TD-DNA Feature Selection for Discriminating WirelessHART IIoT Devices

Christopher M. Rondeau
Department of Electrical and Computer Engineering
US Air Force Institute of Technology, WPAFB OH
christopher.rondeau@afit.edu

Michael A. Temple
Department of Electrical and Computer Engineering
US Air Force Institute of Technology, WPAFB OH
michael.temple@afit.edu

Christine Schubert Kabban
Department of Mathematics and Statistics
US Air Force Institute of Technology, WPAFB OH
christine.schubertkabban@us.af.mil

Abstract

The proliferation of Wireless Highway Addressable Remote Transducer (WirelessHART) communications in support of Industrial Internet of Things (IIoT) applications is accompanied by increased vulnerability concerns that amplify the need for improved pre-attack security and post-attack forensic methods. This paper summarizes demonstration activity aimed at applying Time Domain Distinct Native Attribute (TD-DNA) fingerprinting and improving feature selection to increase computational efficiency and the potential for near-real time operational application. Assessments include both pre-classification and post-classification dimensional reduction using TD-DNA fingerprint features extracted from experimentally collected WirelessHART signals. Results show that pre-classification selection methods are superior, with average percent correct classification differential of $8\% < \%C_{\Delta} < 1\%$ *being maintained using selected feature subsets containing only 24 (10%) of the 243 full-dimensional features.*

1. Introduction

The overwhelming focus in Internet of Things (IoT) growth has been in the so-called "Consumer IoT" [1] subset being used to connect an increasing number of household, personal, and consumer-level devices. While similar growth across the Industrial IoT (IIoT) subspace has been somewhat slower, some improved functionality and efficiency has been realized across nearly all industries [1]. This includes deployment of WirelessHART signaling which is "by far the largest digital communications technology deployed in the process industries with over 40 million field instruments supporting HART technology installed worldwide [2]. This includes support to over 24,000 WirelessHART networks that logged over 5 billion operating hours [3]. While estimates for the number of currently fielded WirelessHART networks and devices vary, there are predictions of exponential growth through 2028 within the oil, gas, chemical, and power generation industries [4].

Whether addressing pre-attack defense or post-attack forensic analysis, the preponderance of threat detection and protection work in process control systems occurs above the PHY layer [5], including bit-level solutions implemented in the upper communication protocol layers [2, 6, 7]. Relative to IIoT vulnerability, the U.S. Industry Control System (ICS) Cyber Emergency Response Team (ICS-CERT) indicated that, "The gateway [of an ICS system] … is where you need to pay the most attention" [8]. This certainly includes all PHY layer communications between a process sensor and the gateway. Therefore, to realize the cross-layer security benefits envisioned in [5], the desirable architecture would include the ability to operate across all IIoT elements by balancing available resources to exploit information at the most vulnerable nodes and achieve an acceptable level of threat warning.

One PHY-based method supporting offensive, defensive, and exploitive network operations is Time Domain Distinct Native Attribute (TD-DNA) Fingerprinting which has been successfully used to discriminate IoT and IIoT communication devices and their operating states [9-18]. The TD-DNA fingerprinting methodology therein is well-suited for consideration here given 1) the observed ZigBee-like signal characteristics of WirelessHART signals, and 2) the ability to perform Dimensional Reduction Analysis (DRA) and identify the minimal subset of features required to achieve a given level of discrimination performance.

The work here extends first-look results presented in [18] that included ZigBee device discrimination and DRA with feature selection using 1) an adopted post-classification Random Forest (RndF) relevance ranking method, and 2) a newly developed pre-classification Wilcoxon Rank Sum (WRS)

URI: https://hdl.handle.net/10125/64524
978-0-9981331-3-3

method based on nonparametric statistical testing. Specific extension includes 1) transition to WirelessHart and ZigBee-Like signal processing, and 2) use of additional DRA methodologies with feature selection based on both the adopted post-classification Generalized Relevance Learning Vector Quantized Improved (GRLVQI) method, and the adopted pre-classification ReliefF method.

This paper is organized as follows. Section 2 provides background information on the IIoT Threat Framework, WirelessHART, TD-DNA Fingerprinting, MDA/ML Discrimination, and DRA. Section 3 details the demonstration methodology used to generate Section 4 DRA Performance Results. The paper concludes with Section 5 summary and conclusions.

2. Background

2.1. IIoT Threat Framework

A majority of IIoT risk mitigation effort has been dedicated to threats posed by actors having malicious intent, especially when those IIoT devices support critical infrastructure (CI) elements. However, when considering CI and IIoT in general, there are threats that do not originate from malicious actors that can have similarly catastrophic effects. As reflected in Fig. 1, the threat categorization in [19] sufficiently embodies IIoT threats under three main categories: accidental, malicious, and natural.

Attack effects within the IIoT threat framework are captured in ICS Impact subcategories that have been added here and shown in Fig. 1. These include both 1) *Incidental* impact (e.g., Slammer Worm at the Davis-Besse nuclear plant [20]), and 2) *By-Design* impact where specific IIoT element vulnerabilities are targeted (e.g., Tehama Colusa Canal sabotage [20], Stuxnet [21], Shamoon [22], and CrashOverride [23]). Both impact categories include insider attacks of given elements that can be directly accessed, with responsible Agent(s) including human, technological, and natural actors [19].

The By-Design impact in Fig. 1 includes creation of malware like TRITON/TRISIS which aims to degrade safety interrupt systems whereby "persons, property, and/or the environment could suffer physical harm" [24]. While malicious threats such as these have potentially catastrophic effects, the threat framework in Fig. 1 also illuminates the fact that the impact of accidental or natural category threats could have equally serious consequences.

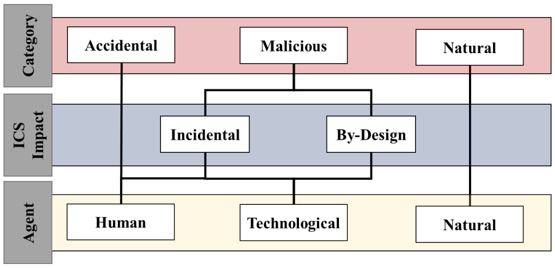


Figure 1. IIoT threat framework with categorization of IIoT elements and agents from [19] and *Incidental* and *By-Design* ICS-impact elements added here.

2.2. WirelessHART Signaling

WirelessHART is a variant of the wired HART protocol used to exchange information via 4-20 mA current loop signaling that exists in nearly all legacy systems. WirelessHART adapters support exchange of the legacy 4-20 mA current information using wireless communications in the 2.4 GHz band. The signaling is compliant with IEEE 802.15.4 PHY layer standards and possess ZigBee-like characteristics that have been successfully exploited in previous work [17,18].

WirelessHART devices from two suppliers were considered here for demonstration, including the Siemens AW210 [25] and Pepperl and Fuchs Bullet [26] devices. The 802.15.4 PHY layer operation for these devices is shown in Fig. 2. As common in many wireless protocols, the transmitted bursts include a preamble response which was observed to be the first 160 μSec (defined as the *PreAmbRgn* in Fig. 2). The *PreAmbRgn* is the primary Region of Interest (ROI) exploited here for TD-DNA Fingerprinting.

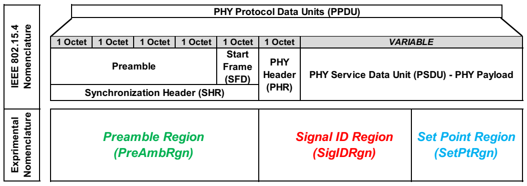

IEEE 802.15.4 Nomenclature
PHY Protocol Data Units (PPDU)
1 Octet
1 Octet
1 Octet
1 Octet
1 Octet
1 Octet
VARIABLE
Preamble
Start Frame (SFD)
PHY Header (PHR)
PHY Service Data Unit (PSDU) - PHY Payload
Synchronization Header (SHR)
Exprimental Nomenclature
Preamble Region (PreAmbRgn)
Signal ID Region (SigIDRgn)
Set Point Region (SetPtRgn)

Figure 2. WirelessHART IEEE 802.15.4 details [27] with 1) IEEE nomenclature and prescribed signal durations, and 2) experimental nomenclature and observed signal durations under this work.

As done previously for ZigBee devices in [14, 17, 18], WirelessHART TD-DNA features were extracted from experimentally collected data for $N_D = 8$ like-model devices. The signals were collected with a National Instruments N2952 Software Defined Radio (SDR) having an RF bandwidth of $W_{RF} = 10$ MHz and operating at a sample frequency of $f_S = 10$ MSps in both the

In-phase and Quadrature-phase (I/Q) channels. Post-collection MATLAB processing was performed on a burst-by-burst basis and included baseband down-conversion, followed by filtering with a 4th-order Butterworth filter having a bandwidth of $W_{BB} = 1$ MHz. The collections were made in a typical office environment (channel conditions consistent with WirelessHART applications) which yielded an average collected Signal-to-Noise Ratio of $SNR_C = 39.0$ dB. The collected signals were SNR-scaled by adding independent, like-filtered, power-scaled Additive White Gaussian Noise to reflect operating conditions for $SNR \in [5.0, 39.0]$.

2.3. TD-DNA Fingerprinting

TD-DNA Fingerprinting utilizes various machine learning algorithms and concepts such as feature selection via DRA which has been the subject of prior related works [9, 18, 28]. The demonstration here focuses on examining DRA methods for use with an MDA/ML classifier, given the MDA/ML classification process has been shown to be computationally efficient while reliably discriminating IIoT signals [9, 13].

Results here are based on TD-DNA fingerprints generated from WirelessHART burst preamble responses and are generated using a methodology consistent with prior related works [9-18]. The instantaneous amplitude (AMP), phase (PHZ), and frequency (FRQ) responses of the PreAmbRgn ROI are divided into $N_R = 26$ subregions. All ROI samples are used for generating features as well, for a total of $N_R = 26{+}1 = 27$ fingerprinting regions. A total of $N_{Stat} = 3$ statistics of variance (σ^2), skewness (γ), and kurtosis (κ) are computed within each region to form,

$$\boldsymbol{F^{Stat}} = [\boldsymbol{\sigma^2} \,\vdots\, \boldsymbol{\gamma} \,\vdots\, \boldsymbol{\kappa}]_{\mathbf{1\times 3}}\,, \tag{1}$$

where $\vdots$ denotes concatenation. Accounting for the $N_R{+}1 = 27$ regions and each instantaneous response, the *Regional Statistic Vector* is given by,

$$\boldsymbol{F^{Rgn}} = \left[\boldsymbol{F^{Stat}_{R1}} \,\vdots\, \boldsymbol{F^{Stat}_{R2}} \,\vdots\, \cdots \,\vdots\, \boldsymbol{F^{Stat}_{N_R+1}}\right]_{\mathbf{1\times[3(N_R+1)]}}\,. \tag{2}$$

The regional vectors are used to form the *Composite TD Fingerprint Vector* given by,

$$\boldsymbol{F_{TD}} = \left[\boldsymbol{F^{Rgn}_{AMP}} \,\vdots\, \boldsymbol{F^{Rgn}_{PHZ}} \,\vdots\, \boldsymbol{F^{Rgn}_{FRQ}}\right]_{\mathbf{1\times N_F}}\,, \tag{3}$$

where N_F is the total number of features. For three instantaneous response with $N_R{+}1 = 27$ and $N_{Stat} = 3$ statistics, the full-dimensional set of fingerprint features considered here includes $N_{FD} = 243$ features.

2.4. MDA/ML Processing

The MDA/ML processing used here is a readily implementable, computationally efficient process that has provided reliable device discrimination in prior TD-DNA works [9, 11-18]. As detailed in [18], MDA is effectively a feature selection process that performs best when input class features and their corresponding projections are Gaussian distributed. The process includes generation of projection matrix **W** with a goal of maximizing between-class separation (projected class means) while minimizing within-class spread (projected class variance). For discriminating N_{Cls} classes using input fingerprints (**F**) having N_F features, **W** has dimension N_F x (N_{Cls}-1) and effectively projects (1xN_F)-dimensional **F** into the (N_{Cls}-1)-dimensional decision space.

Given a trained MDA "model" (projection matrix **W**, input fingerprint scale factors, projected class training means, and projected class training variances) a 1 vs. N_{Cls} called-class estimate (correct or incorrect) for an input "unknown" testing fingerprint $\mathbf{F}_{\mathrm{Tst}}$ is defined as $\boldsymbol{F^{W}_{Tst}} = \boldsymbol{F_{Tst}}\mathbf{W}$, where $\boldsymbol{F^{W}_{Tst}}$ is the projection of $\mathbf{F}_{\mathrm{Tst}}$ in the Fisher space. The classification estimate is based on the conditional probability relationship in the Fisher projection space. Assuming equal probability of class occurrence and equal error costs, the probability relationship becomes a Maximum Likelihood (ML) estimate. The class yielding the highest probability becomes the called-class for $\boldsymbol{F^{W}_{Tst}}$.

Overall *cross-class percent correct classification* (*%C*) is calculated as the percent of correct called-class estimates from the total number of classification decisions. Given that the classification decisions represent independent Monte Carlo trials, 95% Confidence Interval ($\mathrm{CI}_{95\%}$) analysis consistent with [9] is used for comparative (best, same, different, etc.) assessments. For visual clarity, the $\mathrm{CI}_{95\%}$ intervals are intentionally omitted from figures, and the vertical extent of data markers appropriately sized such that they encompass the $\mathrm{CI}_{95\%}$ intervals. Thus, overlapping data markers represent statistically identical and/or indeterminate performance and non-overlapping data markers represent statistically different performance.

2.5. DRA Feature Selection

With the surge of available data for machine learning applications, there has been renewed interest in DRA as a means to reduce the scale of the input data to a manageable size [29]. As depicted in Fig. 3, the feature selection aspect of DRA may be categorized as using label information (supervised, semi-supervised, and unsupervised) and selection strategies (filter, wrapper, and embedded) [29, 30, 31].

The methods here use supervised approaches, i.e., labeled data, whereas semi-supervised and unsupervised approaches use partially-labeled or unlabeled data, respectively [29]. The DRA methods here include two selection strategies that include filter

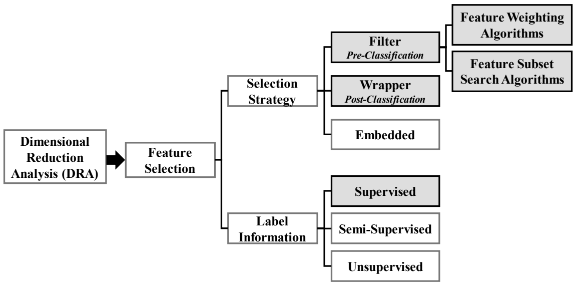


Figure 3. Feature selection categories from [29] and [30]. Shading indicates areas covered here.

(*pre-classification*) and wrapper (*post-classification*). Filter strategies most often employ *pre-classification* ranking and statistical techniques [31], have low computational cost, and are well-suited for higher dimensional data [30].

Wrapper methods differ in that they "use the intended learning algorithm itself to evaluate the features" [29] and are optimized for that learning algorithm [30]. While this "optimization" could be a strength, wrapper methods are limited since they are only intended to work with that same learning algorithm, and therefore may suffer from overfitting [31]. The selection strategies considered here include two *pre-classification* filtering and two *post-classification* wrapper approaches. All methods produce an output vector of N_F sorted features with associated weights. A brief summary of each DRA method is provided in the following sections.

2.5.1. Post-Classification RndF

The Random Forest (RndF) classifier includes an ensemble of single decision tree classifiers that collectively produce a single classification decision and provide a feature relevance metric [32]. Among the hyperparameter tuning for RndF, there are two fundamental parameters that affect classifier performance, including 1) the number of decision trees (classifiers), and 2) the number of predictors (features) sampled at each node. The classifier considers all available features at the initial node, then makes subsequent splits based on a random predictor selection and threshold values at each child node. All features are not considered at each of the child nodes and the selection is done with replacement, thus a feature may be used as a splitting criterion at multiple nodes. The features selected as the splitting criterion include those producing the largest change in Gini-Index (G_I) [9]. The RndF process supports *post-classification* DRA by providing a mean decrease in the G_I metric [9, 10] that is computed for the k^{th} feature by averaging the change in G_I each time the k^{th} feature is used at a splitting decision.

2.5.2. Post-Classification GRLVQI

General Relevance Learning Vector Quantization Improved (GRLVQI) is an extension of Kohonen's Learning Vector Quantization (LVQ) [33] which is in the family of self-organizing Neural Network (NN) approaches using nearest Prototype Vector (PV) optimization. As a classifier, LVQ associates a PV with a given class (typically multiple PVs per class). When an observation is input to the network, the PV closest to the observation "fires" and the prediction accuracy is based on whether or not the firing PV(s) are associated with the correct class for the observation. GRLVQI extends LVQ by incorporating cost functions, learning methods, and logic and operation improvements [33]. There are five fundamental hyperparameters for GRLVQI processing, including the gradient descent learning rate, relevance learning rate, conscience rate(s), and the number of class PVs. The correct model construction requires expertise or appropriate heuristics [33].

2.5.3. Pre-Classification WRS

The Wilcoxon Rank Sum (WRS) method is a nonparametric statistical approach and is therefore unconstrained in terms of the assumptions required for parametric statistical methods, e.g., normality of the underlying distribution. The WRS requires only that the samples are independent and from a continuous distribution [34]. The output of the WRS test is a determination as to whether or not two observations are from distributions with equal medians, regardless of the exact nature of the two underlying distributions. As a DRA method, the WRS test was developed in [18] and is utilized to comapre a feature across all classes for a given classification problem. A feature is considered more relevant the more instances the WRS test concludes that the cross-class comparison has a different median. The feature relevance is computed as a two factor product: 1) the raw count of WRS test failures and 2) the entropy of the repsective p-values (an aggregate measure of confidence of each separate WRS test). The only parameter to modify is the α-value for the statistical test.

2.5.4. Pre-Classification ReliefF

ReliefF processing is derived from the Relief algorithm developed in [35, 36]. Relief is an instance-based learning algorithm that was originally conceived to implement a statistical approach to feature selection (as opposed to a purely heuristic search) resulting in improved learning time and accuracy compared to other feature selection methods [36]. Relief picks a sample of m triplets from the total N_F features and computes a Euclidian distance-based comparison metric. A feature weight vector is routinely updated as the algorithm runs. Similar to the *pre-classification* WRS method, there is one parameter to modify, *k*, which corresponds to the number of nearest neighbors considered during comparison. ReliefF overcomes the noted limitations in [35, 36] by modifying the Relief algorithm to allow for incomplete data sets and $N_{Cls} > 2$ classes [37, 38].

3. Demonstration Methodology

Comparison of the four DRA feature selection methods included the following steps:

3.1. Step 1: Analysis SNR Selection

For all DRA selection methods, proper subsets of $N_{DRA} < N_{FD}$ full-dimensional fingerprint feaures were selected and classification performed. Given that feature relevance is SNR dependent, the specific SNR used for DRA assessments was selected by considering the average percent correct (*%C*) using the N_{FD} feature set with the MDA/ML, RndF, and GRLVQI classifiers. All subsequent classification performance results, with the exception of those presented in Fig. 4, are for MDA/ML classification. The analysis SNR values are selected based on the Fig. 4 results which show *%C* performance for all three classification methods using the full-dimensional N_{FD} features set.

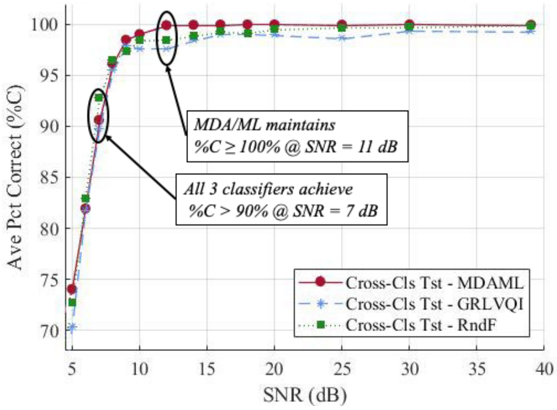


Figure 4. Classification performance of MDA/ML, RndF, and GRLVQI using the full-dimensional $N_{FD} = 243$ feature set.

The results in Fig. 4 did not use "optimized" RndF nor the GRLVQI processes; the results were generated using empirical hyperparameter settings from prior wireless signal discrimination work similar to WirelessHART. Collectively, all three classifiers achieve the $\%C \geq 90\%$ benchmark for $SNR \geq 7$ dB and the MDA/ML classifier maintains $\%C \geq 99\%$ for $SNR \geq 14$ dB. Therefore, results of RndF and GRLVQI processes were used for DRA feature selection based on rank-ordering feature relevances at $SNR = 7$ dB and $SNR = 14$ dB.

3.2. Step 2: Qualitative DRA Application

Qualitative DRA is performed using $N_{DRA} = N_{FD}/3$ (66% reduction) feature subsets denoted as AMP-Only, PHZ-Only, and FRQ-Only DRA sets. Segmentation of the N_{FD} features into response-centric subsets is accomodated by the sequential construct of $\boldsymbol{F_{TD}}$ fingerprint elements in (3). Qualitative DRA assessment was used for identifying the most useful response features and for comparison with quantative DRA selection in Section 3.3. Segmentation of a representative TD-DNA fingerprint into AMP-Only, PHZ-Only, and FRQ-Only subsets is shown in Fig. 5 for fingerprints of $N_{FD} = 234$ features.

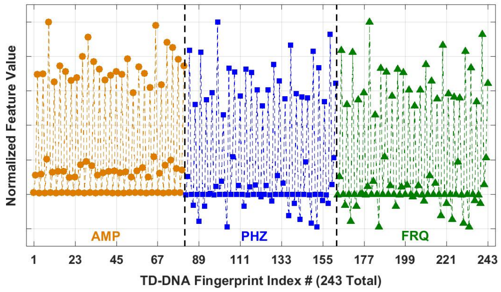
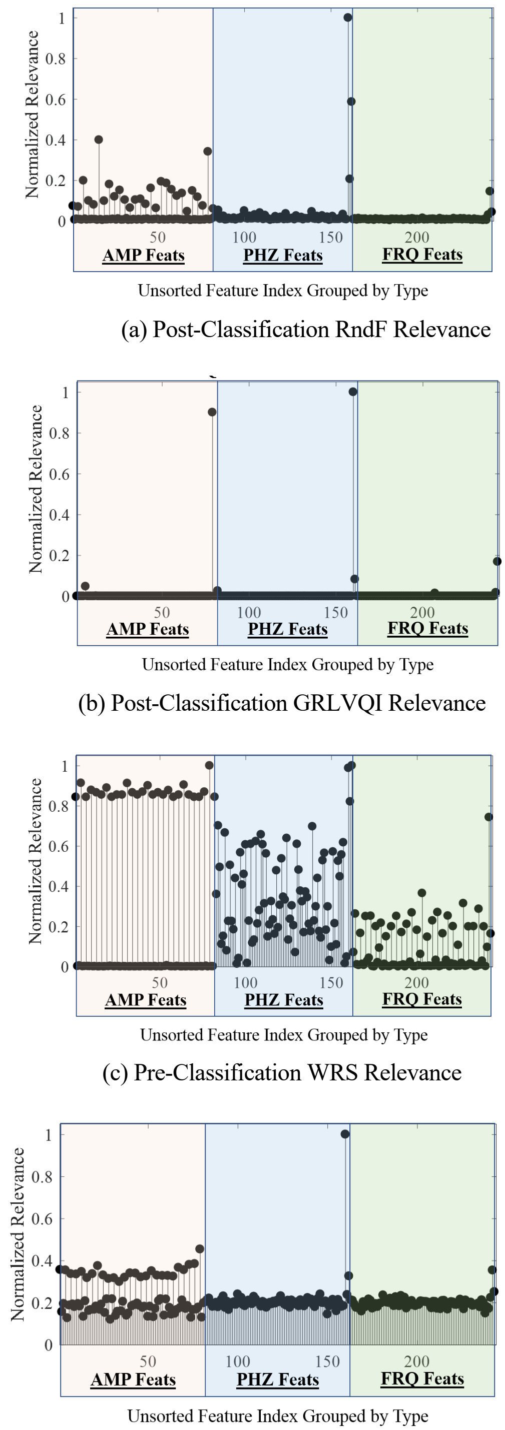


Figure 5. Representative TD-DNA fingerprint from (3) showing the relative location of feature regions using normalized feature values.

3.3. Step 3: Quantative DRA Application

Quantitative DRA is based on feature relevance, such as presented in Fig. 6 for the full-dimensional $N_{FD} = 243$ feature set at $SNR = 7$ dB. This figure shows normalized relevance weighting versus unsorted feature index number for *post-classification* RndF (Fig. 6a) and GRLVQI (Fig. 6b) processes, and *pre-classification* WRS (Fig. 6c) and ReliefF (Fig. 6d) processes. The subplots also show the relationship of quantitative feature relevance to the qualitative index boundaries used for selecting the AMP-Only, PHZ-Only, and FRQ-Only features considered in Section 3.2.

For DRA feature selection, the unsorted relevance metrics in Fig. 6 are rank-ordered (sorted highest to lowest) for each of the DRA methods. These higher-is-more-relevant metrics show that 1) the RndF and GRLVQI plots decrease rapidly (lower number of more relevant features) and become near-zero, whereas 2) the WRS and ReliefF plots decrease less rapidly (higher number of more relevant features) with fewer (WRS) or no (ReliefF) near-zero relevant features.



(a) Post-Classification RndF Relevance

(b) Post-Classification GRLVQI Relevance

(c) Pre-Classification WRS Relevance

(d) Pre-Classification ReliefF Relevance

Figure 6. Normalized relevance versus unsorted feature index number for full-dimensional $N_{FD} = 243$ feature set at $SNR = 7$ dB. The AMP-Only, PHZ-Only, and FRQ-Only qualitative DRA feature regions are shown for comparison.

3.4. Step 4: MDA/ML Classification

MDA/ML classification performance was assessed using the *qualitative* $N_{DRA} = 81$ AMP-Only, PHZ-Only, and FRQ-Only DRA subsets from Step 2 and *quantitative* $N_{DRA} \in \{0.10, 0.15, \ldots, 0.50\} \mathrm{x} N_{FD}$ DRA subsets selected using the sorted feature rankings from Step 3. That is, using the top-ranked 10%, 15%, …, 50% of the $N_{FD} = 243$ of the full-dimensional features. Performance is assessed relative to the $\%C \geq 90\%$ benchmark for individual DRA subsets as well as accumulated averages of $\%C$ for *pre-classification* and *post-classification* methods.

4. DRA Performance Results

4.1. Qualitative DRA Feature Selection

MDA/ML classification using *qualitatively* selected DRA feature sets are shown in Fig. 7 and are useful for identifying the response(s) (AMP, PHZ, or FRQ) that contributes most to classification accuracy. The plots show 1) full-dimensional $N_{FD} = 243$ results from Fig. 4, overlaid with 2) results using the $N_{DRA} = 81$ (66% of the N_{FD}) qualitatively selected AMP-Only, PHZ-Only, and FRQ-Only subsets.

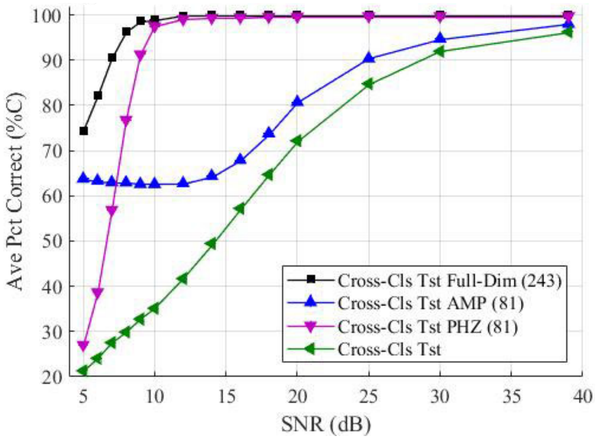


Figure 7. MDA/ML classification performance for $N_{FD} = 243$ full-dimensional and *qualitatively* selected $N_{DRA} = 81$ feature subsets showing that PHZ-Only features are dominant.

The results from Fig. 7 suggest DRA feature selection at $SNR = 7$ dB to be a point of interest since it corresponds to the $\%C \approx 90\%$ point from Fig. 4 and that despite the overall trend, AMP-Only features produce a higher $\%C$ than PHZ-Only. Similarly, $SNR = 14$ dB is shown to be a point of interest because the PHZ-Only features dominate in $\%C$ and produce statistically similar results to the full-dimensional set.

Table 1. Comparison of Fig. 7 *%C* performance at *SNR* = 7 dB and *SNR* = 14 dB for full-dimensional and qualitatively selected DRA feature sets.

Qualitative DRA Subset	*%C* @ *SNR* = 7 dB	*%C* @ *SNR* = 14 dB
AMP	63.00%	64.66%
PHZ	56.92%	99.35%
FRQ	27.55%	49.51%
Full-Dim	90.50%	99.96%

Table 1 shows a comparison of Fig. 7 results at *SNR* = 7 dB and *SNR* = 14 dB. At *SNR* = 7 dB, the best case DRA performance is *%C* ≈ 63.0% for AMP-Only feature subset, and second best *%C* ≈ 56% using the PHZ-Only subset. By comparison with full-dimensional performance of *%C* ≥ 90%, there is a clear loss in discriminating "information" using the DRA feature subsets. This is potentially attributable to a loss in AMP-PHZ-FRQ feature synergism or simply using considerably fewer (66% of the full-dimensional) features.

The noted disparity in Table 1 results is examined further in the quantitative DRA results in Section 4.2. At *SNR* = 14 dB, the best case DRA performance is *%C* ≈ 99.35% for PHZ-Only features, nearly the same as the full-dimensional set with *%C* ≈ 99.96%. This indicates at this SNR, there is little information gained from adding in the remaining features to a subset.

4.2. Quatitative DRA Feature Selection

MDA/ML classification results for *quantitatively* selected DRA feature sets are shown using *SNR* = 14 dB and *SNR* = 7 dB. DRA feature selection was based on the rank-ordered feature relevance plots like those shown in Fig. 6. Of note, the Fig. 6a (RndF) and Fig. 6b (GRLVQI) *post-classification* method plots clearly reflect a steep drop-off in feature relevance with increasing sorted index number. This differs from the *pre-classification* sorted relevances in Fig. 6c (WRS) and Fig. 6d (ReliefF) which do contain observable "breaks" that could serve as DRA selection criteria. As noted in Step 4 of Section 3.4, the *quantitative* $N_{DRA} \in \{0.10, 0.15, \ldots, 0.50\}xN_{FD}$ DRA subsets were selected using sorted feature rankings to identify the top-ranked 10%, 15%, …, 50% of the N_{FD} = 243 of the full-dimensional features at both *SNR* = 7 dB and *SNR* = 14 dB.

Classification performance for $24 \leq N_{DRA} \leq 122$ features (approximately 10% to 50% of N_{FD}) are shown in Fig. 8 for *SNR* = 14 dB. Performance of the N_{FD} = 243 full-dimensional *%C* = 99.96% (dashed line) is also provided for reference. Based on $CI_{95\%}$ analysis, these results show that all methods except GRLVQI achieve statistically similar classification as the full-dimensional set. The RndF method achieves *%C* = 99.96% at N_{DRA} = 97 features whereas the two *pre-classification* methods achieve it at N_{DRA} = 122 features. The largest deviation in *%C* from the full-dimensional set occurs at N_F = 24 features (approximately 10% of N_{FD}). Even with only 10% of the full-dimensional set, all four methods are approximately within *%C* ≈ 1%.

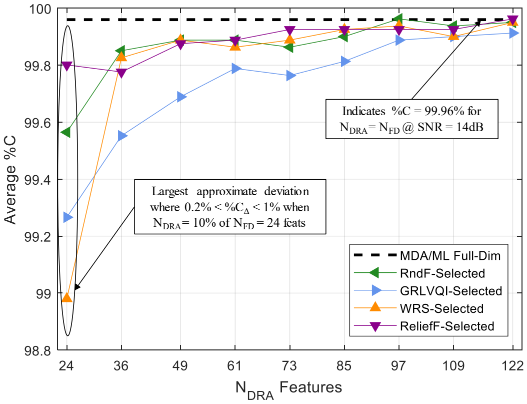


Figure 8. MDA/ML classification versus N_{DRA} with N_{FD} = 243 full-dimensionl reference for *SNR* = 14 dB and *quantitatively* selected DRA subsets.

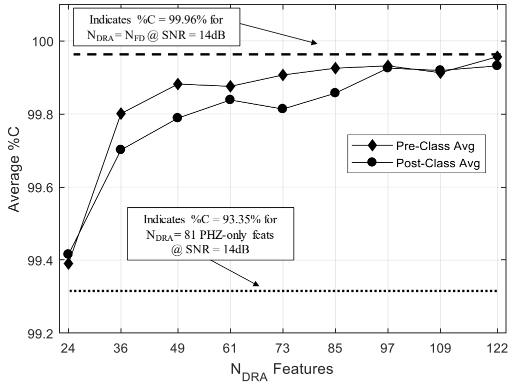


Figure 9. MDA/ML classification using full dimensional features and collective DRA performance aaverages for *SNR* = 14 dB calcuated for 1) two *pre-classification* (WRS and ReliefF) methods, and 2) two *post-classification* (RndF and GRLVQI) methods, with *pre-classification* being generally superior.

The cross-method averages for individual method results in Fig. 8 are presented in Fig. 9 to enable a general assessment of *post-classification* versus *pre-classification* DRA selection methods. For reference, the plot also includes the N_{FD} = 243 *%C* = 99.96% (upper dashed line) and N_{DRA} = 81 best case qualitative PHZ-Only *%C* = 99.35% (lower dashed line) performances. As Fig. 9 results show, the *pre-classification* selection methods outperform the *post-classification* selection methods when using up to 97 of the 122 features.

Classification results for $N_{DRA} = 24$ features (10% of N_{FD}) to $N_{DRA} = 122$ features (50% of N_{FD}) are shown in Fig. 10 for $SNR = 7$ dB. The $N_{FD} = 243$ full-dimensional $\%C = 90.5\%$ (dashed line) is provided for reference. These results show none of the DRA subsets achieve full-dimensional performance and there is no single method that consistently outperforms the others for all N_{DRA} considered. As indicated, the WRS-selected feature performance is 1) best at the two higher N_{DRA} values, and 2) consistent with RndF, WRS, and ReliefF for a majority of N_{DRA} considered. The GRLVQI selected subsets are is the overall poorest.

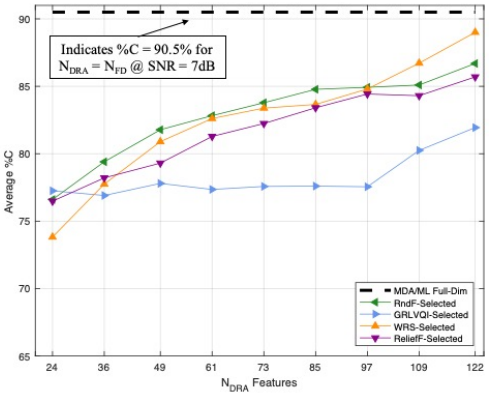


Figure 10. MDA/ML classification versus N_{DRA} showing the $N_{FD} = 243$ full-dimensionl reference for $SNR = 7$ dB and *quantitatively* selected DRA subsets.

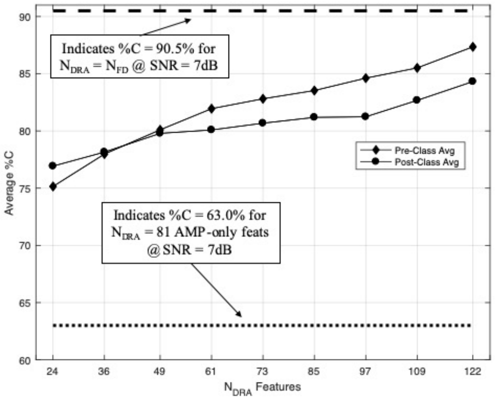


Figure 11. Full-dimensional MDA/ML performance and collective DRA averages for $SNR = 7$ dB calculated across the 1) two *pre-classification* (WRS and ReliefF) methods, and 2) two *post-classification* (RndF and GRLVQI) methods. Averages indicate that *pre-classification* is generally superior.

The cross-method averages for individual method results in Fig. 10 are presented in Fig. 11 to enable a general assessment of *post-classification* versus *pre-classification* DRA selection methods. For reference, the plot also includes the $N_{FD} = 243$ full-dimensional $\%C = 90.5\%$ (upper dashed line) and $N_{DRA} = 81$ best case qualitative AMP-Only $\%C = 63.0\%$ (lower dashed line) performances. As Fig. 11 results show, the *pre-classification* selection methods outperform the *post-classification* selection methods for all but the lowest $N_{DRA} = 24$ subset considered. However, it is obvious from Fig. 10 that GRLVQI performance is the major contributor to the poorer *post-classification* average.

4.3. DRA Method Analysis

Previous section results demonstrate that the quantitative DRA methods used here are valid for feature selection when comparing their performance to qualitatively selected $N_{RA} = 81$ feature subsets (Fig. 7) and corresponding $N_{FD} = 243$ full-dimensional feature set. The collective cross-method average performance of *pre-classification* methods in Fig. 9 and Fig. 11 reflect superior performance relative to the *post-classification* selection methods. Individual method results when features are selected at $SNR = 14$ dB in Fig. 8 show all methods except GRLVQI achieve performance statistically equal to the full-dimensional set. Furthermore, at only 10% of the full-dimensional set, the maximum performance deviation is $\%C_{\Delta} \approx 1\%$ which is still $\%C \geq 99\%$. When selecting features at $SNR = 7$ dB, Fig. 10 shows that no DRA method achieved performance statistically equivalent to $\%C = 90.5\%$ using the $N_{FD} = 243$ set.

Regardless of the N_{DRA} value, the DRA method, or the SNR value considered, all quantitative DRA methods outperform qualitative DRA. This supports the notion that quantitative DRA improves the performance as compared to simple qualitative DRA through the selection of a more relevant feature subset. Between the two sets of quantitative DRA methods, *pre-classification* methods are generally superior.

To fairly represent the *post-classification* methods, it is important to note that the RndF and GRLVQI results presented were generated using empirical hyperparameter values from prior wireless signal discrimination work and not necessarily optimized for the WirelessHART application. Therefore, without hyperparameter tuning it is unknown if the RndF and GRLVQI results are representative of their best performance. In terms of DRA applications, however, the very fact that the *post-classification* methods have hyperparameters that require "tuning" adds a degree of feature selection complexity that the *pre-classification* methods do not possess. With that consideration, the fact that both *pre-classification* methods considered produce *%C* results that are better than, statistically equivalent to, or consistent with (i.e., within $\%C_\Delta \approx 4\%$ for $SNR = 7$dB), the best performing RndF *post-classification* method over the range of N_{DRA} investigated suggests their computational advantage may outweigh any realized performance gain.

5. Summary and Conclusions

Security within the IIoT domain poses certain challenges and PHY-based protection mechanisms remain largely unexploited. This includes security of WirelessHART signaling which is the largest digital communications technology deployed in process control industries, including over 40 million fielded devices [2]. The challenges are further increased when considering projections that predict exponential WirelessHART growth through 2028 within the oil, gas, chemical, and power generation industries [4].

PHY layer information may be reliably extracted from various elements within the IIoT infrastructure and support cross-layer security architectures [5] providing timely and reliable defensive, offensive, and exploitive actions. The extraction of useful PHY layer information is addressed here using Time Domain Distinct Native Attribute (TD-DNA) features from WirelessHART signals. Specific emphasis is place on Dimensional Reduction Analysis (DRA) methods with a goal of improving computation efficiency and moving closer to near-real time implementation by reducing the number of fingerprint features required to achieve desired discrimination performance.

The fingerprint DRA methods considered here include 1) two *post-classification* RndF and GRLVQI processes, and 2) two *pre-classification* WRS and ReliefF statistical analysis methods. Collective performance of *pre-classification* DRA methods was superior to *post-classification* methods, with average correct percent classification (*%C*) being 1) within $8\% < \%C_\Delta < 3\%$ of full-dimensional (243 features) $\%C = 90\%$ performance at $SNR = 7$ dB using only 24 of 243 (~10%) and 122 of 243 (~50%) features, respectively, and 2) within $\%C_\Delta \approx 1\%$ of full-dimensional $\%C = 99\%$ at $SNR = 14$ dB. While some *%C* trade-off is expected and observed, the DRA methods considered enable reliable feature selection (reduction). This in-turn increases computational efficiency and the potential for faster TD-DNA fingerprinting in operational applications.

6. Acknowledgment

The views in this paper are those of the authors and do not reflect the official policy or position of the Air Force Institute of Technology, the Department of the Air Force, the Department of Defense, or the US Government. This paper is approved for public release, Case#: 88ABW-2019-2822.

References

[1] S. Schneider, "Internet of Things and Data Analytics Handbook," Hoboken, NJ: John Wiley & Sons. 2017.

[2] FieldComm Group, "HART-Digital Transformation for Analog Instruments," [Online]. Available: https://fieldcommgroup.org/technologies/hart, Accessed: 21 Aug 2019.

[3] C. Wu, et. al., "Maximizing Network Lifetime of WirelessHART Networks under Graph Routing," *First Int'l Conf on Internet-of-Things Design and Implementation* (IoTDI), Apr. 2016.

[4] Global Control, "WirelessHART: Proven and Growing Technology with a Promising Future, [Online]. Available: https://www.controlglobal.com/articles/2018/ wirelesshart-proven-and-growing-technology-with-a-promising-future/. Accessed: 21 Aug 2019.

[5] C. M. Rondeau, M.A. Temple, J. Lopez, "Industrial IoT Cross-layer Forensic Investigation," *WIRES Forensic Sci*., Vol. 1, No. 1, doi.org/10.1002/wfs2.1322, 2019.

[6] T. Lennvall, S. Svensson, F. Hekland, "A Comparison of Wirelesshart and ZigBee for Industrial Applications," *2008 IEEE Int'l Workshop on Factory Communication Systems*, 2008.

[7] K. Stefanidis, A.G. Voyiatzis, "An HMM-Based Anomaly Detection Approach for SCADA Systems," 2016. [Online]. Available: https://link.springer.com/chapter/ 10.1007/978-3-319-45931-8_6. Accessed: 21 Aug 2019.

[8] M. Edwards, "S4X19 - OnRamp Know Your ICS," presented at S4x19, Miami, FL, 2019. [Online]. Available: https://tinyurl.com/yyapkbxe. Accessed: 21 Aug 2019.

[9] J. Lopez, N.C. Liefer, C.R. Busho, M.A. Temple, "Enhancing Critical Infrastructure and Key Resources (CIKR) Level-0 Physical Process Security Using Field Device Distinct Native Attribute Features," IEEE Trans on Information Forensics and Security, Vol. 13, No. 5, May 2018, pp. 1215-1229.

[10] H.J. Patel, M.A. Temple, R.O. Baldwin, "Introduction of a Random Forrest Classifier to ZigBee Device Network Authentication Using RF-DNA Fingerprinting," Jour of Information Warfare, Vol. 13, No. 3, 2014, pp. 33-45.

[11] D.R. Reising, M.A. Temple, J.A. Jackson, "Authorized and Rogue Device Discrimination Using Dimensionally Reduced RF-DNA Fingerprints," IEEE Trans on Information Forensics and Security, Vol. 10, No. 6, 2015, pp. 1180-1192.

[12] S.J. Stone, M.A. Temple, R.O. Baldwin, "Detecting Anomalous PLC Behavior Using RF-Based Hilbert Transform Features and a Correlation-Based Verification Process," *Int'l Jour of Critical Infrastructure Protection*, Vol. 9, Issue C, Jun. 2015, pp. 41-51.

[13] C.M. Talbot, M.A. Temple, T.J. Carbino, A.J. Betances, "Detecting Rogue Attacks on Commercial Wireless Insteon Home Automation Systems," Jour of Comp. and Sec., Vol. 74, May 2018, pp. 296-307.

[14] C.K. Dubendorfer, B.W. Ramsey, M.A. Temple, "ZigBee Device Verification for Securing Industrial Control and Building Automation Systems," Critical Infrastructure Protection VII, IFIP Advances in Information and Communication Technology, 2013, pp. 47-62.

[15] S.J. Stone, M.A. Temple, "Radio-Frequency-Based Anomaly Detection for Programmable Logic Controllers in in Critical Infrastructure Apps," *Int'l Jour of Critical Infrastructure Protection*, Vol. 5, Issue 2, Jul. 2012, pp. 66-73.

[16] J. Lopez, et al., "Exploitation of HART Wired Signal Distinct Native Attribute (WS-DNA) Features to Verify Field Device Identity and Infer Operating State," Springer LNCS, Vol. 8985, Mar. 2016, pp. 24-30.

[17] C.M. Rondeau, J.A. Betances, M.A. Temple, "Securing ZigBee Commercial Communications Using CB-DNA Fingerprinting," Security and Comm Networks, Jul 2018.

[18] C. M. Rondeau, M.A. Temple, J.A. Betances, "Dimensional Reduction Analysis for Constellation-Based DNA Fingerprinting to Improve Industrial IoT Wireless Security," *52nd Hawaii Int'l Conf on System Sciences (HICSS)*, Wailea, HI, Jan 2019.

[19] E. T. Nakamura, S. L. Ribeiro, "A privacy, security, safety, resilience and reliability focused risk assessment methodology for IIoT systems steps to build and use secure IIoT systems," *2018 Global Internet Things Summit*, 2018.

[20] M. Keefe, "Timeline: Critical Infrastructure Attacks Increase Steadily in Past Decade," Computer World, Nov 2012. [Online.] Available: https://tinyurl.com/y5um5jck, Accessed: 21 Aug 2019.

[21] United States Department of Homeland Security, "Advisory (ICSA-10-201-01C) USB Malware Targeting Siemens Control Software," 2014.

[22] United States Department of Homeland Security, "Joint Security Awareness Report (JSAR-12-241-01B): Shamoon/ DistTrack Malware," 2017.

[23] United States Department of Homeland Security, "Alert (TA17-163A): CrashOverride Malware," 2017.

[24] Department of Homeland Security (DHS) ICS-CERT, "MAR 17 352-01 HatMan - Safety System Targeted Malware," 2019.

[25] Siemens Industry Online Support, "WirelessHART Adapter SITRANS AW210 - 7MP3111," *Siemens*, [Online]. Available: https://tinyurl.com/yyjbgybm. Accessed: 21 Aug 2019/.

[26] WirelessHART Adapter, "BULLET - WirelessHART Adapter," *Pepperl and Fuchs*, [Online]. Available: https://tinyurl.com/y6njpu37. Accessed: 21 Aug 2019.

[27] Institute of Electrical and Electronics Engineers, "Standard for Local and metropolitan area networks Part 15.4: Low-Rate Wireless Personal Area Networks," IEEE Std 802.15.4-2011, Sep 2011.

[28] H. Patel, "Non-parametric feature generation for RF-fingerprinting on ZigBee devices," *2015 IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA 2015)*, 2015.

[29] J. Miao, L. Niu, "A Survey on Feature Selection," *Procedia Computer Science*, Vol. 91, Information Technology and Quantitative Management (ITQM 2016), pp. 919–926, 2016.

[30] S. Khalid, T. Khalil, S. Nasreen, "A Survey of Feature Selection and Feature Extraction Techniques in Machine Learning," *2014 Sci. Info. Conf. SAI 2014*, pp. 372–378, 2014.

[31] S. Visalakshi, V. Radha, "A Literature Review of Feature Selection Techniques and Applications: Review of Feature Selection in Data Mining," *2014 IEEE Int'l Conf. Computer Intelligence Comput. Res. IEEE ICCIC 2014*, no. 1997, pp. 1–6, 2015.

[32] L. Breiman, "Random Forests," *Machine Learning,* Vol. 45, No. 1, 2001, pp. 5-32.

[33] D.W. Steeneck, T. J. Bihl, "Stochastic Approximation for Learning Rate Optimization for Generalized Relevance Learning Vector Quantization," *IEEE Natl. Aero. Elec. Conf. (NAECON)*, pp. 366–371, 2018.

[34] M. Hollander, D.A. Wolfe, E. Chicken, Non-parametric Statistical Methods, Hoboken: Wiley, 2014.

[35] K. Kira, L. A. Rendell, "The Feature Selection Problem: Traditional Methods and a New Algorithm," *AAAI 1992*, pp. 129–134, July 1992.

[36] K. Kira, L. A. Rendell, "A Practical Approach to Feature Selection," *Int'l Conf on Machine Learning*, Jul 1992.

[37] I. Kononenko, E. Šimec, M. Robnik-Šikonja, "Overcoming the Myopia of Inductive Learning Algorithms with RELIEFF," Appl. Intel., vol. 7, no. 1, pp. 35–55, 1997.

[38] M. Wu, Y. Wang, "A Feature Selection Algorithm of Music Genre Classification Based on ReliefF and SFS," *2015 IEEE/ACIS 14th Int'l Conf. Comput. Inf. Sci.*, pp. 539–544, 2015.