S2M: A Lightweight Acoustic Fingerprints-Based Wireless Device Authentication Protocol

Dajiang Chen, *Member, IEEE*, Ning Zhang, *Member, IEEE*, Zhen Qin, *Member, IEEE*, Xufei Mao, *Member, IEEE*, Zhiguang Qin, *Member, IEEE*, Xuemin Shen, *Fellow, IEEE*, and Xiang-Yang Li, *Fellow, IEEE*

***Abstract*—Device authentication is a critical and challenging issue for the emerging Internet of Things (IoT). One promising solution to authenticate IoT devices is to extract a fingerprint to perform device authentication by exploiting variations in the transmitted signal caused by hardware and manufacturing inconsistencies. In this paper, we propose a lightweight device authentication protocol [named speaker-to-microphone (S2M)] by leveraging the frequency response of a speaker and a microphone from two wireless IoT devices as the acoustic hardware fingerprint. S2M authenticates the legitimate user by matching the fingerprint extracted in the learning process and the verification process, respectively. To validate and evaluate the performance of S2M, we design and implement it in both mobile phones and PCs and the extensive experimental results show that S2M achieves both low false negative rate and low false positive rate in various scenarios under different attacks.**

***Index Terms*—Acoustic hardware fingerprinting, device authentication, Internet of Things (IoT), wireless security.**

I. Introduction

A LARGE number of objects are expected to be connected to Internet, such as smart phones, vehicles, sensors, and wearable devices, giving rise to a new era of Internet of Things (IoT) [1]. The IoT enables physical objects to sense, communicate, and perform certain actions on demand by integrating embedded systems, sensors, actuators, communication capabilities, and so on. It can facilitate a multitude of applications, including smart home, smart city [2], and intelligent transportation system [3], which will have a profound effect in almost every aspects of our daily life. Along with the advantages of IoT is the security issue [4]–[6]. With IoT, the security threats are even extended from the cyber world to cyber-physical world. In such a context, the attackers can not only compromise information but also manipulate the physical objects, which will severely influence the physical world.

One of the fundamental mechanisms to secure IoT is device authentication, which can verify the identity of IoT entities and control their access. The main reason lies in the fact that all the identity-based attacks (e.g., MAC address spoofing) are usually considered as the first step of many other attacks [7]–[9], such as man-in-the-middle attack, data modification, and denial of service attack. Traditional solutions are mainly based on cryptographic systems, such as WPA, WPA2 (802.11i), and 802.11w, which however have a track record of being compromised [10]–[12]. Moreover, the IoT devices usually have only limited computing resources, which might not be able to support some complex cryptographic mechanisms.

Alternatively, some promising approaches [13]–[23] have been proposed which exploit physical layer characteristics to enhance device identity authentication. As mentioned in [24], the existing physical-layer device identification schemes can be roughly classified into three categories: 1) software-based [13], [14]; 2) channel fingerprinting-based [15]–[20]; and 3) hardware fingerprinting-based ones [21]–[23]. Unfortunately, the existing approaches suffer from the following major shortcomings: 1) software-based protocols cannot distinguish between different physical devices running the same software; 2) channel fingerprinting-based approaches may not work well in a highly dynamic environment, where the channel state or received signal strength (RSS) changes dramatically over time; and 3) most hardware fingerprinting-based schemes still suffer mimic attacks (details in Section II).

Considering that audio communication (including audio-based near field communication [25]) is increasingly mature and commercialized, many existing wireless devices are equipped with microphones and speakers, such as smart

Manuscript received June 28, 2016; revised September 26, 2016; accepted October 17, 2016. Date of publication October 20, 2016; date of current version February 8, 2017. This work was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada, by the National 973 Program of China under Grant 2013CB834203, by the Major International (Regional) Joint Research Project of China National Science Foundation under Grant 61520106007, by the NSFC under Grant 61502085 and Grant 61133016, and by the China Postdoctoral Science Foundation project under Grant 2015M570775. The work of X.-Y. Li was supported in part by the NSF CNS 1526638, Key Research Program of Frontier Sciences, CAS. No. QYZDYSSWJSC002.

D. Chen is with the School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China, and also with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: dajiang.chen@uwaterloo.ca).

N. Zhang and X. Shen are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: n35zhang@uwaterloo.ca; sshen@uwaterloo.ca).

Z. Qin and Z. Qin are with the School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: zhenqin@uestc.edu.cn; zgqin@uestc.edu.cn).

X. Mao is with the School of Software and TNLIST, Tsinghua University, Beijing, China (e-mail: xufei.mao@gmail.com).

X.-Y. Li is with the Department of Computer Science, University of Science and Technology of China, 230000 Hefei, China (e-mail: xiangyangli@ustc.edu.cn).

Digital Object Identifier 10.1109/JIOT.2016.2619679

phones, tablets, smart appliances, and connected vehicles. To complement the existing authentication solutions for IoT, in this paper, we design a wireless device identity authentication protocol utilizing acoustic hardware (speaker/microphone) fingerprints, named speaker-to-microphone (S2M). The proposed authentication mechanisms can be applied in various scenarios, such as mobile social networking, vehicle to vehicle communication, and smart home. Briefly speaking, S2M utilizes frequency response (FR) between the speaker/microphone of a pair of wireless devices as their acoustic hardware fingerprint [26]. Since the speaker/microphone components are typically designed for human speech, those components are designed for handling low-frequency sound waves rather than high-frequency ones. Moreover, due to the differences in the hardware manufacturing process and other uncontrollable factors, each microphone and speaker have unique defects and characteristics. In other word, there are not two speakers (microphones) are exactly the same. Moreover, we find that the differences are further enlarged when those acoustic hardwares are installed to different wireless devices so that it is possible for us to differentiate reflection in the FR. The key insights of our scheme are summarized as follows: 1) the same speaker/microphone pairs have a similar FR curve at any time and any place when some conditions are satisfied (details in Section IV); 2) different speaker/microphone pairs have different FR curves (details in Section IV); and 3) the properties of FR curve of speaker/microphone pairs cannot be replicated or copied from one device to another (details in Section VII-A).

We consider the following scenario to illustrate the main idea of S2M. Assume that a wireless device (named Alice) would like to authenticate herself to another one (named Bob) in the presence of an active adversary (named Eve). For authentication, the initiator Alice first establishes a connection with the authenticator Bob via the audio handshake method. Second, Alice generates the audio signals for authentication and transmits it to Bob through her speaker. After receiving the audio signals, Bob calculates the FR and extracts the fingerprint. Then Bob stores Alice's ID and the fingerprint. So far, the above series of operations could be considered as learning process. When identity authentication is required (i.e., in authentication process), Alice generates the audio signals for authentication and transmits them to Bob again through her speaker. After Bob receives, it will match the fingerprints from learning and authentication processes. If the output of the matching algorithm (MA) is positive, authentication is successful; otherwise, the authentication fails.

Compared with the state-of-the-art related work, the proposed protocol has higher security and lower requirement of hardware. Moreover, the protocol can be used for many wireless applications, such as wireless network access authentication and device authentication in audio payment system. However, there are four main issues need to be carefully addressed.

A. Generation and Transmission of the Audio Signals for Authentication With High Efficiency

As we know, Alice needs to send some audio signals to Bob for feature extraction. For instance, an intuitive approach is that Alice sends $S(f_i) = \sin(2\pi f_i t)$, $i \in \{1, \ldots, N\}$ to Bob in sequence, where $S(f_i)$ represents the amplitude of the transmitted acoustic pressure at frequency f_i. However, a larger value of N means longer transmission time; otherwise, a smaller value of N means a lower resolution of fingerprint. S2M addresses this problem by leveraging mix-frequency technology, which generates the mixed audio signals by combining various frequency components (details in Section V).

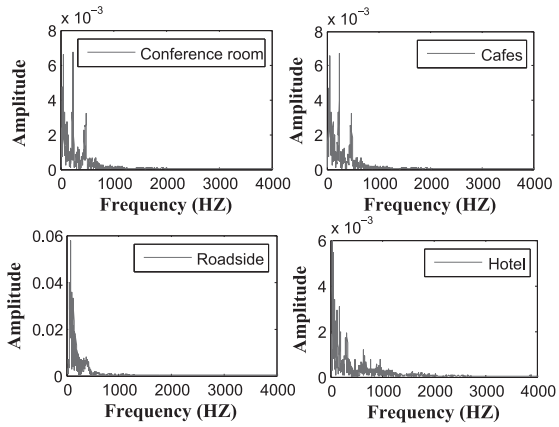


Fig. 1. Spectrum of ambient noise.

B. Elimination of the Negative Effects of Ambient Noise

In order to become an efficient and practical system with rigorous security, S2M must be proved to have a good performance in public spaces, such as conference rooms, cafes, roadsides, and hotels, where the ambient (acoustic) noise can cause significant effect on fingerprints extraction. To characterize the effect, we measure the received acoustic power in some typical environments, including conference rooms, hotels, malls, and cafes during busy hours. As shown in Fig. 1, the received power of ambient noises in all of these environments are negligible when the frequency reaches 3 kHz. Thus, in S2M, we can exploits the frequencies that are lager than 4 kHz for FR extraction using a high-pass filter.

C. Elimination of the Negative Effects of Echoes

Due to the adverse effects of multipaths, the fingerprints between leaning and authentication process are not identical. Thus, an MA is needed to eliminate the negative effects of echoes. To do this, we propose a new method, named deviation-ratio-based MA (D-MA) (details in Section V-C) with which S2M achieves a high successful authentication rate and a low false acceptance rate.

D. Elimination the Negative Effects of Distance and Detection the Active Attack

The S2M can work well when the fingerprint is learned and verified at the similar distance. It is likely to fail when the fingerprint is measured at one distance, while it is tested at a different distance. It also does not work well, when one or more active attackers are transmitting high frequency audio signals from their speakers. To solve these problems, we design an extended protocol based on S2M, named E-S2M,

which leverages a multiposition learning algorithm and an automatic fingerprint MA to realize variable distance authentication, and utilizes an active attack detection algorithm to detect an active attack (details in Section VI).

We summarize our main contributions as follows.

1) We propose a device authentication protocol, named S2M, which leverages the FR of acoustic hardware (speaker/microphone) as a fingerprint on a wireless device to authenticate the other one. Based on S2M, we design an extended protocol E-S2M to achieve variable distance authentication and active attack detection.
2) We evaluate S2M (rep. E-S2M) through extensive experiments using mobile phones in a number of scenarios under different attacks. The results show that the operating distance of S2M (rep. E-S2M) can reach up to 5 m, and the successful authentication rate can achieve 98% (rep. 97%). Experiments also show that both S2M and E-S2M achieve a low false acceptance rate (less than 2%).
3) We design and implement the software application of S2M and E-S2M which are suitable for operation on mobile phones and PCs.

This paper is organized as follows. We review the state-of-the-art related work in Section II. Section III introduces the acoustic attenuation model and adversarial models used in this paper. In Section IV, we analyze the characteristics of speaker and microphone pairs. We present S2M in Section V. An extended protocol E-S2M is presented in Section VI. Section VII gives the experimental studies. We conclude this paper in Section VIII.

II. Related Work

Channel fingerprinting-related device authentication is based on the fact that the channel state information (CSI) is location-specific due to path loss and channel multipath fading [15]–[20], [29]. In [15], the vector of received signal strengths (RSSs), observed from a wireless client's transmission by a predetermined set of nearby APs, is used as a "signalprint" to detect identity-based attacks. This method was further studied in [16]–[18]. However, due to the fact that RSS is a coarser information of CSI and has a high correlation with the transmit power and distance, RSS is naturally more suitable for localization [27] and key distribution [28] rather than device authentication. For instance, Chandrasekaran *et al.* [16] used RSS to detect spoofing which leads to 10% false alarm and missed attack rates. Wang *et al.* [29] leveraged carrier frequency offset and CSI as physical layer signatures to achieve privacy-preserving location authentication.

Jiang *et al.* [19] proposed CSITE by using CSI magnitude measurements averaged in time over multiple frames to form a client signature. Although CSI provides detailed information about the channel, it is not available in the current device driver [24]. In the subsequence work, Xiong and Jamieson [20] proposed SeArray, which uses the angle-of-arrival information to construct highly sensitive signatures to identify a client. However, SeArray must be equipped at the AP with multiple antennas, which limits its application.

Hardware fingerprinting-based schemes [21]–[23] is based on the reflection of defects or unique design of the hardware on the transmitted waveforms. One of them is physically unclonable function [23], which generates fingerprinting based on the complex physical characteristics of the ICs in the wireless devices. As mentioned in [24], the disadvantage of this protocol lies in its requirement of specially manufactured ICs. Clock skew fingerprinting is another hardware fingerprinting-based method [21]. Unfortunately, an attackers can easily mimic the clock skew of legitimate user by manipulating the timestamps. Radiometric fingerprinting is based on the fact that the unique characteristics of a hardware transceiver cannot be replicated or copied from one device to another [22]. However, this scheme is vulnerable to impersonation and replay attacks if the attacker is more powerful [24], [30].

Acoustic fingerprinting has been studied in recent papers [31]–[33]. Zhu *et al.* [33] presented a context-free and geometry-based approach to recover keystrokes by using keyboard acoustic emanations to calculate the relative positions among keystrokes and the microphones equipped in the smartphones. Das *et al.* [31] discussed the feasibility of using microphones and speakers to uniquely fingerprint individual devices. They leverage Gaussian mixture models to classify the device into one of several recorded audio fingerprints. Zhou *et al.*'s work [32] is the most closely related to ours, which propose a scheme to generate stable and unique device ID stealthy for smartphones by exploiting the FR of the speaker. In order to reduce the nonlinear effects of the speaker, Zhou *et al.* [32] adopted the stimulation which consists of a series of cosine wave from 14 to 21 kHz with 100 Hz gap between neighboring frequency points. Different from Zhou *et al.* [32], we consider wireless device authentication with the FR of a speaker and a microphone from two wireless device. We leverage max-frequency technology to reduce the time of acoustic signals transmission and fingerprinting extraction. Moreover, we analyze security of the proposed scheme with mobile phones under different attacks, and find that the fingerprinting of microphone is the key to resist the audio replay attack. Furthermore, we consider the impact of distance between the sender and receiver on S2M, and propose a solution to achieve variable distance authentication.

III. Modeling

Before diving into details on S2M, some basics on acoustic attenuation model and adversarial models are given as follows.

A. Acoustic Attenuation Model

Acoustic attenuation is a measurement of the energy loss of sound propagated in media. Due to the fractal microstructures of media, such an acoustic attenuation typically exhibits a frequency dependency characterized by a power law

$$S(f, x) = S_0(f)e^{-\alpha(f)x} \tag{1}$$

where $\alpha(f) = \alpha_0(2\pi f)^{\eta_f}$, f denotes the frequency; x is wave propagation distance; $S_0(f)$ and $S(f, x)$ represent the amplitude of transmitted and received acoustic pressure at

frequency f, respectively; the tissue-specific coefficients α_0 and η_f are empirically obtained by fitting measured data [33].

Due to the inability of the acoustic components (i.e., speaker and microphone) to faithfully reproduce tones of certain frequencies, the phenomenon of selective attenuation of certain frequencies happens. Consequently, when sound waves are transmitted by a speaker and received by a microphone, the acoustic attenuation model can be rewritten as

$$S(f,x) = L_S(f)L_M(f)S_0(f)e^{-\alpha(f)x} \tag{2}$$

where $L_S(f)$ is the loss of a speaker, $L_M(f)$ is the loss of a microphone.

The FR of a speaker/microphone pair at frequency f is defined as

$$\text{FR} \triangleq R(f) = 20\log\frac{S(f,x)}{S_0(f)} = 20\log\left[L_S(f)L_M(f)e^{-\alpha(f)x}\right] \tag{3}$$

where $\log(\cdot)$ is the logarithmic function with base 10.

Based on (3), we observe that radio frequency characterizes the physical characteristics of speaker and microphone pairs. In order to achieve authentication of wireless devices, an intuitive idea is to use FR as the acoustic hardware fingerprint. However, the receiver (i.e., authenticator) can not calculate the FR as it does not know the transmitted acoustic pressure $S_0(f)$. To overcome this problem, we use

$$\Xi = \{\Xi(f_i)\}_i \qquad i \in \{1,\ldots,N\} \tag{4}$$

to represent the fingerprints of speaker and microphone pairs, where f_i is the selected frequency for feature extraction and $\Xi(f_i)$ is defined as follows:

$$\Xi(f_i) = 20\log S(f_i,x) = R(f_i) + 20\log\left[S_0(f_i)\right]. \tag{5}$$

B. Adversarial Model in Device Authentication

We design S2M based on a strong adversarial model. We summarize our assumptions about the attacker Eve as follows.

1) The adversary power is computationally unbounded, i.e., it can run an infinite number of steps.
2) The adversary is active and can masquerade other nodes by forging their identities.
3) The adversary knows the authentication protocol and its parameters settings.

Based on these assumptions, we consider the following four attack models in this paper.

1) *Audio Replay Attack:* An adversary Eve first records the audio signals from a legitimate sender, and then impersonates the legitimate one by replaying the recordings.
2) *Changing Distance Attack:* Eve intends to improve the successful probability of attack by adjusting the distance between the receiver and him.
3) *Same-Type-Device Attack:* Eve has a device with the same brand (i.e., the same manufacturer) as that of the legitimate transmitter, and he pretends the transmitter by sending the same audio signal as the real transmitter.
4) *Composition Attack:* Eve has a device with the same brand as that of the legitimate transmitter, and he launches a replay record attack with changing distances.

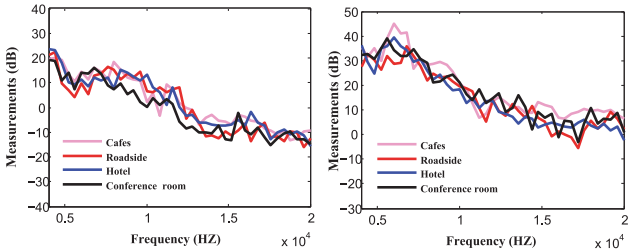


Fig. 2. Fingerprints for the same pair of devices. (a) From Huawei 3C to HTC T328D. (b) From SONY Z1 to HTC T328D.

1) Evaluation Metrics: In this paper, we mainly focus on two evaluation metrics: 1) false negative (FN) error rate and 2) false positive (FP) error rate. Specifically, the FN rate is the number of legitimate sender's audio signals (for authentication) that are incorrectly identified as the attacker's signal over the totally received number of audio signals for authentication. Similarly, the FP error rate is the number of attacker's signals which are incorrectly considered to be sent from the legitimate sender over the totally received number of audio signals for authentication. The main objective is to design a wireless device identification scheme with both the low FP error rate and the FN error rate.

IV. Acoustic Hardware Characteristics

As a valid fingerprint, the physical characteristic of fingerprint must has two properties: 1) the stability and 2) otherness. The stability is that the characteristic does not change along with the changing of time and space; and the otherness means that the characteristic between different devices is big enough to differentiate. In order to verify Ξ has those properties, we conduct extensive experiments as follows.

First of all, we consider the stability of the same speaker and microphone pairs. From (5), we assume that $\Xi(f_i)$ is invariable when f_i, x and $S_0(f_i)$ are fixed. To verify these conclusion, we measure Ξ for the same pair of devices (from Huawei 3C and SONY Z1 to HTC T328D, respectively) at four different locations, including a cafe, a roadside, a hotel, and a conference room. As shown in Fig. 2, the measurement values of Ξ at different locations are similar but not identical. The reason is that multipath (i.e., echo) can affect the received audio signals, and then affects the values of Ξ. Fortunately, as we can see from the figure, the difference is very slight. Thus, we can eliminate the effects of echo by designing a fingerprint MA carefully.

The second experiment considers the difference of speakers between wireless devices. Fig. 3 depicts the fingerprints (i.e., the function Ξ) from different speakers to the same microphone (i.e., from phones: Huawei 3C, ZTE U960, ZTE N880, Samsung S5560, MX3, SONY Z1, Mi2, Huawei honor$^+$, Huawei P1, to phone: HTC T328D). The fingerprints are measured by transmitting tones of frequencies between 4 and 20 kHz (with step size 400), from one device to another, while placing the devices in a fixed distance of 10 cm from

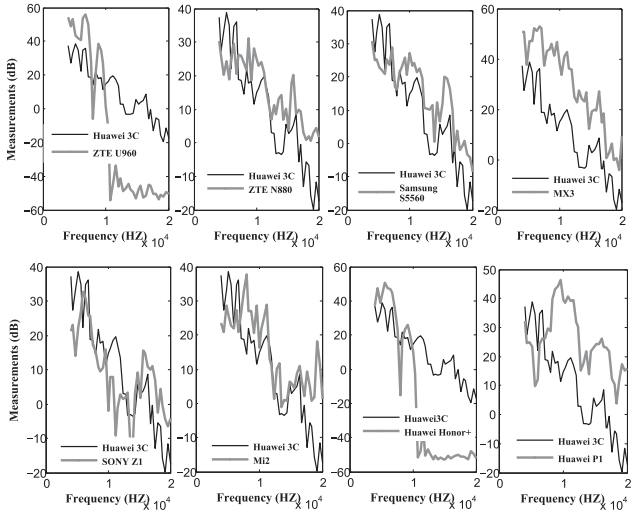


Fig. 3. Fingerprints for different speakers.

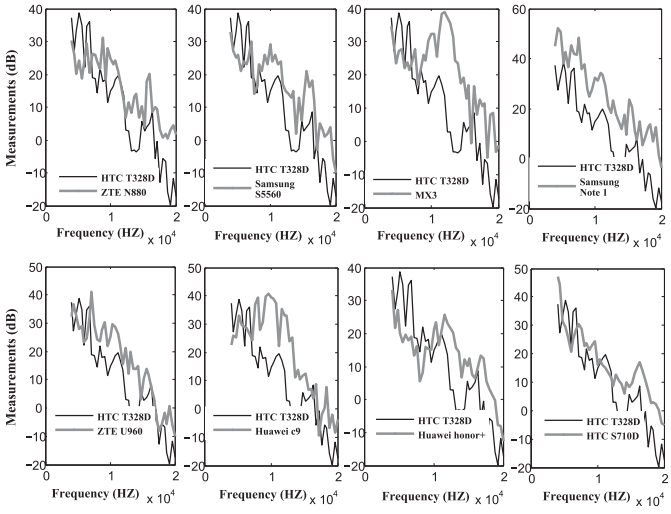


Fig. 4. Fingerprints for different microphones.

each other. The results (Fig. 3) show that the fingerprint differences for two mobile phones are obvious.

Finally, we consider the difference of microphones among wireless devices. Fig. 4 depicts the fingerprints from one speakers to different microphones (i.e., from phone: Huawei 3C, to phones: HTC T328D, Samsung S5560, MX3, Samsung Note1, ZTE U960, Huawei C9, Huawei honor+, and HTC S710D). The fingerprints are also measured by transmitting tones of frequencies between 4 and 20 kHz (with step size 400) in a fixed distance of 10 cm. The results (Fig. 4) show that the microphone differences between mobile phones are obvious.

V. Authentication Protocol

In this section, we present a lightweight device authentication protocol, named S2M consisting of two process: 1) learning process and 2) verification process. The purpose of learning process is to extract and store the fingerprint samples of legitimate wireless devices. Verification process aims to determine whether two fingerprints are from the same device or not. Fig. 5 illustrates the flow of our S2M protocol.

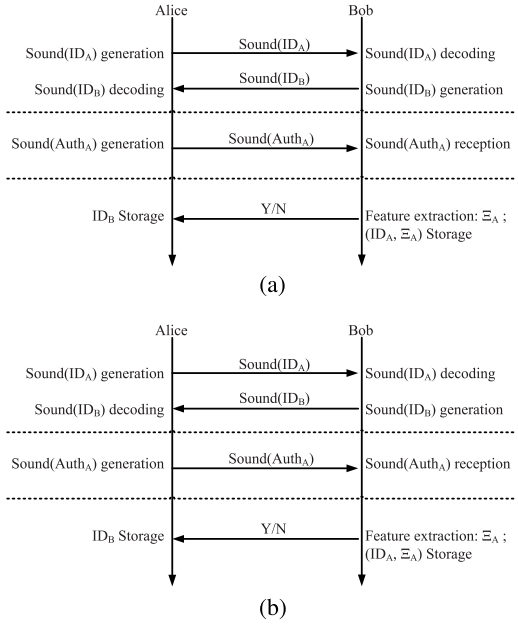


Fig. 5. Timing diagram for S2M. (a) Learning process. (b) Verification process.

A. Learning Process

Learning process consists of three stages: 1) *audio-handshake phase*; 2) *mixed-signal generation phase*; and 3) *feature extraction and storage phase*.

1) Audio-Handshake Phase: During this phase, the identity interaction and synchronization of legitimate nodes (Alice and Bob) must be achieved. With no doubt, it can be achieved by using conventional wireless communication (e.g., Wi-Fi and Bluetooth). However, in order to reduce the dependence on hardware and wireless network infrastructures, as well as expand the application scope of S2M, we realize the identity interaction and synchronization by utilizing of audio data transmission scheme: dual-tone multifrequency [34].

The steps of this phase are described as follows.

1) Alice encodes her ID number into Sound(ID_A) with audio encoding, and sends it to Bob using speaker.
2) After receiving the audio signals, Bob first decodes ID_A with audio decoding, and then encodes ID_B into sound(ID_B) and sends back to Alice by speaker. After that, Bob prepares to receive the audio signals from Alice for feature extraction.
3) If Alice receives the audio signals, she decodes ID_B from the signals sent by Bob. Otherwise, Alice returns step 1.

2) Mixed-Signals Generation Phase: After obtaining Bob's ID, Alice needs to send some audio signals, which are denoted by Sound($Auth_A$), to Bob for feature extraction. To enable Bob to calculate fingerprints Ξ, an simple approach is that Alice sends

$$S_0(f_i) = \sin(2\pi f_i t) \quad i \in \{1, \ldots, N\}$$

to Bob in sequence. However, this method falls into the following dilemma: a larger value of N of frequencies improves the resolution of fingerprint, but increases the transmission time; while a smaller number N of frequencies decreases the transmission time, but deteriorates the resolution of fingerprint at the same time.

In order to obtain a high-resolution fingerprint and reduce the transmission time simultaneously, we leverage mix-frequency technology, which generates the mixed audio signals by combining the various frequency components. Specifically, the mixed-signals $\text{Sound}(\text{Auth}_A)$ can be denoted by

$$\text{Sound}(\text{Auth}_A) = \sum_{i=1}^{N} \frac{1}{N} \sin(2\pi(\varphi_0 + f_\Delta(i-1))t) \tag{6}$$

where φ_0 is the minimum frequency, and f_Δ is the length of step. In our system, we set $\varphi_0 = 4\,\text{kHz}$ and $f_\Delta = 0.4\,\text{kHz}$. That is, $N = 41$, $f_1 = 4\,\text{kHz}$, $f_2 = 4.4\,\text{kHz}$, $\cdots$, and $f_N = 20\,\text{kHz}$.

The steps of this phase are described as follows.

1) Alice generates the mixed-signals $\text{Sound}(\text{Auth}_A)$ with (6), and sends the signals to Bob using speaker.
2) Bob receives the signals $\widehat{\text{Sound}}(\text{Auth}_A)$ with microphone, where

$$\widehat{\text{Sound}}(\text{Auth}_A) = \text{Sound}(\text{Auth}_A) + \text{Noise} + \text{Echoes}.$$

3) Feature Extraction and Storage Phase: To extract the feature of devices pairs, Bob needs to compute Ξ from the record $\widehat{\text{Sound}}(\text{Auth}_A)$. Thus, Bob has to convert the signals $\widehat{\text{Sound}}(\text{Auth}_A)$ from time domain to frequency domain. To do this, we utilize fast Fourier transform (FFT), an algorithm for computing the Fourier transform of a set of discrete data values. Given a finite set of data points, the FFT expresses the data in terms of its component frequencies. By using FFT, the sound pressure amplitude is denoted by

$$< S(f_1), \ldots, S(f_N) >= \text{FFT}_N\left(\widehat{\text{Sound}}(\text{Auth}_A)\right). \tag{7}$$

Here, FFT_N is the function returning the local maximum FFT values $S(f_1), \ldots, S(f_N)$ at frequency ranges $[f_1 - FS, f_1 + FS], \ldots, [f_N - FS, f_N + FS]$, respectively (as shown in Fig. 6). The objective of taking the local maximum is to eliminate the impact of frequency shift. In our system, we set $FS = 5\,\text{Hz}$. Thus, the fingerprint Ξ_A of Alice and Bob can be written as

$$\Xi_A = 20\log(<S(f_1), \ldots, S(f_N)>). \tag{8}$$

The steps of this phase are described as follows.

1) When Bob receives the mixed-signals, he extracts the fingerprint Ξ_A using (7) and (8), and records the pairs $<\text{ID}_A, \Xi_A>$.
2) Bob sends one period sine wave of frequency 10 kHz to Alice (it means success in learning process).
3) After receiving the signals with frequency 10 kHz, Alice records ID_B; otherwise, Alice returns to the mixed-signals generation phase.

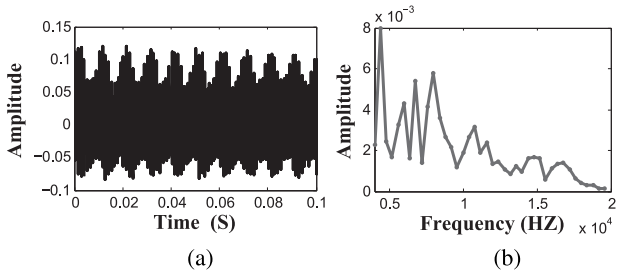


Fig. 6. Signals conversion from time domain to frequency domain. (a) Time domain signals. (b) Signals after FFT_N.

B. Verification Process

Verification process is also divided into three phases: 1) *audio-handshake phase*; 2) *mixed-signal generation phase*; and 3) *feature extraction and matching phase*.

Audio-Handshake Phase: The audio-handshake phase of verification process is similar to that of learning process. The steps of this phase are described as follows.

1) Alice encodes ID_A into $\text{Sound}(\text{ID}_A)$ with audio encoding, and sends it to Bob.
2) After receiving the signals from Alice, Bob decodes ID_A and checks whether the ID_A is already stored locally (if not, the authentication fails).
3) Bob sends $\text{Sound}(\text{ID}_B)$ to Alice, and then is ready to receive the signals for feature extraction from Alice.
4) If Alice receives the audio signals, she decodes ID_B with audio decoding and checks whether the ID_B is already stored locally (if not, the authentication fails); otherwise, Alice goes back to the step 1.

Mixed-Signal Generation Phase: The mixed-signal generation method of this phase is the same as that in the learning process. Thus, we omit the details here.

Feature Extraction and Matching Algorithm Phase: The feature extraction method of this phase is similar to that in learning process. The goal in the fingerprint MA is to determine whether $\Xi_A \approx \Xi'_A$, where Ξ_A is the fingerprint extracted in the learning process and Ξ'_A is the fingerprint extracted in verification process (details of MA is given in Section V-C). Specifically, if $\text{MA}(\Xi_A, \Xi'_A) = Y$, the authentication is successful (i.e., $\Xi_A \approx \Xi'_A$); otherwise, the authentication fails. The steps of this phase are described as follows.

1) Bob extracts the fingerprint Ξ'_A using (7) and (8).
2) Bob calls MA with the input Ξ and Ξ'. If $\text{MA}(\Xi_A, \Xi'_A) = Y$, Bob sends one period sine wave of frequency 10 kHz to Alice; otherwise, the authentication fails.
3) If Alice receives the audio signals with the frequency 10 kHz, the authentication is successful; otherwise, Alice goes back to mixed-signal generation phase.

C. Fingerprint Matching

Low energy consumption of the scheme is one of our main concerns since many wireless devices usually powered by batteries. In order to design a low energy consumption authentication protocol, it is necessary to propose a lightweight fingerprint MA. In this paper, we consider two lightweight

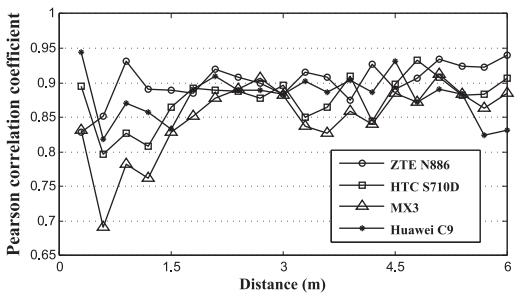


Fig. 7. PCCs between the legitimate sender (ZTE N886) and the attackers (HTC S710D, MX3, and Huawei C9) under different authentication distances.

approaches for fingerprint matching: 1) correlation-coefficient-based MA (C-MA) and 2) D-MA.

An possible approach for fingerprint matching is to use Pearson correlation coefficient (PCC), which is 1 in the case of a perfect positive linear relationship (correlation), -1 in the case of a perfect decreasing (negative) linear relationship, and some value between -1 and 1 in all other cases, indicating the degree of linear dependence between the variables.

Let $\Xi_A = <\xi_1, \ldots, \xi_N>$ be the fingerprint in the learning process, and $\Xi'_A = <\xi'_1, \ldots, \xi'_N>$ be the fingerprint in the verification (authentication) process. The PCC is defined as

$$\mathrm{PCC}\big(\Xi, \Xi'\big) \triangleq \frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{\xi_i - \bar{\xi}}{\sigma}\right)\left(\frac{\xi'_i - \bar{\xi}'}{\sigma'}\right) \tag{9}$$

where $\bar{\xi}$ and $\bar{\xi}'$ are the mean of Ξ_A and Ξ'_A, respectively, $\sigma = \sqrt{(1/(N-1))\sum_{i=1}^{N}(\xi_i - \bar{\xi})^2}$, and $\sigma' = \sqrt{(1/(N-1))\sum_{i=1}^{N}(\xi'_i - \bar{\xi}')^2}$. This method sets a threshold Th. Bob computes $\mathrm{PCC}(\Xi, \Xi')$ with (9), and compares the value of $\mathrm{PCC}(\Xi, \Xi')$ with the threshold Th. That is,

$$\text{C-MA}\big(\mathrm{Th}, \Xi, \Xi'\big) = \begin{cases} Y, & \text{if } \mathrm{PCC}\big(\Xi, \Xi'\big) \geq \mathrm{Th} \\ N, & \text{otherwise.} \end{cases} \tag{10}$$

Fig. 7 plots the PCCs between the legitimate sender and the attackers under different authentication distances, where ZTE N886 is the legitimate sender, HTC T328d is the authenticator, and HTC S710D, MX3 as well as Huawei C9 are the attackers who want to impersonate ZTE N886. It shows that: 1) the PCCs at the legitimate sender are unstable with the variation of authentication distance and 2) there is no a clear dividing line between the legitimate sender and the attackers. Based on the above observations, it is difficult to find a suitable threshold to distinguish between a legitimate sender and an attacker. Thus, C-MA does not work in our scenario.

To overcome the above issues, in the proposed protocol, we design a new MA, named D-MA. The algorithm takes two fingerprints Ξ_A and Ξ'_A as inputs. It first calculates the absolute value of corresponding components of two fingerprints vectors, and then compares each absolute value with a deviation threshold Δ. Specifically, denoting

$$S = \big|\big\{n : \big|\xi_n - \xi'_n\big| \leq \Delta\big\}\big| \tag{11}$$

Algorithm 1 D-MA

Parameter: deviation threshold Δ; matching threshold *Th*.
Input: fingerprint in learning process and in verification process $\Xi_A =< \xi_1, \cdots, \xi_N >$ and $\Xi'_A =< \xi'_1, \cdots, \xi'_N >$, respectively.
Output: Y/N

```
S=0
for round n = 1, ..., N do
    if |ξ_n − ξ'_n| ≤ Δ then
        S=S+1
    end if
end for
DR = (N−S)/S
if DR ≤ Th then
    Output Y
else
    Output N
end if
```

the ratio of $N - S$ to S can be considered as a metric on N-dimensional vector space R^N, named deviation ratio (DR). That is to say, we define

$$\mathrm{DR}\big(\Xi_A, \Xi'_A\big) = \frac{N-S}{S} = \frac{\big|\big\{n : \big|\xi_n - \xi'_n\big| > \Delta\big\}\big|}{\big|\big\{n : \big|\xi_n - \xi'_n\big| \leq \Delta\big\}\big|} \tag{12}$$

where $|\mathcal{A}|$ means the cardinality of set $\mathcal{A}$. Finally, the algorithm compares $\mathrm{DR}(\Xi_A, \Xi'_A)$ with a matching threshold Th. If $\mathrm{DR}(\Xi_A, \Xi'_A) \leq \mathrm{Th}$, the algorithm output Y (Yes); otherwise, output N (No). That is,

$$\text{D-MA}(\Delta, \mathrm{Th}, \Xi, \Xi') = \begin{cases} Y, & \text{if } \mathrm{DR}\big(\Xi, \Xi'\big) \leq \mathrm{Th} \\ N, & \text{otherwise.} \end{cases} \tag{13}$$

Note that $\mathrm{DR}(\Xi_A, \Xi'_A)$ reflects the similarity of two fingerprints. For instance, $\mathrm{DR}(\Xi_A, \Xi'_A) \leq 0.5$ means that there are at least $(2/3)N$ pairs of the corresponding component in the two fingerprints victor such that their absolute values are less than Δ. The details of D-MA are shown in Algorithm 1.

In order to select the appropriate deviation threshold Δ and the matching threshold Th, we conduct extensive experiments to investigate the relationship among Δ, Th and authentication errors (i.e., the FN and FP errors). In our experiments, more than 50 mobile phones are used and more than 20 000 fingerprints are extracted under different authentication distances. Fig. 8 shows the authentication error rate with different deviation thresholds and matching thresholds. We can see that, for each Th, a larger Δ increases the rate of FN error but improves the rate of FP error. Especially, when the matching threshold Th is 0.4 and the deviation threshold Δ is 8, both the FN and FP error rate are less than 0.5%, Thus, in our system, we set deviation threshold $\Delta = 8$ and matching threshold $\mathrm{Th} = 0.4$.

VI. Extended Protocol E-S2M

In this section, we design an E-S2M to realize authentication under variable distance in presence of an active adversary.

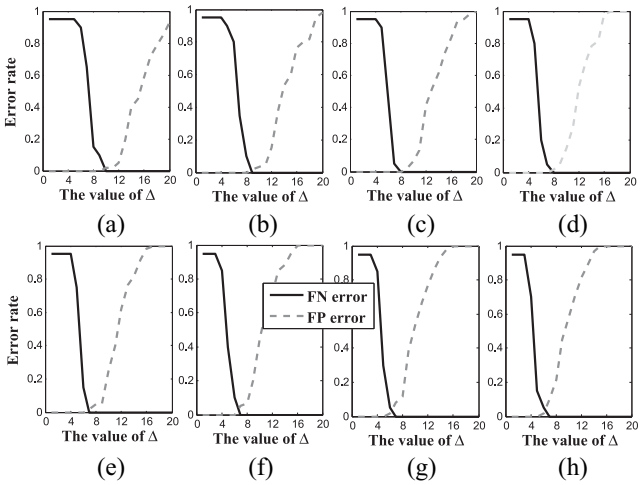


Fig. 8. Relationship between authentication errors and the deviation threshold Δ under different matching threshold Th. (a) Th = 0.3. (b) Th = 0.4. (c) Th = 0.5. (d) Th = 0.6. (e) Th = 0.7. (f) Th = 0.8. (g) Th = 0.9. (h) Th = 1.

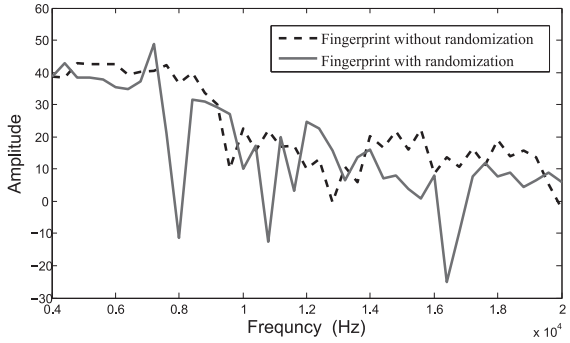


Fig. 9. Fingerprint with and without source randomization.

A. Audio Source Randomization

From Figs. 3 and 4, we can observe that the fingerprints of many phone pairs always decrease with the frequency, which can be explained from (3). Obviously, it is unfavorable to the security of the proposed scheme. Hence, it is necessary to enhance the randomness of fingerprint to improve the security. One feasible way is to add the randomness of the mixed-signals Sound(Auth$_A$). The details of this method is as follows. In the mixed-signals generation phase in learning process, Alice first randomly chooses a set of real numbers $(\varphi_1, \ldots, \varphi_N)$ from standard normal distribution, and then she generates the following mixed-signals:

$$\text{Sound}(R\text{-Auth}_A) = \sum_{i=1}^{N} \frac{\varphi_i}{\sum_{j=1}^{N} \varphi_j} \sin(2\pi(\varphi_0 + f_\Delta(i-1))t). \tag{14}$$

In authentication process, Alice transmits the mixed-signals Sound(R-Auth$_A$) instead of Sound(Auth$_A$) for fingerprints extraction. Fig. 9 shows the fingerprint with and without source randomization. The result shows that the fingerprint with source randomization has more drastic change (i.e., more random) than that without randomization. Therefore, it is more difficult to be forged by adversaries.

B. Variable Distance Authentication Protocol

So far, we assume that the distances between sender and receiver in learning and verification process are the same in the proposed scheme. Hence, this scheme only applies to the scenarios where: 1) the sender and the receiver are fixed at some places and 2) the sender (or receiver) can move such that the distance between them in verification process is equal to that in learning process.

To extend the application of S2M, we design an extended protocol of S2M such that the new protocol can achieve authentication at different distances between the users. Based on S2M, the new protocol includes two algorithms: 1) *mobile learning algorithm* and 2) *automatic fingerprint MA*. The main idea of the former is to acquire the fingerprints at different distances; the idea of the latter is to choose an optimum fingerprint out of the fingerprints at different distances for matching.

We first present the mobile learning algorithm. In learning process, Alice and Bob consult with an *effective range* $[l, r]$ of authentication, and then they proceed the following steps.

1) Alice and Bob first move their positions such that the distance between them is equal to r, and then perform learning process of S2M [i.e., Section V-A, the obtained fingerprint at Bob is denoted by $\Xi_A(0)$].
2) Alice walks toward Bob with *moving distance* ΔL_1, and then they perform the learning process of S2M [the fingerprint at Bob is denoted by $\Xi_A(1)$].
3) For $i > 1$, if $r - \Sigma_i \Delta L_{i-1} > l + (1/2)\Delta L_{i-1}$, and then Alice and Bob repeat step 2 with moving distance ΔL_i [the fingerprint is denoted by $\Xi_A(i)$]; otherwise, stop.

After implementing this algorithm, Bob obtains a list of fingerprints $<\Xi_A(0), \Xi_A(1), \ldots, \Xi_A(M)>$. Denote the *maximum moving distance* by $\Delta L = \max\{\Delta L_1, \ldots, \Delta L_M\}$. The effect of ΔL on the performance of E-S2M will be discussed in the end of this section.

Then, we introduce automatic fingerprint MA. For authentication, Alice and Bob perform the previous scheme in the authentication process (i.e., Section V-B) by replacing the D-MA with *automatic fingerprint MA* as follows. Let $\Xi'_A = <\xi_1, \ldots, \xi_N>$ be the fingerprint obtained by Bob in this process. For each $m \in \{1, \ldots, M\}$, let $\Xi_A(m) = <\xi_1(m), \ldots, \xi_N(m)>$.

1) For each $n \in \{1, \ldots, N\}$, Bob computes $a_n = \min\{|\xi'_n - \xi_n(m)| : m = 1, \ldots, M\}$.
2) For each $m \in \{1, \ldots, M\}$, Bob computes $A_m = \{n : a_n = |\xi'_n - \xi_n(m)|, n = 1, \ldots, N\}$.
3) If m_0 be the index such that $|A_{m_0}| = \max\{|A_m| : m = 1, \ldots, M\}$, Bob computes D-MA(Δ, Th, Ξ_{A^*}, Ξ'_{A^*}) with inputs $\Xi_A(m_0)$ and Ξ'_A.

The details of automatic fingerprint MA are shown in Algorithm 2.

We conduct extensive experiments to investigate the relationship between the maximum moving distance ΔL and the error rates, in which more than 50 different mobile phone groups are used and 20 000 fingerprints are extracted. From Fig. 10, it can be seen that: 1) the rate of FP error is stable as ΔL increases and 2) a larger Δ increases the rate of FN error. Especially, when $\Delta L \leq 50$ cm, both the FN and FP error

Algorithm 2 Automatic Fingerprint MA

input: the fingerprints in learning process $< \Xi_A(m) >_m = \{< \xi_1(m), \cdots, \xi_N(m) >\}_{m=1}^{M}$; the fingerprint in verification process $\Xi'_A =< \xi_1, \cdots, \xi_N >$.

Output: Y/N.

1: **for** round $n = 1, \cdots, N$ **do**
2: $a_n = \min\{|\xi'_n - \xi_n(m)| : m = 1, \cdots, M\}$
3: **end for**
4: **for** round $m = 1, \cdots, M$ **do**
5: $A_m = \{n : a_n = |\xi'_n - \xi_n(m)|, n = 1, \cdots, N\}$
6: **end for**
7: **if** $|A_{m_0}| = \max\{|A_m| : m = 1, \cdots, M\}$ for index m_0 **then**
8: **run** *D-MA*$(\Delta, Th, \Xi_A(m_0), \Xi'_A)$
9: **end if**

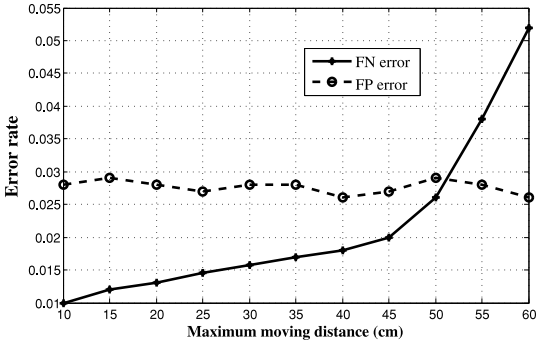


Fig. 10. Impact of change maximum moving distance ΔL on error rate.

rates are less than 0.03, i.e., only 2 or 3 places' fingerprints are required to achieve the FN error rate under 3% when the effective range is $0 - 1$ m. Thus, in our system, we set the maximum moving distance $\Delta L = 50$ cm.

C. Active Attack Detection Algorithm

We consider the scenario of active attacks in which one or more attackers are transmitting high frequency audio signals from their speakers when Alice is sending mixed-signals to Bob in the verification process. Due to the fact that the high frequency signals transmitted by attacker may have negative effect on the reception of the real fingerprints by Bob, the proposed protocol does not work well in such situations. To solve the problem, we design an active attack detect algorithm in which Alice processes as follows.

1) In the learning process, Alice first records the mixed-signals with the microphone [the recording signal denoted by $\overline{\text{Sound}}(R\text{-Auth}_A)$], and then obtains the fingerprint Ξ_{A^*} of speaker and microphone pair of herself from $\overline{\text{Sound}}(R\text{-Auth}_A)$ with (7) and (8).
2) In the verification process, following the same steps as the learning process, Alice first obtains the new fingerprint Ξ'_{A^*}, and then computes D-MA$(\Delta, \text{Th}, \Xi_{A^*}, \Xi'_{A^*})$. If the output is N, then the active attacking is considered as occurring.

To evaluate the performance of the detection algorithm, we carry out an experiment as follows. Two mobile phones Samsung Note1 and Samsung Note2 are considered as the legitimate users, while a Samsung Note3 acts as an attacker who transmits interference with different frequency components during in the authentication process. Setting the deviation threshold $\Delta = 8$ and the matching threshold $\text{Th} = 0.4$, we have $\Pr(\text{Det}) = 58\%$ and $\Pr(\text{Succ}|\text{Unde}) = 98.2\%$, where Det is the event that an attack is detected, Succ is the event that an authentication between Alice and Bob is successful, and Unde is the event that an attack is undetected. The experimental results show that almost every undetected attack can result in to a successful authentication. The reason is that, when the frequency components in interference signals transmitted by an attacker do not include or marginally include the frequency components in signals sent by Alice, the interference signals would have no adverse effect on authentication between legitimate users. Accordingly, the experimental results show that the detection algorithm is extremely effective.

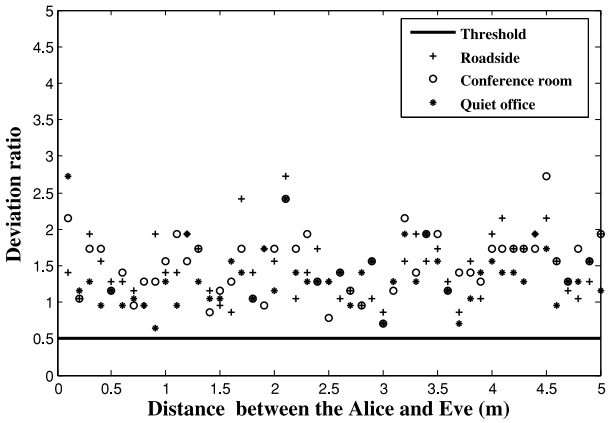


Fig. 11. Reply attack: the DRs (from attacker) for different scenarios.

VII. Performance Evaluation

A. Security Analysis

To evaluate the security strength of E-S2M, we analyze various types of attackers. For each experiment, we set the deviation threshold $\Delta = 8$ and the matching threshold $\text{Th} = 0.4$.

First, we conduct an experiment on E-S2M to evaluate an audio replay attack in a quiet corridor. In this experiment, mobile phone ZTE N880 is the legitimate sender; HTC T328d is the receiver; and MX3 is the attacker. We evaluate E-S2M under different scenarios: roadside, conference room, and quiet office, where the effective range is 0–2 m. For each effective range, the attacker changes the distance between itself and the receiver from 0.1 to 5 m with step length 0.1 m, and launches replay attack at each position. The receiver extracts the fingerprints and computes the DRs. Fig. 11 shows the DRs (from attacker) under different scenarios. It can be seen that the DR under each scenario is larger than the deviation threshold. Thus, E-S2M has excellent performance to resist the audio replay attack.

Second, we consider a changing distance attack in a quiet corridor under different effective ranges (e.g., 0–1 m, 0–2 m, . . . , 0–5 m). In this experiment,

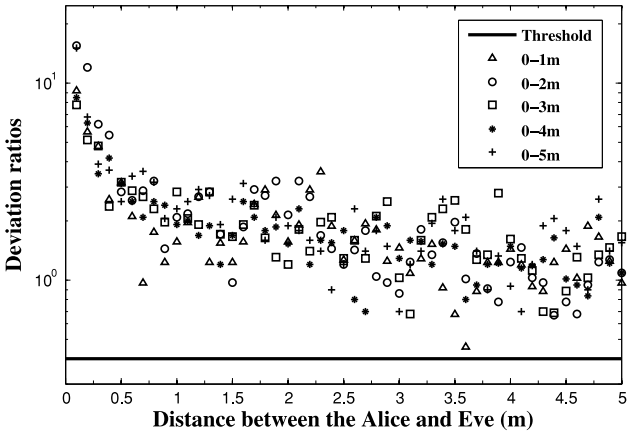


Fig. 12. Changing distance attack: the DRs (from attacker) for each effective range of authentication.

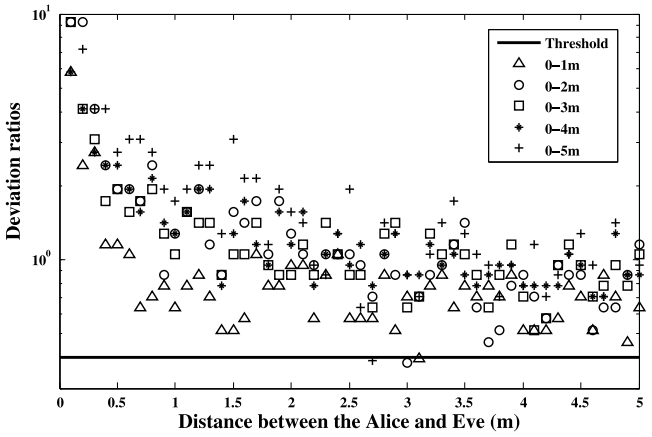


Fig. 14. Composition attack: the DRs from attacker under different positions and effective ranges.

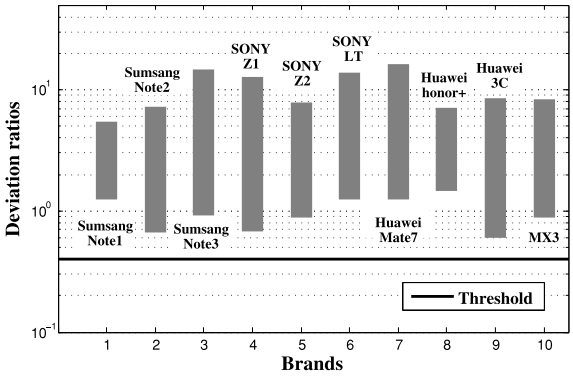


Fig. 13. Same-type-device attack: the value range of DRs (from attacker) for each type of phone.

the mobile phone (Huawei honor3C) is set to be the legitimate sender, the HTC T328d is the receiver and the Samsung S5560 is the attacker. For each effective range, we change the distance between the attacker and the receiver from 0.1 to 5 m with step 0.1 m. In each distance, the attacker follows E-S2M and sends the authentication audio signals, which is generated from (14) (i.e., it is assumed that the attacker know the random numbers $\varphi_1, \ldots, \varphi_N$), to the receiver. Fig. 12 shows the DRs (from attacker) under different authentication distance. It can be seen that all of the DRs are larger than the deviation threshold. Hence, E-S2M has outstanding performance to resist the changing distance attack.

Third, we conduct an experiment for same-type-device attack in a quiet office. In this experiment, we consider ten types of mobile phones: Samsung (Note1, Note2, and Note3), SONY (Z1, Z2, and LT), Huawei (Mate7, honor+, and 3C), and MX3. For each type of cell phones, we have two mobile phones with the same chips, one as a sender and the other as an attacker. We use the HTC T328d as the receiver and set the effective authentication range 0–5 m. Each attacker randomly chooses 200 locations such that the distance form attacker to receiver is less than 5 m, and launches an attack by following E-S2M at each location. Fig. 13 shows the range of DRs (from attacker) under different types. It can be seen that the DRs for each type are larger than the deviation threshold. Thus, E-S2M can effectively resist the same-type-device attack. In addition, Fig. 13 also reveals that two devices with the same type have different fingerprints of speakers.

Finally, a composition attack is considered in a quiet corridor under different effective ranges (e.g., 0–1 m, 0–2 m, ..., 0–5 m), two mobile phones with the same chips (e.g., Huawei Mate7) are used as sender and attacker, respectively, and HTC T328d is considered as a receiver. For each effective ranges, the attacker changes the distance between receiver and him from 0.1 to 5 m with step length 0.1 m and launches replay attack at position. Fig. 14 shows that there are only 3 of the DRs for all positions and all effective ranges are less than the threshold. Accordingly, the proposed scheme has good performance to resist the composition attack.

B. Performance of S2M and E-S2M

To evaluate the Performance of S2M and E-S2M, we conduct extensive experiments under the following settings.

1) For S2M, the distance (in the learning and authentication processes) is fixed from 0.5 to 5 m with step 0.1 m, and the number of authentications is 10 for each fixed distance.
2) For E-S2M, the effective range is fixed $(0, 5$ m$]$, and the distances in the learning process change from 0.25 to 4.75 m with step 0.5 m; the distances in the authentication process change from 0.5 to 5 m with step 0.1 m; and the number of authentication is 10 for each fixed distance.
3) For both S2M and E-S2M, the distance between Eve and Bob is changed from 0.5 to 5 m with step 0.1 m; for each fixed distance, the number of attacks is 10.

We consider the following three scenarios.

1) *Scenario A:* Putting three mobile phones (a sender, a receiver, and an attacker, respectively) in a quiet office.
2) *Scenario B:* Putting three mobile phones at a roadside.
3) *Scenario C:* Putting three mobile phones in a conference room.

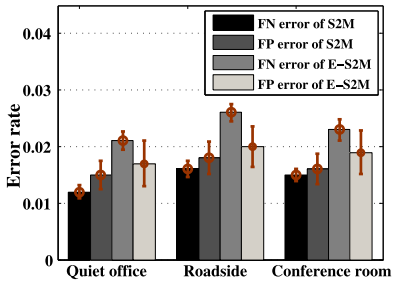


Fig. 15. Error rate of S2M and E-S2M under different scenarios.

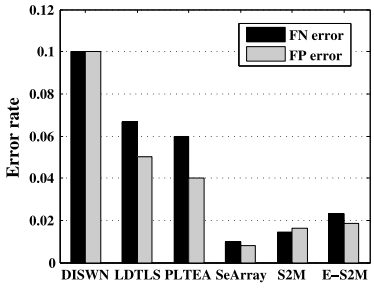


Fig. 16. Comparison between S2M, E-S2M and the exited protocols.

Furthermore, we consider 80 different groups of mobile phone tuples (*sender, receiver,* and *attacker*) in each scenario. Fig. 15 plots the error rate with respect to the authentication distance under different scenarios. This figure also shows the variances of the error rate in terms of the 80 different groups under different scenarios. It can be seen that: 1) for each case, both the FN rate and the FP rate of S2M are less than 2%; 2) the FN rate and the FP rate of E-S2M are less than 3% and 2%, respectively; and 3) the performance of E-S2M is worse than that of S2M, but E-S2M still has the low error rates.

Some famous fingerprinting-based device identification approaches are mentioned in Section II, i.e., DISWN [16], LDTLS [17], PLTEA [18], and SeArray [20]. Fig. 16 plots their performance with S2M and E-S2M. The result shows the following.

1) The proposed schemes have lower error rates compared with DISWN, LDTLS, and PLTEA: in DISWN, both the FN and FP error rates are 10% provided that the distance between attacker and the victim are geographically close [16]; in LDTLS, the FN error rate is 6.66% and the FP one is 5% [17]; in PLTEA, the FN (resp. FP) error rate is 6% (resp. 4%) [18].
2) The error rates of our schemes are larger than that of SeArray: in SeArray, the FN (resp. FP) error rate can achieve 1% (resp. 0.8%) [20]. However, the drawback of SeArray is that it need to equip the wireless devices with eight antennas in order to achieve such excellent performance, which would limit the application of SeArray.

C. Energy Consumption of E-S2M

Since most of wireless devices are power-constrained due to the battery size, it is necessary to evaluate the energy consumption of E-S2M. The total energy consumption E_{sender} at a sender can be given as

$$E_{\text{sender}} = E_{\text{handshake}} + E_{\text{audiogen}} + E_{\text{audiotran}} + E_{\text{dete}}$$

where $E_{\text{handshake}}$ is the energy consumption in audio handshake phase, E_{audiogen} and $E_{\text{audiotran}}$ are the energy consumption during the audio signals Sound(Auth_A) generation and transmission, respectively, and E_{dete} is the energy consumption in the active attack detection algorithm. The total energy consumption E_{receiver} at a receiver can be given as

$$E_{\text{receiver}} = E_{\text{handshake}} + E_{\text{audiorec}} + E_{\text{feature}} + E_{\text{MA}}$$

where $E_{\text{handshake}}$ is the energy consumption in the audio handshake phase, E_{audiorec} is the energy consumption during Sound(Auth_A) reception, while E_{feature} and E_{MA} are the energy consumption during the fingerprint extraction and matching phases, respectively. Obviously, $E_{\text{audiotran}}$ and E_{audiorec} depend on the transmission and reception time. In our system, both the transmission and reception time are set to 1 s.

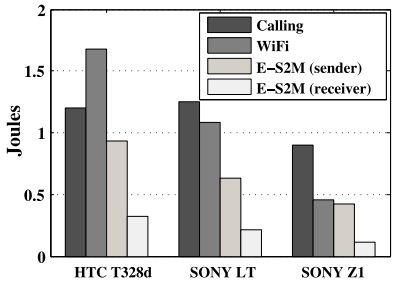


Fig. 17. Energy consumption of S2M, calling and surfing the Internet with Wi-Fi.

We conduct an experiment by using an energy monitor tool (named GSam Battery Monitor Pro V3.22) to compare the energy consumption of E-S2M with the effective range 0–2 m and other common mobile applications, such as calling and surfing Internet with Wi-Fi, in mobile phones (HTC T328d, SONY LT, and SONY Z1). Fig. 17 shows the energy consumption of E-S2M, calling, and surfing Internet with Wi-Fi within the time to complete an authentication with E-S2M. The results show that: the energy consumption of E-S2M at both the sender and the receiver are lower than that of calling and surfing Internet with Wi-Fi. Moreover, the energy consumption of the sender is higher than that of the receiver.

VIII. Conclusion

In this paper, we have proposed S2M and E-S2M, the software-only acoustic device authentication schemes that leverage the FR of speaker and microphone pairs of wireless IoT devices to construct highly specific hardware fingerprints. We have designed and implemented the proposed schemes in real mobile phones to evaluate the performance of S2M and E-S2M. The experimental results show that S2M and E-S2M can achieve both a low FN rate and a low FP error rate in various scenarios for different attacks. For the future work, we will consider to adjust the amplitudes of mixed signals and integrate some low-frequency signals to generate musical audio

signals, to authenticate devices and meanwhile improve the users' experience, for the scenarios where the acoustic sound might be annoying.

REFERENCES

[1] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," *Comput. Netw.*, vol. 54, no. 15, pp. 2787–2805, 2010.

[2] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of Things for smart cities," *IEEE Internet Things J.*, vol. 1, no. 1, pp. 22–32, Feb. 2014.

[3] N. Lu, N. Cheng, N. Zhang, X. Shen, and J. W. Mark, "Connected vehicles: Solutions and challenges," *IEEE Internet Things J.*, vol. 1, no. 4, pp. 289–299, Aug. 2014.

[4] Q. Jing, A. V. Vasilakos, J. Wan, J. Lu, and D. Qiu, "Security of the Internet of Things: Perspectives and challenges," *Wireless Netw.*, vol. 20, no. 8, pp. 2481–2501, 2014.

[5] K. Zhang, X. Liang, R. Lu, and X. Shen, "Sybil attacks and their defenses in the Internet of Things," *IEEE Internet Things J.*, vol. 1, no. 5, pp. 372–383, Oct. 2014.

[6] S. Sicari, A. Rizzardi, L. A. Grieco, and A. Coen-Porisini, "Security, privacy and trust in Internet of Things: The road ahead," *Comput. Netw.*, vol. 76, pp. 146–164, Jan. 2015.

[7] K. Zhang *et al.*, "Security and privacy for mobile healthcare networks: From a quality of protection perspective," *IEEE Wireless Commun.*, vol. 22, no. 4, pp. 104–112, Aug. 2015.

[8] X. Liang, K. Zhang, X. Shen, and X. Lin, "Security and privacy in mobile social networks: Challenges and solutions," *IEEE Wireless Commun.*, vol. 21, no. 1, pp. 33–41, Feb. 2014.

[9] K. Yang, X. Jia, K. Ren, R. Xie, and L. Huang, "Enabling efficient access control with dynamic policy updating for big data in the cloud," in *Proc. IEEE INFCOM*, Toronto, ON, Canada, 2014, pp. 2013–2021.

[10] E. Tews and M. Beck, "Practical attacks against WEP and WPA," in *Proc. ACM WiSec*, Zürich, Switzerland, 2009, pp. 79–86.

[11] B. Bertka, "802.11w security: DoS attacks and vulnerability controls," in *Proc. IEEE Infocom*, Orlando, FL, USA, 2012.

[12] M. Eian and S. F. Mjølsnes, "The modeling and comparison of wireless network denial of service attacks," in *Proc. ACM MobiHeld*, 2011, Art. no. 7.

[13] J. Pang, B. Greenstein, R. Gummadi, S. Seshan, and D. Wetherall, "802.11 user fingerprinting," in *Proc. ACM MobiCom*, Montreal, QC, Canada, 2007, pp. 99–110.

[14] F. Guo and T.-C. Chiueh, "Sequence number-based MAC address spoof detection," in *Proc. 8th Int. Symp. Recent Adv. Intrusion Detection (RAID)*, Seattle, WA, USA, 2005, pp. 309–329.

[15] D. B. Faria and D. R. Cheriton, "Radio-layer security: Detecting identity-based attacks in wireless networks using signalprints," in *Proc. ACM WiSe*, 2006, pp. 43–52.

[16] G. Chandrasekaran, J.-A. Francisco, V. Ganapathy, M. Gruteser, and W. Trappe, "Detecting identity spoofs in IEEE 802.11e wireless networks," in *Proc. IEEE Globecom*, Honolulu, HI, USA, 2009, pp. 1–6.

[17] N. Patwari and S. K. Kasera, "Robust location distinction using temporal link signatures," in *Proc. ACM MobiCom*, Montreal, QC, Canada, 2007, pp. 111–122.

[18] L. Xiao, L. Greenstein, N. Mandayam, and W. Trappe, "A physical-layer technique to enhance authentication for mobile terminals," in *Proc. IEEE ICC*, Beijing, China, 2008, pp. 1520–1524.

[19] Z. Jiang, J. Zhao, X.-Y. Li, J. Han, and W. Xi "Rejecting the attack: Source authentication for Wi-Fi management frames using CSI information," in *Proc. IEEE INFOCOM*, Turin, Italy, 2013, pp. 2544–2552.

[20] J. Xiong and K. Jamieson, "SecureArray: Improving WiFi security with fine-grained physical-layer information," in *Proc. ACM MobiCom*, Miami, FL, USA, 2013, pp. 441–452.

[21] S. Jana and S. K. Kasera, "On fast and accurate detection of unauthorized wireless access points using clock skews," in *Proc. ACM MobiCom*, San Francisco, CA, USA, 2008, pp. 104–115.

[22] V. Brik, S. Banerjee, M. Gruteser, and S. Oh, "Wireless device identification with radiometric signatures," in *Proc. ACM MobiCom*, 2008, San Francisco, CA, USA, pp. 116–127.

[23] G. E. Suh and S. Devadas, "Physical unclonable functions for device authentication and secret key generation," in *Proc. ACM DAC*, San Diego, CA, USA, 2007, pp. 9–14.

[24] K. Zeng, K. Govindan, and P. Mohapatra, "Non-cryptographic authentication and identification in wireless networks [security and privacy in emerging wireless networks]," *IEEE Wireless Commun.*, vol. 17, no. 5, pp. 56–62, Oct. 2010.

[25] R. Nandakumar, K. K. Chintalapudi, and V. Padmanabhan, and R. Venkatesan, "Dhwani: Secure peer-to-peer acoustic NFC," in *Proc. ACM SIGCOMM*, Hong Kong, 2013, pp. 63–74.

[26] D. Chen *et al.*, "Wireless device authentication using acoustic hardware fingerprints," in *Proc. BigCom*, Taiyuan, China, 2015, pp. 193–204.

[27] W. Xi *et al.*, "Locating sensors in the wild: Pursuit of ranging quality," in *Proc. ACM SenSys*, Zürich, Switzerland, 2010, pp. 295–308.

[28] D. Chen *et al.*, "SmokeGrenade: An efficient key generation protocol with artificial interference," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 11, pp. 1731–1745, Nov. 2013.

[29] W. Wang, Y. Chen, and Q. Zhang, "Privacy-preserving location authentication in Wi-Fi networks using fine-grained physical layer signatures," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1218–1225, Feb. 2016.

[30] B. Danev, H. Luecken, S. Capkun, and K. E. Defrawy, "Attacks on physical-layer identification," in *Proc. ACM WiSec*, Hoboken, NJ, USA, 2010, pp. 89–98.

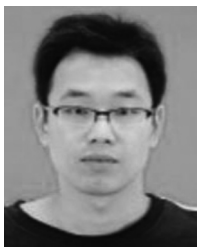
[31] A. Das, N. Borisov, and M. Caesar, "Do you hear what I hear?: Fingerprinting smart devices through embedded acoustic components," in *Proc. ACM CCS*, Scottsdale, AZ, USA, 2014, pp. 441–452.

[32] Z. Zhou, W. Diao, X. Liu, and K. Zhang, "Acoustic fingerprinting revisited: Generate stable device id stealthily with inaudible sound," in *Proc. ACM CCS*, Scottsdale, AZ, USA, 2014, pp. 429–440.

[33] T. Zhu, Q. Ma, S. Zhang, and Y. Liu, "Context-free attacks using keyboard acoustic emanations," in *Proc. ACM CCS*, Scottsdale, AZ, USA, 2014, pp. 453–464.

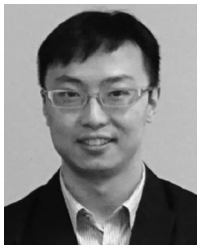
[34] T. L. Szabo, "Time domain wave equations for lossy media obeying a frequency power law," *J. Acoust. Soc. Amer.*, vol. 96, no. 1, pp. 491–500, 1994.

[35] A. Madhavapeddy, R. Sharp, D. Scott, and A. Tse, "Audio networking: The forgotten wireless technology," *IEEE Pervasive Comput.*, vol. 4, no. 3, pp. 55–60, Jul./Sep. 2005.



Dajiang Chen (M'15) received the B.Sc. degree from Neijiang Normal University, Neijiang, China, in 2005, the M.Sc. degree from Sichuan University, Chengdu, China, in 2009, and the Ph.D. degree in information and communication engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, in 2014.

He is currently a Post-Doctoral Fellow with the University of Waterloo, Waterloo, ON, Canada, and also with the School of Information and Software Engineering, UESTC. His current research interests include information theory and channel coding and their applications in wireless network security and wireless communications.

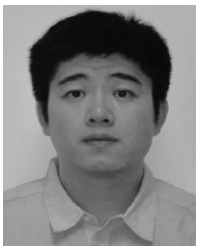


Ning Zhang (S'12–M'16) received the B.Sc. degree from Beijing Jiaotong University, Beijing, China, in 2007, the M.Sc. degree from the Beijing University of Posts and Telecommunications, Beijing, in 2010, and the Ph.D. degree from the University of Waterloo, Waterloo, ON, Canada, in 2015.

He is currently a Post-Doctoral Fellow with the University of Waterloo. His current research interests include dynamic spectrum access, 5G, physical layer security, and vehicular networks.

Zhen Qin (M'15) received the B.Sc. degree in communication engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2005, the M.Sc. degree in electronic engineering from the Queen Mary University of London, London, U.K., in 2007, and the M.Sc. and Ph.D. degrees in communication and information system from UESTC, in 2008 and 2012, respectively.

He is currently a Lecturer with the School of Communication and Information Engineering, UESTC. His current research interests include network measurement, wireless sensor networks, and mobile social networks.



Xufei Mao (M'10) received the bachelor's degree in computer science from the Shenyang University of Technology, Shenyang, China, in 1999, the M.S. degree in computer science from Northeastern University, Shenyang, in 2003, and the Ph.D. degree in computer science from the Illinois Institute of Technology, Chicago, IL, USA, in 2010.

He is with the School of Software and TNLIST, Tsinghua University, Beijing, China. His current research interests include span wireless *ad-hoc* networks, wireless sensor networks, pervasive computing, mobile cloud computing, and game theory.

Zhiguang Qin (S'95–A'96–M'14) is the Dean of the School of Software, University of Electronic Science and Technology of China (UESTC), Chengdu, China, where he is also the Director of the Key Laboratory of New Computer Application Technology and the Director of UESTC-IBM Technology Center. His current research interests include computer networking, information security, cryptography, information management, intelligent traffic, electronic commerce, distribution, and middleware.

Xuemin (Sherman) Shen (M'97–SM'02–F'09) received the B.Sc. degree from Dalian Maritime University, Dalian, China, in 1982, and the M.Sc. and Ph.D. degrees from Rutgers University, New Brunswick, NJ, USA, in 1987 and 1990, respectively, all in electrical engineering.

He is a Professor and the University Research Chair with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. He was an Associate Chair for Graduate Studies from 2004 to 2008. He has authored or co-authored over 600 papers and book chapters in wireless communications and networks, control and filtering. His current research interests include resource management in interconnected wireless/wired networks, wireless network security, social networks, smart grid, and vehicular *ad hoc* and sensor networks.

Dr. Shen served as the Technical Program Committee Chair/Co-Chair for IEEE Infocom'14, IEEE VTC'10 Fall, and the Symposia Chair for IEEE ICC'10. He also serves/served as an Editor-in-Chief for *IEEE Network*, *Peer-to-Peer Networking and Application*, and *IET Communications*; the Founding Area Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS; an Associate Editor for the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, *Computer Networks*, and *ACM/Wireless Networks*; and a Guest Editor for the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, IEEE WIRELESS COMMUNICATIONS, and the *IEEE Communications Magazine*. He is a Registered Professional Engineer of Ontario, Canada, an Engineering Institute of Canada Fellow, a Canadian Academy of Engineering Fellow, and a Distinguished Lecturer of the IEEE Vehicular Technology Society and Communications Society.

Xiang-Yang Li (M'99–SM'08–F'15) received the bachelor's degree in computer science and the bachelor's degree in business management from Tsinghua University, Beijing, China, in 1995, and the M.S. and Ph.D. degrees in computer science from the University of Illinois at Urbana–Champaign, Champaign, IL, USA, in 2000 and 2001, respectively.

He is a Professor with the University of Science and Technology of China, Hefei, China. His current research interests include wireless networking, mobile computing, security and privacy, cyber physical systems, smart grids, social networking, and algorithms.

Dr. Li was a recipient of the China NSF Outstanding Overseas Young Researcher Award (B). He and his students were the co-recipients of five Best Paper Awards, such as the IEEE GlobeCom 2015, the IEEE HPCCC 2014, ACM MobiCom 2014, COCOON 2001, and the IEEE HICSS 2001. He served or is currently serving as an Editor of several journals, including the IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS and the IEEE TRANSACTIONS ON MOBILE COMPUTING. He holds an EMC-Endowed Visiting Chair Professorship with Tsinghua University. He served at various capacities (the Conference Chair, the TPC Chair, the Local Arrangement Chair, and the Technical Program Committee) for a number of conferences such as the TPC Co-Chair of ACM MobiHoc 2014 and the IEEE MASS 2013 and a TPC member of ACM MobiCom from 2014 to 2015. He is an ACM Distinguished Scientist.