FBSleuth: Fake Base Station Forensics via Radio Frequency Fingerprinting

Zhou Zhuang [1], Xiaoyu Ji [1†], Taimin Zhang [1], Juchuan Zhang [1]
Wenyuan Xu [1†], Zhenhua Li [2], Yunhao Liu [2,3]
[1] Zhejiang University [2] Tsinghua University [3] Michigan State University
{zhuangzhou, xji, ztm1992fly, juchuanzhang, wyxu}@zju.edu.cn, {lizhenhua1983, yunhaoliu}@gmail.com

ABSTRACT

Fake base station (FBS) crime is a type of wireless communication crime that has appeared recently. The key to enforcing the laws on regulating FBS based crime is not only to arrest but also to convict criminals effectively. Much work on FBS discovering, localization, and tracking can assist the arresting, but the problem of collecting evidence accurately to support a proper conviction has not been addressed yet.

To fill in the gap of enforcing the laws on FBS crimes, we design FBSleuth, an FBS crime forensics framework utilizing "radio frequency (RF) fingerprints", e.g., the unique characteristics of the FBS transmitters embedded in the electromagnetic signals. Essentially, such fingerprints stem from the imperfections in hardware manufacturing and thus represent a consistent bond between an individual FBS device and its committed crime. We model the RF fingerprint from the subtle variance of the modulation errors, instantaneous frequency, and phases of the RF signals. Our validation of FBSleuth on six FBSes from four cities over more than 5 months shows that FBSleuth can achieve over 99% precision, 96.4% recall, and 97.94% F1 score in a dynamic wild environment.

CCS CONCEPTS

• **Networks → Wireless access points, base stations and infrastructure**; **Mobile and wireless security**; • **Applied computing → Evidence collection, storage and analysis**;

KEYWORDS

Fake Base Station, RF Fingerprinting, Forensics

1 INTRODUCTION

Mobile communication technologies bring us great convenience by allowing us to check sensitive information and make a payment any time anywhere. Unluckily, crimes that utilize fake base station (FBS) make the mobile communication a weak point for attacks. FBS mainly utilizes the vulnerability of Global System for Mobile Communication (GSM) network. Although a few FBSes can attack the 3G, 4G, and 5G networks, they jam the current 3G, 4G or 5G channels and force the smartphones to fall back to GSM networks.

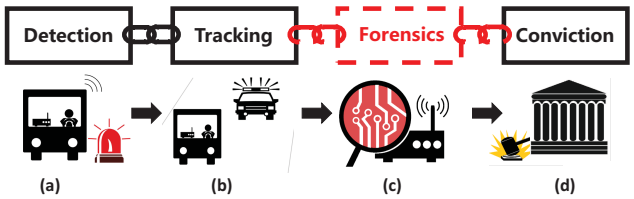


Figure 1: A full loop of law enforcement of FBS crimes consists of four steps, where accurate forensics is missing.

FBS, pretending to be a legitimate base station, can transmit any malicious SMS messages (hereafter FBS messages) that may contain spam advertisements, fishing links, and solicitations for high-fee premium services. These FBSes are pervasively found in the U.S., China, India, Russia, UK, etc. , and are typically cheaply made by underground manufacturing to increase the profit. For instance, an FSB typically cost less than $700 and can earn up to $1400 a day [39]. Not surprisingly, statistics show that even within China, over 5.7 billion fraud messages from FBSes were received in 2015, causing estimated losses of billions of dollars [3–5].

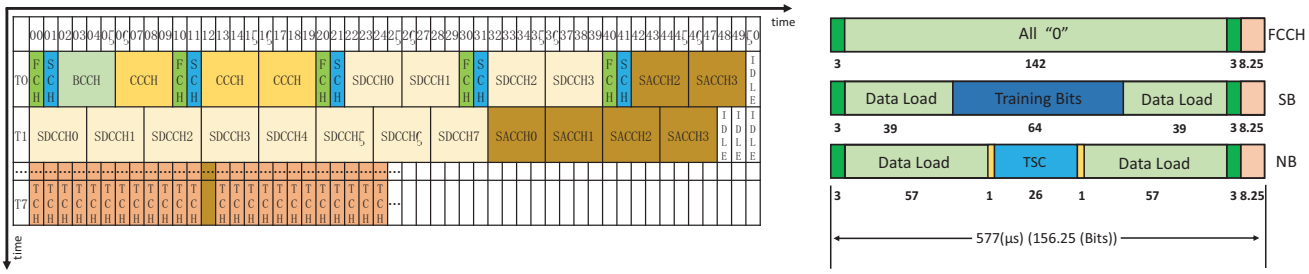
To cope with the aforementioned issues, governments from various countries have passed laws to regulate the FBS devices. In general, it is considered a crime to own and operate FBS devices, and sentences are determined by the level of usage of the devices. For law enforcement, it is critical to both *arrest* and *convict* criminals effectively. To assist arresting criminals, much work has been devoted to assist criminal FBS discovering [10, 18, 27], localization, and tracking [9, 27, 28]. However, to the best of our knowledge, collecting evidence accurately to support a proper conviction has not been addressed yet, as illustrated in Fig. 1. According to the criminal law of China, it is considered to be a serious crime only if the criminal has transmitted more than 5000 FBS messages [6]. Thus, catching a criminal with an FBS in his car may at most lead to a small fine, while a conviction of a serious crime requires proving that over 5000 FBS messages were sent. Thus, in this paper, we investigate the problem of FBS crime forensics, i.e., validating that an FBS message is indeed sent by an illegal device.

For FBSes, existing forensic approaches are either based on log files or inference. The former extracts the log files in the FBS devices [23], which contain the International Mobile Equipment Identity (IMEI) of victims' devices. However, this approach will be invalid once the criminals reset their devices right before being arrested or if they keep erasing them automatically. Alternatively, Telecom providers can infer the existence of an FBS device by estimating the abnormal release report, i.e., the number of users that are disconnected from normal bases stations. Since the FBS transmitters will cause nearby mobile users to disconnect from normal

[†]Corresponding faculty authors.

ASIA CCS '18, June 4–8, 2018, Incheon, Republic of Korea

ACM ISBN 978-1-4503-5576-6/18/06…$15.00
https://doi.org/10.1145/3196494.3196521



(a) GSM Timeslot. (b) Burst Structure.

Figure 2: (a) At each frequency point, signals are divided into slices sequentially and 51 slices form a period. Then each slice is further divided into 8 bursts and 51 bursts assemble one Timeslot, e.g. T0. The Timeslot at a frequency point is a physical channel on which the logical channels are mapped. (b) A GSM burst is the subcomponent of Timeslot and each one contains 156.25 bit, lasting for 577 *us*. 26-bit Training Sequence Code (TSC) from one base station is identical for the Normal Burst and will be utilized for fingerprint extraction.

base stations and connect to themselves by transmitting at a higher power level, an enormous number of disconnections may indicate the existence of an FBS. However, this approach can not directly prove that 5000 FBS messages are sent.

The key to convict FBS crimes is to prove *who transmitted an FBS message*. In this paper, we design `FBSleuth`, an evidence verification framework that utilizes RF (radio frequency) fingerprint to validate the message transmitter. RF fingerprint essentially characterizes the subtle distortion of the messages' electromagnetic signals caused by the transmitters. Given a consistent RF fingerprint for each device, `FBSleuth` fulfills two aspects of evidence verification by (1) matching the RF fingerprint of an FBS message to verify the source of the FBS message and (2) demodulating and decoding FBS messages to count the number of FBS messages. Compared with existing methods, `FBSleuth` utilizes the raw RF signals sent by an FBS to generate RF fingerprints and form the mapping between messages and the FBS. As such, it can even distinguish two FBSes within the same area.

The RF Fingerprint of an FBS is originated from hardware imperfections, e.g., analog circuitry components of the RF front-end, which are introduced during manufacture. In our study, we find that these imperfections lead to subtle but stable variance in the modulation errors, frequencies, and phases of the signals. Different from common RF fingerprints of indoor portable electronic devices, FBS fingerprints should be insensitive to the dynamic ambient noise over a long period, since the distance between the receiver and FBS changes dynamically in the wild and one FBS may keep running for a whole day in practice. Some SNR sensitive features (e.g., amplitude error and error vector magnitude (EVM)), though reflecting the imperfection of hardware and being widely used in fingerprinting indoor devices, are infeasible for FBS fingerprinting. Here, the challenge is to extract RF fingerprints of FBS messages that are consistent for the same devices yet distinct for different ones, regardless of the following challenges.

- How could we extract stable and unique RF fingerprints from mobile FBSes that may travel at various patterns?
- Could the RF fingerprint be consistent while the content of FBS messages, the SNR levels, the supplied batteries, and the receivers are different?
- How to ensure that the RF fingerprints and the matching algorithms lead to a low false positive and a low false negative rate?

To overcome the aforementioned challenges, we extract RF fingerprints from the steady signal of bursts, i.e, utilizing the same segment of signals in FBS messages, unlike many existing transient-signal-based RF fingerprinting methods. Such a method creates an enhanced fingerprint and imposes low requirements on the sampling rate. Our validation on 6 typical real FBSes from four cities during more than five months shows that `FBSleuth` can verify the transmitters with a low false positive rate in various conditions and could effectively assist fighting FBS crimes in the wild. In summary, our contributions are listed as follows.

- To the best of our knowledge, we are the first to propose FBS crime forensics utilizing the RF fingerprint technology.
- We design and implement a framework, `FBSleuth`, for FBS crime evidence verification. Despite the challenges of the ambient noise insensitivity, extended working period and irrelevance of content of FBS fingerprinting, we extract features elaborately from the modulation domain and waveform domain of the raw RF signals to characterize hardware differences of FBS.
- We evaluate `FBSleuth` using 6 real-world FBS devices. Extensive experiments over five months on precision, consistency, and stability of our RF fingerprint demonstrate the high accuracy and robustness of `FBSleuth`.

2 BACKGROUND

To investigate the characteristics of FBSes, we first explore the communication protocol of FBSes, i.e., GSM. The signal emitting of FBS follows the GSM 04.08 specification [26]. Here, we briefly introduce the structure and the modulation errors of GSM bursts and the structure of FBS device.

2.1 Structure of GSM Bursts

The GSM cellular network utilizes Frequency Division Multiple Access (FDMA) and Time Division Multiple Access (TDMA) technologies. For the GSM900 band as an example, the total 25 MHz downlink bandwidth ranging from 935 MHz to 960 MHz is divided into 124 carrier frequency points with 200 kHz bandwidth. Each base station is allocated with one or multiple frequency points.

At one frequency point, time is divided into eight Timeslots (T0-T7) as shown in Fig. 2 (a). In GSM network, the Timeslot at a frequency point is a physical channel on which the logical channels are mapped. A GSM burst is the subcomponent of Timeslot and each one lasts for 577us.

There are mainly three kinds of burst structure in the GSM downlink (shown in the Fig. 2 (b)), FCH Bursts, SCH Bursts, and Normal Bursts. An FCH Burst is used for frequency correction and can be regarded as a pure sinusoid wave. An SCH Burst is used for synchronization with 64 training bits. A Normal Burst (NB) is the most common burst in GSM scheme which is used to transmit SMS messages, system information, control signaling, and voices. Each NB contains Training Code Sequence (TSC) of 26 bits. To extract stable fingerprints, we exploit the constant part of GSM bursts and find that TSC in Normal Bursts is a good candidate.

2.2 Modulation Errors of GSM Bursts

A GSM signal can be decomposed into I component and Q component and thus forms a signal vector (shown in Fig. 3). $\boldsymbol{R}$ represents the reference signal vector (aka. the ideal modulated signal vector) and $\boldsymbol{Z}$ represents the signal measured in practice. $\boldsymbol{E}$ represents the error signal vector that is the subtraction of $\boldsymbol{Z}$ and $\boldsymbol{R}$. Modulation errors include Error Vector Magnitude, Magnitude Error, Phase Error, Frequency Error, IQ Offset, Quadrature Skew Error, IQ Imbalance.

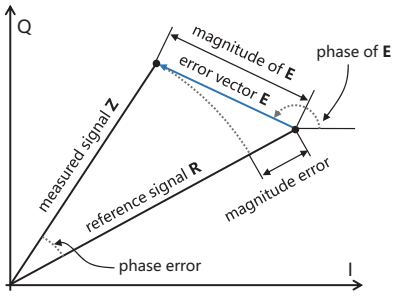


Figure 3: Definition of Modulation Errors.

(a) Error Vector Magnitude (EVM.) is the ratio of the root mean square (RMS) value of average square of magnitude of $\boldsymbol{E}$ and the RMS value of the average power (i.e., square of magnitude) of $\boldsymbol{R}$.

$$\text{EVM}(\%) = \frac{RMS(P_E)}{RMS(P_R)} \times 100\% \tag{1}$$

where RMS is the root mean square, P_E is the average power of $\boldsymbol{E}$ and P_R is the average power of $\boldsymbol{R}$.
(b) Magnitude Error (Mag Err.) is the ratio of the RMS value of the magnitude of $\boldsymbol{E}$ and the RMS value of the magnitude of $\boldsymbol{R}$

$$\text{Mag Err}(\%) = \frac{RMS(Mag_R - Mag_Z)}{RMS(Mag_R)} \times 100\% \tag{2}$$

where Mag_Z is the magnitude of $\boldsymbol{Z}$ and Mag_R is the magnitude of $\boldsymbol{R}$.
(c) Phase Error (Phase Err.) is the intersection angle of $\boldsymbol{Z}$ and $\boldsymbol{R}$.
(d) Frequency Error (Freq Err.) is the frequency offset of the demodulated burst from the center frequency and is averaged over all symbols in the burst.

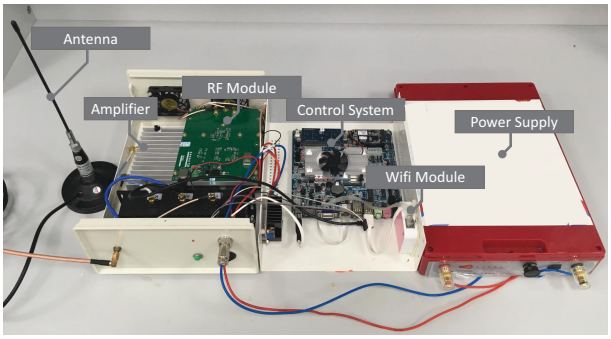


Figure 4: The structure of FBS device (provided by law enforcement agency). FBSes are manufactured by private workshops and sold on the black market illegally.

(e) IQ Offset is the ratio of the power at the center frequency to the average signal power, and indicates the magnitude of the carrier feedthrough signal. When there is no carrier feedthrough, the IQ offset is zero.
(f) Quadrature Skew Error (Quad Err.) indicates the angle error between the I and Q components. Ideally, I and Q should be orthogonal (90 degrees apart). A quadrature skew error of -3 degrees means I and Q are 87 degrees apart.
(g) IQ Imbalance (IQ Imb.) compares the imbalance of the I component with the Q component.

Summary: The aforementioned modulation errors are closely related to the hardware imperfection of GSM devices and can constitute FBS fingerprints.

2.3 Operations on FBS in the crime

Fig. 4 shows the hardware architecture of an FBS, which is mainly composed of a transceiver and a control system. The transceiver is responsible for transmitting the GSM signals and contains an RF module, an amplifier, and an antenna. The control system takes commands from user interface and sets the frequency point, power level, a content of FBS messages, and pattern parameters, etc. When an attacker commits a crime, he has to use the broadcasting channel (BCCH Bursts in T0 Timeslot) at a frequency point. The FBS sends system information on the T0 timeslot to disconnect nearby devices from normal base stations and force them to reconnect to the FBS. Then, the FBS will send FBS messages to the victim device. Note that an FBS may use various frequency points and power levels and it is worth evaluating the feasibility of fingerprinting FBS with various setups.

3 PROBLEM OVERVIEW

In this section, we describe the forensics problem and the threat model.

3.1 Forensics

As the critical step to fight against FBS crimes, forensics for FBS is to collect evidence that can prove the FBS crimes are committed by specific FBS devices. Forensics mainly contains:

- **Source verification**. `FBSleuth` should be able to determine the originality of the collected wireless signals, i.e. which FBSes out of the candidate FBSes commit the crime. For

example, once an FBS is caught, FBSleuth should identify this signal sent by it by comparing the RF fingerprints from the collected signals.

- **Content and quantity verification**. FBSleuth should record the content and the exact number of messages sent by an FBS. This can be achieved with standard signal demodulation technique and is not the focus of this paper.

3.2 Threat Model

According to the operation of FBS, we make the following assumptions on the attackers.

- **High transmitting power.** The FBS is usually operated with a high transmitting power to "kidnap" the victim devices from legitimate base stations. Luckily, the high transmitting power makes it possible to collect FBS's signals from a long distance.
- **Changing frequency point.** An attacker can set the FBS to work at various working frequency points over time. Usually, an attacker will choose the frequency point that is used by the legitimate base station with a weak receiving signal strength.
- **Mobility.** An attacker typically moves the FBS on an automobile and drives around a city to affect as many victim devices as possible.
- **Software manipulation.** The attacker can manipulate the software in an FBS, including the log files, the message contents, and etc.
- **Hardware consistency.** We assume that the attacker won't change the hardware in an FBS unless the devices fail to function. In practice, for the sake of stability and performance, an attacker has no incentive to change the hardware of an FBS.

In summary, the goal of FBSleuth is to achieve high-accuracy forensics by utilizing the RF fingerprints. Specifically, FBSleuth should achieve high precision in fingerprint matching, consistency of the fingerprints from the same individual FBSes and stability under various conditions.

4 FEASIBILITY OF FINGERPRINTING FBS

RF fingerprints are stem from the inherent hardware imperfections or low-grade electronics in analog components. These imperfections are typically introduced at manufacture and assembly stages of analog components of the RF-front-end. The RF-front-end of FBS is a typical superheterodyne transmitter (shown in Fig. 5). In the circuitry of FBS, the IQ quadrature signal is digitized by the DAC (digital to analog converter) and then passes through quadrature modulator, several mixers, and filters for up-conversion and amplification at each stage. Virtually all components on the signal path of the RF-front-end contribute to the distortion of transmitted signals. As shown in Fig. 5, we summarize the distortions associated with the corresponding hardware sources in accordance with the previous work of RF fingerprint [17]. Firstly, when I component and Q component of signals undergo quadrature modulation, quadrature errors and IQ offsets will be introduced to the signals. Then the signals will suffer from self-interference when passing through the filters. Finally, an oscillator will introduce frequency errors, and phase errors to the signals when up-conversion. To verify the

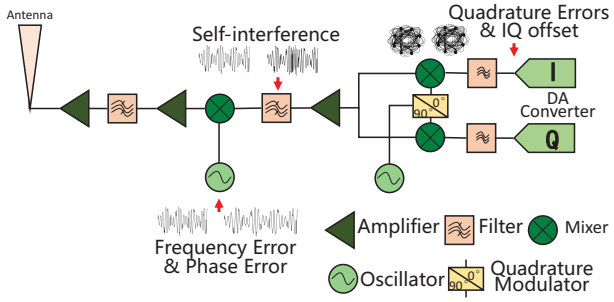


Figure 5: Structure of Superheterodyne Transmitter.

aforementioned signal deviations, we conduct an experiment to analyze the modulation errors of FBS signals.

Experiment Setup. We measured 6 FBSes by a Keysight N9020B spectrum analyzer with VERT900 antenna in the lab environment. The FBSes are powered by a 12 Volt lithium battery to emulate the real scenes. The central frequency is set to 938.2 MHz. The analyzing software VSA89600 inside the spectrum analyzer is used to record the captured signals and calculate the modulation error metrics of each burst. For each FBS, we record 500 sets of modulation errors and calculate the mean values of them.

Experiment Results. We analyzed the errors defined in chapter 2 of the signals collected over the air from the 6 FBSes and two legitimate operator base stations at the same frequency, i.e., 938.2 MHz. Table 1 shows the comparison between legitimate base stations and FBSes on the average modulation errors and their corresponding variances. We found that FBSes have larger modulation errors and variances than the legitimate base stations. Moreover, the differences of the modulation error metrics among FBSes can be discriminated intuitively, which confirms the distance of the deviations among various FBSes.

Due to cost constrains and manufacture limitations, an FBS is often assembled with low-precision oscillators, mixers, and amplifiers. The hardware imperfections of these components even within the same model, introduce various RF characteristics. The variety of RF characteristics can be represented by deviation with respect to ideal signals, e.g, modulation error metrics. Therefore,we constituted FBS fingerprints based on modulation errors, details in the design part.

5 FBSLUETH DESIGN

In this section, we present the design details of FBSleuth.

5.1 Overview

FBSleuth is composed of five modules: raw signal collection, signal processing, evidence database, fingerprint generation, and verification (as shown in Fig. 6). The raw signal collection module is to collect the raw signals from FBSes during crime conduction and after an arrest. The collected signals are marked with both time and location information. The signal processing is divided into modulation domain and waveform domain. Modulation domain processing is to calculate several modulation errors for each burst. Waveform domain processing is to extract bursts (basic processing elements in the RF signals) from the raw signals, select target

Table 1: Modulation Errors of FBS and NBS.

	EVM. (%rms)	Mag Err. (%rms)	Phase Err. (deg)	Freq Err. (Hz)	IQ offset (dB)	Quad Err. (deg)	IQ Imb. (dB)
FBS1	23.3490 ±1.7039	12.9925 ±1.0953	11.4452 ±1.0953	322.0546 ±35.3439	-32.9298 ±5.6213	0.6111 ±1.8032	0.1080 ±0.2417
FBS2	27.2214 ±1.4140	12.7578 ±0.4761	14.0400 ±0.9230	718.6751 ±31.2604	-32.3480 ±5.5010	1.2345 ±1.9753	0.0968 ±0.3021
FBS3	24.6632 ±1.9229	1.4610 ±1.9455	14.2347 ±1.2168	312.5141 ±12.7166	-34.1611 ±5.5262	0.1013 ±0.7220	-0.0114 ±0.2652
FBS4	25.8515 ±1.8104	1.4882 ± 1.9455	14.9477 ±1.1243	714.2193 ±19.6917	-34.0406 ±5.4761	0.1414 ±0.7305	-0.0075 ±0.2374
FBS5	15.1140 ±3.2316	2.4592 ± 1.5677	8.9703 ±3.1295	229.3430 ±21.4874	-38.2389 ±5.8440	0.1158 ±0.5587	0.0035 ±0.1646
FBS6	12.5660 ±3.7198	2.4813 ± 1.7294	7.5617 ±3.6407	241.5086 ±37.3507	-39.5632 ±5.9562	0.1407 ±0.5356	-0.0049 ±0.1234
NBS1	6.2571 ±0.4936	5.3494 ±0.43120	1.9032 ±0.23426	-42.2547 ±2.4104	-65.0432 ±3.8454	-0.0427 ±0.5928	0.0031 ±0.0793
NBS2	6.0003 ±1.6125	4.3314 ±1.6041	3.2696 ±4.9841	44.6279 ±28.0757	-32.5464 ±1.9719	-0.0164 ±0.8697	0.0010 ±0.1014

* In our modulation error analysis, we found that not all of the bursts were well modulated by FBSes and could not be demodulated definitely. We only recorded the modulation errors of FBS bursts whose EVM was under 30%rms and calculated the average and variance value of the modulation errors.

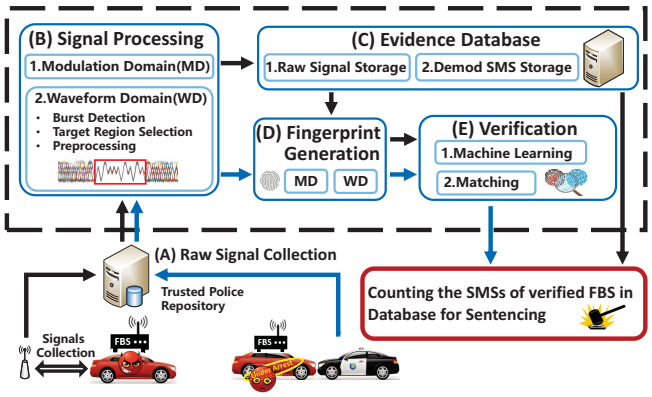


Figure 6: Working Flow of FBSleuth. The black arrow (left) indicates the procedure of processing raw FBS signals collected by the acquisition devices when FBS is committing the crime. The blue arrow (right) indicates the procedure of processing raw FBS signals collected by the police when FBS is caught.

region from the bursts for further fingerprint generation. The evidence database module stores the raw signals, demodulates the FBS signals and records its content, sender and receiver (IMEI for example) information. The fingerprint generation module generates and selects the RF fingerprints from the processed signals both in modulation domain and waveform domain. The verification module utilizes machine learning algorithm to train the model and match the fingerprints with a specific FBS.

With the aforementioned five modules, FBSleuth can help determine when and where the FBS committed crimes before, distinguish two FBSes at one spot and provide the number and content of FBS to the court to assess the accumulated severity of crimes.

5.2 Raw Signal Collection

The signal collection is the first step for FBSleuth, the collection device should have proper sampling rates as well as resolutions for further RF fingerprints generation. In the FBSleuth, we used the USRP N210 equipped with FLEX900 daughterboard as a signal collection device. The USRP N210 is a portable software-defined

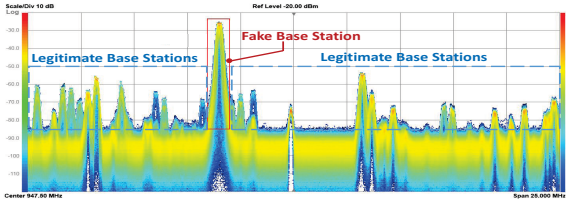


Figure 7: Signal strength of GSM downlink frequency band (from 935 MHz to 960 MHz) under an FBS attack in the urban area. (Measured by Keysight N9020B spectrum Analyzer in the lab environment.)

platform that can collect signals spanning DC-6.0 GHz with tunable 25/50 MHz bandwidth filter. The selected band of collection frequency is first down-converted to an intermediate frequency (IF) and down-converted to near baseband again, sampled by a 14 bit ADC at a rate of 100M samples-per-second. Note that our FBSleuth system can be implemented on devices other than USRP N210, as long as the device performance exceeds the minimum criteria.

A natural question is *how can we distinguish FBS from other legitimate base stations before collecting the signals?* This has been solved by the previous work [19] and we can adopt these methods in FBSleuth for detecting and differentiating FBSes. The signals received by the collection device will be stored as complex in-phase (I) and quadrature (Q) components which preserve amplitude and phase information of the signals.

5.3 Evidence Database

Evidence database is designed to store raw signals with the time/location information, demodulate the signals to extract and store FBS messages. There are also many ways to recover the fraud messages from the recorded raw signals, such as OsmocomBB open source program [2] which implements the GSM protocol stack's three lowest OSI Layers and can be used to parse the FBS fraud messages from the physical layer signals. We implemented the evidence database on top of the OsmocomBB open source program.

5.4 Signal Processing

Filtering. From the received signals shown in (Fig. 7), we can find that the signal frequency points from FBSes are close to the legitimate one in practice. In order to extract the FBS signals from

coupled signals, we designed a six-order complex low pass filter with 250 KHz cut-off frequency to make sure that we collected enough sidelobe information of the FBS signals as well as avoided the interference from adjacent channels.

Modulation Errors Calculation. Recall that in the last chapter, we compared average modulation errors among FBSes and found them closely related to the intrinsic hardware imperfection of FBS. Therefore, we obtain *Error Vector Magnitude, Magnitude Error, Phase Error, Frequency Error, IQ Offset, Quadrature Skew Error, and IQ Imbalance* of each burst. These modulation errors will be evaluated in the next fingerprint generation step.

Burst Detection. In the waveform domain, FCH bursts provide us with calibration for burst detection. Since each FCH burst can be regarded as 37 periods of a pure sinusoid wave, we can generate a standard 37-period sinusoid wave and calculate the correlation to detect the FCH bursts in the raw signals. We leveraged the frame structure sequence of GSM and length of each burst to locate the start points of bursts.

Target Region Selection. To find stable and unique fingerprints, our first mission is to collect a signal sequence that contains enough information to verify different FBSes. As modulation errors represent the average deviation of the whole burst, different content fields of bursts may negatively influence the results. To eliminate the influence of contents of different GSM bursts, we proposed to use the Training Sequence Code (TSC) in Normal Bursts since the contents in TSC bursts are identical. TSC locates in the mid-amble region of Normal Bursts. According to GSM specification [7], there are only 8 kinds of TSC in GSM networks (coded with 3 bits, 000-111). More specifically, the TSC is identical to the value of 3-bit Base station Color Code (BCC), which is constant after a base station starts its service. For an FBS, an attacker cannot change its TSC via software interface in practice. Fig. 8 shows that the instantaneous phases and frequency traces of TSC are almost the same for the Normal Bursts belonging to one FBS. However, identifying an FBS only by TSC is not reliable since different FBSes might have the same TSC because of the limited types of TSC (only 8). In order to be robust, `FBSleuth` should be able to differentiate two FBSes of the same TSC type.

Pre-processing. Instead of extracting features from raw I-Q complex TSC traces directly, `FBSleuth` pre-processes the traces to obtain two intermediate traces: an instantaneous phase trace ϕ and an instantaneous frequency trace f. These traces of the complex signal $s(n) = s_I(n) + js_Q(n)$ are given by:

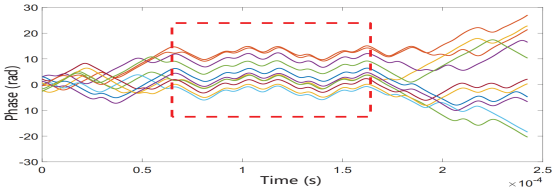
$$\phi(n) = \arctan \frac{s_Q(n)}{s_I(n)} \tag{3}$$

$$f(n) = \frac{1}{2\pi} \frac{\phi(n+1) - \phi(n)}{\Delta t} \tag{4}$$

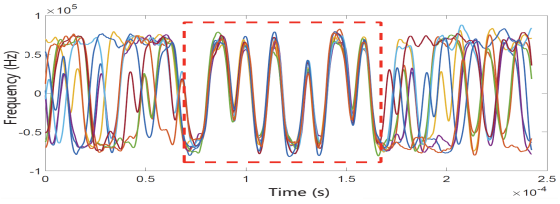
where $n = 1, 2, 3, \ldots, N$, and N is the total number of samples. Δt is the sample interval. Then these traces are centered and normalized prior to fingerprint generation. These traces in Equation (3), (4) are centered into ϕ_c, f_c using:

$$\phi_c(n) = \phi(n) - u_\phi \tag{5}$$

$$f_c(n) = f(n) - u_f \tag{6}$$



(a)Instantaneous phase of TSC in Normal Bursts



(b)Instantaneous frequency of TSC in Normal Bursts

Figure 8: Instantaneous phase (a) and frequency (b) traces of TSC (in the rectangle) in Normal Bursts from one FBS. The phase and frequency patterns of TSC are consistent across time/ bursts.

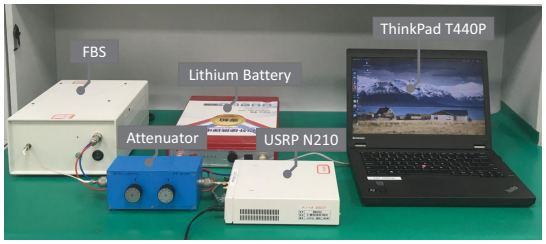


Figure 9: Lab Environment Experiment Setup.

where u_ϕ, u_f are the means of Equation (3),(4) across N samples. The final centered and normalized traces ϕ_{cn}, f_{cn} are given by:

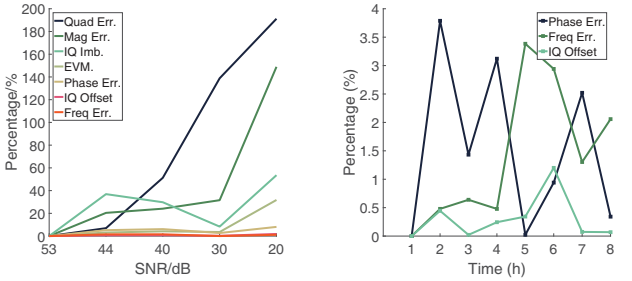
$$\phi_{cn}(n) = \frac{f_c(n)}{max|f_c(n)|} \tag{7}$$

$$f_{cn}(n) = \frac{A_c(n)}{max|A_c(n)|} \tag{8}$$

5.5 Fingerprint Generation

In the fingerprint generation module, we extracted the features in both the modulation domain and the waveform domain to constitute the RF fingerprint of FBS.

Modulation Domain. Considering the stability and consistency requirement of RF fingerprint, the extracted features of FBS signals should be insensitive to the dynamic ambient noise and be consistent during all the working period. Therefore, the stability and consistency of aforementioned seven modulation errors should be evaluated before adopted as fingerprint features. We collected 500 bursts of randomly selected FBS with USRP N210 at 938.2MHz. Then we added Additive White Gaussian Noise to the recorded traces with the signal to noise ratio (SNR) ranging from 20 dB to 53 dB and calculated the $Metrics_{i,j}$, where i represents EVM., Mag Err., Phase Err., Freq Err., Quad Err., IQ Imb. and j represents 20dB, 30dB, 40dB, 44dB, and 53dB. The Relative Metric Change is designed to reflect the fluctuation of modulation error in different ambient



(a) (b)

Figure 10: (a) describes the relative change of different modulation error metrics under different SNR levels. Freq Err., IQ offset, and Phase Err. are more stable than Quad Err., Mag Err., and IQ Imb. under the dynamic ambient noise; (b) describes the relative change of Freq Err., IQ offset, and Phase Err. at different time of FBS working period are no more than 4%.

noise and the formula is as follows.

$$\text{Rel.Change}_{\text{i,j}} = \frac{Metrics_{i,j} - Metrics_{i,53dB}}{Metrics_{i,53dB}} \quad (9)$$

Fig 10 (a) illustrates the relative change of different modulation error metrics with SNR. The Quad Err. and Mag Err. increase rapidly with SNR decreases, which indicates they are sensitive to the dynamic ambient noise. However, the Freq Err., IQ offset, and Phase Err. keep stable with ambient noise and distinct between FBSes. To further evaluate the consistency of Freq Err., IQ offset, and Phase Err., we kept the FBS working continuously, collected 500 bursts and calculated the modulation errors for each hour. Fig 10 (b) shows that the relative change of Freq Err., IQ offset, and Phase Err. overtime are no more than 4%. This indicates the uniqueness, stability, and consistency of Freq Err., IQ offset, and Phase Err. .

Waveform Domain. Although features in modulation domain are typically more robust [50], these metrics are calculated based on the whole burst length and influenced negatively by the content fields of GSM bursts. Thus fingerprinting FBS only with standard modulation error metrics is not enough. Alternatively, instantaneous phases and frequencies of TSC depict the signal characteristics of each FBS and contain rich phase and frequency information of FBS within a short interval. Therefore, We conducted fast Fourier transformation and discrete wavelet transform on each instantaneous trace and obtain their corresponding frequency domain and wavelet domain traces. Thus we obtained 6 final traces in total and a time-domain trace.

We extracted features from modulation errors, i.e., *Frequency Error, Phase Error, and IQ Offset* of each burst and then extracted waveform domain features from 6 traces of TSC in the same burst including *Mean, Standard Deviation, Skewness, Kurtosis, Median, Maximum, Variance, Shannon Entropy* (shown in Tab. 2). Therefore a total of 51 features are available for constituting one fingerprint of FBS.

5.6 Verification

Machine Learning Algorithm. Recall that FBSleuth matches the fingerprints of the arrested FBS with the fingerprints of FBS collected in the wild. To achieve this, we utilized supervised learning to classify each fingerprint. We compared 11 representative classifiers and chose Support Vector Machine (SVM) as the default classification method. SVM separates the labeled set in two areas on a multi-dimensional surface. It is of high efficiency, good accuracy, and robustness against outliers and is less prone to overfitting than other methods.

Table 2: Selected FBS features.

Feature Domain		Feature Name
Modulation Error Metrics	Freq Err.	Frequency Error
	Phase Err.	Phase Error
	IQ Offset	IQ Offset
Instantaneous Phase Metrics	Time.	Mean, Standard Deviation,
	Wavelet.	Skewness, Kurtosis, Median,
	FFT.	Maximum, Variance, Shannon Entropy
Instantaneous Frequency Metrics	Time.	Mean, Standard Deviation,
	Wavelet.	Skewness, Kurtosis, Median,
	FFT.	Maximum, Variance, Shannon Entropy

We utilized n bursts and their corresponding raw TSC traces (each trace lasts 96 *us*) for one FBS to generate fingerprints and train a binary classifier it. For k FBSes in our fingerprint database, we have to train k binary classifiers with $k \times n$ traces.

Matching. FBSleuth collects raw traces from an FBS and extracts their fingerprints. Then FBSleuth tries to match the extracted fingerprint with the existing FBSes in the database. During the forensic stage (i.e., collecting evidence), if a match is found in the database, then newly recorded FBS messages, its time and location and inserted into the database. If none of the existing FBSes matches the new fingerprint, it means the new fingerprint is from a new FBS. Then, a new binary classifier should be trained and incorporated into the system. To be specific, the original k binary-classifier system should be extended to $k+1$. To distinguish a new FBS from the known FBSes, we apply a threshold for the probability score - if the probability score is less than the threshold, the tested FBS should be declared as a new FBS and vice versa. After an FBS is caught by the police, the associated records in the evidence database can be submitted to the court for further trial and sentencing.

6 FBSLEUTH EVALUATION

To evaluate FBSleuth, we conducted extensive experiments with 6 real FBSes provided by the law enforcement agencies of four cities. Particularly, according to the law enforcement agencies, FBS1, FBS2, and FBS3 were manufactured by the same vendor in Zhengzhou, FBS4 was captured in Hangzhou, FBS5 was captured in Shanghai, FBS6 was captured and manufactured in Guangzhou. The various sources of FBSes indicate that they are likely manufactured by different vendors. The key questions we investigated are summarized below:

- What's the performance of FBSleuth in terms of precision, stability, and consistency?
- What classification algorithm is suitable for FBSleuth and how to set the parameters such as training sample size?
- How sensitive is FBSleuth against the supplied voltage of FBS?

- Can FBSleuth identify individual FBS across various working frequency points?
- What's the minimum SNR level that FBSleuth can keep working?
- Does FBSleuth depend on the receivers that record the raw signal?

In summary, the performance of FBSleuth is:

- FBSleuth has over 99% precision, 96.5% recall, and 97.94% F1 score both in the lab and wild environment.
- FBSleuth can maintain relatively good performance with little influence of working frequency point, battery volume, and SNR.
- The portable and low-cost signal acquisition device of FBSleuth make it possible to be deployed in the real world.

6.1 Experiment Setup

FBS Setup. The maximum transmitting power of the 6 FBSes is 50 Watt. These FBSes can work in the whole GSM900 frequency range (935MHz to 960MHz for downlink). The FBSes are connected with a laptop for setting the working mode and parameters such as working frequency configuration.

Collection Device Setup. We used USRP N210 equipped with an RFX900 daughterboard as the collection device. In the wild environments, the USRP N210 is equipped with a VERT900 antenna for signal collection. The USRP is driven by compatible software Gnuradio in Linux platform. We tuned the sample rate to be 25MHz and the RX gain to be 15 in all the following experiments. Additionally, two other USRP N210 devices are tested for evaluating the influence of different receivers.

Power Supply Setup. We supplied the FBS with a 12 V lithium battery of 110 Ah volume and this kind of batteries are widely used in the FBS crimes where criminals are moving around in a city.

Laptop Setup. We utilized two ThinkPad T440p model laptops, with Intel i5 4200M CPU, 4G RAM, and Intel 7260 BGN wireless network adapter in the experiments. One laptop is utilized for connecting to FBS with Wi-Fi and controlling FBS through the user interface. The other one is used for raw signal storage and signal processing.

In the lab environment, the FBS was connected with the USRP over the SMA cable. We added an attenuator between FBS and USRP to adjust the gain of the signal. The signal gain of FBS is set to -10 dBm in the lab experiments. The USRP is connected to the laptop via Ethernet. While in the wild environment, we utilized a USRP N210 equipped with a VERT900 antenna as a receiver. FBSs are powered with a battery carried on a trolley. We conducted extensive experiments and evaluation on 6 real FBSes and the collected data lasts for 5 months. For each experiment, we collected at least 20,000 traces for each FBS at each working frequency point (21 frequency points from total 124 frequency points). To eliminate errors, we randomly chose 1200 traces for each FBS at one working frequency point and calculate the average result over three rounds of the test.

6.2 Metrics and Classifiers

Performance Metrics. Let n be the total number of FBSes and therefore we have n classes. Given a new raw TSC trace from an unknown FBS, FBSleuth identifies whether it is one of the FBSes. We calculate the *true positive (TP_i)* for each classifier i, which measures the cases that an FBS is classified correctly. Similarly, we calculate the *false positive (FP_i)* and *false negative (FN_i)* as the rate of wrongly accepted and wrongly rejected cases for each classifier C_i $(1 \leqslant i \leqslant n)$. We use standard classification metrics: *precision*, *recall*, and *F1 Score* in our evaluation analysis [14]. The standard classification metrics *precision*, *recall*, and *F1 Score* are defined as follows.

$$Precision_i = \frac{TP_i}{TP_i + FP_i}$$

$$Recall_i = \frac{TP_i}{TP_i + FN_i}$$

$$F1\ Score_i = \frac{2 \times Precision_i \times Recall_i}{Precision_i + Recall_i}$$

The average *precision*, *recall*, and *F1 Score* are defined as the average over the n results.

To evaluate the performance of FBSleuth in the presence of new FBS (not in the training set), we chose accuracy as the metric. Given that the classifier trained by n classes and tested by n old classes and m new FBS classes, the accuracy is defined as below.

$$Accuracy = \frac{\sum_{i=1}^{n} TP_i + \sum_{j=1}^{m} TN_j}{n + m}$$

Where TP_i is the true positive for class i and TN_j is the true negative for new FBS class j.

Influence of Classifier. To investigate the influence caused by the classifier selection, we compared 11 most commonly used supervised learning algorithms, including 1) *Logistical Regression* 2) *Support Vector Machine* 3) *Random Forest Classifier* 4) *Gradient Boosting Classifier* 5) *Gaussian Naive Bayes* 6) *AdaBoosting Classifier* 7) *Decision Tree Classifier* 8) *Extra Trees Classifier* 9) *Linear Discriminant Analysis* 10) *Quadratic Discriminant Analysis* 11) *KNeighbors Classifier*. We utilized the default threshold and employed 10-fold cross validation to have a basic understanding which algorithm behaves best.

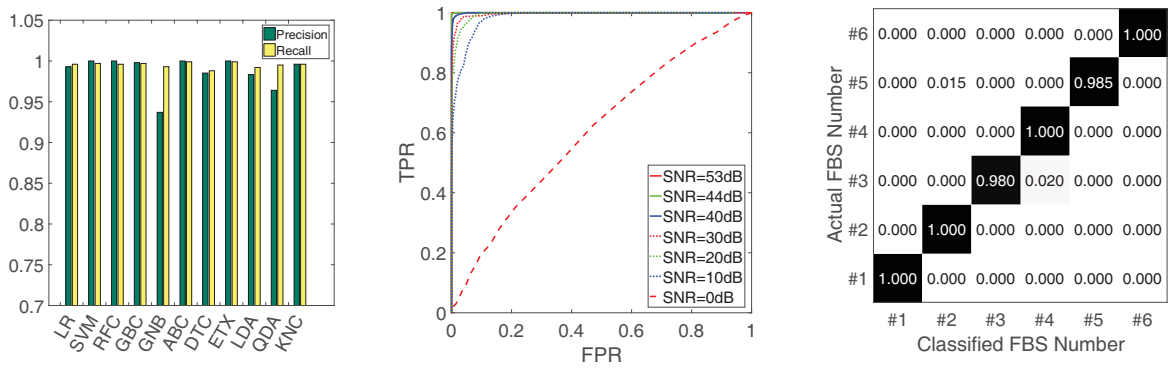
For each classifier, we chose 1200 traces (1000 for training and 200 for testing each time) at $945MHz$ frequency for each FBS, that is 7200 traces in total. We then fed them into the 11 aforementioned classifiers separately. The results in Fig. 11(a) show that Gradient Boosting Classifier, Extra Trees Classifier, and Support Vector Machine are the top 3 classifiers in term of precision, recall, and F1 score. Therefore, we chose them as our candidate classifiers.

Then we evaluated the accuracy of each classifier. Each time, we chose 5 from 6 FBSes for training and all 6 FBSes for testing and calculated the average accuracy in all 5 rounds for each classifier. The maximum accuracy of Extra Trees Classifier was 0.852 with 0.900 as the threshold. The accuracy of Gradient Boosting Classifier was 0.830 with a 0.994 threshold while it could achieve an accuracy of 0.971 with a threshold of 0.84 for Support Vector Machine.

In conclusion, Support Vector Machine achieved good performance in precision, recall, and accuracy as well as its reasonable computation overhead. Therefore in the following experiments, we accepted Support Vector Machine (SVM) with 0.84 threshold as our classifier algorithm used in FBSleuth.

6.3 Overall Performance

ROC in different SNR. To evaluate the performance of the classifier SVM, we calculated the receiver operating characteristics curve, i.e. ROC curve. A ROC curve is a graphical plot that illustrates the



(a) Performance of Classifiers (b) ROC of Various SNR (c) Confusion Matrix of 6 FBSes

Figure 11: (a) describes the precision and recall of 11 common supervised learning algorithms; (b) describes the ROC curve of SVM classifier under different SNR conditions; (c) describes the confusion matrix of 6 FBSes.

diagnostic ability of a binary classifier system with the change of its discrimination threshold. We calculated the true positive rate and false positive rate of SVM with different thresholds and plotted the ROC curve in different SNRs. We collected raw traces under 40dB SNR at 945MHz frequency point and randomly selected 1200 traces from each of FBS, 1000 for training and 200 for testing. Fig 11(b) shows that the ROC curve goes towards the diagonal as the SNR decreases. But even at 10 dB SNR, the classifier can still perform with high precision.

Confusion Matrix. For each confusion matrix, each row of the matrix represents the instances in a predicted class while each column represents the instances in an actual class (or vice versa). We collected raw trace under 40dB SNR at 945MHz frequency point and randomly selected 1200 traces from each of FBS, 1000 for training and 200 for testing. The resulting average classification score for each FBS is shown in the confusion matrix in Fig. 11(c). All 6 FBSes were almost always classified correctly. Even FBS1, FBS2, and FBS3, which were manufactured by the same vendor, can be classified without any misclassification between themselves. In sum, we can conclude that an FBS can be identified with high precision and can rarely be misclassified with each other.

6.4 Micro-benchmark Evaluation

(a) Working Frequency Points. In FBS crimes, the attacker can change the frequency point frequently. For this reason, it is important to identify the FBSes in different working frequency points. To evaluate the effect of different working frequency points, we tested 5 different FBSes and set the FBSes to work at 935MHz, 941MHz, 947MHz, 953MHz, and 959MHz respectively which cover the high, medium, and low frequency with the default settings. Fig. 12 shows that `FBSleuth` can keep over 99% precision, recall, and F1 Score both in the condition the training and testing samples from the same working frequency point and from different working frequency points. It indicates that even attackers change working frequency frequently, `FBSleuth` can still identify them in different working frequencies with high precision.

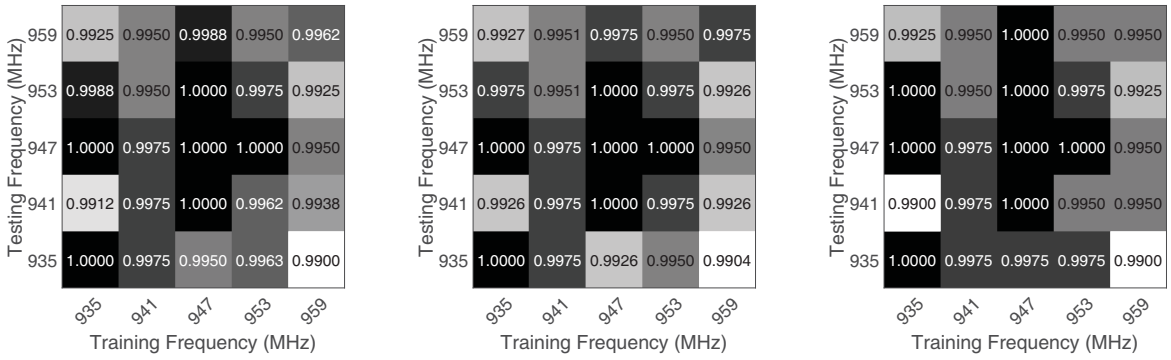
(b) Training Sample. To evaluate the influence of training sample size, we analyzed 7200 traces from 6 FBSes (1200 traces from each) at 947MHz working frequency point under 40dB SNR. We varied the training sample size from 10 to 1000 (samples). Fig. 13(a) shows that as the number of training sample increase, the precision keeps over 99% and the average recall and F1 score all increase. We can get over 99.7% precision as well as 98.6% recall when the trained sample size is 800. This suggests that we can construct a good fingerprinting classifier without requiring a large training sample size.

(c) Ambient Noise. In practice, the SNR of FBS signals varies as an attacker may tune the power amplifier or the distance between the FBS and the receiver changes with time. So it is important that the FBS fingerprint we extracted are stable against the changes of SNR. In addition, the lowest SNR that our fingerprint method can tolerate determine how far away can our method identify the individual FBS.

To evaluate the effects of SNR, we manually added white Gaussian noise on the collected signals. We tested the 6 FBSes under 6 SNR levels. At each SNR level, we collected 1200 traces at 936MHz, 945MHz, and 955MHz respectively and selected 1000 for training and 200 for testing. Then we changed the SNR level and repeated the test. The results are shown in Fig. 13(b). With the increase of SNR, the average precision, recall, and F1 score improve. This is because of the higher the SNR, the less influence from environmental noise on the FBS signals and thus we can obtain more clean features. Fig. 13(b) also shows that the average precision and recall remains over 90% even at 20dB SNR. In real cases, the signal power of FBS can be over 100 Watt with large SNR, therefore we believe that `FBSleuth` can perform well in practice.

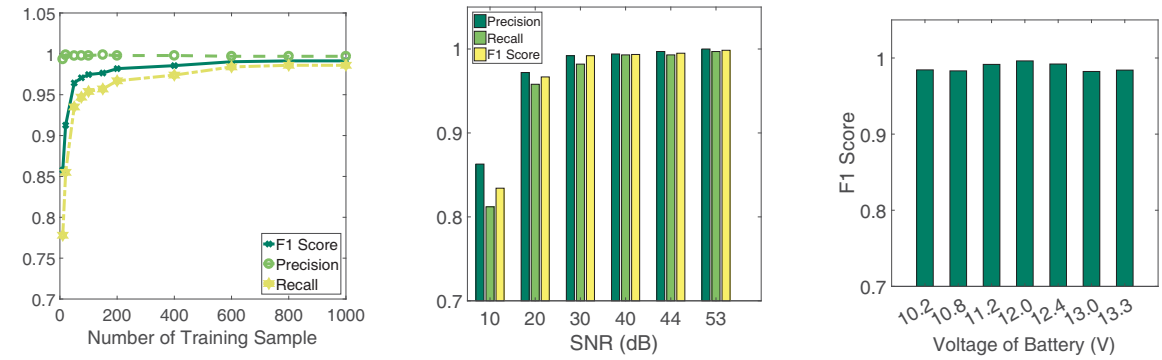
(d) Battery Voltage. In real FBS crimes, the FBS is always supplied by batteries for mobility. The remaining voltage of the battery may have the influence upon the performance of `FBSleuth`. Given the fact that FBS often works continuously for hours, meaning that the remaining voltage will change with time. We collected 1000 traces under the standard 12 V as training set and collect 1000 traces from each voltage ranging from 13.30V (100% charged state) to 10.21V (20% charged state) at 945MHz frequency point for each FBS. Fig. 13(c) shows that variance of power has little effect on the accuracy.

(e) Impact of Receiver. In practice, several receivers will be implemented in the different areas to capture the signals from one



(a) F1 Score (b) Precision (c) Recall

Figure 12: Impact of varying frequency points on the performance of FBSleuth: (a) describes the influence on F1 score; (b) describes the influence on precision; (c) describes the influence on recall.



(a) (b) (c)

Figure 13: (a) describes the impact of varying number of training sample on the performance of FBSleuth; (b) describes the impact of varying SNR on the performance of FBSleuth; (c) describes the impact of varying voltage of battery on the performance of FBSleuth.

FBS in the different time. Therefore, it is crucial to evaluate the portability of receivers. Three USRP N210 devices (U1, U2, U3) have been used for evaluation in total. We collected signals from 6 FBSes with three USRPs separately at 947MHz with 40dB SNR and used signals received from one USRP for training and other USRPs for testing in turns. Experiment results show that there's no evident difference in the performance of different receivers.

6.5 Wild Implementation and Evaluation

We implemented FBSleuth in the wild environment and evaluated its performance. In Fig. 14, the star represents the initial location of the FBS and the dot stands for the receiver, i.e. the USRP. During the experiment, the USRP is fixed and the FBS can move anywhere in the rectangle area. The distance between the USRP and the FBS varies from 20 m to 210 m and the SNR varies from roughly 20dB to 60dB. The transmitting power of the FBS can be tuned as well in the experiment.

In this test, we set 5 different working frequency points for each FBS. We randomly selected 1200 traces in 10 minutes' record, 1000 traces for training and 200 traces for testing at each frequency point for each FBS. Fig. 15 shows that FBSleuth identifies FBSes with over 99% average precision, 96.4% average recall, and 97.94% average F1 score in the wild environment among different working frequency points. Although the F1 score performance drops a little compared with the lab experiment, it still keeps high precision. This indicates that FBSleuth can identify FBSes in the wild effectively.

7 RELATED WORK

Wireless crime. The feasibility of wireless crimes with illegal transceivers to impersonate legitimate radio frequency devices or infrastructures has been studied by many researchers. Xenakis et al. [12, 48, 49] analyzed the vulnerabilities in cellular networks and discussed the feasibility of Man-in-the-middle attacks. Mjolsness et al. [29] showed that even the state-of-art LTE (4G) networks can suffer IMSI catcher attacks. Perez et al. [33] demonstrated that an attacker with a budget of less than 10,000 $ can set up an FBS to launch a practical attack to intercept 2G/3G mobile data. The implementation of FBS attack has been further studied [40, 45] and the attacks in real world have been reported in many countries [1, 3–5, 37].

Countermeasures. To fight against FBS crimes, detecting, and tracking illegal transceivers have been widely studied [8–11, 27]. To protect cellular networks from illegal transceiver attacks, much effort has been put on detecting and tracking the illegal transceivers.

Figure 14: Wild Environment Experiment Setup.

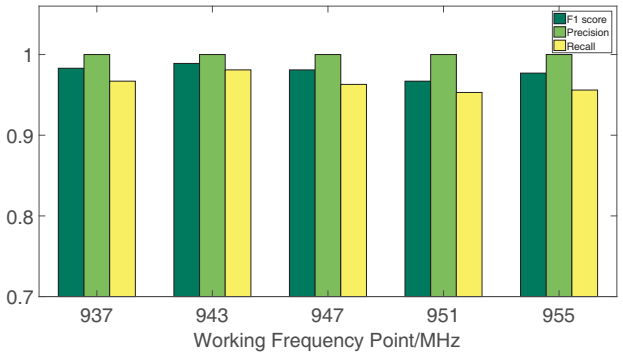


Figure 15: The overall performance of FBSleuth in different working frequency points.

Karsten et al. designed a smartphone application to detect IMSI catchers [9]. Li et al. implemented an FBS detection system named RBS-Radar, which can detect and locate the FBSes in the wild [27]. However, all the above work ignore the forensics for FBS crimes.

RF fingerprinting. The definition of RF fingerprinting was first proposed by Hall et al. at 2003 [20]. They extracted fingerprints from the Bluetooth signals and used it to identify the Bluetooth devices. The key of RF fingerprinting is to extract the fingerprint from wireless signals, and the extraction can be from both transient region and steady region. Transient is the part of the signal that can be observed when the amplitude of the transmitter rises from background noise to the level required for data communication [16, 21, 22, 35, 42–44]. However, due to the short duration, i.e., a few microseconds to tens of milliseconds of transient signals, the extremely high sampling rate is required for acceptable identification performance. On the contrary, Kennedy et al. proposed to extract fingerprints from the steady-state period of signals [24]. Brik et al. developed a hardware fingerprinting approach called PARADIS to extract device signature from modulation domain [17]. Their approach can be used to distinguish wireless cards from the same vendor. Nguyen et al. [31] extended Brik's work by employing unsupervised learning techniques, which gets rid of the training process.

Up to now, abundant RF fingerprinting work have been done on various wireless devices, including Bluetooth devices [20], GSM cell phones [36, 47], IEEE 802.16 (WiMAX) devices [46], UMTS cell phones [38], LTE cell phones [13, 30], CRN devices [25], RFID devices [41], IEEE 802.11 Wi-Fi devices [51], and IEEE 802.15.4 ZigBee devices [15, 32].

Motivated by the work above, FBSleuth extracted FBS fingerprint from the steady region of FBS signals. We extracted the features from FBS signals and explained their meanings in nature. What is more, we dealt with a case where FBSes can change their transmitting power, frequency point, working mode and even the FBSes themselves can be moved, all of which make the RF fingerprinting difficult. Moreover, previous works only focus on using RF fingerprint to authenticate the subscriber in cellphone side which is quite different from our research purpose, object, and methodology. Besides modulation errors, we found the identifiable trace for FBS and proposed a new method to extract the identifiable trace, specifically, TSC in Normal Bursts from the raw RBS signals. We also extracted and selected time domain, spectral domain and wavelet domain features from the identifiable traces. Lastly, We employed low-end devices, i.e., USRP to achieve higher accuracy compared with the previous work [36] on GSM handsets with high-end equipment.

8 LIMITATIONS AND DISCUSSION

(1) Hardware Modification. FBS attackers may attempt to escape from our forensics scheme by modifying hardware. For instance, attackers can modify components in the RF modules of an FBS or just use a different FBS each crime. However, hardware modification increases the cost and causes instability of FBSes.

(2) Signal-level Modification. FBS attackers may attempt to change the RF fingerprints at the signal level. For instance, attackers may add noises before emitting the signal to obscure the RF fingerprints. However, inserting noises will reduce the radius of FBS attack ranges. However, such manipulation requires using arbitrary waveform generators, which typically cost 70 times more than an FBS and are cumbersome to carry around due to its power supply requirement (plug with AC 220V in). In addition, attackers may replay the FBS signals to interfere the forensics, but the timestamps of SMSes can help the police reject the replayed signals.

(3) Scalability. Though Fig. 11 (c) illustrates that FBSleuth is able to classify the 6 real FBSes with high accuracy, it is possible that a larger FBS pool results in lower accuracy. However, there are few chances that hundreds of FBSes existing within one area simultaneously in the real world. According to the Skyeye system of Qihoo 360 Technology Co. Ltd, typically no more than twenty FBSes exist in one city in one month [34]. Moreover, to avoid misclassification, we can combine time and location records of the received signals and confirm the identity of an attacker's vehicle from the images of traffic cameras.

9 CONCLUSION

In this paper, we addressed the issue of forensics in combating the serious FBS crimes in the real world. We designed, implemented, and evaluated FBSleuth, a system that identifies FBS devices based on the minor difference in the emitted signals caused by hardware imperfection. To demonstrate the feasibility of FBS crime forensics, we collected signal traces from 6 real FBSes during 5 months. We conducted our experiments in both lab and wild environments and evaluated FBSleuth under various settings. The results show that FBSleuth can successfully identify the six FBSes with over

99% precision and 96.4% recall under dynamic, low SNR wild environments, limited training samples size across different frequency points. FBSleuth can be promising for fighting against FBS crimes.

ACKNOWLEDGMENTS

We thank our shepherd Kevin R. B. Butler for the valuable comments on the manuscript. Zhuoran Ma, Xuan Ouyang, Tianchen Zhang help a lot in developing the system. This work has been funded in part by NSFC 61472358, NSFC 61702451, the High-Tech Research and Development Program of China ("863-China Cloud" Major Program) 2015AA01A201, NSF CNS-0845671, and the Fundamental Research Funds for the Central Universities 2017 QNA4017 and Qihoo 360 Technology Co. Ltd.

REFERENCES

[1] 2015. Fake Stingray mobile base stations discovered spying on millions of Londoners. http://www.ibtimes.co.uk/. (2015).
[2] 2015. OsmocomBB. https://en.wikipedia.org/wiki/OsmocomBB. (2015).
[3] 2016. Mobile Security Reports by Baidu. http://shoujiweishi.baidu.com/safety.html. (2016).
[4] 2016. Mobile Security Reports by Qihoo 360. http://zt.360.cn/2015/reportlist.html?list=1. (2016).
[5] 2016. A Report of Fake Base Stations in China by Tencent, 2016. http://m.qq.com/securitylab/newsdetail361.html. (2016).
[6] 2017. Explanation of several issues concerning the law in criminal cases of disrupting radio communication and management order. http://www.court.gov.cn/zixun-xiangqing-49322.html. (2017).
[7] July 1993. Eur. Telecommun. Standards Inst., GSM 05.05, Ver 4.6.0. *European Digital Cellular Telecommunication System (phase 2); Radio Transmission and Reception* (July 1993).
[8] Dare Abodunrin et al. 2015. Detection and Mitigation methodology for Fake Base Stations Detection on 3G/2G Cellular Networks. (2015).
[9] Alex, Linus, Jakob, and Luca Karsten. 2017. SnoopSnitch project. https://opensource.srlabs.de/projects/snoopsnitch. (2017).
[10] Zhang Chen. 2014. Malicious base station and detecting malicious base station signal. *China Communications* 11, 8 (2014), 59–64.
[11] Alaaedine Chouchane, Slim Rekhis, and Noureddine Boudriga. 2009. *Defending against rogue base station attacks using wavelet based fingerprinting.*
[12] Xenakis Christos and et.al. 2008. A Qualitative Risk Analysis for the GPRS Technology. In *Ieee/ifip International Conference on Embedded and Ubiquitous Computing*. 61–68.
[13] Frederic Demers and Marc St-Hilaire. 2013. Radiometric identification of LTE transmitters. In *GLOBECOM*. IEEE, 4116–4121.
[14] Sanorita Dey, Nirupam Roy, Wenyuan Xu, Romit Roy Choudhury, and Srihari Nelakuditi. 2014. AccelPrint: Imperfections of Accelerometers Make Smartphones Trackable.. In *NDSS*.
[15] Ramsey Benjamin W Temple Michael A Dubendorfer, Clay K. 2012. An RF-DNA verification process for ZigBee networks. In *MILCOM*. IEEE, 1–6.
[16] KJ Ellis and Nur Serinken. 2001. Characteristics of radio transmitter fingerprints. *Radio Science* 36, 4 (2001), 585–597.
[17] Brik Vladimir et.al. 2008. Wireless device identification with radiometric signatures. In *Proceedings of the 14th ACM international conference on Mobile computing and networking*. ACM, 116–127.
[18] Dabrowski Adrian et.al. 2014. IMSI-catch me if you can: IMSI-catcher-catchers. In *Proceedings of the 30th annual computer security applications Conference*. ACM, 246–255.
[19] Du Gang, Zhang Chen, Zhu Yanyun, Du Xuetao, and Meng Dexiang. 2014. Research on mobile Rogue base station locating and tracking method. *2014 National Wireless and Mobile Communication symposium* (2014).
[20] Jeyanthi Hall, Michel Barbeau, and Evangelos Kranakis. 2003. Detection of transient in radio frequency fingerprinting using signal phase. *Wireless and Optical Communications* (2003), 13–18.
[21] Jeyanthi Hall, Michel Barbeau, and Evangelos Kranakis. 2004. Enhancing intrusion detection in wireless networks using radio frequency fingerprinting.. In *Communications, Internet, and Information Technology*. 201–206.
[22] Jeyanthi Hall, Michel Barbeau, and Evangelos Kranakis. 2005. Radio frequency fingerprinting for intrusion detection in wireless networks. *IEEE Transactions on Defendable and Secure Computing* (2005).
[23] ZHAO Jian-qiang. 2016. Research and Practice of "Pseduo Base Station" Digital Forensic. *Computer Science* (2016).
[24] Irwin O Kennedy, Patricia Scanlon, Francis J Mullany, Milind M Buddhikot, Keith E Nolan, and Thomas W Rondeau. 2008. Radio transmitter fingerprinting: A steady state frequency domain approach. In *Vehicular Technology Conference, 2008. VTC 2008-Fall. IEEE 68th*. IEEE, 1–5.
[25] Kyouwoong Kim, Chad M Spooner, Ihsan Akbar, and Jeffrey H Reed. 2008. Specific emitter identification for cognitive radio with application to IEEE 802.11. In *IEEE GLOBECOM 2008*. IEEE, 1–5.
[26] Mobile Radio Interface Layer. 3. Specification, GSM 04.08 v 4.2. 0, Oct. 1992, Bates Nos. *QBB479485-977* (3).
[27] Zhenhua Li, Weiwei Wang, Christo Wilson, Jian Chen, Chen Qian, Taeho Jung, Lan Zhang, Kebin Liu, Xiangyang Li, and Yunhao Liu. 2017. FBS-Radar: Uncovering Fake Base Stations at Scale in the Wild. *Network and Distributed System Security Symposium* (2017).
[28] Karsten Linus. 2013. CatcherCatcher. https://opensource.srlabs.de/projects/mobile-network-assessment-tools/wiki/CatcherCatcher. (2013).
[29] Stig F. Mjolsness and Ruxandra F. Olimid. 2017. Easy 4G/LTE IMSI Catchers for Non-Programmers. In *International Conference on Mathematical Methods, Models, and Architectures for Computer Network Security*. 235–246.
[30] Riaz Mondal, Jussi Turkka, Tapani Ristaniemi, and Tero Henttonen. 2014. Performance evaluation of MDT assisted LTE RF fingerprint framework. In *Mobile Computing and Ubiquitous Networking (ICMU), 2014 Seventh International Conference on*. IEEE, 33–37.
[31] Nam Tuan Nguyen, Guanbo Zheng, Zhu Han, and Rong Zheng. 2011. Device fingerprinting to enhance wireless security using nonparametric Bayesian method. In *INFOCOM, 2011 Proceedings IEEE*. IEEE, 1404–1412.
[32] Hiren J Patel, Michael A Temple, and Rusty O Baldwin. 2015. Improving zigbee device network authentication using ensemble decision tree classifiers with radio frequency distinct native attribute fingerprinting. *IEEE Transactions on Reliability* 64, 1 (2015), 221–233.
[33] David Perez and Jose Pico. 2011. A practical attack against gprs/edge/umts/hspa mobile data communications. *Black Hat DC* (2011).
[34] Qihoo. 2018. 360 Skyeye. https://skyeye.360safe.com/. (2018).
[35] Kasper Bonne Rasmussen and Srdjan Capkun. 2007. Implications of radio fingerprinting on the security of sensor networks. In *Security and Privacy in Communications Networks and the Workshops, 2007. SecureComm 2007. Third International Conference on*. IEEE, 331–340.
[36] Donald R Reising, Michael A Temple, and Michael J Mendenhall. 2010. Improving intra-cellular security using air monitoring with RF fingerprints. In *IEEE WCNC 2010*. IEEE, 1–6.
[37] Mary Ann Russon. 2014. 19 Fake Mobile Base Stations Found Across US. http://www.ibtimes.co.uk/19-fake-mobile-base-stations-found. (2014).
[38] Patricia Scanlon, Irwin O Kennedy, and Yongheng Liu. 2010. Feature extraction approaches to RF fingerprinting for device identification in femtocells. *Bell Labs Technical Journal* 15, 3 (2010), 141–151.
[39] sohu. 2016. Demystifying the Industrial Chain of Fake Base Stations. http://news.sohu.com/20160412/n443925430.shtml. (2016).
[40] Yubo Song, Kan Zhou, and Xi Chen. 2012. Fake BTS Attacks of GSM System on Software Radio Platform. *JNW* 7, 2 (2012), 275–281.
[41] William C Suski II, Michael A Temple, Michael J Mendenhall, and Robert F Mills. 2008. Using spectral fingerprints to improve wireless network security. In *IEEE GLOBECOM 2008*. IEEE, 1–5.
[42] OH Tekbas, Nur Serinken, and O Ureten. 2004. An experimental performance evaluation of a novel radio-transmitter identification system under diverse environmental conditions. *Canadian Journal of Electrical and Computer Engineering* 29, 3 (2004), 203–209.
[43] J Toonstra and Wintold Kinsner. 1995. Transient analysis and genetic algorithms for classification. In *WESCANEX 95. Communications, Power, and Computing. Conference Proceedings., IEEE*, Vol. 2. IEEE, 432–437.
[44] Oktay Ureten and Nur Serinken. 1999. Bayesian detection of radio transmitter turn-on transients.. In *NSIp*. 830–834.
[45] Kenneth van Rijsbergen. 2016. The effectiveness of a homemade IMSI catcher build with YateBTS and a BladeRF. *University of Amsterdam* (2016).
[46] McKay D Williams, Sheldon A Munns, Michael A Temple, and Michael J Mendenhall. 2010. RF-DNA fingerprinting for airport WiMax communications security. In *Network and System Security (NSS), 2010 4th International Conference on*. IEEE, 32–39.
[47] McKay D Williams, Michael A Temple, and Donald R Reising. 2010. Augmenting bit-level network security using physical layer RF-DNA fingerprinting. In *IEEE GLOBECOM 2010)*. IEEE, 1–6.
[48] Christos Xenakis. 2006. Malicious actions against the GPRS technology. *Journal of Computer Virology and Hacking Techniques* 2, 2 (2006), 121–133.
[49] Christos Xenakis. 2008. Security Measures and Weaknesses of the GPRS Security Architecture. *International Journal of Network Security* 6, 2 (2008).
[50] Qiang Xu, Rong Zheng, Walid Saad, and Zhu Han. 2016. Device fingerprinting in wireless networks: Challenges and opportunities. *IEEE Communications Surveys & Tutorials* 18, 1 (2016), 94–104.
[51] Davide Zanetti, Boris Danev, et al. 2010. Physical-layer identification of UHF RFID tags. In *Proceedings of the sixteenth annual international conference on Mobile computing and networking*. ACM, 353–364.