Improving Intra-Cellular Security Using Air Monitoring with RF Fingerprints

Donald R. Reising, Michael A. Temple and Michael J. Mendenhall

Department of Electrical and Computer Engineering
Air Force Institute of Technology
Wright-Patterson AFB, OH 45433 USA
Email: michael.temple@afit.edu

Abstract—Improved intra-cellular security is addressed using device-specific RF fingerprints to mitigate malicious network activity that can occur through unauthorized use of digital identities. In air monitoring applications where physical equipment constraints are not overly restrictive, RF fingerprinting remains a viable option for providing regional intra-cellular security for systems such as cellular telephone and last mile WiMax networks. Proof-of-concept results are provided for GSM signals given they are readily available in most areas. Recent RF fingerprinting work has demonstrated average device classification accuracies (serial number identification) of 92% using OFDM-based 802.11a preamble responses at SNR = 6 dB. The goal here was to determine if similar performance could be achieved using RF fingerprints extracted from near-transient and midamble regions of GSM signals. This was done using instantaneous phase responses from each region to form RF statistical fingerprints that are subsequently classified using Fisher-based MDA/ML processing. Considering all GSM device permutations from four different manufacturers, near-transient RF fingerprinting provided nearly 13% improvement in classification performance when compared with midamble RF fingerprinting and achieved average classification performance consistent with the 802.11a benchmark of 92% correct classification at SNR = 6 dB.

I. INTRODUCTION

The current ubiquity of Global System for Mobile Communications (GSM) service and anticipated proliferation of last mile Worldwide Interoperability for Microwave Access (WiMax) networks places intra-cellular wireless users at great risk. As such, there remains a critical need to provide improved authentication and security measures. Physical (PHY) layer security using RF signal characteristics is one alternative given that such characteristics are nearly impossible to mimic. Using appropriate characteristics, “RF fingerprints” can be formed, a device’s identity established, and malicious network activity detected and/or mitigated. The goal is to augment bit-level security mechanisms that are based on digital identities and commonly exploited for network spoofing [1]–[3].

The concept for using an “air monitor” to capture PHY layer attributes and provide added security is not new [1], [4]. For work presented here, air monitor functionality (RF fingerprint generation, classification, authentication, etc.) is not envisioned as residing within each network device and is subject to physical size constraints. Rather, the near-term implementation goal is a separate air monitor system that provides regional monitoring and distribution of security status to protected nodes. Such a system could provide intra-cellular security for systems employing regional base stations such as GSM cellular telephone [5], [6] and last mile WiMax networks [7]. Given GSM networks are easily accessed in most areas, proof-of-concept demonstration is conducted using signals collected within a local network. Consistent with earlier work [8], RF fingerprinting is performed using non-transient (near-transient and midamble) regions of GSM normal bursts.

While the transient response of communication signals has been successfully exploited using various RF fingerprinting techniques [9]–[14], the transition to non-transient regions has been successful [4], [15]–[17] and the motivation for doing so remains. The specific non-transient regions of interest (near-transient, preamble, midamble, etc.) are regions where the modulated waveform response is based on a fixed (known or unknown) bit pattern into the modulator. Thus, any induced waveform variation is solely attributable to post-modulator electronics (mixers, filters, amplifiers, etc.) that are unique to a given device. One additional advantage is that the non-transient baseband responses can generally be processed using narrower bandwidth filters and correspondingly lower sampling frequencies relative to transient response processing.

The demonstration methodology here is consistent with [15]–[17] and described in Sect. III. As with results in [8], results here are based on experimentally collected GSM signals from four different manufacturers (Samsung, Nokia, Motorola and Sony Ericsson). However, the results in [8] are expanded here to include classification performance and analysis across multiple device permutations. The unintentional modulation effects (changes in amplitude, phase, frequency and/or statistical characteristics thereof) are generally device specific and dependent upon hardware implementation, component manufacturing materials/processes and/or environmental interaction factors. This work addresses the exploitation of these effects as induced in the near-transient and midamble regions of collected GSM signals. The behavior of these effects is captured through common statistical metrics that collectively form the RF fingerprints used for device classification. Transient and midamble GSM classification performances are compared with each other, as well as 802.11a results in [4], [15], [16] which enable a comparison assessment with alternate signal types.
II. GSM Fundamentals

Signals were collected within a local T-Mobile GSM cellular system for proof-of-concept demonstration. As a PCS-1900 network, mobile station (MS) uplink signals are separated at the base transceiver station (BTS) using Frequency Division Multiple Access (FDMA) across 298 RF channels. The channels occupy 200 KHz of bandwidth and collectively span the 1850.2 to 1909.8 MHz frequency band (59.6 MHz total bandwidth). Additional diversity is provided through frequency hopping and Time Division Multiple Access (TDMA). Eight time slots (TS) are used per TDMA frame and one MS burst is transmitted per channel/frequency hop.

GSM uses Gaussian Minimum Shift Keying (GMSK) modulation with a bandwidth-bit period product of $BT_{Bit} = 0.3$ and Gaussian filter bandwidth of $BW = 81.3$ KHz [5]. The effective GSM bit duration of $T_{Bit} = 3.692$ μsec yields a modulated symbol rate of $R_{Bit} = 270, 833$ Kbps [5]. While several burst structures are used within GSM, the focus here was on MS uplink transmissions using the normal burst structure using bit-to-waveform mapping in Fig. 1 [8]. As shown, the normal burst duration is approximately 547 μsec and followed by an 8.25 μsec guard period. A total of 148 GMSK modulated bits are transmitted per normal burst, including 114 Coded Data bits, 26 Training Sequence Code (TSC) bits, 6 Tail (T) bits, and 2 Stealing Flag (S) bits [5], [6].

The TSC and left-most T bit patterns, which account for the GSM midamble and near transient GMSK responses respectively, are used for RF fingerprinting given they are fixed patterns across multiple MS units within a given cell sector. For example, the T bits are generally set to all zero and there are only eight possible TSC bit patterns that may be assigned by a given BTS to mobiles within its servicing area [5], [18]. Therefore, the GMSK response in the midamble and near transient regions is based on identical input data and would be identical for all devices if their hardware was identical. This is not the case in practice and device specific “coloration” (amplitude, phase, and/or frequency distortion) is induced.

III. Demonstration Methodology

A. Signal Collection and Post-Collection Processing

Signal collection and post-collection processing was performed as detailed in Fig. 2. The RF Signal Intercept and Collection System (RFSICS) is an Agilent E3238S-based system [19] and has a 36.0 MHz RF filter that is tunable across 20.0 MHz to 6.0 GHz. The selected band is down-converted to a 70.0 MHz IF and passed to a digitizer. The digitizing process consists of down-conversion (near baseband), 12 bit analog-to-digital conversion at a rate of 95 M samples-per-second (sps), digital filtering (user defined bandwidth), sub-sampling (Nyquist criteria enforced), and storage as complex in-phase and quadrature (I-Q) components. The GSM cellular telephones under test and the RFSICS were co-located in an office building environment during all collections. Signals were collected using four GSM world cellular telephones: Samsung SGH-a226 (SAM), Nokia 6125 (NOK), Motorola V191 (MOT), and Sony Ericsson W300i (SON).

Observations from local T-Mobile PCS-1900 signal collections included: 1) slow frequency hopping across either two or four channels spanning a maximum of 34 MHz bandwidth (independent of the number of phones activated), 2) a given TSC being assigned by the BTS regardless if one or multiple phones were simultaneously activated, and 3) TSC #5 being assigned most consistently. Given that the phones hopped between either two or four frequencies spanning a maximum bandwidth of 34 MHz, the RFSICS center frequency and bandwidth could be set such that all four GSM channels could be simultaneously collected. Thus, each collected burst was spectrally located near-baseband at one of two or four different frequencies. For these particular RFSICS settings, the collected signals were automatically sub-sampled and stored the complex I-Q data at a rate of $f_c = 47.5$ Msps.

As illustrated in Fig. 2, post-collection processing was accomplished with MATLAB and included: 1) center frequency $f_c$ estimation (either burst-by-burst or across the entire collection of bursts), 2) baseband down-conversion using $f_c$, 3) dig-
ital filtering and 4) power normalization. Like-filtered AWGN was scaled and added to set the desired analysis SNR. For baseband digital filtering, a 6th-order Butterworth filter was implemented with a -3 dB bandwidth of $W_B = 100.0$ KHz, or one-half the 200 KHz GSM channel bandwidth. There was no sub-sampling applied during post-collection processing. Thus, the combination of $W_B = 100.0$ KHz and $f_s = 47.5$ Msps resulted in the final data being heavily over-sampled by as much as 237 times Nyquist.

B. Burst Detection and Feature Selection

Amplitude-based threshold detection was implemented using a −9 dB threshold applied to filtered/smoothed magnitude responses of each collected burst. Following location of the sample number nearest the −9 dB point, samples from each burst (to include a sufficient number of background samples before and after) are stored and a specific region of interest within the collected signals selected for feature generation.

To accurately assess results of this work relative to previous efforts, analytic expressions for instantaneous phase features were adopted from [15], [16] and are presented for completeness. Like a majority of earlier works, instantaneous amplitude, phase and frequency features were initially considered for GSM fingerprinting. However, best case fingerprinting and classification performance was achieved using instantaneous frequency. Thus, only the development of instantaneous phase is considered here included standard deviation ($\sigma$), variance ($\sigma^2$), skewness ($\gamma$), and kurtosis ($k$). Elements of $\mathbf{F}$ are generated by: 1) dividing $\{\phi_{cnl}\}$ into $N_R$ contiguous, equal length subsequences, 2) calculating the four metrics across each subsequence and across all elements in $\{\phi_{cnl}\}$ ($N_R + 1$ total regions), and 3) arranging the metrics as follows:

$$F_{R_i} = [\sigma_{R_i}, \sigma^2_{R_i}, \gamma_{R_i}, k_{R_i}]_{1 \times 4},$$

where $i = 1, 2, ..., N_R + 1$ and

$$\mathbf{F} = [F_{R_1}, F_{R_2}, ..., F_{R_{N_R+1}}]_{1 \times 4(N_R+1)}.$$ (5)

C. Statistical Fingerprint Generation

Statistical fingerprint features are calculated across specific regions of sequence $\{\phi_{cnl}\}$, having elements given by (4), to form the statistical RF fingerprints ($\mathbf{F}$). The statistics considered here included standard deviation ($\sigma$), variance ($\sigma^2$), skewness ($\gamma$), and kurtosis ($k$). Elements of $\mathbf{F}$ are generated by: 1) dividing $\{\phi_{cnl}\}$ into $N_R$ contiguous, equal length subsequences, 2) calculating the four metrics across each subsequence and across all elements in $\{\phi_{cnl}\}$ ($N_R + 1$ total regions), and 3) arranging the metrics as follows:

$$F_{R_i} = [\sigma_{R_i}, \sigma^2_{R_i}, \gamma_{R_i}, k_{R_i}]_{1 \times 4},$$

where $i = 1, 2, ..., N_R + 1$ and

$$\mathbf{F} = [F_{R_1}, F_{R_2}, ..., F_{R_{N_R+1}}]_{1 \times 4(N_R+1)}.$$ (5)

Previous results in [8] for parametric variation in $N_R$ indicated that $N_R = 10$ subregions provided best case, or near best case, performance for variation in other parameters. Thus, all comparative assessment results in this paper are based on $N_R = 10$ subregions. Accounting for $N_R + 1 = 11$ total signal regions, the resultant $\mathbf{F}$ from (5) contained a total of $4 \times 11 = 44$ elements.

The plots in Fig. 3 are based on averages from 200 collected bursts at SNR = 12 dB and are used to help visualize statistical fingerprint behavior across devices and across signal regions. The two signal regions of interest are highlighted in Fig. 3(a) [8] which contains approximately 30,000 samples per burst. The corresponding average fingerprints for the midamble region (4000 samples) and transient region (900 samples) are illustrated in Fig. 3(b) and Fig. 3(c), respectively.

To generate the representations in Fig. 3(b) and Fig. 3(c), the statistics in $\mathbf{F}$ of (5) were grouped to form what are called “Distinct Native Attribute” (DNA) markers. Within each DNA marker, the statistics were scaled, compressed/expanded to span $[0, 1]$ and quantized to a desired number of discrete levels. The quantized markers were stacked to create an “RF DNA Fingerprint” and presented in an electrophoresis-like plot. These reference fingerprints clearly highlight inter-device variability in both the near-transient and midamble regions. The increased variability exhibited in Fig. 3(c) near transient fingerprints relative to Fig. 3(b) midamble fingerprints is consistent with better classification performance as provided in Section IV.

D. MDA/ML Device Classification

The pool of 44-dimensional statistical RF fingerprints, generated using (5) for each collected signal from each device, were input to a Multiple Discriminant Analysis/Maximum Likelihood (MDA/ML) process for device classification. MDA is an extension to Fisher’s Linear Discriminant (FLD) when more than two classes (devices) are being considered. MDA reduces the dimensionality of higher-dimensional input data by projecting it into a lower-dimensional space with the goal of maximizing inter-class separation while reducing intra-class spread [21].
For the three class problem considered here, MDA projects the 44-dimensional input data into a 2-dimensional space. Device classification is performed using a ML classifier as derived from Bayesian Decision Theory, with the 44-dimensional input data classified as being affiliated with one of three possible classes. A Bayesian decision is based on known prior probabilities, probability densities, and relevant costs associated with making a decision [21]. The decision process relies on an accurate representation of the class distribution and its parameters in order to define the likelihood. A sample is evaluated in each class likelihood and the sample is assigned the class label of the class likelihood yielding the maximum response. For ML classification, the prior probabilities are assumed to be equal and the costs uniform. The MDA/ML process was implemented using $K$-fold cross-validation with $K = 5$ to improve the reliability of classification results. While the required value of $K$ can be data dependent, empirical testing with the collected data used here confirmed that $K = 5$ was sufficient to ensure reliability. This finding was consistent with common practice that suggests values of $K = 5$ and $K = 10$ are appropriate [22].

IV. COMPARATIVE CLASSIFICATION PERFORMANCE

Classification performance was assessed using midamble and near-transient responses for all device permutations in Table I and $\text{SNR \in [-235 \text{ dB}]}$. All MDA/ML classification results are based on a total of 5000 classification decisions per input class (device), which accounts for the number of collected bursts that were processed and the number of Monte Carlo noise realizations. The resultant average MDA/ML classification results for GSM midamble and GSM near-transient fingerprinting are provided in Fig. 4. A detailed discussion of these results is provided in the following paragraphs.

Classification performance was first assessed using statistical fingerprints generated from the GSM midamble region highlighted in Fig. 3(a). These results were then compared with previous 802.11a preamble results in [15], [16]. As shown in Fig. 4(a), best case midamble classification performance was achieved with Permutation #4 (least challenging) and worst case classification performance was achieved with Permutation #2 (most challenging). The curve with filled circle markers represents average performance across all permutations. Representative classification per device is provided in Table II which shows the classification confusion matrices for Permutation #4 at three specific SNRs. These results indicate that the Nokia device is the greatest contributor to overall

\begin{table}[h]
\centering
\caption{Permutations of Samsung (SAM), Nokia (NOK), Motorola (MOT) and Sony Ericsson (SON).}
\begin{tabular}{|c|c|c|c|c|}
\hline
Permutation & SAM & NOK & MOT & SON \\
\hline
1 & & & & \\
2 & & & & \\
3 & & & & \\
4 & & & & \\
\hline
\end{tabular}
\end{table}

Fig. 3. Representative GSM burst response and corresponding average RF fingerprints for two highlighted regions and all manufacturers considered.
TABLE II

GSM **MIDAMBLE** FINGERPRINTING: MDA/ML CLASSIFICATION
CONFUSION MATRICES FOR DEVICE PERMUTATION #4 IN TABLE I.

<table>
<thead>
<tr>
<th>SNR = 0 dB</th>
<th>Overall Ave = 52.54%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class Estimate</td>
<td>Actual Class</td>
</tr>
<tr>
<td>NOK</td>
<td>34.94 %</td>
</tr>
<tr>
<td>MOT</td>
<td>21.14 %</td>
</tr>
<tr>
<td>SON</td>
<td>17.28 %</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SNR = 6 dB</th>
<th>Overall Ave = 78.71%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class Estimate</td>
<td>Actual Class</td>
</tr>
<tr>
<td>NOK</td>
<td>61.94 %</td>
</tr>
<tr>
<td>MOT</td>
<td>6.60 %</td>
</tr>
<tr>
<td>SON</td>
<td>11.00 %</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SNR = 12 dB</th>
<th>Overall Ave = 91.99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class Estimate</td>
<td>Actual Class</td>
</tr>
<tr>
<td>NOK</td>
<td>82.78 %</td>
</tr>
<tr>
<td>MOT</td>
<td>2.62 %</td>
</tr>
<tr>
<td>SON</td>
<td>4.16 %</td>
</tr>
</tbody>
</table>

performance degradation, i.e., the Nokia input fingerprints are most often confused (misclassified) with those of the other two devices. The average percentage of correct classification using midamble fingerprints ranges from a low of 52.54% to a high of 91.99%. While 92% correction classification at SNR = 12 dB is promising and may be useful for some applications, this level of performance is generally poorer than previous 802.11a preamble results which demonstrated 92% classification accuracy at SNR = 6 dB [15], [16].

Given the disparity between GSM midamble results and previous 802.11a preamble results, classification performance was next assessed using statistical fingerprints generated from

TABLE III

GSM **NEAR-TRANSIENT** FINGERPRINTING: MDA/ML CLASSIFICATION
CONFUSION MATRICES FOR DEVICE PERMUTATION #4 IN TABLE I.

<table>
<thead>
<tr>
<th>SNR = 0 dB</th>
<th>Overall Ave = 75.95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class Estimate</td>
<td>Actual Class</td>
</tr>
<tr>
<td>NOK</td>
<td>71.32 %</td>
</tr>
<tr>
<td>MOT</td>
<td>24.16 %</td>
</tr>
<tr>
<td>SON</td>
<td>7.30 %</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SNR = 6 dB</th>
<th>Overall Ave = 92.05%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class Estimate</td>
<td>Actual Class</td>
</tr>
<tr>
<td>NOK</td>
<td>90.52 %</td>
</tr>
<tr>
<td>MOT</td>
<td>10.04 %</td>
</tr>
<tr>
<td>SON</td>
<td>0.24 %</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SNR = 12 dB</th>
<th>Overall Ave = 97.89%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class Estimate</td>
<td>Actual Class</td>
</tr>
<tr>
<td>NOK</td>
<td>97.48 %</td>
</tr>
<tr>
<td>MOT</td>
<td>3.24 %</td>
</tr>
<tr>
<td>SON</td>
<td>0.03 %</td>
</tr>
</tbody>
</table>

Fig. 4. Average MDA/ML classification performance for GSM fingerprinting using all device permutations shown in Table I.

the GSM near-transient region highlighted in Fig. 3(a). Classification results for all device permutations in Table I are provided in Fig. 4(b). The curve with filled square markers represents average performance across all permutations. As indicated, Permutation #2 now provides best case overall performance (least challenging). To enable direct numerical
comparison with midamble results in Table II, confusion matrices are provided in Table III for Permutation #4 classification. The average percentage of correct classification using near-transient fingerprints ranges from a low of 75.95% to a high of 97.89% for the same SNRs shown in Table II. Most notably, the average classification performance of 92% at SNR = 6 dB represents nearly 13% improvement over midamble performance and is consistent with previous 802.11a preamble results [15], [16].

A final overall comparison is made in Fig. 4(c) which shows average GSM midamble results from Fig. 4(a) overlaid with average GSM near-transient results from Fig. 4(b). Defining “gain” as the reduction in required SNR to achieve a given percentage of correct classification, these results indicate that near-transient fingerprinting provides 6–15 dB of gain at classification accuracies of 80% or better. A representative gain of approximately 11 dB is indicated in Fig. 4(c) at 90% correct classification.

V. CONCLUSION

For air monitoring applications where physical equipment constraints are not overly restrictive, RF fingerprinting remains promising for uniquely identifying devices and increasing network security. Borrowing concepts from previous RF fingerprinting work that successfully exploited preamble responses of OFDM-based 802.11a signals [15]–[17], this work provides a proof-of-concept demonstration for intra-cellular security enhancement using GSM signals. The MDA/ML classification process was adopted from earlier work and is used here to demonstrate classification performance using statistical RF fingerprints extracted from the instantaneous phase response of GSM midamble and near-transient regions in experimentally collected signals. Inter-manufacturer discrimination is demonstrated using signals collected from Samsung, Nokia, Motorola and Sony Ericsson devices.

Relative to GSM midamble fingerprinting, GSM near-transient fingerprinting proved to be more robust and provided nearly 13% improvement in average classification performance across all device permutations considered. Even more notable is that near-transient fingerprinting achieved 92% correct classification at SNR = 6 dB which is consistent with previous 802.11a benchmark performance [15].

Given previous success with OFDM-based 802.11a signals [15]–[17], and the inter-manufacturer discrimination demonstrated here with intra-cellular GSM signals, considerable interest remains in determining if air monitoring with RF fingerprints can be used for intra-manufacturer device discrimination in other cellular systems such as last mile WiMax networks [7]. Given WiMax is an OFDM-based signal that can be employed in a cellular architecture it is reasonable to expect favorable performance–demonstrating this remains a focus area of ongoing research.

```
The views expressed in this article are those of the author(s) and do not reflect official policy of the United States Air Force, Department of Defense or the U.S. Government.
```

ACKNOWLEDGMENT

This work sponsored by the Sensors Directorate, Air Force Research Laboratory, and the Tactical SIGINT Technology (TST) Program Office.

REFERENCES


