Using Differential Evolution to Optimize 'Learning from Signals' and Enhance Network Security

Paul K. Harmer and Michael A. Temple
Air Force Institute of Technology
2950 Hobson Way
WPAFB, Dayton OH 45433
[Paul.Harmer, Michael.Temple]@afit.edu

Mark A. Buckner and Ethan Farquahar
Oak Ridge National Laboratory
1 Bethel Valley Rd
Oak Ridge, TN 37831
[bucknerma, farquhare]@ornl.gov

ABSTRACT

Computer and communication network attacks are commonly orchestrated through Wireless Access Points (WAPs). This paper summarizes proof-of-concept research activity aimed at developing a physical layer Radio Frequency (RF) air monitoring capability to limit unauthorized WAP access and improve network security. This is done using Differential Evolution (DE) to optimize the performance of a "Learning from Signals" (LFS) classifier implemented with RF "Distinct Native Attribute" (RF-DNA) fingerprints. Performance of the resultant DE-optimized LFS classifier is demonstrated using 802.11a WiFi devices under the most challenging conditions of intra-manufacturer classification, i.e., using emissions of like-model devices that only differ in serial number. Using identical classifier input features, performance of the DE-optimized LFS classifier is assessed relative to a Multiple Discriminant Analysis / Maximum Likelihood (MDA/ML) classifier that has been used for previous demonstrations. The comparative assessment is made using both Time Domain (TD) and Spectral Domain (SD) fingerprint features. For all combinations of classifier type, feature type, and signal-to-noise ratio considered, results show that the DE-optimized LFS classifier with TD features is superior and provides up to 20% improvement in classification accuracy with proper selection of DE parameters.
Track: Real world applications.

Categories and Subject Descriptors

I.2.8 [**Computing Methodologies Artificial Intelligence**]: Problem Solving, Control Methods, and Search—*Heuristic methods*; K.6.5 [**Management of Computing and Information Systems**]: Security and Protection—*Authentication, Unauthorized access*

General Terms

Differential Evolution, Fingerprinting, Network, Security

1. INTRODUCTION

GECCO'11, July 12–16, 2011, Dublin, Ireland.

Computer and communication network attacks occur on a regular basis and will likely continue as ill-intentioned "hackers" attempt to gain unauthorized access [3, 4, 8]. Many of these attacks are orchestrated through wireless access points (WAPs) which have recently been identified as one of the most vulnerable points in an Information Technology (IT) network [10]. The first step to stopping many of these attacks is identifying when they are occurring.

Traditional network security systems have relied heavily on information in upper Open Systems Interconnection (OSI) model layers to provide bit-level security and detect unauthorized users [22]. Unfortunately, these approaches ignore potentially useful information that is inherently "buried" in the Radio Frequency (RF) characteristics of participating network devices. The OSI network model consists of seven layers with different services provided at each layer. Most intrusion detection systems operate at Layer #3, the Network (NET) layer, or above [22]. Thus, inherent RF information that exists in the lowest Physical (PHY) layer remains largely unexploited.

Exploiting PHY layer information is of interest here with a goal of enhancing network user authentication and preventing unauthorized system access. It is envisioned that the bit-level protection provided by higher-layer intrusion and authentication systems could be effectively augmented by RF PHY layer protection hosted in an RF air monitoring device located at specific WAPs [15, 20, 25, 26]. These earlier works demonstrated that RF information, i.e., RF "Distinct Native Attribute" (RF-DNA) fingerprints, are indeed useful for identifying specific devices and augmenting bit-level network security mechanisms. However, overall classification performance in these earlier works decreases as Signal-to-Noise Ratio (SNR) decreases–behavior that is generally expected and observed for all signal detection, estimation and classification algorithms.

Improved classification performance is commonly addressed through two means: 1) finding more robust input features for a given classifier, or 2) finding a more robust classifier for given input features. The second approach is considered here using the same Time Domain (TD) and Spectral Domain (SD) features used in previous related work [15, 20, 25, 26]. Given TD and SD fingerprint features, the goal is to develop a more powerful classification engine that is optimized through Differential Evolution (DE). Recent works have used various Genetic Algorithms (GA) alone as clas-

sifiers [2, 14, 16]. Our approach uses DE to optimize the parameters within an existing classification engine. Success of this so-called "Learning from Signals" (LFS) approach is measured as either 1) improving classification and authentication performance for a given SNR, or 2) achieving a given classification performance at a lower SNR.

Success of the proposed DE-optimized LFS classifier is demonstrated here by way of summarizing proof-of-concept research aimed at improving device classification using an RF-DNA fingerprinting process. This is accomplished by comparative assessment with previous classification results obtained using a Fisher-based Multiple Discriminant Analysis/Maximum Likelihood (MDA/ML) classifier. Most importantly, this comparison is made using *identical* TD and SD RF-DNA fingerprint features input to the classifiers. Relative to inter-manufacturer classification (mix of interoperable devices from different manufacturers), intra-manufacturer classification (like-model devices from the same manufacturer) presents the greatest classification challenge [15, 20, 25, 26]. It also offers the greatest opportunity to demonstrate improvement and is considered here.

The remainder of this paper is organized in the following manner. Section 2 provides background information for key concepts required to conduct the research, including RF-DNA Fingerprinting, LFS Classification, Gaussian Kernel Regression (KR), and the DE-optimized LFS Implementation used here. Section 3 provides the demonstration methodology that details the logical flow of processes used to obtain comparative assessment results in Section 4. Finally, Section 5 summarizes the work and presents conclusions.

2. BACKGROUND

The following subsections provide background information on key technical concepts used to conduct the research.

2.1 RF-DNA Fingerprinting

RF-DNA fingerprinting is a PHY layer technique used to uniquely identify devices based on inherent differences in their transmissions. It has been demonstrated that specific serial-numbered devices possess unique transmission characteristics that may be attributed to minute differences in manufacturing (part type, part lot number, assembly processes, etc.). The goal here is to uniquely identify, by serial number, hardware devices as an aid to network security and user authentication. Various RF fingerprinting techniques have been used previously to demonstrate this for various communication signals, including: 802.11 WiFi signals [5, 15, 17, 21, 24], GSM cell phone signals[19, 26], 802.16 WiMAX signals [25], 802.15 Bluetooth signals [13], and RFID signals [11, 27].

While the earlier cited works have considered several diverse methods for implementing RF fingerprinting, the techniques generally share some common functionality, including: 1) Signal Collection, 2) Signal Detection, 3) Fingerprint Feature Generation, and 4) Signal/Device Classification. The first two steps functionally embody the processes of signal reception, digitization, and post-collection processing to pre-condition the TD signal response for feature extraction. For the next step, Fingerprint Feature Generation, the input classifier features are either generated directly from the TD signal response or generated in an alternate feature domain through transforming the TD response, e.g., to the frequency domain via a Discrete Fourier transform (DFT). The goal of transformation is to project the original TD response into an alternate domain that contains an increased amount of discriminating information. The focus here is on using TD and DFT-based SD responses, with final classification features generated by calculating statistical metrics over selected TD and SD response regions.

In the final RF fingerprinting step, Signal/Device Classification, a given classifier is implemented to separate and identify N_D devices (input classes) using selected input features. Several classification approaches have been considered from within the pattern recognition community, including those based on cross-correlation, vector distance measures, k-nearest neighbor metrics, support vector machines, and Fisher-based MDA/ML processing [13, 15, 23, 25, 27]. The MDA/ML classifier of [15, 25] was adopted here to provide comparative baseline results.

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 [12]. Given N_D devices (input classes), the MDA/ML process projects the input features into an $N_D - 1$ decision space. This is best visualized for the $N_D = 3$ class problem as illustrated in Figure 1 which shows Gaussian class likelihood functions and the resultant 2-dimensional decision space (lower surface) with ML boundaries.

Like similar classification methods, the MDA/ML projection matrix $\mathbf{W}$ is determined using a training process, with previously unseen class members subsequently classified using Bayesian decision theory. Features for previously unseen class members are projected using $\mathbf{W}$ and estimated as coming from one of N_D classes based on ML criteria. While MDA/ML and other classifiers have achieved acceptable classification performance in RF fingerprinting applications, the contributions here are in developing and incorporating a more powerful DE-optimized LFS classifier for applications requiring improved performance.

2.2 LFS Classification

LFS classification is an adaption of Learning From Data

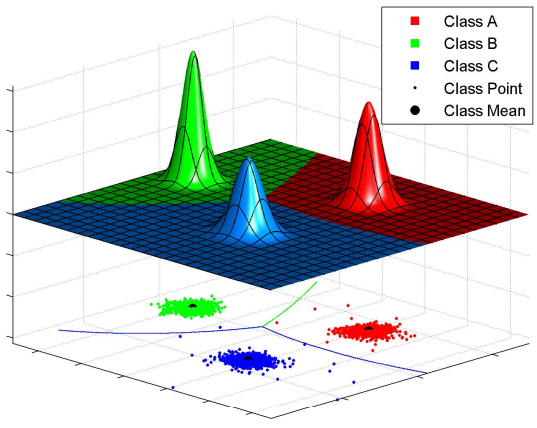

Figure 1: MDA/ML training projections and decision boundaries for $N = 3$ class problem.

(LFD) techniques where the input training data is derived from samples of a given sensor response [7, 9]. 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. The LFD approach is not constrained to using polynomial models and there are several other data fitting methodologies using alternate 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. Therefore, the goal is to find useful information in the input data and exploit that information when acting on future observed data [9].

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 they 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 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. It has been shown that a 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. The GA-optimized approach has been applied using more powerful KR techniques [6, 7] and is adopted here for LFS classification.

2.3 Gaussian KR Processing

KR is a memory-based technique that stores past input data and processes them when a new query is made. So, instead of modeling the entire input set with a model, as in conventional linear regression, the local KR function is estimated over the entire input domain by fitting a simple model at every query point $\mathbf{q}$. Only observations that are close to the query point are used to fit the model. 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 $\mathbf{q}$. Any kernel function can be used for KR provided the following properties are satisfied [9]:

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 satisfying the properties above, a Gaussian kernel was chosen here. For this work, h_i is used to represent the bandwidth parameter for the i^{th} dimension of a multidimensional Gaussian kernel function given by

$$K\left(d^2_{\mathbf{H}}(\mathbf{x_i}, \mathbf{q})\right) = \exp^{-0.5 \cdot d^2_{\mathbf{H}}(\mathbf{x_i}, \mathbf{q})}\ , \tag{1}$$

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 of the off-diagonal elements in $\mathbf{H}$ are zero given by

$$\mathbf{H} = diag\left(h_1, h_2, \ldots, h_{N_F}\right)\ ,\ h_i \geq 0\ \forall\ i\ \ . \tag{2}$$

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

$$d^2_{\mathbf{H}}(\mathbf{x_i}, \mathbf{q}) = (\mathbf{x_i} - \mathbf{q})^T \mathbf{H}^{-1} (\mathbf{x_i} - \mathbf{q})\ . \tag{3}$$

Finally, the kernel regression estimate, $\widehat{y}$, for a previously unseen system input, or query point, $\mathbf{q}$, is given by

$$\widehat{y} = \frac{\sum_{i=1}^{N_S} K(d^2_{\mathbf{H}}(\mathbf{x_i}, \mathbf{q})) \cdot y_i}{\sum_{i=1}^{N_S} K(d^2_{\mathbf{H}}(\mathbf{x_i}, \mathbf{q}))}, \quad N_S = N_D \cdot N_B\ . \tag{4}$$

For conventional KR processing, a given bandwidth of $h \in \Re$ would be used for all input dimensions–elements in $\mathbf{H}$ of (2) are *identical*. The approach here differs in that DE KR optimization, as demonstrated in [6, 7], is able to "learn" the best bandwidth parameter h_i to use for each dimension and improve LFS classifier performance. One can also infer the relative importance of a given dimension/parameter based on h_i, i.e., a "smaller" h_i indicates greater importance.

2.4 DE-Optimized LFS Implementation

The DE-optimized LFS classifier is illustrated in Figure 2 and functionally includes three processes: Input Feature Formatting, DE Optimized KR, and Device Classification.

2.4.1 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 3. 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 right-hand graphic in Figure 2). Details for the classification mapping process are presented in Section 2.4.3.

2.4.2 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

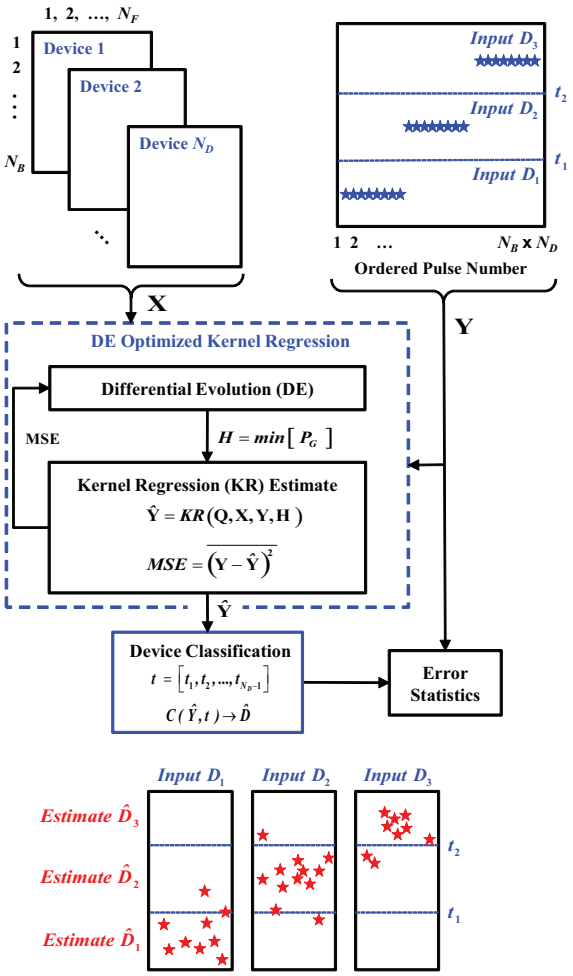

Figure 2: DE-optimized LFS classification process.

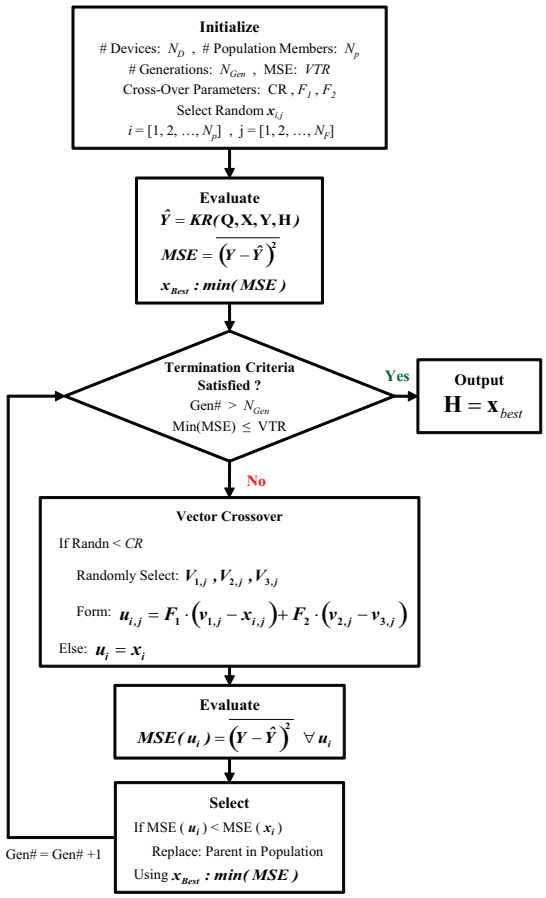

Figure 3: DE process used to optimize Gaussian KR bandwidth parameters.

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. Algorithm details for DE processing differ from conventional GA processing primarily in the manner by which future generations are produced [7, 18].

The DE process for this work was implemented as detailed in Figure 3. The initial population P_{Gen} for $Gen\# = 1$ contains N_P randomly generated members. Each member is represented by a vector of h_i, $i = 1, 2, ..., N_F$, Gaussian kernel bandwidths. 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}, 2) a minimum specified MSE Value To Reach (VTR). If not satisfied during the current generation, vector-based crossover occurs as illustrated in Figure 4. In this case, each population member (parent) X_i is crossed with three other randomly selected individuals (mates) (V_1, V_2 and V_3) based on the crossover threshold CR. As shown in Figure 4, 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. The 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 KR

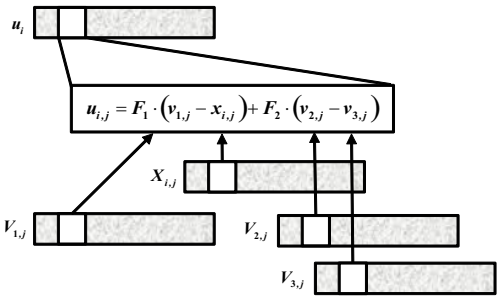

Figure 4: DE vector-based crossover process.

$\vec{H}$, along with the original training data, used for subsequent classification of previously unseen input data.

2.4.3 Device Classification

Device classification in Figure 2 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:

$$
\begin{aligned}
& C\left(\hat{\mathbf{Y}}, \mathbf{t}\right) \rightarrow \hat{\mathbf{D}} \\
& t \in [t_1,\ t_2,\ ...,\ t_{N_D-1}]\ ,\ \hat{D}_i \in [D_1,\ D_2,\ ...,\ D_{N_D}] \\
& \hat{Y}_i \leq t_1\ \rightarrow \hat{D}_1 \\
& t_j < \hat{Y}_i < t_{j+1}\ \rightarrow \hat{D}_j\ \ 2 \leq j \leq N_D - 1 \\
& \hat{Y}_i \geq t_{N_D-1}\ \rightarrow \hat{D}_{N_D}\ .
\end{aligned}
\tag{5}
$$

The mapping process in (5) is graphically illustrated in the bottom portion of Figure 2 for the $N_D = 3$ case. 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.

3. METHODOLOGY

The methodology for collecting, processing, and classifying 802.11a signals of interest is shown in Figure 5. This was adopted from [15] 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." The process begins with signal collection using the RF Signal Intercept and Collection System (RFSICS). This is followed by post-collection processing using MATLAB to generate the desired analysis signal, extract fingerprint features, and perform classification. The RFSICS is based on Agilent's ES238S system and can collect signals from 20.0 MHz to 6.0 GHz [1]. The RF band of interest is selected using a tunable $W_{RF} = 36.0$ MHz filter. The collected signals are down-converted to an intermediary frequency (IF) of $f_{IF} = 70.0$ MHz before being digitized to $b = 12$ bits at a sample rate of $f_s = 95$ M samples-per-second. The sampled IF data is stored as complex In-phase (I) and Quadrature (Q) signal components.

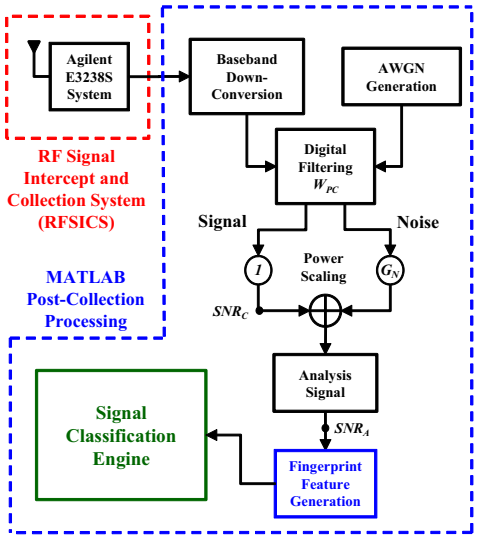

Figure 5: Methodology used for RF signal collection, processing, and classification [15].

The devices under test included three like-model CISCO 802.11a PCMCIA cards. To isolate RF signal features from environmental channel effects and interference, a pair of laptops with PCMCIA cards were set up as a point-to-point (P2P) network in an RF anechoic chamber. File transfer protocol was used to continuously pass files between the laptops during the collections. The two PCMCIA transmit powers were set to two different power levels to facilitate easy association of collected signal responses (occurring in two different, assigned time division duplex intervals) with a given laptop during post-processing. The PCMCIA cards were swapped and collections made for each device of interest. Given the anechoic chamber environment and the relatively close proximity of the P2P laptops, the post-filtered collected SNR for all signals was on the order of $SNR_C = 40$ dB SNR. This high level of SNR_C facilitated direct scaling (G_N in Figure 5) and addition of like-filtered Additive White Gaussian Noise (AWGN) to generate analysis signals at the desired SNR (SNR_A in Figure 5).

The collected signal bursts were detected using a simple amplitude detection method with a threshold of $t_D = -6$ dB. Following detection, the bursts were post-collection filtered using a 6^{th}-order Butterworth filter having a -3 dB bandwidth of $W_{PC} = 7.7$ MHz–a notional bandwidth for receiving 802.11 signals. A total of $N_B = 500$ bursts per device were collected, detected, post-collection filtered, and "fingerprinted" for classification.

For reliable comparative assessment, *identical* fingerprint features were generated and used with each classifier. This was done for both TD and SD signal responses. For TD and SD fingerprinting, there are $N_{SR} = 3$ and $N_{SR} = 1$ signal responses, respectively, with the three available TD responses including instantaneous amplitude, phase and frequency. In both cases, the selected response(s) is parsed into N_R equal length subregions as illustrated in Figure 6 for representative TD and SD responses of an 802.11a WiFi signal. The entire response is included for feature generation as well, yielding a total number of feature regions of $N_R^F = (N_R + 1) \times N_{SR}$. Accounting for N_S^F total statistics per region, the composite RF statistical fingerprint (feature vector) has a total number of elements (feature dimensions) given by

$$
N_F = N_R^F \times N_S^F = N_{SR} \times (N_R + 1) \times N_S^F\ . \tag{6}
$$

While any statistics could be used to characterize signal responses, standard deviation (σ), variance (σ^2), skewness (γ), and/or kurtosis (k) have been generally considered and used to form the *regional fingerprint* given by

$$
F_{R_i} = [\sigma_{R_i}\ \sigma_{R_i}^2\ \gamma_{R_i}\ k_{R_i}]_{1 \times N_S^F}\ , \tag{7}
$$

where $i = 1, 2, ..., N_R + 1$. The vectors from (7) are concatenated to form the *composite statistical fingerprint* for each characteristic and is given by

$$
\mathbf{F}_C = \left[F_{R_1} \vdots F_{R_2} \vdots F_{R_3}\ \ldots\ F_{R_{N_R+1}} \right]_{1 \times N_D^F}\ . \tag{8}
$$

For SD results, the SD signal response is generated using the

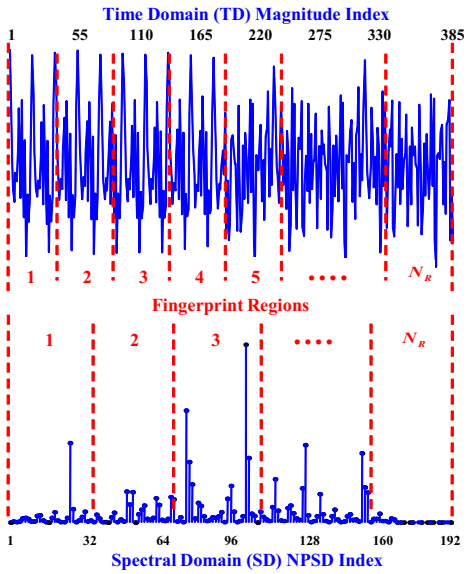

Figure 6: Fingerprint Feature Regions: TD Amplitude Response and Corresponding SD Response Based on NPSD [25].

method in [25]–a Fourier-based Normalized Power Spectral Density (NPSD). For TD results, the TD signal responses are generated per the method in [15] as centered and normalized instantaneous responses.

Classifier performance is first addressed using average *% Correct Classification* versus *SNR* with MDA/ML performance serving as the comparative baseline–note that the subscripted A in SNR_A is suppressed here and SNR used henceforth in the paper. As a first step, an MDA/ML performance baseline is established and results in [15] reproduced and re-verified. Classification was then performed using the DE-optimized LFS classifier with DE parameters *fixed* at a population size of $N_P = 40$, a crossover threshold of $CR = 0.2$, crossover multipliers of $F_1{:}N(0,1)$ and $F_2 = 0.8$, and termination occurring after $N_{Gen} = 200$ generations. It is important to note that this termination criteria differs from conventional DE termination approaches which are generally based on satisfying pre-defined MSE constraints for a given objective function. These initial conditions were empirically determined following a series of TD fingerprinting pilot studies at $SNR = 24$ dB which provided consistent classification performance within reasonable computation times.

4. RESULTS AND ANALYSIS

For initial comparative assessment, the MDA/ML classifier was implemented per Section 2.1. Performance is compared using Monte Carlo simulation with both classifiers trained and tested under identical conditions which included: 1) TD and SD input feature vectors generated from $N_B = 500$ 802.11a bursts per device, 2) $N_z = 10$ independent like-filtered AWGN realizations per burst at each SNR, and 3) $SNR \in [6, 24]$ dB in $\Delta_{dB} = 3.0$ dB increments. The resultant average *% Correct Classification* versus SNR for each classifier and feature type is shown in Figure 7.

Results in Figure 7 are mixed and show that the proposed

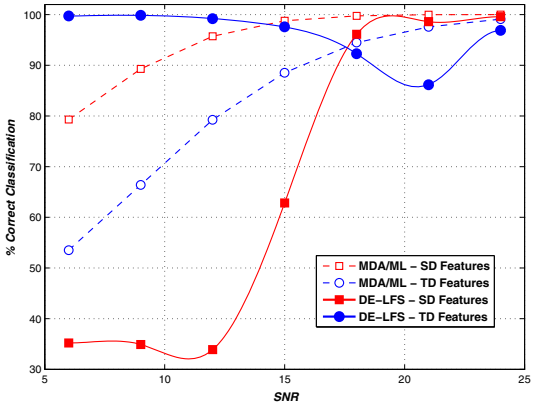

Figure 7: Average *% Correct Classification* versus SNR for 802.11 WiFi signal: TD (circle markers), SD (square markers), MDA/ML classifier (unfilled markers) [15] and DE-LFS classifier (filled markers).

DE-optimized LFS classifier is both superior and inferior to the MDA/ML classifier with performance being highly dependent on feature type. For SD fingerprint features, DE-optimized LFS performance is inferior to MDA/ML for all SNR considered, with performance degradation to random guessing (33%) for $SNR \leq 12$ dB. However, the DE-optimized LFS classifier is superior to MDA/ML in the noise-dominated SNR region using TD fingerprint features. Most notably, DE-optimized LFS with TD fingerprints outperform MDA/ML at $SNR \leq 15$ dB and provides nearly 20% improvement in classification performance at the lowest SNR considered. However, for the signal-dominated SNR region there is a noticeable "dip" in performance for $15 < SNR < 24$ dB; a minimum of approximately 80% occurring at $SNR = 21$ dB. Preliminary analysis of DE-optimized LFS results in Figure 7 focused on answering 1) "Why is SD performance so inferior and degrade rapidly for $SNR \leq 18$ dB?" and 2) "Why does TD performance exhibit a "dip" in performance for $15 < SNR < 24$?" While not analyzed

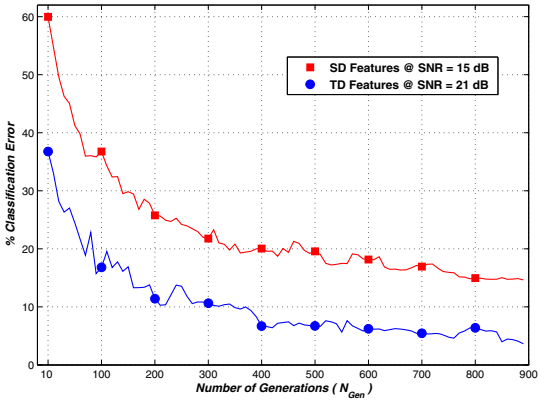

Figure 8: Average *% Classification Error* versus *Number of DE Generations* (N_{Gen}) for the 802.11 WiFi signal: TD (circle markers) at $SNR = 21$ dB and SD (square markers) at $SNR = 15$ dB.

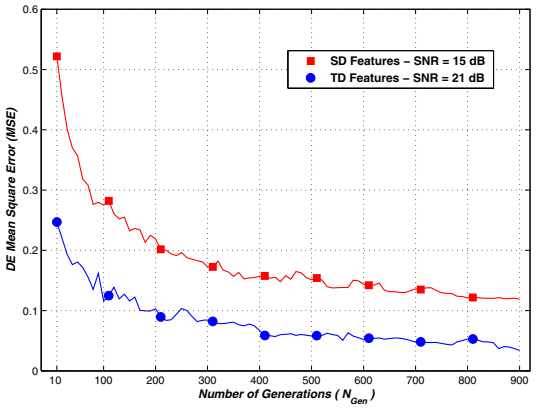

Figure 9: Corresponding *DE MSE* versus *Number of DE Generations* (N_{Gen}) for *% Classification Error* presented in Figure 8.

in detail at this point, the disparity between achievable TD and SD performance with the DE-optimized LFS classifier is believed to be partially attributable to the factor of three disparity between the number of TD ($N^F_{TD} = 99$) and SD features ($N^F_{SD} = 33$). It is interesting to note that this same disparity did not impact the MDA/ML classifier which performed best with the SD features. With regard to answering both questions, the summary discussion that follows suggests that the effects being addressed are attributable to terminating the DE process after a fixed number of $N_{Gen} = 200$ generations–an insufficient number of generations at some SNR for full benefits of DE-optimized LFS to be realized.

Degraded performance for the two noted cases was analyzed by considering classification error (*% Classification Error* = 100 – *% Correct Classification*) as a function of N_{Gen} for $N_{Gen} \in [10, 900]$ in increments of $\Delta_{gen} = 10$. The other parameters (N_B, N_P, CR, F_1, F_2, and N_z) were unchanged from Figure 7 results. The resultant *% Classification Error* versus N_{Gen} is provided in Figure 8 for TD fingerprint features generated at $SNR = 21$ dB and SD fingerprint features generated at $SNR = 15$ dB. As expected, the error exhibits an overall decreasing trend as N_{Gen} increases for both feature types, with DE achieving a *% Classification Error* of approximately 4% for TD and 15% for SD at $N_{Gen} = 900$.

Two things are worth noting in Figure 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–TD behavior in Fig. 7 is believed to be inherent in the DE-optimized LFS RF-DNA fingerprinting process and $N_{Gen} = 200$ 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* $\approx 14\%$.

As used for generating results in Figure 7 and Figure 8, *% Correct Classification* and *% Classification Error* are "operational" classification performance metrics. It is important to note that this operational metric is not the same MSE metric that is generally used for terminating the DE process and characterizing algorithm performance. However, it is expected that the *% Classification Error* response should mimic the DE MSE response. This is confirmed by comparing the DE MSE results in Figure 9 with *% Classification Error* results in Figure 8. In this case, the DE MSE in Figure 9 is that of the best fit member shown for $N_{Gen} \in [10, 900]$ with $\Delta_{Gen} = 10$.

5. SUMMARY AND CONCLUSIONS

This work provides a first look at using Differential Evolution (DE) to optimize parameters of a "Learning From Signals" (LFS) classifier for use in an RF air monitor that uniquely identifies and authenticates wireless devices for network access. It is envisioned that DE-optimized LFS air monitoring would be implemented at Wireless Access Points (WAPs), the vulnerability of which has been recently identified as one of the top security threats to Information Technology (IT) systems. For proof-of-concept demonstration, the LFS classifier input features were based on RF "Distinct Native Attribute" (RF-DNA) fingerprints from IEEE 802.11 WiFi signals. The air interface of existing 3G 802.11 networks is functionally based on Orthogonal Frequency Division Multiplexing (OFDM) which is the foundation of emerging 4G wireless communication systems such as IEEE 802.16 WiMAX and LTE variants. Thus, results here for an existing 3G system directly support a broader research objective to developing improved IT security methods that are generally applicable to emerging 4G OFDM-based systems.

For initial proof-of-concept, the DE-optimized LFS classifier was used as the classification engine in an RF-DNA fingerprinting process adopted from previous work. Demonstration parameters for the DE process included a crossover threshold of $CR = 0.2$, crossover multipliers of $F_1{:}N(0,1)$ and $F_2 = 0.8$, $N_P = 40$ populations, and DE termination occuring after $N_{Gen} = 200$ generations. End-to-end intra-manufacturer device classification was performed using three like-model 802.11 Cisco devices. Monte Carlo simulation was implemented using 1) $N_B = 500$ experimentally collected bursts per device ($N_D = 3$ devices), 2) Time Domain (TD) and Spectral Domain (SD) fingerprint features from each burst, 3) $N_z = 10$ independent like-filtered AWGN realizations per burst at each SNR, and 4) $SNR \in [6, 24]$ dB.

Performance of the DE-optimized LFS classifier was compared with a Multiple Discriminant Analysis/Maximum Likelihood (MDA/ML) classifier as used previously for RF-DNA fingerprinting demonstration. Using *identical* TD and SD input features, DE-optimized LFS classification performance with SD features was inferior to MDA/ML for all SNRs considered and exhibited a sharp decrease in classification performance beginning at $SNR \approx 15$ dB. For TD features, the DE-optimized LFS classifier was generally superior to MDA/ML and achieved near-perfect 98% correct classification at lower noise-dominated SNRs ($SNR < 15$ dB). This included nearly 20% improvement at $SNR = 6$ dB. However, TD classification exhibited an anomalous "dip" in performance at higher signal-dominated SNRs with minimum classification falling to approximately 80% at $SNR = 21$ dB.

Subsequent analysis of *% Correct Classification* and *DE MSE* versus *Number of DE Generations* showed that both the inferior SD performance and anomalous TD performance at signal-dominated $SNRs$ were attributable to fixing DE termination at $N_{Gen} = 200$ generations–an insufficient number at some SNR for benefits of DE-optimized LFS to be fully realized. Preliminary analysis with increasing N_{Gen} suggests that *% Correct Classification* with TD features should approach 100% while performance with SD features will asymptotically approach an upper bound of approxi-

mately 86%. The disparity between achievable TD and SD performance is partially attributed to the factor of three disparity between the number of TD and SD features. A deeper exploration into the effects of N_{Gen}, the number of input features, and other parameters that were fixed for initial demonstration, is warranted and related research continues.

6. ACKNOWLEDGMENT

This work sponsored in part by the Sensors Directorate, Air Force Research Laboratory, Wright-Patterson AFB, OH.

7. REFERENCES

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

[2] M. A. Akbar and M. Farooq. Application of evolutionary algorithms in detection of SIP based flooding attacks. In *Proc of the 11th Annual conference on Genetic and evolutionary computation*, GECCO '09, pages 1419–1426, New York, NY, USA, 2009. ACM.

[3] J. Blau. Open-Source Effort to Hack GSM, spectrum.ieee.org/Telecom/Wireless/Open-Source-Effort-to-Hack-GSM. 2009.

[4] S. Bosworth and M. E. Kabay, editors. *Computer Security Handbook*. Wiley & Sons, fourth edition, 2002.

[5] V. Brik, S. Banerjee, M. Gruteser, and S. Oh. Wireless device identification with radiometric signatures. In *Proc of the 14th ACM international conference on Mobile computing and networking*, MobiCom '08, pages 116–127, New York, NY, USA, 2008. ACM.

[6] M. A. Buckner. *Learning From Data with Localized Regression and Differential Evolution*. PhD thesis, University of Tennessee, Knoxville, May 2003.

[7] 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.

[8] D. D. Capite. *Self-Defending Networks: The Next Generation of Network Security*. Cisco Press, 2006.

[9] V. S. Cherkassky and F. Mulier. *Learning from data: concepts, theory, and methods*. Wiley & Sons, Hoboken, HJ, 2nd edition, 2007.

[10] H. Collins. Top 10 network security threats. *Government Technology, www.govtech.com/security /Top-10-Network-Security-Threats.html*, Sep 2010.

[11] 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*, IPSN '09, pages 25–36, Washington, DC, USA, 2009. IEEE Computer Society.

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

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

[14] K. Kim and B.-R. Moon. Malware detection based on dependency graph using hybrid genetic algorithm. In *Proc of the 12th annual conference on Genetic and evolutionary computation*, GECCO '10, pages 1211–1218, New York, NY, USA, 2010. ACM.

[15] R. W. Klein, M. A. Temple, and M. J. Mendenhall. Application of wavelet-based RF fingerprinting to enhance wireless network security. *Jour of Communications and Networks: Secure Wireless Networking*, 11(6):544–555, Dec 2009.

[16] S. B. Mehdi, A. K. Tanwani, and M. Farooq. IMAD: in-execution malware analysis and detection. In *Proc of the 11th Annual conference on Genetic and evolutionary computation*, GECCO '09, pages 1553–1560, New York, NY, USA, 2009. ACM.

[17] J. Pang, B. Greenstein, R. Gummadi, S. Seshan, and D. Wetherall. 802.11 user fingerprinting. In *Proc of the 13th annual ACM international conference on Mobile computing and networking*, MobiCom '07, pages 99–110, New York, NY, USA, 2007. ACM.

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

[19] D. Reising, M. Temple, and M. Mendenhall. Improved wireless security for gmsk-based devices using RF fingerprinting. *Int. J. Electronic Security and Digital Forensics*, 3(1):41–59, Mar 2010.

[20] 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.

[21] D. Takahashi, Y. Xiao, Y. Zhang, P. Chatzimisios, and H.-H. Chen. Ieee 802.11 user fingerprinting and its applications for intrusion detection. *Computers & Mathematics with Applications*, 60(2):307 – 318, 2010. Advances in Cryptography, Security and Applications for Future Computer Science.

[22] T. D. Tarman and E. L. Witzke. Intrusion detection considerations for switched networks. *Enabling Technologies for Law Enforcement and Security*, 4232(1):85–92, 2001.

[23] J. Toonstra and W. Kinsner. A radio transmitter fingerprinting system ODO-1. In *Electrical and Computer Engineering, 1996. Canadian Conf on*, volume 1, pages 60 –63 vol.1, May 1996.

[24] W.C. Suski II, M.A. Temple, M. J. Mendenhall, and R.F. Mills. Radio frequency fingerprinting commercial communication devices to enhance electronic security. *Int. J. Electron. Secur. Digit. Forensic*, 1:301–322, Oct 2008.

[25] M. D. Williams, S. A. Munns, M. A. Temple, and M. J. Mendenhall. RF-DNA fingerprinting for airport WiMAX communications security. In *4th Int'l Conf on Network and Systems Security*, Sep 2010.

[26] 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.

[27] D. Zanetti, B. Danev, and S. Capkun. Physical-layer identification of UHF RFID tags. In *Proc of the sixteenth annual international conference on Mobile computing and networking*, MobiCom '10, pages 353–364, New York, NY, USA, 2010. ACM.