Nguyen, Ngu; Sigg, Stephan

Secure Context-based Pairing for Unprecedented Devices

Published in:
2018 IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops 2018

DOI:
10.1109/PERCOMW.2018.8480126

Published: 02/10/2018

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Nguyen, N., & Sigg, S. (2018). Secure Context-based Pairing for Unprecedented Devices. In *2018 IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops 2018* (pp. 119-124). Article 8480126 IEEE. https://doi.org/10.1109/PERCOMW.2018.8480126

Secure Context-based Pairing for Unprecedented Devices

Ngu Nguyen
Aalto University
Espoo, Finland
le.ngu.nguyen@aalto.fi

Stephan Sigg
Aalto University
Espoo, Finland
stephan.sigg@aalto.fi

***Abstract*—We introduce context-based pairing protocols that integrate into common distributed device encryption schemes for device management and access control. In particular, we suggest three pairing protocols that integrate implicit proximity-based device pairing to increase convenience and security. From these protocols, we implemented a secure device pairing approach conditioned on natural, unconstrained spoken interaction in a smart environment. In particular, our approach exploits speech recognition to identify devices to pair from free-form spoken interaction and restricts the pairing to the right device in proximity by generating secure keys from audio fingerprints of the same spoken interaction.**

I. Introduction

Recent decades have witnessed the proliferation of mobile systems that can be carried by users, such as smartphones, tablets, and laptops. These systems are equipped with sensing, computational, and networking capability. With increasing device penetration on-body and in the environment, there is a demand to establish secure spontaneous networks of unprecedented devices. For example, imagine a person arriving at a building, where she has not been before. Her smartphone can connect to the local wireless network. She would like to securely access a printer or connect to a projector she observes in the same room with her. Since the location is new to the user, she does not know specific device name or identity. One solution is to leverage environmental information in order to facilitate the secure connection [1, 2, 3, 4].

In this paper, we propose a secure key generation protocol which is based on the surrounding context of partner devices. Specifically, secure keys are generated from ambient audio recorded by embedded microphones. Our approach allows proximate devices to derive shared cryptographic keys without any key distribution center. As depicted in Figure 1, we envision the following use case. In order to connect to a device in proximity of a specific device class, a user will express the desired connection by speaking out loud the pairing intention. In this interaction, the user is not restricted to any format or convention but needs to mention the name of the intended device-class in the request. We then exploit speech recognition in order to extract the device class as the first unique identifier and, in addition, generate an implicit secure key from the same spoken audio command as the second unique identifier. Only the devices in proximity and in the correct device class match both unique identifiers and are thus identified for device pairing by a remote device manager.

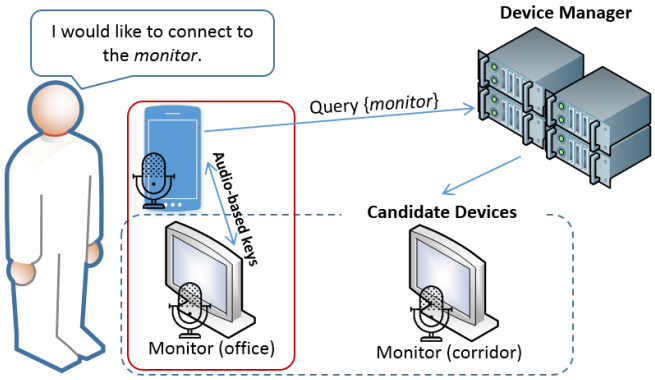

Fig. 1. Proposed concept: Voice command for secure spontaneous device pairing

In the following sections, after introducing the related work in section II we first briefly introduce contextual data and fuzzy cryptography concepts (section III). Then, we describe key generation methods for audio data in section IV. In particular, an audio fingerprinting algorithm is used to extract contextual characteristics of ambient sounds. We design a communication protocol to derive shared secret keys among proximate devices. In section V, we discuss a case study conducted with an Android application we developed to implement the proposed speech recognition and audio-fingerprinting-based device pairing mechanism. Section VI concludes our discussion.

II. Related Work

With increasing density of mobile and smart device deployment, spontaneous interaction patterns in which environmental devices or services are accessed ad-hoc in the context of use are expected to increase. Since manual pairing for this multitude of interactions is not feasible, co-presence based pairing in the same context can be exploited to generate common secrets among devices [5].

In recent work, a popular sensor to detect co-presence has been the accelerometer. For instance, [6] present a process to generate shared keys based on shaking processes. A similar approach has been followed by Mayrhofer et al. [7], who demonstrated that an authentication is possible when devices

are shaken simultaneously by a single person. However, simultaneous shaking of devices is limited to lightweight device classes that can be easily lifted and carried by individuals.

For authentication based on sensor data from arbitrary co-aligned devices, [8] propose the candidate key protocol, which interactively exchanges hashes from feature sequences as short secrets and concatenates the key from the secrets with matching hashes. An alternative related protocol has been proposed in [3]. Sensor modalities suited for unattended co-presence-based device pairing include magnetometer [9], RF-signals [10, 11] luminosity [12] or audio [2]. The authors of [13] investigated the performance of four commonly available sensor modalities (WiFi, Bluetooth, GPS, and audio) for co-presence detection and find that WiFi is better than the rest. Also, they show that, compared to any single modality, fusing multiple modalities improves resilience while retaining a high level of usability. In our case, however, we decide for audio over WiFi and Bluetooth, since it features better room-level recognition due to the longer wavelength and hence less drastically changing environment of the channel. GPS is not feasible in our case since it ceases to work indoors. Miettinen et al. [12] use co-presence and a continuous authentication scheme to pair devices. Their underlying assumption states that only devices that are worn together or are located nearby will *in the long run* measure the same luminosity or ambient audio. A further example of proximity-based device pairing related to our work but which requires manual user interaction is presented in [14]. The system uses fuzzy cryptography to generate a shared secret on two devices from correlated drawings on the devices' displays [15].

A conceptional challenge with all context-based authentication approaches is that due to sensing inaccuracies, different hardware and noise the sensed signals are likely not identical but only similar. Fuzzy cryptography presents a methodology to obtain identical keys from similar patterns [16]. In particular, by mapping the patterns into the codespace of an error correcting code, mismatches can be mitigated without disclosing the pattern over a potentially insecure channel. These approaches have been applied to various noisy data traces for authentication, such as face biometrics [17].

We propose to utilize ambient audio to pair a mobile device with an environmental device or service in proximity, such as, for instance, in the same room. For this, we extract audio fingerprints from ambient audio to obtain a binary sequence which is then mapped into the codespace of an error correcting code. With Fuzzy cryptography, remaining bit errors are then removed from the bitstring so that devices in proximity, which observe similar audio and, hence, generate similar audio fingerprints, arrive at identical secure secrets.

For audio fingerprinting, various approaches have been proposed in the literature. For instance, [18] presented a robust audio recognition approach for ambient audio sequences. In particular, the authors consider 22 daily activities in the bath and kitchen and distinguish between their audio patterns. After MFCC computation, the signal is clustered utilising random forest classifier before applying an ensemble-voting recognition. [19] presented an audio fingerprinting approach which is robust to modification of the original audio. The authors detect salient spectral points and compare these pairwise to other points in the audio spectrum applying a distance-based template on the spectrum graph in order to derive a representative binary pattern of the audio sequence. A robust audio fingerprinting approach has been proposed in [20], in which the fingerprint is constructed from the vector of characteristic peaks in the frequency-time domain. [21] enhanced this method by proposing a scalable and robust audio fingerprinting method tolerable to time-stretching. Features are, in this approach considered only with respect to frequency and independent from the time domain. Similarly, [22] defines representative feature combinations for audio identification systems. The algorithm works on translation- and scale-invariant hashes of combinations of spectral peaks.

The approach to utilize co-presence for the pairing of devices in a smart office space has been investigated by other authors before, prominently by [23], who link pervasive displays to a mobile device. The approach displays a URL on the screen of the device (alternatively in the form of QR-Code or NFC) and the mobile phone accesses the device by accessing the url.

In contrast, in our case, we propose an implicit pairing of devices and condition the pairing on spoken audio for the double purpose of (1) identifying the device class to pair with via speech recognition (e.g. display, printer, cloud storage, ...) and (2) identifying co-presence from identical audio-fingerprints generated on the respective devices. In particular, in order to access and pair with a device in proximity, the user articulates the device-class she would like to access in arbitrary free-form spoken language. We do not restrict the spoken text or sentence to any format but require that the sentence specifies the device class (such as '*I would like to connect to the printer*', '*please connect me to the beamer*', '*Use the cloud storage for this*', ...). While the device class is identified via speech recognition, the audio of the spoken sentence is exploited to generate an audio fingerprint and from this a secure key on devices in proximity. Only devices in proximity are thus able to generate the key and our system will connect the device of the user with the device of the correct class that can prove proximity by presenting the correct key.

III. Background

A. Contextual Data

Proximate mobile devices share similar contexts, such as ambient audio. Hence, we propose to leverage contextual data as a source to generate secret keys for ad-hoc device-to-device communication. To initialize a secure connection, each device independently record contextual data and extract characteristics of the data, called fingerprints. Each fingerprint f is represented as a binary sequence. The entropy of audio fingerprints (i.e. material to generate keys in our case study) has been proved to be sufficiently secure for cryptographic communication [2].

B. Fuzzy Cryptography

To derive unique shared secrets on two participating devices without exchanging additional information for comparison, error correcting codes are utilized. Error correcting codes are normally used to encode messages from the messagespace $m \in \mathcal{M}$ into codewords of the (larger) codespace $c \in \mathcal{C}$ introducing redundancies

$$m \xrightarrow{Encode} c \tag{1}$$

This process allows to correct errors introduced when transmitting c over a lossy channel before decoding it back to m from c:

$$c \xrightarrow{Decode} m \tag{2}$$

We apply error correcting codes in a different way. In a sense, our fingerprints $\boldsymbol{f}$ are lossy as they are not entirely equal on the devices trying to pair. Here, the codespace $\mathcal{C}$ is chosen in a way that we can directly map a fingerprint $\boldsymbol{f}$ into this codespace and apply the $Decode$-method $\boldsymbol{f}$:

$$\xrightarrow{Decode} \boldsymbol{k} \tag{3}$$

to derive a binary key $\boldsymbol{k}$ that is error-corrected. Due to the usage of binary fingerprints, we used Reed-Solomon codes [24]. A Reed-Solomon code can be parameterized to correct up to t errors, which depends on the similarity of audio fingerprints.

IV. Audio-based Secure Device Pairing Protocol

Ambient audio, including human voice, music, and sound of walking steps, characterizes the contextual information of a certain location (e.g. meeting room). Originally, audio-fingerprinting [25] was leveraged in music identification. We use binary fingerprints from ambient audio in order to establish secure communication of proximate devices.

A. Audio Fingerprinting

After recording sufficient audio samples, we leverage frequency domain representation of audio to extract audio fingerprints following the approach specified in [2]. An audio fingerprint is calculated with the following steps. The audio file is split into short windows of samples. Fast Fourier Transform is applied on each window to generate the frequency domain representation. The frequencies are divided into equally-long bands. The energy of each frequency band is computed. Then, energy differences of consecutive bands are calculated. We compare the values of energy variation between short windows to form the binary audio fingerprint. Figure 2 visualizes the audio fingerpriting scheme.

However, proximate devices can obtain similar but not identical fingerprints. Thus, we apply fuzzy cryptography to generate the shared secret keys. Specifically, if the Hamming distance between two fingerprints satisfies a pre-defined threshold t, they can be corrected to form the key independently on each device. The threshold is parameterized on the physical distance between the devices (see Section III-B).

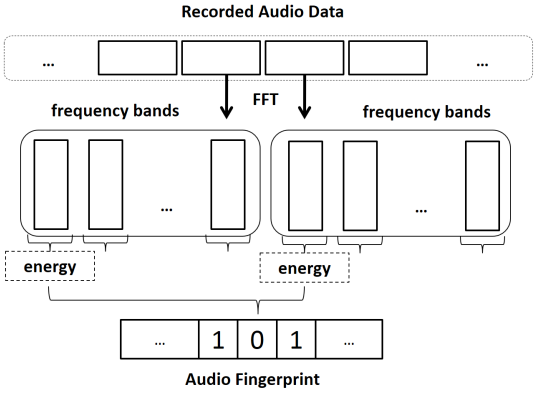

Fig. 2. Audio fingerprinting scheme

B. Audio-based Pairing Protocol

We introduce the following scheme. A set of mobile devices are willing to establish a common secret key extracted from ambient audio data. Each device records a number of audio samples and then independently compute an audio fingerprint. The fingerprints are binary sequences that are designed to fall into the code-space of a Reed-Solomon error correcting code. Audio fingerprints generated from similar ambient audio resemble. However, due to noise and inaccuracy in the audio-sampling process (i.e. caused by hardware and software diversity), it is unlikely that two fingerprints are identical. Devices therefore exploit the error correction capabilities of the error correcting code utilized to map fingerprints to codewords. For fingerprints with a Hamming distance within the configured threshold t of the error correcting code, the codewords are identical and then can be utilized as shared secret keys.

Leveraging the audio-based approach, we can establish a secure communication session between a user's device and a local device with or without a central authority (we will refer to this as the Device Manager). We assume that both partners are in the same network, and the approach increases the connection security with contextual information. Our proposed mechanism ensures that the devices can instantiate a secure session if they are proximate to each other (e.g. in a physical room). We denote $KGF()$ as an audio-based key generation function which derives one cryptographic key from an audio fingerprint of ambient sounds. Then, the user's device D establishes a secure session with the local device L according to the following steps (cf. figure 3).

1) D sends the initialization message to L to start the key generation process
2) D and L start recording a sequence of ambient audio whose length is specified a priori.
3) D and L locally compute the audio fingerprints f_D and f_L, respectively.
4) D transforms f_D to the secret key $K_D = KGF(f_D)$ while L transforms K_L to the secret key $K_L = KGF(f_L)$. Using an error correcting code (e.g. Reed-Solomon scheme [24]), both partners can derive the

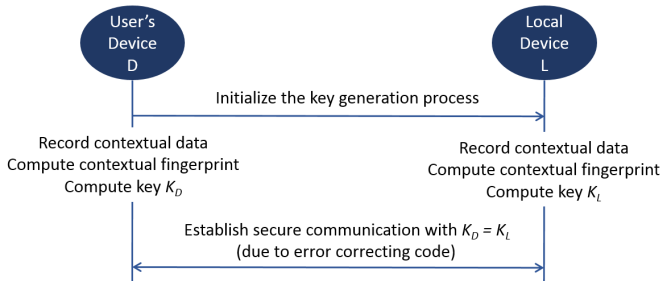

Fig. 3. Protocol for key exchange and management without a central authority

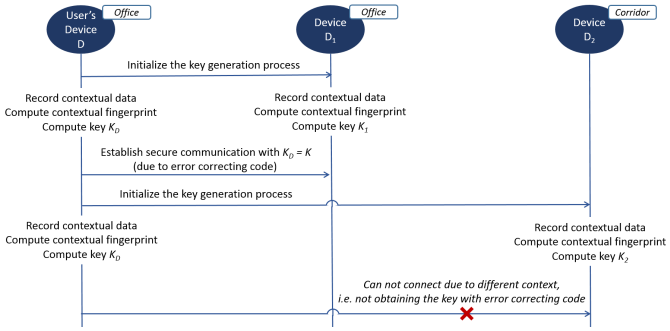

Fig. 4. Protocol for device group formation based on context

same secret key K if the Hamming distance between K_D and K_L satisfies a predefined threshold t.

This is represented in our basic scenario:

- Scenario 1 (cf. Figure 3): *Key exchange and management without a central authority*. In case there is no central key authority, proximate devices can leverage contextual information to form ad-hoc device groups. The registration and de-registration process (i.e. joining and leaving a group) rely on context-based secret keys only.

The protocol can also be integrated into further scenarios that may featuer a central authority (e.g. Device Manager):

- Scenario 2 (cf. Figure 4): *Device group formation based on context*. Our mechanism adds a fine-grained layer to the conventional group key management framework. For example, the users smartphone S is connected with the printer P_1 (at the corridor) and the printer P_2 (at the user's office) in the local wireless network. That means S, P_1, and P_2 share the group key. With our context-based key generation technique, the framework can issue a new secret key only for S and P_1 (based on proximity). S does not need to leave the former group. Furthermore, if attackers compromise P_2, they can not access the new group formed by S and P_1.
- Scenario 3 (cf. Figure 5): *Context-based device discovery*. A user's device U is in the same group with multiple local devices L_i, which may belong to different contexts (e.g. in different rooms). U wants to access an unprecedented L_i in the same room. It can obtain the device information at a certain location, i.e. L_i. After that, U can connect to the specific L_i.

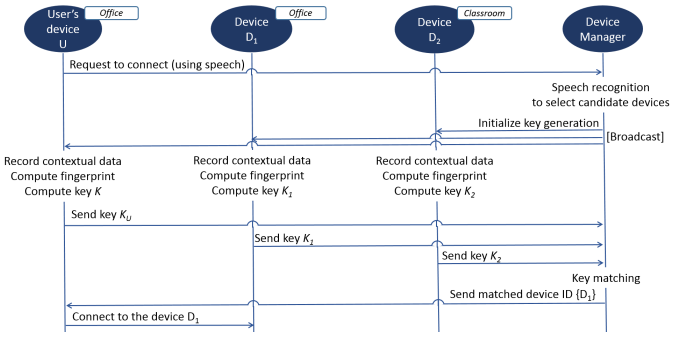

Fig. 5. Protocol for context-based device discovery

These scenarios are appropriate for Intenet-of-Things devices, including mobile and wearable appliances. Moreover, these mechanisms strengthen the traditional device pairing mechanisms (e.g. Bluetooth or ad-hoc wireless network) in terms of security and usability. The configuration of our proposed protocol (e.g. Reed-Solomon-Code parameters, cryptographic primitives, key length, etc) can be customized depending on specific systems and security policies.

V. Case Study and Result

A. Audio Fingerprinting of Vocal Commands

In order to verify and demonstrate the feasibility of the proposed use case specified in Figure 1, we conducted a case study in two environments of our university. This case study implements the protocol specified in figure 5 The case study shall demonstrate that audio-based pairing based on free-text input of a user is feasible. For this, we developed an Android application which extracted and compared audio fingerprints of proximate devices. In particular, while the subjects specified the device they want to pair with unconstrained free speech, the application recorded the ambient audio, extracted audio fingerprints as specified in [2] and compared the fingerprint similarity. As detailed in [2], fuzzy cryptography can then be exploited in order to correct bit errors in the generated fingerprints.

For the experiment, two android mobile phones running the audio-based ad-hoc pairing were placed in the same room in distances of at least one meter. The subject then chose one out of a given five possible device classes (printer, projector, monitor, speaker, TV) and spoke out a free form request containing the name of the particular device (for instance, the command might be *"I would like to connect to the speaker"* if a user wants to pair her smartphone with the Bluetooth speaker in the room) while the android application was running, recording the ambient audio to generate the secure keys separately on the two devices. The users were asked to issue natural voice commands. Spoken audio sentences have been in the order of 2-4 seconds depending on speed and length of the sentence. Users are located at two distinct locations in a university campus: a meeting room and an office. The former had more background ambient sounds than the latter. In each environment, the users have repeated the experiment

TABLE I
AVERAGE SIMILARITY (%) OF AUDIO FINGERPRINTS. VALUES IN BRACKETS ARE STANDARD DEVIATION.

	Meeting room	Office
Printer	68.2 (14.8)	60.4 (10.5)
Projector	74.6 (5.2)	67.6 (16.5)
Monitor	69.6 (7.6)	74.2 (4.4)
Speaker	75.0 (13.2)	75.4 (6.6)
TV	73.8 (8.8)	63.6 (9.1)

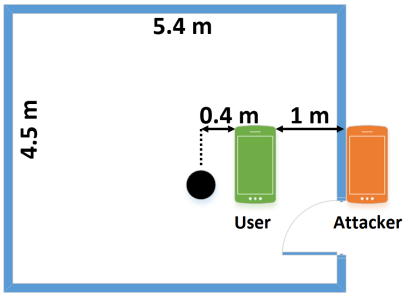

Fig. 6. Attacking scenario

10 times for each device class. Table I shows the results of the case study. The table depicts similarity of audio fingerprints generated for the various sentences containing different device classes and conducted in different locations. We observe that the similarity is in most cases above 65% and therefore suggest to utilize Reed-Solomon codes [24] to correct 35% of the bits in the generated audio fingerprint in order to arrive at the same secure key for device pairing.

B. Attacking Scenario

We installed an attacking scenario in which users stayed inside an office while adversaries were outside (i.e. at the corridor). The attacker aims to obtain the same shared secret key with the user, the local device (e.g. monitor), or the device manager. In this situation, encrypted data can be stolen when transmitting between these partners.

Figure 6 visualizes the set-up. The user, i.e. source of vocal commands, and devices were located as in the figure. We investigated two situations with open and closed door. Table II summarizes the comparison of audio fingeprints obtained on the user's and attacker's device. Due to different audio contexts, the fingerprints are less similar than those captured by devices in the same room (see Table I). Consequently, our system can be configured (e.g. modifying threshold of the error correcting code) to derive keys for devices in the same audio context only. If the adversary tries to sneak into the room, it is possible for the users to discover. Other strategies such as stealing contextual data are not in the scope of this paper. In addition, Schürmann and Sigg [2] performed an extensive analysis of the audio-based pairing scheme in various locations and setting distances, which solidifies usability and security of our proposed approach.

TABLE II
AVERAGE SIMILARITY (%) OF AUDIO FINGERPRINTS BETWEEN THE USER AND THE ATTACKER. VALUES IN BRACKETS ARE STANDARD DEVIATION.

	Open door	Closed door
Similarity	51.7 (7.2)	47.6 (11.2)

VI. CONCLUSION

In this paper, we presented three context-based implicit pairing protocols and described how these can be integrated into distributed encryption schemes for distributed IoT devices. In particular, we implemented and evaluated a spontaneous device pairing protocol based on ambient audio information. In the protocol, free-form spoken interaction is interpreted by speech recognition to identify the device class to pair with while audio-fingerprints are generated from the same spoken interaction in order to generate secure keys via fuzzy cryptography. Both, the device class and the secure key are then utilized as unique identifier to pair a personal device with a proximate device of the requested class. We performed a case study in which users select partner devices through natural voice commands. An Android application was implemented to evaluate our mechanism in two distinct locations.

ACKNOWLEDGMENT

We appreciate partial funding in the frame of an EIT Digital HII Active project, as well as from Academy of Finland and from the German Academic Exchange Service (DAAD).

REFERENCES

[1] D. Schuermann, A. Bruesch, S. Sigg, and L. Wolf, "Bandana body area network device-to-device authentication using natural gait," in *2017 IEEE International Conference on Pervasive Computing and Communication*, 2017.

[2] D. Schürmann and S. Sigg, "Secure communication based on ambient audio," *IEEE Transactions on Mobile Computing*, vol. 12, no. 2, pp. 358–370, 2013.

[3] S. Sigg, D. Schuermann, and Y. Ji, "Pintext: A framework for secure communication based on context," in *Proceedings of the Eighth Annual International ICST Conference on Mobile and Ubiquitous Systems:Computing, Networking and Services (MobiQuitous 2011)*, 2011.

[4] S. Sigg, L. N. Nguyen, A. Huynh, and Y. Ji, "Adhoc-pairing: Spontaneous audio based secure device pairing for android mobile devices," in *Proceedings of the 4th International Workshop on Security and Privacy in Spontaneous Interaction and Mobile Phone Use, in conjunction with Pervasive 2012*, 2012.

[5] M. K. Chong, R. Mayrhofer, and H. Gellersen, "A survey of user interaction for spontaneous device association," *ACM Computing Surveys*, vol. 47, no. 1, 2014.

[6] D. Bichler, G. Stromberg, M. Huemer, and M. Löw, "Key generation based on acceleration data of shaking processes," in *International Conference on Ubiquitous Computing*. Springer, 2007, pp. 304–317.

[7] R. Mayrhofer and H. Gellersen, "Shake well before use: Authentication based on accelerometer data," in *Pervasive computing*. Springer, 2007, pp. 144–161.

[8] R. Mayrhofer, "The candidate key protocol for generating secret shared keys from similar sensor data streams," in *European Workshop on Security in Ad-hoc and Sensor Networks*. Springer, 2007, pp. 1–15.

[9] R. Jin, L. Shi, K. Zeng, A. Pande, and P. Mohapatra, "MagPairing: Pairing Smartphones in Close Proximity Using Magnetometers," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 6, 2016.

[10] A. Varshavsky, A. Scannell, A. LaMarca, and E. De Lara, *Amigo: Proximity-Based Authentication of Mobile Devices*. Berlin, Heidelberg: Springer, 2007.

[11] D. A. Knox and T. Kunz, "Wireless fingerprints inside a wireless sensor network," *ACM Transactions on Sensor Networks (TOSN)*, vol. 11, no. 2, p. 37, 2015.

[12] M. Miettinen, N. Asokan, T. D. Nguyen, A.-R. Sadeghi, and M. Sobhani, "Context-based zero-interaction pairing and key evolution for advanced personal devices," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2014.

[13] H. T. T. Truong, X. Gao, B. Shrestha, N. Saxena, N. Asokan, and P. Nurmi, "Using contextual co-presence to strengthen zero-interaction authentication: Design, integration and usability," *Pervasive and Mobile Computing*, vol. 16, pp. 187–204, 2015.

[14] M. Antikainen, M. Sethi, S. Matetic, and T. Aura, "Commitment-based device-pairing protocol with synchronized drawings and comparison metrics," *Pervasive and Mobile Computing*, vol. 16, pp. 205–219, 2015.

[15] M. Sethi, M. Antikainen, and T. Aura, "Commitment-based device pairing with synchronized drawing," in *2014 IEEE International Conference on Pervasive Computing and Communications (PerCom'14)*. IEEE, 2014.

[16] A. Juels and M. Wattenberg, "A fuzzy commitment scheme," in *Proceedings of the 6th ACM conference on Computer and communications security*. ACM, 1999.

[17] Y. Wang and K. Plataniotis, "Fuzzy vault for face based cryptographic key generation," in *Biometrics Symposium, 2007*. IEEE, 2007, pp. 1–6.

[18] J. A. Stork, L. Spinello, J. Silva, and K. O. Arras, "Audio-based human activity recognition using non-markovian ensemble voting," in *RO-MAN, 2012 IEEE*. IEEE, 2012, pp. 509–514.

[19] X. Anguera, A. Garzon, and T. Adamek, "Mask: Robust local features for audio fingerprinting," in *Multimedia and Expo (ICME), 2012 IEEE International Conference on*. IEEE, 2012, pp. 455–460.

[20] A. Wang *et al.*, "An industrial strength audio search algorithm." in *ISMIR*, vol. 2003. Washington, DC, 2003, pp. 7–13.

[21] J. George and A. Jhunjhunwala, "Scalable and robust audio fingerprinting method tolerable to time-stretching," in *Digital Signal Processing (DSP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 436–440.

[22] R. Sonnleitner and G. Widmer, "Quad-based audio fingerprinting robust to time and frequency scaling." in *DAFx*, 2014, pp. 173–180.

[23] E. Oat, M. Di Francesco, and T. Aura, "Mocha: Augmenting pervasive displays through mobile devices and web-based technologies," in *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2014 IEEE International Conference on*. IEEE, 2014, pp. 506–511.

[24] I. S. Reed and G. Solomon, "Polynomial Codes Over Certain Finite Fields," *Journal of the Society for Industrial and Applied Mathematics*, vol. 8, no. 2, pp. 300–304, 1960. [Online]. Available: http://dx.doi.org/10.1137/0108018

[25] J. Haitsma and T. Kalker, "A highly robust audio fingerprinting system," in *International Society for Music Information Retrieval Conference*, 2002, pp. 107–115.