm can take advantage of additional features—in this case mainly the I/Q offset of different transmitters—that offer increased ability to discriminate between devices. Finally, while the improvement obtained using MIMO for identification is significant, it is also important to recognize from these results that using more than five features offers little performance advantage, even for SISO classification. This means that unused features in the measurements either are highly correlated with used features or provide little ability to discriminate between devices.

Table I shows that the SYNC correlation is an important feature for classification, but as mentioned in Section III-A, this feature could become less reliable if multipath propagation is present. To analyze the impact of such a scenario, we assume that the SYNC correlation and OFDM common phase error become too unreliable for identification, and we perform the analysis of Table I and Fig. 1 under this constraint. Under these circumstances, SYNC Corr is removed from its # 2 position in Table I and features # 3–7 move up one position for SISO ($N_{\mathrm{Tx}} = 1$) and MIMO ($N_{\mathrm{Tx}} = 3$). Fig. 2 shows the identification performance as a function of the number of features included. Comparison of these results with those in

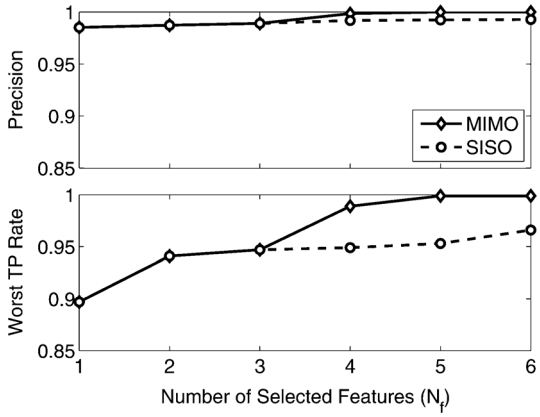

Fig. 1. Identification performance as a function of the number of features N_f for SISO and MIMO devices.

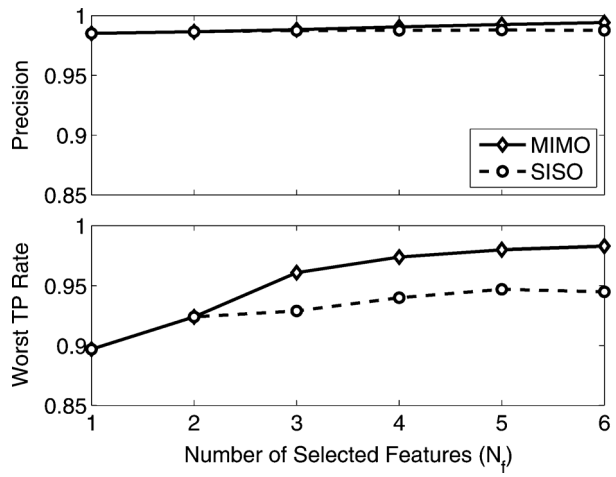

Fig. 2. Identification performance as a function of the number of features N_f for SISO and MIMO devices when SYNC correlation and OFDM common phase error are excluded.

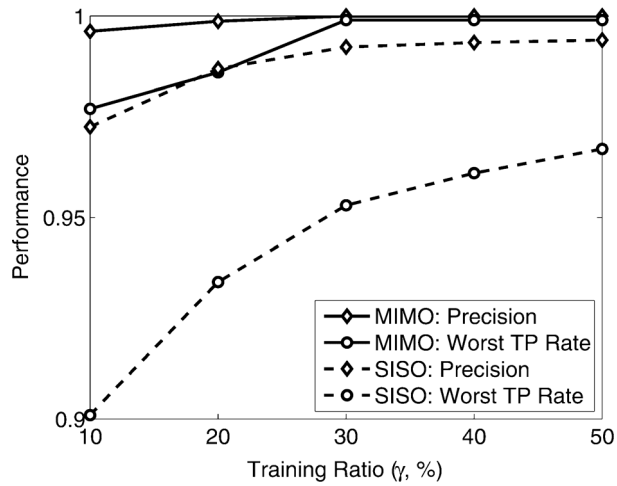

Fig. 3. Identification performance as a function of the training ratio γ for SISO and MIMO devices.

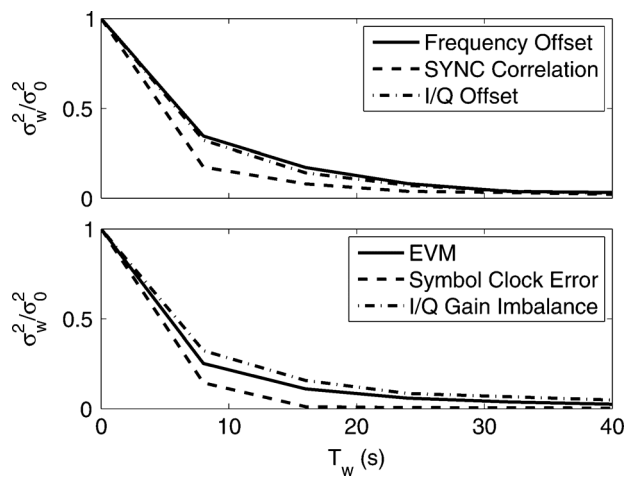

Fig. 4. Normalized change in feature variance as a function of the window size T_w used for data averaging.

Fig. 1 show that loss of an important feature such as SYNC correlation degrades the identification accuracy. However, the results continue to show the improvement in identification enabled by MIMO, particularly for the worst TP rate.

We also recognize that the reliability of a feature impacted by the channel may be improved if that feature is averaged over the MIMO transmitters. Naturally, the effectiveness of such a technique depends on whether or not the observations of the feature over the multiple transmitters are uncorrelated. While some speculation on the effectiveness of such an approach for different features could be entertained, definitive decisions must rely upon observations or careful system analysis, and therefore, such a concept will be reserved for future work where the appropriate measurements are available.

C. Training

Intuitively, increasing the fraction of data used to train the mapping function F should produce a better model that achieves enhanced identification performance. However, training is simplest if it is done with a small amount of data. Therefore, it is interesting to explore the amount of training data required to achieve reliable identification performance. Fig. 3 plots the identification performance as a function of the training ratio γ (indicated as a percent). While the identification of MIMO and SISO devices is influenced by this parameter, clearly the impact is more dramatic for SISO signaling. Our decision to use $\gamma = 30\%$ as the default value of the training ratio stems simply from the observation that little additional benefit is obtained as γ is increased beyond this value.

D. Windowing

Next, we consider the impact of the averaging used to reduce the variance of the data before its use in training or identification. We recall that this averaging is performed using a simple moving average over a window of length T_w, as suggested in (1). We examine the first six features listed under $N_{\mathrm{Tx}} = 1$ in Table I, with the variance of a feature with and without averaging generically represented as σ_w^2 and σ_0^2, respectively. Fig. 4 plots σ_w^2/σ_0^2 as a function of the window size T_w, revealing a dramatic reduction in variance with window size for $T_w \leq 20$ s. Fig. 5 plots the identification performance as a function of the window size. As can be observed, for MIMO devices, the impact of the variance reduction is arguably insignificant, while

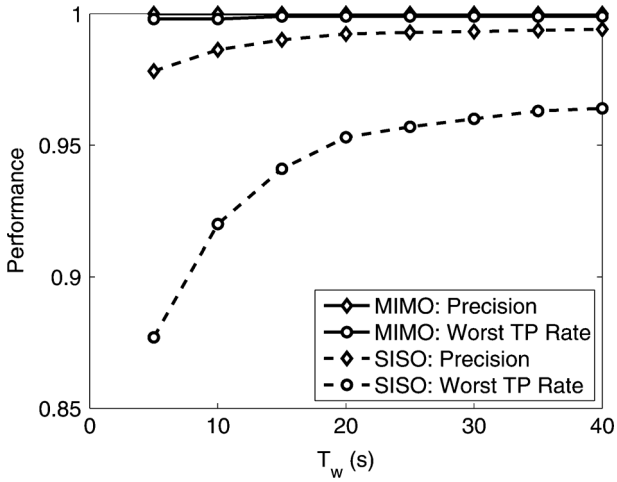

Fig. 5. Identification performance as a function of the window size T_w used for data averaging for SISO and MIMO devices.

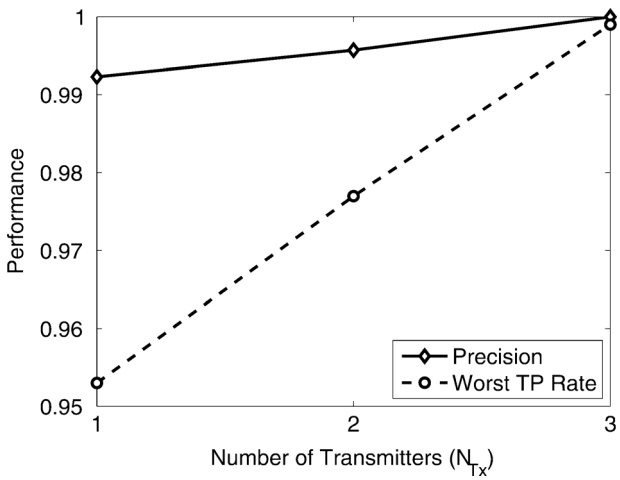

Fig. 6. Identification performance as a function of the number of antennas for SISO and MIMO devices.

for SISO devices a more pronounced trend is observable, with most of the gains for SISO identification achieved using the default value of $T_w = 20$ s.

E. Number of Transmitters

The results so far have assumed that all three transmitters are used for MIMO device identification. Fig. 6 plots the identification performance as a function of the number of transmitters used (N_{Tx}). We emphasize that for each value of N_{Tx}, the actual features used differ, as detailed in Section III-B. Based on what we have observed in our prior analysis, it is not surprising to see that use of additional transmitters improves the identification accuracy, with the most pronounced benefit being the increase in the worst TP rate, indicating once again that exploiting MIMO transmissions enables identification of devices that are difficult to identify when relying only on SISO transmissions.

F. Bagging versus SVM

Given our understanding of the impact of these algorithmic parameters on classification, we now briefly compare the performance of bagging and SVM approaches for identification. In our analysis using $N_{\mathrm{Tx}} = 3$ transmitters, bagging and SVM achieve nearly identical precisions of 99.23% and 99.17%, respectively. However, on our computational platform, bagging and SVM respectively require 11.14 s and 133.73 s to construct the identification model. While in reality model construction is done infrequently, this comparison shows that the bagging algorithm with the DT kernel performs as well as SVM, and offers an advantage of reduced computational complexity and, as suggested in Section II-C, reduced memory requirements.

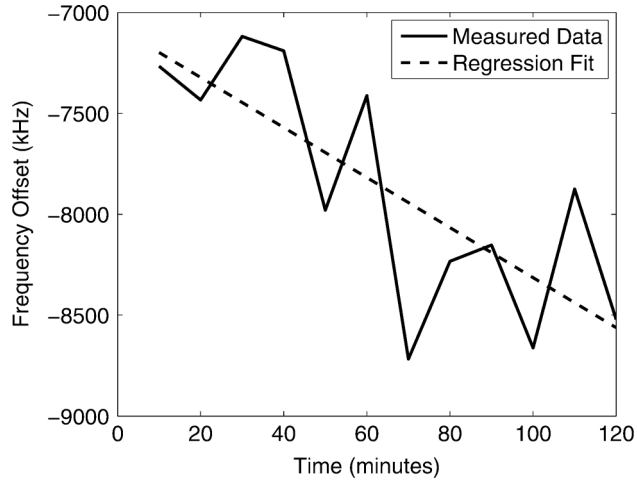

Fig. 7. Drift of the frequency offset as a function of time for a single transmitter as well as a linear regression fit to the observed data.

IV. Time Drift

The preceding analysis demonstrates how use of MIMO transmissions can enhance the accuracy of identification and reveals the impact of different algorithmic parameters on the identification performance. However, what has not been considered is that radiometric features can drift in time based on different environmental conditions (with temperature being a particularly significant factor) [15]. As a result, if the mapping function F is trained using features observed at one time, it may not offer precise identification when applied to data observed at a later time.

As an example of this drift, consider that the transmit frequency of a device generally depends on the frequency of a reference oscillator, often realized using a crystal oscillator and analog components in a loop filter within a phase-locked loop. These components can be particularly sensitive to temperature or other environmental factors, and therefore, this radiometric feature can vary dramatically with time. Fig. 7 plots the change in the frequency offset as a function of time for a single transmit device observed over a two hour period. The linear regression fit to the data shows that over this time period, the frequency offset changes approximately 20%. Such variation clearly has the potential to dramatically influence the identification accuracy.

Arguably, one way to overcome the impact of this feature drift is to ensure that the mapping function F is trained using data over a long time period so that the drift is included in the model. Unfortunately, measuring the time drift of all devices over multiple hours to capture this behavior accurately poses practical difficulties. Specifically, a comprehensive analysis requires that

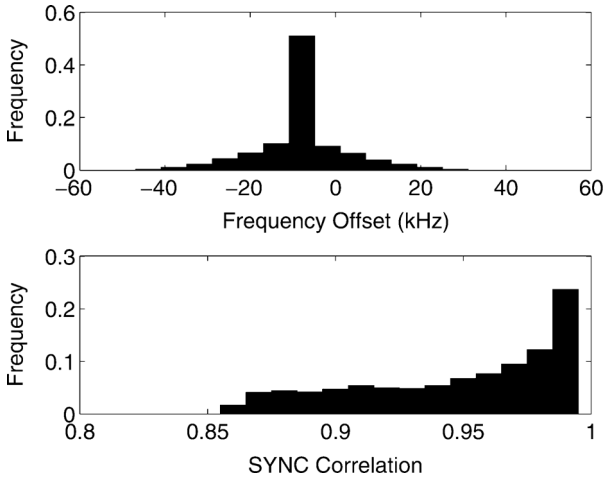

Fig. 8. Normalized histograms of the frequency offset and SYNC correlation from a single transmitter derived from data observed over a window of two hours.

we systematically probe the behaviors of all features over the appropriate range of environmental conditions, determine the statistical distribution of these conditions under normal operating scenarios, and finally apply the identification techniques developed to the observed data. However, based on observations from the data captured for this work, we can gain some insight into the impact of feature drift, with a specific goal of understanding how the mRMR feature selection approach and the enriched feature set associated with MIMO radios can be instrumental in improving identification accuracy under this scenario of drifting features.

As a first step in this analysis, consider the statistical behavior of our three dominant parameters measured over a time period of two hours. We emphasize that during this two-hour time period, the environmental conditions are not explicitly controlled but rather fluctuate with heating/cooling cycles as in a typical office setting, thus making the observations consistent with what might be observed in typical device operation. Fig. 8 shows normalized histograms of the frequency offset and SYNC correlation derived from these measurements. The histogram of the I/Q offset (in decibels), not shown for the sake of brevity, is also a single-mode symmetric distribution, much like that for the frequency offset, but with a smaller variance. These results suggest that the frequency offset and I/Q offset (in decibels) can be modeled approximately as real Gaussian random variables with mean μ and variance σ^2, or $\epsilon \sim N(\mu_\epsilon, \sigma_\epsilon^2)$ for frequency offset and $D \sim N(\mu_D, \sigma_D^2)$ for I/Q offset. The SYNC correlation on the other hand is more accurately modeled using a Beta distribution, or $R \sim \text{Beta}(\alpha_R, \beta_R)$.

Based on this representation, we model the features as random processes that satisfy these distributions, recognizing that the extent of the drift for a feature increases with the variance of the descriptive distribution. Specifically, we generate data for each of the features as a set of random realizations based on the underlying distributions, train the model based on this data, and then apply the model to additional random realizations of the features. This simple approach allows us to explore the impact of feature drift across all devices without explicitly measuring their long-term performance, and by parameterizing the features through their statistical distribution we can use the models to extend the severity of the drift beyond that observed in our particular experimental scenario. To simplify the analysis of the underlying behavior and based on our observation that frequency offset is highly sensitive to environmental conditions, we examine the identification performance as a function of the variance of this feature while leaving its mean as well as the distributions of the remaining features fixed. Furthermore, because we are modeling these three features as random variables, their correlations are important. We have computed the correlation coefficient between the three parameters as: 0.036 for frequency offset and I/Q offset, -0.023 for frequency offset and SYNC correlation, and 0.011 for I/Q offset and SYNC correlation. Because these correlations are low, we treat them as uncorrelated random variables in the simulations.

TABLE II
ORDERED SIGNIFICANCE OF THE FEATURES AS A FUNCTION OF THE RELATIVE INCREASE IN THE VARIANCE OF THE FREQUENCY OFFSET FOR SISO DEVICES

Rank	$\Delta\sigma_\epsilon^2$			
	0%	10%	20%	30%
1	**Freq Offset**	**Freq Offset**	SYNC Corr	SYNC Corr
2	SYNC Corr	SYNC Corr	**Freq Offset**	I/Q Offset
3	I/Q Offset	I/Q Offset	I/Q Offset	**Freq Offset**

To systematically change the variance of the distribution for frequency offset, we define the reference value of σ_ϵ^2 as the value obtained using the short-duration measurements (used for the results in Section III). The variance used for each simulation is then denoted as $\overline{\sigma}_\epsilon^2$, with the relative change in the variance given by

$$\Delta\sigma_\epsilon^2 = \frac{\bar{\sigma}_\epsilon^2}{\sigma_\epsilon^2} - 1. \tag{9}$$

As a first study, consider application of the mRMR approach to determine the significance of each feature for device identification. Table II lists the ordered significance of each of the first three features as a function of the relative change in the variance $\Delta\sigma_\epsilon^2$ for SISO devices. The frequency offset is printed in bold text to enable visual tracking of its change in significance. As can be seen, as the variance in the frequency offset increases, it loses its position as the primary feature for reliable device identification, mainly because it loses relevance (correlation with the device identifier).

Fig. 9 plots the identification performance achieved using this simulation as a function of $\Delta\sigma_\epsilon^2$, where the $N_f = 5$ features used are again prioritized using the mRMR algorithm for each value of $\Delta\sigma_\epsilon^2$. These results are plotted over a limited range of $\Delta\sigma_\epsilon^2$ to clearly show the behavior. The results show that for SISO devices, there is a dramatic reduction in the identification reliability even for a modest increase in the variance. Fig. 10 plots the results over a much wider range of variance values, revealing that as the variance becomes large, the MIMO identification performance also degrades. The vertical dashed line in Fig. 10 shows the value of $\Delta\sigma_\epsilon^2$ for the variance observed

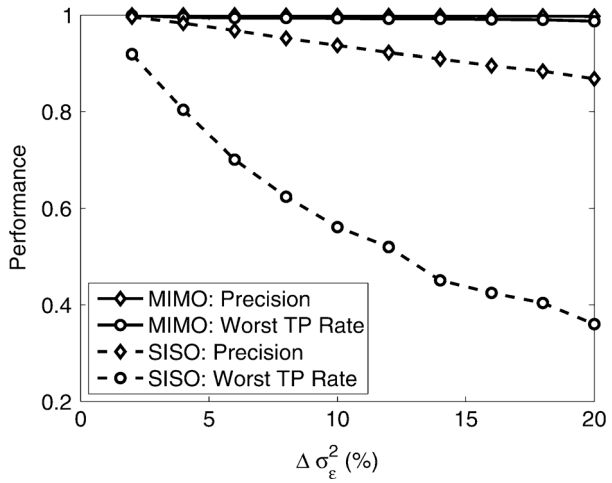

Fig. 9. Identification performance as a function of the relative change $\Delta\sigma_\epsilon^2$ in the variance of the frequency offset for SISO and MIMO devices.

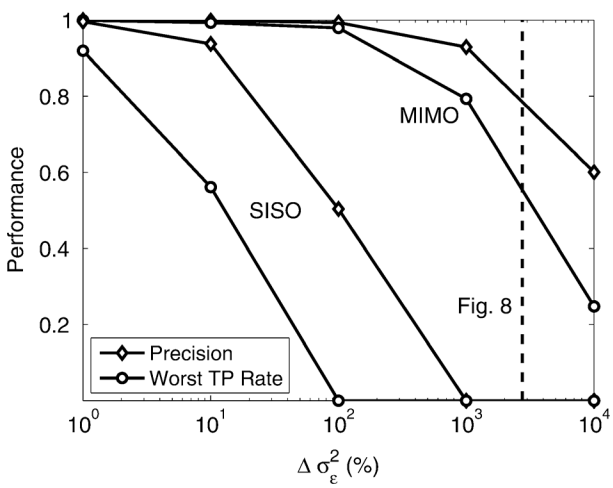

Fig. 10. Identification performance as a function of the relative change $\Delta\sigma_\epsilon^2$ in the variance of the frequency offset for SISO and MIMO devices.

over the two-hour observation period represented by the frequency offset histogram in Fig. 8. These results clearly demonstrate the necessity of thoroughly understanding the behavior of the features before hastily determining the best feature set and time window for training the model. They also show that while identification performance suffers with feature variance, MIMO classification provides increased robustness due to the enriched feature set.

We reiterate that this simple analysis is preliminary, and much more work should be performed on the impact of feature drift on identification accuracy. Prior work on device identification based on radio transients has shown that identification accuracy can be dramatically improved by ensuring that the classifier is trained at approximately the same signal-to-noise ratio (SNR) as the observed waveform [26]. It is perhaps possible to extend this idea to the present work by using an environmentally sensitive feature such as frequency offset to provide an indication of the environmental conditions and then training the classifier to use this information to assist in identification. Alternatively, by correcting the data based on observations of this feature, perhaps other important features could be made more reliable. Such approaches could for the basis for further studies on this topic.

V. Conclusion

This paper demonstrates identification of devices based on their radiometric features in the modulation domain, with specific emphasis on the performance improvements achievable for MIMO devices. The technique first uses an information theoretic approach to identify the radiometric features that are most effective at discriminating between different devices. The discussion then turns to use of a nonlinear model constructed based on training data that can uniquely identify each device based on observations of the selected radiometric features. Parametric studies reveal that while the technique is effective at identifying both MIMO and SISO devices, MIMO identification performance is less sensitive to specific parameters used in the model construction. Furthermore, an analysis of the impact of feature drift with time reveals that MIMO device identification demonstrates increased resilience to this feature variation.

References

[1] W. Stallings, *Cryptography and Network Security: Principles and Practice*, 3rd ed. Englewood Cliffs, NJ: Prentice-Hall, 2003.

[2] C. Chen, "Secret Key Establishment Using Wireless Channels as Common Randomness in Time-Variant MIMO Systems" Ph.D. dissertation, Brigham Young University, Provo, UT, 2010 [Online]. Available: http://contentdm.lib.byu.edu/u?/ETD,2083

[3] M. Barbeau and J.-M. Robert, "Perfect identity concealment in UMTS over radio access links," in *Proc. IEEE Int. Conf. Wireless and Mobile Computing, Networking and Communications (WiMob)*, Montreal, Canada, Aug. 22–24, 2005, vol. 2, pp. 72–74.

[4] M. Barbeau, J. Hall, and E. Kranakis, "Detecting impersonation attacks in future wireless and mobile networks," in *Proc. Mobile Ad-hoc Networks and Sensors (MADNES)—Workshop on Secure Mobile Ad-hoc Networks and Sensors (Singapore, Sep. 20–22, 2005)*, 2006, vol. 4074, pp. 80–95, Springer Lecture Notes on Computer Science.

[5] N. Patwari and S. K. Kasera, "Temporal link signature measurements for location distinction," *IEEE Trans. Mobile Comput.* vol. 10, no. 3, pp. 449–462, Mar. 2011.

[6] D. B. Faria and D. R. Cheriton, "Detecting identity-based attacks in wireless networks using signalprints," in *Proc. ACM Workshop on Wireless Security (WiSe)*, Los Angeles, CA, Sep. 29, 2006, pp. 43–52.

[7] H. C. Choe, C. E. Poole, A. M. Yu, and H. H. Szu, "Novel identification of intercepted signals from unknown radio transmitters," in *Proc. SPIE, Wavelet Applications II*, H. H. Szu, Ed., 1995, vol. 2491, pp. 504–517.

[8] J. Hall, M. Barbeau, and E. Kranakis, "Enhancing intrusion detection in wireless networks using radio frequency fingerprinting," in *Proc. 3rd IASTED Int. Conf. Communications, Internet and Information Technology (CIIT)*, St. Thomas, U.S. Virgin Islands, Nov. 22–24, 2004, pp. 201–206.

[9] K. B. Rasmussen and S. Capkun, "Implications of radio fingerprinting on the security of sensor networks," in *Proc. IEEE Third Int. Conf. Security and Privacy in Communications Networks (SecureComm)*, Nice, France, Sep. 17–21, 2007, pp. 331–340.

[10] O. Ureten and N. Serinken, "Wireless security through RF fingerprinting," *Can. J. Elect. Comput. Eng.*, vol. 32, no. 1, pp. 27–33, 2007.

[11] J. Franklin, D. McCoy, P. Tabriz, V. Neagoe, J. Van Randwyk, and D. Sicker, "Passive data link layer 802.11 wireless device driver fingerprinting," in *Proc. 15th USENIX Security Symp.*, Vancouver, B.C., Canada, Jul. 31–Aug. 4 2006, pp. 167–178.

[12] V. Brik, S. Banerjee, M. Gruteser, and S. Oh, "Wireless device identification with radiometric signatures," in *Proc. 14th ACM Int. Conf. Mobile Computing and Networking (Mobicom)*, San Francisco, CA, Sep. 14–19, 2008.

[13] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.

[14] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," in *Proc. IEEE Bioinformatics Conf. (CSB2003)*, Stanford, CA, Aug. 11–14, 2003, pp. 523–528.

[15] Ö. H. Takbaş, N. Serinken, and O. Üreten, "An experimental performance evaluation of a novel radio-transmitter identification system under diverse environmental conditions," *Can. J. Elect. Comput. Eng.*, vol. 29, no. 3, pp. 203–209, Jul. 2004.
[16] R. E. Bellman, *Adaptive Control Processes: A Guided Tour*. Princeton, NJ: Princeton Univ. Press, 1961.
[17] R. E. Bellman, *Dynamic Programming*. Mineola, NY: Dover, 2003.
[18] I. V. Tetko, D. J. Livingstone, and A. I. Luik, "Neural network studies. 1. Comparison of overfitting and overtraining," *J. Chem. Inf. Comput. Sci.*, vol. 35, pp. 826–833, 1995.
[19] P. L. Bartlett, "For valid generalization, the size of the weights is more important than the size of the network," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, Denver, CO, Dec. 2–5, 1996, vol. 9, pp. 134–140.
[20] R.-H. Li and G. G. Belford, "Instability of decision tree classification algorithms," in *Proc. Eighth ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, Edmonton, Canada, Jul. 23–26, 2002, pp. 570–575.
[21] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, Aug. 1996.
[22] J. R. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann, 1993.
[23] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: An update," *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, pp. 10–18, Jun. 2009.
[24] H. Blockeel and J. Struyf, "Efficient algorithms for decision tree cross-validation," *J. Mach. Learning Res.*, vol. 3, pp. 621–650, Dec. 2002.
[25] J. C. Platt, Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines Microsoft Research, Tech. Rep. MSR-TR-98-14, Apr. 21, 1998.
[26] Ö. H. Takbaş, O. Üreten, and N. Serinken, "Improvement of transmitter identification system for low SNR transients," *Electron. Lett.*, vol. 40, no. 3, pp. 182–183, Feb. 5, 2004.

Yan Shi (S'10) received the B.S. degree in electrical engineering from Shanghai Jiao Tong University, in 2007, and the Ph.D. degree in electrical engineering from Brigham Young University, Provo, UT, in 2011.

His research interests focus on robust signaling, scheduling, and authentication in the multiuser MIMO channel.

Michael A. Jensen (S'93–M'95–SM'01–F'08) received the B.S. and M.S. degrees from Brigham Young University (BYU), Provo, UT, in 1990 and 1991, respectively, and the Ph.D. degree from the University of California, Los Angeles (UCLA), in 1994, all in electrical engineering.

Since 1994, he has been at the Electrical and Computer Engineering Department, BYU, where he is currently a Professor and Department Chair. His research interests include antennas and propagation for communications, microwave circuit design, and multiantenna signal processing.

Dr. Jensen is currently the Editor-in-Chief of the IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION. Previously, he was an Associate Editor for the IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION and the IEEE ANTENNAS AND WIRELESS PROPAGATION LETTERS. He has been a member and Chair of the Joint Meetings Committee for the IEEE Antennas and Propagation Society, a member of the society AdCom, and Cochair and Technical Program Chair for five society-sponsored symposia. In 2002, he received the Harold A. Wheeler Applications Prize Paper Award in the IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION in recognition of his research on multiantenna communication.