7-11-2018

Securing ZigBee Commercial Communications Using Constellation Based Distinct Native Attribute Fingerprinting

Christopher M. Rondeau

J. Addison Betances
Air Force Institute of Technology

Michael A. Temple
Air Force Institute of Technology

Recommended Citation

Rondeau, C. M., Betances, J. A., & Temple, M. A. (2018). Securing ZigBee Commercial Communications Using Constellation Based Distinct Native Attribute Fingerprinting. Security and Communication Networks, 2018, Article 1489347. https://doi.org/10.1155/2018/1489347

Hindawi
Security and Communication Networks
Volume 2018, Article ID 1489347, 14 pages
https://doi.org/10.1155/2018/1489347

WILEY | Hindawi

Research Article

Securing ZigBee Commercial Communications Using Constellation Based Distinct Native Attribute Fingerprinting

Christopher M. Rondeau, J. Addison Betances, and Michael A. Temple

Department of Electrical and Computer Engineering, US Air Force Institute of Technology, Wright-Patterson AFB, Dayton, OH 45433, USA

Correspondence should be addressed to Michael A. Temple; michael.temple@afit.edu

Received 19 April 2018; Revised 8 June 2018; Accepted 20 June 2018; Published 11 July 2018

Academic Editor: Bela Genge

This work provides development of Constellation Based DNA (CB-DNA) Fingerprinting for use in systems employing quadrature modulations and includes network protection demonstrations for ZigBee offset quadrature phase shift keying modulation. Results are based on 120 unique networks comprised of seven authorized ZigBee RZSUBSTICK devices, with three additional like-model devices serving as unauthorized rogue devices. Authorized network device fingerprints are used to train a Multiple Discriminant Analysis (MDA) classifier and Rogue Rejection Rate (RRR) estimated for 2520 attacks involving rogue devices presenting themselves as authorized devices. With MDA training thresholds set to achieve a True Verification Rate (TVR) of TVR = 95% for authorized network devices, the collective rogue device detection results for SNR ≥ 12 dB include average burst-by-burst RRR ≈ 94% across all 2520 attack scenarios with individual rogue device attack performance spanning 83.32% < RRR < 99.81%.

1. Introduction

The need to establish reliable and secure communications remains a challenge across commercial Industrial Internet of Things (IIoT) applications that support Critical Infrastructure (CI) elements (water treatment, petroleum product distribution, transportation, etc.) that are commonly operated through Industrial Control System (ICS) architectures. ZigBee networks are common within the IIoT and CI/ICS domains and remain a mainstay for implementing wireless sensor and automation networks supporting medical, smart home and building automation, and consumer electronics [1–3]. The degree of required ZigBee antihacking security varies with application criticality and will increase as the number of deployed ZigBee devices under 802.15.4 market expansion grows to 1 billion units being shipped annually by 2022 and the next generation multiprotocol 802.15.4/Bluetooth/WiFi hardware becomes available [4]. As device makers strive to take advantage of market opportunity and satisfy consumer wants for the next "greatest" interface device, it remains unclear that they have taken necessary prudent steps to address legacy security concerns.

In light of vital asset vulnerability, protection of IIoT CI and ICS elements has become a national-level priority for both the public and private sectors [5–7]. Mitigation strategies against cyberattacks have traditionally focused on bit-level solutions targeting the higher communication protocol layers and until recently there has been minimal emphasis on physical (PHY) layer development [8–10]. This work addresses hardware device identity (ID) verification as a means to enhance network security by preventing unauthorized access through the PHY doorway through which a preponderance of malicious cyberattacks occur. The focus on ZigBee device security is motivated by two factors, including the following: (1) ZigBee and related 802.15.4 communication systems are deployed world-wide and (2) ZigBee serves as a representative protocol for broader IIoT applications [11, 12]. This work expands previous wireless device ID discrimination activity that has successfully exploited various Distinct Native Attribute (DNA) features extracted from selected signal responses to reliably discriminate transmitting hardware devices.

The Constellation Based DNA (CB-DNA) development here is motivated by concepts introduced in [13] used to

discriminate Ethernet cards with features extracted from a contrived (nonconventional) binary constellation. The extension to this earlier work includes (1) formal analytic development of CB-DNA Fingerprinting for systems using conventional M-ary Quadrature Amplitude Modulation (M-QAM) signaling, (2) demonstration of CB-DNA Fingerprinting applicability to ZigBee and related 802.15.4 communication protocols, and (3) proposition of a network device ID process that incorporates mechanisms of localised RF air monitors that have been vetted for other wireless networks [14–17] while achieving security benefits of verification-based Multifactor Authentication (MFA). This proposition includes use of wireless MFA processing with success of the first "something you have" (network compliant device) and second "something you know" (authorized device bit-level ID) checks followed by a final "something you are" (biometric-like CB-DNA fingerprint) check to boost overall security [18, 19]. While comparison of the proposed verification-based rogue detection process with fielded and/or emerging commercial approaches is certainly of interest, a meaningful comparison is not viable given that (1) implementation details of commercial methods are generally proprietary and (2) the statistical effectiveness of such methods is generally unpublished. Regardless, the computational efficiency and speed of biometric-based MFA [18] make it a top-ranked choice for communication device discrimination [19] and it is reasonable to expect similar advantages in MFA-based CB-DNA security applications.

2. Background

2.1. Quadrature Amplitude Modulation (QAM). The general development for the class of complex M-ary QAM modulated signals having in-phase/quadrature-phase (I/Q) components includes the mth complex data modulated symbol given by

$$S_m(t) = I_{S_m} + jQ_{S_m}, \tag{1}$$

for $0 < t < T_{Sym}$ where T_{Sym} is the total symbol duration, $m = 1, 2, \ldots, \mathrm{M}$, and I_{S_m} and Q_{S_m} are real-valued modulation components in the I/Q constellation space with $I_{S_m} \in [I_{S_1}, I_{S_2}, \ldots, I_{S_M}]$ and $Q_{S_m} \in [Q_{S_1}, Q_{S_2}, \ldots, Q_{S_M}]$. For complex symbols given by (1), a transmitted (Tx) burst of N_{Sym} QAM modulated symbols is given by

$$\begin{aligned} \mathrm{S}_{Tx}(t) &= \left[\sum_{m=1}^{N_{Sym}} S_m\left(t - kT_{\mathrm{Sym}}\right)\right] \exp\left(2\pi f_c t + \phi_{Tx}\right), \\ \mathrm{S}_{Tx}(t) &= \left[\sum_{m=1}^{N_{Sym}} S_m\left(t - kT_{Sym}\right)\right] \cos\left(2\pi f_c t + \phi_{Tx}\right) \\ &\quad + j\left[\sum_{m=1}^{N_{Sym}} S_m\left(t - kT_{Sym}\right)\right] \sin\left(2\pi f_c t + \phi_{Tx}\right), \end{aligned} \tag{2}$$

for $0 < t < N_{Sym}\mathrm{X}T_{Sym}$ with f_c being the transmitted carrier frequency and $\phi_{Tx} = \phi/2$ accounting for quadrature-phase error induced by hardware components [21]. The sequence of *ideal transmitted* QAM symbols in $\mathrm{S}_{Tx}(t)$ is denoted by vector $\mathbf{S}_m = (S_1, S_2, \ldots, S_m, \ldots, S_{N_{Sym}-1}, S_{N_{Sym}})$ where $S_m \in [S_1, S_2, \ldots, S_{\mathrm{M}}]$. For the case of M = 4-ary signaling, the QAM $S_{Tx}(t)$ expression in (2) can be used to effectively represent the 4-ary Offset Quadrature-Phase Shift Keyed (O-QPSK) used here for ZigBee demonstration.

Considering channel amplitude A_{Ch} and transmitter-to-receiver propagation delay τ_{Ch} factors, the received (Rx) burst corresponding to $S_{Tx}(t)$ in (2) is given by

$$S_{Rx}(t) = A_{Ch} S_{Tx}\left(t - \tau_{Ch}\right) \tag{3}$$

which has baseband received $I_{Rx}(t)$ and $Q_{Rx}(t)$ components that can be expressed as

$$I_{Rx}(t) = G_{I/Q}\left[\sum_{k=1}^{N_{Sym}} I_{S_k}\left(t - kT_{Sym} - \tau_D\right)\right] + O_I(t), \tag{4}$$

$$Q_{Rx}(t) = G_{I/Q}\left[\sum_{k=1}^{N_{Sym}} Q_{S_k}\left(t - kT_{Sym} - \tau_D\right)\right] + O_Q(t), \tag{5}$$

where $G_{I/Q}$ is the I/Q gain imbalance, τ_D accounts for τ_{Ch} and relative time delay between receiver I/Q channels, and $O_I(t)$ and $O_Q(t)$ represent I/Q offset factors [21]. The $G_{I/Q}$, τ_D, $O_I(t)$, and $O_Q(t)$ factors in (4) and (5) collectively account for transmitter ϕ_{Tx} error in (2) and additional receiver imperfections. The sequence of *corrupted received* QAM symbols in $\mathrm{S}_{Rx}(t)$ is denoted by vector $\mathbf{S}_k = (S_1, S_2, \ldots, S_k, \ldots, S_{N_{Sym}-1}, S_{N_{Sym}})$.

The cumulative effect of transmitter-receiver imperfections and channel errors captured in $I_{Rx}(t)$ and $Q_{Rx}(t)$ components is a degradation in received QAM symbol estimates, denoted here as $\widehat{\mathbf{S}}_k = (\widehat{S}_1, \widehat{S}_2, \ldots, \widehat{S}_k, \ldots, \widehat{S}_{N_{Sym}-1}, \widehat{S}_{N_{Sym}})$ for a given $\mathbf{S}_k$, induced by a location shift of *received* $\mathbf{C}^{S_k} = I_{S_k} + jQ_{S_k}$ QAM constellation points relative to the corresponding *ideal transmitted* $\mathbf{C}^{S_m} = I_{S_m} + jQ_{S_m}$ constellation points. In addition to potential QAM symbol estimation error induced by received $\mathbf{C}^{S_k}$ deviation, there are two other receiver processes that are key for achieving reliable QAM symbol estimation, including (1) received carrier frequency offset f_{Rx} estimation and (2) phase recovery for symbol constellation derotation.

2.1.1. Received Carrier Estimation. Following downconversion by f_c and baseband filtering, samples of the received M-QAM signal at the receiver's Matched Filter (MF) output can be modeled as [22]

$$S_{MF}(n) = \mathrm{K_R} S_k(n) \exp\left(\mathrm{j}2\pi f_{Rx} t\right) + \mathrm{N_B}(n), \tag{6}$$

where $n = 1, 2, \ldots N_{MF}$, K_R is a real-valued scalar, S_k are the transmitted QAM symbols in (2), f_{Rx} is relative received carrier frequency offset, and $\mathrm{N_B}$ is communication channel background noise [22]. The residual f_{Rx} in $S_{Rx}(\mathrm{t})$ can be estimated by raising $S_{MF}(n)$ in (6) to the Mth power to remove the modulation effects. This effectively creates a multitone spectral response with a dominant (highest power) tone

Table 1: Constellation phase derotation algorithm.

Require: Received Constellation Projection $\mathbf{C}^{S_k}$
  RotationVariances $\longleftarrow \infty$
  for $N_\Delta = 1$ to 100 **do**
    $\theta \longleftarrow (N_\Delta \cdot \pi)/(2 \times 100)$
    $\text{Rot}(\mathbf{C}^{S_k}) \longleftarrow \mathbf{C}^{S_k} \cdot e^{j\theta}$
    $\text{Temp} \longleftarrow |\text{Re}[\text{Rot}(\mathbf{C}^{S_k})]| + j|\text{Im}[\text{Rot}(\mathbf{C}^{S_k})]|$
    $\text{RotationVariances}\,(N_\Delta) \longleftarrow \text{Variance}(\text{Temp})$
  end for
  $N_\Delta \longleftarrow \text{argmin}_{N_\Delta}[\text{RotationVariances}]$
  return $\mathbf{C}^{S_k} \cdot e^{j(N_\Delta \cdot \pi)/(2\times 100)}$

occurring at $\text{M}\times f_{Rx}$ [23]. This is illustrated for 4-QAM where $S^4_{MF}(n)$ can be expanded as

$$\begin{aligned} S^4_{MF}(n) = {} & \left[\text{K}_\text{R} S_m(n)\right]^4 \exp\left(\text{j}8\pi f_{Rx} t\right) \\ & + 4\left[\text{K}_\text{R} S_m(n)\right]^3 \exp\left(j6\pi f_{Rx} t\right) \text{N}_\text{B}(n) \\ & + 4\left[\text{K}_\text{R} S_m(n)\right] \exp\left(j2\pi f_{Rx} t\right) \text{N}^3_\text{B}(n) \\ & + 6\left[\text{K}_\text{R} S_m(n)\right]^2 \exp\left(j4\pi f_{Rx} t\right) \text{N}^2_\text{B}(n) \\ & + \text{N}^4_\text{B}(n) \end{aligned} \tag{7}$$

which includes a dominant $8\pi f_{Rx} = 2\pi(4f_{Rx})$ frequency component. The estimated received carrier frequency offset is given by $\widehat{f}_{Rx} = 4[\arg\max_n(\mathcal{F}|S^4_{MF}(n)|)]$ where $\mathcal{F}(\cdot)$ denotes the discrete Fourier transform.

2.1.2. Constellation Phase Recovery. Receivers commonly use a Phase Locked Loop (PLL) to reconstruct the suppressed carrier via dynamic feedback that autocompensates for phase errors [24]. While generally beneficial, this within-burst autocompensation can potentially obscure subtle DNA feature differences that may help discriminate transmitters. Therefore, burst-by-burst discrete phase estimation and constellation derotation was implemented here using an algorithm that rotates the received $\mathbf{C}^{S_k}$ constellation points for each burst from 0 to $\pi/2$ radians in $N_\Delta = 100$ increments and selects the phase rotation angle yielding the minimum variance between the incrementally rotated pool of received $\mathbf{C}^{S_k}$ and the ideal $\mathbf{C}^{S_m}$ constellation points. The pseudocode for implementing this algorithm is presented in Table 1.

There are four different phase angle ambiguities that can exist after derotating the constellation using the algorithm in Table 1. These are resolved using estimated rotation angles of known preamble (training) symbols. The rotated constellation projections can also be normalized by scaling (dividing) each $\text{Rot}(\mathbf{C}^{S_k})$ point by the mean $(|\text{Rot}(\mathbf{C}^{S_k})|)$ which locates the center of all constellation clusters on the unit circle.

2.2. ZigBee Communications. The ZigBee Communication protocol includes a Medium Access Control (MAC) layer, where device IDs are verified using bit-level credentials, that interfaces with the RF communications channel through the PHY layer using RF hardware and firmware [25]. The PHY layer is implemented according to the IEEE 802.15.4 standard for low data-rate, low-power, and short range RF communications [20]. It is estimated that more than one billion 802.15.14 compliant components will be sold by the end of this decade with a majority of them supporting localised smart home networks [4]. One such component is the Atmel AT86RF230 radio transceiver that is hosted on RZUSBSTICK devices [26]. These are small low-power devices that support ZigBee operation at 2.4 GHz through an integrated folded dipole antenna with a net peak gain of $G_A = 0$ dB. Accounting for $G_A = 0$ dB and maximum AT86RF230 output power of $\text{P}_{\text{Out}} = +3.0$ dBm [27], the effective transmit power of the RZUSBSTICK is $\text{P}_{\text{Tx}} = +3.0$ dBm which make it a viable alternative for not only smart home networks but other wireless sensor networks, industrial control system, and building automation [27]. Details for the specific RZUSBSTICK devices used for demonstration are provided in Table 2 which shows the unique ZigBee Communication (ZC) device IDs assigned for experimentation.

Table 2: ZigBee RZUSBSTICK device details showing the device ID, the digital MAC address, and two unique physical markings appearing on the device AT86RF230 transceiver chips.

ID	MAC	Mark 1	Mark 2
ZC1	A0:F6:9F:E7	1442 PH	1R8338-7
ZC2	A0:01:43:70	0923 PH	8P0772
ZC3	A0:01:5D:34	0936 PH	9P0187-2
ZC4	A0:F6:A0:68	1442 PH	1R8338-7
ZC5	A0:F6:A0:4E	1442 PH	1R8338-7
ZC6	A0:F6:9F:FF	1442 PH	1R8338-7
ZC7	A0:F6:A0:0C	1442 PH	1R8338-7
ZC8	A0:F6:A0:04	1442 PH	1R8338-7
ZC9	A0:F6:9F:EA	1442 PH	1R8338-7
ZC10	A0:F6:9F:E0	1442 PH	1R8338-7

The use of PHY layer O-QPSK modulation is mandatory for ZigBee operation at 2.4 GHz, with the O-QPSK modulator preceded by a 4-to-32 (information bit-to-spread chip) Pseudorandom Noise (PN) mapping such that the information bits are transmitted at an effective rate of (2M Chips/Sec) × (4/32 Bits/Chip) = 250K Bits/Sec [20, 25]. Accounting for I/Q channel offset processing in the modulator, the corresponding output O-QPSK communication symbol rate for a transmitted $\text{S}_{Tx}(t)$ burst given by (2) is $R_{Sym} = 1/T_{Sym} =$ (250K Bits/Sec)/(2 Bits/Sym) = 125K Sym/Sec.

The required 4-to-32 PN mapping for 2.4 GHz ZigBee operation is shown in Table 3 [20]. Given this mapping, there are specific transmitted O-QPSK $\mathbf{S}_m$ symbol sequences that occur with varying probability. For example, the bold highlighted {1 0 0 1 0 0} 6-bit pattern in the output chip sequences in Table 3 is among the most frequently occurring ones (appears in 13 of 16 chip sequences) and produces the O-QPSK transmitted symbol sequence $\mathbf{S}_m = (S_2, S_2, S_3, S_3, S_3)$. This 5-symbol $\mathbf{S}_m$ vector is denoted in Table 4 by an $*$ and

Table 3: Input-output sequences for ZigBee 4-to-32 premodulation PN mapping [20]. Bold entries highlight one of 30 highest probability 6-bit sequences.

Input $\{b_0, b_1, b_2, b_3\}$	Output Chip Sequence $\{c_0, c_1, c_2, \ldots, c_{31}\}$
0000	11011001110000110101**100100**0101110
1000	11101101100111000011010**100100**010
0100	00101110110110011100001101010010
1100	00100010111011011001110000110101
0010	010**100100**010111011011001110000011
1010	00110101**100100**010111011011001100
0110	11000011010**100100**01011011011001
1110	10011100001101010**100100**01011011101
0001	10001**100100**10110000001110111011
1001	10111000011**100100**10110000001110111
0101	01111011100011**100100**1011000000111
1101	01110111011100011**100100**101100000
0011	00000111011110111000**100100**10110
1011	01100000011101111011000**100100**1
0111	10010110000001110111011100011100
1111	1**100100**10110000001110111110111000

Table 4: 30 highest probability 5-symbol $\mathbf{S}_m$ for Table 3 mapping with $*$ denoting $\mathbf{S}_m$ for the output bit sequence $\{1\ 0\ 0\ 1\ 0\ 0\}$ highlighted in Table 3.

$(S_1, S_1, S_1, S_1, S_2)$	$(S_3, S_1, S_1, S_3, S_3)$
$(S_1, S_3, S_3, S_4, S_4)$	$(S_3, S_3, S_4, S_4, S_4)$
$(S_2, S_2, S_2, S_4, S_3)$	$(S_4, S_2, S_2, S_4, S_4)$
$(S_2, S_4, S_4, S_4, S_3)$	$(S_1, S_3, S_3, S_1, S_2)$
$(S_3, S_3, S_4, S_4, S_4)^*$	$(S_2, S_2, S_1, S_3, S_3)$
$(S_4, S_2, S_2, S_2, S_2)$	$(S_2, S_4, S_4, S_2, S_1)$
$(S_1, S_1, S_1, S_3, S_4)$	$(S_3, S_1, S_1, S_3, S_4)$
$(S_1, S_3, S_4, S_2, S_1)$	$(S_4, S_2, S_1, S_1, S_2)$
$(S_2, S_4, S_3, S_1, S_1)$	$(S_4, S_4, S_3, S_1, S_2)$
$(S_3, S_3, S_1, S_1, S_2)$	$(S_1, S_3, S_3, S_1, S_2)$
$(S_3, S_3, S_4, S_2, S_2)$	$(S_2, S_2, S_2, S_2, S_1)$
$(S_4, S_2, S_2, S_4, S_3)$	$(S_2, S_4, S_4, S_2, S_2)$
$(S_1, S_1, S_2, S_4, S_4)$	$(S_3, S_1, S_2, S_2, S_2)$
$(S_1, S_3, S_4, S_2, S_2)$	$(S_4, S_2, S_1, S_3, S_3)$
$(S_2, S_4, S_3, S_1, S_2)$	$(S_4, S_4, S_3, S_3, S_4)$

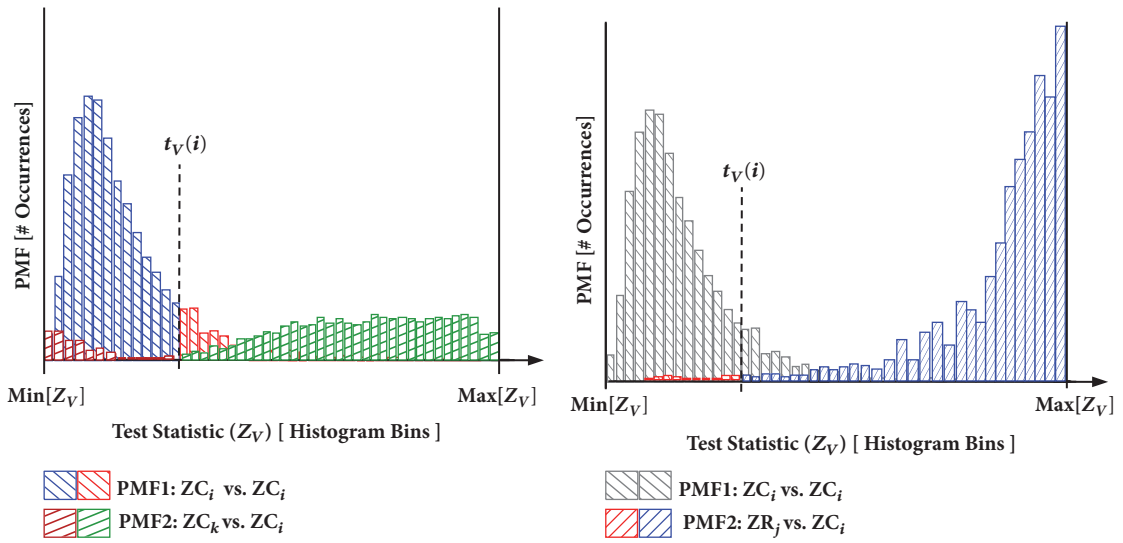
is among the 30 highest probability transmitted O-QPSK $\boldsymbol{S}_m$ used for conditional CB-DNA demonstration.

2.3. Device Classification and Device ID Verification. Device discrimination (classification and ID verification) is performed using DNA fingerprints with a Multiple Discriminant Analysis/Maximum Likelihood (MDA/ML) process adopted from [11]. This includes MDA model training for N_{Cls} classes (ZC devices) with components of (1) an N_F x N_{Cls}-1 dimensional matrix $\mathbf{W}$ for projecting 1x N_F dimensional input fingerprints ($\mathbf{F}$) into the N_{Cls}-1 discrimination space containing fingerprint projection $\mathbf{P_F} = \mathbf{FW}$; (2) an 1x N_F dimensional fingerprint scaling vector $\boldsymbol{\alpha}$; and (3) the N_{Cls} training means ($\boldsymbol{\mu}$) and covariances ($\boldsymbol{\Sigma}$). MDA models are generated using a pool of 4400 total fingerprints per class that are equally divided into N_{TNG} = 2200 Training (even indexed fingerprints) and N_{TST} = 2200 Testing (odd indexed fingerprints) subsets. The even-odd indexing assignment ensures the models account for temporal channel variation, collection bias, etc., effects occurring during the course of emission collection.

The TNG fingerprints at a given SNR are used for MDA model training that includes K = 5-fold cross-validation [Dud1] with the best projection matrix $\mathbf{W}_{\text{Best}}$ selected as the fold $\mathbf{W}$ producing the highest cross-validation accuracy. The TST fingerprints are then input to the model and a 1 versus N_{Cls} best match ML classification decision is made based on a selected classification test statistic (Z_{Cls}). The trained class yielding highest conditional probability P(ZCi | Z_{Cls}) for all $i = 1, 2, \ldots, N_{Cls}$ is the called class (right or wrong) assigned to the unknown input fingerprint $\mathbf{F}$. Classification performance at a given SNR is presented in an N_{Cls} x N_{Cls} (input versus called) classification confusion matrix, with (1) average *cross-class* percent correct classification (%C) calculated as the sum of diagonal (correct) matrix entries divided by the total number of classification trials (N_{Cls} x N_{TST}) and (2) individual class %C for each class C_i calculated as the sum of ith row entries divided by N_{TST}. Alternately, classification performance is presented in %C versus SNR plots.

The device ID verification process uses the selected MDA model components ($\mathbf{W}$, $\boldsymbol{\alpha}$, $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$) and device TST fingerprints to estimate both (1) *authorized* network device True Verification Rate (TVR) (true positive) and (2) *unauthorized* device Rogue Rejection Rate (RRR) (true negative). For a given claimed (unknown) authorized device ID to be verified, the process includes the following: (1) projecting TST $\mathbf{F}$ fingerprints for the device under test into the N_{Cls}-1 discrimination space using $\mathbf{P_F} = \boldsymbol{\alpha} \otimes \mathbf{FW}$ where $\otimes$ denotes element-by-element vector multiplication, (2) calculating the selected verification test statistic (Z_V) for N_{TST} total fingerprints using training $\boldsymbol{\mu}$ and/or $\boldsymbol{\Sigma}$ for the claimed authorized device ID, (3) forming a normalized (unit area) Probability Mass Function (PMF) using N_{TST} total Z_V, (4) overlaying a desired training verification threshold (t_V), and (5) calculating the PMF area above/below t_V to estimate the desired verification rate. Common Z_V measures of similarity include (1) distance-based metrics such as the Euclidean distance between projected $\mathbf{P_F}$ and the claimed training class mean $\boldsymbol{\mu}$ and (2) probability-based metrics that map the calculated $\mathbf{P_F}$ Euclidean distance to a normalized multivariate Gaussian probability distribution having mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. Euclidean distance is perhaps the most easily conceptualised and was chosen here for proof-of-concept demonstration.

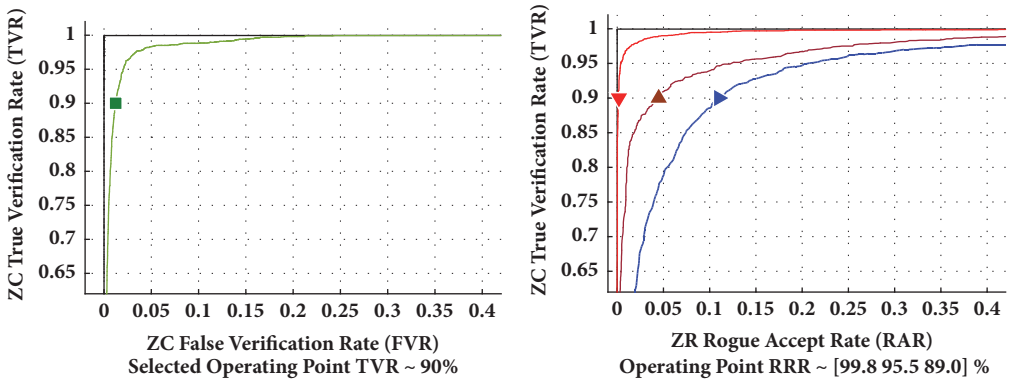
The PMFs in Figure 1 are used to illustrate *Device ID verification* for Euclidean distance "lower-is-better" measure of similarity [11]. Given these PMFs, the ID verification process includes (1) using network ZC TNG fingerprint Z_V to set the training verification threshold $t_V(i)$ shown in Figure 1(a) to achieve the desired TVR (blue PMF1 area) where PMF1 is for ZCi TNG and PMF2 is based on accumulated TNG Z_V for



(a) Network ZC*k* versus Network ZC*i* (ZC*k*:ZC*i*): PMF1 for ZC*i* TNG and PMF2 for all "other" network ZC*k*

(b) Rogue ZR*j* versus Network ZC*i* (ZR*j*:ZC*i*): PMF1 for ZC*i* TNG and PMF2 for ZR*j* versus ZC*i*

FIGURE 1: PMFs showing device dependent $t_V(i)$ set to achieve desired network ZC*i* TVR (true positive) given by blue PMF1 area in (a) and resultant RRR (true negative) for ZR*j* device given by blue PMF2 area in (b) [11].



(a) Network ZC *Training* ROC

(b) Rogue ZR *Testing* ROCs

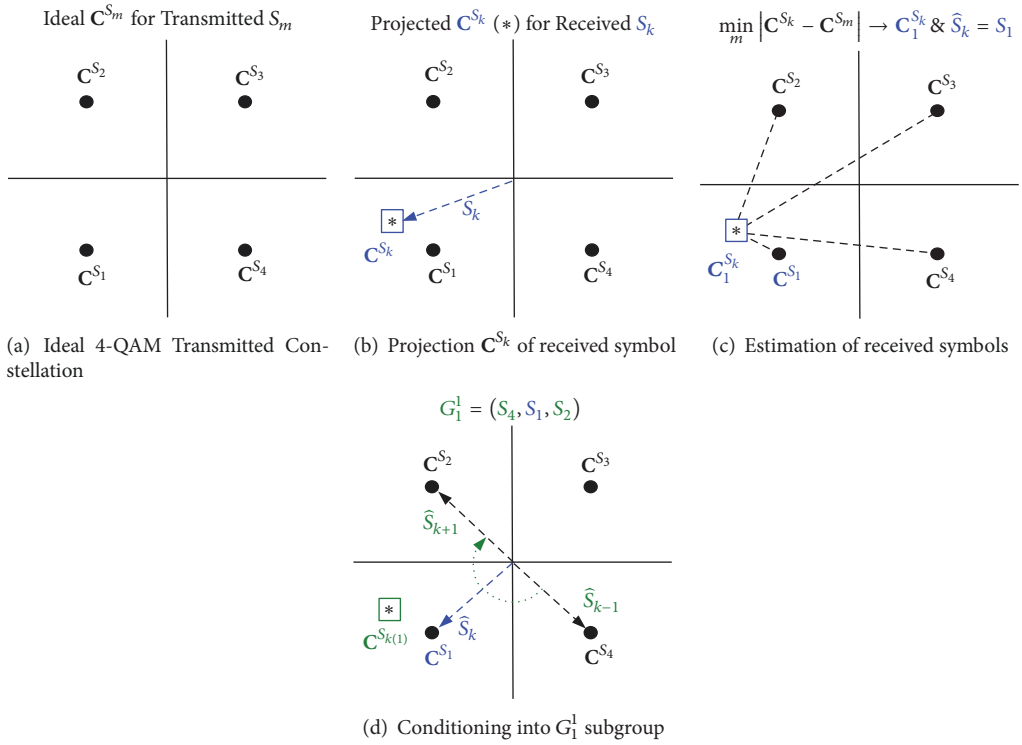
FIGURE 2: ROC curves for Figure 1 PMFs with indicated operating points based on desired TVR = 90%.

all "other" network ZC*k* ($k = 1, 2, \ldots, N_{Cls}$ and $k \neq i$) and (2) calculating the corresponding RRR (true negative, blue PMF2 area) in Figure 1(b) where PMF1 is the same and PMF2 is based on TST fingerprint Z_V for the rogue ZR*j* device. ID verification performance can be based on TNG $t_V(i)$ set for either (1) equal error rate conditions with False Verification Rate (FVR) given by FVR = 1-TVR or (2) a specific desired authorized TVR.

The authorized TVR (true positive) versus FVR (false positive) trade-off is effectively captured in a Receiver Operator Characteristic (ROC) curve [Faw1] as shown in Figure 2 using Figure 1 PMFs with varying the TNG verification threshold t_V varied from Min[Z_V] to Max[Z_V]. Figure 2(a) shows TVR versus FVR with the indicated operating point (■) corresponding to desired TVR = 90% and yielding FVR ≈ 1.2%. Figure 2(b) shows TVR versus RAR where Rogue Accept Rate (false positive) is used to estimate the RRR ≈ 1-RAR shown along the x-axis for three arbitrary ZR devices (▼, ▲, and ▶) and the TVR = 90% operating point.

3. CB-DNA Fingerprinting Development

Time domain RF-DNA Fingerprinting has historically exploited statistical features extracted from partial-burst responses where *invariant* (data independent) synchronisation and channel estimation (preamble, midamble, etc.) symbols are transmitted [15, 28–30]. The CB-DNA Fingerprinting method developed here differs considerably and exploits features extracted from full-burst responses, including regions where *variant* (data dependent) symbols are transmitted. The CB-DNA Fingerprinting development here is motivated by concepts first used in [13] to discriminate



(a) Ideal 4-QAM Transmitted Constellation

(b) Projection $\mathbf{C}^{S_k}$ of received symbol

(c) Estimation of received symbols

(d) Conditioning into G_1^1 subgroup

Figure 3: Illustration of unconditional and conditional 4-QAM constellation processing.

Ethernet cards but it fundamentally differs in that work in [13] is based on features extracted from a contrived (nonconventional) binary constellation while the development here is for any application using conventional M-QAM signaling as introduced in Section 2.1. The development for *unconditional* and *conditional* fingerprinting is supported by the process depicted in Figure 3.

For ideal transmitted symbols having constellation projections $\mathbf{C}^{S_m}$ such as those shown in Figure 3(a), the kth received QAM symbol in burst $S_{Rx}(t)$ of (3) is denoted as S_k for $t_k < t < t_k + T_{Sym}$ where t_k is the symbol start time, T_{Sym} is the symbol duration, and $k = 1, 2, \ldots, N_{Sym}$ where N_{Sym} is the total number of symbols in a received burst. Following synchronisation to the kth symbol interval, the QAM receiver extracts symbol S_k and projects it to a single point $\mathbf{C}^{S_k}$ in the QAM constellation space (Figure 3(b)). The corresponding estimated transmitted symbol is determined as $\widehat{S}_k = S_m$: $\arg\min_m |\mathbf{C}^{S_k} - \mathbf{C}^{S_m}|$ for $S_m \in [S_1, S_2, \ldots, S_M]$ (Figure 3(c)). For generating *unconditional* CB-DNA statistical fingerprint features, the N_{Sym} received $\mathbf{C}^{S_k}$ in each $S_{Rx}(t)$ burst are grouped based on their corresponding $\widehat{S}_k = S_m$ estimate with the group of $\mathbf{C}^{S_k}$ yielding the mth QAM symbol estimate denoted by the sequence $\{\mathbf{C}_m^{S_k}\}$ for $m = 1, 2, \ldots, M$.

While some prior works have investigated constellation error differences as a means for device discrimination [31], e.g., mean and variance, of Euclidean distances between received $\mathbf{C}^{S_k}$ and ideal $\mathbf{C}^{S_m}$, the approach here exploits constellation spatial statistical differences in $\{\mathbf{C}_m^{S_k}\}$ groups which are induced by channel propagation and hardware variability (e.g., I/Q imbalance) resulting from component differences (oscillator phase noise, spurious mixer tones, manufacturing processes, etc.) [21]. The exploitation of these differences was first demonstrated for the contrived binary constellation work in [13] which showed that the statistical distribution of $\{\mathbf{C}_m^{S_k}\}$ elements around the corresponding ideal $\mathbf{C}^{S_m}$ point is *conditional*, i.e., the location of a given $\mathbf{C}_m^{S_k}$ for S_k in the received QAM constellation space is dependent upon symbols received just prior to and immediately following S_k; these two symbols are denoted as $S_{k\text{-}1}$ and S_{k+1}, respectively.

The device discrimination improvement in [13] using conditional fingerprint features from the contrived binary constellation motivated formal development of the *multisymbol constellation conditioning* (subgrouping) method for M-QAM signaling. For the $\widehat{S}_k$ dependent $\{\mathbf{C}_m^{S_k}\}$ group sequences, the basic process includes considering multiple consecutive received QAM symbols in a $S_{Rx}(t)$ burst which are denoted here by vector $\mathbf{S}_k = (\ldots, S_{k-2}, S_{k-1}, S_k, S_{k+1}, S_{k+2}, \ldots)$ where S_k is the central reference symbol. These received symbols have corresponding estimates that are used to form vector $\widehat{\mathbf{S}}_k = (\ldots, \widehat{S}_{k-2}, \widehat{S}_{k-1}, \widehat{S}_k, \widehat{S}_{k+1}, \widehat{S}_{k+2}, \ldots)$ where $\widehat{S}_k$ is the estimate for reference symbol S_k. Multisymbol constellation conditioning involves parsing each of the *unconditional*$\{\mathbf{C}_m^{S_k}\}$ groups into *conditional*$\{\mathbf{C}_m^{S_{k(n)}}\}$ subgroups for $n = 1, 2, \ldots, N_{SG}$ total subgroups with $S_{k(n)}$ denoting the nth subgroup. The parsing of *unconditional*$\{\mathbf{C}_m^{S_k}\}$ sequences and selection of N_{SG} subgroups is somewhat arbitrary but performed with a goal of maximising cross-subgroup distribution differences that will be captured in statistical fingerprint features.

The subgrouping of $\{\mathbf{C}_m^{S_k}\}$ is illustrated (as shown in Figure 3(d)) by considering three received symbols of $\mathbf{S}_k = (S_{k-1}, S_k, S_{k+1})$ and a set of N_{SG} desired subgroup conditioning vectors $\mathbf{G}^n$ of equivalent dimension and denoted by $\mathbf{G}^n = (G_1^n, G_2^n, G_3^n)$ where $G_i^n \in [S_1, S_2, \ldots, S_M]$. The process for assigning each element of the mth $\{\mathbf{C}_m^{S_k}\}$ group to one of N_{SG} subgroups based on $\mathbf{G}^n$ conditions includes (1) taking each received S_k producing $\mathbf{C}_m^{S_k}$, (3) estimating received $\widehat{S}_{k-1}$ and $\widehat{S}_{k+1}$ and forming $\widehat{\mathbf{S}}_k = (\widehat{S}_{k-1}, \widehat{S}_k, \widehat{S}_{k+1})$, and (4) comparing the resultant $\widehat{\mathbf{S}}_k$ with each desired $\mathbf{G}^n$. If $|\widehat{\mathbf{S}}_k - \mathbf{G}^n| = 0$ for some $n = 1, 2, \ldots, N_{SG}$ the $\mathbf{C}_m^{S_k}$ under evaluation is assigned to the nth conditional $\{\mathbf{C}_m^{S_{k(n)}}\}$ subgroup. If $|\widehat{\mathbf{S}}_k - \mathbf{G}^n| \neq 0$ for all n the $\mathbf{C}_m^{S_k}$ under evaluation is assigned to an "other" conditional subgroup. Formation of the $N_{SG} + 1$ "other" subgroup is required when all possible combinations of estimated $\widehat{\mathbf{S}}_k$ symbols are not included as desired $\mathbf{G}^n$ conditions and ensures that all elements of $\{\mathbf{C}_m^{S_k}\}$ are accounted for. Accounting for all possible M-QAM symbols, the total number of conditional subgroups formed for fingerprint generation is either $M \times N_{SG}$ or $M \times N_{SG}+1$ if an "other" subgroup is required.

There are many possible symbol combinations that could be used for conditioning $\mathbf{G}^n$ vectors and formation of conditional subgroups. In light of noted M-QAM I/Q phase imbalance effects, there are some specific $\mathbf{G}^n$ that may accentuate cross-subgroup differences based on how the phase in consecutive $S_{Rx}(t)$ symbols changes during QAM signaling. The two extreme phase changes are captured using (1) $\mathbf{G}^n = (\widehat{S}_k, \widehat{S}_k, \widehat{S}_k)$ which represents the case of no symbol-to-symbol phase change across $\widehat{\mathbf{S}}_k$ symbols and (2) $\mathbf{G}^n = (-\widehat{S}_k, \widehat{S}_k, -\widehat{S}_k)$ which represents the case of maximum ±180 degrees' symbol-to-symbol phase change across $\widehat{\mathbf{S}}_k$ symbols. Considering 4-QAM and accounting for all possible symbol combinations in the 1x3-dimensional $\mathbf{G}^n$ vectors, there are a total of $N_{SG} = 16$ conditional $\{\mathbf{C}_m^{S_{k(n)}}\}$ subgroup sequences for $m = 1, 2, 3, 4$ with no "other" subgroup formed. The effect of conditional subgrouping is illustrated with the aid of Figure 4 which shows an unconditioned QAM received constellation for an $S_{Rx}(t)$ burst at SNR = 12 dB and containing approximately $N_{Sym} \approx 3400$ total symbols (approximately 850 $\mathbf{C}_m^{S_k}$ projections per quadrant).

Considering the *S1* quadrant and selected conditional $\mathbf{G}^n$ symbol vectors yields the pairwise conditional $\{\mathbf{C}_1^{S_{k(n)}}\}$ projections plotted in Figure 5. Of note in Figure 5 is that all plots are presented on the same scale over the same I-Value and Q-Value ranges. Thus, the observable similarities and/or differences in the illustrated conditional $\{\mathbf{C}_1^{S_{k(n)}}\}$ subgroups exhibit behavior that is indicative of I/Q imbalance and increase the potential for device characterisation. Assuming identical channel conditions and receiver imperfection effects (I/Q imbalance, etc.) during the signal collection interval, the visually discernable differences in conditional $\{\mathbf{C}_1^{S_{k(n)}}\}$ subgroup distributions in Figure 5 are attributable to transmitter component differences and aid in uniquely identifying transmitting devices using conditional CB-DNA Fingerprinting.

Statistical features of *unconditional*$\{\mathbf{C}_m^{S_k}\}$ sequences and *conditional*$\{\mathbf{C}_m^{S_{k(n)}}\}$ sequences are used to form CB-DNA fingerprints. The construction processes for *unconditional*$(\mathbf{F}_{CB}^{UNC})$ and *conditional*$(\mathbf{F}_{CB}^{CND})$ CB-DNA fingerprint vectors are identical and presented for an arbitrary complex sequence {X} having N_X elements. The fingerprint statistics are calculated using (1) *polar* magnitude (*Mag*) and angle (*Ang*) components and (2) *rectangular* real (*Re*) and imaginary (*Im*) components of {X}. While any number of statistics could be used, the specific statistical CB-DNA features used for polar representation include variance (σ^2), skewness (γ), and kurtosis (κ) statistics of both the magnitude $\{Mag[X]\}$ and angle $\{Ang[X]\}$ sequences for a total of 6 polar statistics. For the rectangular $[Re\{X\}:Im\{X\}]_{2\times N_X}$ matrix representation, the calculated statistics include three unique covariance $\sigma^2\sigma^2{}_{(1:3)}$ values, two nontrivial coskewness moments $\gamma\gamma_{(1:2)}$, and three nontrivial cokurtosis $\kappa\kappa_{(1:3)}$ moments [32]. Accounting for all possible statistics, the *Statistical Fingerprint* vector for complex sequence {X} is formed as

$$\mathbf{F}^X = \left[\sigma^2_{Mag(X)} \;\; \gamma_{Mag(X)} \;\; \kappa_{Mag(X)} \;\; \sigma^2_{Ang(X)} \;\; \gamma_{Ang(X)} \;\; \kappa_{Ang(X)} \;\; \sigma\sigma^2_{X(1:3)} \;\; \gamma\gamma_{X(1:2)} \;\; \kappa\kappa_{X(1:3)}\right]_{1\times N_{Stat}}, \tag{8}$$

where $N_{Stat} = 14$ if all indicated statistics are included.

For *unconditional* CB-DNA Fingerprinting $\mathbf{F}^X$ in (8) is calculated for all $m = 1, 2, \ldots, M$ constellation symbols with $\{X\} = \{\mathbf{C}_m^{S_k}\}$ and the resultant $\mathbf{F}_m^X$ concatenated to form the final composite *unconditional CB-DNA Fingerprint* vector $\mathbf{F}_{CB}^{UNC}$ given by

$$\mathbf{F}_{CB}^{UNC} = \left[\mathbf{F}_1^X \vdots \mathbf{F}_2^X \cdots \mathbf{F}_M^X\right]_{1\times N_F^{UNC}}, \tag{9}$$

where $N_F^{UNC} = N_{Stat} \times M$ is the total number of *unconditional* CB-DNA features.

For *conditional* CB-DNA Fingerprinting $\mathbf{F}^X$ in (8) is calculated for all $n = 1, 2, \ldots, N_{SG}$ subgroups of each $m = 1, 2, \ldots, M$ constellation symbol using $\{X\} = \{\mathbf{C}_m^{S_{k(n)}}\}$. The resultant $\mathbf{F}_{SG(m,n)}^{CND}$ vectors are used form the mth *Conditional CB-DNA Fingerprint* vector $\mathbf{F}_m^{CND}$ given by

$$\mathbf{F}_m^{CND} = \left[\mathbf{F}_{SG(m,1)}^{CND} \vdots \mathbf{F}_{SG(m,2)}^{CND} \vdots \cdots \vdots \mathbf{F}_{SG(m,N_{SG})}^{CND}\right]_{1\times(N_{Stat}\times N_{SG})}, \tag{10}$$

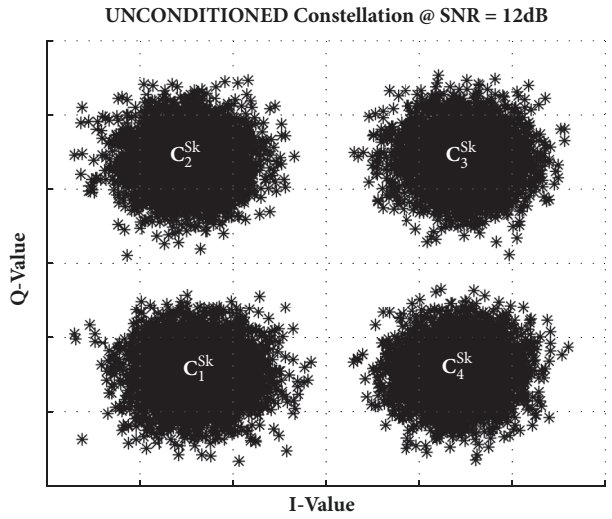
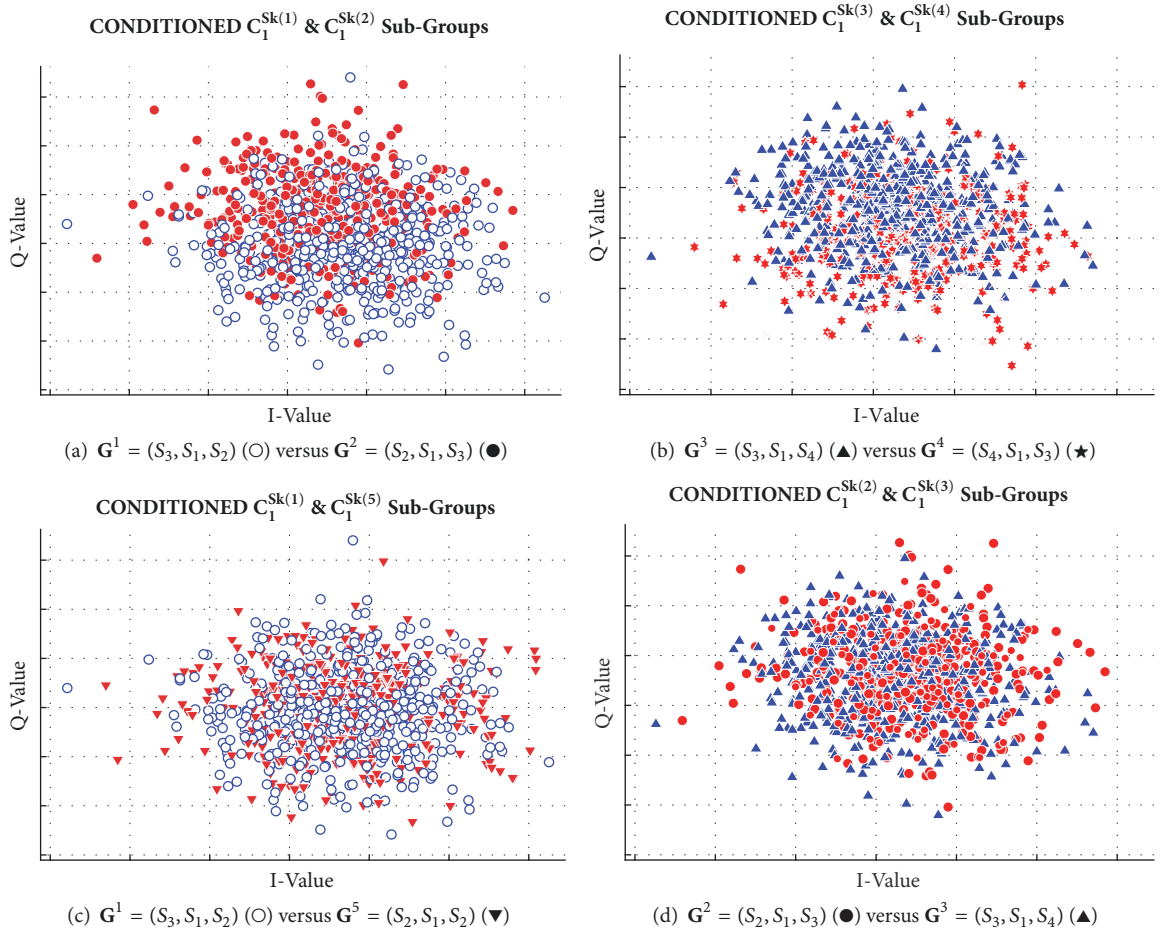


FIGURE 4: Received *unconditioned* QAM constellation at SNR = 12 dB for a burst of $N_{Sym} \approx 3400$ symbols producing approximately 850 total projections in the indicated $\mathbf{C}_1^{S_k}$, $\mathbf{C}_2^{S_k}$, $\mathbf{C}_3^{S_k}$, and $\mathbf{C}_4^{S_k}$ quadrant groups.



(a) $\mathbf{G}^1 = (S_3, S_1, S_2)$ (○) versus $\mathbf{G}^2 = (S_2, S_1, S_3)$ (●)

(b) $\mathbf{G}^3 = (S_3, S_1, S_4)$ (▲) versus $\mathbf{G}^4 = (S_4, S_1, S_3)$ (★)

(c) $\mathbf{G}^1 = (S_3, S_1, S_2)$ (○) versus $\mathbf{G}^5 = (S_2, S_1, S_2)$ (▼)

(d) $\mathbf{G}^2 = (S_2, S_1, S_3)$ (●) versus $\mathbf{G}^3 = (S_3, S_1, S_4)$ (▲)

FIGURE 5: Received *conditional* QAM constellation points for S_1 quadrant projections in Figure 4 showing pairwise relationship of five conditioned $\{\mathbf{C}_1^{S_{k(n)}}\}$ subgroups with elements assigned using the indicated $\mathbf{G}^n$ conditions.

which are concatenated for all $m = 1, 2, \ldots, \mathrm{M}$ to form the composite *conditional CB-DNA Fingerprint* vector

$$\mathbf{F}_{\mathrm{CB}}^{\mathrm{CND}} = \left[\mathbf{F}_1^{\mathrm{CND}} \vdots \mathbf{F}_2^{\mathrm{CND}} \vdots \cdots \vdots \mathbf{F}_{\mathrm{M}}^{\mathrm{CND}}\right]_{1\times N_F^{CND}}, \quad (11)$$

where $N_F^{CND} = N_{Stat} \times N_{SG} \times \mathrm{M}$ is the total number of *conditional* CB-DNA features. In general, unconditional and conditional CB-DNA fingerprint features can be generated using all or a subset of noted statistics, calculated for all or a subset of available projected $\{\mathbf{C}_m^{S_k}\}$ groups or $\{\mathbf{C}_m^{S_{k(n)}}\}$ subgroups. The choice of which statistics and which groups to use may vary with the specific communication application (fixed, mobile, urban, city, etc.) and determines the final number of N_F^{UCB} and N_F^{CCB} features generated.

4. CB-DNA Fingerprinting Demonstration

ZigBee transmissions were collected for all RZUSBSTICK devices listed in Table 2 using an X310 Software Defined Radio (SDR) having an RF bandwidth of $W_{RF} = 10$ MHz and operating at a sampling rate of $f_S = 10$ MSps in both the I/Q channels. Subsequent postcollection signal processing was performed using MATLAB and included burst-by-burst (1) center frequency estimation, (2) baseband (BB) downconversion and filtering using a 16th-order Butterworth filter having a -3 dB bandwidth of $W_{BB} = 2$ MHz, (3) constellation phase derotation, and (4) unconditional and conditional CB-DNA fingerprint generation per Section 3. The CB-DNA fingerprints were used to generate demonstration results for a total of $N_{NC} = 10$-choose-3 = 120 unique network configurations with the $N_{ZR} = 3$ chosen devices serving as *unauthorized* attacking ZigBee Rogue (ZR) devices and the remaining $N_{ZC} = 7$ devices serving as *authorized* ZC network devices.

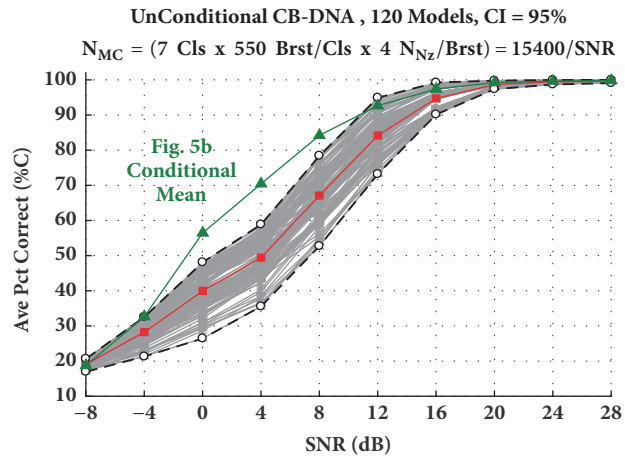
For each network configuration, the RRR was estimated for the $N_{ZR} = 3$ rogue devices using the device ID verification process detailed in Section 2.3. For each network configuration, each of the $N_{ZR} = 3$ ZR devices presents false ID credentials for all $N_{Cls} = 7$ authorized ZC network devices for a total of $7 \times 3 = 21$ ZR*j*:ZC*i* assessments per network configuration. Considering all networks, a total of $120 \times 21 = 2520$ ZR*j*:ZC*i* device ID verification (rogue detection) assessments were completed. Alternately, each ZC device in Table 2 served as an attacking ZR device 36 times for a total of $36 \times 7 = 252$ ZR*j*:ZC*i* device ID verification assessments per RZUSBSTICK device. The RRR estimates are based on a total of 4400 fingerprints per ZR device that are presented on a fingerprint-by-fingerprint basis for ID verification; the assessments here do not include nor account for envisioned benefits to be realised by averaging fingerprints, features, etc., prior to making a final authorized versus rogue verification decision. For presentation brevity, limited results are presented herein that are representative of the poorest (lowest RRR) and best (highest RRR) results obtained across all $N_{NC} = 120$ network configurations and are sufficient for supporting proof-of-concept demonstration conclusions.

4.1. Authorized Network Device Classification. Device classification is first required to generate the MDA/ML models ($\mathbf{W}$, $\boldsymbol{\alpha}$, $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$) required for device ID verification. The CB-DNA Fingerprinting results in Figure 6 were generated using unconditional and conditional features for all $N_{NC} = 120$ networks. Results show %C versus SNR for all 120 networks along with cross-network average %C (solid lines) and extreme bounds (dashed lines with ○ markers) for highest and lowest %C. The benefit of constellation conditioning is evident by comparing cross-network averages which show that the %C = 90% benchmark is achieved for *conditional* features (■) at SNR ≈ 11 dB and *unconditional* features (▲) at SNR ≈ 14 dB. For presentation brevity, additional results in this section are presented for conditional CB-DNA Fingerprinting only given its superiority.

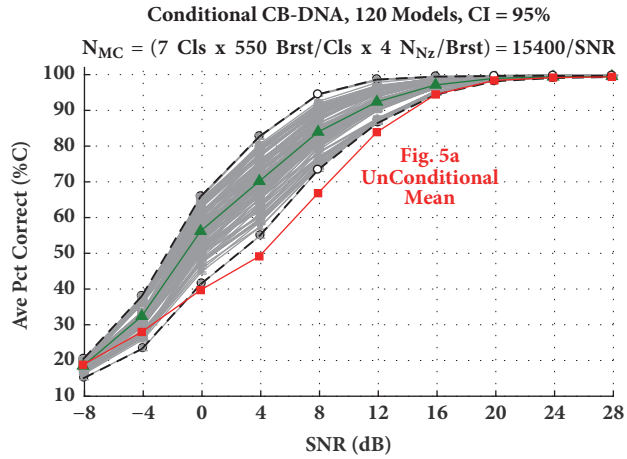
For conditional CB-DNA Fingerprinting at SNR = 12 dB in Figure 6(b), the extreme results include (1) lowest %C ≈ 86.78% performance for Model #1 (excludes ZC1, ZC2, and ZC3 devices) and (2) highest %C ≈ 98.75% performance for Model #90 (excludes ZC4, ZC5, and ZC10 devices). The classification confusion matrices for these extreme cases are provided in Tables 5 and 6 and suggest that the inclusion of ZC4, ZC5, ZC6, and ZC10 devices in Model #1 is most detrimental (italic entries in Table 5). Of note from Table 2 is that package markings for the ZC2, ZC3 pair differs from all other package markings. Thus, Model #1 versus Model #90 performance is consistent with historical DNA discrimination given that the ZC2, ZC3 pair is (1) *excluded* in the poorest Table 5 results (model includes all like-model, *similarly marked* devices) and (2) *included* in the highest Table 6 results (model includes a higher number of like-model *dissimilarly marked* devices).

4.2. Authorized Network Device ID Verification. SNR dependent MDA/ML model components ($\mathbf{W}$, $\boldsymbol{\alpha}$, $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$) from Section 4.1 are used to assess authorized network ZC device ID verification at selected verification $\mathrm{SNR_V}$. Results are presented for c*onditional* CB-DNA fingerprints at $\mathrm{SNR_V}$ = 12 dB where average MDA/ML performance in Figure 6(b) achieves the %C ≈ 90% benchmark. For each network, device TNG fingerprints are used to set device dependent $t_{\mathrm{V}}(i)$ for all authorized devices to achieve TVR ≈ 95%. $t_{\mathrm{V}}(i)$ for the worst and best performing MDA/ML models in Figure 6(b) are shown in Figure 7(a) (Model #1) and Figure 7(b) (Model #90). $t_{\mathrm{V}}(i)$ are overlaid with Euclidean distance TNG statistics (Z_{V}) and ID verification identified as either accept (○) or reject (X) decisions. The accept/reject decisions and final performance are based on Z_{V} for $N_{TNG} = 2200$ fingerprints per authorized device with $Z_{\mathrm{V}} < t_{\mathrm{V}}(i)$ (○ markers) representing *correct* ID verification (proper access granted) and $Z_{\mathrm{V}} > t_{\mathrm{V}}(i)$ (X markers) representing *incorrect* ID verification (improper access denial). The resultant TVR for individual ZC devices is shown along the x-axis and yields an overall cross-ZC average TVR ≈ 94.84% for both models.

4.3. Unauthorized Rogue Device Detection. Accounting for all $N_{NC} = 120$ network configurations with each of the $N_{ZR} = 3$ held-out ZR*j* ($j = 1, 2, \ldots, 10$, $j \neq i$) devices serving in an attacking ZR*j*:ZC*i* role a total of 252 times (including multiple



(a) Results for $N_F^{UNC} = 36$ *unconditional* features with conditional mean from Figure 6(b) overlaid for comparison



(b) Results for $N_F^{CND} = 270$ *conditional* features with unconditional mean from Figure 6(a) overlaid for comparison

FIGURE 6: Classification for 120 networks with (a) *unconditional* and (b) *conditional* CB-DNA features. Mean results show that the %C = 90% benchmark is achieved at SNR ≈ 14 dB (unconditional) and SNR ≈ 11 dB (conditional).

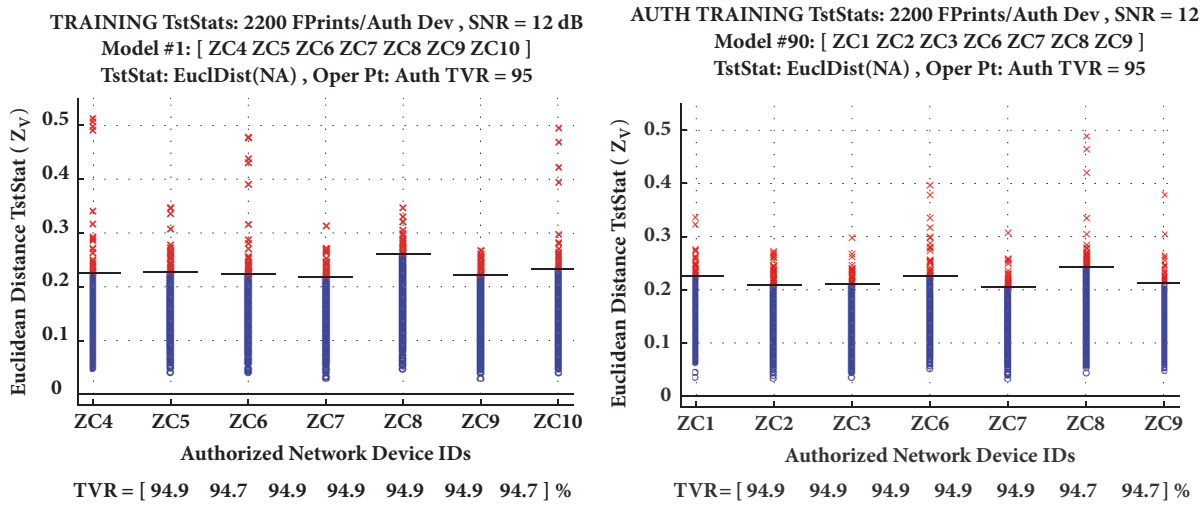
TABLE 5: Confusion matrix for *lowest* performing Model #1 in Figure 6(b) at SNR = 12 dB with %C ≈ 86.78% (sum of diagonals divided by 15,400 trials) and italic to highlight the largest error contributors (ZC4, ZC5, ZC6, and ZC10).

		CALLED CLASS						
		ZC4	ZC5	ZC6	ZC7	ZC8	ZC9	ZC10
INPUT CLASS	ZC4	**1868**	*152*	*64*	16	0	0	*100*
	ZC5	*152*	**1700**	*236*	4	0	0	*108*
	ZC6	*128*	*232*	**1608**	12	0	0	*220*
	ZC7	4	0	0	**2188**	8	0	0
	ZC8	0	0	0	24	**2172**	4	0
	ZC9	0	0	0	8	4	**2120**	68
	ZC10	*100*	*108*	*264*	8	0	12	**1708**

TABLE 6: Confusion matrix for *highest* performing Model #90 in Figure 6(b) at SNR = 12 dB with %C ≈ 98.75% (sum of diagonals divided by 15,400 trials) and italic to highlight the largest error contributors (ZC6 and ZC7).

		CALLED CLASS						
		ZC1	ZC2	ZC3	ZC6	ZC7	ZC8	ZC9
INPUT CLASS	ZC1	**2184**	4	0	*12*	0	0	0
	ZC2	0	**2132**	0	*16*	*40*	4	8
	ZC3	0	0	**2200**	0	0	0	0
	ZC6	*12*	*24*	0	**2164**	0	0	0
	ZC7	0	*36*	0	0	**2156**	8	0
	ZC8	0	0	0	0	4	**2192**	4
	ZC9	0	4	0	8	8	0	**2180**

attacks against a given ZC*i* device present in multiple networks), the cumulative per ZR*j* RRR performance averaged across all networks for $8 \leq \text{SNR}_V \leq 20$ dB is shown in Table 7. Of note here is the average cross-ZR*j* RRR ≈ 89.42% at $\text{SNR}_V = 12$ dB which is approximately the same SNR where MDA/ML device classification in Figure 6(b) achieves the %C = 90% benchmark. As shown in Table 7 $\text{SNR}_V = 12$ dB results, the lowest RRR occurs for ZR4 and ZR6 devices and the highest RRR occurs for ZR1 and ZR3 devices. Excluding $\text{SNR}_V = 8$ dB performance, collective rogue device results for $\text{SNR}_V \geq 12$ dB include (1) cumulative cross-ZR RRR ≈ 94% across all ZR:ZC attack scenarios and (2) individual cross-ZR performance across 252 attacks spanning 83.32% < RRR < 99.81%.



(a) Model #1 Authorized Network Devices

(b) Model #90 Authorized Network Devices

FIGURE 7: Authorized device ID verification for *Conditional CB-DNA Fingerprinting* at SNR_V = 12 dB for (a) worst case Model #1 and (b) best case Model #90 MDA/ML classification in Figure 6(b). Device dependent training thresholds t_V (horizontal lines) set for TVR = 95% with resultant per device TVR shown along the x-axis.

TABLE 7: *Conditional* rogue ID verification performance showing cumulative average RRR (%) for indicated ZR*j*. The highest and lowest RRR per SNR_V and row/column averages are denoted by bold text and italic, respectively.

SNR_V (dB)	ZR Rogue ID										Cross-ZR Ave
	ZR1	ZR2	ZR3	ZR4	ZR5	ZR6	ZR7	ZR8	ZR9	ZR10	
8	**95.55**	*65.76*	**97.90**	*67.74*	72.40	69.69	85.82	94.86	83.60	71.38	*80.47*
12	**99.24**	83.17	**99.60**	*79.17*	82.16	*76.29*	97.78	98.80	95.78	82.24	*89.42*
16	99.64	92.78	**99.84**	88.72	*89.49*	*82.77*	**99.83**	99.80	99.37	92.96	**94.45**
20	99.66	98.64	**99.99**	95.93	*95.65*	*90.90*	**99.99**	**99.99**	99.95	98.34	**97.90**
Cross- SNR_V Ave	99.51	91.53	**99.81**	*87.94*	89.10	*83.32*	99.20	**99.53**	98.37	91.18	93.95

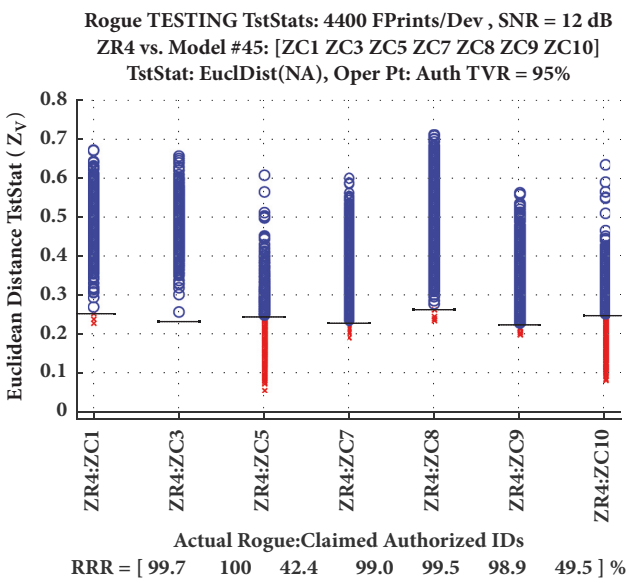
For the overall poorest ZR4 and ZR6 results in Table 7 at SNR_V = 12 dB there are eight network models (#17, #45, #66, #86, #91, #92, #93, and #94) that include both ZR4 and ZR6 serving as rogue devices. Considering only these models, the cumulative ZR4 and ZR6 results include RRR ≈ 85.25% and RRR ≈ 82.03%, respectively. The overall poorest ZR4 and ZR6 RRR results for these eight models at SNR_V = 12 dB are presented in Figure 8 and occur for Model #45 with ZC1, ZC3, ZC5, ZC7, ZC8, ZC9, and ZC10 authorized devices. As estimated by averaging individual ZR*j*:ZC*i* RRR presented along Figure 8 x-axes, the average performance for ZR4:ZC*i* is RRR ≈ 84.14% and for ZR6:ZC*i* is RRR ≈ 77.56%. These are higher than the cumulative 120 model averages in Table 7 and thus do not represent the overall poorest ZR4 and ZR6 device results.

For completeness, the overall poorest ZR4 and ZR6 RRR results across all 120 models are presented in Figure 9 which shows that the lowest RRR results are obtained for separate models and include average RRR ≈ 73.27% in Figure 9(a) for ZR4 with Model #19 and average RRR ≈ 64.84% in Figure 9(b) for ZR6 with Model #4. While it is not immediately obvious why these are the two poorest cases, these ID verification results are consistent with the increased MDA/ML classification challenge noted in Section 4.1 for models based on *similarly marked* authorized devices. Specifically, the poorest RRR < 80% results in Figure 9 are all attributable to ZC*j*:ZC*i* combinations of *similarly marked* ZC4, ZC5, ZC6, and ZC10 devices.

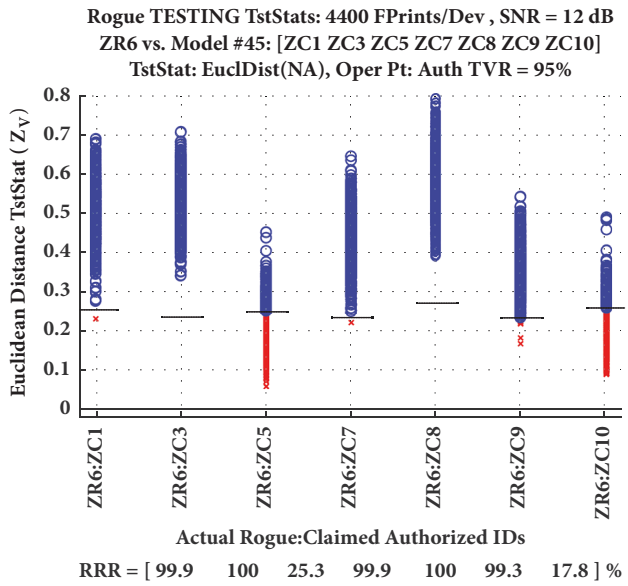
For the overall best RRR ZR1 and ZR3 results in Table 7 at SNR_V = 12 dB there are eight network models (#1, #9, #10, #11, #12, #13, #14, and #15) that include both ZR1 and ZR3 serving as rogue devices. The overall best rogue ZR1 and ZR3 detection results for these models at SNR_V = 12 dB are presented in Figure 10 and include assessments for Model #11 with ZC2, ZC4, ZC5, ZC7, ZC8, ZC9, and ZC10 authorized devices. As estimated by averaging the individual ZR*j*:ZC*i* RRR indicated along Figures 10(a) and 10(b) x-axes, the average RRR performance across best case ZR1:ZC*i* is RRR ≈ 99.31% and across all ZR3:ZC*i* is RRR ≈ 99.98%; this best case cross-ZR*j* RRR was observed for a majority of models and ZR*j*:ZC*i* considered.

5. Conclusion

An analytic development of CB-DNA Fingerprinting for conventional QAM features is presented as well as its application to verification-based rogue detection demonstrated using ZigBee RZSUBSTICK communication devices. Results are

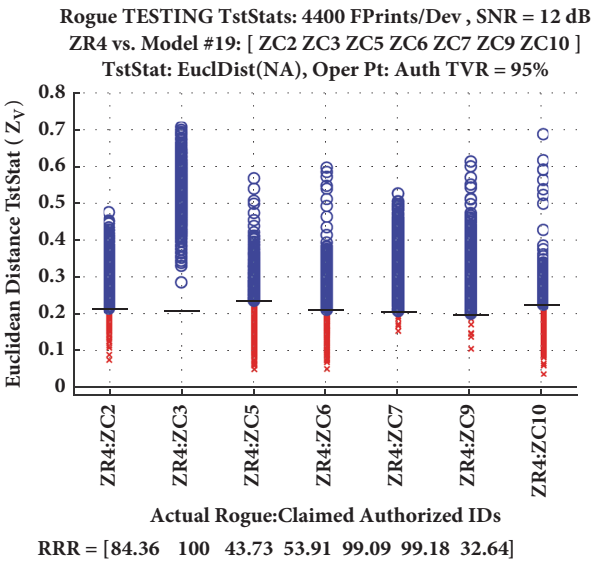


(a) ZR4 versus Model #45: Average RRR ≈ 84.14%

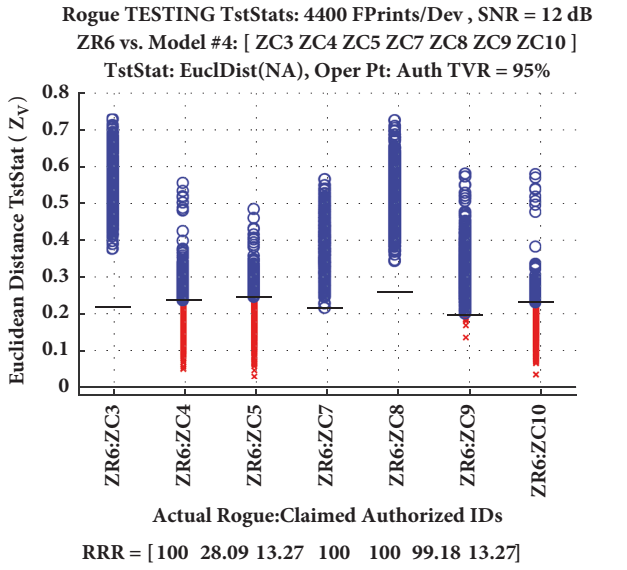


(b) ZR6 versus Model #45: Average RRR ≈ 77.46%

Figure 8: Rogue ID verification for (a) ZR4 and (b) ZR6 devices attacking Model #45 with ZC1, ZC3, ZC5, ZC7, ZC8, ZC9, and ZC10 network devices and contributing to poorest (minimum) RRR shown in Table 7 at SNR_V = 12 dB.



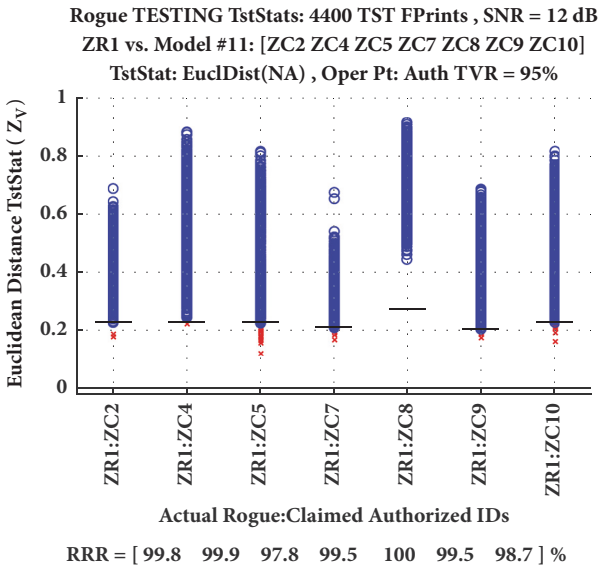
(a) ZR4 versus Model #19: Average RRR ≈ 73.27%



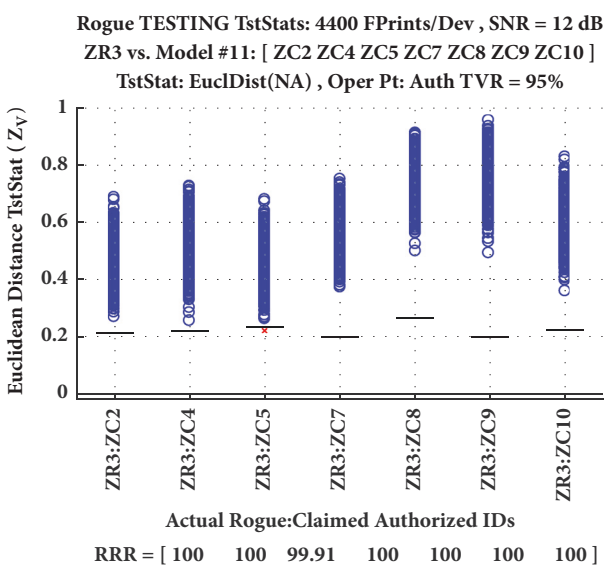
(b) ZR6 versus Model #4: Average RRR ≈ 64.84%

Figure 9: Overall poorest Rogue ID verification performance across 120 models for (a) ZR4 and (b) ZR6 with indicated network devices and contributing to poorest (minimum) RRR shown in Table 7 at SNR_V = 12 dB.

based on experimentally collected signals with postcollection fingerprint generation and authorized versus rogue device ID verification performed for 120 unique networks consisting of seven authorized and three unauthorized attacking rogue devices. Collective authorized device discrimination results for all 120 network configurations using an MDA classifier included (1) average cross-class percent correct classification of %C > 90% achieved for SNR ≥ 12 dB and (2) identification of device dependent verification thresholds yielding True Verification Rates (true positive) of TVR = 95% for all authorized network devices. The MDA network models were used for rogue device ID verification and Rogue Rejection Rate (RRR) (true negative) estimated for all rogues presented to the networks. Collective rogue device detection results for SNR ≥ 12 dB included (1) cumulative average burst-by-burst RRR ≈ 94% across 2520 total rogue attack scenarios and (2)



(a) ZR1 versus Model #11: Average RRR ≈ 99.31%



(b) ZR3 versus Model #11: Average RRR ≈ 99.98%

FIGURE 10: Rogue ID verification for (a) ZR1 and (b) ZR3 devices attacking Model #11 with ZC2, ZC4, ZC5, ZC7, ZC8, ZC9, and ZC10 network devices and contributing to best (maximum) RRR shown in Table 7 at SNR_V = 12 dB.

performance across 252 attacks per individual devices spanning 83.32% < RRR < 99.81%. As a first successful proof-of-concept demonstration using CB-DNA Fingerprinting with conventional communication constellation features, these results are promising and further research is warranted.

Data Availability

The data used to support the findings is generally unavailable due to public releasability constraints. However, please contact the corresponding author for special release consideration.

Disclosure

The views expressed 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-2018-2040.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] N. Goldenberg and A. Wool, "Accurate modeling of Modbus/TCP for intrusion detection in SCADA systems," *International Journal of Critical Infrastructure Protection*, vol. 6, no. 2, pp. 63–75, 2013.

[2] Z. Zheng and A. L. Reddy, "Safeguarding Building Automation Networks: THE-Driven Anomaly Detector Based on Traffic Analysis," in *Proceedings of the 26th International Conference on Computer Communication and Networks (ICCCN)*, pp. 1–11, Vancouver, BC, Canada, July 2017.

[3] J. Jiang and L. Yasakethu, "Anomaly Detection via One Class SVM for Protection of SCADA Systems," in *Proceedings of the International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC '13)*, pp. 82–88, Beijing, China, October 2013.

[4] 802.15.4 IoT Markets: A Market Dynamics Report," Research and Markets, Market Report, ID: 4392927, Jul 2017.

[5] Homeland Security, "Recommended Practice: Improving Industrial Control Systems Cybersecurity with Defense-In-Depth Strategies," *ICS-CERT*, pp. 1–56, Sep 2016, https://ics-cert.us-cert.gov.

[6] B. Erinle, *Cyber Security for National Defense*, 2010.

[7] A. Mehta, *Could an Air Conditioner Take Down a Military Base? The Pentagon is Worried*, 2017.

[8] T. Lennvall, S. Svensson, and F. Hekland, "A comparison of wirelessHART and ZigBee for industrial applications," in *Proceedings of the 7th IEEE International Workshop on Factory Communication Systems*, pp. 85–88, May 2008.

[9] Field Communications Group, Connecting the World of Process Automation, 2017.

[10] K. Stefanidis and A. G. Voyiatzis, "An HMM-Based Anomaly Detection Approach for SCADA Systems," in *Information Security Theory and Practice*, vol. 9895 of *Lecture Notes in Computer Science*, pp. 85–99, Springer International Publishing, Cham, 2016.

[11] C. M. Talbot, M. A. Temple, T. J. Carbino, and J. A. Betances, "Detecting rogue attacks on commercial wireless Insteon home automation systems," *Computers & Security*, vol. 74, pp. 296–307, 2018.

[12] C. Dubendorfer, B. Ramsey, and M. Temple, "ZigBee device verification for securing industrial control and building automation systems," *IFIP Advances in Information and Communication Technology*, vol. 417, pp. 47–62, 2013.

[13] T. J. Carbino, M. A. Temple, and T. J. Bihl, "Ethernet card discrimination using unintentional cable emissions and constellation-based fingerprinting," in *Proceedings of the International Conference on Computing, Networking and Communications (ICNC '15)*, pp. 369–373, Garden Grove, CA, USA, Feburary 2015.

[14] D. R. Reising, M. A. Temple, and M. J. Mendenhall, "Improving Intra-Cellular Security Using Air Monitoring with RF Fingerprints," in *Proceedings of the Networking Conference (WCNC)*, pp. 1–6, Sydney, Australia, April 2010.

[15] M. D. Williams, M. A. Temple, and D. R. Reising, "Augmenting Bit-Level Network Security Using Physical Layer RF-DNA Fingerprinting," in *Proceedings of the IEEE Global Communications Conference (GLOBECOM '10)*, pp. 1–6, Miami, FL, USA, December 2010.

[16] M. D. Williams, S. A. Munns, M. A. Temple, and M. J. Mendenhall, "RF-DNA fingerprinting for airport WiMax communications security," in *Proceedings of the 4th International Conference on Network and System Security (NSS '10)*, pp. 32–39, Melbourne, Australia, September 2010.

[17] M. A. Buckner et al., "Enhancing Network Security Using 'Learning-from-Signals' and Fractional Fourier Transform Based RF Fingerprints," in *Proceedings of Wireless Innovation Forum Conference on Communications Technologies and Software Defined Radio (SDR 11-WInnComm)*, pp. 317–325, 2011.

[18] U. Gupta, "Application of Multi Factor Authentication in Internet of Things Domain," *International Journal of Computer Applications*, vol. 123, no. 1, pp. 29–31, 2015.

[19] PCI Security Standards Council, Information Supplement: Multi-Factor Authentication, 2017.

[20] IEEE Computer Society, IEEE Std 802.15.4, Sep 2011.

[21] F. zhuo, Y. Huang, and J. chen, "Radio Frequency Fingerprint Extraction of Radio Emitter Based on I/Q Imbalance," *Procedia Computer Science*, vol. 107, pp. 472–477, 2017.

[22] N. S. Alagha, "Cramer-Rao bounds of SNR estimates for BPSK and QPSK modulated signals," *IEEE Communications Letters*, vol. 5, no. 1, pp. 10–12, 2001.

[23] F. G. Stremler, *Introduction to Communication Systems*, Addison-Wesley Publishing Company, Reading, MA, 3rd edition, 1990.

[24] C. R. Johnson Jr, W. A. Sethares, and A. G. Klein, *Software Receiver Design*, Cambridge University Press, Cambridge, 2011.

[25] G. Shi and K. Li, "Fundamentals of ZigBee and WiFi," in *Signal Interference in WiFi and ZigBee Networks*, Wireless Networks, pp. 9–27, Springer International Publishing, Cham, 2017.

[26] Atmel Corporation, *R2016: RZRAVEN Hardware User's Guide, Rev. 8117D-AVR-04/08*, 2008.

[27] Atmel Corporation, AVR Low Power 2.4 GHz Transceiver for ZigBee, IEEE 802.15.4, 6LoWPAN, RF4CE and ISM Applications, AT86RF230 Spec Sheet, 5131E-MCU Wireless-02/09, 2009.

[28] D. R. Reising, M. A. Temple, and M. J. Mendenhall, "Improved wireless security for gmsk-based devices using rf fingerprinting," *International Journal of Electronic Security and Digital Forensics*, vol. 3, no. 1, pp. 41–59, 2010.

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

[30] S. U. Rehman, K. W. Sowerby, and C. Coghill, "Radio-frequency fingerprinting for mitigating primary user emulation attack in low-end cognitive radios," *IET Communications*, vol. 8, no. 8, pp. 1274–1284, 2014.

[31] Y. Huang and H. Zheng, "Radio frequency fingerprinting based on the constellation errors," in *Proceedings of the 18th Asia-Pacific Conference on Communications: "Green and Smart Communications for IT Innovation", APCC 2012*, pp. 900–905, Republic of Korea, October 2012.

[32] B. Miller Michael, *Mathematics and Statistics for Financial Risk Management*, 2nd., Ed., John Wiley & Sons, Inc, Hoboken, New Jersey, 2014.