Proximity-Based Security Techniques for Mobile Users in Wireless Networks

Liang Xiao, *Senior Member, IEEE*, Qiben Yan, *Student Member, IEEE*, Wenjing Lou, *Senior Member, IEEE*, Guiquan Chen, *Student Member, IEEE*, and Y. Thomas Hou, *Senior Member, IEEE*

***Abstract*— In this paper, we propose a privacy-preserving proximity-based security system for location-based services in wireless networks, without requiring any pre-shared secret, trusted authority, or public key infrastructure. In this system, the proximity-based authentication and session key establishment are implemented based on spatial temporal location tags. Incorporating the unique physical features of the signals sent from multiple ambient radio sources, the location tags cannot be easily forged by attackers. More specifically, each radio client builds a public location tag according to the received signal strength indicators, sequence numbers, and media access control (MAC) addresses of the ambient packets. Each client also keeps a secret location tag that consists of the packet arrival time information to generate the session keys. As clients never disclose their secret location tags, this system is robust against eavesdroppers and spoofers outside the proximity range. The system improves the authentication accuracy by introducing a nonparametric Bayesian method called infinite Gaussian mixture model in the proximity test and provides flexible proximity range control by taking into account multiple physical-layer features of various ambient radio sources. Moreover, the session key establishment strategy significantly increases the key generation rate by exploiting the packet arrival time of the ambient signals. The authentication accuracy and key generation rate are evaluated via experiments using laptops in typical indoor environments.**

***Index Terms*— Authentication, encryption, wireless networks, Gaussian mixture model.**

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]–[4]. In order to support the business or financial oriented LBS services, proximity tests have to provide location privacy protection and location unforgeability [5]–[9].

Consequently, privacy-preserving proximity tests have recently drawn considerable research attention [10]–[16]. Based on the received signal strength (RSS) of a single radio source, many of the proximity tests suffer from the limited proximity range and the authentication accuracy is not high in both stationary and fast changing radio environments [14], [15]. Moreover, a recent study has shown that the RSS-based strategies are vulnerable to man-in-the-middle attacks [17]. To address this problem, Zheng et al. have proposed a location tag-based proximity test, which exploits the contents of ambient radio signals to improve the authentication accuracy and provides flexible range control [16]. However, the extraction of the packet contents in the proximity test not only engenders privacy leakage, but also increases the overall system overhead.

In typical indoor environments, a radio client can usually access multiple ambient radio sources, such as WiFi access points (APs), bluetooth devices and FM radios. Many off-the-shelf radio devices, such as laptops and smartphones, can readily extract the physical-layer features of the ambient signals, including the received signal strength indicator (RSSI) and the packet arrival time. Field tests have shown that clients in the same geographic area can observe a certain shared ambient signals, with approximately the same normalized packet arrival time and similar RSSIs. These physical-layer features do not directly disclose the client location and cannot be easily estimated and forged by a client outside the proximity [18]. Therefore, users can exploit the ambient radio signals to establish spatial temporal location tags and use the location tags to enhance security for LBS services.

In this paper, we propose a proximity-based authentication and key generation strategy for radio clients, without involving any trusted authority, pre-shared secret or public key infrastructure. For simplicity, we assume that a radio client called Alice initiates the authentication and pairwise session key generation with clients in her proximity. A peer client called Bob responds to her request.[1] Both clients monitor their ambient radio signals at the frequency band during the time specified by Alice.

According to the physical-layer features of the signals sent by multiple ambient radio sources, Bob constructs and

Manuscript received June 16, 2013; revised October 3, 2013; accepted October 11, 2013. Date of publication October 18, 2013; date of current version November 13, 2013. The work by L. Xiao was supported in part by NCETFJ, in part by the Fundamental Research Funds for the Central Universities under Grant 2012121028, in part by NSFC under Grants 61271242 and 61001072, and in part by the Natural Science Foundation of Fujian Province of China under Grant 2010J01347. The work of Y. T. Hou was supported by NSF under Grants CNS-1217889, CNS-1156318, and CNS-1156311. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Jiangtao Li.

L. Xiao and G. Chen are with the Department of Communication Engineering, Xiamen University, Xiamen 361005, China (e-mail: lxiao@xmu.edu.cn; guiq_chen@yahoo.cn).

Q. Yan, W. Lou, and Y. T. Hou are with Virginia Polytechnic Institute and State University (Virginia Tech), Blacksburg, VA 24061 USA (e-mail: qbyan@vt.edu; wjlou@vt.edu; thou@vt.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIFS.2013.2286269

[1] This system can be directly extended to the case with Alice connecting to multiple peer clients.

informs Alice his public location tag, which incorporates the RSSIs, sequence numbers (SN) and media access control (MAC) addresses of the packets. Bob also builds and keeps a secret location tag, which consists of the packet arrival time sequence. Based on Bob's public location tag and her own measurements, Alice identifies their shared ambient packets and uses their features to derive the proximity evidence of Bob for both authentication and session key generation. Meanwhile, Alice informs Bob the indices of their shared packets in his secret location tag and helps him to generate his copy of the session key.

The authentication utilizes a nonparametric Bayesian method (NPB) called infinite Gaussian mixture model (IGMM) [19] to classify the RSSI data. This method avoids the "overfitting" problem and thus addresses the challenging issue of adjusting model complexity. The NPB method has shown its strength in the design of device fingerprints [20] and the detection of primary user emulation attacks in cognitive radio networks [21]. As an important alternative to deterministic inference such as expectation-maximization algorithm [22], the IGMM model is implemented in the proximity test to authenticate radio clients.

The proximity-based security system takes into account the packet loss due to the channel fading and interference, and can counteract various types of attacks. By hiding the packet arrival time sequence in the secret location tag, which is the basis of the session key and cannot be forged by malicious users, this scheme can efficiently address eavesdropping and spoofing attacks [34] who are located outside the proximity range. Moreover, as public location tags do not disclose the client locations, location privacy is preserved for radio clients.

Involving multiple ambient radio sources, the proximity test improves the authentication accuracy and obtains more flexible range control than those based on a single RSSI trace [14]. Unlike the content-based location tag [16], the tag in this work consists of the physical-layer features of ambient signals, and thus avoids decoding the ambient signals. Therefore, this work is applicable to the case that the ambient packet decoding is not available or desirable, significantly reduces the computational overhead, and prevents privacy leakages.

A. Contributions

The contributions of this paper are summarized as follows:

(1) We exploit the arrival time sequence of the shared ambient radio packets to establish pairwise session keys for proximity clients. This scheme achieves a faster and more reliable key generation than the RSSI-based strategies [23].

(2) Unlike the work [16] whose location tag incorporates the contents of the ambient packets, this strategy depends on the physical-layer features, including the packet arrival time and RSSI. Without checking the packet contents, this system provides better privacy protection and is more robust against spoofing, eavesdropping, replay attacks and man-in-the-middle attacks.

(3) By applying the nonparametric Bayesian method called IGMM and exploiting the packet arrival time information, the proximity test is more accurate than [13]–[15]. Moreover, this strategy also provides more flexible proximity range control and larger coverage area by combining the packet arrival time and RSSI information for appropriately chosen ambient radio signals.

B. Related Work

As a location sharing method, proximity test enables the information sharing between users within a certain range. Related security issues have recently received significant attentions among researchers [12]–[15], [24]–[27]. In [12], a practical solution exploits the measured accelerometer data resulting from hand shaking to determine whether two smartphones are held by one hand.

For the proximity range exceeding a single hand, RSSI-based proximity tests were proposed in [13]–[15]. The proximity test in [13] calculates the Euclidean distance between the RSSIs of the shared ambient WiFi signals and applies a classifier called MultiBoost. The test in [14] relies on the feature of the peer client's signal. In [15], a secure pairing strategy exploits the amplitudes or phases of the shared ambient TV/FM radio signals to generate bits for the client pairs with longer proximity range. However, these methods are limited to the case where the distance between the radio clients is no more than a half wavelength away [15].

To achieve flexible range control, Zheng et al. proposed a private proximity test and secure cryto protocol, which applies the fuzzy extractors to extract secret keys and bloom filters to efficiently represent the location tags [16]. Inspired by this work, we propose a location tag-based security technique to further improve the performance, and some preliminary results were given in [28]. In this paper, we move forward to present the proximity-based security protocol that incorporates the proximity range control with fine granularity. We analyze the range control and security performance, and perform in-depth experiments to evaluate its performance such as the key generation rate and session key matching rate in typical indoor scenarios.

The remainder of the paper is organized as follows. We describe the system model in Section II. Then we present the proximity-based authentication method in Section III, and the session key generation method in Section IV. We discuss the proximity range control and other important issues in Section V and provide experimental results in Section VI. Finally, we conclude in Section VII.

II. System Model

In this paper, we consider two radio mobile clients called Alice and Bob, respectively, who are located in a certain geographic region. Without sharing any secret, trusted authority or public key infrastructure with Bob, Alice aims at initiating a proximity test and establishing a session key with him.

Both clients apply off-the-shelf radio devices, such as laptops and smartphones to extract the features of ambient radio signals, including the RSSIs, arrival time, source MAC addresses and sequence numbers (SN) of the packets. For simplicity, we take the 802.11 systems as an example in this section and consider the other types of radio sources in the later sections. In this system, each client monitors the ambient packets, which can be sent by access points (APs), over the frequency channel during the time specified by Alice, yielding a feature trace with N records.

As shown in [13], [16], [28], a radio client in typical indoor environments can usually receive signals from *multiple* APs. For example, a stationary laptop in an experiment as later shown in Fig. 3 received signals from four APs in the 0.24s time duration. Unlike [14], we utilize the ambient signals sent by multiple APs instead of the testing packets sent by the clients or a single neighboring AP. In addition, clients have small chances to receive the same ambient packet sequence in the presence of multiple APs due to the path-loss and small-scale fading in radio propagation in typical indoor environments. Therefore, an attacker outside the proximity can rarely obtains all the shared ambient packets between Alice and Bob, and thus has difficulty predicting the exact arrival time sequence for their shared ambient packets.

We assume that Alice initiates the proximity test, while Bob can be either an honest client to be tested or an attacker outside the area. In this work, Bob sends his temporal spatial location tag incorporating the trace information to Alice, and hence Alice obtains the RSSIs, MAC addresses and SN information of Bob's ambient signals. Let $rssi_i^A$, t_i^A, MAC_i^A and SN_i^A denote the RSSI, arrival time, MAC address and sequence number of the i-th ambient packet in Alice's feature trace, respectively, with $i = 1, \cdots, N$. Similarly, let $rssi_i^B$, t_i^B, MAC_i^B and SN_i^B represent the corresponding information monitored by Bob.

A. Proximity-Based Security Protocol

By integrating the authentication and key generation process, we build a proximity-based security protocol for mobile users in wireless networks. As illustrated in Fig. 1, this protocol consists of the following steps:

1. According to the desired proximity range, Alice decides and broadcasts her proximity test policy, including the frequency channel, the time duration and the features to monitor the ambient signals.

2. Upon receiving Alice's request, Bob measures the features of the packets as Alice specified. Both clients extract and store the RSSIs, arrival time, MAC addresses and sequence numbers of their ambient packets, i.e., $\mathrm{rssi}_i^X, t_i^X, \mathrm{MAC}_i^X$ and SN_i^X, with $1 \le i \le N$.

3. Bob builds a location tag, sends Alice his public location tag, and keeps his secret location tag.

4. Alice authenticates Bob.

5. Alice compares Bob's public location tag with her trace to identify their shared packets. Following a key generation algorithm, Alice builds a session key, $\mathbf{K}_A$, and informs Bob the indices of their shared packets in his trace, $\mathbf{J}$.

6. Based on his secret location tag and the indices $\mathbf{J}$, Bob generates his session key, $\mathbf{K}_B$.

In the above handshake process, error correction coding such as BCH can be applied to counteract the transmission errors due to channel fading and interference. In addition, because of the different ambient radio environments and packet loss rates, clients usually take different time to obtain a given number of ambient packets. Due to this problem, the proposed key generation strategy solely relies on the same shared packets between Alice and Bob and thus provides a certain degree of robustness against packet loss. More details of this protocol are presented in Section III and Section IV.

TABLE I

SUMMARY OF SYMBOLS AND NOTATIONS

$rssi_i^X$ t_i^X MAC_i^X SN_i^X $\mathbf{X}_i = [MAC_i^X, SN_i^X]$	RSSI of Packet i received by Client X Arrival time of Packet i received by X MAC addr. of Packet i received by X SN of Packet i received by X MAC addr. & SN of Packet i received by X
N D $\mathbf{x} = [x_i]_{1 \le i \le n}$ $\mathbf{c} = [c_i]_{1 \le i \le n}$	Length of the trace recorded by Alice Number of the ambient radio sources Feature records obtained by Alice Classification results of $\mathbf{x}$
Θ ν Δ Υ Ξ	Threshold to evaluate the Euclidean distance Proximity passing rate of Bob's records Threshold to evaluate ν in the authentication Rounding precision Threshold to evaluate the key generation rate in the authentication
$\mathbf{K}_X$	Session key generated by Client X

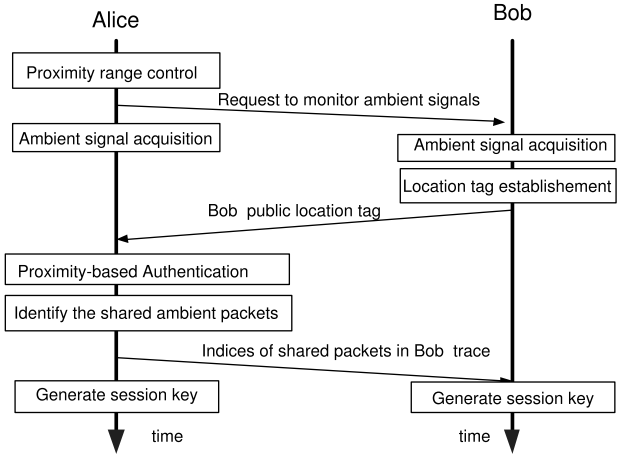

Alice
Bob
Proximity range control
Request to monitor ambient signals
Ambient signal acquisition
Ambient signal acquisition
Location tag establishement
Bob public location tag
Proximity-based Authentication
Identify the shared ambient packets
Indices of shared packets in Bob trace
Generate session key
Generate session key
time
time

Fig. 1. Flowchart of the proximity-based security system based on ambient radio signals.

The proximity-based security techniques have to address the following types of adversary clients: (1) third-party eavesdroppers whose goal is to obtain the session key between Alice and Bob, (2) third-party attackers located outside the proximity range, who inject spoofed or replay signals in hopes of leading to a mismatched session key between Alice and Bob, and (3) Bob as an attacker who aims at illegally passing the proximity test although he is outside the proximity range specified by Alice. We will investigate the impacts of the other attackers in our future work. For ease of reference, the commonly used notations are summarized in Table I.

III. PROXIMITY-BASED AUTHENTICATION

Receivers in the proximity have similar RSSIs and approximately the same arrival time regarding their shared ambient radio signals. Without directly disclosing the clients' locations, these physical-layer features cannot be easily estimated and thus be forged by clients outside the neighborhood [18]. Therefore, we propose a proximity-based authentication strategy for peer clients in wireless networks, where Alice decides whether Bob is in her proximity without violating his location privacy.

The proximity-based authentication is based on the similarity between the physical features of the shared ambient radio signals obtained by the radio clients. More specifically, Alice compares her trace with Bob's measurements extracted from his public location tag, according to a nonparametric Bayesian method (NPB) called infinite Gaussian mixture model (IGMM). Unlike the hypothesis tests such as maximum likelihood estimation, IGMM does not rely on the *a priori* knowledge of the input data model and works well even with uncertainty regarding the number of hidden classes and the data model [19]. In this authentication strategy, Alice classifies the RSSI information of the ambient signals from D APs to authenticate clients such as Bob.

A. IGMM-Based Proximity Test

According to Bob's location tag and her own feature trace, each with N records, Alice obtains a record vector $\mathbf{x}$ with $n = 2N$ feature records. For simplicity of denotation, we assume in this section that each record has only $D = 1$ dimension and $\mathbf{x} = [x_i]_{1\le i\le n} \triangleq [rssi_1^A, \dots, rssi_N^A, rssi_1^B, \dots, rssi_N^B]$, where the first N elements correspond to Alice's trace. However, this method can be extended straightforwardly to the multivariate case with D features, where the gamma variables are replaced by Wishart random matrices and the normal variables become multinormal random vectors. As an example, the experiments that will be presented in Section VI took into account the RSSI data of the signals sent by two ambient radio sources with $D = 2$.

The proximity test is based on the implementation of the IGMM model of $\mathbf{x}$ with the Markov chain Monte Carlo method called Gibbs sampling [22]. More specifically, first, we can use the finite Gaussian mixture model (FGMM) with k basis Gaussian distributions [19] to model the RSSI data x_i in Alice's record vector. In this model, the probability distribution function (pdf) of x_i is given by

$$p(x_i) = \sum_{l=1}^{k} \pi_l \mathrm{N}\big(\mu_l, s_l^{-1}\big), \quad \forall 1 \le i \le n, \tag{1}$$

where μ_l and s_l are the mean and precision of the l-th Gaussian distribution, respectively, and π_l is the mixing proportion [22] with $0 \le \pi_l \le 1$ and $\sum_{l=1}^{k} \pi_l = 1$.

The component means μ_l in Eq. (1) have the following Gaussian priors,

$$p(\mu_l|\lambda, r) \sim \mathrm{N}\big(\lambda, r^{-1}\big), \tag{2}$$

where $\sim$ means "to be proportional to". Both the mean, λ, and precision, r, are hyperparameters with the same values for all the k components in FGMM. They have the following normal and gamma priors:

$$p(\lambda) \sim \mathrm{N}\big(\mu_x, \sigma_x^2\big), \tag{3}$$

and

$$p(r) \sim \mathrm{G}\big(1, \sigma_x^{-2}\big), \tag{4}$$

where μ_x and σ_x^2 are the mean and variance of the RSSI value x_i, respectively.

Let $\mathbf{c} = [c_i]_{1\le i\le n}$ denote the classification labels of Alice's record vector $\mathbf{x}$, where c_i is the classification result of x_i, and $\mathbf{c}_{-i}$ incorporate the labels for the observations other than x_i. Following Bayesian principle, by (1) and (2), we can derive the posterior distribution of μ_l, conditioned on the classification results $\mathbf{c}$,

$$p(\mu_l|\mathbf{c}, \mathbf{x}, s_l, \lambda, r) \sim \mathrm{N}\left(\frac{\bar{x}_l n_l s_l + \lambda r}{n_l s_l + r}, \frac{1}{n_l s_l + r}\right), \tag{5}$$

where $\bar{x}_l$ is the mean of the observations belonging to Class l that has n_l elements and is given by

$$\bar{x}_l = \frac{1}{n_l} \sum_{j:c_j=l} x_j. \tag{6}$$

Similar to the derivation in [19], according to (2)-(5), the posteriors of the hyperparameters, λ and r, are given by

$$p(\lambda|\mu_1, \cdots, \mu_k, r) \sim \mathrm{N}\left(\frac{\mu_x \sigma_x^{-2} + r\sum_{l=1}^{k} \mu_l}{\sigma_x^{-2} + kr}, \frac{1}{\sigma_x^{-2} + kr}\right), \tag{7}$$

$$p(r|\mu_1, \cdots, \mu_k, \lambda) \sim \mathrm{G}\left(k+1, \frac{k+1}{\sigma_x^2 + \sum_{l=1}^{k} (\mu_l - \lambda)^2}\right). \tag{8}$$

Similarly, the component precisions s_l in Eq. (1) have the Gamma priors as follows,

$$p(s_l|\beta, \omega) \sim \mathrm{G}(\beta, \omega^{-1}), \tag{9}$$

whose shape β and mean ω^{-1} are hyperparameters in the FGMM model. Their priors have the following inverse Gamma and Gamma forms,

$$p(\beta^{-1}) \sim \mathrm{G}(1, 1), \tag{10}$$

$$p(\omega) \sim \mathrm{G}(1, \sigma_x^2). \tag{11}$$

By (1) and (9), we obtain the posterior of s_l as

$$p(s_l|\mathbf{c}, \mathbf{x}, \mu_l, \beta, \omega) \sim \mathrm{G}\left(\beta + n_l, \frac{\beta + n_l}{\beta\omega + \sum_{j:c_l=l} (x_j - \mu_l)^2}\right). \tag{12}$$

Then, by combining Eqs. (9)-(11) and after simplification, we have the following posteriors

$$p(\omega|s_1, \cdots, s_k, \beta) \sim \mathrm{G}\left(k\beta + 1, \frac{k\beta + 1}{\sigma_x^{-2} + \beta\sum_{j=l}^{k} s_j}\right), \tag{13}$$

$$p(\beta|s_1, \cdots, s_k, \omega) \sim \Gamma\big(\frac{\beta}{2}\big)^{-k} e^{\frac{-1}{2\beta}} \big(\frac{\beta}{2}\big)^{\frac{k\beta-3}{2}} \prod_{j=1}^{k} (\omega s_j)^{\frac{\beta}{2}} e^{-\frac{\beta s_j \omega}{2}}. \tag{14}$$

According to [19], the mixing proportion $\hat{\pi} = [\pi_l, \cdots, \pi_k]$ in Eq. (1) follows Dirichlet distribution, whose joint pdf is given by

$$p(\pi_1, \cdots, \pi_k|\alpha) = \frac{\Gamma(\alpha)\prod_{l=1}^{k} \pi_l^{\alpha/k-1}}{\Gamma(\alpha/k)^k}, \tag{15}$$

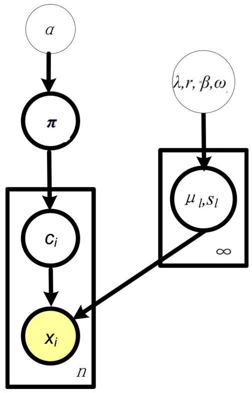


Fig. 2. Directed graph with plate notations for the infinite Gaussian mixture model in the proximity test.

where $\Gamma(\cdot)$ is the Gamma function. The concentration parameter α in Eq. (15) has an inverse Gamma shape, and its prior and posterior can be written as

$$p(\alpha) \sim \alpha^{-3/2} \exp\big(-\frac{1}{2\alpha}\big), \tag{16}$$

$$p(\alpha|k,n) \sim \frac{\alpha^{k-3/2} \exp(-\frac{1}{2\alpha})\Gamma(\alpha)}{\Gamma(n+\alpha)}. \tag{17}$$

By using the standard Dirichlet integral and integrating out the mixing proportions, we have the prior of the indicators as the following,

$$p(c_1,\cdots,c_n|\alpha) = \int p(c_1,\cdots,c_n|\hat{\pi})p(\hat{\pi})d\pi_1\cdots d\pi_k \tag{18}$$

$$= \frac{\Gamma(\alpha)}{\Gamma(n+\alpha)} \prod_{j=1}^{k} \frac{\Gamma(\alpha/k+n_j)}{\Gamma(\alpha/k)}, \tag{19}$$

where n_j is the number of data labelled with Class j.

Let $n_{-i,j}$ represent the number of data before x_i belonging to Class j, and $p(c_i = j|\mathbf{c}_{-i}, \alpha, n_{-i,j})$ denote the conditional prior probability for x_i in Class j. The infinite Gaussian mixture model can be viewed as an extreme case of FGMM with k in Eq. (19) approaching infinity. Consequently, if $n_{-i,j} > 0$, the conditional probability of c_i in the IGMM model can be simplified into

$$p(c_i = j|\mathbf{c}_{-i}, \alpha, n_{-i,j}) = \frac{n_{-i,j}}{n-1+\alpha}. \tag{20}$$

Otherwise, if no data has been assigned to Class j yet, i.e., $n_{-i,j} = 0$, the conditional probability of c_i in IGMM becomes

$$p(c_i = j|\mathbf{c}_{-i}, \alpha) = \frac{\alpha}{n-1+\alpha}. \tag{21}$$

According to Bayesian principle, we obtain the conditional posterior of the classification indicator as

$$p(c_i = j|\mathbf{c}_{-i}, \alpha, \mu_j, s_j) \sim p(c_i = j|\mathbf{c}_{-i}, \alpha)p(x_i|\mathbf{c}_{-i}, \mu_j, s_j). \tag{22}$$

The relationships among the hyperparameters (λ, r, β and ω), the input data $\mathbf{x}$ and the variables in the infinite Gaussian mixture model can be illustrated in the directed graph with plate notations in Fig. 2, where the rectangular block represents the repeated structure.

Algorithm 1 IGMM-Based Authentication

```
Input: RSSI measurements $\mathbf{x} = [x_i]_{1 \le i \le n}$
Output: Authentication result
  $k \leftarrow 1$
  $\mu_x \leftarrow E[\mathbf{x}]$, $\sigma_x^2 \leftarrow Var[\mathbf{x}]$
  $\lambda \leftarrow$ Eq. (3), $r \leftarrow (4)$, $\mu_l \leftarrow (2)$
  $\beta \leftarrow (10)$, $\omega \leftarrow (11)$, $s_l \leftarrow (9)$
  for $iter \leftarrow 0$ to $NU$ do
    for $l \leftarrow 1$ to $k$ do
      $\mu_l \leftarrow (5)$, $s_l \leftarrow (12)$
    end for
    $\lambda \leftarrow (7)$, $r \leftarrow (8)$
    $\beta \leftarrow (14)$, $\omega \leftarrow (13)$
    $\alpha \leftarrow (17)$
    for $i = 1$ to $n$ do
      $c_i \leftarrow (22)$
      if $c_i > k$ then
        Generate a new class $c_i$
        $\mu_{c_i} \leftarrow (2)$, $s_{c_i} \leftarrow (9)$
      end if
      Update $\mathbf{c}$ by deleting empty classes
      $k \leftarrow$ Number of distinct components in $\mathbf{c}$
    end for
  end for
  Update $\mathbf{c}$ by combining the classes whose centroid Euclidean distance is less than $\Theta$
  $C_A \leftarrow (23)$
  $j \leftarrow 0$
  for $i \leftarrow N+1$ to $2N$ do
    if $c_i = C_A$ then
      Alice accepts the packet, $j++$
    end if
  end for
  Proximity passing rate $\nu \leftarrow j/N$
  if $\nu > \Delta$ then
    Bob passes the authentication
  else
    Bob fails the authentication
  end if
```

In the proximity test, we can apply the Gibbs sampling method to generate the random samples for the joint probability distributions given by the above formulas of the IGMM model. The classification indicators $\mathbf{c}$ can be calculated according to the observations $\mathbf{x}$. The number of distinct values in the resulting c_i indicates whether the recipient of the ambient signal is in the proximity of Alice. Ideally, all c_i take the same value if Bob is in the proximity with Alice, and take two different values if otherwise.

Detailed steps of the IGMM-based proximity test are illustrated in Algorithm 1, where NU is an integer that has to be set large enough to ensure accurate sampling for the IGMM model. In addition, the system parameter N has to be less than the maximum value of sequence number of the specified ambient signals to avoid packet aliasing.

B. Post-IGMM Process

As radio signals in typical radio environments usually have time-variant RSSIs, a post-IGMM process is proposed to address slight channel time variations. This process combines the classes resulting from the IGMM-based proximity test, if they are close to each other. More specifically, if the Euclidean distance of the centroids of two classes is below a threshold denoted as Θ, these two classes are joined together. We now take Class i and j for instance. If $\| E_{c_l=i}[x_l] - E_{c_l=j}[x_l] \| < \Theta$, we combine these data and update the labels $c_l \in \{i, j\}$ with $\min(i, j)$, $\forall 1 \leq l \leq n$. Then the empty class is deleted by reducing c_l by one if their original value $c_l > \max(i, j)$.

Next, we apply the majority rule to process Alice's trace with N records and calculate their new label C_A by the following,

$$C_A = \arg\max_{c \in \mathbf{c}} \sum_{i=1}^{N} \delta(c_i - c), \tag{23}$$

where $\delta(\cdot)$ is the discrete delta function. Alice accepts the data whose label equals C_A. We define the proximity passing rate denoted with ν as the ratio of Bob's records that pass the proximity test after the majority rule. Bob passes the proximity test, if the passing rate of his monitored ambient packets exceeds a threshold Δ, i.e., $\nu > \Delta$.

The above IGMM-based authentication strategy is summarized in Algorithm 1. Besides this RSSI-based strategy, we also provide another authentication strategy that exploits the packet arrival time to achieve a larger proximity range. More specifically, as the key generation rate of the strategy given by Algorithm 2 contains proximity information, we can utilize this information for authentication purpose. More details will be provided in Section V.

IV. Session Key Establishment

Note that clients receive the shared ambient radio packets approximately at the same time. Hence they can exploit the arrival time of the packets to establish pair-wise session keys without requiring any pre-shared secret, trusted authority or public key infrastructure. To this end, Alice initiates the process by broadcasting her key establishment policy. Upon receiving the policy, radio clients in the proximity including Bob monitor the ambient signals accordingly and build their spatial temporal location tags by extracting the physical-layer features of the signals.

Each location tag consists of two parts: a secret location tag that incorporates the packet arrival time information and is kept by the client, and the public location tag that informs Alice the RSSIs for authentication and the MAC addresses and SNs to identify ambient packets.[2] To counteract the difference between the secret location tag between clients due to the transmission over air, the measured packet arrival time is rounded according to a properly chosen rounding precision. The rounding precision denoted with Υ is a tradeoff between the key generation speed and the key matching rate between clients.

Algorithm 2 Session Key Generation

```
Input:
  $\underline{A} = [\mathbf{A}_i]^T_{1 \leq i \leq N}$, $\mathbf{A}_i = [MAC^A_i, SN^A_i]$
  $\underline{B} = [\mathbf{B}_i]^T_{1 \leq i \leq N}$, $\mathbf{B}_i = [MAC^B_i, SN^B_i]$
  $t^A_i$ and $t^B_i$: packet arrival time, $1 \leq i \leq N$
  $\Upsilon$: Rounding precision
Output: Session Key, $\mathbf{K}_A$ and $\mathbf{K}_B$
  $\mathbf{I} \leftarrow \{i | \exists j, 0 \leq i, j \leq N, \mathbf{A}_i = \mathbf{B}_j\}$
  $\mathbf{J} \leftarrow \{j | \exists i, 0 \leq i, j \leq N, \mathbf{A}_i = \mathbf{B}_j\}$
  Alice sends $\mathbf{J}$ to Bob
  $t_a \leftarrow t^A_1$, $t_b \leftarrow t^B_1$
  for $i \leftarrow 1$ to $N$ do
    $T^A_i \leftarrow round(t^A_i - t_a, 10^{-\Upsilon})$
    $T^B_i \leftarrow round(t^B_i - t_b, 10^{-\Upsilon})$
  end for
  $\mathbf{K}_A \leftarrow [T^A_i]_{i \in \mathbf{I}}$
  $\mathbf{K}_B \leftarrow [T^B_i]_{i \in \mathbf{J}}$
```

For simplicity of notation, we take the key establishment between Alice and Bob as an example. Define $\underline{A} \triangleq [\mathbf{A}_i]_{1 \leq i \leq N}$ and $\underline{B} \triangleq [\mathbf{B}_i]_{1 \leq i \leq N}$, where $\mathbf{A}_i \triangleq [MAC^A_i, SN^A_i]$ and $\mathbf{B}_i \triangleq [MAC^B_i, SN^B_i]$. Bob's secret location tag contains t^B_i, $1 \leq i \leq N$, and his public location tag consists of $\underline{B}$, i.e., the MAC addresses and SNs of his ambient signals.

To address the transmission time, both Alice and Bob round the packet arrival time according to Υ. As Alice and Bob are asynchronous in general, we take the first packet received by both clients as the reference packet and utilize the packet arrival time offset in terms of the arrival time of the reference packet in Algorithm 2. More specifically, let $t_a = t^A_1$ and $t_b = t^B_1$ denote the arrival time of the reference packet at Alice and Bob, respectively. Alice and Bob take the rounded packet arrival time offsets, $T^A_i \triangleq round(t^A_i - t_a, 10^{-\Upsilon})$ and $T^B_i \triangleq round(t^B_i - t_b, 10^{-\Upsilon})$ in the session key generation to address the clock difference between the radio devices. The selections of $\Upsilon = 1$, 2 and 3 correspond to the rounding of the time information to the order of 0.1s, 0.01s and 1ms, respectively. Experimental results show that $\Upsilon = 2$ is a reasonable choice for ambient WiFi signals.

The session key generation process is presented in Algorithm 2. Upon receiving Bob's public location tag, Alice compares it with her trace to identify their shared ambient packets. As a result, Alice obtains their indices in her trace and Bob's trace, given by $\mathbf{I} = \{i | \exists j, 0 \leq i, j \leq N, \mathbf{A}_i = \mathbf{B}_j\}$, and $\mathbf{J} = \{j | \exists i, 0 \leq i, j \leq N, \mathbf{A}_i = \mathbf{B}_j\}$, respectively. Then Alice sends $\mathbf{J}$ to Bob.

In the next step, Alice generates her session key $\mathbf{K}_A$ based on the arrival time of their shared packets, i.e., $\mathbf{K}_A = [T^A_i]_{i \in \mathbf{I}}$. Similarly, Bob uses $\mathbf{J}$ to find their shared packets in his secret location tag and derives his session key with $\mathbf{K}_B = [T^B_i]_{i \in \mathbf{J}}$. The proposed key establishment process is summarized in Algorithm 2. We can see that this strategy has low complexity and is easy to implement.

[2] The duration is assumed to be short enough to avoid the reuse of SN for a given radio source.

V. Proximity Range Control and Security Analysis

In this section, we discuss related issues of the proposed security techniques, including the proximity range control and the security performance against various types of attackers.

A. Proximity Range Control

In this system, Alice can control the proximity range by choosing appropriate ambient radio sources and signal features at multiple levels. First, as shown in Table II, radio devices such as smartphones and laptops can access multiple radio sources with various coverage ranges and frequency bands. By switching her frequency bands, Alice chooses the radio sources whose coverage ranges are larger than the proximity range. For example, Alice can use FM radio signals for the proximity range of several miles, and choose WiFi or bluetooth signals if contacting with clients within the same room.

Second, the range control can also be achieved by selecting suitable physical-layer features, since the features have different coherent spacial distances. For example, Alice and Bob usually obtain different RSSIs if their distance is greater than a half wavelength, which is around several centimeters for WiFi sources. On the other hand, two clients can receive a shared packet approximately at the same time, even if they are more than 30m away. Therefore, we perform a fine-range proximity test by taking into account the RSSIs of the ambient signals and implement a large-range test based on the normalized packet arrival time.

The RSSI-based proximity test has been given in Algorithm 1, where the range granularity is determined by the thresholds in the post-IGMM process. In general, the range granularity decreases with the threshold Θ. The thresholds are determined according to the proximity range via training in the similar environments.

As comparison, we also propose a simplified version of the proximity-based authentication strategy. As described in Algorithm 3, this strategy is based on the RSSI information of the ambient radio signals and applies the Euclidean distance method for classification. By skipping the IGMM process of Algorithm 1, this strategy reduces the system overhead and complexity.

Moreover, we also propose an authentication strategy by exploiting the packet arrival time feature of the ambient signals. As shown in Fig. 7(b), the key generation rate of Algorithm 2 decreases smoothly and approximately monotonically with the client distance. Therefore, Alice can evaluate the key generation performance of Algorithm 2 to perform the proximity-based authentication. More specifically, Alice compares her key generation rate with a threshold denoted as Ξ: she believes that Bob is in her proximity if her key generation rate is higher than Ξ, and rejects Bob if otherwise.

As will be shown in the experimental results in Section VI, the packet arrival time-based authentication strategy can control the proximity range more flexibly. In that strategy, the coverage range that is more than 50 meters for WiFi signals is much larger than the proximity range of the method in [15], which is around several centimeters. Thus for a large proximity range (e.g., a WiFi-based proximity test with 50m proximity range), Alice chooses the key generation rate of Algorithm 2 instead of Algorithm 1 in the proximity-based authentication. On the other hand, if Alice's proximity range is short, Algorithm 1 that is based on RSSIs achieves a higher authentication accuracy.

Algorithm 3 Simplified Proximity-Based authentication

```
Input: RSSI measurements $\mathbf{x} = [x_i]_{1 \le i \le n}$
Output: Authentication result
  $c_i = i, \forall 1 \le i \le n$
  Update $\mathbf{c}$ by combining the classes whose centroid Euclidean distance is less than $\Theta$
  $C_A \leftarrow (23)$
  $j \leftarrow 0$
  for $i \leftarrow N+1$ to $2N$ do
    if $c_i = C_A$ then
      Alice accepts the packet, $j++$
    end if
  end for
  Proximity passing rate $\nu \leftarrow j/N$
  if $\nu > \Delta$ then
    Bob passes the authentication
  else
    Bob fails the authentication
  end if
```

TABLE II
Range Control by Selecting Different Ambient Radio Sources in the Proximity-Based Security System

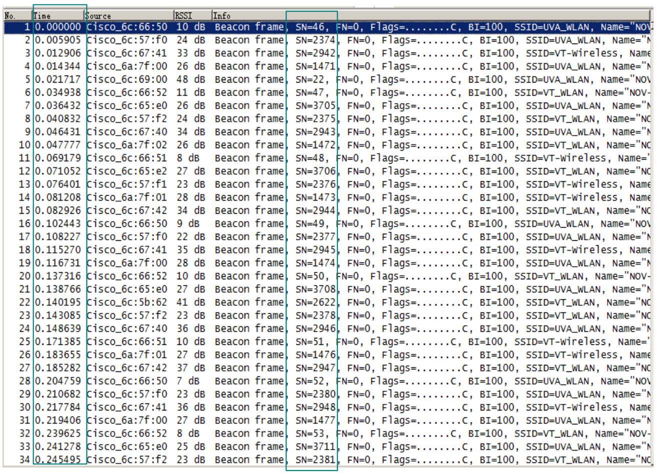
System	Bluetooth	WLAN	GSM	FM radio
Frequency (Hz)	2.4G	2.4,5G	.9/1.8G	87.5-108M
Range (m)	∼10	∼35	∼30k	> 100 k

B. Security and Performance Analysis

The proximity-based security technique is robust against the eavesdropper whose goal is to locate clients. As shown in Fig. 1, all that eavesdroppers can capture are the indices **J** and Bob's public location tag that consists of the RSSIs, SNs and MAC addresses of the ambient packets. Since neither of them directly discloses Bob's location, this system can protect the location privacy.

As shown in [17], existing key generation strategies that are based on the RSSI and channel impulse response (CIR) [23], [29]–[32] or the phase [33] are vulnerable to the man-in-the-middle attacks. For instance, eavesdroppers can reveal 40% to 50% of the keys, and attackers can sabotage the key agreements with 95% confidence by injecting spoofing signals during less than 4% of the overall communication duration [17].

Fortunately, man-in-the-middle attacks out of the proximity can be addressed in the proposed key establishment system by exploiting the packet arrival time. Because of the packet loss due to the channel fading that decorrelates fast over space, it is highly challenging for an attacker outside the proximity



```
No.  Time      Source             RSSI   Info
  1 0.000000 Cisco_6c:66:50  10 dB  Beacon frame, SN=46, FN=0, Flags=........C, BI=100, SSID=UVA_WLAN, Name="NOV
  2 0.005905 Cisco_6c:57:f0  24 dB  Beacon frame, SN=2374, FN=0, Flags=........C, BI=100, SSID=UVA_WLAN, Name="N
  3 0.012906 Cisco_6c:67:41  33 dB  Beacon frame, SN=2942, FN=0, Flags=........C, BI=100, SSID=VT-Wireless, Name
  4 0.014344 Cisco_6a:7f:00  26 dB  Beacon frame, SN=1471, FN=0, Flags=........C, BI=100, SSID=UVA_WLAN, Name="N
  5 0.021717 Cisco_6c:69:00  48 dB  Beacon frame, SN=22, FN=0, Flags=........C, BI=100, SSID=UVA_WLAN, Name="NOV
  6 0.034938 Cisco_6c:66:52  11 dB  Beacon frame, SN=47, FN=0, Flags=........C, BI=100, SSID=VT_WLAN, Name="NOV-
  7 0.036432 Cisco_6c:65:e0  26 dB  Beacon frame, SN=3705, FN=0, Flags=........C, BI=100, SSID=UVA_WLAN, Name="N
  8 0.040832 Cisco_6c:57:f2  24 dB  Beacon frame, SN=2375, FN=0, Flags=........C, BI=100, SSID=VT_WLAN, Name="NC
  9 0.046431 Cisco_6c:67:40  34 dB  Beacon frame, SN=2943, FN=0, Flags=........C, BI=100, SSID=UVA_WLAN, Name="N
 10 0.047777 Cisco_6a:7f:02  26 dB  Beacon frame, SN=1472, FN=0, Flags=........C, BI=100, SSID=VT_WLAN, Name="NC
 11 0.069179 Cisco_6c:66:51  8 dB   Beacon frame, SN=48, FN=0, Flags=........C, BI=100, SSID=VT-Wireless, Name='
 12 0.071052 Cisco_6c:65:e2  27 dB  Beacon frame, SN=3706, FN=0, Flags=........C, BI=100, SSID=VT_WLAN, Name="NC
 13 0.076401 Cisco_6c:57:f1  23 dB  Beacon frame, SN=2376, FN=0, Flags=........C, BI=100, SSID=VT-Wireless, Name
 14 0.081208 Cisco_6a:7f:01  28 dB  Beacon frame, SN=1473, FN=0, Flags=........C, BI=100, SSID=VT-Wireless, Name
 15 0.082926 Cisco_6c:67:42  34 dB  Beacon frame, SN=2944, FN=0, Flags=........C, BI=100, SSID=VT_WLAN, Name="NC
 16 0.102443 Cisco_6c:66:50  9 dB   Beacon frame, SN=49, FN=0, Flags=........C, BI=100, SSID=UVA_WLAN, Name="NOV
 17 0.108227 Cisco_6c:57:f0  22 dB  Beacon frame, SN=2377, FN=0, Flags=........C, BI=100, SSID=UVA_WLAN, Name="N
 18 0.115270 Cisco_6c:67:41  35 dB  Beacon frame, SN=2945, FN=0, Flags=........C, BI=100, SSID=VT-Wireless, Name
 19 0.116731 Cisco_6a:7f:00  28 dB  Beacon frame, SN=1474, FN=0, Flags=........C, BI=100, SSID=UVA_WLAN, Name="N
 20 0.137316 Cisco_6c:66:52  10 dB  Beacon frame, SN=50, FN=0, Flags=........C, BI=100, SSID=VT_WLAN, Name="NOV-
 21 0.138766 Cisco_6c:65:e0  27 dB  Beacon frame, SN=3708, FN=0, Flags=........C, BI=100, SSID=UVA_WLAN, Name="N
 22 0.140195 Cisco_6c:5b:62  41 dB  Beacon frame, SN=2622, FN=0, Flags=........C, BI=100, SSID=VT_WLAN, Name="NC
 23 0.143085 Cisco_6c:57:f2  23 dB  Beacon frame, SN=2378, FN=0, Flags=........C, BI=100, SSID=VT_WLAN, Name="NC
 24 0.148639 Cisco_6c:67:40  36 dB  Beacon frame, SN=2946, FN=0, Flags=........C, BI=100, SSID=UVA_WLAN, Name="N
 25 0.171385 Cisco_6c:66:51  10 dB  Beacon frame, SN=51, FN=0, Flags=........C, BI=100, SSID=VT-Wireless, Name='
 26 0.183655 Cisco_6a:7f:01  27 dB  Beacon frame, SN=1476, FN=0, Flags=........C, BI=100, SSID=VT-Wireless, Name
 27 0.185282 Cisco_6c:67:42  37 dB  Beacon frame, SN=2947, FN=0, Flags=........C, BI=100, SSID=VT_WLAN, Name="NC
 28 0.204759 Cisco_6c:66:50  7 dB   Beacon frame, SN=52, FN=0, Flags=........C, BI=100, SSID=UVA_WLAN, Name="NOV
 29 0.210682 Cisco_6c:57:f0  23 dB  Beacon frame, SN=2380, FN=0, Flags=........C, BI=100, SSID=UVA_WLAN, Name="N
 30 0.217784 Cisco_6c:67:41  36 dB  Beacon frame, SN=2948, FN=0, Flags=........C, BI=100, SSID=VT-Wireless, Name
 31 0.219406 Cisco_6a:7f:00  27 dB  Beacon frame, SN=1477, FN=0, Flags=........C, BI=100, SSID=UVA_WLAN, Name="N
 32 0.239625 Cisco_6c:66:52  8 dB   Beacon frame, SN=53, FN=0, Flags=........C, BI=100, SSID=VT_WLAN, Name="NOV-
 33 0.241278 Cisco_6c:65:e0  25 dB  Beacon frame, SN=3711, FN=0, Flags=........C, BI=100, SSID=UVA_WLAN, Name="N
 34 0.245495 Cisco_6c:57:f2  23 dB  Beacon frame, SN=2381, FN=0, Flags=........C, BI=100, SSID=VT_WLAN, Name="NC
```

Fig. 3. Sequence numbers and MAC addresses of the ambient WiFi signals captured by wireless adapters *AirPcap Nx* and open-source packet analyzers *Wireshark* in an experiment.

to estimate the exact ambient packet arrival time sequence of a client, if there are *multiple ambient radio sources*, which is true in most indoor environments. For example, Fig. 3 presents a packet arrival sequence captured by a client with a wireless adapter in an experiment, showing the difficulty in estimating the exact SN sequence over time and thus the corresponding packet arrival time. This system never broadcasts the packet arrival time information over the air. Therefore, eavesdroppers outside the proximity cannot derive the pairwise session key between Alice and Bob.

Next, we consider attackers who spoof ambient radio sources by injecting faked or replay signals in hopes of significantly increasing the key disagreement rate between Alice and Bob in Algorithm 2. Note that the actual ambient radio source and the attacker usually result in different RSSIs in their signals due to distinct locations. Therefore, the faked packets can hardly pass the proposed proximity-based authentication, and thus are discarded in the session key generation using Algorithm 2. In addition, even with the knowledge of the past RSSI information, attackers still have difficulty in estimating the current RSSI obtained by the radio client due to the random time variation of RSSIs. Consequently, the proposed authentication strategy can also filter out the relayed messages.

Finally, compared with the time-variant RSSI or CIR, the packet arrival time has much higher entropy and is less sensitive to the radio propagation pattern. Therefore, as will be shown in the experimental results, this system can generate session keys much faster, and control the proximity range more flexible than the RSS-based key generation strategies such as [23]. Moreover, by introducing the IGMM method and exploiting the packet arrival time information, this security system provides more accurate authentication with flexible range control for larger coverage area than the strategies in [13]–[15]. More in-depth analysis of the security performance will be performed in our future work.

VI. Experimental Results

We performed experiments in Virginia Tech Northern Virginia Center to evaluate the performance of this system.

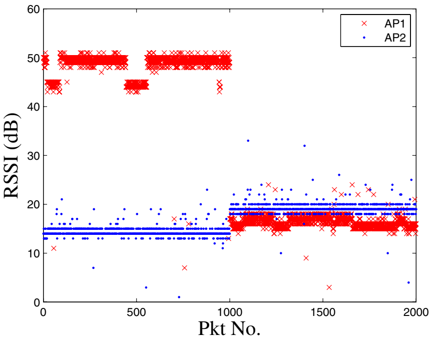
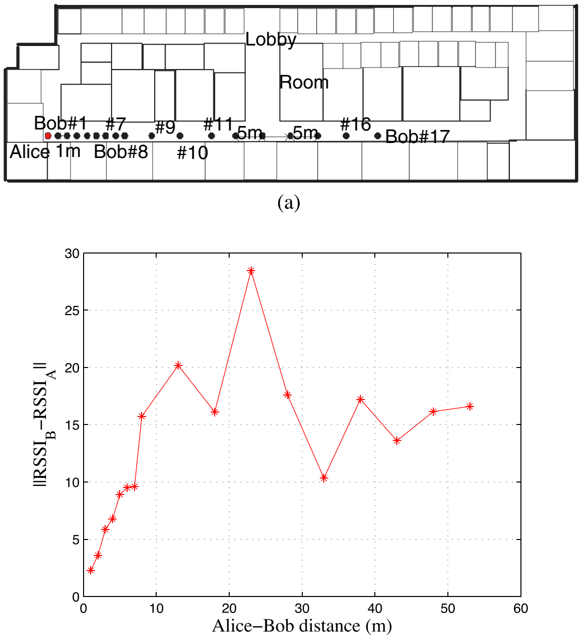


Fig. 4. RSSI trace with $D = 2$ and $N = 2$, as the input of Algorithm 1.



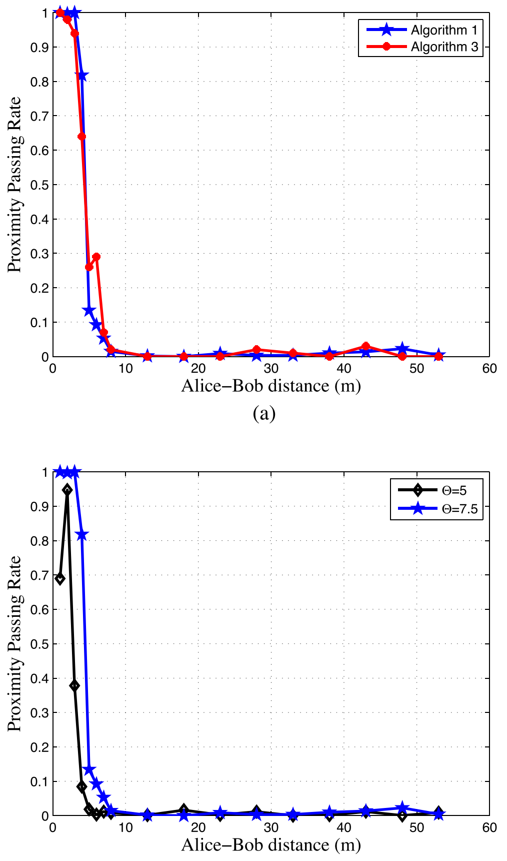
(a)

(b)

Fig. 5. (a) Client placements. (b) Example of the average difference between the RSSI vectors observed by Alice and Bob. Settings of Experiment 1 performed in Virginia Tech Northern Virginia Center.

As shown in Fig. 5 and Fig. 8, two laptops acting as Alice and Bob, respectively, were placed in different locations in the 2nd floor of the building. Utilizing wireless adapters *AirPcap Nx* and open-source packet analyzers *Wireshark*, both laptops simultaneously captured the ambient WiFi signals. Although the experiments were based on WiFi, the proposed strategy can be easily extended to the case with multiple types of radio sources such as FM and Bluetooth ambient signals.

In each scenario, clients extracted the RSSI, packet arrival time, SN and MAC addresses of the ambient beacon frames at 2.417 GHz, and recorded the trace for one minute. Both clients recorded the RSSIs from $D = 2$ ambient WiFi APs. An example of the measured RSSI vectors is presented in Fig. 4, where the first $N = 1000$ data were observed by Alice, while the following 1000 vectors were reported by Bob. Clearly, the RSSI vectors variant over time.



(a)

(b)

Fig. 6. (a) Proximity passing rate of Algorithm 1 and 3 with $\Theta = 7.5$. (b) Proximity passing rate of Algorithm 1 with $\Theta = 5$ and 7.5. Performance of the proximity-based authentication in Experiment 1.

A. Proximity-Based Authentication

The settings of the first experiment with 17 scenarios are shown in Fig. 5, where Bob was placed in different locations along the hallway. Both clients recorded the RSSI from 2 ambient WiFi APs. An example of the difference between the ambient RSSI vectors obtained by Alice and Bob is presented in Fig. 5(b), showing that the average RSSI difference often increases with the distance between Alice and Bob, especially when the distance between Alice and Bob is less than 15m. On the other hand, their relationship is in general complicated, as RSSI also depends on the transmitter location and the specific radio environment.

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

We present the probability for Bob to pass the proximity test by Alice in different scenarios for both Algorithm 1 with the threshold $\Theta = 7.5$ and Algorithm 3. As illustrated in Fig. 6(a), Alice can accurately determine whether Bob is in her proximity with the 4m proximity range. For example, the false rejection rate of Algorithm 1 is very small if the Alice-Bob distance is less than 3m. In this case, the false acceptance rate is less than 5% when the distance between Alice and Bob is larger than 6m, and is very small when the Alice-Bob distance is more than 10m. We also provide the performance of Algorithm 1 with different Θ in Fig. 6(b), showing that $\Theta = 7.5$ is a good heuristic choice for the authentication with the 4m proximity range.

Compared with Algorithm 3, the NPB-based strategy, Algorithm 1, is more stable in both the rejection region and the passing region, and has a narrower transition region. For example, the Type 1 error rate of Algorithm 1 is more than 5% lower than Algorithm 3, when the Alice-Bob distance is 2m and the proximity range is 3m. Meanwhile, Algorithm 1 rejects the clients outside the proximity more accurately. For instance, the Type 2 error of Algorithm 1 is about 20% lower than Algorithm 3, when the Alice-Bob distance is 6m and the proximity range is 3m. On the other hand, Algorithm 3 also works well when the Alice-Bob distance is much larger than the proximity range (e.g., the proximity range and Alice-Bob distance are 3m and 40m, respectively), as shown in Fig. 6(a).

B. Key Generation Performance and Range Control

We use two criteria to evaluate the performance of the session key establishment: (1) the key generation rate that is the speed for Alice to generate $\mathbf{K}_A$ in bits per second, and (2) the key disagreement rate defined as the percentage of bits in Alice's key ($\mathbf{K}_A$) that are different from Bob's ($\mathbf{K}_B$).

Fig. 7 provides the performance of Algorithm 2 in Experiment 1, with the time rounding parameter $\Upsilon = 1, 2$ and 3. It is shown in Fig. 7 that $\Upsilon = 2$ achieves both a high key generation rate and low key mismatching rate for all 17 scenarios. For instance, the lowest key generation rate is about 100 bps and the key disagreement rate is no more than 4%, if the Alice-Bob distance ranges between 1m and 55m. With such a low error rate, the key disagreement can be conveniently addressed by the error correction codes such as BCH.

Next, as shown in Fig. 7(b), the key generation rate decreases smoothly and slowly with the Alice-client distance. That is because clients in different areas see different ambient packet arrival sequences and thus packet arrival time sequences, in presence of multiple ambient radio sources as is typical in indoor environments. For instance, the key generation rate is above 100 bps even when Bob is about 50m away from Alice and the key disagreement rate is less than 4%. Therefore, the key generation rate of Algorithm 2 can be used by Alice to determine whether Bob is in her proximity.

The maximum proximity range of the authentication based on the packet arrival time is much larger than that of the RSSI-based strategies. For example, Alice can authenticate clients as far as 50m away by comparing the key generation rate of Algorithm 2 with the threshold Ξ. The parameter settings in Experiment 1 are listed in Table III, with proximity range changing from 3m to 50m. The system parameters, Δ and Ξ, are chosen according to the specified proximity range in the experiment.

Compared with most existing work, the proposed strategy provides a much larger maximum proximity range than most

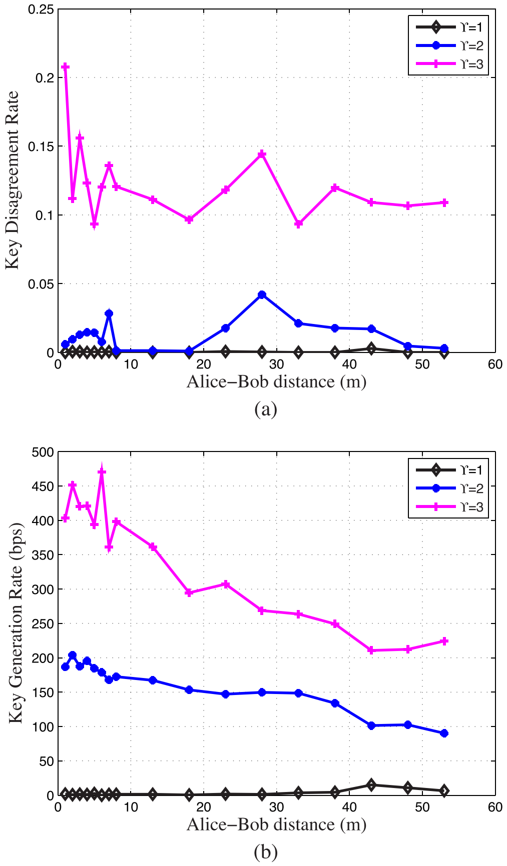


Fig. 7. (a) Key disagreement rate between $\mathbf{K}_A$ and $\mathbf{K}_B$. (b) Key generation rate of $\mathbf{K}_A$. Performance of the key generation algorithm (Algorithm 2), whose locations are shown in Fig. 5, with $\Upsilon = 1$, 2 and 3 (rounding to 0.1s, 0.01s and 1 ms).

TABLE III
PROXIMITY CONTROL IN THE PROPOSED AUTHENTICATION METHOD IN EXPERIMENT 1

Proximity range (m)	3	4	5	6
Threshold Δ in Alg. 1 ($\Theta = 7.5$)	.9	.5	.1	.05

Proximity range (m)	10	20	40	50
Threshold Ξ	160	150	125	100

existing work. More specifically, considering 2.4GHz WiFi ambient signals, the maximum proximity range of this strategy is around 50m, while the maximum proximity ranges supported by ProxiMate in [15] and Ensemble in [14] are only 6.25 cm and 2m, respectively. Moreover, this scheme provides higher key generation rates than ProxiMate. For example, in typical indoor environments, the key generation rate of this scheme, which is around 200bps, is much higher than the 13 bps rate of ProxiMate in [15]. In addition, this scheme also provides more accurate authentication than Ensemble. For example, as shown in Section VI, this scheme has a small false rejection rate for clients within 3m from Alice and false acceptance rate for clients more than 10 m away, which outperforms the 0.19 false rejection rate of Ensemble [14].

C. Room-Based Proximity Test

Experiment 2 contained six scenarios, with topology illustrated in Fig. 5(a). In this experiment, Alice performed Algorithm 1 to decide whether Bob is in the same office. The performance of the room-based proximity test is presented in Fig. 8(b), showing that the error rates for Alice to find a same-room client are mostly below 15%. We have also found that the ambient packet matching ratio is mostly above 40% when Alice and Bob are in the same room, or above 25% when they are in different rooms. The results indicate that both clients have plenty of shared ambient packets to build the session key. Finally, we can see that the lowest session key generation rate is approximately as high as 248 bps. More details are given in [28].

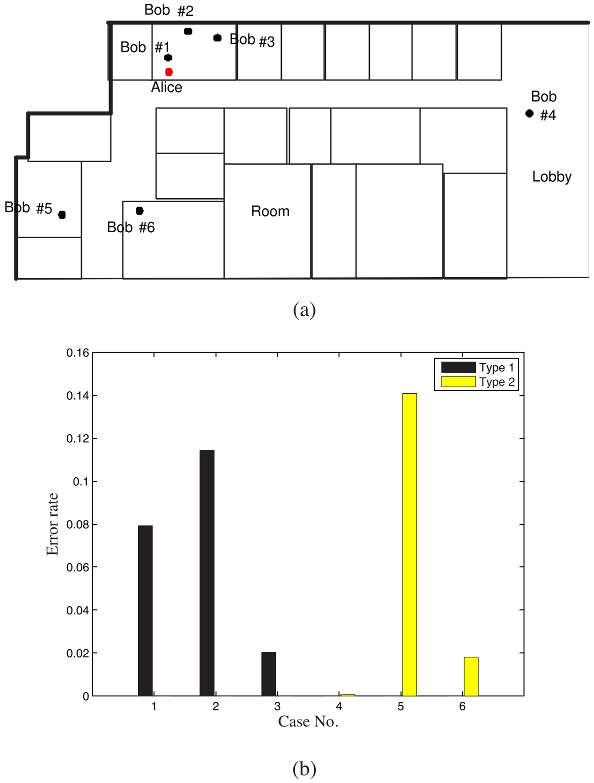


Fig. 8. (a) Client placements in Virginia Tech Northern Virginia Center. (b) Error rates of the proximity test with $\Theta = 7.5$. Performance of the proximity-based authentication in Experiment 2.

Finally, we note that this work cannot achieve zero error rates, just like the other PHY-layer security schemes due to the properties of radio propagation. However, it can be used to enhance the security of LBS in wireless networks. For example, the proposed strategy provides a lightweight security protection for the LBS applications that do not require zero error rates in a wireless network without any pre-shared secret, trusted authority or public key infrastructure. On the other hand, for the applications with strict security requirements, the proposed scheme can serve as the bootstrap for the establishment of secure connections among the clients in the proximity and be incorporated with existing traditional security methods to achieve “100% security”.

VII. CONCLUSION

We have proposed a proximity-based authentication and key establishment scheme by exploiting the physical-layer features of ambient radio signals for LBS services in wireless networks, without requiring any pre-shared secret. Flexible range control is achieved by selecting the appropriate radio sources, such as ambient WiFi access points (APs), bluetooth devices and FM radios and choosing their suitable physical-layer features.

The system applies the Markov chain Monte Carlo implementation of the infinite Gaussian mixture model (IGMM) to classify the RSSIs of multiple ambient signals and thus determines whether a client is in the proximity. In the key establishment, clients generate session keys based on the normalized arrival time of their shared ambient packets.

The system does not disclose the client locations, and is robust against eavesdropping, spoofing, replay attacks and man-in-the-middle attacks outside the proximity. Experiments using laptops with WiFi packet analyzers in typical indoor environments have verified the efficacy of the security technique. By applying the IGMM model, the authentication is more accurate and is less sensitive to the radio propagation pattern than existing RSS and CIR-based authentication strategies. The key generation rate that can be as high as 248 bps in ideal cases is much higher than that of the RSS-based strategies. In the future, we will further evaluate the performance of the proposed strategy with experiments based on FM, Bluetooth and WiFi ambient signals and study how to incorporate this PHY-layer security strategy with the existing traditional security protocols to address the man-in-the-middle attacks inside the proximity.

References

[1] M. Li, W. Lou, and K. Ren, "Data security and privacy in wireless body area networks," *IEEE Wireless Commun.*, vol. 17, no. 2, pp. 51–58, Feb. 2010.

[2] X. Liang, R. Lu, C. Le, X. Lin, and X. Shen, "PEC: A privacy-preserving emergency call scheme for mobile healthcare social networks," *J. Commun. Netw.*, vol. 13, no. 2, pp. 102–112, Apr. 2011.

[3] A. Narayanan and V. Shmatikov, "De-anonymizing social networks," in *Proc. 30th IEEE Symp. Sec. Privacy*, May 2009, pp. 173–187.

[4] J. Tsai, P. Kelley, L. Cranor, and N. Sadeh, "Location-sharing technologies: Privacy risks and controls," *ISJLP*, vol. 6, pp. 119–317, Aug. 2009.

[5] G. Ghinita, P. Kalnis, A. Khoshgozaran, C. Shahabi, and K. Tan, "Private queries in location based services: Anonymizers are not necessary," in *Proc. ACM SIGMOD Int. Conf. Manag. Data*, Jun. 2008, pp. 121–132.

[6] W. He, X. Liu, and M. Ren, "Location cheating: A security challenge to location-based social network services," in *Proc. IEEE ICDCS*, Feb. 2011, pp. 1–10.

[7] W. Chang, J. Wu, and C. Tan, "Enhancing mobile social network privacy," in *Proc. IEEE Global Telecommun. Conf.*, Dec. 2011, pp. 1–5.

[8] Z. Zhu and G. Cao, "Applaus: A privacy-preserving location proof updating system for location-based services," in *Proc. IEEE INFOCOM*, Apr. 2011, pp. 1889–1897.

[9] L. Siksnys, J. Thomsen, S. Saltenis, M. Yiu, and O. Andersen, "A location privacy aware friend locator," in *Advances in Spatial and Temporal Databases*. New York, NY, USA: Springer-Verlag, 2009, pp. 405–410.

[10] A. Narayanan, N. Thiagarajan, M. Lakhani, M. Hamburg, and D. Boneh, "Location privacy via private proximity testing," in *Proc. NDSS*, 2011, pp. 1–17.

[11] N. Talukder and S. Ahamed, "Preventing multi-query attack in location-based services," in *Proc. 3rd ACM Int. Conf. Wireless Netw. Sec.*, 2010, pp. 25–36.

[12] R. Mayrhofer and H. Gellersen, "Shake well before use: Intuitive and secure pairing of mobile devices," *IEEE Trans. Mobile Comput.*, vol. 8, no. 6, pp. 792–806, Jun. 2009.

[13] A. Varshavsky, A. Scannell, A. LaMarca, and E. Lara, "Amigo: Proximity-based authentication of mobile devices," in *Proc. Int. Conf. Ubiquitous Comput.*, 2007, pp. 1–18.

[14] A. Kalamandeen, A. Scannell, E. de Lara, A. Sheth, and A. LaMarca, "Ensemble: Cooperative proximity-based authentication," in *Proc. ACM Int. Conf. Mobile Syst., Appl. Services*, Jun. 2010, pp. 331–344.

[15] S. Mathur, R. Miller, A. Varshavsky, and W. Trappe, "ProxiMate: Proximity-based secure pairing using ambient wireless signals," in *Proc. ACM Int. Conf. Mobile Syst., Appl., Services*, 2011, pp. 1–14.

[16] Y. Zheng, M. Li, W. Lou, and T. Hou, "Sharp: Private proximity test and secure handshake with cheat-proof location tags," in *Proc. ESORICS*, 2012, pp. 361–378.

[17] S. Eberz, M. Strohmeier, M. Wilhelm, and I. Martinovic, "A practical man-in-the-middle attack on signal-based key generation protocols," in *Proc. ESORICS*, 2012, pp. 235–252.

[18] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005, ch. 3.

[19] C. Rasmussen, "The infinite Gaussian mixture model," in *Advances in Neural Information Processing Systems*. Cambridge, MA, USA: MIT Press, 2000, pp. 554–560.

[20] N. Nguyen, G. Zheng, Z. Han, and R. Zheng, "Device fingerprinting to enhance wireless security using nonparametric Bayesian method," in *Proc. IEEE INFOCOM*, Apr. 2011, pp. 1404–1412.

[21] N. Nguyen, R. Zheng, and Z. Han, "On identifying primary user emulation attacks in cognitive radio systems using nonparametric Bayesian classification," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1432–1445, Mar. 2012.

[22] C. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer-Verlag, 2006.

[23] S. Mathur, W. Trappe, N. Mandayam, C. Ye, and A. Reznik, "Radio-telepathy: Extracting a secret key from an unauthenticated wireless channel," in *Proc. ACM 14th Annu. Conf. Mobile Comput. Syst.*, 2008, pp. 128–139.

[24] Z. Lin, D. Kune, and N. Hopper, "Efficient private proximity testing with GSM location sketches," in *Financial Cryptography Data Security* (Lecture Notes in Computer Science). New York, NY, USA: Springer-Verlag, 2012, pp. 73–88.

[25] S. Mascetti, C. Bettini, D. Freni, X. Wang, and S. Jajodia, "Privacy-aware proximity based services," in *Proc. Int. Conf. Mobile Data Manag., Syst., Services Middlew.*, May 2009, pp. 441–444.

[26] J. Meyerowitz and R. Roy Choudhury, "Hiding stars with fireworks: Location privacy through camouflage," in *Proc. Int. Conf. Mobile Comput. Netw.*, 2009, pp. 345–356.

[27] L. Siksnys, J. Thomsen, S. Saltenis, and M. Yiu, "Private and flexible proximity detection in mobile social networks," in *Proc. 11th Int. Conf. Mobile Data Manag.*, May 2010, pp. 75–84.

[28] L. Xiao, Q. Yan, W. Lou, and T. Hou, "Proximity-based security using ambient radio signals," in *Proc. IEEE ICC*, Jul. 2011, pp. 211–224.

[29] B. Azimi, A. Kiayias, A. Mercado, and B. Yener, "Robust key generation from signal envelopes in wireless networks," in *Proc. 14th ACM Conf. Comput. Commun. Sec.*, 2007, pp. 401–410.

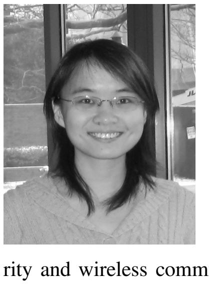
[30] C. Ye, S. Mathur, A. Reznik, Y. Shah, W. Trappe, and N. Mandayam, "Information-theoretically secret key generation for fading wireless channels," *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 2, pp. 240–254, Jun. 2010.

[31] T. Aono, K. Higuchi, T. Ohira, B. Komiyama, and H. Sasaoka, "Wireless secret key generation exploiting reactance-domain scalar response of multipath fading channels," *IEEE Trans. Antennas Propag.*, vol. 53, no. 11, pp. 3776–3784, Nov. 2005.

[32] J. Croft, N. Patwari, and S. Kasera, "Robust uncorrelated bit extraction methodologies for wireless sensors," in *Proc. ACM/IEEE Int. Conf. IPSN*, Apr. 2010, pp. 70–81.

[33] Q. Wang, H. Su, K. Ren, and K. Kim, "Fast and scalable secret key generation exploiting channel phase randomness in wireless networks," in *Proc. IEEE INFOCOM*, Apr. 2011, pp. 1422–1430.

[34] L. Xiao, L. Greenstein, N. B. Mandayam, and W. Trappe, "Channel-based spoofing detection in frequency-selective Rayleigh channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 12, pp. 5948–5956, Dec. 2009.



Liang Xiao (M'09–SM'13) received the B.S. degree in communication engineering from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2000, the M.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 2003, and the Ph.D. degree in electrical engineering from Rutgers University, New Brunswick, NJ, USA, in 2009. She is currently an Associate Professor in the Department of Communication Engineering, Xiamen University, Xiamen, China. Her research interests include network security and wireless communications.

Qiben Yan (S'11) received the B.E. and M.E. degrees in electrical engineering from Fudan University, Shanghai, China, in 2007 and 2010, respectively. He is currently pursuing the Ph.D. degree with the Computer Science Department, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA. His current research interests include wireless network security and privacy, network monitoring and forensics, botnet detection, intrusion and anomaly detection, and cloud and software-defined networking security.

Wenjing Lou (M'03–SM'08) is an Associate Professor with the Virginia Polytechnic Institute and State University (Virginia Tech), Blacksburg, VA, USA. Prior to joining Virginia Tech in 2011, she was a faculty member at the Worcester Polytechnic Institute from 2003 to 2011. She received the Ph.D. degree in electrical and computer engineering from the University of Florida, Gainesville, FL, USA, in 2003. Her current research interests are in cyber security, with emphases on wireless network security and data security and privacy in cloud computing. She was a recipient of the U.S. National Science Foundation CAREER Award in 2008.

Guiquan Chen (S'13) received the B.S. degree in communication engineering from Jimei University, Xiamen, China, in 2012. He is currently pursuing the Graduate degree with the Department of Communication Engineering, Xiamen University, Xiamen, China. His research interests include network security and wireless communications.

Y. Thomas Hou (S'91–M'98–SM'04) is a Professor in the Bradley Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA. His research interests are cross-layer optimization for wireless networks and wireless security. He has published extensively in leading journals and top-tier conferences and received five Best Paper Awards from the IEEE (including the IEEE INFOCOM 2008 Best Paper Award and IEEE ICNP 2002 Best Paper Award) and one Distinguished Paper Award from ACM. He is currently serving as an Area Editor of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, an Associate Editor of IEEE TRANSACTIONS ON MOBILE COMPUTING, an Editor of IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS (Cognitive Radio Series), and an Editor of IEEE WIRELESS COMMUNICATIONS. He is the Chair of the IEEE INFOCOM Steering Committee.