A Nonlinearity-Based Secure Face-to-Face Device Authentication for Mobile Devices

Xiaoyu Ji, *Member, IEEE*, Xinyan Zhou, *Member, IEEE*, Chen Yan, *Member, IEEE*, Jiangyi Deng, *Member, IEEE*, and Wenyuan Xu, *Member, IEEE*

Abstract—With the proliferation of mobile devices, face-to-face device-to-device (D2D) communication has been applied to a variety of daily scenarios such as mobile payment and short distance file transfer. In D2D communications, a critical security problem is to verify the device legitimacy when they share no secrets in advance. Previous research proposed device authentication schemes based on pre-built database or exploiting physical properties. However, a remaining challenge is to secure face-to-face D2D communication even in the middle of a crowd, within which an attacker may hide. In this paper, we present NAuth, a nonlinearity-enhanced, location-sensitive authentication mechanism. Especially, we target at the secure authentication within a limited range such as 20 cm, which is typical for face-to-face scenarios. NAuth designs a *verification scheme* based on the nonlinear distortion of speaker-microphone systems and a location-based *validation model*. The verification scheme guarantees device authentication consistency by extracting acoustic nonlinearity patterns (ANP) while the validation model ensures device legitimacy by measuring the time difference of arrival (TDOA) at two microphones. We analyze the feasibility and security of NAuth theoretically and evaluate its performance experimentally. Results demonstrate that NAuth can verify the device legitimacy in the presence of nearby attackers.

Index Terms—Device authentication, nonlinear distortion, mobile devices, face-to-face

1 Introduction

Mobile devices are becoming increasingly prevalent in our daily life. It is reported that over 63 percent of the network traffic came from mobile devices in 2017 [1]. With this growing trend, face-to-face Device-to-Device (D2D) communication has emerged and involves a pair of devices nearby to communicate directly, e.g., face-to-face mobile payment [2] and short distance file transfer.

In many scenarios, one may launch secure communication with the help of a trusted management center or a negotiated password. However, in D2D communication, a common case usually occurs between two devices sharing no secrets in advance, and it is important to ensure that they are indeed communicating with each other even if many other devices are around. Considering the mobile payment in Fig. 1, the payer device should authenticate the legitimacy of the payee device (cashing machine), under the risk of nearby attackers (fake cashing machines). Typically, standard protocols such as Bluetooth ask the payee to input a "code" provided by the payer, thereby ensuring the authentication of the payee. Such an approach mandates user intervention and the security cannot be guaranteed [3], [4].

- *Xiaoyu Ji, Chen Yan, Jiangyi Deng, and Wenyuan Xu are with the College of Electrical Engineering Zhejiang University, Hangzhou, Zhejiang 310027, China. E-mail: {xji, yanchen, jydeng, xuwenyuan}@zju.edu.cn.*
- *Xinyan Zhou is with the Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo, Zhejiang 315211, China. E-mail: zhouxinyan@nbu.edu.cn.*

Manuscript received 9 Feb. 2020; revised 14 July 2020; accepted 4 Sept. 2020. Date of publication 18 Sept. 2020; date of current version 4 Mar. 2022. (Corresponding author: Xinyan Zhou.)
Digital Object Identifier no. 10.1109/TMC.2020.3025023

To eliminate such levels of user intervention, alternative solutions are proposed for device authentication in D2D communications. Typical approaches are extracting reciprocity physical layer information, including RSS (received signal strength) [5], [6], [7] and CSI (channel state information) [8], [9], [10], [11]. However, RSS-based mechanisms have a limited bit generation rate while CSI-based mechanisms rely on specified hardware equipment (Intel 5300 Wifi card). Another type of work utilizes the physical randomness of the environment to extract symmetric keys [12], [13], [14]. Besides, Xie *et al.* [15] proposed a device authentication and key agreement mechanism that extracts ACR (acoustic channel response) as device features and utilizes a response interval to verify the device legitimacy.

Although the aforementioned approaches improve the convenience by reducing user intervention, they may be bypassed by an attacker located close to the device to be authenticated [4]. In Fig. 1, for example, the fake cashing machine can impersonate the genuine one and trigger mistakenly transfer money from the payer device. Such a threat is made possible because of the low location-sensitivity of the medium for key extraction, i.e., devices nearby may extract similar keys from the radio channel or the acoustic channel.

In this paper, we focus on the device authentication problem in face-to-face D2D communication in the presence of nearby attackers. Specifically, face-to-face indicates a short communication range, e.g., 20 cm for common mobile payment scenarios. For such a scenario, we propose NAuth, a nonlinearity-enhanced, location-sensitive authentication mechanism for secure authentication. The key insight of NAuth is to utilize the nonlinear distortions for authentication. Nonlinear distortions are essentially fine-grained and location-sensitive because they are combinations of multiple frequency harmonics. In

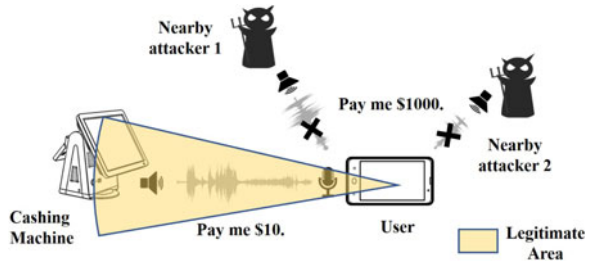


Fig. 1. A NAuth-based mobile payment scenario. NAuth can authenticate the legitimate device and detect nearby hidden attackers.

particular, we extract the acoustic nonlinear distortions of the speaker-microphone system (SMS), which commonly exist among current mobile devices. Moreover, a location-based security model is designed to reduce the legitimate area and to decrease the chances of attacks. The high-resolution nonlinearity feature together with the location-based security model eliminate any attacks within the legitimate area, and hereafter we name the two components the nonlinearity-based *verification scheme* and the location-based *validation model*.

The design of NAuth depends on exploring the following questions. First, *can nonlinear distortions be utilized for device authentication?* The basic requirement for device authentication is that the nonlinear distortion should be unique and device-dependent. Moreover, the nonlinear distortion characteristic should be hard to imitate, otherwise, the attacker can replay the signals easily. Second, *how to extract sufficient nonlinear distortion characteristics for device authentication?* Even if nonlinear distortion can be used for device authentication, it is unknown whether and how it can be applied for real D2D applications. Last but not the least, *how to guarantee the extracted nonlinear distortion characteristics come from the legitimate device?* If the source legitimacy cannot be guaranteed, the extracted characteristics are thus invalid.

To tackle the questions above, we first explore the nonlinear distortions for speaker-microphone systems and validate the fact that nonlinear distortions are both device and location dependent, which are essential for device authentication. We derive unique patterns, i.e., the *acoustic nonlinearity patterns* (ANP), with an elaborately designed amplitude modulation (AM) signal. The verification scheme verifies device consistency during the authentication process. Besides, we design a lightweight location-based model to validate the source location by measuring the time difference of arrival (TDOA) at two microphones. NAuth can be utilized in various application scenarios, including mobile payment, data transmission, etc.

We summarize our main contributions as follows:

- We propose and validate that nonlinearity can be used as a fine-grained feature for device authentication with a speaker-microphone system.
- We design NAuth, a secure and location-sensitive device authentication mechanism for face-to-face D2D communications built on a nonlinearity-based verification scheme and a location-based validation model.
- We evaluate the performance and analyze the security of NAuth. Theoretical and experimental results prove the efficiency and security of our mechanism.

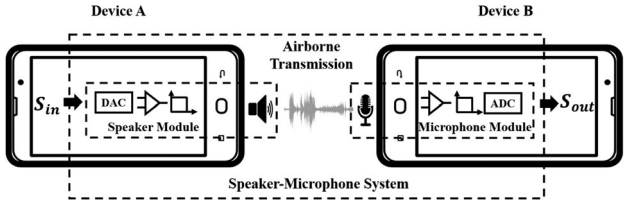


Fig. 2. A speaker-microphone system.

The rest of the paper is organized as follows. We first introduce the nonlinear distortion characteristic and investigate its feasibility for device authentication in Section 2. We then present an acoustic nonlinear pattern (ANP) in Section 3. We describe the threat model and application scenarios in Section 4. We present the detailed design of NAuth in Section 5 and analyze the security of NAuth in Section 6. Section 7 evaluates the performance of NAuth. We make some complementary discussion in Section 8 and the related work is summarized in Section 9. We conclude our work in Section 10.

2 Nonlinearity of Speaker-Microphone Systems For Device Authentication

2.1 Nonlinearity in Speaker-Microphone System

Microphones and speakers are transducers that convert signals between acoustic and electrical states. For the purpose of user experience, stereo effect and noise canceling, most smart devices (iPhone, Echo, etc.) are built with two or more modules of both microphones and speakers. For example, even early versions of smartphones (e.g., iPhone 5) have three microphones and two speakers [16]. Multiple signal processing circuits are utilized in microphone and speaker modules. Taking the microphone module as an example, the converted electrical signals are processed by multiple stages of amplifiers and low-pass filters (LPF) before being sampled by the analog-to-digital converter (ADC).

For a *speaker-microphone System (SMS)*, the signal goes through three stages in the speaker-microphone channel in sequence—a speaker module, airborne transmission, and a microphone module, as shown in Fig. 2. Ideally, one can expect the speaker-microphone system to be linear, which means for a given input signal S_{in} at the speaker module, the output S_{out} at the microphone module is

$$S_{out} = AS_{in}. \tag{1}$$

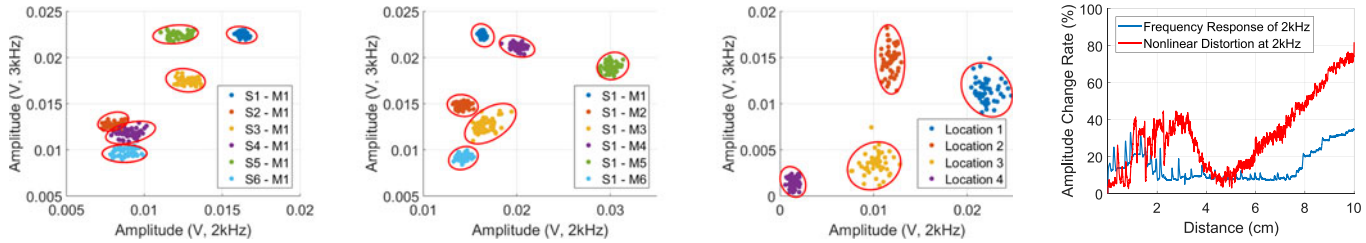
where A is the amplification factor.

However, real speaker-microphone systems are nonlinear because the signal processing circuits are made of nonlinear electronic components, e.g., transistors and the transducers are nonlinear [17], [18]. In general, a nonlinear system can be modeled as the following polynomial equation:

$$S_{out} = a_0 + a_1 S_{in} + a_2 S_{in}^2 + a_3 S_{in}^3 + \cdots = \sum_{i=0}^{\infty} a_i S_{in}^i, \tag{2}$$

where a_i is the corresponding polynomial coefficient.

Speaker-Microphone System Nonlinearity. Besides the linear component $a_1 S_{in}$ in Eq. (2), S_{out} contains nonlinear distortions including a DC signal a_0 and $\{a_i S_{in}^i\}(i > 1)$, which are exponents of the input. Nonlinearity can deteriorate the



(a) Speaker dependence. (b) Microphone dependence. (c) Location dependence. (d) Location sensitivity comparison.

Fig. 3. The amplitudes of harmonics on $2kHz$ and $3kHz$ ($f_b = 1kHz$) when sending signals (a) from 6 stand-alone speakers to the same microphone, (b) from the same speaker to 6 different microphones, (c) with the same SMS at 4 locations. (d) The amplitude change rates for nonlinear distortion and fundamental frequency response at different speaker-microphone distances.

output signals and has unexpected consequences. Despite the manufacturers' efforts in designing linear electronic components especially within the commonly used 100Hz to 10kHz frequency range, nonlinear distortion is still a common phenomenon among microphone and speaker modules.

2.2 Distinct Nonlinearity of Devices

The speaker-microphone system demonstrates inevitable nonlinearity and one can formulate the relationship between S_{out} and S_{in} by a vector, named the *nonlinear coefficient vector* $V = [a_0, a_1, a_2, \ldots, a_n]$. Essentially, V is determined by the physical structures of the nonlinear components, i.e., the CMOS chips [19] in both speaker and microphone modules (the nonlinearity caused by the air is ignored). As a result, the nonlinearity of an SMS varies among devices. Moreover, the nonlinearity can be easily observed and quantified because the nonlinear distortions are at different frequency bands from the original input signals. For example, let $S_{in} = \sin(2\pi f_0 t)$, the output of the speaker-microphone system S_{out} would have $2f_0, 3f_0, \ldots, nf_0$ frequency harmonics. Therefore, it is feasible for us to utilize the nonlinearity to identify a device (either the speaker or the microphone) in the speaker-microphone system. In this paper, we authenticate a device (Device A in Fig. 2) by looking at the nonlinearity of the signals received on the microphone side of Device B.

2.3 Feasibility Validation and Results

Experimentally, we validate the feasibility and effectiveness of speaker-microphone nonlinearity for device authentication. We experiment on 6 stand-alone microphone modules and 6 speakers, both of which are of the same model, and the details are shown in Fig. 11a. The parameters of the speakers are 8Ω and $0.5W$, and each microphone module consists of a MEMS microphone chip ADMP401 [20], an impedance converter and an output amplifier. We stimulate the speakers with a 1kHz tone of $1.5Vpp$ from a function generator and collect the output signals of the microphone modules with a Keysight U2541A data acquisition card (DAQ) sampling at 100kHz. We conduct the experiment in a quiet meeting room, and the distance between the microphones and speakers is $3cm$. In the following, we examine whether the nonlinearity is device dependent and location dependent, which are basic requirements for location-sensitive device authentication.

2.3.1 Device Dependence

We examine the nonlinearity behaviors of both the speaker and the microphone modules separately. a) We utilize 6 speaker modules (same model) to stimulate an identical microphone, and b) we use the same speaker to stimulate 6 microphone modules (same model) under the above settings and record the frequency response at the microphone (s) side. For each SMS, we collect 50 $10ms$-long samples of the microphone output, perform Fast Fourier Transform (FFT) analysis on the 50 samples, and extract the amplitudes of the $2nd$ and $3rd$ harmonics, i.e., at 2kHz and 3kHz for simplicity (higher harmonics can also be utilized).

The results are shown in Figs. 3a and 3b. Generally, samples from the same speaker-microphone system show well-marked clustering characteristics, and samples from different SMSes can be easily separated in a two-dimensional plane (i.e., 2kHz Amplitude as X-Axis and 3kHz Amplitude as Y-Axis). This confirms our assumption that both speakers and microphones share nonlinearity-specific properties. Though the samples from speaker 2 and 4 (S2-M1 and S4-M1) in Fig. 3a partly overlap, they can be distinguished in a higher dimensional space by including more harmonics.

2.3.2 Location Dependence

To validate the location-dependence of nonlinearity, we transmit a 1kHz tone from the built-in speaker of a Huawei P10 Plus smartphone and record the frequency responses with an iPhone 6s at 4 different locations in a line. The two devices are lying on a table, with one's bottom speaker opposite to the bottom microphone of another. For each location, the distance between the Huawei smartphone (speaker) to the iPhone (microphone) is $1cm$, $3cm$, $5cm$ and $8cm$ respectively by moving the Huawei smartphone (we use commercial smartphones here because it is convenient to get moved). We extract the 2kHz and 3kHz harmonics at the microphone side. The results are shown in Fig. 3c. We can find that at different locations, the nonlinear distortions are also clearly clustered. Moreover, the location with larger distance results in weaker harmonics, nevertheless, they can be classified by involving more dimensions, i.e., higher harmonics.

2.3.3 Location Sensitivity

One may argue that the fundamental frequency response of a speaker-microphone system can also be utilized for authentication, i.e., by measuring the frequency response at

2kHz with a 2kHz input [15]. We demonstrate the advantage of a nonlinearity-based approach over fundamental-frequency-based in terms of location-sensitivity, which enhances security. With the same setup in the location dependence experiment, we extract both amplitudes of 1) the harmonic signal at 2kHz stimulated by a 1kHz signal, and 2) the 2kHz fundamental frequency with a 2kHz input, and gradually increase the distance between the two devices from $0cm$ to $10cm$.

We formally define the *Amplitude Change Rate* of frequency f at distance d as:

$$\Upsilon(d)_f = \frac{|A(d)_f - A(d_0)_f|}{A(d_0)_f} * 100\%, \tag{3}$$

where $A(d)_f$ is the amplitude of frequency f at distance d and in our case $d_0 = 0cm$. The results are shown in Fig. 3d. Compared with the fundamental frequency response, the $\Upsilon(d)_f$ of nonlinear distortion is higher, which coincides the location-sensitive property, and thus can be more secure.

2.4 Summary

The nonlinearity of speaker-microphone systems is demonstrated to be speaker/microphone specific as well as location-sensitive, which makes it a natural candidate for device authentication.

In the following sections, we first describe how to extract the nonlinearity of speaker-microphone systems efficiently, and then we elaborate the design details of using nonlinearity for device authentication in Section 5.

3 Acoustic Nonlinear Patterns

We evaluate the feasibility of utilizing speaker-microphone systems nonlinearity for device authentication in Section 2. In this section, we discuss how to express and extract the nonlinearity by answering the following questions: (1) How to formulate and express the nonlinearity of the system as the acoustic nonlinear patterns (ANP) for device authentication, considering accuracy and the computation cost, (2) In what way and how to extract ANP, in an efficient way especially.

For a clear presentation, we first summarize the symbols utilized in ANP in Table 1.

3.1 Acoustic Nonlinearity Patterns (ANP)

Basically, we attempt to extract the nonlinear characteristic of the speaker-microphone system to verify an identity, and we call it the acoustic nonlinearity patterns (ANP) in the following. Recall that in Eq. (2), we denote the nonlinear coefficient vector $V = [a_0, a_1, \ldots, a_i, \ldots]$ to describe the relationship of the input and the output signals. a_i is the gain of the ith harmonic and is observable in the frequency domain of the output signal. Intuitively, one can use V for nonlinearity pattern extraction. However, obtaining the value of a_i is difficult. First, a_i cannot be separated in the time domain because harmonics at a certain frequency is a combination of multiple nonlinear components ($a_i S^i_{in}$). Therefore, we propose acoustic nonlinearity patterns (ANP) as an alternative to the nonlinear coefficient vector V.

TABLE 1
Summary of Symbols in ANP

Symbols	Description
ANP	The acoustic nonlinearity patterns.
A	The signal amplitude.
f_c	The frequency of the carrier in the AM signal.
f_b	The frequency of the baseband in the AM signal.
f_{LPF}	The cut-off frequency of the low-pass filter.
A_{f_c}	The amplitude of the carrier signal.
A_{f_b}	The amplitude of the baseband.
$f_{S^i_{in}}$	The frequency components for ith exponent in S_{out}.
$c_{(i,j)}$	The corresponding coefficient for $sin(2\pi j f_b t)$ after the trigonometric expansion of S^i_{in}.

Considering an input signal $S_{in} = sin2\pi f_0 t$, the new frequency components in S_{out} contain $\{f_0, 2f_0, \ldots, nf_0\}, n \in N^+$. Despite of the ambient noise, the amplitudes of these new frequency components are linear combinations of the nonlinear coefficient vector, which can be presented as:

$$A(nf_0) = \sum_{i=1}^{\infty} A^i a_i c_{(i,n)}, \tag{4}$$

where $A(nf_0)$ is the amplitude of the nf_0, A^i is the signal gain and $c_{(in)}$ is a constant determined by the input signal, which can be calculated by trigonometric expansion. For example, if $S_{in} = \sin(2\pi 1000t)$, $nf_0 = \{1kHz, 2kHz, \ldots, nkHz\}, n \in N^+$, and for $i = 3$, we have $c_{(3,1)} = 0.75$, $c_{(3,2)} = 0$ and $c_{(3,3)} = -0.25$.

Based on the analysis above, the amplitudes of harmonics are the linear combination of a_i. Thus, we define the acoustic nonlinearity patterns (ANP) as:

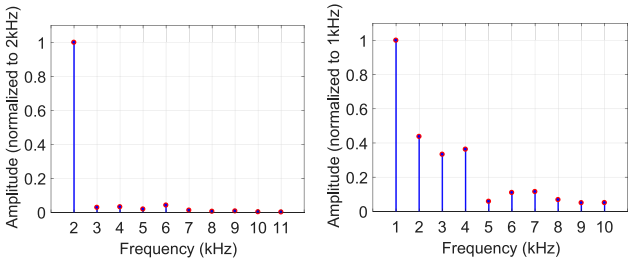
$$ANP = [\ A(f_{nd})\ ]. \tag{5}$$

Taking the same input above as an example, $ANP = [A(1kHz), A(2kHz), \ldots, A(nkHz)]$. One key advantage of selecting amplitudes of harmonics as the ANP is the low computation cost. Extracting ANP is feasible for mobile devices while they only need to apply FFT and extract the amplitudes of new frequency components after nonlinear distortions from authentication signals. However, extracting ANP for device authentication still faces several challenges. In order to derive a fine-grained ANP efficiently, we elaborately design the authentication signal in the following.

3.2 ANP Extraction

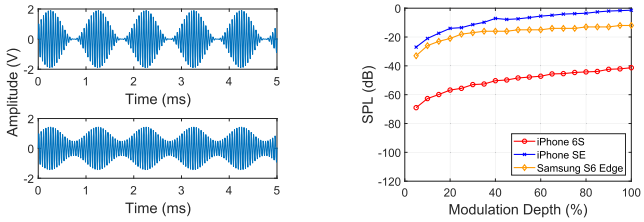
Nonlinear distortion is an unexpected phenomenon since it leads to audio quality degradation and affects the user experience. To mitigate this, nonlinear distortions of signals at frequencies within $100Hz$ to 10kHz are elaborately relieved by manufacturers, which brings challenges for ANP extraction.

To investigate, we measure the frequency response of a microphone module (the same as the one used in Section 2) with a 1kHz input signal. The distance between the speaker and the microphone is $5cm$, and we show the normalized amplitudes in Fig. 4a. The experiment result reveals higher harmonics (e.g., 3kHz, 4kHz, etc.) are even unobservable when compared with the second harmonic (2kHz). Extracting amplitudes of such weak harmonics may bring



(a) From a $1kHz$ signal. (b) From an AM signal.

Fig. 4. The frequency response of (a) a signal at $1kHz$ and its harmonics and (b) the nonlinear distortions of an AM signal with $f_c = 20kHz$ and $f_b = 1kHz$. Higher harmonics can be extracted when the signal is modulated.



(a) The input signal in time domain with different modulation depth: 100% (top) and 50% (bottom).

(b) Amplitudes of nonlinear distortion at f_b with different modulation depths on three smartphones.

Fig. 5. Evaluation of modulation depth upon the efficiency of the nonlinear patterns. ($f_c = 20kHz$, $f_b = 1kHz$).

unpredictable measurement errors and impact authentication accuracy.

One may consider sending signals with frequencies above 10kHz to obtain a stronger nonlinear distortion. However, the commodity microphones and speakers are embedded with low-pass filters (LPF) to filter the high-frequency interference. Therefore, signals with frequencies higher than the cut-off frequency of the LPF (e.g., 20kHz) will be removed, including harmonics. Thus, sending an unmodulated high-frequency signal shows inefficiency in ANP extraction.

To tackle above challenges, we specifically design the signal and discuss the signal parameters in the following.

3.2.1 Extracting ANP by AM Modulation

To pass through the LPF while preserving high harmonics, we use amplitude modulation (AM) to produce intentional nonlinear distortion. To be specific, we modulate a baseband signal upon a carrier signal whose frequency is far above than 10kHz (e.g., at 20kHz) to produce significant harmonics, as is used in [17]. We notate the carrier and baseband frequencies f_c and f_b, and the AM signal is presented as:

$$S_{in} = A_{f_c} sin(2\pi f_c t)(1 + A_{f_b} sin(2\pi f_b t)), \tag{6}$$

where A_{f_c} and A_{f_b} are the amplitudes of the carrier and baseband signals. Considering the nonlinear relationship of S_{in} and S_{out} in Eq. (2), when the input signal is an AM signal, the new frequency components for the ith exponent in S_{out} are:

$$f_{S_{in}^i} = \begin{cases} kf_c,\ mf_c \pm nf_b, & k \in \{1,3,\ldots,i\},\ i\ is\ odd \\ & m \in \{1,3,\ldots,i\} \\ & n \in \{1,2,\ldots,i\} \\ kf_b,\ mf_c \pm nf_b, & k \in \{1,2,\ldots,i\},\ i\ is\ even \\ & m \in \{2,4\ldots,i\} \\ & n \in \{0,1,2,\ldots,i\} \end{cases}, \tag{7}$$

where $f_{S_{in}^i}$ is the frequency components in S_{in}^i.

From Eq. (7), we can find that there are abundant harmonic components whose frequencies can be below 20kHz. To name a few, kf_b and $(f_c - nf_b)$ can produce frequencies less than 20kHz. The ANP is actually extracted from those new low-frequencies ($<$ 20kHz) composed of harmonic components of other frequencies under AM modulation.

To validate, we send an AM signal with f_c=20kHz and f_b=1kHz to a microphone module and measure the amplitudes of harmonics. The frequency response is shown in Fig. 4b with the same setting as in Fig. 4a. Compared to Fig. 4a, the 1kHz-generated harmonics at 2kHz, 3kHz and even 8kHz demonstrate strong power than non-modulated input signals. Therefore, an AM signal with a high-frequency carrier can enhance the nonlinear distortions, which is beneficial to the ANP extraction.

3.2.2 AM Parameters

In the modulation process, three parameters, i.e., f_c, f_b and modulation depth (A_{f_b}/A_{f_c}), should be carefully selected to improve the effectiveness of ANP extraction.

f_c and f_b. The maximum value of f_c is constrained by the sampling rate of DAC in the speaker module. Based on the Nyquist sampling theorem, f_c should be less than the half of the sampling rate. Besides, sending an inaudible signal is a user-friendly choice while audible signals are annoying and inconvenient in some occasions. Typically, the upper bound frequency of human hearing is 20kHz, and the value decreases to 16kHz for adults. Furthermore, to capture enough nonlinear distortions, intuitively f_b should be as small as possible while f_c should be the opposite. Due to the constraint of low-pass filters, we should have:

$$\begin{gathered} N_{fh} \cdot f_b \le f_{LPF} \\ f_c - f_b \le f_{LPF} \\ 16kHz \le f_c \le \frac{f_{sp}}{2}, \end{gathered} \tag{8}$$

where f_{sp} is the sampling rate of the microphone, N_{fh} is the space of available harmonics, and typically we prefer a larger N_{fh} to extract efficient nonlinear patterns. The second condition should be satisfied because $f_c - f_b$ also contributes relatively strong harmonics than others. Typically, the sampling rates of devices are higher than $44.1kHZ$, and in our implementation, we select f_c=20kHz and f_b=1kHz empirically.

Modulation Depth. The modulation depth is defined as the ratio of the baseband amplitude and the carrier amplitude, i.e., A_{f_b}/A_{f_c}, which impacts the strength of nonlinear distortions. Fig. 5a is an example of two modulated signal in time domain with two modulation depths: 50 and 100 percent ($f_c = 20kHz$, $f_b = 1$kHz). The modulation depth impacts the nonlinear distortion while the low-frequency leakages are generated from the baseband. To figure out how modulation depth impacts the nonlinear distortion, we conduct experiments on 3 smartphones: iPhone 6S, iPhone SE and Samsung S6 Edge. We transmit a modulated signal with modulation depth from 5 to 100 percent ($f_c = 20$kHz, $f_b = 1$kHz) from a

speaker and measure the nonlinear distortion at 1kHz. The experiment results in Fig. 5b demonstrate that the efficiency of the nonlinear distortion improves with the increasing of the modulation depth. Therefore, we set the modulation depth to 100 percent to achieve the best nonlinear distortion.

In conclusion, ANP is a set of amplitudes of harmonics. To extract a fine-grained ANP, we design the authentication signal as an AM-modulated signal and select f_c=20kHz, f_b=1kHz and $A_{f_b}/A_{f_c} = 1$ respectively.

4 Threat Model and Overview

With a thorough description of the ANP above, we design NAuth. Basically, NAuth is designed to secure the face-to-face D2D communication through the nonlinear characteristic of the microphone-speaker-system and a location-sensitive authentication mechanism. In this section, we first illustrate the threat model and assumptions and provide a system overview of NAuth in the following.

4.1 Threat Model and Assumptions

The threat model involves two parties namely Alice and Bob that need to authenticate each other and an attacker named Eve. For simplicity, we only consider the case that Alice authenticates Bob. Eve's purpose is to make Alice believe she is Bob while Alice and Bob share no common secrets in advance.

Without loss of generality, we have the following assumptions for Alice and Bob:

- Alice and Bob are physically close to each other, namely within $50cm$ or closer. The distance can vary across D2D application scenarios, e.g., mobile payment (within $20cm$) or secret information sharing (within $50cm$).
- Alice is for sure that any device within a restricted "legitimate area" is trustworthy and she can control the orientation of her device to guarantee Bob is in the "legitimate area", as illustrated in Fig. 1. The design details of this "legitimate area" can be referred to Section 5.
- Both parties' devices have speakers and microphones. The party who initiates an authentication, e.g., Alice here, should have two microphones at least.
- Both parties' devices could be relatively stationary during the authentication process.

For Eve, she has the following capabilities and assumptions:

- Eve is free to move anywhere around Alice and Bob. She can even hide her attack equipment in the pocket or under a book. However, neither Eve or her equipment can be between Alice and Bob in face-to-face scenarios.
- Eve is able to capture and inject signals at any stage of the authentication process, and thereby launch replay or man-in-the-middle attacks.
- Eve may be aware of the authentication mechanism.

From the above threat model, a successful authentication relies on three important characteristics: 1) the consistency of the authenticated device can be guaranteed, 2) there is a "legitimate area", and 3) the area is reliable to differentiate attackers such as Eve. In the next section, we elaborate the location-based validation model and the nonlinearity-based verification scheme which satisfy the above requirements.

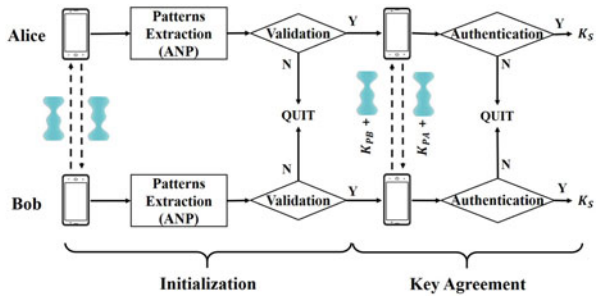


Fig. 6. A NAuth-based key establishment procedure.

4.2 System Overview With Key Establishment as an Example

NAuth is a location-sensitive device authentication mechanism built on two key components: a *verification scheme* based on the nonlinearity of speaker-microphone systems and a location-based *validation model*. They mainly address two challenges respectively:

1) How to authenticate a device from the sound it generates?
2) Is the received sound generated by a legitimate device?

To illustrate, we give an example of a secure key establishment process implemented with NAuth. As shown in Fig. 6, Alice and Bob are two legitimate users who need to establish a session key between their devices. The process consists of two steps: initialization and key agreement. To initialize, the two devices send acoustic *authentication signals* to each other and extract nonlinearity patterns from the sounds they receive. Besides, they independently verify the legitimacy of received sounds with the location-based validation model. After that, they can exchange their public keys K_{PA}, K_{PB} via acoustic signals and derive the same session key K_s while constantly sending *declaration signals* to verify the consistency of nonlinearity patterns and validate the source legitimacy.

In the following, we focus on the design of the verification scheme and the validation model in Section 5.

5 Design

In this section, we describe the design of the nonlinearity-based verification scheme and the location-based validation model in detail. We first summarize all the notations in Table 2 for a clear presentation.

5.1 Device Verification

5.1.1 ANP-Based Verification

To verify the authentication consistency of a device, NAuth requires devices to send declaration signals proactively. During the initialization process in Fig. 6, both Alice and Bob extract ANPs from the received declaration signals. The extracted ANP during the initialization process is regarded as the identity of the signal source. In the following key agreement process, devices should send authentication signals for further verification. The authentication signal is the

TABLE 2
Notations

Parameter	Description
d_{ij}	The euclidean distance between ANP_i and ANP_j.
σ	The threshold of the d_{ij} from the same device.
τ	The maximum interval to send an authentication signal.
$TDOA$	Time difference of arrival at two microphones.
L_2, L_1	The distances from the source device to the bottom microphone and the top microphone.
L_m	The distance between the top and bottom microphones.
c	The speed of sound, $340m/s$.
ε	The threshold of a legitimate device's TDOA.
η	The threshold of a legitimate device's pass rate.
$L_{shoulder}$	The width of the user's shoulder.
D_{u2m}	The distance between the user and the bottom microphone.

same one as the declaration signal, thus ANPs of these two signals should be similar for the same speaker-microphone system at the same location. For clarity, we utilize ANP_0 to represent the ANP of the declaration signal while using ANP_i as the ANP of the ith authentication signal.

To judge whether two ANPs are from the same device, a straightforward way is comparing the distance. In NAuth, we exploit the euclidean distance to determine whether two ANPs are consistent, specifically, the distance (d) between ANP_0 and ANP_1 is defined as:

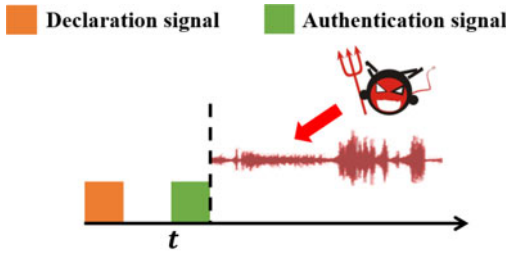
$$d_{01} = \sqrt{\sum_{i=1}^{N}(ANP_0(i) - ANP_1(i))^2}, \tag{9}$$

where N is the dimension of ANP. If these two ANPs are close enough, we consider they belong to the same speaker-microphone system. Therefore, if d is smaller than a predefined threshold σ, one can accept the authentication consistency, otherwise, a new authentication should be performed. We discuss the selection of σ in Section 7.

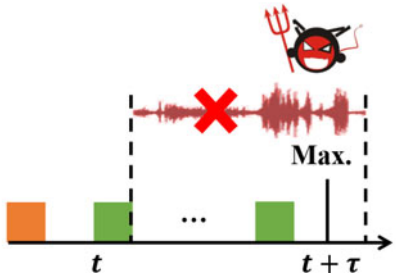
5.1.2 Verification Scheme

However, verifying ANPs during the key agreement still faces another challenge. Sensitive information including the session key cannot be encoded upon authentication signals. In NAuth, the data transmission mechanism is based on Dolphin [21], which adopts orthogonal frequency-division multiplexing (OFDM) to encodes digital signals to acoustic signals. Since the data transmission mechanism is not the main contribution of NAuth, the detailed design is omitted. The key point here is that the acoustic signal with the session key reveals significantly different characteristics with authentication signals. Therefore, by monitoring the channel, the attacker can recognize whether the previous signal is an authentication signal or data transmission, which provides a chance for attackers.

As shown in Fig. 7a, we use an orange block and a green block to represent the declaration signal and the authentication signal. If Alice only sends the authentication signal once, the attacker Eve can realize when the authentication process ends and replace the legitimate device (e.g., Alice) after the authentication process and send session key to Bob. By doing so, Eve may still have the chance to communicate with Bob as a legitimate identity and the key agreement process is still vulnerable to impersonation attack. To tackle this problem, we request devices actively claim their identities by sending authentication signals randomly multiple times.



(a) Sending authentication signal once is vulnerable since the attacker can inject any signal after hearing it.



(b) Sending authentication signals randomly with a maximum interval τ to avoid injection attack.

Fig. 7. Illustration of verification scheme in NAuth. Devices should send an authentication signal randomly multiple times in the key agreement process.

In NAuth, we set a maximum interval between two authentication signals. As shown in Fig. 7b, the first authentication signal is sent at time t and the maximum interval between authentication signals is τ. We request Alice to send the next authentication signal within the timeslot $(t, t+\tau)$. If Bob doesn't receive any authentication before $t+\tau$, he will restart the authentication process immediately. The selection of τ should also be carefully considered. A small τ will decay the key agreement efficiency while a large τ provides the opportunity for attackers to inject fraud information without being noticed. For the sake of security, τ should be at least smaller than the time duration that used for sending the public key, thus we have:

$$\tau < N_k * \frac{1}{v}, \tag{10}$$

where N_k is the length of the public key and v is the transmission rate. Considering the length of the public key in Diffie-Hellman protocol [22] is at least 128 bits and the data transmission rate in Dolphin [21] is $240bps$, we select τ as $250ms$ in NAuth.

To conclude, NAuth verifies the device's identity by comparing the distance between the ANPs extracted from the declaration signal and authentication signals. If $d_{0i} > \sigma$, the receiver should quit the authentication process and restart it again. During the key agreement process, the device should send authentication multiple times randomly with a maximum interval of $250ms$.

5.2 Location-Based Validation Model

The nonlinearity-based verification scheme can ensure authentication consistency. However, it cannot confirm the legitimacy

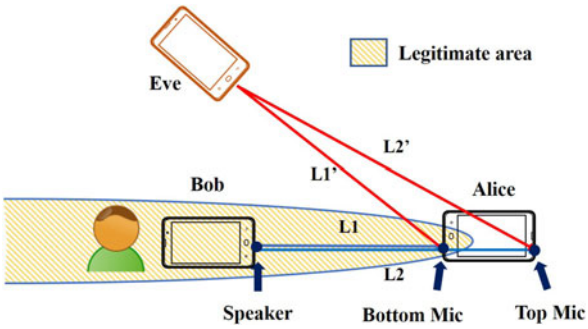


Fig. 8. A device (Bob) inside the legitimate area shows higher TDOA at the two microphones of the authenticator device (Alice) than a device (Eve) outside the legitimate area because $(L_2 - L_1) > (L_2^{'} - L_1^{'})$.

of the device. If an attacker (Eve) sends the authentication signal before a legitimate device (Bob) does, the attacker (Eve) can impersonate Bob because Alice fails to differentiate them. Therefore, the location-based validation model is introduced to distinguish legitimate devices.

5.2.1 TDOA-Based Validation

The location-based validation model utilizes two microphones embedded in devices. Typically, devices like smartphones and intelligent speakers are designed with at least two microphones to support various applications. We notice that when we record acoustic signals with both microphones, there is always a time difference, a.k.a. TDOA (time difference of arrival) because the distances between the signal source and two microphones are different.

An illustration of the location-based validation model is shown in Fig. 8. Signals from the speaker at the legitimate sender side are assumed to be along the connecting line of the two microphones at the receiver side. By measuring the TDOA at two microphones, we can approximately estimate the location of the signal source.

$$TDOA = \frac{L_2 - L_1}{c} = \frac{L_m}{c}, \tag{11}$$

where L_1 and L_2 are the distances from the source device to the bottom and top microphones respectively, c is the speed of sound and L_m is the distance between the two microphones. Both L_m and c are constants.

5.2.2 Legitimate Area

According to Eq. (11), the TDOA of an attacker (Eve) is smaller if she is not located on the connecting line of two microphones because $L_2^{'} - L_1^{'} < L_2 - L_1$, where $L_2^{'}$ and $L_1^{'}$ are distances from Eve's speaker to the two microphones. Therefore, $TDOA_{Eve}$ is smaller than $\frac{L_m}{c}$.

`NAuth` measures the TDOA by comparing the signal arrival time at two microphones, and the precision of the TDOA is constrained by the device sampling rate f_{sp}, i.e., the measurement of TDOA may have a maximum error of $1/f_{sp}$. Taking the accuracy error into consideration, `NAuth` validates the source device as a legitimate device if its TDOA satisfies the following constraint:

$$TDOA \geq \frac{L_m}{c} - \frac{1}{f_{sp}}, \tag{12}$$

For a specific device, both L_m and f_{sp} are constants. We notate $\varepsilon = \frac{L_m}{c} - \frac{1}{f_{sp}}$, and we then simplify Eq. (12) as:

$$TDOA \geq \varepsilon, \tag{13}$$

Therefore, points (P) in the legitimate area satisfy:

$$|PM_{top}| - |PM_{bottom}| \geq \left(L_m - \frac{c}{f_{sp}}\right), \tag{14}$$

where M_{top} and M_{bottom} are the top and the bottom microphones of the receiver. The boundary of the legitimate area is the equality condition of Eq. (14), which is the left branch of a hyperbola[1] with bottom and top microphones as the two foci. Thus, the device with TDOA satisfying Eq. (12), i.e., located inside the left branch of the hyperbola (the shaded area in Fig. 8), is considered as a legitimate device.

5.2.3 Legitimacy Validation Scheme

Ideally, the TDOA of legitimate users inside the legitimate area should maintain Eq. (12). However, the multipath effect cannot be ignored in practice. Obstacles in the environment may impact the propagation path of the acoustic signal, which may change the TDOA for both legitimate users and attackers. Fortunately, the ultrasound reveals naturally directionality, which significantly reduces the multipath effect. To improve the robustness of the validation model, we design the legitimacy validation scheme. To be specific, the receiver samples the TDOA multiple times to prevent accidental error for both users and attackers. Considering the dual microphones embedded in the device share the same system clock, the receiver can easily calculate TDOAs with minor overhead. In `NAuth`, the receiver should measure 10 TDOAs in $100ms$ and legitimate users should have a higher number of passes than attackers. As shown in Eq. (15), if the pass rate is higher than a predefined threshold η, one can accept the signal as a legitimate one.

$$Pr\{TDOA \geq \varepsilon)\} \geq \eta. \tag{15}$$

We select η as 70 percent in `NAuth`, and the evaluation of η is discussed in Section 7.

5.2.4 User Experience of Legitimacy Validation

The size of the legitimate area is a tradeoff between user experience and security. If the area is too small, legitimate users need to put two devices on a strictly straight line to pass the validation, which is hard for users. On the contrary, a bigger area may leave space for attackers.

Based on the analysis above, the size of the legitimate area is covered by the left branch of a hyperbola. To quantify the legitimate area, we introduce a Cartesian coordinate such that the origin is the center of two microphones and the x-axis is the main axis. We have the bottom microphone as $F_1 = (-\frac{L_m}{2}, 0)$ and the top microphone as $F_2 = (\frac{L_m}{2}, 0)$. With two fixed foci, the hyperbola approaches two asymptotes (red dash lines in Fig. 9) and the shape of the hyperbola is bounded by its asymptotes.

1. A hyperbola is a set of points (P) that have a constant absolute difference between $|PF_1|$ and $|PF_2|$, where F_1 and F_2 are two fixed points (the foci).

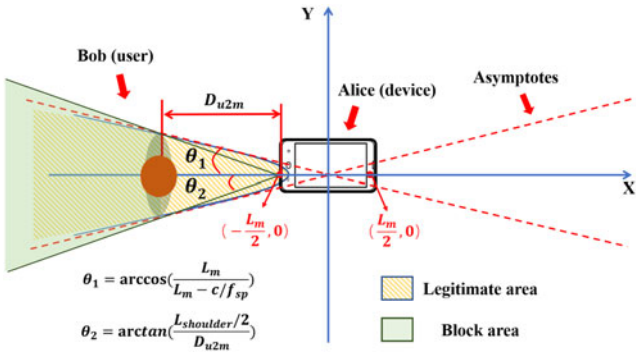


Fig. 9. The boundary of the legitimate area can be approximated to the red dashed asymptotes of the hyperbola. A user (Bob) can essentially block the legitimate area behind him if $\theta_2 \geq \theta_1$.

As shown in Fig. 9, we can approximately consider the legitimate area to be within the two asymptotes (the shaded area), and θ_1 is the tolerance of the speaker-to-microphone angle. With basic geometric knowledge, we have:

$$\theta_1 = arccos\left(1 - \frac{c}{f_{sp} L_m}\right). \tag{16}$$

where c is the speed of sound (approximately $340m/s$), L_m is a constant related to the device size and f_{sp} is higher than 44.1kHz for most of devices. Taking a mobile device with $L_m = 14cm$ as an example (L_m is $13cm$ for iPhone 6s and $15cm$ for iPhone 6s Plus), θ_1 is $19.1°$. Therefore, the tolerance range for the speaker-to-microphone angle is $[-\theta_1, \theta_1]$, which is $[-19.1°, 19.1°]$ in this case. We believe this range is big enough for users when we require them to put the speaker and the microphone on a straight line.

6 Security Analysis

In this section, we analyze the security of NAuth from the following perspectives.

- Can attackers bypass the location-based validation model and impersonate a legitimate user?
- Can attackers deceive the verification scheme and launch replay, man-in-the-middle attacks or injection attacks?

6.1 Security of the Location-Based Validation Model

If the attacker is outside the legitimate area, she cannot satisfy the requirement imposed by Eq. (12). Even if she exploits multiple speakers, it is extremely difficult to beamform a sound at one microphone without getting it received by another one close by, e.g., $15cm$ away for a smartphone.

Therefore, a more threatening scenario is when the attacker is inside the legitimate area. Since attackers cannot locate between the legitimate users due to the risk of getting visually exposed, we only consider the situation that attackers are behind a legitimate user. In this scenario, the user interaction is considered. Since users are required to manually align the devices, naturally we can assume that they sit or stand behind their devices. Users can block all line-of-sight transmissions of acoustic signals behind them because very few acoustic energies can penetrate through the human body. We highlight the block area in Fig. 9, and the boundary of the block area is the line that connects the bottom microphone and the user's shoulder. The angle between the boundary and the x-axis is:

$$\theta_2 = arctan\left(\frac{0.5 * L_{shoulder}}{D_{u2m}}\right), \tag{17}$$

where $L_{shoulder}$ is the width of the user's shoulder and D_{u2m} is the distance between the user and the bottom microphone. When $\theta_2 \geq \theta_1$, the user can block all attackers behind her even if the attacker is located in the legitimate area. With Eqs. (16) and (17), we have:

$$D_{u2m} \leq \frac{0.5 * L_{shoulder}}{tan\theta_1}. \tag{18}$$

Given $\theta_1 = 19.1°$ and consider a shoulder width of $36cm$ (an average for adult females), we derive $D_{u2m} \leq 51.98cm$, which is typical for face-to-face scenarios. Thus, the location-based validation model is efficient to detect attackers outside and even inside the legitimate area behind users.

6.2 Replay Attack

Typically, the replay attack is a security issue that a third party deceives a legitimate device by sending a captured authentication signal. In NAuth, we assume the attacker Eve has full knowledge of the authentication mechanism. Therefore, the modulation and the generation methods of the authentication signal is public. To launch the replay attack, Eve may attempt to: 1) send an authentication signal directly or 2) replay an elaborately designed authentication signal with previously extracted ANPs. We discuss in the following why a replay attack is not feasible.

For the first case, Eve sends a modulated authentication signal with $f_b = 1$kHz and $f_c = 20$kHz, for example. As discussed in Section 2, the ANP is device-dependent and location-dependent, therefore Eve cannot generate the same ANP as Alice with a different speaker while locating at different locations.

For the second case, Eve first needs to obtain Alice's ANP by listening to the public acoustic channel and then elaborately compensates nonlinear distortion by adjusting the harmonic amplitudes. Considering the acoustic signal attenuates exponentially during propagation, Eve cannot extract the same ANPs as the one from a legitimate user. Even though Eve may compromise Bob and eavesdrop Alice's ANP, she still cannot be authenticated by Bob with a crafted authentication signal. The reason is that Bob's microphone reveals unpredictable nonlinear distortion, and the ANP eventually extracted by Bob still does not match Alice's.

In conclusion, due to the location and device-dependent characteristics of the acoustic nonlinear distortion in NAuth, Eve cannot successfully launch a replay attack.

6.3 Man-in-the-Middle Attack

We consider the scenario that an attacker attempts to participate in the pairing of two legitimate devices and launches man-in-the-middle attacks by impersonating both Alice and Bob at the same time. The goal of the attacker is trying to make both Alice and Bob believe they are paired with the

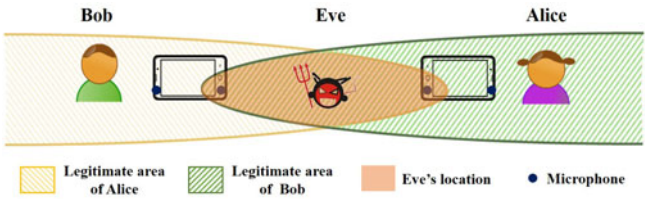


Fig. 10. The only location to launch a man-in-the-middle attack is between two users.

legitimate device. To do this, the attacker should simultaneously pair with both legitimate devices successfully.

According to the location-based validation model, the legitimate area of Alice and Bob is covered with a left-branch-hyperbola in yellow and a right-branch-hyperbola in green as shown in Fig. 10. If the attacker Eve attempts to achieve the above goal, she must be located in the overlapped legitimate area for both users, i.e., between them (the orange area in Fig. 10). Thus, a man-in-the-middle attack is not feasible as Eve will be noticed immediately once she locates between the two legitimate devices.

6.4 Injection Attack

The injection attack includes two cases: 1) the attacker keeps injecting fraud information to Bob, and 2) the attacker attempts to avoid overlapping with legitimate authentication signals and inject fraud information.

For the first case, the authentication process will be stopped by Bob unilaterally since no legitimate authentication signal is received and the attacker can't inject fraud keys without sending any legitimate authentication signal. We take this case as a deny-of-service attack, which is out of the scope of NAuth.

For the second case, we consider the probability of the attacker to avoid all authentication signals and inject information extremely low. For example, we consider the case that Alice transmits a 1024 bit information to Bob with Dolphin [21] as the data transmission scheme. Alice should encode the information into 18 symbols with at least 9 additional authentication signals based on the design of NAuth. The probability of Eve to inject signals successfully while avoiding all authentication signals is:

$$Pr = C_{27}^{9} = 2.1 * 10^{-7}. \tag{19}$$

The probability is extremely low and thus we conclude the attacker has little chance to avoid all authentication signals and bypasses the legitimate verification.

7 Evaluation

In this section, we conduct extensive experiments to evaluate the efficiency of NAuth. For the nonlinearity-based verification scheme, we emulate a mobile payment scenario under different experiment settings. Besides, we evaluate the location-based validation model with two smartphones. The experiment setups and tested devices are summarized in Table 3.

7.1 Efficiency of the Verification Scheme

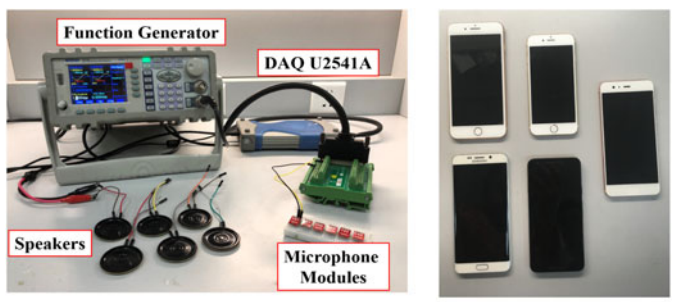
We evaluate the verification scheme with a mobile payment scenario shown in Fig. 1, where a smartphone (the receiver) needs to authenticate a cashing machine (the sender). We envision that NAuth can be applied to various types of acoustic D2D communications, including the trending ultrasonic communication. Therefore, we emulate 4 ultrasound-capable cashing machines with 4 ultrasonic speakers and choose the authentication signal to be an AM signal with $f_c = 20\text{kHz}$ and $f_b = 1\text{kHz}$. The signals are received on 4 smartphones (iPhone 8P, iPhone 6S, Galaxy S6 Edge and Nexus 5X) shown in Fig. 11b. For each sender-receiver pair (SMS) in each setting, we collect 300 sets of ANPs and compare the euclidean distances. We consider four settings that may affect the performance—different receivers, senders, distances and noise levels, which correspond to four assumptions: 1) different customers at the same store, 2) a customer at different stores, 3) a customer pays multiple times at the same store, and 4) payments are performed under different background noises. We investigate the four settings separately in the following.

TABLE 3
Summary of Experiment Setups

	Model	L_m (mm)	f_{sp} (kHz)	$2\theta_1$
Devices	Apple iPhone 8 Plus	148	48	35.6°
	Apple iPhone 6S	130	48	38°
	Samsung Galaxy S6 Edge	149	48	35.5°
	Google Nexus 5X	140	48	36.6°
	Huawei P10 Plus	145	44.1	37.5°
Scenario	(1) Quiet meeting room; (2) restaurant; and (3) the street.			
Distances	1cm, 3cm, 5cm and 8cm.			

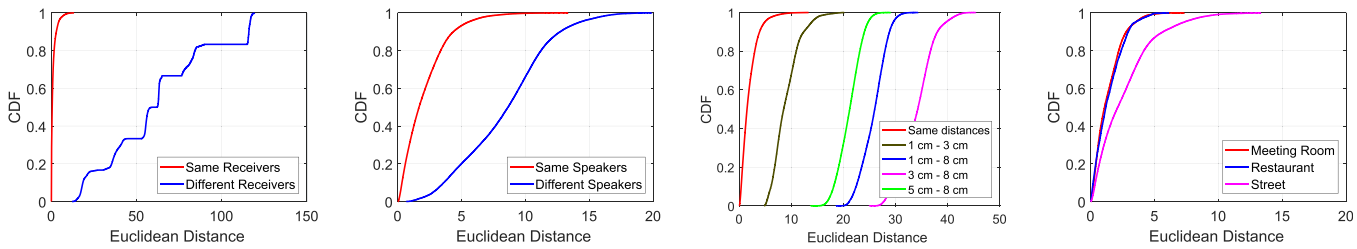
7.1.1 The Impact of Receivers

We send authentication signals from the same speaker and utilize four smartphones as receivers respectively at a distance of $3cm$. We calculate the euclidean distances of ANPs from the same SMSes ($d(i,i)$, $i \in [1,4]$) and different SMSes ($d(i,j)$, $i,j \in [1,4]$ & $i \neq j$). We show the CDF of euclidean distances of both cases in Fig. 12a. The euclidean distances between the same SMSes are significantly smaller than between different SMSes.



(a) Equipment and stand-alone modules. (b) Tested smartphones.

Fig. 11. Experiment settings for (a) feasibility validation and (b) evaluation.



(a) Varying receivers. (b) Varying senders. (c) Varying distances. (d) Varying noise levels.

Fig. 12. The CDF of euclidean distances (a) between the same and different receivers; (b) between the same and different senders; (c) between the same and different distances; (d) between the same SMSes under different noise levels.

7.1.2 The Impact of Senders

Similarly, we send authentication signals from four speakers of the same model and receive signals with an iPhone 6S. Results in Fig. 12b show that over 95 percent of euclidean distances from the same SMSes are smaller than 5 while it is 14 for different SMSes. We find the euclidean distances of different senders reveal higher similarity in ANPs when compared with the different receivers' case. The reason is that the receivers in Fig. 12a are of different models, while we use four speakers of the same model in the case of Fig. 12b. We expect the euclidean distances between different senders to be bigger if different models of ultrasonic speakers are used. On the other hand, we test the ANPs of different speakers at the same location in Fig. 12b, which is not possible to exist in real scenarios (attackers cannot locate at the same point as the legitimate user). Therefore, the euclidean distances could be bigger even if the attackers use the speaker of the same model in real cases.

7.1.3 The Impact of Distances

We send signals from the same speaker to four smartphones at four locations, and the distance between the four locations to the sender are 1cm, 3cm, 5cm and 8cm respectively. We show the ANP euclidean distances of the same SMSes at the same and different distances in Fig. 12c. The results reveal that distance can significantly affect the ANPs. For ANPs at the same distance, the euclidean distances are smaller than 10, while they increase significantly even if the device moves slightly. The experiment results also indicate that a bigger movement of devices does not necessarily represent a higher euclidean distance. Thus, we suggest that users do not move the device during the `NAuth` authentication.

7.1.4 The Impact of Noise Levels

We conduct experiments on the same SMS at three places including a quiet meeting room, a restaurant, and the street. The average noise levels at the three places are 38.8, 58.2 and $73.7dB$ SPL. As shown in Fig. 12d, the ANP euclidean distances on the street are higher than in the meeting room and restaurant, therefore the ambient noise can interfere with the ANPs. Nevertheless, the ANP euclidean distances are no more than 10 for all scenarios, which indicates that the efficiency will not be affected.

7.1.5 The Selection of σ

Experiment results in Fig. 12 demonstrate that the euclidean distances of ANPs from the same SMSes in stable test scenarios are much smaller than other cases, no matter the change of the receiver, the sender, or the location. We select the value of σ based on the experiment results, and the selection of σ is a trade-off between the user experience and the system security.

We evaluate the user experience by calculating the overall true positive rate (TPR) of the verification scheme, and we quantify the security of the verification scheme by calculating the false positive rate (FPR) of the verification scheme. Hereby we define $TPR = \frac{TP}{TP+FN}$ and $FPR = \frac{FP}{TN+FP}$. We plot the ROC curve of the verification scheme with the FPR as the x-axis and TPR as the y-axis in Fig. 13. The red circles on the ROC curve correspond to the cases where the σ goes from 1 to 10. A small σ may bring poor user experience (low TPR) because legitimate users may be rejected by the verification scheme, while a large σ increase the FPR of the system. By considering both the user experience and the security of the verification scheme, we introduce the EER (equal error rate) to select an appropriate σ in `NAuth`. EER corresponds to a trade-off σ at which the two indicators are equal, i.e., EER(σ) = 1-TPR(σ) = FPR (σ). The intersection of the blue line and the ROC curve in Fig. 13 is the EER of the verification scheme. As shown in Fig. 13, we select the σ as 5 in `NAuth`.

7.2 Efficiency of the Validation Model

We record acoustic signals with a Huawei P10 Plus and measure the TDOA at the top and bottom microphones separated by $145mm$. According to Eq. (12), the TDOA for legitimate devices should be higher than $0.381ms$, which takes approximately 17 sample points at 44.1kHz.

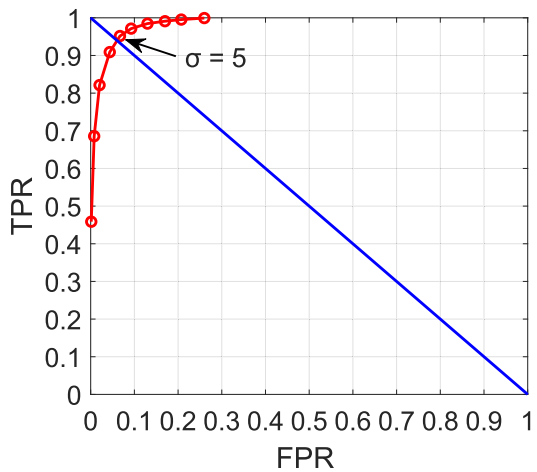


Fig. 13. The ROC curve of the verification scheme (σ from 1 to 10).

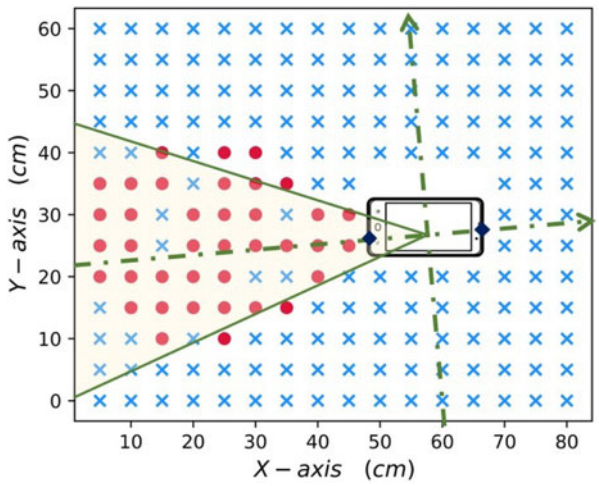


Fig. 14. Results of the validation model running on a Huawei P10 Plus receiver (with two marked microphones near coordinates (50,25) and (65,30)) tested with the speaker of an iPhone 6S at 196 locations around it. The passed and rejected locations are marked with red dots and blue crosses.

7.2.1 Legitimacy Validation

We send a $500Hz$ tone with an iPhone 6S at 196 locations around the receiver as illustrated in Fig. 14. We mark a red dot on the locations that pass the validation model and mark a cross for those that fail. We mark the bottom and top microphones of the receiver with diamonds and plot the theoretical legitimate area ($37.5°$) with yellow shadow. Experiment results show that the passed locations concentrate in a small area, which approximates the theoretical legitimate area. With a second experiment, most false negatives (crosses in the legitimate area) can be eliminated.

To investigate the overall efficiency when multiple attempts are possible, we measure the TDOA for 200 times at 196 locations in Fig. 14 and calculate the pass rates. We divide the area into three categories, the legitimate area, the boundary area, and the rejection area. The experiment results show that the average pass rates of the legitimate, the boundary area, and the rejection area are 87.38, 39.72, and 2.38 percent individually. We select 9 locations from the legitimate area, boundary line, and rejection area as representatives and report their pass rates in Fig. 15. The results demonstrate that devices in the legitimate area can pass the validation model easily while it is hard on the boundary and almost impossible in the rejection area.

7.2.2 The Selection of η

We also conduct experiments to assess the value of η in Eq. (15). We separate all 196 locations in Fig. 14 into three categories: the legitimate area, the boundary line, and the rejection area, and we evaluate the number of passes of each category individually. We count the number of passes within 10 TDOA measurements and repeat the experiment 200 times at each location. We show the CDF of the number of passes in Fig. 16. A small number of passes represents a lower probability of passing the validation model.

As shown in Fig. 16, over 90 percent of experiments conducted in the legitimate area can pass the validation model for at least 7 times within 10 measurements, while the number of passes is lower than 4 for all experiments conducted

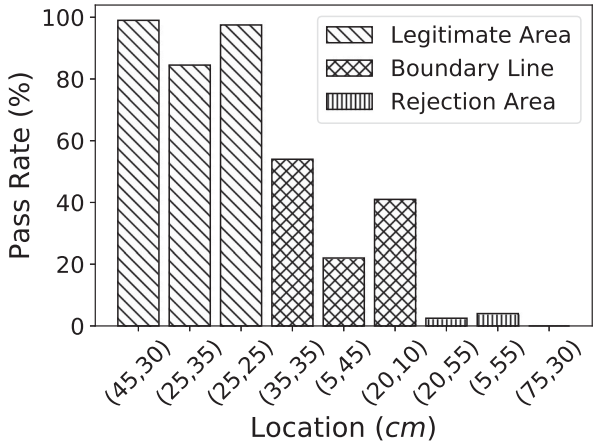


Fig. 15. Rates of passing the validation model at 9 locations in the legitimate area, boundary line and rejection area.

in the rejection area. We also notice the number of passes is distributed uniformly in the boundary line. To enhance the reliability of the validation model, we select $\eta = 70\%$. Though selecting $\eta = 70\%$ may lead to a few legitimate users failing, we consider they can easily tackle the problem after a slight adjustment of the device.

7.2.3 User Intervention

We evaluate the efficiency when a user blocks the sound from an attacker behind her but in the legitimate area. A user with a $36cm$ shoulder width sits between the receiver (a Huawei P10 Plus) and an attacker (an iPhone 6S). We place the attacker at 5 locations (listed in Table 4) in the legitimate area and calculate the pass rates with and without the user as an obstacle. Results show that with the user as an obstacle, the pass rates drop significantly, therefore the user intervention in `NAuth` is sufficient to prevent attackers in the legitimate area from passing the validation model.

8 Discussion

8.1 ANP Entropy

We need to discuss whether the ANP has enough entropy to be used as a device identity. Taking the MEMS microphone

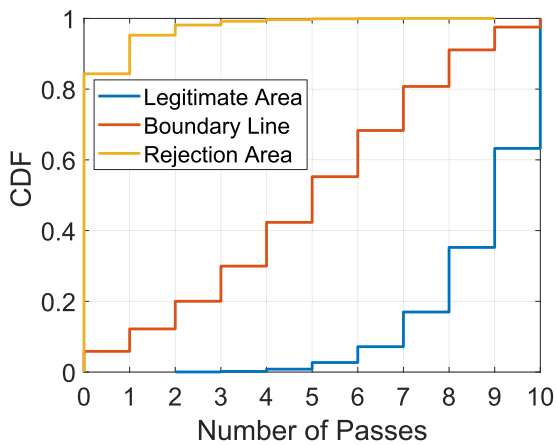


Fig. 16. The CDF of the number of passes within 10 attempts. A small number of passes represents a lower probability of passing the validation model.

TABLE 4
Pass Rates in the Legitimate Area With and Without a User

Locations (cm)	Pass Rate (%) w/o User	Pass Rate (%) w/ User
(15,20)	91.5	8.5
(10,30)	89	0.5
(10,25)	85	0
(5,30)	88	0.5
(5,25)	96	0

module ADMP401 [20] as an example, we analyze the amplitudes of nonlinearity distortion from 25 modules and find the range of each element in the ANP ($ANP(i)$) different. Assuming $ANP(i)$ is normally distributed, we summarize the entropy of $ANP(i)$ for ADMP401 in Table 5. Based on the analysis above, the ANP for ADMP401 has an estimated entropy of 55 bits. Compared with the same type of modules, different types of modules reveal significant differences, which may expand the ANP's entropy. Therefore, we consider the entropy of the ANP is sufficient enough for authentication.

8.2 ANP Stability and Uniqueness

We further discuss the stability and uniqueness of the ANP among more devices and modules to demonstrate the feasibility of utilizing ANP for authentication. We evaluate the ANP among 9 popular commercial smartphones (6 models), 28 microphone modules (3 models) and 30 speaker modules (3 models), which include both ANPs of different models and the same model. The number and the model of the devices and modules are listed in Tab 6. As for the device set and the speaker set, we receive signal with an ADMP401 microphone, and we send signal with an M38 speaker when evaluate the microphone set.

We extract 300 ANPs on each device/module at the same location (8cm) with the same amplitude ($10dBm$). We compare ANPs in each category (device, microphone modules and speaker modules), and the prediction results are shown in the head maps (aka. the confusion matrix) in Fig. 17. A darker square means a higher probability to be predicted as the corresponding device/module. Evidently, the diagonal cells are the highest ones for each device/module, which implies that the ANPs of device/module i are indeed predicted as device/module i. We also notice that a few light gray cells appear outside the diagonal cells, which indicates such cases are rare. We also mark the ANPs from the same model with dotted boxes, and the results demonstrate that ANPs from the same model of devices/modules can also be easily distinguished. Therefore, we can conclude that the ANPs of devices/modules are stable and unique for device authentication.

TABLE 5
Entropy of Each Element of ANP for ADMP401[1]

Element	$ANP(1)$	$ANP(2)$	$ANP(3)$	$ANP(4)$
Entropy (bits)	13.7	8.9	8.1	6.4
Element	$ANP(5)$	$ANP(6)$	$ANP(7)$	$ANP(8)$
Entropy (bits)	7.4	6.2	5.6	5.8

[1]*The resolution of the ADC is 16-bit.*

TABLE 6
The Number and the Model of Devices and Modules

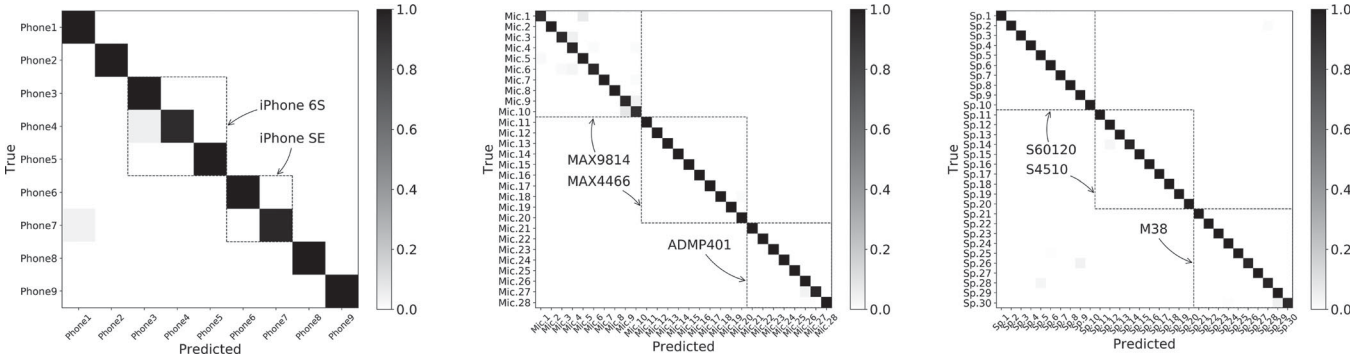
Type	Model	Number
Devices	iPhone 8 Plus	1
	iPhone X	1
	iPhone 6s	3
	iPhone SE	2
	Huawei P10 Plus	1
	Huawei Honor V8	1
Microphone Modules	MAX4466	10
	MAX9814	10
	ADMP401	8
Speaker Modules	S60120	10
	S4510	10
	M38	10

8.3 Comparison With Existing Authentication Methods

Traditional device authentication methods like the digital certificate rely on a trust management center, and not suitable for IoT devices when they do not have Internet access (mobile payment can be launched in off-line environment). The barcode or QR code scanning are convenient in mobile payment scenarios, however, such methods face security threats like the man-in-the-middle attack. For the devices that have no prior knowledge, the major information they can obtain in authentication is the physical proximity, and many authentication methods are proposed basing on such characteristic.

We compare `NAuth` with existing device authentication methods, which include acoustic-based methods (S2M [23] and GeneWave [15]) and non-acoustic-based methods (Bluetooth [24], AccelPrint [25], KEEP [8], TDS [10] and ProxiMate[26]). We summarize the performance, implementation conditions, and user experience of the above methods in Table 7.

Traditional authentication methods like Bluetooth generally need users to input a pin code or pre-negotiate a session key during the device pairing and authentication, which requires more user intervention (we define the user intervention as medium in Table 7). Other methods only require user to put device at a fixed position, and we define such level of user intervention as mild in Table 7. AccelPrint extracts hardware fingerprint from built-in accelerometers, which needs to build a fingerprint database and train a classifier in advance and may not be suited for real-time authentication without pre-knowledge. Compared with KEEP, TDS, and ProxiMate, `NAuth` doesn't need extra hardware and can be easily promoted among mobile devices. Besides, `NAuth` shows higher efficiency when compared with S2M. The values of the authentication time are directly obtained from papers or calculated basing on the data provided in papers. GeneWave reveals similar performance when compared with `NAuth`. However, the legitimate area in GeneWave is a $3.57\ m$ radius area, which means users need to identify attackers within a $40m^2$ space by themselves, and `NAuth` only requires users to guarantee that no attacker stays between two authentication entities. It is worth to mention that none of the methods (except Bluetooth) support device movement during the authentication.



(a) Devices. (b) Microphone Modules. (c) Speaker Modules.

Fig. 17. The confusion matrix of devices, microphone modules and speaker modules. A darker square means a higher probability to be predicted as the corresponding device/module.

8.4 Ambient Noisec

Random ambient noises can interfere with the ANP and may affect its consistency over time. However, the application scenarios for NAuth can generally be finished in 1 second, e.g., mobile payment and key establishment, therefore the impact of the ambient noises can be limited.

8.5 Time Overhead

NAuth requires users to send authentication and declaration signals and can impose time overhead on the D2D communication. Considering a frequency resolution of $100Hz$ in performing the Fourier Transform (as in our experiment), a $10ms$ sample is required. If we average the ANPs of 5 authentication signals for initialization and use 5 declaration signals for verification, the total time overhead is $100ms$, which is acceptable for most application scenarios.

8.6 Device Requirement

The authentication-initiating device should have two microphones in order to measure the TDOA. NAuth is inapplicable to devices with only one microphone. Besides, microphones locate at different positions in different types of devices, which brings challenges for users to place the devices properly. Fortunately, the positions of microphones are fixed for a certain device and we can display the connecting line of the top and the bottom microphones on the screen when launching NAuth. In this way, the user can place the device accurately by aligning the connecting line with the speaker without knowing the specific locations of microphones.

8.7 User Requirement

NAuth requires users to put the devices into the legitimate area and hold them still during the authentication process, which sometimes might be tricky. The authentication process could be fast that users only need to hold the device for less than 1 second. To be specific, launching NAuth and exchanging a 128-bit public key with Dolphin [21] takes $633ms$ (less than 1 second), which is acceptable for most users. Compared with methods that require users to input a given code, NAuth achieves the authentication security with less user intervention or less time overhead. On the other hand, for the users that are unable to hold the device

TABLE 7
The Comparison With Existing Authentication Methods

Method	Authentication Basis	User Intervention	Still hold (Y/N)	Hardware Requirement	Extra Hardware (Y/N)	Security				Pre-training (Y/N)	Authentication Time[§]
						Replay Attack	MITM Attack	Spoofing Attack	DoS Attack		
Bluetooth [23]	PIN code	medium	Y	bluetooth chip	N	√	×[†]	√	×	N	N/A
S2M [24]	audible sound,	mild	N	speaker, microphone	N	√	√	√	×	Y	>2s
GeneWave [15]	inaudible sound,	mild	N	speaker, microphone	N	√	√	√	×	N	<2s
AccelPrint [25]	hardware fingerprint	mild	N	accelerometer	N	√	√	√	×	Y	N/A
KEEP [8]	radio signal, CSI	mild	N	off-the-shelf 802.11n devices	Y*	√	√	√	×	N	N/A
TDS [10]	radio signal, CSI	mild	N	off-the-shelf 802.11n devices	Y*	√	√	√	×	N	>2.5s
ProxiMate [26]	radio signal, RSS	mild	N	USRP[‡]	Y	√	√	√	×	N	>60s
NAuth	**inaudible sound**	**mild**	**N**	**speaker, microphone**	**N**	√	√	√	×	**N**	**<1.5s**

[§]*Including the device authentication and a key agreement session (256-bit).*
[†]*The JUSTWORK mode cannot prevent the MIMT attack.*
[*] *Intel 5300 network card.*
[‡]*USRP (Universal Software Radio Peripheral).*

stably even for 1 second (dynamic scenarios or physical problems), we may suggest users put the mobile device on a fixed plane for better user experience.

8.8 ANP Extension

All microphones embedded in devices can be utilized for ANP extraction, which shows potentiality to increase the entropy of the ANP. For instance, we can measure the ANPs of both the top and the bottom microphones when calculating the TDOA of the received signal. The newly extracted ANP set has a double entropy, which provides better robustness of the NAuth with limited computing overhead. Smart devices now are generally embedded with even more than 2 microphones, e.g., Apple designs 3 microphones on iPhone since the iPhone 5 series. Therefore, we take it as an extension of the ANP extraction and allow developers to decide how many microphones should be utilized for ANP extraction.

8.9 Limitations

With the elaborate design, NAuth achieves high performance most of the time. However, NAuth has the following limitations.

First, NAuth shows a deficiency in an extremely dynamic environment. Since the ANP is location-sensitive, NAuth requires a relatively static environment. In an extremely dynamic environment, the device could be rocked wildly, which impacts the extraction and the authentication of the ANP. Therefore, we do not recommend users launch NAuth when staying in a dynamic environment.

Second, the ANP is location-related. The location sensitivity brings benefits in the design of NAuth, e.g., it prevents attackers from impersonating a legitimate user with the same model of device at different locations. However, this characteristic also restricts the working conditions of NAuth. Users should run NAuth in a fixed position, moreover, the extracted ANP cannot be utilized repeatedly if the user runs NAuth again at another position.

9 Related Work

Extensive research has been proposed for establishing secure D2D communications mainly from three perspectives: proximity, hardware fingerprint, and covert channel.

The proximity-based approaches extract symmetric keys from properties of the wireless channels such as RSS (received signal strength) [5], [6], [7], [26], [27] and CSI (channel state information) [8], [9], [10], [11]. The RSS-based key agreement algorithms typically include three steps: quantization,reconciliation and privacy amplification[7]. In the quantization step, the receiver samples the signal with a predefined sampling scheme which converts the physical signal into a digital signal. The reconciliation step calibrates the mismatch bits which caused by the ambient noise while privacy amplification protects secret keys from being eavesdropped by attackers. Though RSS-based mechanisms have been intensively investigated, they still suffer from restricted efficiency because RSS can only provide coarse-grained information. The key generation rates are relatively slow in most of the RSS-based key agreement algorithms, for example, the bit generation rate is less than $5bps$ in ProxiMate[26]. Compared with the RSS-based mechanisms, CSI-based ones are more efficient because they can derive fine-grained physical layer information, e.g., the channel response from multiple subcarriers of Orthogonal Frequency-Division Multiplexing (OFDM). However, CSI-based mechanisms are hard to promote because such methods rely on dedicated hardware (Intel 5300 Wi-Fi card) and cannot be widely implemented on mobile devices. Another challenge is that CSI is so sensitive to location and environment that the bit error rate is usually considerable, which may lead to key agreement inefficiency in generalized application scenarios.

A number of studies have shown that mobile devices can be fingerprinted with inherent hardware modules including clocks[28], accelerometers [25], microphones [23], [29], [30] and speakers [30], [31]. Though the same type of electronic components shares the same design and material, the manufacturing process introduces uncontrollable production error, which leads to subtle response errors with the same stimulation. Utilizing the naturally unclonable and unique characteristics of the embedded electronic components, the hardware-based fingerprinting schemes can authenticate devices efficiently. Although these hardware fingerprints are inimitable, undeniable and stable, the authentication requires prior extraction of features and trained classifiers, thus they are inapplicable to D2D communication when no secret is shared in advance.

Roeschlin *et al.* [32] and Chang *et al.* [33] exploit secure body channels for key establishment. Roeschlin's work [32] assumes the devices touched by the same person at the same can perform device pairing. They require the user to touch two devices' electrodes, therefore, the human body can play as a transmission medium for intra-body communication. Such intra-body-channel cannot be impersonated by attackers and it is read-only for attackers. This body channel is used as part of a pairing protocol which allows the devices to agree on a mutual secret and, at the same time, extract physical features to verify that they are being held by the same person. However, extra hardware like electrodes and on-body sensors are required while NAuth only relies on built-in microphones and speakers.

Besides, Xie *et al.* [15] proposed GeneWave, a key establishment mechanism which is based on the acoustic channel response of devices. Xie's work is the most closely related to ours. Their method finds the frequency response for a certain pair of speaker and microphone reveals unique patterns, and they extracted such patterns as ACR which could be an alternative of device identity. GeneWave also designs a key agreement scheme and implements it on mobile devices. We investigate ACR and ANP in speaker-microphone systems and find ANP is more sensitive to the location. On the other hand, GeneWave assumes the attackers to be outside a certain range and may not suffice to detect hidden attackers nearby in face-to-face D2D communications.

10 Conclusion

We propose NAuth, a nonlinearity-enhanced, location-sensitive authentication mechanism for secure face-to-face D2D communication. NAuth consists of two main components: a nonlinearity-based verification scheme and a location-based validation model. We extract acoustic nonlinear patterns

(ANP) to verify device consistency in the verification scheme and measure the TDOA at two microphones to guarantee device legitimacy in the validation model. Theoretical analysis and experiment results demonstrate NAuth can authenticate devices efficiently in the presence of nearby attackers.

Acknowledgments

This Work was supported by the National Natural Science Foundation of China (No.61941120, 61702451, 62002183, and 61925109), the Natural Science Foundation of Zhejiang Province (No.LGG19F020020 and No.LQ20F020012).

References

[1] Eric Enge, "Mobile vs desktop usage in 2018: Mobile takes the lead," 2018. [Online]. Available: https://www.stonetemple.com/mobile-vs-desktop-usage-study/
[2] J. Roberts, "Different types of mobile payments explained," 2018. [Online]. Available: https://www.mobiletransaction.org/different-types-of-mobile-payments/
[3] Y. Jiang, Z. Li, and J. Wang, "Ptrack: Enhancing the applicability of pedestrian tracking with wearables," *IEEE Trans. Mobile Comput.*, vol. 18, no. 2, pp. 431–443, Feb. 2019.
[4] Y. Liu and Z. Li, "aleak: Privacy leakage through context - free wearable side-channel," in *Proc. IEEE Conf. Comput. Commun.*, 2018, pp. 1232–1240.
[5] Y. Luo, L. Pu, Z. Peng, and Z. Shi, "RSS-based secret key generation in underwater acoustic networks: Advantages, challenges, and performance improvements," *IEEE Commun. Magazine*, vol. 54, no. 2, pp. 32–38, Feb. 2016.
[6] H. Liu, J. Yang, Y. Wang, Y. J. Chen, and C. E. Koksal, "Group secret key generation via received signal strength: Protocols, achievable rates, and implementation," *IEEE Trans. Mobile Comput.*, vol. 13, no. 12, pp. 2820–2835, Dec. 2014.
[7] T. Wang, Y. Liu, and A. V. Vasilakos, "Survey on channel reciprocity based key establishment techniques for wireless systems," *Wireless Netw.*, vol. 21, no. 6, pp. 1835–1846, 2015.
[8] W. Xi *et al.*, "KEEP: Fast secret key extraction protocol for D2D communication," in *Proc. IEEE 22nd Int. Symp. Quality Service*, 2014, pp. 350–359.
[9] Y. Liu, S. C. Draper, and A. M. Sayeed, "Exploiting channel diversity in secret key generation from multipath fading randomness," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 5, pp. 1484–1497, Oct. 2012.
[10] W. Xi *et al.*, "Instant and robust authentication and key agreement among mobile devices," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2016, pp. 616–627.
[11] H. Liu, Y. Wang, J. Yang, and Y. Chen, "Fast and practical secret key extraction by exploiting channel response," in *Proc. IEEE INFOCOM*, 2013, pp. 3048–3056.
[12] F. Qiu, Z. He, L. Kong, and F. Wu, "MAGIK: An efficient key extraction mechanism based on dynamic geomagnetic field," in *Proc. IEEE Conf. Comput. Commun.*, 2017, pp. 1–9.
[13] D. Bichler, G. Stromberg, M. Huemer, and M. Löw, "Key generation based on acceleration data of shaking processes," in *Proc. Int. Conf. Ubiquitous Comput.*, 2007, pp. 304–317.
[14] X. Liang, T. Yun, R. A. Peterson, and D. Kotz, "Lighttouch: Securely connecting wearables to ambient displays with user intent," in *Proc. IEEE Conf. Comput. Commun.*, 2017, pp. 1–9.
[15] P. Xie, J. Feng, Z. Cao, and J. Wang, "GeneWave: Fast authentication and key agreement on commodity mobile devices," *IEEE/ACM Trans. Netw.*, vol. 26, no. 4, pp. 1688–1700, Aug. 2018.
[16] Apple, "Test the microphones on your device," 2018. [Online]. Available: https://support.apple.com/en-us/HT203792
[17] G. Zhang, C. Yan, X. Ji, T. Zhang, T. Zhang, and W. Xu, "Dolphinattack: Inaudible voice commands," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2017, pp. 103–117.
[18] N. Roy, H. Hassanieh, and R. R. Choudhury, "Backdoor: Making microphones hear inaudible sounds," in *Proc. 15th Annu. Int. Conf. Mobile Syst. Appl. Services*, 2017, pp. 2–14.
[19] K. Koli and K. A. Halonen, *CMOS Current Amplifiers: Speed Versus Nonlinearity*. Berlin, Germany: Springer, 2002.
[20] Analog Devices, "Admp401: Omnidirectional microphone with bottom port and analog output," *AnInd. ICs Solutions Bull.*, vol.10, no. 8, pp. 1–12, 2013.
[21] Q. Wang, K. Ren, M. Zhou, T. Lei, D. Koutsonikolas, and L. Su, "Messages behind the sound: Real-time hidden acoustic signal capture with smartphones," in *Proc. 22nd Annu. Int. Conf. Mobile Comput. Netw.*, 2016, pp. 29–41.
[22] W. Diffie and M. Hellman, "New directions in cryptography," *IEEE Trans. Inf. Theory*, vol. 22, no. 6, pp. 644–654, Nov. 1976.
[23] D. Chen *et al.*, "S2M: A lightweight acoustic fingerprints-based wireless device authentication protocol," *IEEE Internet Things J.*, vol. 4, no. 1, pp. 88–100, Feb. 2017.
[24] J. Haartsen, "The bluetooth radio system," *IEEE Pers. Commun.*, vol. 7, no. 1, pp. 28–36, Feb. 2000.
[25] S. Dey, N. Roy, W. Xu, R. R. Choudhury, and S. Nelakuditi, "Accelprint: Imperfections of accelerometers make smartphones trackable," in *Proc. Netw. Distrib. Syst. Secur. Symp.*, 2014, pp. 1–16.
[26] S. Mathur, R. D. Miller, A. Varshavsky, W. Trappe, and N. B. Mandayam, "Proximate: proximity-based secure pairing using ambient wireless signals," in *Proc. 9th Int. Conf. Mobile Syst. Appl. Services*, 2011, pp. 211–224.
[27] A. Varshavsky, A. Scannell, A. LaMarca, and E. de Lara, "Amigo: Proximity-based authentication of mobile devices," in *Proc. Int. Conf. Ubiquitous Comput.*, 2007, pp. 253–270.
[28] S. Jana and S. K. Kasera, "On fast and accurate detection of unauthorized wireless access points using clock skews," *IEEE Trans. Mobile Comput.*, vol. 9, no. 3, pp. 449–462, Mar. 2010.
[29] Z. Zhou, W. Diao, X. Liu, and K. Zhang, "Acoustic fingerprinting revisited: Generate stable device ID stealthily with inaudible sound," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2014, pp. 429–440.
[30] X. Zhou, X. Ji, C. Yan, J. Deng, and W. Xu, "Nauth: Secure face-to-face device authentication via nonlinearity," in *Proc. IEEE Conf. Comput. Commun.*, 2019, pp. 2080–2088.
[31] A. Das, N. Borisov, and M. Caesar, "Do you hear what I hear?: Fingerprinting smart devices through embedded acoustic components," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2014, pp. 441–452.
[32] M. Roeschlin, I. Martinovic, and K. B. Rasmussen, "Device pairing at the touch of an electrode," in *Proc. Netw. Distrib. Syst. Secur. Symp.*, 2018, pp. 1–15.
[33] S. Chang, Y. Hu, H. Anderson, T. Fu, and E. Y. L. Huang, "Body area network security: Robust key establishment using human body channel," in *Proc. USENIX Workshop Health Secur. Privacy*, 2012, pp. 5–14.

Xiaoyu Ji (Member, IEEE) received the BS degree in electronic information and technology and instrumentation science from Zhejiang University, Hangzhou, China, in 2010, and the PhD degree from the Department of Computer Science, Hong Kong University of Science and Technology, in 2015. He is currently an associate professor at the Department of Electrical Engineering, Zhejiang University. From 2015 to 2016, he was a researcher with Huawei Future Networking Theory Lab in Hong Kong. His research interests include IoT security, wireless communication protocol design, especially with cross-layer techniques. He won the best paper awards of ACM CCS 2017, ACM ASIACCS 2018, and IEEE Trustcom 2014.

Xinyan Zhou (Member, IEEE) received the BS degree in electrical engineering from Zhejiang University, Hangzhou, China, in 2014, and the PhD degree from the College of Electrical Engineering, Zhejiang University, in 2019. She is currently an assistant professor at the Faculty of Electrical Engineering and Computer Science, Ningbo University. Her research interests include IoT security and wireless communication protocol design.

Chen Yan (Member, IEEE) received the BE degree in electrical engineering from Zhejiang University, Hangzhou, China, in 2015. He is currently working toward the PhD degree in the College of Electrical Engineering, Zhejiang University. His research interests include sensor security, wireless security, and IoT security. He received a Best Paper Award from the 2017 ACM CCS. He was acknowledged by Tesla Motors in the Security Researcher Hall of Fame in 2016.

Jiangyi Deng (Member, IEEE) received the bachelor's degree from the College of Electrical Engineering, Zhejiang University, Hangzhou, China, in 2019. He is currently working toward the doctoral degree from the College of Electrical Engineering, Zhejiang University.

Wenyuan Xu (Member, IEEE) received the BS degree in electrical engineering from Zhejiang University, in 1998, the MS degree in computer science and engineering from Zhejiang University, in 2001, and the PhD degree in electrical and computer engineering from Rutgers University, in 2007. She is currently a professor at the College of Electrical Engineering, Zhejiang University. Her research interests include wireless networking, network security, and IoT security. She received the NSF Career Award in 2009, a CCS Best Paper Award in 2017, and an ASIACCS Best Paper Award in 2018. She was granted tenure (an associated professor) with the Department of Computer Science and Engineering, University of South Carolina in the U.S. She has served on the technical program committees for several IEEE/ACM conferences on wireless networking and security, and is an associated editor of the TOSN and TIOT.