Abstract—Improved network security is addressed using device dependent physical-layer (PHY) based fingerprints from Ethernet cards to augment traditional MAC-based ID verification. The investigation uses unintentional Ethernet cable emissions and device fingerprints comprised of Constellation-Based, Distinct Native Attribute (CB-DNA) features. Near-field collection probe derivative effects dictated the need for developing a two-dimensional (2D) binary constellation for demodulation and CB-DNA extraction. Results show that the 2D constellation provides reliable demodulation (bit estimation) and device discrimination using symbol cluster statistics for CB-DNA. Bit Error Rate (BER) and Cross-Manufacturer Discrimination (CMD) results are provided for 16 devices from 4 different manufacturers. Device discrimination is assessed using both Nearest Neighbor (NN) and Multiple Discriminant Analysis, Maximum Likelihood (MDA/ML) classifiers. Overall results are promising and include CMD average classification accuracy of $C = 91.38\%$ (MDA/ML).

I. INTRODUCTION

In environments where devices connect and disconnect from a network at will, there are unique security challenges and increased difficulty in mitigating unauthorized access. This is further compounded as the sophistication of network Medium Access Control (MAC) based attacks increase and alternate security measures are needed. Physical-layer (PHY) augmentation of MAC-based authentication processes provides one means to improve security and reliably authenticate network devices. PHY-based security exploits unique PHY features that can be extracted from device emissions (intentional and unintentional). The PHY-based features and/or statistics thereof are used to form unique fingerprints that enable reliable device discrimination. The uniqueness is a function of tolerances and limits inherent in device manufacturing processes [1] which impart sufficient variability even when supposedly identical parts are used [2], [3].

Multiple PHY-based fingerprinting approaches have been considered over the past couple of decades to improve device identification, with the main emphasis on transient-based and constellation-based features. Additional approaches “target specific wireless technologies and/or exploit additional properties from the signal and logical layer” [2]. These approaches extract fingerprint features from a predefined Region of Interest (ROI) that may include the transient, invariant or entire burst responses. Restrictions are placed on the applicability of transient-based methods given that the signal transient response is influenced by the device hardware and the wireless communication channel [2].

The work here focuses on Constellation-Based demodulation and Distinct Native Attribute (CB-DNA) fingerprinting using experimentally collected Ethernet signals. The derivative effect of near-field probe collections dictated the need for developing a unique 2D received signal constellation. Validating constellation demodulation capability was essential for establishing validity of the CB-DNA fingerprinting approach that to the authors’ knowledge has not been previously considered for unintentional Ethernet emissions. The approach here differs from more common approaches that exploit dependent modulation error in the constellation space, i.e., differences (error) between received projected symbols and the ideal transmitted signaling constellation [1], [2], [4], [5]. Regardless of the method employed, constellation responses are influenced by variation in hardware components (e.g., amplifiers, capacitors, inductors, and oscillators) [1], [4], [6], [7]. Component variation is present in wireless network devices and can cause deviation in symbol rate, frequency, and AM/FM/PM conversion [1].

The remainder of the document is organized as follows. Sect. II provides background information on the experimental setup and PHY-based CB-DNA device fingerprinting. This is followed by device discrimination results in Sect. III and a summary and conclusions in Sect. IV.

II. BACKGROUND

Collecting intentional or unintentional Radio Frequency (RF) device emissions is essential to feature-based device fingerprint creation. The majority of recent fingerprinting work investigates intentional RF emissions from wireless network cards [1], [2], [4]–[9]. These works perform device discrimination through feature extraction using two fundamentally different approaches, including: 1) Constellation-Based (CB) features [1], [2], [4]–[7], [10] and 2) RF Distinct Native Attribute (RF-DNA) features [8], [9], [11]–[13]. Fingerprinting wired network cards is covered by [14], [15]. The approach in [14], [15] differs from CB-DNA and RF-DNA approaches in two ways 1) the need for direct contact with network card...
pins to collect intentional voltage responses and 2) device discrimination is accomplished using a matched filter.

The need for a constellation signaling representation inherently limits CB approaches to intentional emission applications which leaves device fingerprints developed from unintentional emission features largely unexplored. The use of unintentional emissions here is not entirely new given that expansion of the less constrained RF-DNA approach has included the use of unintentional emissions to discriminate between Integrated Circuit (IC) components and IC operational state [3], [11], [16], [17].

The collection technique in [18] was adopted here to collect unintentional RF emissions from Ethernet cables and create a 2D signal constellation representation. This provides the ability to remotely fingerprint and discriminate wired network devices without requiring direct access to the hardware as required in [14], [15]. Additionally, the presented work adopts the statistical approach of RF-DNA [3], [8], [9], [11]–[13], [16], [17] to capture statistical variation of symbol clusters in the constellation space to form CB-DNA fingerprints.

A. Experimental Emission Collection

The emission collection setup includes a laptop and a Dell Precision T7500 desktop computer that collects trace information from the LeCroy WavePro 760Zi-A 6.0 GHz oscilloscope operating at a sample frequency of \( f_s = 250 \) M Samples/Sec (MSPS). An in-line baseband filter with bandwidth of \( W_{BB} = 32 \text{MHz} \) and a Riscure 205HS “High Sensitivity” near-field probe are connected to the oscilloscope to capture the RF emissions. The T7500 desktop computer also hosts the network card being fingerprinted. The network connection between host computer and laptop is configured for 10BASE-T full duplex Ethernet signaling via a Category 6 Ethernet cable.

A standard Ethernet cable has four Twisted Wire Pairs (TWP) and the low-level power of unintentional RF emissions makes RF probe placement critical to ensure reliability and repeatability. Emissions from the multiple TWP will contribute to collected emission responses given that the network card under evaluation can transmit and receive simultaneously over multiple TWP. To minimize interference, the connected laptop only responded to requests that were necessary to maintain the 10BASE-T connection which results in minimal traffic being sent on the received wire.

Of greatest importance to the collected probe responses is the physical orientation between the probe and wire of interest. Fig. 1 illustrates three possible probe locations (A, B, C) to highlight the orientation between the probe and wire of interest as the probe moves axially along the wire. When the probe is placed at location A or C, the collected signal response is most affected by the Electromagnetic (EM) field generated by the solid color and multicolored wires, respectively. However, when the probe is at location B the collected signal response is influenced equally by EM signals generated by both the solid and multicolored wires. The use of a jig during testing mitigated the collected signal variation due to RF probe placement while swapping out Ethernet cards.

As shown in Table I, a total of 16 network cards were tested, including 4 cards from 4 manufacturers identified as: D-Link (DL), TRENDnET (TN), Intel (IN), and StarTech (ST). The specific cards are identified using the last four alpha numeric digits of their MAC address. The four unique LAN transformer markings are also provided and used for analyzing results. The LAN transformer is the last card component that the signal goes through prior to reaching the RJ45 output jack.

![Fig. 1. Relationship of Near-Field RF Probe Location and Wire of Interest Within the TWP.](image)

B. 2D Binary Constellation Development

This section provides technical details for developing a two-dimensional (2D) signaling constellation. This development was required given that the near-field probe response represents the time derivative of signals passing through the Ethernet cable. The resultant 2D binary constellation is used for demodulation (bit estimation) and CB-DNA device fingerprinting using symbol cluster statistics. The symbol estimation boundary is presented and its use for emission demodulation explained. The process used for locating and synchronizing to individual burst responses in collected probe traces was adopted directly from [18] without modification.

### TABLE I

<table>
<thead>
<tr>
<th>Manufacturer</th>
<th>Reference</th>
<th>MAC Address</th>
<th>LAN Transformer Markings</th>
</tr>
</thead>
<tbody>
<tr>
<td>D-Link</td>
<td>DL1</td>
<td>D966</td>
<td>Bi-Tek IM-1178LLF 12471</td>
</tr>
<tr>
<td></td>
<td>DL2</td>
<td>DA06</td>
<td>Bi-Tek IM-1178LLF 12471</td>
</tr>
<tr>
<td></td>
<td>DL3</td>
<td>DA07</td>
<td>Bi-Tek IM-1178LLF 12471</td>
</tr>
<tr>
<td></td>
<td>DL4</td>
<td>60E0</td>
<td>Bi-Tek IM-1178LLF 12471</td>
</tr>
<tr>
<td>TRENDnET</td>
<td>TN1</td>
<td>9B55</td>
<td>Bi-Tek IM-1178LLF 12471</td>
</tr>
<tr>
<td></td>
<td>TN2</td>
<td>9334</td>
<td>Bi-Tek IM-1178LLF 12471</td>
</tr>
<tr>
<td></td>
<td>TN3</td>
<td>9B54</td>
<td>Bi-Tek IM-1178LLF 12471</td>
</tr>
<tr>
<td></td>
<td>TN4</td>
<td>9B56</td>
<td>Bi-Tek IM-1178LLF 12471</td>
</tr>
<tr>
<td>Intel</td>
<td>IN1</td>
<td>1586</td>
<td>BI HS00-06037LF 12471</td>
</tr>
<tr>
<td></td>
<td>IN2</td>
<td>1A93</td>
<td>BI HS00-06037LF 12471</td>
</tr>
<tr>
<td></td>
<td>IN3</td>
<td>1A59</td>
<td>BI HS00-06037LF 12471</td>
</tr>
<tr>
<td></td>
<td>IN4</td>
<td>1A9E</td>
<td>BI HS00-06037LF 12471</td>
</tr>
<tr>
<td>StarTech</td>
<td>ST1</td>
<td>32CB</td>
<td>FPE G24102MK 1250a1</td>
</tr>
<tr>
<td></td>
<td>ST2</td>
<td>32B4</td>
<td>FPE G24102MK 1250a1</td>
</tr>
<tr>
<td></td>
<td>ST3</td>
<td>96F4</td>
<td>FPE G24102MK 1320G1</td>
</tr>
<tr>
<td></td>
<td>ST4</td>
<td>3048</td>
<td>FPE G24102MK 1250a1</td>
</tr>
</tbody>
</table>
1) Symbol Response Mapping: Generation of the 2D constellation for 10BASE-T binary signal reception is described with the aid of Fig. 2. First, consider a sequence of symbol samples \( \{ s(k) \} \) for \( 1 \leq k \leq N_{TS} \) where \( N_{TS} \) is the total number of samples spanning symbol interval \( T_S \). Given sequence mid-point \( s(k_m) \) at index \( k_m \), gradient-based test statistics are generated using two sub-sequences, \( \{ T_G^{-}(k) \} \) and \( \{ T_G^{+}(k) \} \), on either side of \( s(k_m) \) according to (1) through (4), where each sub-sequence contains \( N_{\Delta} + 1 \) samples. The Gradient operation in (2) and (4) is a numerical gradient used to represent the slope between adjacent sequence samples \( (N_{\Delta} \text{ total slope values}) \). The resultant \( Z_G^{-} \) from (2) and \( Z_G^{+} \) from (4) are used to form the 2D \( Z_G^{-} \cdot Z_G^{+} \) constellation. This is illustrated in Fig. 3 which shows a representative received symbol constellation for each device manufacturer.

\[
\begin{align*}
T_G^{-}(k) &= s(k) \text{ for } (k_m - N_{\Delta} \leq k \leq k_m) \quad (1) \\
Z_G^{-} &= \text{Mean}[\text{Gradient}\{T_G^{-}(k)\}] \quad (2) \\
T_G^{+}(k) &= s(k) \text{ for } (k_m \leq k \leq k_m + N_{\Delta}) \quad (3) \\
Z_G^{+} &= \text{Mean}[\text{Gradient}\{T_G^{+}(k)\}] \quad (4)
\end{align*}
\]

C. Device Constellations

A constellation diagram provides a pictorial representation of communication symbols contained within a digitally modulated signal. This is illustrated in Fig. 3 which provides a collected signal constellation for each manufacturer in Table I based on approximately 380,000 symbols. Each constellation includes a composite cluster representing a Binary 0 (Blue) and Binary 1 (Red). It is visually evident that the composite cluster points are non-Gaussian distributed and actually comprised of multiple sub-clusters.

The presence of multiple sub-clusters is evident in Fig. 4 which shows that each unconditional composite cluster is actually comprised of multiple conditional sub-clusters. The constellation points have been color coded according to bit values preceding and succeeding the bit being estimated, i.e.,

For binary demodulation, symbols mapped into the lower left hand quadrant are estimated as a Binary 0 while symbols mapped to into the upper right hand quadrant are estimated as a Binary 1.

D. Constellation-Based Fingerprinting

CB-DNA fingerprint generation begins by dividing constellation points into their respective unconditional composite and conditional sub-cluster groupings (a total of \( N_{CR} = 2 + 8 = 10 \) cluster regions). Statistical CB-DNA features are then calculated as the mean \( (\mu) \), variance \( (\sigma^2) \), skewness \( (\gamma) \), and kurtosis \( (\kappa) \) along the \( Z_G^{-} \) and \( Z_G^{+} \) dimensions for each

Fig. 2. Near-Field Probe Response for a Binary 1 Symbol with Gradient Calculation Sub-Sequences \( \{ T_G^{-}(k) \} \) and \( \{ T_G^{+}(k) \} \) Highlighted. For \( f_s = 250 \text{ MSFPS} \) Each Symbol Includes \( N_{TS} = 25 \text{ Total Samples} \) and Midpoint Index \( k_m = 13 \).

Fig. 3. Device Constellations Based on Approximately 380,000 Symbols for Each Card Manufacturer. The Binary 0 (Blue) and Binary 1 (Red) Composite Clusters are Comprised of Multiple Sub-Clusters Corresponding to Bit Combinations Preceding and Succeeding the Bit Being Estimated.

Fig. 4. 2D Binary Constellation for D-Link Card Showing Non-Gaussian Multimodal Symbol Sub-Clusters and Linear Bit Estimation Boundary (ZC).
cluster region. The resultant statistics form a Regional Cluster Fingerprint \( F_{RB}^{CB} \) given by (5), where the superscripted \(-/+\) sign denotes constellation dimension and \( i = 1, 2, \ldots, N_{CR} \). The final Composite CB-DNA Fingerprint \( F_C^{CB} \) is of dimension \( 1 \times (8 \times N_{CR}) \) and constructed by concatenating \( F_{RB}^{CB} \) from (5) as shown in (6).

\[
F_{RB}^{CB} = [\mu_R^+, \mu_R^-, \sigma_R^+, \sigma_R^-, \gamma_R^+, \gamma_R^-, \kappa_R^+, \kappa_R^-]_{1 \times 8} \tag{5}
\]

\[
F_C^{CB} = [F_{RB}^{CB} : F_{RB}^{CB} : \cdots : F_{RB}^{CB}]_{1 \times (8 \times N_{CR})} \tag{6}
\]

Device discrimination was assessed using a simple Nearest Neighbor (NN) and Multiple Discriminant Analysis, Maximum Likelihood (MDA/ML) classifier. Performance was assessed using \( N_{Tng} \) Training and \( N_{Tst} \) Testing fingerprints, based on only the two unconditional composite clusters \( N_{CR} = 2 \), and respectively denoted by \( F_{Tng}^{CB} \) and \( F_{Tst}^{CB} \).

For MDA/ML classifier implementation [3], [8], [9], each of the \( N_C = 4 \) classes was represented by \( N_{Tng} = N_{Tst} = 2000 \) fingerprints (500 fingerprints per card for a given manufacturer with each fingerprint based on approximately 1,400 symbols). Training was performed using \( K \)-fold cross-validation \( K = 5 \). This involves [19]: 1) dividing the \( N_{Tng} \) set of \( F_{Tng}^{CB} \) into \( K \) equal size disjoint blocks of \( N_{Tng}/5 \) fingerprints, 2) holding out one block and training on \( K-1 \) blocks to produce projection matrix \( W \), and 3) using the holdout block and resultant \( W \) for validation. Final Testing is then performed using \( W \) from the best training iteration and the \( N_{Tst} \) set of \( F_{Tst}^{CB} \).

The NN classifier was implemented using reduced dimension \( F_C^{CB} \) from (5) that only consisted of mean (\( \mu \)) and standard deviation (\( \sigma \)) features; \( F_C^{CB} \) was constructed as in (6). For NN classifier implementation, each of the \( N_C = 4 \) classes was represented by \( N_{Tng} = 48 \) fingerprints (3 fingerprints per card for a given manufacturer with each fingerprint based on approximately 32,000 symbols) and \( N_{Tst} = 2,000 \) fingerprints were used for assessment. Euclidean distance between \( F_{Tng}^{CB} \) and \( F_{Tst}^{CB} \) was used as the measure of similarity to produce test statistic \( z_{ED} \). For a given \( F_{Tst}^{CB} \) under evaluation, the Euclidean distance is calculated between \( F_{Tst}^{CB} \) and \( F_{Tng,j}^{CB} \) according to (7), where \( j = 1, 2, 3, 4 \) denotes the Class ID number assigned to a given card manufacturer. Next, the index number associated with the minimum value in the test statistic vector \( z_{ED} \) dictates the predicted card \( PC_j \) assignment as described in (8).

\[
z_{ED_{4 \times 1}} = \text{sum} \left( \sqrt{\left( F_{Tst}^{CB} - F_{Tng,j}^{CB} \right)^2} \right) \tag{7}
\]

\[
PC_j = \min(z_{ED_{4 \times 1}}) \tag{8}
\]

III. RESULTS

This section provides demodulation and device discrimination results for the 16 cards considered. Symbol demodulation (bit estimation) assessment is based on approximately 1.7 Billion collected symbols per manufacturer (pooled symbols from 4 cards). The Bit Error Rate (BER) results are presented in Table II where the Single Slope (SSLP) results were generated using the technique introduced in [18] and CB results are based on the technique introduced here in Sect. II-B. The overall BER for each method is approximately the same, which provides valid evidence that the 2D binary constellation development is appropriate for generating CB-DNA fingerprints.

Device discrimination assessment is based on \( N_{Tst} = 2,000 \) independent testing fingerprints using the NN and MDA/ML classifiers. Each testing fingerprint was classified using the \( W \) matrix for MDA/ML and the NN process given by (7) and (8). The fingerprint under test is classified (rightly or wrongly) as being associated with the training class that it most closely resembles. The correct and incorrect classifications are tracked and used to form a classification confusion matrix.

The classification confusion matrices are structured similar to [14] with matrix rows representing true card IDs and columns representing estimated card IDs. The average percentage of correct classification (\%C) is derived from matrix diagonal entries and incorrect classifications are represented in off-diagonal entries. Confusion matrix results for Cross-Model Discrimination (CMD) classification using the 16 devices under consideration are presented in Table III for NN and CB-DNA. The table entries are presented as \% C NN / \% C MDA/ML with bold entries denoting best or statistically equivalent performance based on 95% confidence intervals. Relative to NN versus MDA/ML performance in Table III, results are as expected with the more robust MDA/ML method yielding overall cross-manufacturer \% C = 91.38% and NN yielding only \% C = 76.73%. Of particular note is that DL and TN devices account for a majority of the misclassification error. The higher degree of DL and TN confusion is attributed to the devices using identical LAN transformers as indicated in Table I.

IV. SUMMARY & CONCLUSIONS

Constellation-based fingerprinting has been predominantly applied to wireless device discrimination using intentional radiated emissions and discriminating features derived from signal constellation errors, i.e., differences between ideal transmitted and received constellation responses. Work here expands constellation-based methods to include unintentional
RF emissions from Ethernet cables with discriminating features extracted from a newly developed 2D binary signaling constellation. As verified, the 2D binary constellation enables reliable demodulation and Constellation-Based, Distinct Native Attribute (CB-DNA) fingerprinting based on symbol cluster statistics.

Cross-Model Discrimination (CMD) was assessed using NN and MDA/ML classifiers with input fingerprints generated from 16 different Ethernet devices (4 like-model cards from D-Link, TRENDnET, Intel, and StarTech manufacturers). Relative to NN versus MDA/ML implementation, overall CMD performance was as expected with the more robust MDA/ML classifier yielding CMD average percent correct classification of $\%C = 91.38\%$ and NN only yielding $\%C = 76.73\%$. For both classifiers, a majority of the misclassification error occurred between D-Link and TRENDnET devices.

The lack of consistent performance across all manufactures is attributed to 1) D-Link and TRENDnET cards using identical LAN transformers (Table I) and having similar constellation point distributions (Fig. 3), 2) non-Gaussian multimodal composite constellation point distributions (Fig. 4) producing non-Gaussian CB-DNA features that are sub-optimal for MDA/ML, and 3) failure to fully exploit discriminating information in conditional constellation point sub-clusters (Fig. 4). These issues are being addressed in ongoing research, as well as extension to address Like-Model Discrimination (LMD) which historically presents a greater classification challenge.

### Table III

<table>
<thead>
<tr>
<th>Manufacturer</th>
<th>Predicted Network Card</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DL</td>
</tr>
<tr>
<td>D-Link</td>
<td>54.35 / 85.75</td>
</tr>
<tr>
<td>IN</td>
<td>0.0 / 0.025</td>
</tr>
<tr>
<td>TN</td>
<td>43.00 / 19.94</td>
</tr>
<tr>
<td>ST</td>
<td>1.6 / 0.0</td>
</tr>
</tbody>
</table>

### References


