Protocol Switching Mechanism for Countering Radiometric Signature Exploitation

Wassila Lalouani, Mohamed Younis, Danila Frolov
Department of Computer Science and Electrical Engineering
University of Maryland, Baltimore County
Baltimore, Maryland, USA
lwassil1, younis, dan50@umbc.edu

Uthman Baroudi
Computer Engineering Department
King Fahd Univ. of Petroleum & Minerals
Dhahran, 31261, Saudi Arabia
ubaroudi@kfupm.edu.sa

Abstract*— Manufacturing variations introduce features that distinguish radio transceivers even those of the same vendor. Such distinction is often referred to as radiometric signature and is found to be useful in conducting device authentication and crime forensics. Yet, radiometric signatures could constitute a privacy threat. Particularly, in the realm of wireless networks, an adversary may exploit RF fingerprinting to identify devices and conduct traffic analysis in order to uncover the topology and categorize the role of various nodes. In this paper, we show that RF fingerprinting could be a major tool for the adversary to distinguish among nodes and bypass the provisioned anonymity protection in the network. We analyze the accuracy of RF fingerprinting and highlight how the accuracy affects the success of adversary attacks. We further develop a novel countermeasure to degrade the adversary's ability in exploiting RF fingerprinting. The proposed countermeasure is based on switching among preset communication protocols and employs adversarial machine learning to select the protocol for a transmission so that the accuracy of the RF fingerprinting diminishes. We demonstrate the effectiveness of our scheme through simulation and prototype experiments.

Keywords: ***Radiometric signature, RF Fingerprinting, Traffic analysis, Adversarial machine learning.***

I. Introduction

The continual advancement in wireless technologies has enabled networked solutions for many unconventional civil and military applications. In addition, practitioners have been viewing radio based links as inexpensive and versatile option for interconnecting computational and control devices. Nowadays, wireless links are used in major infrastructure, such as power grids, manufacturing facilities, residential buildings, etc. In addition, ad-hoc networking of smart devices have been attracting increased attention from the research and engineering community, motivated by applications like situational awareness, asset tracking, air-borne safety, digital battlefield, and border protection. Wireless sensor networks, and more recently the notion of Internet of Things (IoT), constitute popular examples of ad-hoc networks with application-centric design and ability to operate autonomously in both friendly and hostile environments. Nonetheless, the great flexibility of wireless networking comes at the price of an increased vulnerability to security attacks and performance tradeoffs become unavoidable to address security concerns [1].

The broadcast nature of radio transmissions makes wireless networks more susceptible to attacks than their wireline based counterparts. In essence, it is easy for an adversary to eavesdrop and intercept wireless transmissions by just deploying antennas in the region of interest. Successful eavesdropping could then be followed by a wide range of attacks, such as packet replay, impersonation, data manipulation, etc. The conventional means for tackling such vulnerability is to pursue packet encryption in order to safeguard the data and the anonymity of the source. While packet encryption is effective in sustaining privacy, ensuring data integrity, and enabling authentication, it cannot prevent an adversary from conducting traffic analysis to uncover the network topology, and determining the roles that nodes play. To elaborate, an adversary can intercept transmissions and localize the sender, even if the packet is encrypted [2][3]. By correlating the intercepted frames from various nodes, the adversary can conduct flow analysis to determine the packet dissemination paths [4], or apply statistical analysis of the traffic intensity to identify critical nodes [2].

Numerous techniques have been proposed in the literature to counter the traffic analysis threat in the realm of wireless networks [2][5][6]. The key design principle of these countermeasures is to conceal the relationship between communicating nodes by provisioning anonymity, e.g., using pseudonyms, and confusing the link between transmitter-receiver pairs, e.g., by avoiding handshake protocols. Thus, the effectiveness of the countermeasure will be dramatically degraded if an adversary can distinguish between the nodes through other means such as Radio frequency (RF) fingerprinting. It has been shown that radio transceivers, even those built by the same manufacturer, exhibit some unique characteristics for the transmitted signal [7]-[10]. In essence, these unique characteristics become a radiometric signature for the device. An adversary could intercept transmissions and apply machine learning techniques to classify the features and differentiate between the various RF emitters. Thus, RF fingerprinting could identify the transceiver, and consequently distinguish its presence among the communicating nodes in a network and strengthen the adversary analysis.

This paper characterizes the use of RF fingerprinting as an adversarial means and proposes an effective countermeasure. We first investigate the accuracy of the RF fingerprinting by studying the various transmitter-specific features. We show that for accurate radiometric signature detection, the RF features used for device classification are mainly extracted from the modulation domain; we report on experiments with sample wireless protocols in 2.4 ISM GHz, namely Zigbee

978-1-7281-5089-5/20/$31.00 ©2020 IEEE

where contemporary machine learning algorithms are applied to associate the radiometric signature to the involved devices. We then highlight how traffic analysis attack models can leverage RF fingerprints to boost the attack success. Finally, we devise an effective technique for degrading the accuracy of RF fingerprinting and consequently nullifying its contribution to traffic analysis. Our proposed countermeasure exploits the feasibility of switching among multiple wireless protocols in order to confuse the adversary about the modulation schemes and introduce high error in the classification process of the device-specific RF features. Adversarial machine learning techniques are employed to determine the protocol used for each packet transmission in order to introduce the most classification error for the attacker. Our proposed countermeasure is validated through extensive simulation as well as prototype experiments. To the best of our knowledge, our work is the first to tackle possible malicious use of RF fingerprinting without imposing hardware modifications.

The paper is organized as follows. The next section discusses related work in the literature. Section III summarizes the system and attack models. Section III focuses on RF fingerprinting and reports our findings through lab experiments. Section IV shows how RF fingerprinting can be exploited by an adversary in conducting traffic flow and statistical signature analyses. Section V presents our proposed countermeasure. The validation results are reported in Section VI and the paper is concluded in Section VII.

II. Related Work

Radiometric can be classified as location dependent and location independent. For example, the RSSI at the receiver could serve as signature for the transmitter assuming that the two communicating nodes are stationary. Obviously such a type of radiometric is location dependent, is sensitive to the environment condition, lacks flexibility, and is prone to errors. On the other hand, location-independent radiometric relies on features related to the transceiver hardware. As pointed out earlier, variations among the transceiver characteristics enable the distinction among them where each will have a radiometric signature that can serve as an identifier. These characteristic variations are caused by the manufacturing process and do not hinder the transceivers from successful operation. Numerous characteristics such as power amplifier imperfections, clock offset, amplitude and phase errors, etc., have been explored in the literature [7]. Location-independent radiometric techniques fall into two categories based on the used characteristics into waveform-based and modulation-based. The former exploits time and frequency representation of the transmitted signals [11]-[13]. Meanwhile, modulation-based techniques utilize sampled I/Q data and are shown to be more robust [7]-[9].

Other than investigating exploitable signal features, published work on the radiometric signatures either studies how to increase the device identification accuracy and robustness [9][11][14][15], or explores their use in device authentication for both security and crime-forensics applications [16]-[18]. To yield accurate signatures, Peng el al. [9] improve the classification by factoring in the channel conditions in determining feature weights, while Bihl et al. [11] point out that phase based features are more important. Meanwhile, Baldini et al. [14] use multiple receivers to define "golden reference" in order to make the RF fingerprint more robust and less sensitive to the receiver circuit so that the radiometric of a node remains associated with such a node for all receivers. On the other hand, Wu et al. [15] opt to boost the accuracy and robustness of the RF fingerprinting process through the data processing algorithm where deep learning techniques are employed to do the classification.

We note that very few studies have considered radiometric manipulation. The approach of [17] is geared for forensics where a criminal is to be identified within a pool of suspects. The criminal is assumed to be able to exert control on the radio circuit to fool the investigator. Meanwhile, in [19] a radiometric is superimposed on a wireless transmission to enable authentication. Unlike such work, we do not introduce any changes to the radio circuit; instead we pursue protocol switching, guided by an adversarial machine learning methodology in order to prevent the exploitation of the radiometric in violating the anonymity of nodes and in launching traffic analysis attacks. To the best of our knowledge, our work is the first to tackle such a challenge.

III. Accuracy of Radiometric Singnature

To assess the effectiveness of RF fingerprinting and also check its dependency on the pursued wireless protocols, we have experimented with Zigbee and Bluetooth devices. All three protocols operate in the 2.4 ISM GHz Band, and are widely used in practice. The data is collected using a real-time spectrum analyzer that can handle radio frequencies up to 8 GHz with a maximum vector span of 36 MHz. It also easily captures continuous, intermittent or random signals generated in the close proximity of the antenna. The collected data is then preprocessed to extract all the important characteristics of the captured RF signals. The preprocessed data is used as input to two popular classifiers, namely SVM (Super Vector Machines) and KNN (K-Nearest Neighbors), both written in Python. All considered features, enlisted in Table 1, are from the modulation domain and extracted using the "digital demodulation mode" of the spectrum analyzer. The features in the table are ordered from the most effective to the least effective, based on the experiment results.

Here, we summarize the results for Zigbee devices; the experiments with Bluetooth have yielded consistent results and findings. Three groups of wireless motes operating under the Zigbee protocol have been used in the experiments: (i) 4 Xbee S2C wireless motes, (ii) 2 Xbee SMT Development Board (which also can be used as transmitters and receivers), and (iii) 10 Memsic IRIS Motes of the same model and manufacturer. The entire dataset was split into training set and testing set in order to enable effective training for the SVM classifier. The accuracy is calculated as $T_p/(T_p+F_p)$, where T_p is true positive, and F_p is false positive. Overall, the obtained classification accuracy ranges between 65-84%. The experiments with Iris motes have demonstrated that I/Q origin offset could be helpful in identifying the correct node; nevertheless, the accuracy in this case stays at approximately

65% compared to 85% achieved with Xbee Motes. In addition, it has been observed that not all wireless motes of the same model and manufacturer are equally recognizable by the classifier, that is, two of the four motes used in Xbee experiments have shown an identification rate of average 85% throughout the experiments, while the identification of the other two motes has only reached an average accuracy of 70%.

Table 1: Radiometric features used for classification

Feature	Description
Error vector magnitude	The vector of the magnitude difference between the ideal and measured phasors
Magnitude Error	The difference in magnitudes between the ideal and measured phasors
Phase Error	The angle between the ideal and measured phasors
Frequency error	The difference between the ideal and observed carrier frequencies, reflecting the amount by which the receiver's frequency had to be adjusted from the channel center in order to achieve carrier lock
I/Q origin offset	The distance between the origin of the ideal I/Q plane and the origin of the observed symbols

We briefly report on how the selection of different radiometric features impacts the overall classification accuracy. The highest accuracy (~84%) has been reached with SVM when using only modulation error features, i.e., magnitude error, phase error and frequency error. Another observation is that each classifier (KNN and SVM) uses a different set of features to achieve the highest accuracy; KNN performs at best (70%) when the error vector magnitude is added to modulation error features, while for the same feature SVM drag the performance of SVM to 60%. In the same manner, the best feature used for SVM yields the worst accuracy (44%) in case of KNN. Overall, it can be stated that all motes can be correctly classified with the maximum accuracy of 84%. Based on the observations, the most effective features leading to the lowest misclassification rate are modulation errors, i.e., magnitude error, phase error and frequency error, and consequently are used in our performance analysis. Since the feature space is high-dimensional (more than two features are used for classification) SVM obtains more accurate classification results than KNN and hence the latter will not be considered in the balance for the paper. Overall, the experiment results have confirmed the effectiveness of radiometric signatures in identifying devices, which as we discuss in the next section can be exploited by adversaries in launching privacy and traffic analysis attacks.

IV. RF Fingerprinting Exploitation

The major advances in wireless communication devices have led to the prevalence of networked system solutions for civil and military applications. In many of these applications, security is a core requirement and consequently the design and operation of the involved networks are to be concealed. For example, a networked set of sensor nodes could be deployed to operate unattended in a combat zone or along a border, where devices are to be camouflaged to avoid being captured and tampered with. Yet, the broadcast nature of wireless communications allows an adversary to eavesdrop in order to intercept transmissions. Although employing encryption will deprive the adversary from extracting relevant information from the intercepted packets, it cannot prevent traffic analysis [2]. The goal of the traffic analysis is to uncover the network topology and identify key players, such as active sources, data sink, and critical relays that could be targeted by pinpointed attacks, e.g., radio jamming. Similar scenarios could be enumerated in the context of smart cities, internet of things, etc. Sustaining node anonymity, in terms of both identity and role, is the main defense strategy against traffic analysis. However, RF fingerprinting degrades the resilience of the network and diminishes the anonymity of nodes, and consequently makes the network vulnerable. In the following, we elaborate on the possible adverse effects of RF fingerprinting through two examples of traffic analysis attacks, namely, traffic statistical profiling and flow correlation. .

A. *Traffic Statistical Profiling*

When the network deployment area is physically inaccessible to the adversary, an effective means for conducting traffic analysis is to track the spatial distribution of transmissions. By intercepting transmissions, e.g., through highly sensitive antennas, and then estimating the location of transmitters, e.g., using trilateration algorithms, the adversary would form a statistical distribution of traffic intensity that can be further analyzed to uncover the network topology. Given the localization error and computational complexity, the analysis is typically based on a grid overlay of the deployment area where the number of transmissions per cell is used as an indication of the presence of important nodes within such a cell. For example, in sensor networks the vicinity (cell) of the base-station experiences high traffic volume and consequently the cell with the most transmission count could be where the base-station is located. Let N be the number of cells, and p_i be the probability at time t that a cell i contains the base-station. Then, the entropy of the system at time t is defined as:

$$entr_t = -\sum_{i=0}^{N-1}[p_i \times \log_2 p_i] \qquad (1)$$

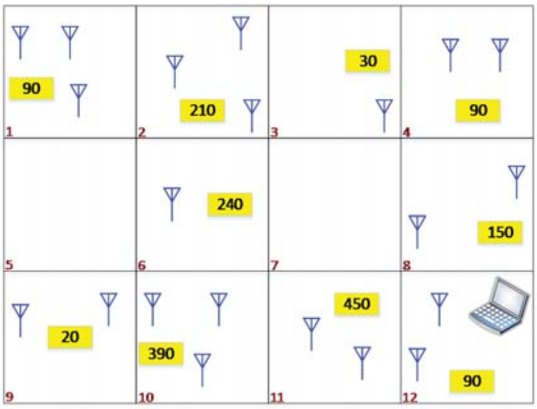


Figure 1: An articulation of a traffic statistical profiling attack scenario. The number in each cell reflects the count of intercepted transmissions. An antenna denotes a node, while the laptop mimics the base-station, i.e., data sink.

Initially, the probability for each cell will be 1/*N*, which corresponds to the maximum entropy. With successful interceptions, the distribution of packets over the area changes and consequently cells' probabilities are adjusted. Such a process takes time and is still subject to errors. Consider the example scenario in Figure 1. The cells with high transmission count do not correspond to the base-station. However, using radiometric signatures the adversary can estimate the node density and refine the analysis and/or identify highly active nodes, e.g., critical relays such as the one in cell #6, in order to distinguish the cells with high relaying rather than packet generation rates. Packet relaying will be more indicative of proximity to the base-station. Thus, RF fingerprinting could enable traffic analysis to converge rather fast.

B. Flow Correlation Attack

In this attack, the packet flow going to and out of the individual nodes is analyzed to determine where data from certain sources ends up being delivered. The objective is to determine the communicating parties over multipath routes. Such send/receive association enables the adversary of uncovering the role of the various network nodes and determining critical relays for successful network operation. Basically, the packet inter-arrival times are monitored and analyzed to estimate the traffic volume between node pairs and assess the portion of outflow from a node *i* reaching another node *j*. Such an attack scenario stays effective even if that path is intentionally extended with the incorporation of relays [4]. In unattended setups, concealing the identity of nodes is a key mitigation measure for such correlated flow attacks since the adversary cannot accurately associate a transmission with a node without being in the vicinity, given the usually-high RF-based localization errors. By exploiting radiometric signatures, an attacker can succeed in identifying a transmission source and conducting flow correlation.

V. COUNTERING RF FINGERPRINTING

Overall, lowering the radiometric accuracy will degrade the above-mentioned attacks and diminish their success. In this section we present our approach for decreasing the achievable accuracy of RF fingerprinting. The main idea is to inject data that negatively affects the classification algorithm. Such a process is often referred to as poisoning and evasion, based on whether the training or test data is targeted, respectively. The general methodology is called "Adversarial Machine Learning". In essence, adversarial machine learning (AML) methodically generates confusing data samples based on the actual data in order to results in misclassification. In the context of RF fingerprinting, applying AML will require means to manipulate the transceiver circuitry to produce the confusing data. We argue that such transceiver manipulation is complex and necessities the use of custom-made radio circuits, due to the excessive cost, which is generally unwarranted and often blatantly unacceptable in most applications. Our approach avoids circuit level realization of AML; instead it pursues switching among multiple protocols to achieve the AML goal. Many commercial RF transceivers nowadays support multiple protocols. For example, the DIGI XBee3 wireless module supports multiple protocols, including Zigbee, Bluetooth, WiFi, LTE, etc. Our objective is to transmit the packet while introducing intentional perturbation into the RF fingerprinting data used by the classifier; we do so by mixing multiple protocols and selecting the nearest possible perturbation to the AML-derived values in order to make the radiometric inconclusive, meaning that the adversary cannot distinguish among the nodes with sufficient fidelity.

More formally, given a supervised classifier, the aim is to learn the model *f(x,θ)* from a labeled set of data $\{(x_i,y_i)\}$, where *x* and *y* can be in one of the *k* classes that represent the identity of the nodes. A machine learning model needs to predict probability *P(y/x,θ)* using the loss function *L(x,y,θ)*, where θ is the weighting vector of the classification parameters. AML opts to introduce a small perturbation, Δx, in the input to cause a large change in model output and maximize the loss, i.e., for $\delta>0$: $\|\Delta x\|<\delta$, $\|f(x+\Delta x)-f(x)\|>\varepsilon$. Many techniques have been proposed in the literature for adversarial machine learning [20]. We use the fast gradient method for illustrating the operation of our approach and validating its performance. The fast gradient method utilizes the L2 norm bound ($\|\Delta x\|_2 \leq \varepsilon$) and the derivative of the loss function with respect to the input, i.e., $\Delta x = \varepsilon . \frac{\nabla_{\mathrm{x}} L(x,y)}{\|\nabla_{\mathrm{x}} L(x,y)\|_2}$. By providing the adversarial sample $x_i^* = x_i + \varepsilon . \frac{\nabla_{\mathrm{x}} L(x,y)}{\|\nabla_x L(x,y)\|_2}$, rather than x_i, the goal is to maximally diminish the classifier accuracy. However, in our case an AML technique is constrained in the data it can introduce since no physical layer alteration of the transmitted signals is deemed possible, which is a more practical assumption, given the dominant use of commercial-off-the-shelf transceivers and standard protocols. In other words, it is not possible to control the transmitter in order to provide x_i^*. Therefore, our approach sends a packet with the protocol that yields the closest value to x_i^*.

To counter RF fingerprinting, our approach pursues protocol switching and employs AML to determine what protocol to use at a certain instance of time, specifically when sending some specific packets from a set of buffered packets. Again the AML technique cannot generate the exact transceiver imperfection that corresponds to the adversarial (poisoned) data, due to lack of physical layer control. Therefore, the transmission protocol for a subset of packets is picked such that the resulting features are highly discriminative to mislead the classification. Assuming a set of packets $\Psi=\{\psi_1, .., \psi_n\}$, and a set of optional secondary protocols $P=\{p_1, .., p_m\}$, our approach strives to determine for a node *k* the set of transmissions $T = \{(\psi_i , p_j)\} \ \forall \ 1 \leq i \leq n$, where $p_j \in P$, such that the accuracy of the radiometric signature of node *k* is minimized. This is accomplished in fast gradient by picking for each ψ_i a protocol p_j for which Δx_i is the least, i.e., smallest perturbation. Let Δx_i^j be the perturbation introduced by using protocol p_j for sending ψ_i. The set of *T* can be defined as follows:

$$(\psi_i,\ p_j) \mid \Delta x_i^j = \min\,(\Delta x_i^1, \Delta x_i^2, .., \Delta x_i^m) \ \ \forall \ 1 \leq i \leq n \quad (2)$$

In practice, the receiver and the transmitter need to agree on a deterministic protocol of communication to avoid any mismatch between the protocol used for packet sending, and what is expected by the receiver. This issue can be tackled by either: (i) employing transmission preamble based on which the receiver can determine the protocol that the sender will soon use. Such preamble is usually small and is sometimes used by MAC protocols that support sleep modes [21]; (ii) using a control channel that is distinct from the data channel; and (iii) pre-agreeing on a specific protocol mix, i.e., protocol use frequency within a batch of packets, and allowing packet ordering to make the protocol use pattern predictable. Given that the third option achieves instantaneous agreement and is the easiest to implement, we study its performance through simulation in the next section. Moreover, there will obviously be a trade-off between packet buffering and security. On the one hand, it is better to buffer more packets in order for the optimization formulation above to yield better solution (i.e., expand the solution space). On the other hand, buffering will not only require storage but also impose packet delivery latency. Therefore, we expect the buffering aspect to be application dependent. In the next section, we study the implication of buffering on the performance of our approach.

VI. Validation Experiments

To validate the effectiveness of our approach, we use experimental and synthetic data. The former is collected while experimenting with the Xbee and IRIS nodes, as discussed in Section III. We have also developed a simulation environment to generate synthetic data. We have considered two prominent machine learning algorithms, namely, SVM and Neural networks (NN). We evaluate the effectiveness of both protocol switching and AML strategies in terms of the variation in the accuracy of the nodes identification. Due to the absence of any contemporary approach that tackles the same problem without any specific hardware for the manipulation of the quality of the packets transmitted, we compare our approach to the baseline case where no countermeasure is applied. The "sklearn" python library and the adversarial-robustness-toolbox from IBM [22], specifically the fast gradient method, have been used for generating AML data.

A. Simulation Environment

In order to study the performance of our approach under varying frequencies of protocol usage and device counts, we have developed a simulator using MATLAB to generate synthetic datasets. The simulator generates waveforms using Zigbee (Zbee) and Bluetool Low Power (BLE) libraries of MATLAB. Although Zbee and BLE are used in the evaluation, our methodology is generic and applies to other protocols. Imperfections in terms of amplitude and magnitude noise are generated and associated with each node, using uniform random distributions. The specific imperfections for node "*i*" are added to the RF transmissions when such a node sends a packet. Then, Gaussian channel noise is included in order to simulate the loss due to the communication. The receiver (adversary) measures digital modulation quality parameters in order to identify the transmitting node, as explained in Section III. In the simulation, we track the error vector magnitude (EVM), magnitude error (MagErr), phase error (PhaseErr), and modulation error ratio (MER). These measurements can be represented by the points on the constellation diagram.

To collect these measurements, the adversary estimates the errors by comparing the received signals and those of an ideal noise-free transmission. The latter may be approximated using a filter such as Root Raised Cosine, Gaussian and Half Sine. In the simulation, we apply the raised cosine filter since it is frequently used for pulse-shaping in digital modulation and due to its ability to minimize inter-symbol interference. Once the ideal signal is determined, the vector of modulation error can be produced using the actual constellation points and the corresponding reference constellation points in every symbol period. The magnitude error, phase error and EVM are estimated using the following formulas [23]:

$$MagErr_{rms} = \sqrt{\frac{\sum_{i=0}^{i+N-1}(|S_i|-|R_i|)^2}{\sum_{i=0}^{i=N-1}|R_i|^2}} \tag{3}$$

$$PhaseErr_{rms} = \sqrt{\frac{1}{N}\sum_{i=0}^{i=N-1}(argS_i - argR_i)} \tag{4}$$

$$EVM_{rms} = \sqrt{\sum_{i=0}^{i=N-1}\frac{|S_i-R_i|^2}{\sum_{i=0}^{i=N-1}|R_i|^2}} \tag{5}$$

where *S* is the received signal and *R* is the estimated (ideal) reference signal. MER is another important feature that is considered in the simulation, but the error is calculated from the signal's average power. MER includes all imperfections including deterministic amplitude imbalance, quadrature error and distortion, while noise is random by nature.

Using the aforementioned features, the adversary applies machine learning algorithms, namely, SVM and NN. We used the following parameters for SVM: RBF kernel with coefficient is set to 1/number of features, regulazier C equal to 1.0, and the shape of the decision function is one versus one. For neural networks, we used Rectified Linear Unit as an activation function with 15 hidden layers and automatic batch size. L-BFGS is uses as a solver and the learning_rate_init, maximum iteration and momentum are set to 0.001, 200 and 0.9, respectively. We have evaluated the effectiveness of both protocol switching and unconstrained AML strategies in terms of the variation in the node identification accuracy. To nullify the effect of RSSI, we have set the power of the transmitter based on the proximity to the adversary. The density of nodes (transmitters) has been varied from 4 to 14. The protocol switching frequency has been variated from 0 to 50% by first assuming unconstrained packet buffering in terms of space and latency in order to grow the set of available packets for the AML-based selection strategy. Additional simulations have been conducted to capture the effect of restricted buffering.

B. Simulation Results

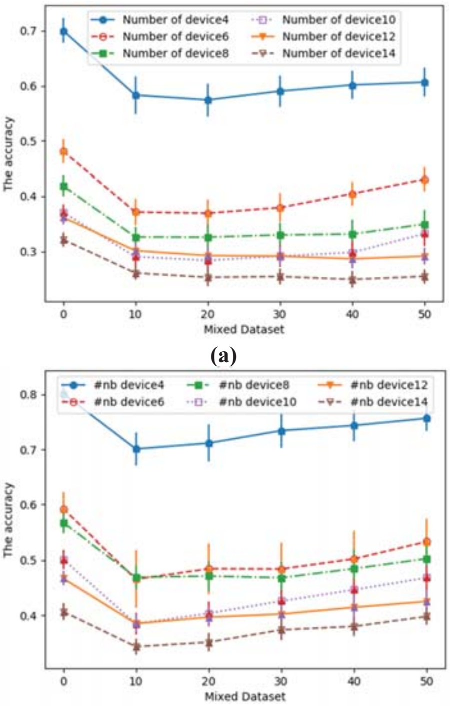
Protocol switching: Figure 2 reports the performance of SVM and NN when Zigbee is used as the primary protocol for communication and BLE is used for poisoning (confusing the classifier). Clearly, increasing the device density will diminish the accuracy as the anonymity set is larger and it becomes harder to distinguish among the involved device. It is important to note that the accuracy is high for both classifiers when no poisoning data is involved, which corresponds in the figure to "zero" mixed dataset. The accuracy decreases significantly when the percentage of poisoned data grows by having 20%-50% of the packets being sent using Bluetooth. Such performance confirms the effectiveness of AML-based protocol switching mechanism in countering RF fingerprinting exploitation. It should be noted that it does not make sense for the poisoning data to exceed 50% since it will cause Bluetooth to be the primary protocol that determines the radiometric signature. Similar conclusion can be drawn when BLE is used as a primary protocol and Zigbee for poisoning as shown in Figure 3. Significant decrease in the accuracy is achieved by transmitting just 10% of the packet using Zigbee. However, the accuracy stabilizes and even grows as more packets being sent using Zigbee since the radiometric accuracy of Zigbee is found to be much higher than BLE, and given that packets using both protocols are fed to the classifier at training time.

Adversarial protocol switching: Figure 4 reports how the AML-based protocol switching strategy is affected by packet buffering, when Zigbee is used as the primary protocol. The entry for "Pure Adv." reflects the case when large packet volume (400 packets) are buffered and the best mix of Zigbee and BLE based transmissions of these packets are picked such that the accuracy is diminished the most. The other entries correspond to buffering 2, 4, 6, 8 and 10 packets. The ratio of poisoning data is also varied. The fast gradient method is deemed to be a very effective adversarial technique against the SVM classifier, which is evident from the results in Figures 4(a). The accuracy is also dramatically decreased when a NN classifier is used (see Figures 4(b)). Both figures illustrate the significant impact of using protocol switching even with limited buffering. To determine the packet mix, the ceiling of the poisoning percentage of the buffer is used. That is for a buffer of 8 and 33.3% poisoning rate, 3 (rather than 2.33) secondary-protocol packets are included every 8 primary-protocol packet transmissions. Therefore, there is a slight leap in the 33.3% curve at buffer of 8. Again, the higher the poisoning rate is, the more accuracy reduction becomes. Same trends have been observed when BLE is used as the primary protocol (results are not shown due to space constraints).

C. Experiments Results

We have assessed the effectiveness of our proposed AML-based protocol switching strategy using the data collected while experimenting with Xbee and IRIS nodes. The results shown in Figure 5 are based on four Xbee nodes, where the IRIS data are mixed with the Xbee data to confuse the classifiers. Three mixes are tried, namely, 50%, 33.3%, and 20%, reflecting the cases of sending an IRIS packet after every one, two and four Xbee transmissions, respectively. Basically, the characteristics of each packet are captured based on transmission using IRIS and Xbee. For the unconstrained scenario, all packets generated during the experiments are considered at the same time and each packet is associated with IRIS or Xbee transmissions such that the radiometric accuracy is maximally diminished. The results for the SVM classifier are shown in Figure 5 under "pure Adv.".

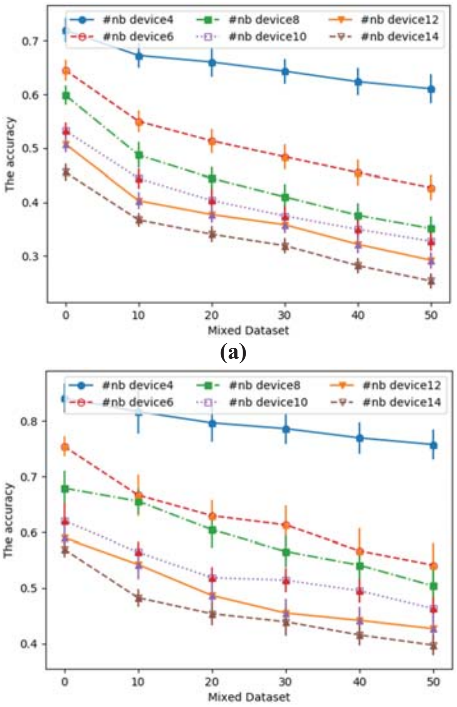
When packet buffering is constrained, the effect on



(a)

(b)

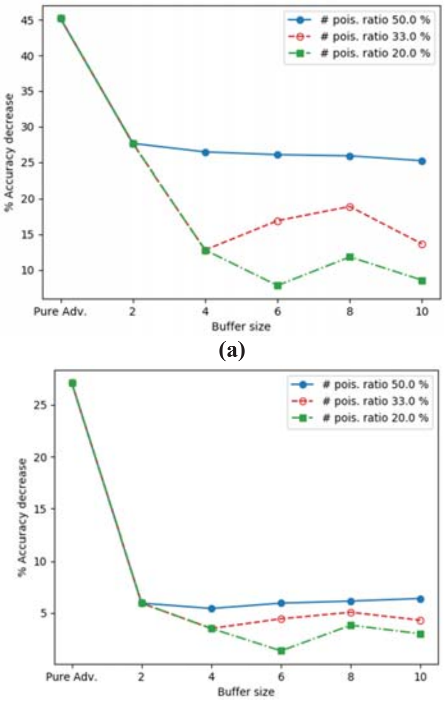
Figure 2: FR fingerpringing accuracy of Zigbee when poisoned with BLE data while using (a) SVM, and (b) NN.



(a)

(b)

Figure 3: The effect of device density and AML on BLE based radiometric while using (a) SVM, and (b) NN.



(a)

(b)

Figure 4: Effect of packet buffering on accuracy when Zigbee is the primary protocol while using (a) SVM, and (b) NN.

accuracy declines since fewer options become available to the AML based protocol selection strategy; in such a case poisoning at a high rate is advisable. Again the leap for poisoning rates 20% and 33.3% is due to rounding, as discussed earlier in Section VI-B. Otherwise, increasing the buffer size have nominal effect unless a large number of packets can be buffered. On the other hand, an increased poisoning rate is more advantageous, where a rate of 50%, i.e., equal split of packets among the two protocols, achieves about 40% reduction in accuracy. We note that similar trends have been observed when NN is applied (plot is omitted due to space limitations). Finally, we note that the goal is to show the effect of AML on the radiometric accuracy and confirm the simulation results, although the IRIS and Xbee transmissions are not based on the same node.

VII. Conclusions and Future Work

This paper has presented a novel adversarial strategy to prevent the exploitation of RF fingerprinting to identify devices and conduct traffic analysis. We first have demonstrated the effectiveness of RF fingerprinting in determining radiometric signatures using machine learning techniques. We have further analyzed the accuracy of RF fingerprinting and highlighted how the accuracy affects the success of adversary attacks. Then, we have developed a novel countermeasure based on switching among preset communication protocols and employing adversarial machine learning to determine the protocol selection for a transmission so that the accuracy of the RF fingerprinting diminishes. In the future, we plan to investigate the integration of the physical-layer based radiometric signatures with other mechanisms at higher layers in the communication protocol stack for both authentication of devices and for supporting anonymity.

Acknowledgment: Dr. Younis and Dr. Baroudi would like to acknowledge the support of the Deanship of Scientific Research at King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia under the grant IN171025.

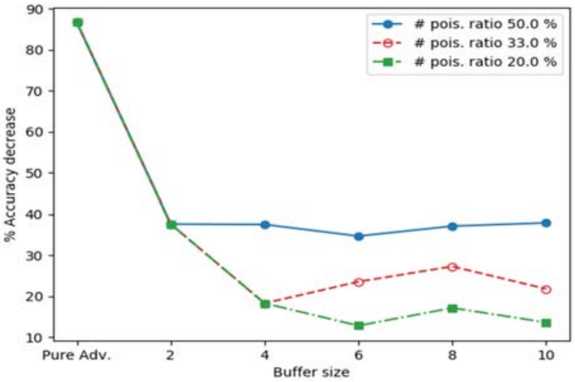


Figure 5: Results when applying AML-based protocol switching using data collected for the Xbee and IRIS devices. Four nodes are assumed to be deployed. IRIS transmissions are used to poison Xbee data for the SVM classifier.

Reference

[1] Y. Zou, J. Zhu, X. Wang and L. Hanzo, “A Survey on Wireless Security: Technical Challenges, Recent Advances, and Future Trends,” *Proceedings of the IEEE*, Vol. 104, No. 9, pp. 1727-1765, Sept. 2016.

[2] N. Baroutis, and M. Younis, “Location Privacy in Wireless Sensor Networks,” Book Chapter, *The Philosophy of Mission-Oriented Wireless Sensor Networks*, Ed. Habib Ammari, Springer, 2019.

[3] M. Mahmoud and X. Shen, “A Novel Traffic-Analysis Back Tracing Attack for Locating Source Nodes in Wireless Sensor Networks,” *Proc. of IEEE Int’l Conf. on Comm. (ICC 2012)*, Ottawa Canada, June 2012.

[4] Z. Ye, F. Xinwen, B. Graham, R. Bettati, and Z. Wei, “Correlation-based Traffic Analysis Attacks on Anonymity Networks,” *IEEE Trans. on Parallel and Distributed Systems*, Vol. 21, No. 7, pp. 954-967, 2010.

[5] A. Proaño, et al, “Traffic Decorrelation Techniques for Countering a Global Eavesdropper in WSNs,” *IEEE Transactions on Mobile Computing*, Vol. 16, No. 3, pp. 857-871, March 1 2017.

[6] R. Manjula, R. Datta, A Novel Source Location Privacy Preservation Technique to Achieve Enhanced Privacy and Network Lifetime in WSNs,” *Computer Networks*, Vol. 44, pp. 58-73, Feb. 2018.

[7] Q. Xu, R. Zheng, W. Saad, and Z. Han, “Device fingerprinting in wireless networks: Challenges and opportunities,” *IEEE Comm. Surveys Tutorials*, Vol. 18, No. 1, pp. 94–104, First quarter 2016.

[8] V. Brik, S. Banerjee, M. Gruteser, and S. Oh, “Wireless device identification with radiometric signatures,” *Proc. of MOBICOM 2008*, San Francisco, California, Sept. 2008.

[9] L. Peng, A. Hu, J. Zhang, Y. Jiang, J. Yu and Y. Yan, “Design of a Hybrid RF Fingerprint Extraction and Device Classification Scheme,” *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 349-360, Feb. 2019.

[10] 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 Transactions on Information Forensics and Security*, Vol. 11, No. 8, pp. 1862-1874, Aug. 2016.

[11] M. M. U. Rahman, A. Yasmeen, and J. Gross, “Phy-layer authentication via drifting oscillators,” *Proc. of the IEEE Global Communications Conference (GLOBECOM 2014)*, Austin, TX, December 2014.

[12] M. Köse, S. Taşcioğlu and Z. Telatar, "RF Fingerprinting of IoT Devices Based on Transient Energy Spectrum," *IEEE Access*, vol. 7, pp. 18715-18726, 2019.

[13] A. C. Polak, S. Dolatshahi and D. L. Goeckel, “Identifying Wireless Users via Transmitter Imperfections,” *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 7, pp. 1469-1479, August 2011.

[14] G. Baldini, R. Giuliani, C. Gentile and G. Steri, “Measures to Address the Lack of Portability of the RF Fingerprints for Radiometric Identification,” *Proc. of the 9th IFIP International Conference on New Techn., Mobility and Security (NTMS)*, Paris, France, February 2018.

[15] Q. Wu et al., “Deep learning based RF fingerprinting for device identification and wireless security,” *Electronics Letters*, vol. 54, no. 24, pp. 1405-1407, 29 11 2018.

[16] M. Ezuma et al., “Micro-UAV Detection and Classification from RF Fingerprints Using Machine Learning Techniques,” in the *Proceedings of the IEEE Aerospace Conference*, Big Sky, Montana, March 2019.

[17] A. C. Polak and D. L. Goeckel, “Identification of Wireless Devices of Users Who Actively Fake Their RF Fingerprints with Artificial Data Distortion,” *IEEE Transactions on Wireless Communications*, 14(11), pp. 5889-5899, Nov. 2015.

[18] Q. Yang, H. Qin, X. Liang, and T. A. Gulliver “An Improved Unauthorized Unmanned Aerial Vehicle Detection Algorithm Using Radiofrequency-Based Statistical Fingerprint Analysis,” *Sensors*, 19(2), pp. 274-295, 2019.

[19] P. L. Yu, G. Verma and B. M. Sadler, "Wireless physical layer authentication via fingerprint embedding," *IEEE Communications Magazine*, vol. 53, no. 6, pp. 48-53, June 2015.

[20] S. Qiu, Q. Liu, S. Zhou, and C. Wu, “Review of Artificial Intelligence Adversarial Attack and Defense Technologies,” Appl. Sci. Vol. 9, Vol. 5, pp. 909-937, 2019.

[21] Y. Z. Zhao, et al. “A survey and projection on medium access control protocols for wireless sensor networks,” *ACM Comput. Surv.*, Vol. 45, No. 1, Article 7, December 2012.

[22] https://adversarial-robustness-toolbox.readthedocs.io/en/latest/

[23] F. Hong, B. Xin, Z. Xin, and L. Ke, “The Numerical Simulation and Experiment Research for Measurement of Error Vector Magnitude (EVM),” *Applied Mechanics and Materials*, Vol.103, pp199-204, 2012.