An Optimization Framework for Generalized Relevance Learning Vector Quantization with Application to Z-Wave Device Fingerprinting

Trevor J. Bihl
Air Force Institute of Technology
Trevor.Bihl@afit.edu

Michael A. Temple
Air Force Institute of Technology
Michael.Temple@afit.edu

Kenneth W. Bauer Jr.
Air Force Institute of Technology
Kenneth.Bauer@afit.edu

Abstract

Z-Wave is low-power, low-cost Wireless Personal Area Network (WPAN) technology supporting Critical Infrastructure (CI) systems that are interconnected by government-to-internet pathways. Given that Z-wave is a relatively unsecure technology, Radio Frequency Distinct Native Attribute (RF-DNA) Fingerprinting is considered here to augment security by exploiting statistical features from selected signal responses. Related RF-DNA efforts include use of Multiple Discriminant Analysis (MDA) and Generalized Relevance Learning Vector Quantization-Improved (GRLVQI) classifiers, with GRLVQI outperforming MDA using empirically determined parameters. GRLVQI is optimized here for Z-Wave using a full factorial experiment with spreadsheet search and response surface methods. Two optimization measures are developed for assessing Z-Wave discrimination: 1) Relative Accuracy Percentage (RAP) for device classification, and 2) Mean Area Under the Curve (AUCM) for device identity (ID) verification. Primary benefits of the approach include: 1) generalizability to other wireless device technologies, and 2) improvement in GRLVQI device classification and device ID verification performance.

1. Introduction

The Information Technology (IT) centric focus of the 2002 E-Government Act was appropriate at that time and highlighted the importance of information security and privacy [1]. Since then, wireless communication and Critical Infrastructure (CI) control technologies have changed considerably and e-government connectivity now exists well below the IT internet backbone. Although remaining IT-centric, the US government has more recently acknowledged the importance of addressing "rapidly evolving and persistent cyber threats" [2]. Perhaps of greatest concern from a protection perspective is that the cyber threat and attack surface increases as government-to-internet connectivity (exposure) increases through supporting sub-internet pathways comprised of wireless WiFi, Z-Wave, and Bluetooth devices.

Cyber physical systems (CPS) include Wireless Personal Area Network (WPAN) devices supporting the Internet of Things (IoT) and CI systems [3], [4], [5]. Low-cost, low-power Z-Wave devices are among the sub-internet WPAN support technologies that enable mesh networks comprised of smart devices [6], [7]. These networks support data collection and control [8] via Supervisory Control and Data Acquisition (SCADA) systems [9]. Mesh networks are used, for instance, in hospital [10] and electrical smartgrid [11] applications, both of which are CI elements within e-government and private sectors. Of particular risk in WPAN applications is that a security compromise of one device can threaten the security of the entire network. Thus, vetting of individual Z-Wave device identities is critical for ensuring robust security. This criticality extends beyond CI, with e-government CPS applications including interactive public displays and urban intervention systems (participatory and interactive) that relay information of interest to the public [12].

Of interest here are CPS implementations using Z-Wave WPAN devices that 1) lack robust security and 2) which are readily exploitable (hackable) [13], [14]. Device hardware ID and operating state discrimination for CI security applications has been reliably demonstrated using Physical (PHY) layer security enhancement [4], [9], [15]. As discussed in [16], PHY layer security involves either 1) adding physically traceable objects to devices [17] or 2) Radio Frequency Distinct Native Attribute (RF-DNA) fingerprinting based on PHY device emissions which overcome limitations of encryption key-based measures [18]. RF-DNA differs from typical WPAN defense and security strategies that target higher bit-level network layers [19], i.e., the Network (NWK) and Media Access Control (MAC) layers [20]. Using underutilized PHY information [18] with NWK and MAC information yields a more robust biometric-like wireless security strategy that includes [18], [21]:

1. "Something you know" (NWK – encryption keys)
2. "Something you have" (MAC – MAC address)
3. "Something you are" (PHY – RF Fingerprints).

The inclusion of PHY-based information is most important given that replication of known bit-level ID credentials is relative easy and enables unauthorized network access [21]. The device dependent PHY

URI: http://hdl.handle.net/10125/41444
ISBN: 978-0-9981331-0-2

features capture intrinsic device differences resulting from component and production variation [21] and are nearly impossible to replicate. The degree of device discrimination is captured through statistical methods of feature extraction, device classification (one vs. many), and device ID verification (one vs. one) [21] [22]. Discrimination is assessed using 1) an eigenspace based Multiple Discriminant Analysis (MDA) classifier, and 2) a nonlinear Generalized Relevance Learning Vector Quantization-Improved (GRLVQI) neural network based classifier [23].

Prior research considered GRLVQI for Z-Wave devices [21] but used empirical parameter settings derived from related ZigBee work [23]. Given that algorithmic settings directly impact LVQ performance [24], with setting determination being a balance between science and art with "no hard-and-fast rules" [25], the focus here was on optimizing GRLVQI settings for Z-Wave. This was done using a 5-factor full factorial experimental design and Analysis of Variance (ANOVA), methods commonly used for industrial process improvement [26]. Optimal settings were determined using both a spreadsheet search [27] and Response Surface Methodology (RSM) [28], [29], [30], [31] with nonlinear optimization [32]. The experimental design assessment was aided by introducing two performance measures, including: 1) Relative Accuracy Percentage (RAP) for classification, and 2) Mean Area Under the Curve (AUC_M) for device ID verification.

The remainder of this paper is organized as follows. Section 2 provides a summary of Z-Wave devices, RF-DNA Fingerprinting, and the MDA and GRLVQI classifiers. Section 3 addresses ANOVA, RSM and RAP and AUC_M performance measure development. Section 4 presents performance results and Section 5 concludes the paper.

2. RF-DNA and Z-Wave

Z-Wave wireless communication devices are small, low-cost hardware devices and are generally considered less secure than competing WPAN technologies given 1) originally lacked built in encryption [33] and 2) the proprietary nature of the standard making it difficult for third parties to provide enhancements [34]. Z-Wave follows a similar ISO architecture to ZigBee, and similarly has a predefined preamble and Start of Frame (SoF) [35]. General Z-Wave signal characteristics are known and presented in Table 1 along with a conceptualization of Z-Wave PHY packet structure in Figure 1. The preamble response (the first 8.3 ms of Z-Wave bursts) was considered the Region Of Interest (ROI) for RF-DNA extraction. Z-Wave also includes a payload-based home identification (32-bits) and source identification (8-bits) [34]. Due to their proprietary nature, further knowledge of Z-Wave signal characteristics is limited and thus digital forensic analysis, c.f. [36], [37], of Z-Wave devices remains an emerging area of interest [38].

Table 1. Z-Wave Characteristics

FREQUENCY	906 MHz
BIT RATE	40 Kbits/s
SECURITY	None (200 and 300 series models) AES 128 (400 series models)
LATENCY	~1000 ms
RANGE	30-100 m
MESSAGE SIZE (BYTES)	64 (max)

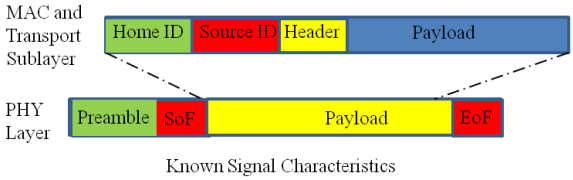

Figure 1. Z-Wave device signal characteristics [33] [34] [35].

2.1. Z-Wave Signal Collection

The work herein considered the Z-Wave dataset first presented in [21], where $N_D = 3$ Aeotec Z-Stick S2 transmitter devices were considered for analysis with all devices serving in "authorized" roles. A total of 230 preamble responses were collected for each device at a sample frequency of $f_s = 2$ Msps. Burst detection was accomplished using an amplitude-based leading edge detector with a -6 dB threshold [21]. For signal collection, each Z-Stick transmitter was located 10 cm from a vertically-oriented LP0410 log periodic antenna, which was connected via a Gigabit Ethernet cable directly to the USRP-2921 RF input [21]. The resultant collected Signal-to-Noise Ratio (SNR) was at $SNR_C = 24.0$ dB. Independent, like-filtered Additive White Gaussian Noise (AWGN) was added to collected signals to achieve desired operating conditions of $SNR \in [0\ 24.0]$ dB [21], [22].

2.2. RF-DNA Fingerprint Generation

Consistent with [21], $N_S = 3$ RF-DNA fingerprint features (statistics) of variance (σ^2), skewness (γ), and kurtosis (κ) were computed for $N_R = 20$ subregions spanning the ROI within $N_C = 3$ Z-Wave instantaneous time domain responses of amplitude (a), phase (ϕ), and frequency (f). RF-DNA fingerprints were generated, as in [21], [22], by 1) dividing each response into N_R contiguous equal length bins, 2) calculating N_s features within each bin

and across the entire response (N_R + 1 total bins), and 3) computing regional fingerprint vectors as,

$$F_{Ri} = [\sigma_{Ri}^2, \gamma_{Ri}, \kappa_{Ri}]_{1\times 3} \ , \quad (1)$$

where $i = 1,2,\ldots,N_R+1$. A fingerprint vector for each of the N_C characteristics is formed from (1) as,

$$F^C = \left[F_{R_1} \vdots F_{R_2} \cdots F_{R_{(N_R+1)}}\right]_{1\times N_s(N_R+1)} \ , \quad (2)$$

which are concatenated to form the final fingerprint vector:

$$\boldsymbol{F} = \left[\boldsymbol{F^a} \vdots \boldsymbol{F^\phi} \vdots \boldsymbol{F^f}\right]_{1\times N_s(N_R+1)\times N_C} \ . \quad (3)$$

For Z-Wave device discrimination assessments, a total of N_F = 189 features are computed with N_{TRN} = 115 Training (TNG) and N_{TST} = 115 Testing (TST) observations per device. The TNG and TST data was sequestered during model development to avoid the possibility of overfitting.

2.3. Classifier Models

2.3.1. GRLVQI Classifier Model. The GRLVQI classifier employed herein is based on the work in [21], [23]. GRLVQI extends the squared-Euclidean distance based gradient descent process of Learning Vector Quantization (LVQ) with embellishments of a sigmoidal cost function [39], [40], relevance learning [41], [42], and conscience learning [25], [43], which are employed to train prototype vectors to a given class label [21], [23]. GRLVQI extends GRLVQ [42] with the conscience learning of DeSieno [44], improved PV update logic, and a frequency based maximum input update strategy [25].

As with LVQ and various embellishments, GRLVQI has five different factors to consider: 1) Factor A, gradient descent learning rate (ϵ), 2) Factor B, relevance learning rate (ξ), 3) Factor C, conscience rate 1 (γ), 4) Factor D, conscience rate 2 (β), and 5) Factor E), the number of prototype vectors (N_{PV}) instantiated per class. For all devices used herein, prior probabilities were considered equal between devices, with the update logic and GRLVQI classifier model as described in [18], [23].

2.3.2. Multiple Discriminant Analysis (MDA). MDA is both readily interpretable and is computationally inexpensive. Furthermore, MDA has shown significant performance advantages over GRLVQI for many RF-DNA Fingerprinting problems and it is thus included to provide a baseline performance reference, consistent with [23]. MDA is a multi-class extension of Fisher's two class linear classifier [23]. MDA considers input fingerprint matrix $\boldsymbol{F}$ and N_C classes and involves an eigenvector-based projection of the data relative to a ratio of between-group to within-group sum-of-squares, the Fisher criterion [45].

2.4. Quantifying Classification Performance

Classification is considered for "one vs. many" scenarios as in [21], [22]. Two performance measures are considered: 1) *SNR* (dB) "Gain" (G_{SNR}) defined as the reduction in *SNR* for two methods to achieve a given average percent correct classification (*%C*) [23], and 2) Relative Average Percentage (RAP). Both G_{SNR} and RAP measures consider figures with *%C* on the y-axis and *SNR* (dB) on the x-axis, as seen in Figure 2 for both training (TNG) and testing (TST) performance of MDA and GRLVQI using the Z-Wave dataset.

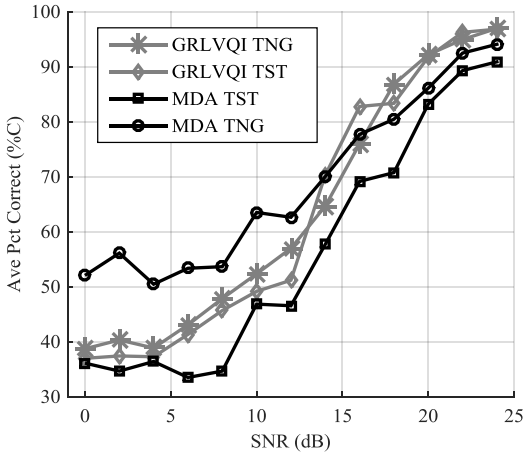

Figure 2. Z-Wave *Testing (TST)* and *Training (TNG)* Classification performance for MDA and GRLVQI classifier models.

2.4.1. SNR Gain. G_{SNR} is computed for authorized device TNG and TST datasets [23]. For results herein, performance using the full dimensional (N_F = 189) baseline feature set serves as the reference with an arbitrary *%C* ≥ 90% benchmark as in [21], [23]. G_{SNR} is interpreted as:

1) G_{SNR} < 0.0 (negative), a given method achieves the same *%C* as the baseline at a higher *SNR*, i.e. the method underperforms the baseline method.

2) G_{SNR} = 0.0, a given method achieves the same *%C* as the baseline at the same *SNR*

3) G_{SNR} > 0.0 (positive), a given method achieves the same *%C* as the baseline at a lower SNR, i.e. the method outperforms the baseline method.

For Z-Wave results in Figure 2 at *%C* = 90%, GRLVQI outperforms MDA with G_{SNR} = +3.32 dB (TST) and G_{SNR} = +3.72 dB (TNG). Therefore, when considering classification performance, GRLVQI is a superior classifier for Z-Wave relative to MDA.

2.4.2. Relative Accuracy Percentage (RAP). In cases where $\%C \geq 90\%$ is not achieved, G_{SNR} is not computable and is thus insufficient for some complete analysis. Since determining algorithmic settings by examining possible setting combinations is of interest herein, the possible lack of G_{SNR} can introduce instabilities and the RAP measure was introduced herein as an alternative measure.

RAP is generated by 1) computing the Area Under Classification Curve (*AUCC*) values for each method via a trapezoidal approximation, and 2) computing the RAP of a given method's $AUCC_{M(i)}$ relative to the baseline $AUCC_{Base}$ method according to

$$RAP = AUCC_{M(i)}/AUCC_{Base} \ . \qquad (4)$$

RAP provides the fraction of $AUCC_{M(i)}$ with respect to $AUCC_{Base}$ and 1) enables a comparison for methods not achieving $\%C \geq 90\%$, and 2) reflects performance across all *SNR*. RAP is interpreted as:

1) *RAP* < *1.0*, a given method achieves overall lower *%C* than the baseline

2) *RAP* = 1.0, a given method achieves overall *%C* comparable to the baseline

3) *RAP* > *1.0*, a given method exceeds overall baseline *%C* performance.

Applying the RAP process to results in Figure 2 yields MDA $AUCC_{Base} = 13.32$ (TST) and $AUCC_{GRLVQI} = 15.06$ (TST), with a *RAP* = 1.13. Thus, GRLVQI TST performance is better, on average, across all operating points when compared to MDA (consistent with a visual assessment of Figure 2).

2.4.3. Classification Performance Results Table 2 presents overall results for the Z-Wave data for both MDA and GRLVQI using AUCC, RAP and Gain. Overall, Table 2 shows that GRLVQI performs consistently better across all operating points when compared to MDA for Z-Wave.

Table 2. Baseline Classification Results

ALG.	SET	AUCC	*SNR* (dB) AT *%C* = 90%	RELATIVE MDA (TST) RAP	RELATIVE MDA G_{SNR} (TST) AT *%C* = 90%
MDA	TNG	16.39	21.23	1.23	+1.68
	TST	13.32	22.91	1.00	0.00
GRLVQI	TNG	15.23	19.19	1.14	+3.72
	TST	15.06	19.59	1.13	+3.32

2.5. Quantifying Verification Performance

Device ID verification is considered in a "one versus one" (claimed vs actual) ID assessment. Here, a trained classifier is considered along with probability mass functions (PMFs) for authorized devices [46]. Computed for ID verification are True Verification Rate (TVR) and False Verification Rate (FVR) for the y-axis and x-axis, respectively, of authorized device Receiver Operating Characteristic (ROC) curves [46]. Two measures are considered herein to quantify verification performance: percent authorized [21], [22], and mean AUC (AUC_M).

2.5.1. Percentage Authorized (*%Aut*). Consistent with [21], [22], ID verification performance is commonly evaluated by a percentage correctly authorized from a binary grant/deny network access decisions based on a verification criteria, e.g. TVR ≥ 90% at FVR ≤ 10%. Figure 3 presents analysis based on this threshold (denoted by dashed red lines) for GRLVQI at *SNR* = 20 dB, with solid black lines indicate successfully achieving this threshold and dashed grey lines indicate a failure to achieve this threshold. Overall results in Figure 3 show *%Aut* = 1/3 = 33.33% success.

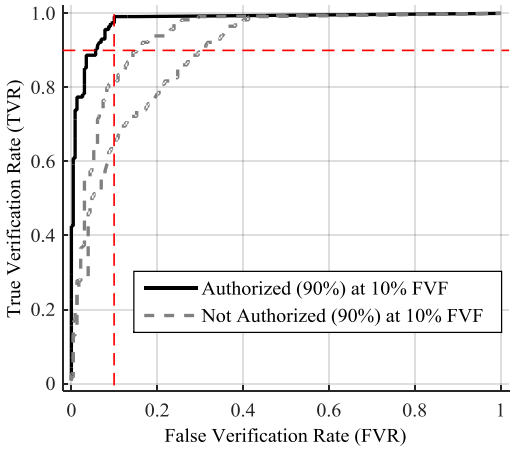

Figure 3. Example Z-Wave Authorized Device ID Verification performance at *SNR* = 20 dB for GRLVQI.

2.5.2. Verification Mean AUC (AUC_M). Percentage authorized (*%Aut*) reflects coarse sampling, e.g. $N_D = 3$ devices *%Aut* ∈ [0, 33, 66, 100], and does not distinguish between perfectly verified results (a ROC Area Under the Curve (AUC) of AUC = 1.0) and results that merely achieve the TVR > 90% at FVR < 10%. Therefore, AUC_M is proposed as an alternative verification performance measure. This involves computing the AUC for each ROC curve, one curve is associated with one device, and then computing the mean of all curves considered.

2.5.3. Verification Performance Results

Table 3 presents verification results via *%Aut*, AUC and AUC_M at *SNR* = [18, 20, 22] dB; further verification results will only be considered herein for *SNR* = 20dB (the *SNR* at which GRLVQI achieves *%C* = 90%). As seen in the *%Aut* column Table 3, the *%Aut* rate involves dichotomization, c.f. [47],

whereby the continuous response of the ROC curve is made discrete, which introduces issues related to resolution, c.f. [48], [49], [50], and optimization of a dichotomous response variable is nontrivial, c.f. [51], [52]. When examining the continuous AUC values, one can notice slight differences and thus optimization relative to a continuous variable is preferred.

Table 3. Baseline Verification Results

METHOD	*SNR* (DB)	% AUT	AUC_1	AUC_2	AUC_3	AUC_M
MDA	18	100	0.978	0.974	0.978	0.977
	20	100	0.978	0.957	0.978	0.971
	22	100	0.978	0.978	0.978	0.978
GRLVQI	18	0	0.849	0.902	0.945	0.899
	20	33	0.890	0.937	0.981	0.936
	22	66	0.944	0.961	0.992	0.966

3. Analysis of Variance and GRLVQI Optimization Considerations

Due to the small size of the Z-Wave dataset, it is intuitive that linear methods (MDA) underperform nonlinear methods (GRLVQI). However, determining appropriate settings is critically important for Z-Wave analysis via GRLVQI since this data is associated with unknown operating characteristics. However, determining appropriate LVQ algorithm settings is a largely unexplored domain; herein, a second-order RSM model will be considered to solve for optimal algorithmic settings where the target are the dependent variables (RAP or AUC_M).

3.1. Analysis of Variance (ANOVA)

General linear models, e.g. ANOVA and linear regression, work to understanding variability of data through sums of squares [28]. Factorial experiments consider all combinations of different factors and levels to understand significance of factors relative to the response and the interaction of factors [28]. Herein, a factorial experiment for the 5 GRLVQI algorithmic settings is proposed, ANOVA responses will be considered as RAP_{TNG}, RAP_{TST} and AUC_M.

3.2. Response Surface Methodology (RSM)

RSM extends ANOVA by considering an ANOVA model with both squared terms and two-way interactions:

$$J(x) = B_0 + \sum_{i=1}^{s} B_i x_i + \sum_{i,j,i=1}^{s} B_{i,j} x_i x_j + \sum_{i,j,i=1}^{s} B_{i,i} x_i^2 , \quad (5)$$

where s represents the number of factors, B terms are coefficients solved for via a general linear model, and x represents a given factor [53].

3.3. GRLVQI Algorithmic Settings

To consider a full factorial model, appropriate minimum and maximum values for GRLVQI algorithmic settings must be developed. However, little has been published about LVQ algorithmic settings beyond 1) the general hierarchy of $0 \leq \xi(t) \leq \epsilon(t) \leq 1$ for relevance-based LVQ methods [54], 2) specific guidelines for specific applications, e.g. [55], [56], and 3) learning rate convergence methods, e.g. [55]. Additionally, appropriately specifying N_{PV} is also critical to avoid overfitting and/or poor performance [57].

A few considerations were made in determining ranges for appropriate settings. The operational design points for each factor appear in Table 4 where the baseline settings, coded as "0", are baseline GRLVQI settings used by [23], [58]. The high (+) and low (–) settings in Table 4 were determined by 1) taking magnitudes of 10 times above (+) and below (–) the learning (Factor A) and relevance (Factor B) rates, 2) conscience rate limits were determined by considering the extreme settings explored in [59], and 3) PV limits were found by going 30% above and below the baseline.

Table 4. Experimental Design Region for GRLVQI

	FACTORS				
	A	B	C	D	E
LEVEL	LEARN. RATE (ϵ)	REL. RATE (ξ)	CONSC. RATE 1 (γ)	CONSC. RATE 2 (β)	N_{PV}
–	0.0025	0.0005	0.5	0.15	7
0	0.025	0.005	2.0	0.35	10
+	0.25	0.05	4.5	0.55	13

4. GRLVQI Optimization Framework

Results were generated using GRLVQI with THE Z-Wave datasets for all $3^5 = 243$ combinations of a full factorial design using values in Table 4. To determine optimal settings, two approaches were considered: 1) a spreadsheet search of the full factorial results to find the maximum classification and verification performance and 2) employing nonlinear optimization methods to find potential optimal settings within the full factorial settings. Sequestered TST data was used to validate the settings in a process similar to that of [60].

4.1. Spreadsheet Search

A spreadsheet search, consistent with [27], was performed to find the highest performing results, and resultant settings, from the experimental design. The highest performing results were found for: 1) TNG

results, 2) TST results, and 3) AUC_M results. When considering TNG results, the highest performance was found with Factor A, Factor B, and Factor C at the highest setting, Factor D at its mid-range setting, and Factor E its lowest setting. For TST results, the highest performance was found with Factor A and Factor C at their highest settings, Factor B at its mid-range setting, and Factor D and Factor E at their lowest settings. For verification AUC_M results, the highest performance was found with Factor A and Factor B at their highest settings, Factor C and Factor D at mid-range settings, and Factor E at its lowest setting.

4.2. Constrained Nonlinear Optimization

One limitation of the spreadsheet search is that it only finds best results in explored combinations. To find optimal algorithmic settings within the design space, the full factorial results were considered with RSM. First, an ANOVA model was computed for classification (RAP_{TST} and RAP_{TNG}) and verification (AUC_M). Next, the statistically significant ($\alpha = 10$) features were selected and a second ANOVA model, with only the selected factors, was then created. The second ANOVA model was optimized consistent with [32]; the optimization process employed constrained nonlinear optimization (interior point optimization), consistent with [61], and the results provided optimal GRLVQI settings.

4.2.1. ANOVA and RSM Results. Table 5 presents the ANOVA table for the model, error or residuals, and total Sum of Squares (SoS). The model is further broken down by each model factor (main effects and 2nd order interaction) along with its SoS and p-value. Although Degrees of Freedom (DoF), Mean Squares, and F_0 are not shown in Table 4, these are easily recomputed due to the underlying relationships: 1) each main effect and each interaction has one DoF each, 2) Mean Squares (MS) for a factor are $MS_{Factor} = SoS / DoF$, and 3) factor F_0 is computed as $F_0 = MS_{Factor} / MS_{Error}$ [28].

4.2.2. Sequential Quadratic Programming (SQP). A majority voting approach was applied to the ANOVA models in Table 5 to determine which features to consider for further analysis. Thus, features that were on majority statistically significant (dark or light gray shading) were retained and further ANOVA models were computed. Constrained minimization (target values were negated since maximization is possible by minimizing a negation) was considered where a finite-difference approximation was computed by starting with an initial estimate (the baseline GRLVQI settings). The relationship between variables was optimized via SQP wherein a line search was employed [61].

The minimization was constrained between the minimum and maximum values seen in Table 4 to avoid computing values outside those explored (e.g. unbounded optimization yielded settings far outside the design space, with magnitudes ranging from 10^{13} to 10^{42}). The optimal solution was then computed for each factor level with the resultant optimal algorithmic settings for each factor are presented in Table 6 along with performance results. Optimization was considered individually for maximum RAP_{TST}, RAP_{TNG}, and AUC_M. Of note, some optimization solutions had results that were identical to lower or upper bounds, denoted with + or –; otherwise, the resultant uncoded setting is presented.

Table 5. Analysis of Variance Table from Full Factorial Data. Dark Grey indicates a variable significant at 5%, Light Grey indicates a variable is significant at 10% and * indicates a p-value < 0.001

	RAP_{TNG}		RAP_{TST}		AUC_M	
Source of Variance	SoS	p	SoS	p	SoS	p
Total Model	0.4459	*	0.3631	*	1.5549	*
ϵ	0.1595	*	0.1474	*	0.7093	*
ξ	0.0089	*	0.0113	*	0.0007	0.854
γ	0.0190	*	0.0294	*	0.0001	0.886
β	0.0055	*	0.0100	*	0.0023	0.745
N_{PV}	0.0067	*	0.0389	*	0.1333	0.014
ϵ^2	0.0743	*	0.0953	*	0.6106	*
ξ^2	0.0113	*	0.0154	*	0.0001	0.952
γ^2	0.0025	*	0.0012	0.035	0.0592	0.098
β^2	0.0004	0.153	0.0029	0.001	0.0576	0.103
N_{PV}^2	0.0033	*	0.0236	*	0.0162	0.386
$\epsilon \times \xi$	0.0013	0.009	0.0046	*	0.4835	*
$\epsilon \times \gamma$	0.0011	0.020	0.0053	*	0.0105	0.485
$\epsilon \times \beta$	0.0001	0.508	0.0001	0.815	0.0191	0.347
$\epsilon \times N_{PV}$	0.0001	*	0.0053	*	0.0219	0.314
$\xi \times \gamma$	0.0003	0.234	0.0001	0.682	0.0001	0.960
$\xi \times \beta$	0.0001	0.485	0.0004	0.223	0.0095	0.508
$\xi \times N_{PV}$	0.0001	0.479	0.0007	0.119	0.0001	0.968
$\gamma \times \beta$	0.0001	0.813	0.0001	0.650	0.0019	0.761
$\gamma \times N_{PV}$	0.0001	0.479	0.0005	0.189	0.0004	0.893
$\beta \times N_{PV}$	0.0023	0.001	0.0032	*	0.0032	0.699
Error	0.0423	*	0.0589	*	4.773	*
Total	0.4883	*	0.4221	*	6.328	*
R^2	0.913		0.860		0.245	
R^2 Adj	0.905		0.848		0.178	

Table 6. Algorithm Optimization Results

Method	Max. Objective	Factors Levels					Performance Results					
							Classification				Verification at *SNR* = 20dB	
		A	B	C	D	E	G_{SNR} (dB) at %C = 90%		RAP		TVR	AUC_M
							TNG	TST	TNG	TST		
Spreadsheet Search	RAP_{TNG}	+	+	+	0	–	+5.30	+5.78	1.22	1.18	66%	0.974
	RAP_{TST}	+	+	+	–	–	+4.99	+5.63	1.22	1.20	66%	0.977
	AUC_M	+	+	0	0	–	+5.30	+5.77	1.22	1.18	66%	0.979
Constrained Nonlinear Optimization	RAP_{TNG}	0.1573	+	4.3746	–	–	+4.51	+5.27	1.20	1.20	33%	0.961
	RAP_{TST}	0.1501	+	+	–	–	+5.23	+5.26	1.20	1.19	66%	0.967
	AUC_M	0.1509	+	+	–	–	+4.79	+5.28	1.20	1.17	66%	0.974
None	Baseline GRLVQI	0	0	0	0	0	+3.72	+3.32	1.14	1.13	33%	0.936
	MDA	N/A	N/A	N/A	N/A	N/A	+1.68	0.00	1.23	1.0	100%	0.971

Evident in Table 6 is that both classification and verification performance improve with the spreadsheet search and optimized settings when compared to baseline GRLVQI settings. Consistency across results indicates that using too many PVs is detrimental to performance, logically this could facilitate over-fitting, and thus the LVQ architecture does not need to be too cumbersome. Although TVR = 100% for MDA, overall AUC_M is consistent between MDA and GRLVQI optimized results.

5. Summary and Conclusions

From an e-government cyber security and protection perspective, sub-internet pathways that are comprised of common wireless WiFi, Z-Wave and Bluetooth devices increase the cyber attack surface and risk of service degradation or disruption. Risk mitigation is a top priority when considering that hospital, electrical power grid and other CI systems are vulnerable. The focus here is on demonstrating measures to enhanced Z-Wave security, with results being generally applicable to other WPANs.

There are four contributions for improving Z-Wave device discrimination using RF-DNA Fingerprints, including: 1) introduction of RAP and AUC_M performance measures, 2) formalization of a DOE approach for classifier model development, 3) demonstration of a GRLVQI optimization framework for classification and verification, and 4) a GRLVQI and MDA/ML comparative assessment for Z-Wave PHY device identification.

Herein, a process was presented to find optimal algorithm settings by first performing a designed experiment (full factorial) and then employing both a spreadsheet search and nonlinear optimization. The results collectively illustrate that 1) determining appropriate GRLVQI algorithm settings is critical (the optimized learning rates differed by no more than 5% yet produced larger variations in RAP and AUC_M), and 2) the viability of DOE methods for RF-DNA Fingerprinting algorithm optimization.

6. References

[1] *E-Government Act of 2002, Public Law 107–347, 116 Stat. 2899, 44 U.S.C. § 101, 107th Congress,* 2002.

[2] "FACT SHEET: Enhancing and Strengthening the Federal Government's Cybersecurity," Executive Office of The President, Office of Management and Budget, June 2015. [Online]. Available: https://www.whitehouse.gov/omb. [Accessed 29 Aug. 2016].

[3] Y. Sharaf-Dabbagh and W. Saad, "On the authentication of devices in the Internet of things," *IEEE 17th Int. Symp. on A World of Wireless, Mobile and Multimedia Networks (WoWMoM),* pp. 1-3, 2016.

[4] S. Stone and M. Temple, "RF-Based Anamoly Detection for PLCs in Critical Infrastructure Apps," *Int. J. on Critical Infrastructure Protection,* vol. 5, no. 2, pp. 66-73, 2012.

[5] E. Shakshuki, H. Malik and T. Sheltami, "WSN in cyber physical systems: Enhanced energy management routing approach using software agents," *Future Generation Computer Systems,* vol. 91, pp. 93-104, 2014.

[6] A. Wong, "Case Study: Simulated deployment of a mesh network in Honolulu," *43rd Hawaii Int. Conf. of System Sciences (HICSS),* pp. 1-9, 2010.

[7] R. Bruno, M. Conti and E. Gregori, "WLAN Technologies for mobile ad hoc networks," *Hawaii Int. Conf. on System Sciences (HICSS),* pp. 1-11, 2001.

[8] A.-M. Chang, P. K. Kannan and S. Fellow, "Preparing for wireless and mobile technologies in government," *E-government ,* pp. 345-393, 2003.

[9] S. Stone, M. Temple and R. Baldwin, "Detecting Anomalous PLC Behavior Using RF-Based Hilbert Transform Features and a Correlation-Based Verification Process," *Int. J. on Critical Infrastructure Protection,* vol. 9, pp. 41-51, 2015.

[10] H. Cao, V. Leung, C. Chow and H. Chan, "Enabling technologies for wireless body area networks: A survey and outlook.," *IEEE Commun. Mag.*, vol. 47, no. 12, pp. 84-93, 2009.

[11] V. Güngör, D. Sahin, T. Kocak, S. Ergüt, C. Buccella, C. Cecati and G. Hancke, "Smart grid technologies: communication technologies and standards," *IEEE Trans. on Industrial Informatics,* vol. 7, no. 4, pp. 529-539, 2011.

[12] F. Salim and U. Haque, "Urban computing in the wild: A survey on large scale participation and citizen engagement with ubiquitous computing, cyber physical systems, and Internet of Things," *Int. J. of Human-Computer Studies,* vol. 81, pp. 31-48, 2015.

[13] B. Fouladi and S. Ghanoun, "Security evaluation of the Z-Wave wireless protocol," *Black hat USA,* 2013.

[14] J. Fuller and B. Ramsey, " Rogue Z-Wave controllers: A persistent attack channel," *IEEE Local Computer Networks Conf. Workshops,* pp. 734-741, 2015.

[15] J. Lopez, M. Temple and B. E. Mullins, "Exploitation of HART Wired Signal Distinct Native Attribute (WS-DNA) Features to Verify Field Device Identity and Infer Operating State," *Springer LNCS,* vol. 8985, pp. 24-30, Mar. 2016.

[16] T. J. Bihl, K. W. Bauer and M. A. Temple, "Feature Selection for RF Fingerprinting With Multiple Discriminant Analysis and Using ZigBee Device Emissions," *IEEE Trans. on Information Forensics and Security,* vol. 11, no. 8, pp. 1862-1874, 2016.

[17] F. Amsaad and M. Niamat, "Reliable and reproducible PUF based cryptographic keys under varying environmental conditions," *IEEE National Aerospace & Electronics Conf. (NAECON),* 2016.

[18] B. W. Ramsey, M. A. Temple and B. E. Mullins, "PHY foundation for multi-factor ZigBee node authentication," *IEEE Global Commun. Conf. (GLOBECOM),* pp. 795-800, 2012.

[19] Y. Sheng, K. Tan, G. Chen, D. Kotz and A. Campbell, "Detecting 802.11 MAC layer spoofing using received signal strength," *27th Conf. on Computer Commun.*, 2008.

[20] M. Dener, "Security analysis in wireless sensor networks," *Int. J. of Distributed Sensor Networks,* vol. 2014, pp. 1-9, 2014.

[21] T. J. Bihl, M. A. Temple, K. Bauer and B. Ramsey, "Dimensional Reduction Analysis for Physical Layer Device Fingerprints with Application to ZigBee and Z-Wave Devices," *IEEE Military Commun. Conf. (MILCOM),* pp. 360-365, 2015.

[22] W. E. Cobb, E. D. Laspe, R. O. Baldwin, M. A. Temple and Y. C. Kim, "Intrinsic physical-layer authentication of integrated circuits," *IEEE Trans. on Information Forensics and Security,* vol. 7, no. 1, pp. 14-24, 2012.

[23] P. K. Harmer, D. R. Reising and M. A. Temple, "Classifier selection for physical layer security augmentation in Cognitive Radio networks," *IEEE Int. Conf. on Commun. (ICC),* pp. 2846-2851, 2013.

[24] M.-T. Vakil-Baghmisheh and N. Pavesic, "Premature clustering phenomenon and new training algorithms for LVQ," *Pattern Recognition,* vol. 36, pp. 1901-1912, 2003.

[25] M. J. Mendenhall, *A Neural Relevance Model for Feature Extraction from Hyperspectral Images, and its Application in the Wavelet Domain*, PhD Dissertation: Rice University, 2006.

[26] P. Gamage, N. Jayamaha, N. Grigg and N. Nanayakkara, "Comparative analysis of Taguchi's crossed array approach vs combined array approach to robust parameter design: A study based on apparel industry," *IEEE International Conference on Industrial Engineering and Engineering Managemen,* pp. 749-753, 2014.

[27] W. Zhang and A. T. Goh, "Reliability assessment on ultimate and serviceability limit states and determination of critical factor of safety for underground rock caverns," *Tunnelling and Underground Space Technology,* vol. 32, pp. 221-230, 2012.

[28] J. Bellucci, T. Smetek and K. Bauer, "Improved hyperspectral image processing algorithm testing using synthetic imagery and factorial designed experiments," *IEEE Trans. on Geoscience and Remote Sensing,* vol. 48, no. 3, pp. 1211-1223, 2010.

[29] C.-C. Chiu, D. F. Cook, J. P. Pignatiello and A. D. Whittaker, "Design of a radial basis function neural network with a radius-modification algorithm using response surface methodology," *J. of Intelligent Manufacturing,* vol. 8, no. 2, pp. 117-124, 1997.

[30] A. Glyk, D. Solle, T. Scheper and S. Beutel, "Optimization of PEG–salt aqueous two-phase systems by design of experiments," *Chemometrics and Intelligent Laboratory Systems,* vol. 149, pp. 12-21, 2015.

[31] L. Wu, K. Yick, S. Ng and J. Yip, "Shape characterization for optimisation of bra cup moulding," *J. of Fiber Bioengineering and Informatics,* vol. 4, no. 3, pp. 235-243, 2011.

[32] X. Yang, J. Li, Z. Fang and C. Wang, "The Optimum Design of Gear Transmission Based on MATLAB," *Int. Conf. Measuring Technology and Mechatronics Automation (ICMTMA),* vol. 3, pp. 925-928, 2010.

[33] M. Knight, "How safe is Z-Wave?[Wireless standards]," *Computing and Control Engineering,* vol. 17, no. 6, pp. 18-23, 2006.

[34] C. Gomez and J. Paradells, "Wireless home automation networks: A survey of architectures and technologies," *IEEE Commun. Magazine,* pp. 92-101, June 2010.

[35] M. Galeev, "Catching the Z-Wave," *Electronic Engineering Times India,* pp. 1-5, Oct. 2006.

[36] J. Sammons, *The basics of digital forensics: the primer for getting started in digital forensics*, Elsevier, 2012.

[37] S. Garfinkel, "Digital Forensics," *American Scientist,* vol. 101, no. 5, pp. 370-377, 2013.

[38] C. Badenhop, B. Ramsey, B. Mullins and L. Mailloux, "Extraction and analysis of non-volatile memory of the ZW0301 module, a Z-Wave transceiver," *Digital Investigation,* vol. 17, pp. 14-27, 2016.

[39] A. S. Sato and K. Yamada, "Generalized learning vector quantization," in *Advances in neural information processing systems*, Cambridge, MA, MIT Press, 1995, pp. 423-429.

[40] A. I. Gonzalez, M. Grana and A. D'Anjou, "An analysis of the GLVQ algorithm," *IEEE Trans. on Neural Networks,* vol. 6, no. 4, pp. 1012-1016, 1995.

[41] T. Bojer, B. Hammer, D. Schunk and K. Tluk von Toschanowitz, "Relevance determination in learning vector quantization," *Proc. European Symposium on Artificial Neural Networks (ESANN),* pp. 271-276, 2001.

[42] B. Hammer and T. Villmann, "Generalized relevance learning vector quantization," *Neural Networks,* vol. 15, no. 8-9, pp. 1059-1068, 2002.

[43] M. J. Mendenhall and E. Merenyi, "Relevance-Based Feature Extraction for Hyperspectral Imagery," *IEEE Trans. on Neural Networks,* vol. 19, no. 4, pp. 658-672, 2008.

[44] D. DeSieno, "Adding a conscience to competitive learning," *Proc. of the IEEE Int. Conf. on Neural Networks,* pp. 117-124, 1988.

[45] T. J. Bihl, M. A. Temple and K. Bauer, "Feature Selection Fusion (FSF) for Aggregating Relevance Ranking Information with Application to ZigBee Radio Frequency Device Identification," *IEEE National Aerospace and Electronics Conf. (NAECON),* 2016.

[46] C. K. Dubendorfer, B. W. Ramsey and M. A. Temple, "ZigBee device verification for securing industrial control and building automation systems," *Int. Conf. Critical Infrastructure Protection,* vol. 417, pp. 47-62, 2013.

[47] R. MacCallum, S. Zhang, K. Preacher and D. Rucker, "On the practice of dichotomization of quantitative variables," *Psychological methods,* vol. 7, no. 1, pp. 19-40, 2002.

[48] F. Ortiz, "Dealing with Categorical Data Types in a Designed Experiment Part I: Why You Should Avoiding Using Categorical Data Types," *STAT T&E Center of Excellence*, Wright Patteron AFB, OH, 2012.

[49] F. Ortiz, "Dealing with Categorical Data Types in a Designed Experiment Part II: Sizing a Designed Experiment When Using a Binary Response," *STAT T&E Center of Excellence*, Wright Patteron AFB, OH, 2014.

[50] J. Cohen, "The cost of dichotomization," *Applied Psychological Measurement,* vol. 7, no. 3, pp. 249-253, 1983.

[51] D. Berry and L. Pearson, "Optimal designs for clinical trials with dichotomous responses," *Statistics in Medicine,* vol. 4, no. 4, pp. 497-508, 1985.

[52] P. L. Canner, "Selecting one of two treatments when the responses are dichotomous," *J. of the American Statistical Association,* vol. 65, no. 329, pp. 293-306, 1970.

[53] T. J. Paciencia, *Improving Non-Linear Approaches to Anomaly Detection, Class Separation, and Visualization,* Air Force Institute of Technology: PhD Dissertation, 2014.

[54] M. Strickert, U. Seiffert, N. Sreenivasulu, W. Weschke, T. Villmann and B. Hammer, "Generalized relevance LVQ (GRLVQ) with correlation measures for gene expression analysis," *Neurocomputing,* vol. 69, no. 7-9, pp. 651-659, 2006.

[55] T. Kohonen, J. Kangas, J. Laaksonen and K. Torkkola, "LVQ_PAK: A program package for the correct application of Learning Vector Quantization algorithms," *Proc. Int. Joint Conf. on Neural Networks,* pp. 725-730, 1992.

[56] T.-S. Lim, W.-Y. Loh and Y.-S. Shih, "A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms," *Machine Learning,* vol. 40, pp. 203-229, 2000.

[57] P. Schneider, M. Biehl and B. Hammer, "Distance learning in discriminative vector quantization," *Neural Computation,* vol. 21, no. 10, pp. 2942-2969, 2009.

[58] D. Reising, M. Temple and J. Jackson, "Authorized and Rogue Device Discrimination Using Dimensionally Reduced RF-DNA Fingerprints," *IEEE Trans. on Info. Forensics and Security,* vol. 10, no. 6, pp. 1180-1192, 2015.

[59] S. Bischoff, M. J. Mendenhall, A. C. Rice and J. R. Vasquez, "Adapting learning parameter transition in the generalized learning vector quantization family of classifiers," *Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS),* pp. 1-4, 2010.

[60] L. Wu, K. Yick, S. Ng and J. Yip, "Application of the Box–Behnken design to the optimization of process parameters in foam cup molding," *Expert Systems with Applications,* vol. 39, no. 9, pp. 8059-8065, 2012.

[61] P. Venkataraman, *Applied Optimization with MATLAB Programming*, 2 ed., John Wiley & Sons, 2009.