Proximity-based Security Using Ambient Radio Signals

Liang Xiao*, Qiben Yan†, Wenjing Lou†, and Y. Thomas Hou†
* Dept. Communication Engineering, Xiamen Univ., 361005 China. Email: lxiao@xmu.edu.cn
† Virginia Polytechnic Institute and State University, VA. Email:{qbyan,wjlou,thou}@vt.edu

***Abstract*—In this paper, we propose a privacy-preserving proximity-based security strategy for location-based services in wireless networks, without requiring any pre-shared secret, trusted authority or public key infrastructure. More specifically, radio clients build their location tags according to the unique physical features of their ambient radio signals, which cannot be forged by attackers outside the proximity range. The proximity-based authentication and session key generation is based on the public location tag, which incorporates the received signal strength indicator (RSSI), sequence number and MAC address of the ambient radio packets. Meanwhile, as the basis for the session key generation, the secret location tag consisting of the arrival time interval of the ambient packets, is never broadcast, making it robust against eavesdroppers and spoofers. The proximity test utilizes the nonparametric Bayesian method called infinite Gaussian mixture model, and provides range control by selecting different features of various ambient radio sources. The authentication accuracy and key generation rate are evaluated via experiments using laptops in typical indoor environments.**

I. Introduction

The pervasion of smartphones and social networks has boosted the rapid development of location-based services (LBS), such as the request of the nearest business and the location-based mobile advertising. Reliable and secure location-based services demand secure and accurate proximity tests, which allow radio users and/or service providers to determine whether a client is located within the same geographic region [1]. In order to support business or financial oriented LBS services, proximity tests have to provide location privacy protection and location unforgeability.

Consequently, privacy-preserving proximity tests have recently drawn considerable research attention [1]–[7]. However, based on the received signal strength of a single radio source, many of them have a limited proximity range and provide inaccurate authentication for both stationary and fast changing radio environments [4]–[6]. To address this problem, Zheng et al proposed a location tag-based proximity test that exploits the contents of ambient radio signals to improve the authentication accuracy and provide flexible range control [7].

In this paper, we propose a proximity-based authentication and key generation strategy using ambient radio signals for LBS services in wireless networks. In the Alice-Bob model, a radio client Bob builds a spatial temporal location tag consisting of physical-layer features of his ambient radio environment, which cannot be forged by malicious users outside the proximity range, and sends his public location tag to Alice. Upon receiving Bob's tag, the peer client Alice derives the proximity evidence of Bob and generates the session key without involving any trusted authority, pre-shared secret or public key infrastructure.

In the proximity test, Bob constructs his public location tag that incorporates the received signal strength indicator (RSSI), sequence number (SN) and MAC address of the ambient radio packets, as well as his secret location tag consisting of the packet arrival time during the monitoring period. Bob keeps his secret location tag and only sends his public location tag to Alice, addressing eavesdropping and spoofing attacks. As Bob's location tag does not disclose his location, the proximity test preserves his location privacy.

Nonparametric Bayesian method (NPB) that avoids the "overfitting" problem and thus the challenging job of adjusting model complexity, has recently shown strength in the design of device fingerprints [8] and the detection of primary user emulation attacks in cognitive radio networks [9]. Therefore, we utilize the NPB method called infinite Gaussian mixture model (IGMM) [10] to design the proximity-based authentication, taking into account packet loss due to deep channel fading and/or strong interference, radio environment changes and attacks by adversary clients.

Involving multiple ambient radio sources, our proximity test can achieve more accurate and flexible proximity range control, compared with that relies on a single RSSI trace [4], [5]. Unlike the content-based location tag [7], our tag incorporates the physical-layer features of ambient signals. Thus clients avoid decoding all the ambient radio signals, which significantly reduces the computational overhead and makes it applicable to the case that the ambient packet decoding is not available or desirable. Moreover, compared with [7], our proximity test is more robust against spoofing and eavesdropping by not broadcasting the secret location tag.

The remainder of the paper is organized as follows. We overview related work in Section II. In Section III, we present the proximity-based security system based on ambient radio signals. Then we analyze its performance in Section IV and provide experimental results in Section V. Finally, we conclude in Section VI.

The work by Xiao is partly supported by NSFC (61271242,61001072), the Natural Science Foundation of Fujian Province of China (No.2010J01347), NCETFJ, SRF for ROCS, SEM, and Fundamental Research Funds for the Central Universities (2012121028). The work of W. Lou and Y.T. Hou was supported by NSF grants CNS-1155988 and CNS-1217889, and ONR grant N000141310080.

II. Related Work

Privacy-preserving proximity test has received plenty of attention recently due to the proliferation of smartphones and LBS services. We only overview those closely related to this work. In [3], a practical solution exploits the measured accelerometer data due to hand shaking to determine whether two smartphones are held by one hand.

For the proximity range not limited to a single hand, the RSSI-based proximity test was proposed [4]–[6]. The proximity test in [4] calculates the Euclidean distance between the RSSIs of the shared ambient WiFi environment and applies a classifier called MultiBoost, while the test in [5] relies on the feature of the peer client's signal. In [6], a proximity-based secure pairing strategy exploits the amplitude or phase of the shared ambient TV/FM radio environment to generate bits for the client pairs with longer proximity range. However, these proximity-based methods do not provide flexible range control. To address this problem, Zheng et al proposed a private proximity test and secure cryto protocol, which applies the fuzzy extractor to extract secret keys and bloom filter to efficiently represent the location tags [7]. As mentioned in Section I, we aim at further improving its performance.

III. Proximity-based Security Using Ambient Radio Signals

We present a proximity-based security strategy, including the authentication and session key generation for peer clients in wireless networks. A temporal spatial location tag is built for the radio client, consisting of the physical features of multiple ambient radio sources, which cannot be forged by attackers and does not disclose location privacy of the client.

A. System Model

Consider two radio clients, Alice and Bob, who do not share any secret, trusted authority or public key infrastructure. Alice initiates the proximity test to determine whether Bob is in her proximity and establish a session key with him if they are in the same area. As is available by many off-the-shelf radio devices, both clients can extract the physical features of radio signals, such as RSSI, arrival time, MAC address and SN of the packets sent by the ambient WiFi access points (APs), and/or features of the other ambient signals such as bluetooth and FM radios.

Alice and Bob usually do not receive the same number of ambient packets following the same signal acquisition policy, due to their different ambient radio environments and the packet loss due to channel fading and strong interference. For similar reasons, the clients have different RSSI for the same ambient packet, unless their distance is less than a half wavelength of the signal. On the other hand, clients can extract the same arrival time (minus a constant), SN and MAC address from a given ambient packet. These facts can be utilized to provide range control in the proximity test.

We consider two types of attackers: (1) eavesdroppers whose goal is to obtain the session key between Alice and Bob, and (2) attackers located outside the proximity range, who inject spoofed or replayed signals in hopes of making Alice and Bob generate a mismatched session key. We will investigate the impacts of the other type of attackers in our future work.

B. Proximity-based Authentication

As closely located radio clients have similar ambient radio environments, Alice can decide whether Bob is in her proximity by investigating their ambient radio signals following a nonparametric Bayesian method called infinite Gaussian mixture model (IGMM) [10]. Unlike hypothesis tests such as maximum likelihood estimation, NPB method does not rely on the *a priori* knowledge of input data model and works well even with uncertainty regarding the number of hidden classes and the data model.

In the proximity test, both Alice and Bob monitor their ambient packets over the same frequency channel during the same time period. According to the acquired ambient packets, each client extracts a M-length trace of D-dimensional physical features, such as the RSSI information from D radio sources over time. Bob builds a public location tag based on his ambient feature trace, and sends a message incorporating his public location tag to Alice. Upon receiving this message and checking her own trace with length M, Alice forms $n = 2M$ D-dimensional feature vectors $\mathbf{x}_i$, with $i = 1, \cdots, 2M$. For simplicity, the first M vectors correspond to Alice's trace, while the latter M are extracted from Bob's location tag.

Next, Alice implements IGMM with a Markov chain Monte Carlo method called Gibbs sampling [10]. In this proximity test, $\mathbf{x}_i$ is modelled with the finite Gaussian mixture model (FGMM) with k components, whose probability distribution function (pdf) is given by

$$p(\mathbf{x}_i) = \sum_{l=1}^{k} \pi_l \mathcal{N}(\mu_l, s_l^{-1}), \tag{1}$$

where k is the number of basis Gaussian distributions, μ_l and s_l are the mean and precision of the l-th Gaussian distribution, respectively, and π_l is the mixing proportion with $0 \leq \pi_l \leq 1$ and $\sum_{l=1}^{k} \pi_l = 1$.

In the finite Gaussian mixture model, the prior of μ_l follows Gaussian distribution, whose mean and precision have normal and gamma priors, respectively. Similarly, the prior of s_l in FGMM has Gamma distribution, whose shape and mean follow inverse Gamma and Gamma form. The mixing proportion π_l in (1) follows the Dirichlet distribution, whose joint pdf is given by

$$p(\pi_1, \cdots, \pi_k) = \frac{\Gamma(\alpha)\Pi_{l=1}^{k} \pi_l^{\alpha/k-1}}{\Gamma(\alpha/k)^k}, \tag{2}$$

where $\Gamma(\cdot)$ is the Gamma function. The concentration parameter α in (2) has an inverse Gamma shape, and its pdf is proportional to the following,

$$p(\alpha) \propto \alpha^{-3/2} \exp(-\frac{1}{2\alpha}). \tag{3}$$

Infinite Gaussian mixture model is actually an extreme case of FGMM with k approaching infinity. Let c_i and $\mathbf{c}_{-i}$

denote the classification result of $\mathbf{x}_i$ and the labels for the data other than $\mathbf{x}_i$, respectively. With $n_{-i,j}$ representing the number of data before $\mathbf{x}_i$ belong to Class j, we denote the conditional prior probability for $\mathbf{x}_i$ to belong to Class j as $p(c_i = j|\mathbf{c}_{-i}, \alpha, n_{-i,j})$.

If $n_{-i,j} > 0$, similar to the analysis in [10], the conditional probability can be written as

$$p(c_i = j|\mathbf{c}_{-i}, \alpha, n_{-i,j}) = \frac{n_{-i,j}}{n-1+\alpha}. \quad (4)$$

Otherwise, if no data is assigned to Class j yet, i.e., $n_{-i,j} = 0$, the conditional probability becomes

$$p(c_i = j|\mathbf{c}_{-i}, \alpha) = \frac{\alpha}{n-1+\alpha}. \quad (5)$$

Following Bayesian principle, we have the conditional posterior of the classification indicator as given by

$$p(c_i = j|\mathbf{c}_{-i}, \alpha, \mu_j, s_j) \propto p(c_i = j|\mathbf{c}_{-i}, \alpha)p(\mathbf{x}_i|\mathbf{c}_{-i}, \mu_j, s_j). \quad (6)$$

According to (1)-(6), we can use Gibbs sampling [11] to obtain the classification indicators c_i from $\mathbf{x}_i$. The number of distinct values in the resulting c_i indicates whether Bob is in the proximity of Alice. Ideally, all c_i take the value 1 if Bob is in the proximity, and take two different values if otherwise.

As shown later in Fig. 3.a, RSSI of the same AP's signals monitored by a client changes slightly over time in typical indoor radio environments. Consequently, two classes resulting from the NPB method that are close to each other should be assigned with the same label. The final authentication result is made after a post processing process, which combines the classes resulting from NPB, whose centroid data have the Euclidean distance belowing a threshold.

C. Session Key Establishment

As we know, radio clients in the same geographic area share some ambient radio signals, and can extract the same arrival time[1], MAC and SN information from a given packet. In the key establishment, Alice and Bob generate the session key based on the arrival time interval of their shared ambient packets.

Alice initiates the session key establishment process by requesting the peer client in her proximity to monitor the ambient radio signals. Upon receiving Alice's request, Bob monitors his ambient signals according to the policy specified in the request, and then builds a location tag incorporating the physical features of the received ambient packets.

Each location tag consists of two parts: the public location tag that incorporates the MAC and SN of the received ambient packets, together with their RSSI for the authentication purpose as illustrated in Section III.B, and the secret location tag that includes the corresponding packet arrival time sequence.

Alice utilizes Bob's public location tag to identify their shared ambient packets according to their MAC and SN, and then generates the session key based on the arrival time interval of these packets. Meanwhile, Alice informs Bob the location of their shared packets in his tag, helping him to locate the related information in his secret location tag and derive the session key. Finally, error correction coding, e.g., BCH, is applied to the messages in the handshake process to counteract deep channel fading and strong interference.

[1]The difference of the transmission time is neglected.

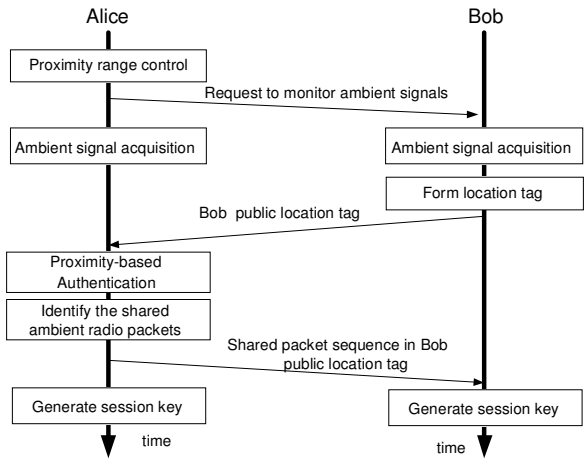


Fig. 1. Steps taken by clients in the key generation process based on the shared ambient radio signals.

As illustrated in Fig. 1, the key establishment process consists of the following steps:

1. Alice first decides her range control policy, and then sends a request to her peer clients to specify the frequency channel, the type of features, and the time duration to monitor the ambient radio signals.

2. Upon receiving the request, Bob, together with Alice, acquires the ambient packets according to the specified policy. Both Alice and Bob store the arrival time, MAC and SN of their received ambient packets in order.

3. Bob builds his location tag, transmits a message to Alice informing his public location tag, and keeps his secret location tag.

4. After authenticating Bob's message, Alice compares his public location tag with her own trace to identify their shared packets, and then builds the session key based on their arrival time intervals.

5. Alice locates their shared packets in Bob's public location and sends the position information to Bob.

6. Bob identifies their shared packets in his secret location tag and generates the session key based on the corresponding arrival time intervals.

IV. Performance of Proximity-Based Security

In this section, we analyze the range control of the proximity-based security based on the selection of different ambient radio sources and/or signal features, briefly discuss its performance against spoofing and eavesdropping attacks, and present the metrics to evaluate the authentication accuracy.

System	Bluetooth	WLAN	GSM	FM radio
Freq (Hz)	2.4G	2.4,5G	.9/1.8G	87.5-108M
Range (m)	$\sim$10	$\sim$35	$\sim$30k	$>$ 100 k

TABLE I
RANGE CONTROL BY SELECTING DIFFERENT AMBIENT RADIO SOURCES IN THE PROXIMITY-BASED SECURITY SYSTEM.

A. Range Control

Radio devices such as smartphones can access multiple radio systems each with different coverage range, as illustrated in Table I. By switching her frequency bands, Alice monitors different radio systems and thus controls her proximity range. For example, Alice, together with Bob, acquires FM radio signals for the proximity range of several miles. If Alice is only interested in the neighbors within the same room, she chooses WiFi or bluetooth signals.

Another way to control the proximity range is to select different features of the ambient radio signals. More specifically, clients have different RSSI, if their distance is greater than a half wavelength. On the other hand, they have the same arrival time, SN and MAC addresses for a given signal, even when their distance is larger than 30 m for WiFi signals. Therefore, a fine-range proximity test can take into account the RSSI information, while a coarse-range test should consider the packet arrival time.

Finally, in the RSSI-based proximity test, the range granularity is controlled by the threshold in the post processing step. In general, the range granularity increases with the threshold. We will provide detailed design of the range control with fine granularity in our future work.

B. Security Performance

We briefly discuss the performance of the proposed proximity-based security against two types of attackers. First is the eavesdropper who aims at deriving the session key between Alice and Bob, or obtaining their locations. As shown in Section III.C, eavesdroppers can only obtain Bob's public location tag, and thus the RSSI, SN and the MAC address of his ambient packets. Obviously that does not disclose Bob's location. Moreover, eavesdroppers cannot derive the session key either, as the arrival time information of the shared ambient packets is never broadcast over the air.

Besides eavesdroppers, our proposed strategy is also robust against attackers outside the proximity range, who inject faked or replayed signals to spoof an ambient radio source, in hopes of leading to the mismatched session key. Since the actual ambient radio source and the attacker usually have different RSSI in their signals, the faked packets can hardly pass the proximity-based authentication, and thus are ignored in the session key generation. In addition, the relayed message cannot pass the authentication, due to the time variation of RSSI. More in-depth analysis of the security performance will be performed in our future work.

Finally, we calculate two metrics to evaluate the authentication performance: (1)Type 1 error rate, also known as false alarm rate, is the probability that Alice rejects the packet from the client in her proximity by mistake; and (2) Type 2 error rate is the probability to falsely accept a packet sent by a client outside her proximity.

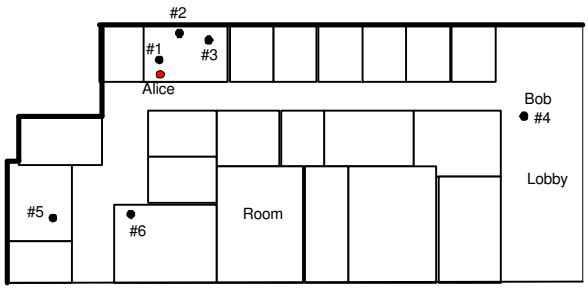


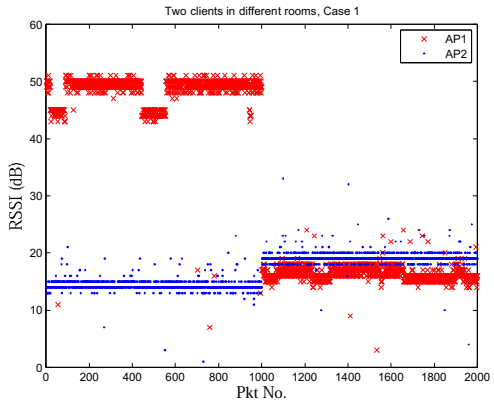
Fig. 2. Placement of Alice and Bob in six experiments performed in Virginia Tech Northern Virginia Center.

V. EXPERIMENTAL RESULTS

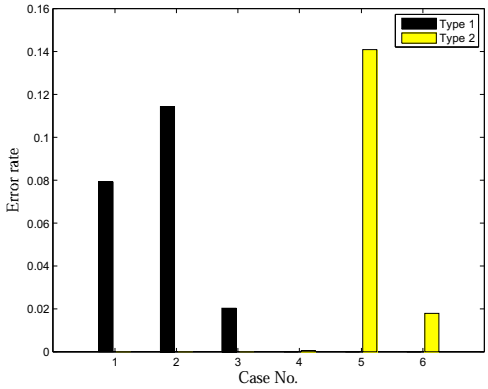
In order to evaluate the performance of the proposed scheme, we performed experiments by using two laptops, each equipped with a wireless adapter AirPcap Nx and an open-source packet analyzer Wireshark. Placed in different locations in Virginia Tech Northern Virginia Center as shown in Fig. 2, the laptops acted as Alice and Bob. They simultaneously captured the ambient WiFi signals, extracted the RSSI, arrival time, SN and MAC address of the beacon frames at 2.417 GHz (Channel 2), and recorded the trace with duration of one minute.

Alice performed the proximity-based authentication based on the RSSI trace of the signals sent by two APs, with MAC addresses as Cisco:6a:7f:41 and Cisco:6c:67:40, respectively. The authentication was performed based on the RSSI trace consisting of 2000 frames, where the first 1000 frames are Alice's record and the second 1000 come from Bob's public location tag. An example of the input to the proximity-based authenticator is illustrated in Fig. 3.a. We performed proximity test to decide whether Bob is in the same office with Alice, and computed the resulting error rates as defined in Section IV.C, for six experiment scenarios as illustrated in Fig. 2. The error rates of the authentication results in the experiments are presented in Fig. 3.b, which is mostly below 15%, verifying the efficacy of the proposed authentication.

We also evaluate the performance of the key generation. Levenshtein distance or edit distance, defined as the minimum number of changes in spelling required to change one word into another, is a metric for measuring the amount of difference between two sequences and widely used for pairwise string alignments. Therefore, we calculate the packet matching rate of the beacon frames from two clients, defined as $(1-D)/L$, where L is the trace length and D is Levenshtein distance. Fig. 4 presents the packet matching rate for six typical indoor scenarios, including three cases where both clients are in the same room and three with clients in different rooms. It is shown that the packet matching ratio is mostly above 40% for the same room case, or above 25% for the different room case,



(a) RSSI trace regarding two APs as the basis of proximity test.



(b) Error rates of the proximity test for six experiments, with topology shown in Fig. 2.

Fig. 3. Experiment results of the proximity-test based on IGMM method.

indicating that clients have plenty of shared ambient packets to build the session key.

We calculated the probability of each value of the packet arrival time interval in the measured traces, and then computed its entropy by definition. Results show that the entropy of the arrival time interval is 12.8 bits. As the average packet arrival interval is 0.0129 s in the experiments, the ideal session key generation rate is approximately as high as $12.8 \times 25\% \div 0.0129$ =248 bps, if the packet matching ratio is 25%.

VI. CONCLUSION

We have proposed a proximity-based authentication and key establishment scheme by exploiting the physical features of ambient radio signals for LBS services in wireless networks. Based on the RSSI information of the ambient radio packets, the authentication utilizes the Markov chain Monte Carlo implementation of a nonparametric Bayesian method called infinite Gaussian mixture model to determine whether a client is in the proximity. In the key establishment scheme, each client pair generate a session keys based on the arrival time interval of their shared ambient packets. Without disclosing the client's location, the proposed scheme is robust against eavesdroppers and spoofers outside the proximity range. Experiments using laptops with WiFi packet analyzer show that the authentication error rate is mostly below 15% for the same-room proximity test and the key generation rate is as high as 248 bps approximately in typical indoor environments.

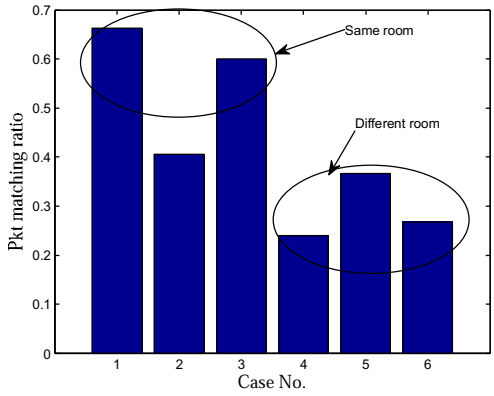


Fig. 4. Packet matching rate of the arrival time information of the beacon frames acquired by Alice and Bob, whose locations are shown in Fig. 2.

Moving forward, further investigation is needed to evaluate the key generation rate and to design in detail the range control with fine granularity. Another interesting topic is to study the computational overhead and communication overhead of our proposed strategy. Finally, we are working to thoroughly evaluate its performance under a wide range of network scenarios against various types of attacks and compare it with existing proximity-based security strategies.

REFERENCES

[1] A. Narayanan, N. Thiagarajan, M. Lakhani, M. Hamburg, and D. Boneh., "Location privacy via private proximity testing," in *Proc. Network and Distributed System Security Symposium (NDSS)*, 2011.

[2] N. Talukder and S. Ahamed, "Preventing multi-query attack in location-based services," in *Proc. ACM conference on Wireless network security*, 2010.

[3] R. Mayrhofer and H. Gellersen, "Shake well before use: intuitive and secure pairing of mobile devices," *IEEE Trans. Mobile Computing*, vol. 8, pp. 792 – 806, June 2009.

[4] A. Varshavsky, A. Scannell, A. LaMarca, and E. Lara, "Amigo: Proximity-based authentication of mobile devices," in *Proc. UbiComp*, 2007.

[5] A. Kalamandeen, A. Scannell, E. de Lara, A. Sheth, and A. LaMarca, "Ensemble: cooperative proximity-based authentication," in *Proc. ACM 8th international conference on Mobile systems, applications, and services*, 2010.

[6] S. Mathur, R. Miller, A. Varshavsky, and W. Trappe, "Proximate: Proximity-based secure pairing using ambient wireless signals," in *Proc. ACM MobySys*, 2011.

[7] Y. Zheng, M. Li, W. Lou, and T. Hou, "Sharp: Private proximity test and secure handshake with cheat-proof location tags," in *Proc. European Symposium on Research in Computer Security (ESORICS)*, 2012.

[8] N. Nguyen, G. Zheng, Z. Han, and R. Zheng, "Device fingerprinting to enhance wireless security using nonparametric bayesian method," in *Proc. IEEE INFOCOM*, 2011.

[9] N. Nguyen, R. Zheng, and Z. Han, "On identifying primary user emulation attacks in cognitive radio systems using nonparametric bayesian classification," *IEEE Trans. Signal Processing*, vol. 60, pp. 1432– 1445, March 2012.

[10] C. Rasmussen, "The infinite gaussian mixture model," *Advances in neural information processing systems*, pp. 554– 560, 2000.

[11] C. Bishop, *Pattern recognition and machine learning*, Springer Press, 2006.