4G Security Using Physical Layer RF-DNA with DE-Optimized LFS Classification

Paul Harmer[1], Michael Temple[1], Mark Buckner[2], and Ethan Farquhar[2]
Air Force Institute of Technology[1], Oak Ridge National Laboratory[2]
Email: paul.harmer@afit.edu

***Abstract*—Wireless communication networks remain under attack with ill-intentioned "hackers" routinely gaining unauthorized access through Wireless Access Points (WAPs)–one of the most vulnerable points in an information technology system. The goal here is to demonstrate the feasibility of using Radio Frequency (RF) air monitoring to augment conventional bit-level security at WAPs. The specific networks of interest are those based on Orthogonal Frequency Division Multiplexing (OFDM), to include 802.11a/g WiFi and 4G 802.16 WiMAX. Proof-of-concept results are presented to demonstrate the effectiveness of a "Learning from Signals" (LFS) classifier with Gaussian kernel bandwidth parameters optimally determined through Differential Evolution (DE). The resultant DE-optimized LFS classifier is implemented within an RF "Distinct Native Attribute" (RF-DNA) fingerprinting process using both Time Domain (TD) and Spectral Domain (SD) input features. The RF-DNA is used for intra-manufacturer (like-model devices from a given manufacturer) discrimination of IEEE compliant 802.11a WiFi devices and 802.16e WiMAX devices. A comparative performance assessment is provided using results from the proposed DE-optimized LFS classifier and a Bayesian-based Multiple Discriminant Analysis/Maximum Likelihood (MDA/ML) classifier as used in previous demonstrations. The assessment is performed using identical TD and SD fingerprint features for both classifiers. Finally, the impact of Gaussian, triangular, and uniform kernel functions on classifier performance is demonstrated. Preliminary results of the DE-optimized classifier are very promising, with correct classification improvement of 15% to 40% realized over the range of signal to noise ratios considered.**

***Index Terms*—Wireless, Security, Fingerprinting, Differential Evolution, Genetic, Algorithm, 4G, 802.16, WiMAX, 802.11, WiFi, Learning from Signals**

I. INTRODUCTION

As fourth generation (4G) communication systems such as last mile Worldwide Interoperability for Microwave Access (WiMAX) and Long Term Evolution (LTE) systems evolve, so does consumer exposure and risk of attack. The relative ease by which ill-intentioned "hackers" access these systems is enabled by a couple factors, including 1) the availability of relatively inexpensive, high power hacking equipment (workstations, servers, etc.) and 2) the fact that these systems fundamentally operate through Wireless Access Point (WAPs) which are readily accessible and remain one of the most vulnerable points in an Information Technology (IT) network [1].

Manuscript received August 15, 2011; revised October 17, 2011; accepted November 8, 2011.

Network WAP vulnerability has been traditionally addressed through bit-level security mechanisms in upper Open Systems Interconnection (OSI) layers with the majority of intrusion detection systems operate at OSI Layer #3, the Network layer, or higher [2]. While providing some measure of security, these methods generally ignore potentially useful information that is in device Radio Frequency (RF) emissions. Thus, potential security benefits available within the lower OSI Physical (PHY) layer remains largely unexploited.

The task at hand is to exploit PHY information and improve 4G communication security by providing more robust device authentication for mitigating unauthorized system access. The envisioned implementation is to augment bit-level protection mechanisms using RF air monitoring devices located at network access points [3]–[8]. Given the computational power required for air monitoring, typical WAP locations seem ideal as they tend to have the necessary resources (physical space, prime power, etc.) available. This application was targeted in [4], [5] for GSM signals and is thought to be directly applicable for similarly configured WiMAX and LTE systems.

Earlier related works demonstrated that RF "Distinct Native Attribute" (RF-DNA) features, as identified using various terminology, are indeed useful for discriminating between specific wireless devices [3]–[13]. The effectiveness of RF fingerprinting has already drawn the attention of counter-measure researchers who are taking the next step of assessing RF PHY layer security robustness [14]. But, as typically expected with RF signal processing techniques, overall device classification performance with RF-DNA fingerprints decreases as Signal-to-Noise Ratio (SNR) decreases. This is commonly addressed by either finding 1) more robust input features for a given classifier, or 2) a more robust classifier for given input features.

The second of these approaches is considered here using Time Domain (TD) and Spectral Domain (SD) signal features that have been successfully exploited in previous work [3]–[8], [15]. Given these features, the goal is to demonstrate a more powerful "classification engine" that is optimized through Differential Evolution (DE). Success of the resultant DE-optimized "Learning from Signals" (LFS) classifier is measured as either 1) improving device classification accuracy for a given SNR, or by 2) maintaining a given classification accuracy at a lower SNR.

doi:10.4304/jcm.6.9.671-681

DE-optimized LFS classifier capability is demonstrated through comparative assessment of its classification accuracy with that of a Bayesian-based Multiple Discriminant Analysis/Maximum Likelihood (MDA/ML) classifier. Assessment reliability is ensured by inputting *identical* TD and SD fingerprint features into the classifiers. The features are extracted from experimentally collected 802.11a WiFi and 802.16e WiMax signals under *intra-manufacturer* conditions (same manufacturer and model, different serial numbers). Relative to *inter-manufacturer* conditions (inter-operable devices from different manufacturers), *intra-manufacturer* classification poses the greatest classification challenge [3]–[6] while presenting the greatest opportunity for technical contribution.

The remainder of the paper includes: Sect. II *Technical Background* on key technical aspects; Sect. III *Comparative Assessment Methodology* used to obtain results and conduct analysis; Sect. IV *Results* of classification performance; and Sect. V *Summary and Conclusions* for accomplishments.

II. Technical Background

The following subsections provide a summary of key technical concepts that were employed in the methodology to generate desired results. This includes a discussion of RF-DNA Fingerprinting in Sect. II-A and DE-optimized LFS Implementation in Sect. II-B.

A. RF-DNA Fingerprinting

RF-DNA fingerprinting is a PHY technique for discriminating devices based on their inherent emission differences. It has been shown that specific serial-numbered devices possess unique characteristics that result from minute differences in manufacturing (part type, part lot number, assembly processes, etc.). The goal is to capture these differences in fingerprints that can be used to uniquely identify, by serial number, hardware devices as an aid to network security and user authentication. Various RF fingerprinting techniques have been demonstrated previously using various communication signals, including: 802.11 WiFi signals [3], [7], [8], [16]–[19], GSM cell phone signals [5], [15], 802.16 WiMAX signals [6], [8], 802.15 Bluetooth signals [9], and RFID signals [12], [20].

While these earlier cited works have considered several diverse methods for implementing RF fingerprinting, the techniques generally share some common functionality, including: 1) Signal Collection and Post-Collection Processing, 2) Fingerprint Feature Generation, and 3) Signal/Device Classification. Based on processes in [3]–[5], these functions are collectively embodied in the RF-DNA fingerprinting process overview shown in Fig. 1 and described in the following subsections. This approach was adopted here to facilitate direct comparison of previous MDA/ML classification results with new DE-optimized LFS results to assess the impact of introducing an alternate "Signal Classification Engine."

1) Signal Collection and Post-Collection Processing: The first step includes signal reception, digitization, and post-collection processing to prepare the TD signal response for feature extraction. Relative to the process overview in Fig. 1, this includes all processes up to the point where the desired analysis SNR (SNR_A) is established and the analysis signal is passed on for statistical fingerprint generation.

All signals here were collected using an RF Signal Intercept and Collection System (RFSICS) based on Agilent's E3238 system [21]. The devices under test were isolated from the RFSICS to minimize the introduction of unrepeatable environmental and interference effects. This is achieved by placing 1) some devices in an RF anechoic chamber, 2) some devices in separate rooms, 3) some RF absorbing material in strategic locations, and/or 4) combinations thereof. Data transfer is easily accomplished using the conventional File Transfer Protocol (FTP) to pass files between devices. When possible, device transmit powers are controlled to enable association of collected data with specific transmitting devices.

Accounting for all collection factors, the post-filtered SNR for signals collected under controlled conditions is on the order of $SNR_C \in [30, 40]$ dB. This enables direct scaling (G_N in Fig. 1) and addition of like-filtered Additive White Gaussian Noise (AWGN) to generate analysis signals at the desired SNR_A. The resultant analysis signals are used for RF-DNA fingerprinting and device classification

2) Fingerprint Feature Generation: For fingerprint feature generation the classifier input features are generated directly from either the TD signal response or within an alternate feature domain through transformation, e.g., to the frequency domain via a Discrete Fourier Transform (DFT). Transformation is used to exploit additional discriminating features that may be present in an alternate domain. This work considers the collected TD response and DFT-based SD response. In both cases, the final classification features are generated by calculating statistical metrics over selected response regions.

Three signal responses are used for TD fingerprinting, including instantaneous amplitude, phase and frequency (N_{SR}=3). For SD fingerprinting only the normalized Power Spectral Density (PSD) response is used (N_{SR}=1). In both cases, the response(s) is parsed into N_R equal length subregions as illustrated in Fig. 2 for representative TD and SD responses of an 802.11a WiFi signal. Features for the entire signal response are included as well, yielding a total number of feature regions of N_R^F=$(N_R + 1)$. Each of the N_{SR} signal responses are characterized using N_{SM} statistical measures of standard deviation (σ), variance (σ^2), skewness (γ), and/or kurtosis (κ). These statistics are used to form the i^{th} *regional fingerprint* given by

$$F_{R_i} = [\sigma_{R_i}\ \sigma^2_{R_i}\ \gamma_{R_i}\ \kappa_{R_i}]_{1\times(N_{SM}\times N_{SR})}\ , \qquad (1)$$

where $i = 1, 2, ..., N_R^F$ and only selected σ, σ^2, γ, and κ elements are included. The fingerprints from each region

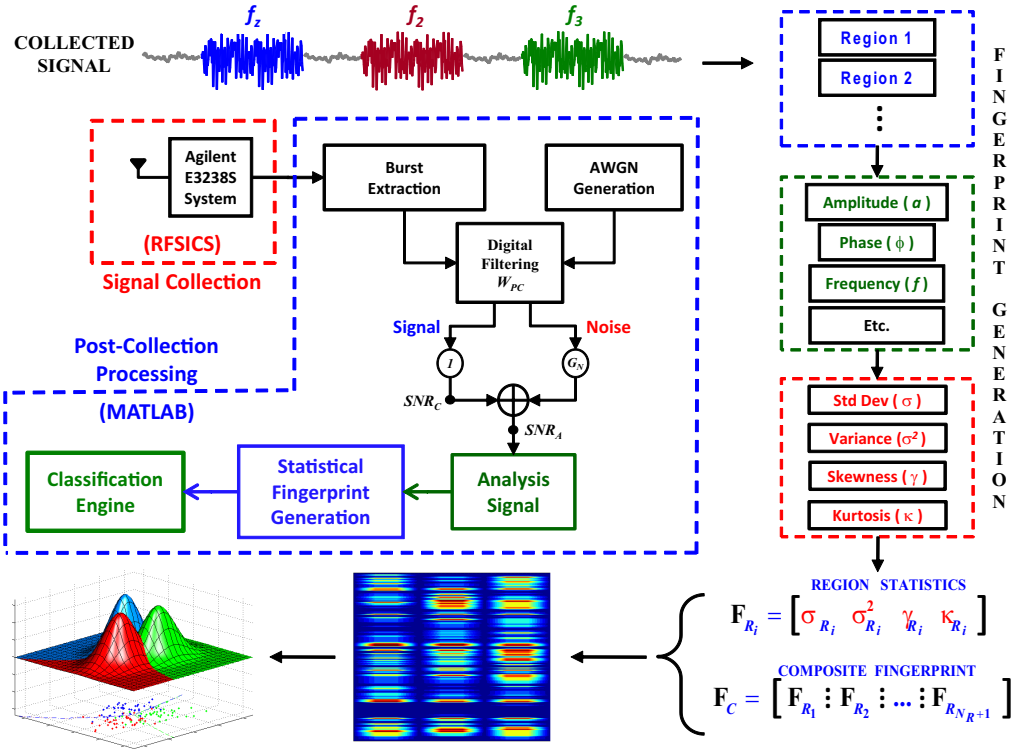


Fig. 1. RF-DNA Fingerprinting Process [8].

per (1) are concatenated to form the *composite statistical fingerprint* given by

$$\mathbf{F}_C = \left[F_{R_1} \vdots F_{R_2} \vdots F_{R_3} \ \ldots \ F_{R_{(N_R+1)}} \right]_{1 \times N_F} , \quad (2)$$

where N_F is the total number of fingerprint features (dimensions) input to the classifier and given by

$$N_F = N_{SM} \times N_{SR} \times (N_R + 1) \ . \quad (3)$$

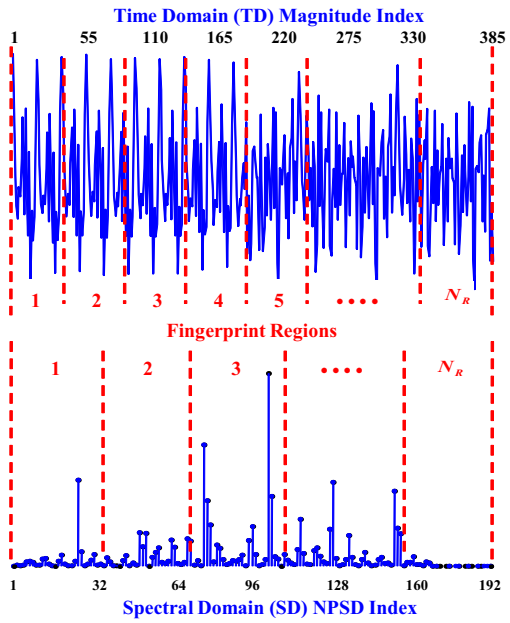


Fig. 2. Representative 802.11a preamble fingerprint feature regions: TD Amplitude Response and Corresponding SD Response Based on NPSD [6].

3) Signal/Device Classification: For signal/device classification a given classifier is implemented to separate and identify N_D devices (input classes) using selected input features. Classification approaches vary across the pattern recognition community and generally include methods based on cross-correlation, vector distance measures, k-nearest neighbor metrics, support vector machines, and Fisher-based MDA/ML processing [3], [6], [9], [20], [22].

As adopted from [3], [5], [15] and used here, the MDA/ML classifier is an extension of Fisher's Linear Discriminant that is used when more than two input devices are to be classified. MDA uses a projection matrix ($\mathbf{W}$) to reduce the input dimensionality. The MDA/ML process is that of finding $\mathbf{W}$ such that projected inter-class separation is maximized and intra-class spread is minimized [23]. Given N_D devices (input classes), the MDA/ML process projects the input features into an $N_D - 1$ decision space.

Device classification is performed using a ML classifier derived from Bayesian Decision Theory, with the multi-dimensional input data classified as being affiliated with one of N_D possible classes. A Bayesian-based decision uses known prior probabilities, probability densities, and relevant costs associated with making a decision. The decision process relies on an accurate representation of the class distribution and its parameters in order to define the likelihood. A sample is assigned the class label of the

class likelihood yielding the maximum response. In this demonstration, the ML prior probabilities are assumed to be equal and the costs uniform. This is best visualized for the $N_D = 3$ class problem as illustrated in the lower left-hand corner of Fig. 1 that shows Gaussian class likelihood functions and the resultant 2-dimensional decision space (lower surface) with ML boundaries.

B. LFS Classification

LFS classification is an adaption of Learning From Data (LFD) techniques where the input training data is derived from samples of a given sensor response [7], [8], [24], [25]. LFD is an algorithm that approximates an unknown relationship between a system's inputs and outputs using known available data. Most scientists and engineers are familiar with this as a form of regression, e.g., a least squares fit using polynomial models. But, LFD modeling is not constrained to using polynomial functions. Once a model of the data is "learned," the model can be applied to previously unseen data to provide an approximation of the modeled system's output. The goal is to find useful information in the input data and exploit that information when acting on future observed data [25].

The LFD concept functionally includes three steps: 1) preprocessing (transformation to feature space), 2) learning or training, and 3) operation or classification. The learned model can be applied to accomplish three basic tasks: classification, regression, or probability density estimation. Classification is estimation of class association based on modeled decision boundaries. This is used in pattern recognition systems and is of greatest utility for RF-DNA fingerprinting. Using N_F input features, the device classification goal is to find a mapping from input sample $\mathbf{x_i} = (x_1, \ldots, x_{N_F})$ to one of N_D devices (classes) where $D \in \{D_1, D_2, \ldots, D_{N_D}\}$. This final classification decision is based on a set of learned boundaries or threshold values, $\boldsymbol{t} = (t_1,\ t_2,\ \ldots,\ t_{N_D-1})$. Once established, this mapping function provides the decision rule by which subsequent operation/classification decisions are made for future samples.

LFD problems are inherently ill-posed given there are more unknowns than available data to describe them. Therefore, there is no unique solution to, or single model of, the system under consideration. In such cases, a search or optimization approach is required to minimize some predefined error function to find the "best" solution among possible solutions. Mean Square Error (MSE) is a commonly used error metric and is adopted here because the training set includes both input signals and associated known class membership.

Many LFD approaches include parameters on the search and fitness functions. These parameters are usually set to common, or default values. However, the defaults may not be the optimum for a specific set of data or a given problem domain. It has been shown that a Genetic Algorithm (GA) can be used to improve LFD modeling. The concept is to improve the regression process using a GA to optimize the regression parameters for each input dimension, rather than using a single, global value for all dimensions. This GA-optimized approach has been applied using the more powerful kernel regression (KR) technique [24], [26] and is adopted here for LFS classification.

C. Gaussian KR Processing

KR is a memory-based technique that stores past input data and processes them when a new data point is observed. So, instead of modeling the entire input/output set (x, y) with a model, as in conventional polynomial regression, the local KR function is estimated over the entire input domain by fitting a simple model at every new sample, or query point $\mathbf{q} = (q_1, q_2, \ldots, q_{N_F})$. The local models are built using a distance weighting kernel function, $K(d^2(\mathbf{x_i}, \mathbf{q}))$, that assigns a weight based on the distance between $\mathbf{x_i}$ and $\mathbf{q}$. Because of this weighting, only observations that are close to the query point are used to fit the model. Any kernel function can be used for KR provided the following properties are satisfied [25]:

1) $K(x_i, q) \geq 0$ (non-negative)
2) $K(\parallel x_i, q \parallel)$ is radially symmetric
3) $K(x_i, q)$ is maximum for $q = x_i$
4) $K(x_i, q)$ decreases monotonically with $|x_i - q|$

While there are virtually an unlimited number of possible functions that satisfy the noted properties, a Gaussian kernel is most commonly used. Two additional kernel functions are also of interest, both of which limit the influence of query point neighbors during model development. The three kernels considered here are shown in Fig. 3 and include the Gaussian (4), triangular (5), and uniform (6) functions [27].

$$K\left(d^2_{\mathbf{H}}(\mathbf{x_i}, \mathbf{q})\right) = \exp^{-0.5 \cdot d^2_{\mathbf{H}}(\mathbf{x_i}, \mathbf{q})} \ , \tag{4}$$

$$K\left(d^2_{\mathbf{H}}(\mathbf{x_i}, \mathbf{q})\right) = 1 - \sqrt{d^2_{\mathbf{H}}(\mathbf{x_i}, \mathbf{q})} \ , \tag{5}$$

$$K\left(d^2_{\mathbf{H}}(\mathbf{x_i}, \mathbf{q})\right) = 1 \ , \quad d^2_{\mathbf{H}}(\mathbf{x_i}, \mathbf{q}) \leq 1 \ . \tag{6}$$

These kernel function have a spread, or bandwidth, that relates to feature cluster size within in a specific dimension. For this work, h_i represents the bandwidth parameter for the i^{th} dimension of a multidimensional kernel function where $\mathbf{H}$ is an $N_F \times N_F$ diagonal matrix. To reduce computational complexity, the inter-dimensional cross-correlations are not considered. Therefore, all off-diagonal elements in $\mathbf{H}$ are zero and $\mathbf{H}$ is given by

$$\mathbf{H} = diag\left(h_1, h_2, \ldots, h_{N_F}\right) \ , \ h_i \geq 0 \ \forall \ i \ . \tag{7}$$

Distance function $d^2(\mathbf{x_i}, \mathbf{q})$ defines the neighborhood of points surrounding $\mathbf{q}$, which is implemented here as the squared Euclidean distance parameterized by $\mathbf{H}$

$$d^2_{\mathbf{H}}(\mathbf{x_i}, \mathbf{q}) = (\mathbf{x_i} - \mathbf{q})^T \mathbf{H}^{-1} (\mathbf{x_i} - \mathbf{q}) \ . \tag{8}$$

Finally, the kernel regression estimate, $\widehat{y}$, for a previously unseen system input, or query point, $\mathbf{q}$, is determined by summing the model contributions for each of

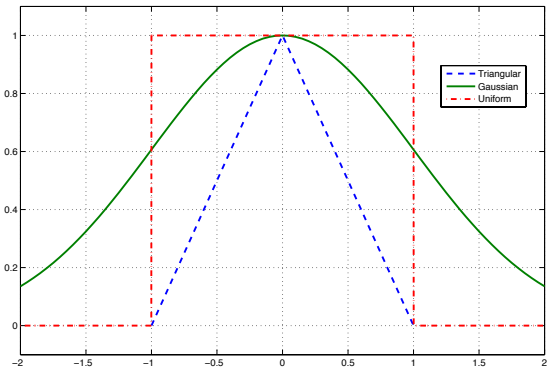


Fig. 3. Relative normalized magnitude of triangular, Gaussian, and uniform kernel functions

N_B signal bursts for each of N_D devices as given by

$$\widehat{y} = \frac{\sum_{i=1}^{N_S} K(d^2_{\mathbf{H}}(\mathbf{x_i}, \mathbf{q})) \cdot y_i}{\sum_{i=1}^{N_S} K(d^2_{\mathbf{H}}(\mathbf{x_i}, \mathbf{q}))}, \quad N_S = N_D \cdot N_B \ . \tag{9}$$

For conventional KR processing, a given bandwidth of $h \in \Re$ would be used for all input dimensions–elements of $\mathbf{H}$ in (7) are *identical*. The approach here differs in that DE KR optimization, as demonstrated in [24], [26], is able to "learn" the best bandwidth parameter h_i to use for each feature dimension and improve LFS classifier performance. A by product of this process is that one can also infer the relative importance of a given dimension based on the magnitude of h_i. A "smaller" h_i indicates greater importance as relevant features in that dimension cluster closer together.

D. DE-Optimized LFS Implementation

The DE-optimized LFS classifier is illustrated in Figure 4 and functionally includes three processes: Input Feature Formatting, DE Optimized KR, and Device Classification.

1) Input Feature Formatting: Given N_D devices to be classified, the classifier input data includes N_B fingerprint vectors per device with each fingerprint containing N_F features (dimensions). Specific details for the RF-DNA fingerprints used here are provided in Section II-A.

2) DE Optimized Kernel Regression: DE is a form of GA processing that performs a population-based global search to optimize a given objective function. With any GA, a group of solutions are retained in the current population which is iteratively updated until specific termination criteria are satisfied. Upon termination, the population member with the best fitness is the one that best optimizes the objective function and it is selected as the solution. The uniqueness of DE lies in its use of real-valued genes within the population members and vector-based operations to produce future generations [24], [28]. The ability to operate on real-valued vectors makes DE

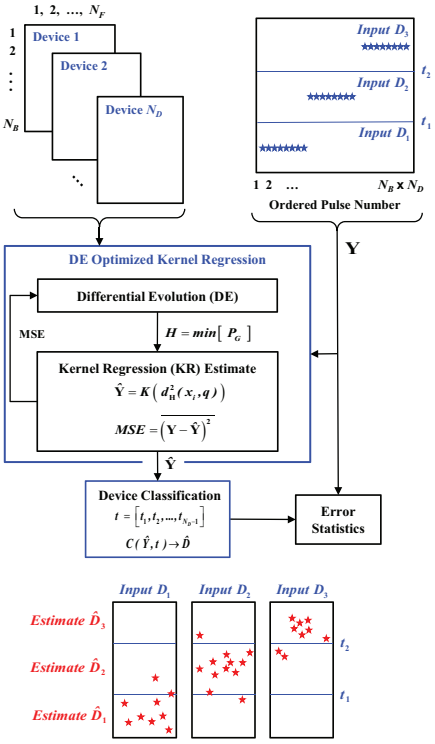


Fig. 4. DE-optimized LFS classification process [7], [8].

ideally suited for operation with our linear algebra kernel regression process.

The DE process for this work was implemented as detailed in Figure 5. The initial population P_{Gen} for $Gen\# = 1$ contains N_P randomly generated members. Each member is represented by a vector of h_i, $i = 1, 2, ..., N_F$, kernel bandwidth values. The initial population is evaluated for fitness using KR and the MSE calculated for each member. Termination criteria can be based on reaching either 1) a maximum number of generations N_{Gen} and/or, 2) a minimum specified MSE Value To Reach (VTR). If not satisfied during the current generation, vector-based mutation occurs as illustrated in Figure 6.

There are many variants in how the vector-based mutation and crossover can be implemented. To describe these there exists a nomenclature of four positions, DE/X/Y/Z. The one chosen here is the DE/current-to-rand/1/bin. This indicates that DE is used, children in the population are created by mutating a current parent population member with a linear combination of random mates. The result is a single child vector. Finally, binary crossover between the child and parent is allowed to occur [29].

In this case, each population member (parent) X_i is mutated using the vector values in three other randomly selected individuals (mates) (V_1, V_2 and V_3). As shown in Figure 6, the child's final value u_i in the j^{th} feature dimension is a linear combination of weighted parent and mate differences using crossover multipliers of F_1 and F_2. Finally, if a crossover threshold CR probability is exceeded for each position in the vector, a binomial

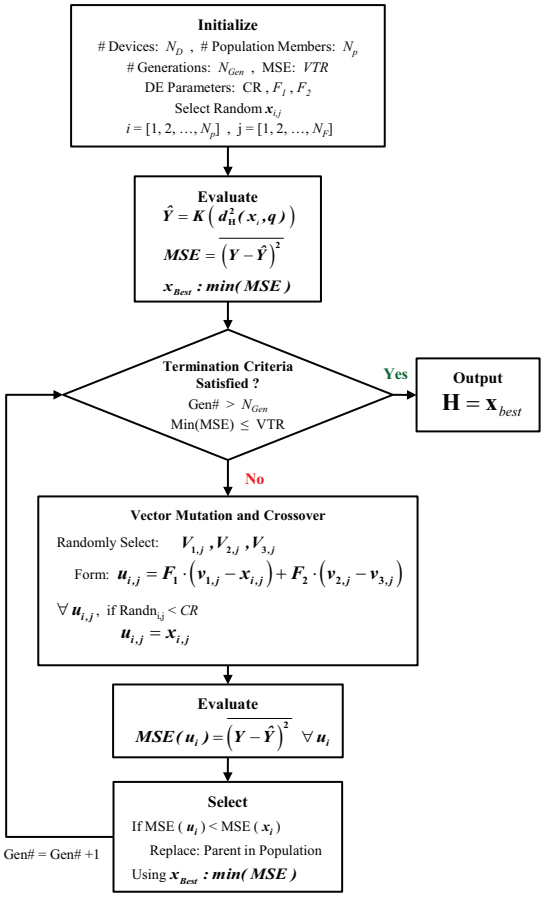


Fig. 5. DE process used to optimize Gaussian KR bandwidth parameters [7].

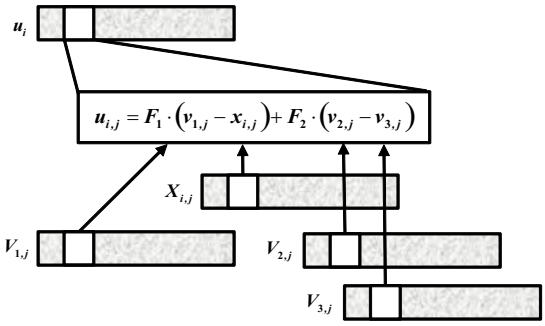


Fig. 6. DE vector-based crossover process [7].

crossover occurs which switches the values between the parent and child in that position. The final result is a child population containing N_P members that are then each assigned a fitness value based on their KR MSEs. Selection of surviving members for the next population is based on the lowest MSE values. The fitness of each child, u_i, is compared with its parent, x_i. The one with the lowest MSE is selected for the next generation. The iterative mating, crossover, and selection process continues until termination criteria is satisfied. Upon termination, the member of the final population with the lowest MSE is deemed "best" and its corresponding $\mathbf{H}$, along with the original training data, is used for subsequent classification of previously unseen input data.

3) Device Classification: Device classification in Figure 4 is implemented here as a non-linear mapping between the "best" $\hat{\mathbf{Y}}$ output from the DE process and possible input devices (classes). The mapping process is implemented via a simple comparison of each $\hat{Y}_i$ in $\hat{\mathbf{Y}}$ with threshold values and an estimated device $\hat{D}_i$ assigned as follows:

$$
\begin{aligned}
& C\left(\widehat{\mathbf{Y}}, \mathbf{t}\right) \rightarrow \widehat{\mathbf{D}} \\
& t \in [t_1,\ t_2,\ ...,\ t_{N_D-1}]\ ,\ \widehat{D}_i \in [D_1,\ D_2,\ ...,\ D_{N_D}] \\
& \widehat{Y}_i \leq t_1\ \rightarrow \widehat{D}_1 \\
& t_j < \widehat{Y}_i < t_{j+1}\ \rightarrow \widehat{D}_j\ \ 2 \leq j \leq N_D - 1 \\
& \widehat{Y}_i \geq t_{N_D-1}\ \rightarrow \widehat{D}_{N_D}\ .
\end{aligned}
\tag{10}
$$

If perfect model training occurs, i.e., the DE-optimized process results in an ideal model that perfectly represents the input data, the training data would be classified perfectly (see classification mapping in the upper righthand graphic in Figure 4 for the $N_D = 3$ case). A more general case of the mapping process in (10) is graphically illustrated in the bottom portion of Figure 4. These plots indicate less-than-perfect classification performance with the actual *Input* device number shown on the top and resultant *Estimated* device number on the left hand side.

III. Comparative Assessment Methodology

For reliable comparative assessment, *identical* fingerprint features were generated per Sect. II-A2 and input to each classifier. This was done for both TD and SD signal responses. The TD signal responses were generated per the method in [3] as centered and normalized instantaneous responses. The SD signal response was generated using the method in [6]–a Fourier-based Normalized Power Spectral Density (NPSD).

Comparison is based on Monte Carlo simulation with both classifiers trained and tested under identical conditions, including: 1) *identical* TD and SD input features extracted from N_B=500 bursts per device, 2) N_z=10 independent like-filtered AWGN realizations per burst at each SNR, and 3) SNR increments of Δ_{dB}=3.0 dB.

MDA/ML classification was implemented with *K-fold cross-validation* (Sect. II-A3) using K=5 to enable a statistically significant assessment. The required value of K can be data dependent and pilot studies confirmed that K=5 was sufficient to ensure reliability. This value is consistent with common practice that suggests values of K=5 and K=10 are appropriate [30].

The DE-optimized LFS classifier was implemented per Sect. II-B using initial parameter values of N_P=40 population members; a cross-over threshold of CR=0.2; mutation multipliers of random F_1:$N(0,1)$ and fixed F_2=0.8; and a Gaussian kernel. Initially, the DE optimization was terminated after reaching N_{Gen}=200. This

termination strategy differs from conventional DE termination that is generally based on satisfying pre-defined MSE constraints. These initial DE parameter values were empirically determined using a series of pilot studies at a given SNR and provide consistent classification performance within reasonable computation times. For legacy reasons, the LFS DE engine was implemented with K=4 fold cross validation to search for the best $\mathbf{H}$. The LFS results use the best $\mathbf{H}$ at each SNR to classify the training data.

It is important to note that *none* of the demonstration parameter values are based on optimal selection but chosen for computational efficiency to enable reliable proof-of-concept demonstration. Optimization and characterization of algorithmic and computational intensity trade-offs between MDA/ML and DE-optimized LFS classifiers remains an area of interest for ongoing research.

Classifier performance is first assessed using average *% Correct Classification* versus *SNR*, with MDA/ML performance serving as the comparative baseline for subsequent DE-optimized LFS results. This enables a one-on-one assessment of overall "classification engine" power. Performance of the DE-optimized LFS classifier is then analyzed using *Classification Error* versus number of DE generations N_{Gen} where *% Classification Error* is calculated as *100% – % Correct Classification*. This analysis enables efficient selection of N_{Gen} for achieving a desired *% Correct Classification* while at the same time reducing computation time.

IV. Results

Classifier assessment results are presented here for each of the OFDM-based signals of interest: 802.11a WiFi in Sect. IV-A and 802.16e WiMAX in Sect. IV-B. The presentation order of results and analysis are identical in the sections, with *% Correct Classification* versus *SNR* results provided first and followed by *Classification Error* versus N_{Gen}. The effect of various kernel functions on classifier performance is provided at the end.

A. 802.11a WiFi Devices

Signals for 802.11a WiFi demonstration were collected from like-model Cisco Aironet PCMCIA adapters using a pair of laptops in point-to-point (P2P) mode in an RF anechoic chamber. The collected 802.11a bursts were detected using a simple amplitude detection method with a threshold of t_D=−6 dB. The detected bursts were post-collection filtered using a 6^{th}-order Butterworth filter having a −3 dB bandwidth of W_{PC}=7.7 MHz. This same filter was used for generating the like-filtered AWGN required for SNR scaling.

For WiFi TD fingerprinting, N_{SR}=3 signal responses were used (instantaneous amplitude, phase and frequency) with N_R=10 subregions per response and N_{SM}=3 statistics per region (σ^2, γ, and κ). Therefore, the resultant number of fingerprint features (classifier input dimensions) was N_F=99 per (3). The same number of subregions and statistics were used for WiFi SD fingerprinting (N_{SR}=1), resulting in N_F=33 fingerprint features–a three-fold reduction in features relative to TD.

As presented in Fig. 7, 802.11a WiFi device classification is highly dependent on fingerprint type and somewhat unexpected. The DE-optimized LFS classifier with SD input features performs much poorer than 1) the MDA/ML classifier with equivalent input SD features, and 2) itself when using TD input features. It is hypothesized that the poorer DE-optimized LFS performance with SD features can be partially attributed to the fact that there are one-third fewer SD features than TD features (N_F = 33 SD versus N_F = 99 TD). SD classification performance actually degrades to that of random guessing (33%) for $SNR \leq 12$ dB.

The DE-optimized LFS classifier provided notable improvement with TD fingerprints and outperformed MDA/ML for $SNR \leq 15$ dB. This includes greater than 40% better classification at the lowest SNR considered. The anomalous decrease in TD performance at SNR = 21 dB was unexpected and warranted further investigation to determine if this was due to simulation error or whether it was inherent in the DE-optimized LFS RF-DNA fingerprinting process. This was first addressed by considering the effect of setting the DE termination criteria to a fixed number of generations, N_{Gen} = 200. Recall that N_{Gen} is only one of several parameters that were empirically selected for initial proof-of-concept demonstration.

The effect of fixing N_{Gen} was addressed by considering *% Classification Error* versus N_{Gen} for $N_{Gen} \in [10, 900]$ with other simulation parameters (N_B, N_P, CR, F_1, F_2, and N_z) the same as used to generate Fig. 7 results. The *% Classification Error* results are provided in Fig. 8 for SD fingerprinting at SNR = 15 dB and TD fingerprinting at SNR = 21 dB. As expected, the error exhibits an overall decreasing trend as N_{Gen} increases, with DE achieving a *% Classification Error* of approximately 4% for TD and 15% for SD at N_{Gen}=900.

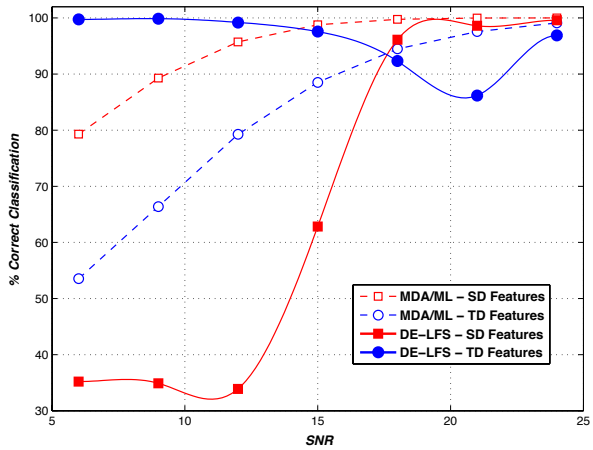


Fig. 7. Average *% Correct Classification* vs. SNR for the **802.11a WiFi** signal: TD (circle markers) and SD (square markers) fingerprinting. Previous MDA/ML classifier (unfilled markers) [3] and new DE-optimized LFS classifier (filled markers) [7], [8].

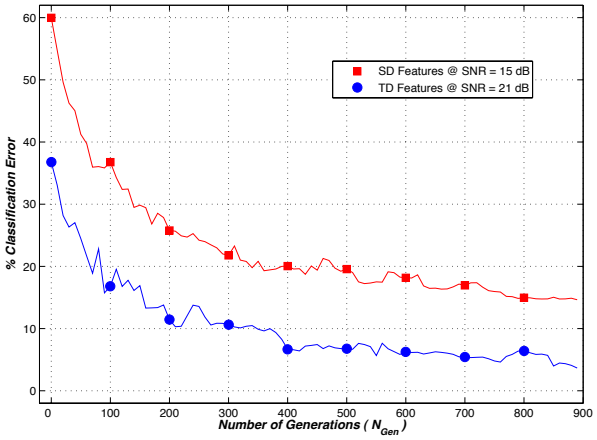


Fig. 8. Average *% Classification Error* vs. *Number of DE Generations* (N_{Gen}) for the **802.11a WiFi** signal using TD features (circle markers) at $SNR = 21$ dB and SD features (square markers) at $SNR = 15$ dB [7], [8].

Two things are worth noting in Fig. 8. First, the TD response at N_{Gen}=200 shows *% Classification Error* $\approx$ 12% which corresponds directly to the minimum *% Correct Classification* $\approx$ 87% anomaly in Fig. 7. Therefore, the TD behavior in Fig. 7 is believed to be inherent in the DE-optimized LFS RF-DNA fingerprinting process, i.e., N_{Gen}=200 iterations is simply insufficient at some SNR to realize potential DE-optimized LFS benefits. Second, it appears that SD is asymptotically approaching a lower bound of *% Classification Error* $\approx$ 14%. Investigating the effects of varying N_{Gen} and other parameters remains an area of interest in ongoing research.

B. 802.16e WiMAX Devices

Signals for 802.16e WiMAX demonstration were collected from an Alvarion-based test bed that included one Base Station (BS) transceiver (model XTRM-BS-1DIV-5.4-90D) and six like-model Mobile Subscriber (MS) station transceivers (model XTRM-SU-OD-1D-4.9-UL-A). This is commercially available equipment that provides unlicensed operation in two bands: $f_c \in [4900, 5350]$ MHz and $f_c \in [5470, 5950]$ MHz [31]. As specified in the governing IEEE 802.16 standards [32], the experimental system supports channel bandwidths of 5 MHz and 10 MHz. Results here are based on a 5 MHz channel bandwidth at f_c=5475 MHz with time division duplexing (TDD) providing separation of BS and MS transmissions.

Consistent with the goal of RF air monitoring at WAPs, the WiMAX MS-to-BS transmissions were of interest here. The observed signal structure within each WiMAX TDD frame spanned approximately t_{TDD}=5.0 ms and included a BS subframe of t_{BS}=3.0 ms followed by an S subframe of t_S=2.0 ms. In addition, the S subframe contained two distinct responses, with the first response of $t_{Rng} \approx$300 μs used for ranging (dynamic network maintenance) and the second region used for user data transfer. Considering the various TDD subframe responses that are available for RF-DNA fingerprinting, empirical studies showed that the MS subframe *ranging only* response was most promising for RF air monitoring and thus it was considered here for demonstration.

The WiMAX signals were collected using the RFSICS and detected using a simple amplitude detection method with a threshold of t_D=$-$12 dB. Once detected, the MS *ranging only* region of the subframes was extracted and post-collection filtered using a 6^{th}-order Butterworth filter having a -3 dB bandwidth of W_{PC}=3.0 MHz. This same filter was used for generating the like-filtered AWGN required for SNR scaling.

For WiMAX TD fingerprinting, N_{SR}=3 signal responses (amplitude, phase and frequency) were used with N_R=12 subregions per response and N_{SM}=3 statistics/region (σ^2, γ, and κ). Therefore, the resultant number of TD fingerprint features (classifier input dimensions) is N_F=117 per (3). The same number of subregions and statistics were used for WiMAX SD fingerprinting (N_{SR}=1), resulting in N_F=39 SD fingerprint features–again, a three-fold reduction in features relative to TD.

As presented in Fig. 9, results show that the DE-optimized LFS classifier performed well using WiMAX S *ranging-only* responses. Most notably, the DE-optimized LFS classifier outperformed the MDA/ML classifier for $SNR \leq 6$ dB using TD features. In addition, the DE-optimized LFS classifier with TD features yielded nearly 100% *Correct Classification* at lower SNR, with as much as 30% improvement noted at SNR=$-$6 dB. DE-optimized LFS classifier was less effective with SD features as classification but out performed MDA/ML over the range 0 dB $< SNR <$ 15 dB. As with WiFi device classification, the poorer performance with SD features is partially attributed to the reduces number of features (N_F=39 SD versus N_F=117 TD). Future work is planned to address the effects of dimensional TD-SD differences.

As with earlier 802.11a results in Fig. 7, the effect of

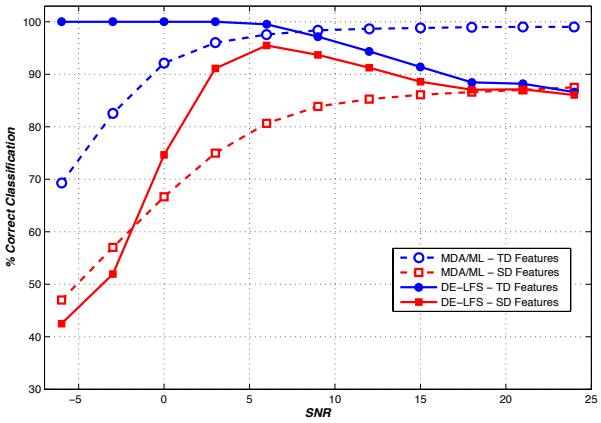


Fig. 9. Average *% Correct Classification* vs. SNR for the **802.16e WiMAX** signal using the S *range-only* response. Results shown for TD (circle markers) and SD (square markers) features using the MDA/ML classifier (unfilled markers) and the DE-optimized LFS classifier (filled markers).

fixing N_{Gen} with WiMAX signals was addressed using *% Classification Error* versus N_{Gen} for $N_{Gen} \in [10, 900]$. The other simulation parameters (N_B, N_P, CR, F_1, F_2, and N_z) remained the same. In this case, SNR=18 dB was used for both TD and SD features. The results of this analysis are presented Fig. 10. As shown, performance with SD features converged very quickly to approximately 13% average *% Classification Error*. The number of generations, N_{Gen}=200 was clearly sufficient for generating results in Fig.9.

For TD features, the *% Classification Error* versus N_{Gen} trend in Fig. 10 for the WiMAX signal is similar to what was observed in Fig. 8 with WiFi signals, with *% Classification Error* of TD features beginning to approach an asymptotic lower bound of approximately 3% after N_{Gen}=800 generations. Given this lower bound, N_{Gen}=200 was not sufficient for maximizing performance at SNR=18 dB. Clearly, in some cases, increased "learning" through more generations can improve classification performance. The limits of this approach as well as the effects of other parameters on system performance remains an area of future research.

C. Impact of Kernel Selection: WiFi Signals

The DE-optimized LFS simulation results for 802.11a signals shown in Fig. 7 were accomplished using a Gaussian kernel function. The simulation was repeated using the same parameters (N_z=10, N_{Gen}=200, SD and TD features) but alternately using a triangular and uniform kernel functions. The results of using three different kernel weighting functions can be seen in Fig. 11.

Alternate kernel performance with TD features is shown in Fig. 11(a). Relative to Gaussian kernel performance, the two alternate kernel functions produce nearly identical or better classification performance using the same number of "learning" generations (N_{Gen}=200). Specifically, the triangular kernel was able to produce at, or near, 100% correct classification for all SNRs. Of particular interest, the use of the triangular kernel eliminated the anomalous "dip" in performance at higher SNRs without requiring an increase in learning generations. The uniform kernel performed similarly across all SNRs except at SNR=24 dB. At this point, the uniform kernel produced reduced performance over both the Gaussian and triangular kernels. Further research is required to fully understand the nature of this anomaly.

Alternate kernel performance with SD features is shown in Fig. 11(b). In this case the triangular and uniform kernels outperformed the Gaussian kernel across the lower half of the simulated SNR range. However, unlike TD feature performance there is no clear top performing kernel over the range of SNR considered. For fixed N_{Gen}=200 "learning" the Gaussian kernel produced the best performance for approximately SNR>16 dB, with both the alternate kernels performing better at lower SNRs. Most notable here is that MDA/ML generally out performed LFS for all kernels and SNR considered. Relative to better TD performance in Fig. 11(a), poorer LFS SD performance in Fig. 11(b) is believed to be

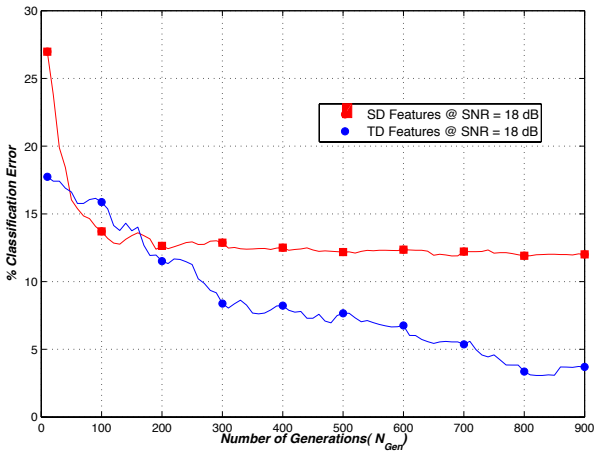
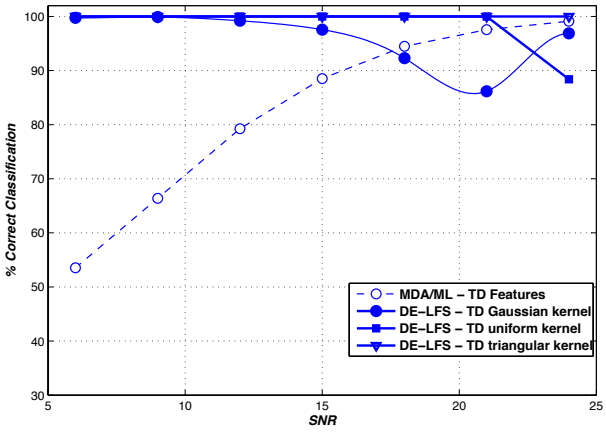
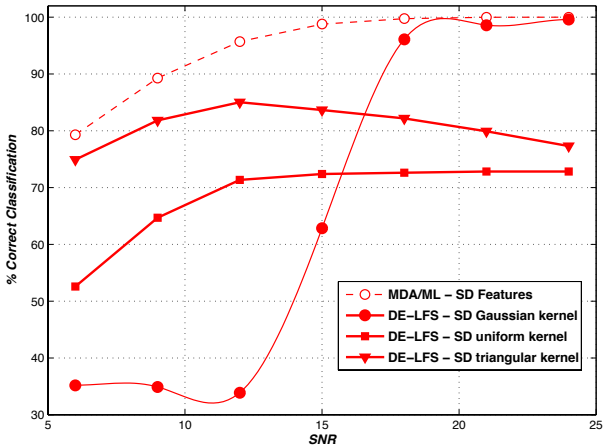


Fig. 10. Average *% Classification Error* vs. *Number of DE Generations* (N_{Gen}) for the **802.16e WiMAX** signal using TD features (circle markers) and SD features (square markers) both at SNR=18 dB.



(a) 802.11a WiFi signal fingerprinting using TD features



(b) 802.11a WiFi signal fingerprinting using SD features

Fig. 11. Average *% Correct Classification* vs. SNR for 802.11a WiFi signal fingerprinting using Gaussian (circle markers), triangular (triangle markers), and uniform (square markers) kernel functions based on a) TD features and b) SD features.

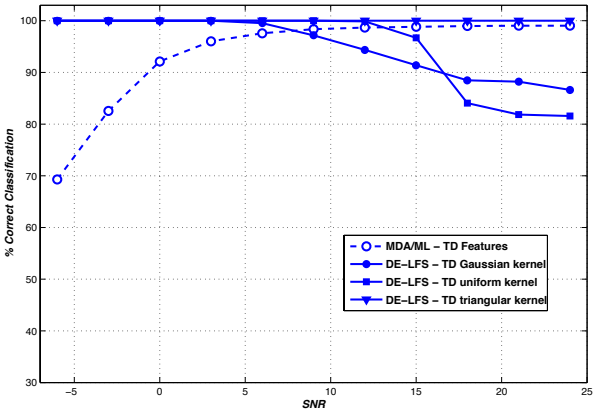
partially attributable to the reduced number of SD features (one third the number used for TD processing). The information contained in the reduced SD feature set is simply insufficient for LFS to effectively classify the 802.11a signals. Work continues on assessing the effect of feature dimensionality on overall RF-DNA fingerprinting performance.

D. Impact of Kernel Selection: WiMAX Signals

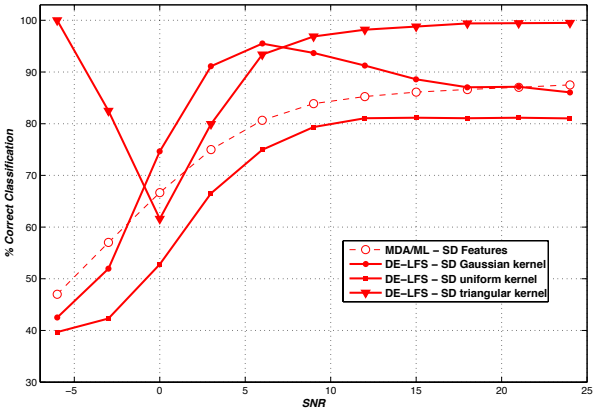
The process used in Sect. IV-C for assessing the impact of kernel function selection with WiFi signals was repeated for WiMAX signals. The results can be seen in Fig. 12, with the alternate kernel performance using TD features with 802.16e signals shown in Fig. 12(a).

The TD results with WiMax signals exhibited similar performance trends as WiFi signals with improved performance by all kernels at approximately $SNR < 8$ dB. Additionally, the triangular kernel again performed the best with 100% correct classification across all SNRs. Again, the triangular weighting function effectively removed the performance dips occurring at approximately $SNR > 8$ dB for other kernel functions without increasing the number of "learning" generations.

Alternate kernel performance using WiMax signals and SD features is shown in Fig. 12(b). For 802.11a signals, LFS had poorer performance than MDA/ML for all simulated conditions. However, for WiMax signals, LFS with the Gaussian, uniform, and triangular kernels shows mixed improvement. In the previous results, and repeated here, the Gaussian kernel performed worse than MDA/ML across all SNRs considered. However, the uniform and triangular kernel shapes resulted in *% Correct Classification* that was better than or equal to the MDA/ML results for almost all SNRs considered. A unique aspect of the triangular kernel results is the anomalous dip in performance centered at $SNR = 0$ dB. It is hypothesized that increased "learning" with more generations of training could reduce the anomalous dip while simultaneously realizing the increased performance of the triangular kernel at higher SNRs.



(a) 802.16e WiMAX signal fingerprinting using TD features



(b) 802.16e WiMAX signal fingerprinting using SD features

Fig. 12. Average *% Correct Classification* vs. SNR for 802.16e WiMAX signal fingerprinting using Gaussian (circle markers), triangular (triangle markers), and uniform (square markers) kernel functions based on a) TD features and b) SD features.

V. Summary and Conclusions

Differential Evolution (DE) is used to optimize the performance of a Learning From Signals (LFS) classification engine operating within an RF "Distinct Native Attribute" (RF-DNA) fingerprinting process. The DE-optimized LFS classifier is envisioned for use in RF air monitoring at 4G communication Wireless Access Points (WAPs) which remain as one the most vulnerable points within an information technology network. The goal is to provide additional Physical layer (PHY) based security at WAPs to augment existing bit-level mechanisms. Of particular interest here are systems based on Orthogonal Frequency Division Multiplexing (OFDM), to include existing 802.11a/g WiFi and emerging 4G 802.16 WiMAX and LTE variants.

Comparative classification performance assessment is provided for the DE-optimized LFS classifier relative to a common Bayesian-based Multiple Discriminant Analysis/Maximum Likelihood (MDA/ML) classifier. The assessment is performed using *identical* classifier input features extracted from experimentally collected 802.11a WiFi and 802.16 WiMAX signals. Classification features are derived from statistical RF-DNA extracted from Time Domain (TD) and Spectral Domain (SD) responses.

Relative to MDA/ML classification, DE-optimized LFS classification performance was generally superior at lower SNR and provided considerable improvement of over 40% in classification accuracy in some cases. Best case classification improvement was realized with TD fingerprint features. While not quite as effective, DE-optimized LFS classification with SD fingerprinting was notable with the difference between TD and SD performance initially attributed to feature dimension differences as there were one-third fewer SD features than TD features.

Analysis based on *% Classification Error* versus the number of DE generations N_{Gen} showed that the anomalous behavior of TD and SD fingerprinting at higher signal-dominated SNR, as well as the poorer performance with SD fingerprinting, is inherent in the DE-optimized LFS RF-DNA fingerprinting process and directly at-

tributable to forced DE termination after N_{Gen}=200 generations. Alternatively, improved performance over specific SNR regions was also demonstrated using alternate kernel functions. A triangular kernel was able to eliminate the TD feature classification performance "dip" at higher SNR without requiring additional generations of "learning." Investigation continues into the effects of varying N_{Gen}, kernel function shape, and other parameters that were fixed for initial proof-of-concept demonstration.

ACKNOWLEDGMENT

Research sponsored by the Laboratory for Telecommunications Sciences (LTS), US Department of Defense, and the Sensors Directorate, Air Force Research Laboratory (AFRL), Wright-Patterson AFB, OH.

The views expressed in this paper are those of the authors and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government.

REFERENCES

[1] H. Collins, "Top 10 Network Security Threats," Sep 2010.

[2] T. D. Tarman and E. L. Witzke, "Intrusion Detection Considerations for Switched Networks," vol. 4232, no. 1, pp. 85–92, 2001. [Online]. Available: http://link.aip.org/link/?PSI/4232/85/1

[3] R. W. Klein, M. A. Temple, and M. J. Mendenhall, "Application of Wavelet-Based RF Fingerprinting to Enhance Wireless Network Security," vol. 11, no. 6, pp. 544–555, Dec 2009.

[4] D. Reising, M. Temple, and M. Mendenhall, "Improving Intra-Cellular Security Using Air Monitoring with RF Fingerprints." IEEE Wireless Communications and Networking Conf (WCNC10), Apr 2010.

[5] M. D. Williams, M. A. Temple, and D. R. Reising, "Augmenting Bit-Level Network Security Using Physical Layer RF-DNA Fingerprinting," in *IEEE Global Communications Conf*, Dec 2010.

[6] M. D. Williams, S. A. Munns, M. A. Temple, and M. J. Mendenhall, "RF-DNA Fingerprinting for Airport WiMAX Communications Security," in *Proc of 4th Int'l Conf on Network and Systems Security (NSS10)*, Sep 2010.

[7] P. K. Harmer, M. A. Temple, M. A. Buckner, and E. Farquahar, "Using Differential Evolution to Optimize 'Learning from Signals' and Enhance Network Security," in *Genetic and Evolutionary Computation Conf (GECCO11)*, Jul 2011.

[8] P. K. Harmer, M. D. Williams, and M. A. Temple, "Using DE-Optimized LFS Processing to Enhance 4G Communication Security," in *Proc of Int'l Conf on Computer Communication Networks (ICCCN11)*, 2011.

[9] J. Hall, M. Barbeau, and E. Kranakis, "Detecting rogue devices in bluetooth networks using radio frequency fingerprinting," in *Communications and Computer Networks*, 2006, pp. 108–113.

[10] ——, "Detection of Transient in Radio Frequency Fingerprinting Using Signal Phase." 2003 IASTED Int'l Conf on Wireless and Optical Communications (WOCC03), May 2003.

[11] ——, "Using Transceiverprints for Anomaly Based Intrusion Detection." 2004 IASTED Int'l Conf on Communications, Internet and Information Technology (CIIT), Nov 2004.

[12] B. Danev and S. Capkun, "Transient-based identification of wireless sensor nodes," in *Proc of the 2009 Int'l Conf on Information Processing in Sensor Networks*, ser. IPSN '09. Washington, DC, USA: IEEE Computer Society, 2009, pp. 25–36. [Online]. Available: http://portal.acm.org/citation.cfm?id=1602165.1602170

[13] B. Danev, T. S. Heydt-Benjamin, and S. Čapkun, "Physical-layer identification of RFID devices," in *Proc of the 18th Conf on USENIX security symposium*, ser. SSYM'09. Berkeley, CA, USA: USENIX Association, 2009, pp. 199–214. [Online]. Available: http://portal.acm.org/citation.cfm?id=1855768.1855781

[14] B. Danev, H. Luecken, S. Čapkun, and K. El Defrawy, "Attacks on Physical-Layer Identification," in *Proc of the 3rd ACM Conf on Wireless Network Security*, ser. WiSec'10. ACM, 2010.

[15] D. Reising, M. Temple, and M. Mendenhall, "Improved wireless security for GMSK-based devices using RF fingerprinting," vol. 3, no. 1, pp. 41–59, Mar 2010.

[16] V. Brik, S. Banerjee, M. Gruteser, and S. Oh, "Wireless device identification with radiometric signatures," in *Proc of the 14th ACM Int'l Conf on Mobile computing and networking*, ser. MobiCom '08. New York, NY, USA: ACM, 2008, pp. 116–127. [Online]. Available: http://doi.acm.org/10.1145/1409944.1409959

[17] J. Pang, B. Greenstein, R. Gummadi, S. Seshan, and D. Wetherall, "802.11 user fingerprinting," in *Proc of the 13th annual ACM Int'l Conf on Mobile computing and networking*, ser. MobiCom '07. New York, NY, USA: ACM, 2007, pp. 99–110. [Online]. Available: http://doi.acm.org/10.1145/1287853.1287866

[18] D. Takahashi, Y. Xiao, Y. Zhang, P. Chatzimisios, and H.-H. Chen, "IEEE 802.11 user fingerprinting and its applications for intrusion detection," vol. 60, no. 2, pp. 307 – 318, 2010, advances in Cryptography, Security and Applications for Future Computer Science. [Online]. Available: http://www.sciencedirect.com/science/article/B6TYJ-4YBVN48-1/2/02fc08e080dbf58edcc440f8db4ed9f3

[19] W.C. Suski II, M.A. Temple, M. J. Mendenhall, and R.F. Mills, "Radio frequency fingerprinting commercial communication devices to enhance electronic security," vol. 1, pp. 301–322, Oct 2008. [Online]. Available: http://portal.acm.org/citation.cfm?id=1454744.1454749

[20] D. Zanetti, B. Danev, and S. Capkun, "Physical-layer identification of UHF RFID tags," in *Proc of the sixteenth annual Int'l Conf on Mobile computing and networking*, ser. MobiCom '10. New York, NY, USA: ACM, 2010, pp. 353–364. [Online]. Available: http://doi.acm.org/10.1145/1859995.1860035

[21] Agilent, *Agilent E3238 Signal Intercept and Collection Solutions: Family Overview*, Publication 5989-1274EN, Agilent Technologies Inc., USA, 2004.

[22] J. Toonstra and W. Kinsner, "A Radio Transmitter Fingerprinting System ODO-1," in *Electrical and Computer Engineering, 1996. Canadian Conf on*, vol. 1, May 1996, pp. 60 –63 vol.1.

[23] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*, 2nd ed. Wiley, Nov 2001.

[24] M. A. Buckner, A. M. Urmanov, A. V. Gribok, and J. W. Hines, "Application of Localized Regularization Methods for Nuclear Power Plant Sensor Calibration Monitoring," Technical Correspondence, 2002.

[25] V. S. Cherkassky and F. Mulier, *Learning From Data: Concepts, Theory, and Methods*, 2nd ed. Hoboken, HJ: Wiley & Sons, 2007.

[26] M. A. Buckner, "Learning From Data with Localized Regression and Differential Evolution," Ph.D. dissertation, University of Tennessee, Knoxville, May 2003.

[27] C. Goutte and J. Larsen, "Adaptive Metric Kernel Regression," vol. 26, no. 1/2, pp. 155–167, 2000.

[28] K. Price, R. M. Storn, and J. A. Lampinen, *Differential Evolution: A Practical Approach to Global Optimization (Natural Computing Series)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2005.

[29] E. Mezura-Montes, J. Velazquez-Reyes, and C. A. C. Coello, "A comparative study of differential evolution variants for global optimization," in *Proc of Genetic and Evolutionary Computation Conference (GECCO)*, July 2006.

[30] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning; Data Mining, Inference, and Prediction*. Springer-Verlag, New York, New York, USA, 2001.

[31] Alvarion Ltd, *Alvarion BreezeMAX Extreme 5000: WiMAX 16e Pioneer for the License-Exempt Market*, Pub #215373, Rev. A, 2009.

[32] *IEEE Std 802.16-2009, Local and Metropolitan Area Networks, Part 16: Air Interface for Broadband Wireless Access Systems*, Inst of Electrical and Electronics Engineers, New York, New York, USA, May 2009.