Secure communication based on ambient audio

Dominik Schürmann, and Stephan Sigg, *Member, IEEE*

Abstract—We propose to establish a secure communication channel among devices based on similar audio patterns. Features from ambient audio are used to generate a shared cryptographic key between devices without exchanging information about the ambient audio itself or the features utilised for the key generation process. We explore a common audio-fingerprinting approach and account for the noise in the derived fingerprints by employing error correcting codes. This fuzzy-cryptography scheme enables the adaptation of a specific value for the tolerated noise among fingerprints based on environmental conditions by altering the parameters of the error correction and the length of the audio samples utilised. In this paper we experimentally verify the feasibility of the protocol in four different realistic settings and a laboratory experiment. The case-studies include an office setting, a scenario where an attacker is capable of reproducing parts of the audio context, a setting near a traffic loaded road and a crowded canteen environment. We apply statistical tests to show that the entropy of fingerprints based on ambient audio is high. The proposed scheme constitutes a totally unobtrusive but cryptographically strong security mechanism based on contextual information.

Index Terms—J.9.d Pervasive computing, E.3 Data Encryption, G.3.j Random number generation, H.5.5.c Signal analysis, synthesis, and processing, I.5.4.m Signal processing, J.9.a Location-dependent and sensitive

1 Introduction

An important factor in the set of security risks is typically the human impact. People are occasionally careless or incompletely understanding the underlying technology. This is especially true for wireless communication. For instance, the communication range or the number of potential communication partners might be underestimated. This is natural since humans typically base trust on the situation or context they perceive [1]. Nevertheless, the range of a communication network most likely bridges devices in various contexts.

As context, proximity and trust are related [1], a security scheme that utilises common contextual features among communicating devices might provide a sense of security which is perceived as natural by individuals and reduce the number of human errors related to security.

Consider, for instance, a meeting with co-workers of a specific project. Naturally, workers trust the others based on working agreements. Every group member needs the permission to access common information like mobile phone numbers or shared files. Communication between group members, however, should be guarded against access from external devices or individuals. The meeting room defines the borders which shall not be crossed by any confidential data. Context information that is unique inside these borders, such as ambient audio, can be exploited as the seed to generate a common secret for the secure information exchange and authentication.

Mobile phones can then synchronise their ID-cards ad-hoc without user interaction and secured by their physical proximity. Similarly, access to shared files on computers of co-workers and communication links among co-workers can be secured.

Another reason why security cautions might be discarded occasionally is the effort required and inconvenience to establish a secure connection. This is especially true between devices that communicate seldom or for the first time.

We propose a mechanism to unobtrusively (zero interaction) establish an ad-hoc secure communication channel between unacquainted devices which is conditioned on the surrounding context. In particular, we consider audio as a source of spatially centred context. We exploit the similarity of features from ambient audio by devices in proximity to create a secure communication channel exclusively based on these features. At no point in the protocol the secret itself or information that could be used to derive audio feature values is made public. In order to do so, we generate synchronised audio-fingerprints from ambient sounds and utilise error correcting codes to account for noise in the feature vector. On each communicating device the feature vector is then used to create an identical key. The proposed protocol is non-interactive, unobtrusive and does not require specific or identical hardware at communication partners.

The remainder of this document is structured as follows. In section 2 we introduce related work on context-based security mechanisms and security with noisy input data. Section 3 discusses the algorithmic background required for ambient audio-based key generation and implementation details. In section 4 we discuss the noise and entropy of audio-fingerprints achieved in an offline-experiment with sampled audio sequences. We show that the similarity in audio-fingerprints is sufficient for authentication but can not be utilised as secure key directly. In particular, we utilise fuzzy-cryptography schemes to account for noise in the input data. Section 5

- *D. Schürmann is with the TU Braunschweig, Braunschweig, Germany. E-mail: d.schuermann@tu-braunschweig.de*
- *S. Sigg is with the Information Systems Architecture Science Research Division, National Institute of Informatics (NII), Tokyo, Japan. E-mail: sigg@nii.ac.jp*

presents four case-studies in different environments that exploit the feasibility of the approach in various settings. The general feasibility of the approach is demonstrated in section 5.1 in a controlled office environment. Section 5.2 then shows that the audio context can be separated between two offices even when a synchronised audio source is located in both places. Additionally, we studied the feasibility of ambient audio-based key generation at the side of a heavily trafficked road in section 5.4 and in a canteen environment in section 5.3. The entropy of the ambient audio-based characteristic binary sequences generated by our method is discussed in section 6. In section 7 we draw our conclusion.

2 Related work

In the literature, several authors consider spontaneous authentication or the establishing of a secure communication channel among mobile and ad-hoc devices based on environmental stimuli [2], [3], [4], [5]. So far, shaking processes from accelerometer data and RF-channel measurements have been utilised as unique context source that contains shared characteristic information.

This concept was presented 2001 by Holmquist et al. [4]. The authors propose to utilise the accelerometer of the Smart-It [6] device to extract characteristic features from simultaneous shaking processes of two devices. Later, Mayrhofer et al. presented an authentication mechanism based on this principle [7]. The authors demonstrated, that an authentication is possible when devices are shaken simultaneously by a single person, while an authentication was unlikely for a third person trying to mimic the correct movement pattern remotely. Also, Mayrhofer derived in [8] that the sharing of secret keys is possible with a similar protocol. The proposed protocol that can be utilised with arbitrary context features repeatedly exchanges hashes of key-sub-sequences until a common secret is found. In this instrumentation, exponentially quantised fast Fourier transformation (FFT) coefficients of a sequence of accelerometer samples are utilised. In contrast, Bicher et al. describe an approach in which noisy acceleration readings can be utilised to establish a secure communication channel among devices [9], [3]. They utilise a hash function that maps similar acceleration patterns to identical key sequences. However, their approach suffers from the required exact synchronisation among devices so that the authors computed the correct hash-values offline. Additionally, the hash function utilised required that the keys computed exactly match and that the neighbourhood around these keys is precisely defined. When patterns are located at the border of one of the region's neighbourhoods, the tolerance for noise in the input is biased in the direction of the centre of this region. Additionally, key generation by simultaneous shaking is not unobtrusive.

We utilise an error correction scheme to account for noise in the input data which can be fine-tuned for any Hamming distance desired which is centred around the noisy characteristic sequences generated instead of an artificially defined centre value. We implement a Network Time Protocol (NTP) based synchronisation mechanism that establishes sufficient synchronisation among nodes.

Another sensor class utilised for context-based device authentication is the RF-channel. Varshavsky et al. present a technique to authenticate co-located devices based on RF-measurements since channel measurements from devices in near proximity are sufficiently similar to authenticate devices against each other [5]. Hershey et al. utilise physical layer features to derive secret keys for a pair of devices [10]. In the absence of interference and non-linear components, transmitter and receiver experience identical channel response [11]. This information is utilised to generate a secret key among a node pair. Since channel characteristics are spatially sharply concentrated and not predictable at a remote location [12], an eavesdropper is not capable of guessing information about the secret. This scheme was validated in an indoor environment in [13]. Although we consider the keys generated by this scheme as strong, it does not preserve spatial properties. A device at arbitrary distance could pretend to be a nearby communication partner.

Kunze and Lukowicz recently demonstrated, that audio information indeed suffices to derive spatial information [14]. They combine audio readings with accelerometer data to classify locations of mobile devices. In their work, the noise emitted by a vibrating mobile phone was utilised to distinguish among 35 specific locations in three different rooms with over 90 % accuracy.

Instead, we utilise purely ambient noise to establish a secure communication channel among devices in spatial proximity. We record NTP-synchronised audio samples at two locations, generate a characteristic audio-fingerprint and map this fingerprint to a unique secret key with the help of error correcting codes.

The last step is necessary since the similarity between fingerprints is typically not sufficient to establish a secure channel. With fuzzy-cryptography schemes, the generation of an identical key based on noisy input data [15] is possible. Li et al. analyse the usage of biometric or multimedia data as part of an authentication process and propose a protocol [16]. Due to the use of error-tolerant cryptographic techniques, this protocol is robust against noise in the input data. The authors utilise a secure sketch [17] to produce public information about an input without revealing it. The input can then be recovered given another value that is close to it. A similar study is presented by Miao et al. [18]. The authors establish a key distribution based on a fuzzy vault [19] using data measured by devices worn on the human body. The fuzzy vault scheme, also utilised in [20], enables the decryption of a secret with any key that is substantially similar to the key used for encryption.

3 Ad-hoc audio-based encryption

Originally, audio-fingerprinting was proposed to classify music or speech. In our work binary fingerprints

from ambient audio are used to establish an encrypted connection based on the surrounding audio context. Due to differences between fingerprints generated by participating devices, a cryptographic protocol is needed that tolerates a specific amount of noise in these keys.

We propose the following scheme. A set of devices willing to establish a common key conditioned on ambient audio take synchronised audio samples from their local microphones. Each device then computes a binary characteristic sequence for the recorded audio: An audio-fingerprint (cf. section 3.1). This binary sequence is designed to fall onto a code-space of an error correcting code (cf. section 3.2). In general, a fingerprint will not match any of the codewords exactly. Fingerprints generated from similar ambient audio resemble but due to noise and inaccuracy in the audio-sampling process, it is unlikely that two fingerprints are identical. Devices therefore exploit the error correction capabilities of the error correcting code utilised to map fingerprints to codewords (cf. section 3.3). For fingerprints with a Hamming-distance within the error correction threshold of the error correcting code the resulting codewords are identical and then utilised as secure keys (cf. section 3.4). This scheme is in principle not limited in the number of devices that participate. When devices are synchronised in their local times, they agree on a point in time when audio shall be recorded and proceed with the fingerprint creation and error correction autonomously as described above. All similar fingerprints will map to an identical codeword. As detailed in section 5.3, the Hamming distance tolerated in fingerprints rises with increasing distance of devices.

The following sections provide an overview over audio-fingerprinting, our fuzzy commitment implementation, problems we experienced and possible solutions.

3.1 Audio-fingerprinting

Audio-fingerprinting is an approach to derive a characteristic pattern from an audio sequence [21]. Generally, the first step involves the extraction of features from a piece of audio. These features are usually isolated in a time-frequency analysis after application of Fourier or Cosine transforms. Some authors also utilise wavelet-transforms [22], [23], [24]. Common applications include the retrieval of a specific music file in an audio database [25], duplicate detection in such a database [26] as well as identification of music based on short samples [27]. The capabilities of detecting similar audio sequences in the presence of heavy signal distortion are prominently demonstrated by applications such as query by humming [28]. The authors utilise autocorrelation, maximum likelihood and Cepstrum analysis to describe the pitch of an audio sequence as a Parsons encoded music contour [29]. Similar audio sequences are detected by approximate string matching [30]. McNab et al. added rhythm information by analysing note duration to match the beginning of a song [31]. A similar approach is presented by Prechelt et al. [32]. They achieved more accurate results for query by whistling since the frequency range of whistling is much lower than for humming or singing. In 2002, Chai et al. computed a rough melodic contour by counting the number of equivalent transitions in each beat [33]. Notes are detected by amplitude-based note segmentation. Later, Shiffrin et al. showed that songs can be described by Markov-chains [34] where states represent note transitions. Retrieval of songs is then achieved by the HMM Forward algorithm [35] so that no database query is required. In 2003, Zhu et al. addressed practical problems of recently proposed approaches such as the accuracy of the derived description by utilising a dynamic time-warping mechanism [36].

Most of these studies are based on music-specific properties such as rhythm information, pitch or melodic contour. Since such features might be missing in ambient audio, these methods are not applicable in our case. Haitsma et al. presented in [37] an approach applicable for the classification of general audio sequences by extracting a binary representation of audio from changes in the energy of successive frequency bands. This system was later shown to be highly robust to noise and distortion in audio data [25]. Due to its reported robustness, several authors employ slightly modified versions of this approach [27]. Lebossé et al, for instance, add further redundant sub-samples taken from the beginning and the end of an overlapping time window in order to reduce the number of bits in the fingerprint representation [38]. Alternatively, Burges et al. enhance the former approach by utilising a distortion discriminant analysis [39]. Generally, time frames taken from the audio source are mapped successively on smaller time windows in order to generate a condensed characteristic representation of the audio sequence. An alternative approach based on spectral flatness of a signal is proposed Herre et al. [40].

Also, Yang presented a method to utilise characteristic energy peaks in the signal spectrum in order to extract a unique pattern [41]. A general framework that supports this scheme was later presented by Yang et al. [42]. Building on these ideas, a similar algorithm was then successfully applied commercially by Avery Wang on a huge data base of audio sequences [43], [44].

To create audio-fingerprints for our studies, we split an audio sequence S with length $|S| = l$ and sample rate r up into n frames $S_1, \ldots, S_n$ of identical length $d = |S_i| = r \cdot \frac{l}{n}$. On each frame a discrete Fourier transformation (DFT) weighted by a Hanning window (HW) is applied:

$$\begin{aligned} &\forall i \in \{0, \ldots, n-1\}, \\ &\overline{S_i} = DFT\,(HW(F_i)) \end{aligned} \tag{1}$$

The frames are divided into m non-overlapping frequency bands of width

$$b = \frac{\text{maxfreq}(S_i) - \text{minfreq}(S_i)}{m}. \tag{2}$$

On each band the sum of the energy values is calculated and stored to an energy matrix E with energy per frame

per frequency band.

$$\forall j \in \{0, \dots, m-1\},$$
$$S_{ij} = \text{bandfilter}_{b \cdot j, b \cdot (j+1)}(S_i) \quad (3)$$
$$E_{ij} = \sum_k S_{ij}[k] \quad (4)$$

Using the matrix E, a fingerprint f is generated, where $\forall i \in \{1, \dots, n-1\}, \forall j \in \{0, \dots, m-2\}$ each bit describes the difference between the energy on frequency bands between two consecutive frames:

$$f(i,j) = \begin{cases} 1, & (E(i,j) - E(i,j+1)) - \\ & (E(i-1,j) - E(i-1,j+1)) > 0 \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

The complete algorithm is detailed in the appendix.

For each synchronisation, we sampled $l = 6.375$ seconds of ambient audio at a sample rate of $r = 44100$ Hz. We split the audio stream into $n = 17$ frames of $d = 0.375$ seconds each and divide every frame into $m = 33$ frequency bands, to obtain a 512 bit fingerprint. Due to the extensive recording duration, the generated fingerprints show great robustness in real world experiments (cf. section 4 and section 5). We used a Fast Fourier Transform (FFT) with fixed values on the length of the segments as detailed above.

This audio-fingerprinting scheme utilised in our studies utilises energy differences between frequency bands, as proposed by Haitsma et al. [25]. However, we take a more general approach of classifying ambient audio instead of music. Commonly, in the literature, the characteristic information is found in a smaller frequency band and a logarithmic scaling is suggested to better represent properties of the human auditory system. Since our system is not restricted to musical recordings, we expect that all frequency bands are equally important. Therefore, we divide frames into frequency bands at a linear scale rather than a logarithmic one. Additionally, we do not use overlapping frames since this has not shown improvements in our case. Also, the entropy and therefore the security features of the generated fingerprint is likely to become impaired with overlapping frames [45], [46].

3.2 Audio-fingerprints as cryptographic keys

To use the audio-fingerprints directly as keys for a classic encryption scheme the concurrence of fingerprints generated from related audio sequences has to be 1 with a considerably high probability [47]. Since we experienced a substantial difference in the audio-fingerprints created (cf. section 4) we consider the application of fuzzy-cryptography schemes. Note that a perfect match in fingerprints is unlikely since devices are spatially separated, not exactly synchronised and utilise possibly different audio hardware.

The proposed cryptographic protocol shall be feasible unattended and ad-hoc with unacquainted devices. For an eavesdropper in a different audio context it shall be computationally infeasible to use any intercepted data to decrypt a message or parts of it. Additionally, we want to control the threshold for the tolerated offset between fingerprints based on contextual conditions of different physical locations.

With fuzzy encryption schemes, a secret ς is used to hide the key κ in a set of possible keys $\mathcal{K}$ in such a way that only a similar secret ς' can find and decrypt the original key κ correctly. In our case, the secrets which ought to be similar for all communicating devices in the same context are audio-fingerprints.

A Fuzzy Commitment scheme can, for instance, be implemented with Reed-Solomon codes [48]. The following discussion provides a short introduction to these codes.

Given a set of possible words $\mathcal{A}$ of length m and a set of possible codewords $\mathcal{C}$ of length n, Reed-Solomon codes $RS(q, m, n)$ are initialised as:

$$\mathcal{A} = \mathbb{F}_q^m, \quad (6)$$
$$\mathcal{C} = \mathbb{F}_q^n, \quad (7)$$

with $q = p^k, p$ prime, $k \in \mathbb{N}$. These codes are mapping a word $a \in \mathcal{A}$ of length m uniquely to a specific codeword $c \in \mathcal{C}$ of length n:

$$a \xrightarrow{Encode} c, \quad (8)$$

This step adds redundancy to the original words with $n > m$, based on polynomials over Galois fields [48].

Decoding utilises the error correction properties of the Reed-Solomon-based encoding function to account for differences in the fingerprints created. The decoding function maps a set of codewords from one group $C = \{c, c', c'', \dots\} \subset \mathcal{C}$ to one single original word. It is

$$\tilde{c} \xrightarrow{Decode} a \in \mathcal{A}. \quad (9)$$

The value

$$t = \left\lfloor \frac{n-m}{2} \right\rfloor \quad (10)$$

defines the threshold for the maximum number of bits between codewords that can be corrected in this manner to decode correctly to the same word a [49]. In the following algorithms the fingerprints f and f' are used in conjunction with codewords to make use of this error correction procedure. Dependent on the noise in the created fingerprints, t can then be chosen arbitrarily.

3.3 Commit and Decommit algorithms

We utilise Reed-Solomon error correcting codes in the following scheme to generate a common secret among devices. A fingerprint f is used to hide a randomly chosen word a as the basis for a key in a set of possible words $a \in \mathcal{A}$. This is a commit method. A decommit method is constructed in such a way that only a fingerprint f' with maximum Hamming distance

$$\text{Ham}(f, f') \leq t \quad (11)$$

can find a again. We use Reed-Solomon $RS(q, m, n)$ codes, with $q = 2^k$, $k \in \mathbb{N}$ and $n < 2^k$, for our commit and decommit methods. After initialisation, a private word $a \in \mathcal{A}$ is randomly chosen. It is then encoded following the Reed-Solomon scheme to a specific codeword c. For a subtract-function $\ominus$ in $\mathcal{C} = \mathbb{F}_{2^k}^n$, the difference to the fingerprint is calculated as

$$\delta = f \ominus c. \tag{12}$$

Then, a SHA-512 hash [50] $\mathtt{h}(a)$ is generated from a. Afterwards, the tuple $(\delta, \mathtt{h}(a))$ containing the difference and the hash is made public. Note that the transmission of $\mathtt{h}(a)$ is optional and is only required to check whether the decommitted a' on the receiver side equals a. However, provided a sufficiently secure hash function, an eavesdropper does not learn additional information about the key a within reasonable time provided that she is ignorant of a fingerprint sufficiently similar to f.

The decommitment algorithm uses the public tuple $(\delta, \mathtt{h}(a))$ together with the secret fingerprint f' to verify the similarity between f and f' and to obtain a shared word a. A codeword c' is calculated by subtracting f' by δ in $\mathbb{F}_{2^k}^n$.

$$c' = f' \ominus \delta. \tag{13}$$

Afterwards c' is decoded to a' as

$$a' \in \mathcal{A} \xleftarrow{\text{Decode}} c' \in \mathcal{C}. \tag{14}$$

From $\mathtt{h}(a) = \mathtt{h}(a')$ we can conclude $a = a'$ with high probability. This procedure is capable of correcting up to t (cf. equation (10)) differing bits between the fingerprints. The decommitment was then successful and differences between f and f' are t at most. The decommitted word a' is privately saved.

Participants can use their private words to derive keys for encryption. A simple example for using $a = a' = (a_0, \ldots, a_{m-1})$ to generate an encryption key for the Advanced Encryption Standard (AES) [51] is to sum over blocks of values of a. For example, when $m = 256$ we would sum over blocks with the length 8 and take these values modulo $2^8 - 1$ to represent characters for a string with the length 32, that can be used as a key κ:

$$\text{Let } \kappa = (\kappa_0, \ldots, \kappa_{31}), \text{ whereas}$$
$$\kappa_i = \left(\sum_{j=0}^{7} c_{(i*8)+j} \right) \mod 2^8 - 1$$

In our study, for fingerprints of 512 bits we apply Reed-Solomon codes with $RS(q = 2^{10}, m, n = 512)$. Given a maximum acceptable Hamming distance t^* (cf. equation (10)) between fingerprints we can then set m flexibly to define the minimum required fraction u of identical bits in fingerprints as

$$t^* = \lceil (1 - u) \cdot n \rceil, \tag{15}$$
$$m = n - 2 \cdot t^*. \tag{16}$$

Experimentally, we found $u = 0.7$ as a good trade-off for common audio environments to allow a sufficient amount of differences among the used fingerprints to pair devices successfully while at the same time providing sufficient cryptographic security against an eavesdropper in a different audio context (cf. section 4).

$$\begin{aligned} m &= 512 - 2 \cdot \lceil (1 - 0.7) \cdot 512 \rceil \\ &= 204 \end{aligned} \tag{17}$$

We therefore use Reed-Solomon codes with

$$RS(2^{10}, 204, 512). \tag{18}$$

The commit and decommit algorithms are further detailed in the appendix.

3.4 Synchronising communicating devices

Since audio is time-dependent, a tight synchronisation among devices is required. In particular, we experienced that fingerprints created by two devices were sufficiently similar only when the synchronisation offset among devices was within tens of milliseconds. For synchronisation, any sufficiently accurate time protocol such as the Network Time Protocol (NTP) [52], [53], the Precision Time Protocol (PTP) [54] or a similar time protocol can be utilised. Also, synchronisation with GPS time might be a valid option.

When two participants, Alice and Bob, are willing to communicate securely with each other, Alice starts the protocol by requesting a pairing with Bob. Then, they synchronise their absolute system times using a sufficiently accurate time protocol. Afterwards, Alice sends a start time τ_{start} to Bob. When their clocks reach τ_{start}, the recording of ambient audio is initiated and audio-fingerprinting is applied.

In our case-studies, synchronisation of devices was a critical issue. Since the approach bases the binary fingerprints on energy differences of sub-samples of 0.375 seconds width, a misalignment of several hundreds of milliseconds results in completely different fingerprints. For best results, the start times of the audio recordings should not differ more than about 0.001 seconds. We successfully tested this with a remote NTP-server and also with one of the devices hosting the server.

Still, since NTP is able to synchronise clocks with an error of several milliseconds [55], [52], some error in the synchronisation of audio samples remains. For instance, the usage of sound subsystems, like GStreamer [56], to record ambient audio introduces new delays.

Figure 1 illustrates this aspect in the frequency spectrum of two NTP-synchronised recordings.

As a solution, we had the decommiting node create 200 additional fingerprints by shifting the audio sequence in both directions in steps of 0.001 seconds. The device then tried to create a common key with each of these fingerprints and uses the first successful attempt. In this way, we could compensate for an error of about 0.2 seconds in the clock synchronisation among nodes.

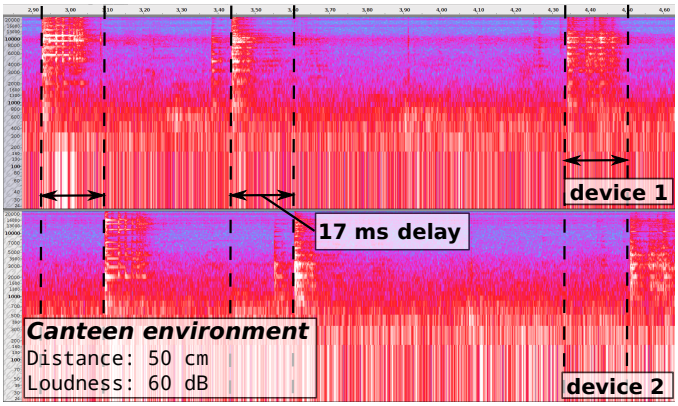


Fig. 1: Synchronisation offset of NTP synchronised audio recordings

3.5 Security Considerations and Attack Scenarios

Privacy leakages translate to leaking partial information about the used audio-fingerprints. This can simplify the attack when further details of the ambient audio of Alice and Bob is available.

Possible attacks on fuzzy-cryptography are reviewed by Scheirer et al. [57]. In particular, fuzzy commitment is evaluated regarding information leakage by Ignatenko et al. [58]. It was found that the scheme can leak information about the secret key. However, this is attributable to helper data, a bit sequence at random distance to the secret key, which is made public in traditional fuzzy commitment schemes. In our case, we do not utilise helper data and only optionally provide the hash of a data sequence with similar purpose.

The publicly available distance δ between f and c might, however leak information when either the fingerprints f, the code-sequences $c \in \mathcal{C}$ or the random word $a \in \mathcal{A}$ are not distributed uniformly at random or have insufficient entropy. Generally, it is important that

1) the random function to generate a has a sufficiently high entropy
2) the codewords $c \in \mathcal{C}$ are independently and uniformly distributed over all possible bit sequences of length n
3) The entropy of the generated fingerprints is high

We address these issues in the following.

1) The choice of $a \in \mathcal{A}$ has to be done by using a random source with sufficient entropy. In Linux-based systems `/dev/urandom` should provide enough entropy for using the output for cryptographic purposes [59]. For generating $\mathrm{h}(a)$ a one-way-function has to be chosen to make sure that no assumptions on a can be made based on $\mathrm{h}(a)$. We utilise SHA-512 which is certified by the NIST and was extensively evaluated [50].

2) We are using 512 bit fingerprints and the Reed-Solomon code $RS(2^{10}, 204, 512)$. Consequently, sets of words and codewords are defined as $\mathcal{A} = \mathbb{F}_{2^{10}}^{204}$ and $\mathcal{C} = \mathbb{F}_{2^{10}}^{512}$. A word a out of $2^{10^{204}} = 1024^{204}$ possible words is randomly chosen and encoded to c.

3) In order to test the entropy of generated fingerprints we applied the dieHarder [60] set of statistical tests. Generally, we could not find any bias in the fingerprints created from ambient audio. Section 6 discusses the test results in more detail.

A relevant attack scenario valid in our case is that the attacker is in the same audio context as Alice and Bob. In this case, no security is provided by the proposed protocol. Although this is a plausible threat, it can hardly be avoided that the leaking of contextual information poses a thread to a protocol that is designed to base the secure key generation exclusively on exactly this information. This principle is essential for the desired unobtrusive and ad-hoc operation. An overview over possible attack scenarios when the attacker is not inside the same context is listed below.

3.5.1 Brute force

The set of possible words $\mathcal{A}$ has to be large enough. It should be computationally infeasible to test every combination to get the used word a. The probability to guess the right a is 1024^{-204} in our implementation. Note that even with $u = 0.6$, this probability is still 1024^{-102}.

3.5.2 Denial-of-service (DoS)

An attacker could stress the communication while Alice and Bob are using the fuzzy pairing. The pairing would fail if $(\delta, \mathrm{h}(a))$ is not transmitted correctly. DoS preventions should be implemented to provide an accurate treatment. As part of these preventions a maximum number of attempts to pair two devices should be defined. Generally, this type of attack is only possible when $(\delta, \mathrm{h}(a))$ or δ is transmitted. As mentioned in section 3.3, with a careful choice of the fingerprint mechanism the exchange of data can be avoided.

3.5.3 Man-in-the-middle

An Eavesdropper Eve could be located in such a way, that she can intercept the wireless connection but is not located in the same physical context as Alice and Bob. When Eve intercepts the tuple $(\delta, \mathrm{h}(a))$, she must generate an audio-fingerprint $\overline{f}$ that is sufficiently close to the fingerprints f and f' of Alice and Bob to intercept successfully. With no knowledge on the audio context, a brute force attack is then required. This has to be done while Alice and Bob are currently in the phase of pairing. Therefore Eve is limited by a strict time frame. Again, this attack can be prevented by avoiding the transmission of $(\delta, \mathrm{h}(a))$ or δ.

3.5.4 Audio amplification

An Eavesdropper Eve could be located in physical proximity where the ambient audio used by Alice and Bob to generate their fingerprints is replicated. Eve can utilise a directional microphone to amplify these audio signals. In fact, this is a security threat which increases the chance that Eve can reconstruct the fingerprint partly to have

TABLE 1: Approximate mean loudness experienced for several sample classes at 1.5 m distance

	loud	median	quiet
Clap	40 dB	35 dB	25 dB
Music	35 dB	25 dB	15 dB
Snap	30 dB	25 dB	10 dB
Speak	25 dB	20 dB	15 dB
Whistle	45 dB	35 dB	25 dB

TABLE 2: Percentage of identical bits between fingerprints

	matching samples	non-matching samples
Median	0.7617	0.5332
Mean	0.7610	0.5322
Variance	0.0014	0.00068342
Min	0.6777	0.4414
Max	0.8750	0.6484

a greater probability of guessing the secure secret. Since our scheme inherently relies on contextual information we can not completely eliminate this threat. However, we show in section 5.2 that the acoustic properties in two rooms are at least sufficiently different to prevent a device with access to the dominant audio source to be successful in more than 50 % of all cases.

4 Fingerprint-based authentication

In a controlled environment we recorded several audio samples with two microphones placed at distinct positions in a laboratory. The samples were played back by a single audio source. Microphones were attached to the left and right ports of an audio card on a single computer with audio cables of equal lengths. They were placed at 1.5 m, 3 m, 4.5 m and 6 m distance to the audio source. For each setting, the two microphones were always located at non-equal distances. In several experiments, the audio source emitted the samples at quiet, medium and loud volume. The audio samples utilised consisted of several instances of music, a person clapping her hands, snapping her fingers, speaking and whistling. Dependent on the specific sample, the mean dB for these loudness levels varied slightly. The loudness levels for several sample classes experienced in 1.5 m distance are detailed in table 1.

For these samples recorded by both microphones we created audio-fingerprints and compared their Hamming distances pair-wise. We distinguish between fingerprints created for audio sampled simultaneously and non-simultaneously. Overall, 7500 distinct comparisons between fingerprints are conducted in various environmental settings. From these, 300 comparisons are created for simultaneously recorded samples.

Figure 2 depicts the median percentage of identical bits in the fingerprints for audio samples recorded simultaneously and non-simultaneously for several positions of the microphones and for several loudness levels. The error bars depict the variance in the Hamming distance.

First, we observe that the similarity in the fingerprints is significantly higher for simultaneously sampled audio in all cases. Also, notably, the similarity in the fingerprints of non-simultaneously recorded audio is slightly higher than 50 %, which we would expect for a random guess. The small deviation is a consequence of the monotonous electronic background noise originated by the recording devices consisting of the microphones and the audio chipsets.

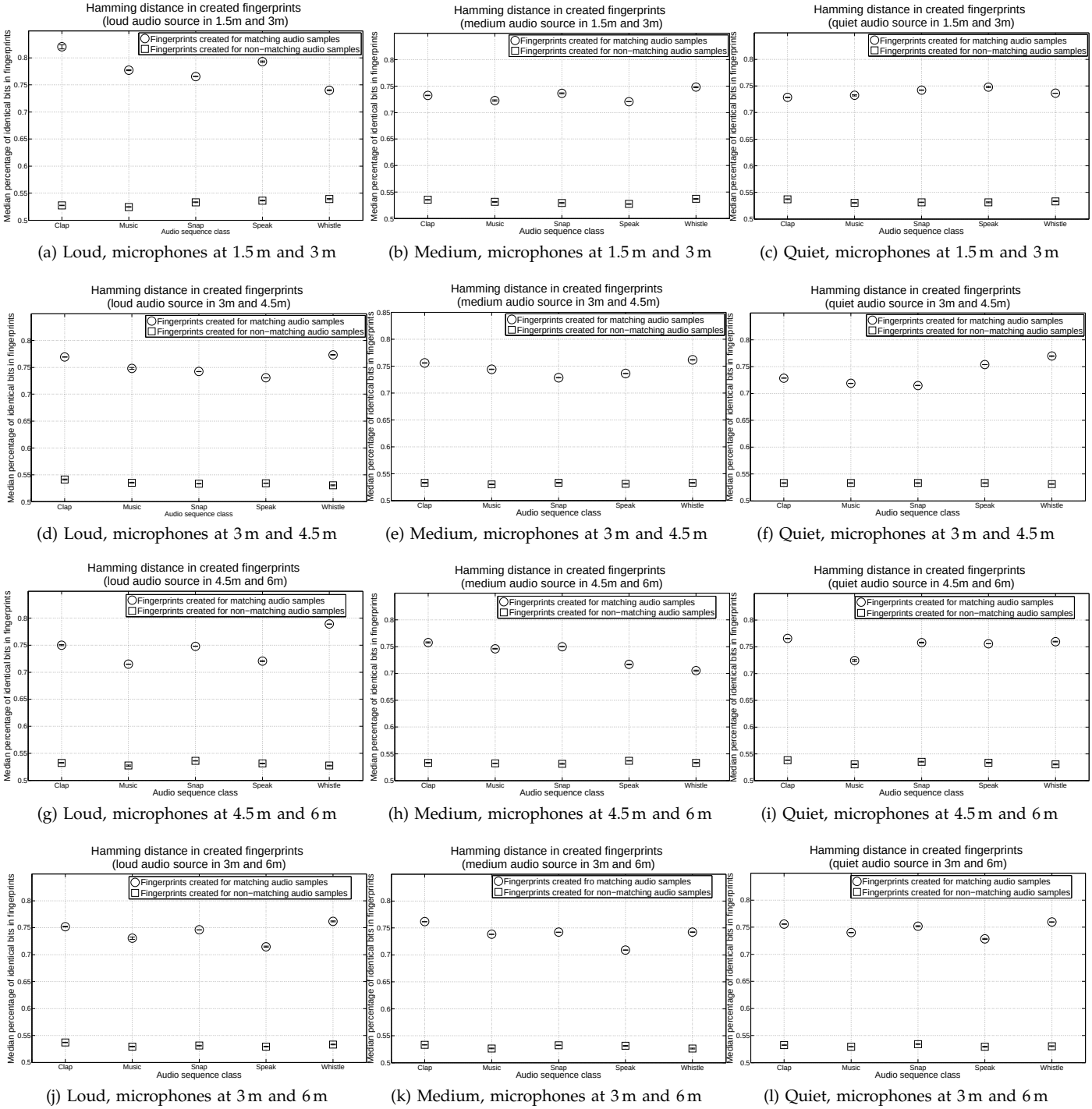
Additionally, the distance of the microphones to the audio source has no impact on the similarity of fingerprints. Similarly, we can not observe a significant effect of the loudness level. This confirms our expectation since for the fingerprinting approach not the absolute energy on frequency bands but changes in energy over time were considered (cf. section 3.1). Therefore, changes in the loudness level as, for instance, by altering the distance to the audio source or by changing the volume of the audio, have minor impact on the fingerprints.

Table 2 depicts the maximum and minimum Hamming distance among all experiments. We observe that one of the comparisons of fingerprints for non-simultaneously recorded audio yielded a maximum similarity of 0.6484. This value is still fairly separated from the minimum bit-similarity observed for fingerprints from simultaneously recorded samples. Also, this event is very seldom in the 7200 comparisons since the mean is sharply concentrated around the median with a low variance. Therefore, by repeating this process for a small number of times, we reduce the probability of such an event to a negligible value. For instance, only about $3.8\,\%$ of the comparisons between fingerprints from non-matching samples have a similarity of more than 0.58; only $0.4583\,\%$ have a similarity of more than 0.6. Similarly, only $2.33\,\%$ of the comparisons of synchronously sampled audio have a similarity of less than 0.7.

With these results, we conclude that an authentication based on audio-fingerprints created from synchronised audio samples in identical environmental contexts is feasible. However, since it is unlikely that the fingerprints match in all bits, it is not possible to utilise the audio-fingerprints directly as a secret key to establish a secure communication channel among devices. We therefore considered error correcting codes to account for the noise in the fingerprints created.

5 Case-studies

We implemented the described ambient audio-based secure communication scheme in Python and conducted case-studies in four distinct environments. The experiments feature differing loudness levels, different background noise figures as well as distinct common situations. In section 5.1, we observe how the proposed method can establish an ad-hoc secure communication



(a) Loud, microphones at 1.5 m and 3 m (b) Medium, microphones at 1.5 m and 3 m (c) Quiet, microphones at 1.5 m and 3 m

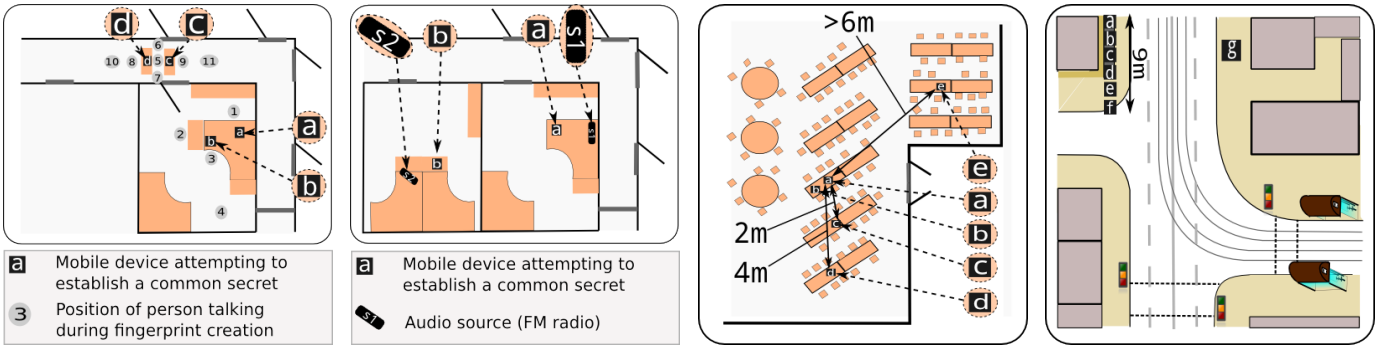
(d) Loud, microphones at 3 m and 4.5 m (e) Medium, microphones at 3 m and 4.5 m (f) Quiet, microphones at 3 m and 4.5 m

(g) Loud, microphones at 4.5 m and 6 m (h) Medium, microphones at 4.5 m and 6 m (i) Quiet, microphones at 4.5 m and 6 m

(j) Loud, microphones at 3 m and 6 m (k) Medium, microphones at 3 m and 6 m (l) Quiet, microphones at 3 m and 6 m

Fig. 2: Hamming distance observed for fingerprints created for recorded audio samples at distinct loudness levels and distances between microphones and the audio source

based on audio from ongoing discussions in a general office environment. Since an adversary able to sneak into the audio context of a given room might be better positioned to guess the secure key, we demonstrate in section 5.2 that even for an adversary device that is able to establish a similar dominant audio context in a different room by listening to the same FM-radio-channel, the gap in the created fingerprints is significant. In these two experiments, we utilised artificial audio sources in a sense that they were specifically placed to create the ambient audio context. In section 5.3 and section 5.4 we describe experiments in common environments where ambient audio was utilised exclusively. In section 5.3 we placed devices at distinct locations in a canteen and studied the success probability based on the distance between devices. In section 5.4 we study the feasibility



(a) Office setting. Devices and speakers located at distinct positions.

(b) Office setting. Synchronised dominant audio source established in both rooms via FM radios.

(c) Canteen setting. Devices located at various distances. No dominant noise source.

(d) Roadtraffic setting. Devices arranged alongside a road. No dominant noise source.

Fig. 3: Environmental settings of the case-studies conducted.

TABLE 3: Configuration of the four scenarios considered

Microphones (external)	
Impedance	$\leq 22\,\mathrm{k\Omega}$
Current consumption	$\leq 0.5\,\mathrm{mA}$
Frequency response	$100\,\mathrm{Hz} \sim 16\,\mathrm{KHz}$
Sensitivity	$-38\,\mathrm{dB} \pm 2\,\mathrm{dB}$
Microphones (internal)	
Device A	Intel G45 DEVIBX
Device B	Intel 82801I

of establishing a secure communication channel with road-traffic as background noise. Figure 3 summarises all settings considered. To capture audio we utilised the build-in microphones of the computers. The only exception is the reference scenario 3a in which simple off-the-shelf external microphones have been utilised. For both devices, the manufacturer and audio device types differed. Table 3 details further configuration of the scenarios conducted and the hardware utilised.

5.1 Office environment

In our first case-study, we position two laptops in an office environment. Ambient audio was originated from individuals speaking inside or outside of the office room. We conducted several sets of experiments with differing positions of laptop computers and audio sources as depicted in figure 3a. We distinguish four distinct scenarios

- $3a_1$ Both devices inside the office at locations a and b. 1-2 Individuals speaking at locations 1 to 4.
- $3a_2$ One device inside and one outside the office in front of the open office door at locations a and c. 1-2 Individuals speaking at locations 1 and 5.
- $3a_3$ Both devices in the corridor in front of the office at locations c and d. 1-2 Individuals speaking at locations 5 to 11.
- $3a_4$ One device inside and one outside the office in front of the closed office door at locations a and c. 1-2 Individuals speaking (damped but audible behind closed door) at locations 1 and 5.

In all cases the devices were synchronised over NTP. For each synchronisation, one device indicated at which point in time it would initiate audio recording. Both devices then sample ambient audio at that time and create a common key following the protocol described in section 3.1. For each scenario the key synchronisation process was repeated 10 times with the persons located at different locations. From these persons, either person 1, person 2 or both were talking during the synchronisation attempts in order to provide the audio context.

The settings $3a_1$ and $3a_3$ represent the situation of two friendly devices willing to establish a secure communication channel. The setting $3a_2$ could constitute the situation in which a person passing by is accidentally witnessing the communication and part of the audio context. In setting $3a_4$, the communication partners might have closed the office door intentionally in order to keep information secure from persons outside the office.

In scenario $3a_1$, where both devices share the same audio context a fraction of 0.9 of all synchronisation attempts have been successful. Also, for scenario $3a_3$, the fraction of successful synchronisation attempts was as high as 0.8. Consequently, when both devices are located in the same audio context, a successful synchronisation is possible with high probability.

For scenario $3a_2$, where the device outside the ajar door could partly witness the audio context, we had a success probability of 0.4. Although this means that less than every second approach was successful, this is clearly not acceptable in most cases. Still, this low success probability it is remarkable since the person speaking in the office or on the corridor was clearly audible at the respective other location.

In scenario $3a_4$, however, when the audio context was separated by the closed door, no synchronisation attempt was successful. Remarkably in this case, the person speaking was, although hardly comprehensible, still audible at the other side of the door.

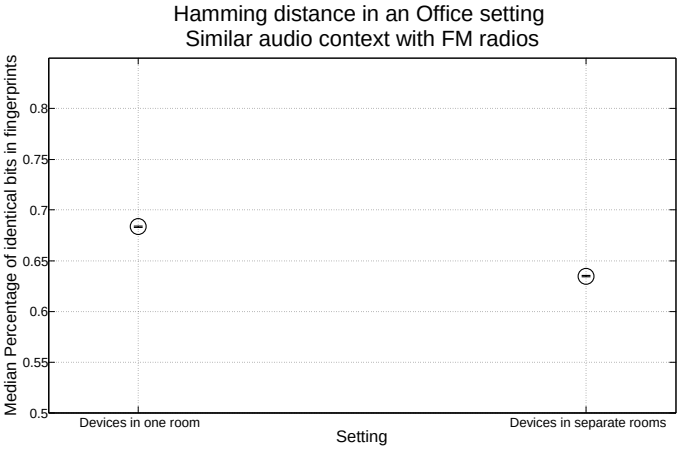


Fig. 4: Median percentage of bit errors in fingerprints generated by two mobile devices in an office setting. The audio context was dominated by an FM radio tuned to the same channel.

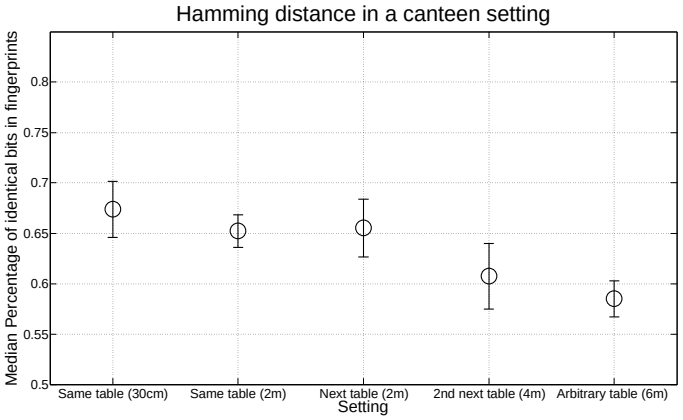


Fig. 5: Median percentage of bit errors in fingerprints generated by two mobile devices in a canteen environment.

Finally, we attempted to establish a synchronisation in the scenarios $3a_1$, $3a_2$ and $3a_3$ when only background noise was present. This means that no sound was emitted from a source located in the same location as one of the devices. Some distant voices and indistinguishable sounds could occasionally be observed. After a total of twelve tries in these three scenarios, not a single one resulted in a successful synchronisation between devices. We conclude that a dominant noise source or at least more dominant background noise needs to be present in the same physical context as the devices that want to establish a common key.

5.2 Context replication with FM-radio

A straightforward security attack for audio-based encryption could be for the attacker to extract information about the audio context and use this in order to guess the secret key created. We studied this threat by trying to generate a secret key between two devices in different rooms but with similar audio contexts. In particular, we placed two FM-radios, tuned to the same frequency in both rooms (cf. figure 3b).

The audio context was therefore dominated by the synchronised music and speech from the FM-radio channel. No other audio sources have been present in the rooms so that additional background noise was negligible. We conducted two experiments in which the devices were first located in the same room and then in different rooms with the same distance to the audio source. The loudness level of the audio source was tuned to about 50 dB in both rooms. Figure 4 depicts the median bit-similarity achieved when the devices were placed in the same room and in different rooms respectively.

We observe that in both cases the variance in the bit errors achieved is below $0.01\,\%$. When both devices are placed in the same room, the median Hamming distance between fingerprints is only $31.64\,\%$. We account this high similarity and the low variance to the fact that background noise was negligible in this setting since the FM-radio was the dominant audio source.

When the devices are placed in different rooms, the variance in bit error rates is still low with $0.008\,\%$. The median Hamming distance rose in this case to $36.52\,\%$.

Consequently, although the dominant audio source in both settings generated identical and synchronised content, the Hamming distance drops significantly when both devices are in an identical room. With sufficient tuning of the error correction method conditioned on the Hamming distance, an eavesdropper can be prevented from stealing the secret key even though information on the audio context might be leaking.

5.3 Canteen environment

We studied the accuracy of the approach in the canteen of the TU Braunschweig (cf. figure 3c). At different tables, laptop computers have been placed. For each configuration we conducted 10 attempts to establish a unique key based on the fingerprints. We conducted all experiments during 11:30 and 14:00 on a business day in a well populated canteen. The ambient noise in this experiment was approximately at 60 dB. Apart from the audible discussion on each table, background noise was characterised by occasional high pitches of clashing cutlery.

Figure 5 depicts the results achieved. The figure shows the median percentage of bit errors between the fingerprints generated by both devices.

We observe that generally the percentage of identical bits in the fingerprint decreases with increasing distance. With about 2 m distance the percentage of identical bits is still quite similar to the similarity achieved when devices are only 30 cm apart. This is also true when one of the devices is placed at the next table. However, with a distance of about 4 meters and above, the percentage of bit errors are well separated so that also the error correction could be tuned such that a generation of a unique key is not feasible at this a distance.

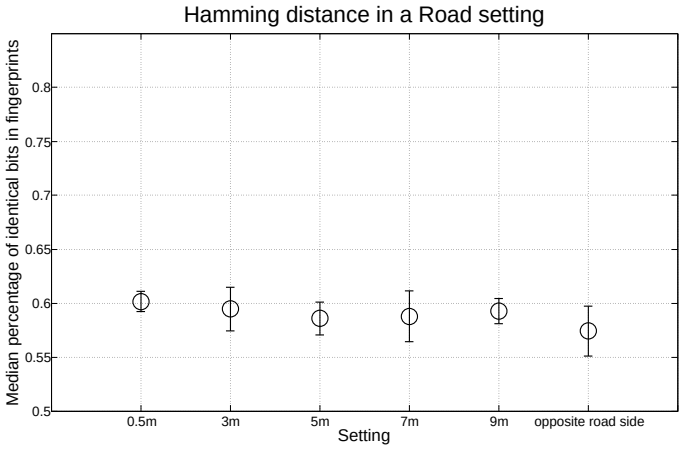


Fig. 6: Median percentage of bit errors in fingerprints from two mobile devices beside a heavily trafficked road.

5.4 Outdoor environment

In this instrumentation the two computers were located at the side of a well trafficked road. The study has been conducted during the rush hour between 17:00 and 19:00 at a regular working day. The road was frequented by pedestrians, bicycles, cars, lorries and trams. The data was measured not far off a headlight so that traffic occasionally stopped with running motors in front of the measurements. The loudness level was about 60 dB for both devices. The setting is depicted in figure 3d. We gradually increased the distance among devices. Devices have been placed with a distance between their microphones of 0.5 m, 3 m, 5 m, 7 m and 9 m at one side of the road. Additionally, for one experiment devices are placed at opposite sides of the road. For each configuration 10 to 13 experiments have been conducted. The results are depicted in figure 6. The figure depicts the median Hamming distance and variance for the respective configurations applied.

Not surprisingly, we observe that the Hamming distance between fingerprints generated by both devices is lowest when devices are placed next to each other. With increasing distance, the Hamming distance increases slightly but then stays similar also for greater distances.

At the opposite side of the road, however, the Hamming distance drops more significantly. When both devices are at the same side of the road, the probability to guess the secret key is high even for greater distances between the devices. We believe that this property is attributable to the very monotonic background noise generated by the vehicles on the road. The audio-context is therefore similar also in greater distances.

Only when one of the devices is located at the opposite side of the road, a more significant distinction between the generated fingerprints is possible. This may account to the different reflection of audio off surrounding buildings and to the fact that vehicles on the other lane generate a different dominant audio footprint.

Generally, these results suggest that audio-based key generation is hardly feasible in this scenario. Audio-based generation of secret keys is not well suited in an environment with very monotonic and unvaried background noise. Although a light protection from intruders on the different side of the road is possible, the radius in which similar fingerprints are generated on one side of the road is unacceptably high.

6 Entropy of Fingerprints

Although these results suggest that it is unlikely for a device in another audio context to generate a fingerprint which is sufficiently similar, an active adversary might analyse the structure of fingerprints created to identify and explore a possible weakness in the encryption key. Such a weakness might be constituted by repetitions of subsequences or by an unequal distribution of symbols. A message encrypted with a key biased in such a way may leak more information about the encrypted message than intended.

We estimated the entropy of audio-fingerprints generated for audio-sub-sequences by applying statistical tests on the distribution of bits. In particular, we utilised the dieHarder [60] set of statistical tests. This battery of tests calculates the p-value of a given random sequence with respect to several statistical tests. The p-value denotes the probability to obtain an input sequence by a truly random bit generator [61]. All tests are applied to a set of fingerprints of 480 bits length. We utilised all samples obtained in section 4 and section 5.

From 7490 statistical-test-batches consisting of 100 repeated applications of one specific test each, only 173, or about 2.31% resulted in a p-value of less than 0.05[1]. Each specific test was repeated at least 70 times. The p-values are calculated according to the statistical test of Kuiper [61], [62].

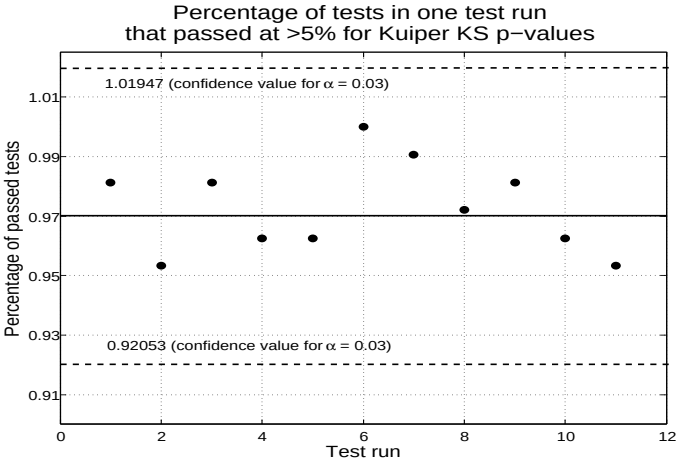
Figure 7 depicts for all test-series conducted the fraction of tests that did not pass a sequence of 100 consecutive runs at $> 5\%$ for Kuiper KS p-values [61] for all 107 distinct tests in the DieHarder battery of statistical tests. Generally, we observe that for all test-runs conducted, the number of tests that fail is within the confidence interval with a confidence value of $\alpha = 0.03$. The confidence interval was calculated for $m = 107$ tests as

$$1 - \alpha \pm 3 \cdot \sqrt{\frac{(1-\alpha)\cdot\alpha}{m}}. \tag{19}$$

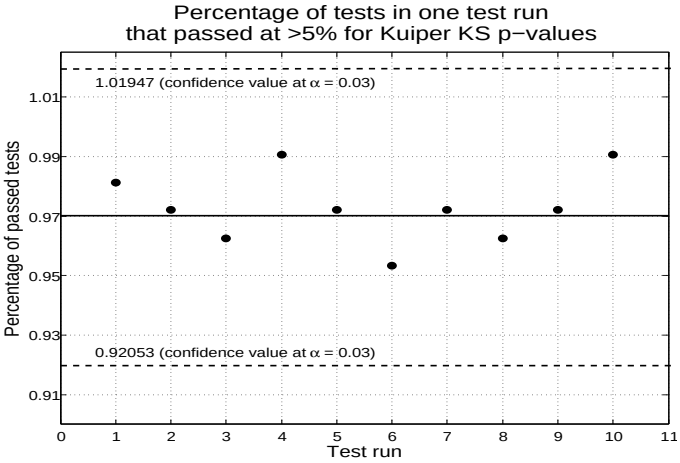
Alternatively, we could not observe any distinction between indoor and outdoor settings (cf. figure 7a and figure 7b) and conclude that also the increasing noise figure and different hardware utilised [2] does not impact the test results. Since music might represent a special case due to its structured properties and possible repetitions in an audio sequence, we considered it separately from the

1. All results are available at http://www.ibr.cs.tu-bs.de/users/sigg/StatisticalTests/TestsFingerprints_110601.tar.gz

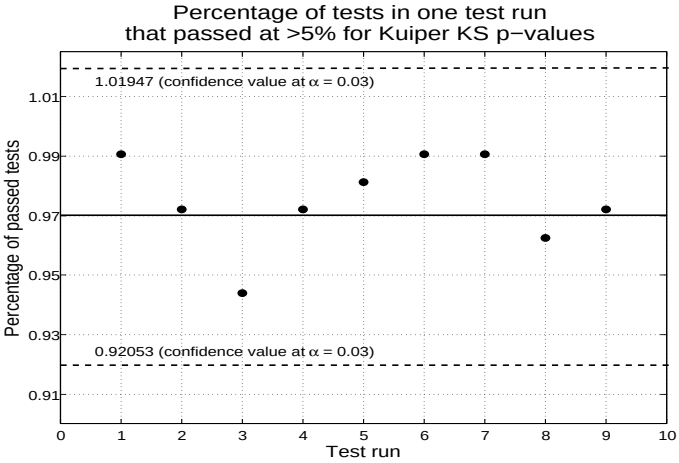
2. Overall, the microphones utilised (2 internal, 2 external) were produced by three distinct manufacturers



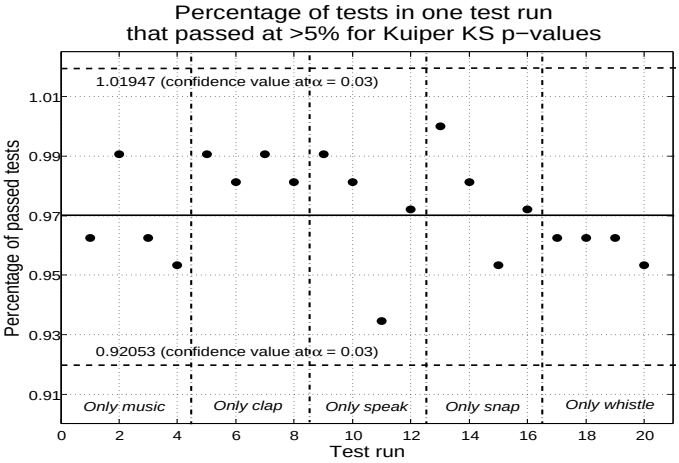
(a) Proportion of sequences from an indoor laboratory environment passing a test



(b) Proportion of sequences from various outdoor environments passing a test



(c) Proportion of sequences from all but music samples passing a test



(d) Proportion of sequences belonging to a specific audio class passing a test

Fig. 7: Illustration of P-Values obtained for audio-fingerprints by applying the DieHarder battery of statistical tests.

other samples. We could not identify a significant impact of music on the outcome of the test results (cf. figure 7c).

Additionally, we separated audio samples of one audio class and used them exclusively as input to the statistical tests. Again, there is no significant change for any of the classes (cf. figure 7d).

We conclude that we could not observe any bias in fingerprints based on ambient audio. Consequently, the entropy of fingerprints based on ambient audio can be considered as high. An adversary should gain no significant information from an encrypted message eavesdropped.

7 Conclusion

We have studied the feasibility to utilise contextual information to establish a secure communication channel among devices. The approach was exemplified for ambient audio and can be similarly applied to alternative features or context sources. The proposed fuzzy-cryptography scheme is adaptable in its noise tolerance through the parameters of the error correcting code utilised and the audio sample length.

In a laboratory environment, we utilised sets of recordings for five situations at three loudness levels and four relative positions of microphones and audio source. We derived in 7500 experiments the expected Hamming distance among audio-fingerprints. The fraction of identical bits is above 0.75 for fingerprints from the same audio context and below 0.55 otherwise. This gap in the Hamming distance can be exploited to generate a common secret among devices in the same audio context. We detailed a protocol utilising fuzzy-cryptography schemes that does not require the transmission of any information on the secure key. The common secret is instead conditioned on fingerprints from synchronised audio-recordings. The scheme enables ad-hoc and unobtrusive generation of a secure channel among devices in the same context. We conducted a set of common statistical tests and showed that the entropy of audio-fingerprints based on energy differences in adjacent frequency bands is high and sufficient to implement a cryptographic

scheme.

In four case-studies, we verified the feasibility of the protocol under realistic conditions. The greatest separation between fingerprints from identical and non-identical audio-contexts was observed indoor with low background noise and a single dominant audio source. In such an environment we could distinguish devices in the same and in different audio contexts. It was even possible to clearly identify a device that replicated dominant audio from another room with an equally tuned FM-radio at similar loudness level.

In a case-study conducted in a crowded canteen environment, we observed that the synchronisation quality was generally impaired due to the absence of a dominant audio source. However, it was still possible to establish a privacy area of about 2 m inside which the Hamming distance of fingerprints was distinguishably smaller than for greater distances. The worst results have been obtained in a setting conducted beside a heavily trafficked road. In this case, when the noise component becomes dominant and considerably louder, the synchronisation quality was further reduced. Additionally, due to the increased loudness level, a similar synchronisation quality was possible also at distances of about 9 m. We conclude that in this scenario, a secure communication channel based purely on ambient audio is hard to establish.

We claim that the synchronisation quality in scenarios with more dominant noise components can be further improved with improved features and fingerprint algorithms. Currently, most ideas are lent from fingerprinting algorithms and features designed to distinguish between music sequences. Although algorithms have been adapted to better capture characteristics of ambient audio, we believe that features and fingerprint generation to classify ambient audio might be further improved. Additionally, the consideration of additional contextual features such as light or RF-channel-based should improve the robustness of the presented approach.

In our implementation we faced difficulties to achieve sufficiently accurate (in the order of few milliseconds) time-synchronisation among wireless devices. In our current studies we tested several sample windows of NTP-synchronised recordings in order to achieve a feasible implementation on standard hardware. However, a more exact time synchronisation would further reduce the accuracy and computational complexity of the approach.

ACKNOWLEDGMENTS

This work was supported by a fellowship within the Postdoc-Programme of the German Academic Exchange Service (DAAD)

REFERENCES

[1] C. Dupuy and A. Torre, *Local Clsuters, trust, confidence and proximity*, ser. Clusters and Globalisation: The development of urban and regional economies. Edward Elgar, 2006, ch. 5, pp. 175–195.

[2] R. Mayrhofer and H. Gellersen, "Spontaneous mobile device authentication based on sensor data," *information security technical report*, vol. 13, no. 3, pp. 136–150, 2008.

[3] D. Bichler, G. Stromberg, M. Huemer, and M. Loew, "Key generation based on acceleration data of shaking processes," in *Proceedings of the 9th International Conference on Ubiquitous Computing*, J. Krumm, Ed., 2007.

[4] L. E. Holmquist, F. Mattern, B. Schiele, P. Schiele, P. Alahuhta, M. Beigl, and H. W. Gellersen, "Smart-its friends: A technique for users to easily establish connections between smart artefacts," in *Proceedings of the 3rd International Conference on Ubiquitous Computing*, 2001.

[5] A. Varshavsky, A. Scannell, A. LaMarca, and E. de Lara, "Amigo: Proximity-based authentication of mobile devices," *International Journal of Security and Networks*, 2009.

[6] H.-W. Gellersen, G. Kortuem, A. Schmidt, and M. Beigl, "Physical prototyping with smart-its," *IEEE Pervasive computing*, vol. 4, no. 1536-1268, pp. 10–18, 2004.

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

[8] R. Mayrhofer, "The Candidate Key Protocol for Generating Secret Shared Keys from Similar Sensor Data Streams," *Security and Privacy in Ad-hoc and Sensor Networks*, pp. 1–15, 2007.

[9] D. Bichler, G. Stromberg, and M. Huemer, "Innovative key generation approach to encrypt wireless communication in personal area networks," in *Proceedings of the 50th International Global Communications Conference*, 2007.

[10] J. Hershey, A. Hassan, and R. Yarlagadda, "Unconventional cryptographic keying variable management," *IEEE Transactions on Communications*, vol. 43, pp. 3–6, 1995.

[11] G. Smith, "A direct derivation of a single-antenna reciprocity relation for the time domain," *IEEE Transactions on Antennas and Propagation*, vol. 52, pp. 1568–1577, 2004.

[12] M. G. Madiseh, M. L. McGuire, S. S. Neville, L. Cai, and M. Horie, "Secret key generation and agreement in uwb communication channels," in *Proceedings of the 51st International Global Communications Conference (Globecom)*, 2008.

[13] S. T. B. Hamida, J.-B. Pierrot, and C. Castelluccia, "An adaptive quantization algorithm for secret key generation using radio channel measurements," in *Proceedings of the 3rd International Conference on New Technologies, Mobility and Security*, 2009.

[14] K. Kunze and P. Lukowicz, "Symbolic object localization through active sampling of acceleration and sound signatures," in *Proceedings of the 9th International Conference on Ubiquitous Computing*, 2007.

[15] P. Tuyls, B. Skoric, and T. Kevenaar, *Security with Noisy Data*. Springer-Verlag, 2007.

[16] Q. Li and E.-C. Chang, "Robust, short and sensitive authentication tags using secure sketch," in *Proceedings of the 8th workshop on Multimedia and security*. ACM, 2006, pp. 56–61.

[17] Y. Dodis, R. Ostrovsky, L. Reyzin, and A. Smith, "Fuzzy extractors: How to generate strong keys from biometrics and other noisy data," *EUROCRYPT 2004*, pp. 79–100, 2004.

[18] F. Miao, L. Jiang, Y. Li, and Y.-T. Zhang, "Biometrics based novel key distribution solution for body sensor networks," in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*. IEEE, 2009, pp. 2458–2461.

[19] A. Juels and M. Sudan, "A Fuzzy Vault Scheme," *Proceedings of IEEE Internation Symposium on Information Theory*, p. 408, 2002.

[20] Y. Dodis, J. Katz, L. Reyzin, and A. Smith, "Robust fuzzy extractors and authenticated key agreement from close secrets," *Advances in Cryptology-CRYPTO 2006*, pp. 232–250, 2006.

[21] P. Cano, E. Batlle, T. Kalker, and J. Haitsma, "A Review of Algorithms for Audio Fingerprinting," *The Journal of VLSI Signal Processing*, vol. 41, no. 3, pp. 271–284, 2005.

[22] S. Baluja and M. Covell, "Waveprint: Efficient wavelet-based audio fingerprinting," *Pattern Recognition*, vol. 41, no. 11, 2008.

[23] L. Ghouti and A. Bouridane, "A robust perceptual audio hashing using balanced multiwavelets," in *Proceedings of the 5th IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2006.

[24] S. Sukittanon and L. Atlas, "Modulation frequency features for audio fingerprinting," in *Proceedings of the 2nd IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2002.

[25] J. Haitsma and T. Kalker, "A highly robust audio fingerprinting system," in *Proceedings of the 3rd International Conference on Music Information Retrieval*, October 2002.
[26] C. Burges, D. Plastina, J. Platt, E. Renshaw, and H. Malvar, "Using audio fingerprinting for duplicate detection and thumbnail generation," in *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on*, vol. 3, march 2005, pp. iii/9–iii12 Vol. 3.
[27] C. Bellettini and G. Mazzini, "A framework for robust audio fingerprinting," *Journal of Communications*, vol. 5, no. 5, 2010.
[28] A. Ghias, J. Logan, D. Chamberlin, and B. C. Smith, "Query by humming," in *Proceedings of the ACM Multimedia*, 1995.
[29] D. Parsons, *The directory of tunes and musical themes*. Cambridge University press, 1975.
[30] R. A. Baeza-Yates and C. H. perleberg perleberg perleberg perleberg, "Fast and practical approximate string matching," *Third annual symposium on combinatorial pattern matching*, 1992.
[31] R. J. McNab, L. A. Smith, I. H. Witten, C. L. Henderson, and S. J. Cunningham, "Towards the digital music library: tune retrieval from acoustic iinput," *Proceedings of the ACM*, 1996.
[32] L. Prechelt and R. Typke, "An interface for melody input," *ACM Transactions on Computer Human Interactions*, vol. 8, 2001.
[33] W. Chai and B. Vercoe, "Melody retrieval on the web," in *Proceedings of the ACM/SPIE conference on Multimedia Computing and Networking*, 2002.
[34] L. Shifrin, B. Pardo, and W. Birmingham, "Hmm-based musical query retrieval," in *Proceedings of the joint conference on digital libraries*, 2002.
[35] L. Rabiner, "A tutorial on hidden markov models and selected accplications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, 1989.
[36] Y. Zhu and D. Shasha, "Warping indexes with envelope transforms for query by humming," in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, 2003.
[37] J. Haitsma and T. Kalker, "Robust audio hashing for content identification," in *In Content-Based Multimedia Indexing (CBMI*, 2001.
[38] J. Lebossé, L. Brun, and J.-C. Pailles, "A robust audio fingerprint's based identification method," in *Proceedings of the 3rd Iberian conference on Pattern Recognition and Image Analysis, Part I*, ser. IbPRIA '07. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 185–192.
[39] C. Burges, J. Platt, and S. Jana, "Distortion discriminant analysis for audio fingerprinting," *Speech and Audio Processing, IEEE Transactions on*, vol. 11, no. 3, pp. 165–174, may 2003.
[40] J. Herre, E. Allamanche, and O. Hellmuth, "Robust matching of audio signals using spectral flatness features," in *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*, 2001, pp. 127–130.
[41] C. Yang, "Macs: music audio characteristic sequence indexing for similarity retrieval," in *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*, 2001, pp. 123–126.
[42] ——, "Efficient acoustic index for music retrieval with various degrees of similarity," in *Proceedings of the tenth ACM international conference on Multimedia*, ser. MULTIMEDIA '02. New York, NY, USA: ACM, 2002, pp. 584–591. [Online]. Available: http://doi.acm.org/10.1145/641007.641125
[43] A. Wang, "The Shazam music recognition service," *Communications of the ACM*, vol. 49, no. 8, p. 48, 2006.
[44] ——, "An Industrial Strength Audio Search Algorithm," in *International Conference on Music Information Retrieval (ISMIR)*, 2003.
[45] F. Hao, "On using fuzzy data in security mechanisms," Ph.D. dissertation, Queens College, Cambridge, April 2007.
[46] A. C. Ibarrola and E. Chavez, "A robust entropy-based audio-fingerprint," in *Proceedings of the 2006 International Conference on Multimedia and Expo (ICME 2006)*, 2006.
[47] B. Schneider, *Applied Cryptography: Protocols, Algorithms, and Source Code in C*, 2nd ed. John Wiley and Sons, Inc., 1996.
[48] I. Reed and G. Solomon, "Polynomial codes over certain finite fields," *Journal of the Society for Industrial and Applied Mathematics*, pp. 300–304, 1960.
[49] A. Juels and M. Wattenberg, "A Fuzzy Commitment Scheme," *Sixth ACM Conference on Computer and Communications Security*, pp. 28–36, 1999.
[50] National Institute of Standards and Technology, "180-3, Secure Hash Standard (SHS)," *Federal Information Processing Standards Publications (FIPS PUBS)*, Oct. 2008.
[51] ——, "197, Advanced Encryption Standard (AES)," *Federal Information Processing Standards Publications (FIPS PUBS)*, 2001.
[52] D. Mills, J. Martin, J. Burbank, and W. Kasch, "Network Time Protocol Version 4: Protocol and Algorithms Specification," RFC 5905 (Proposed Standard), Internet Engineering Task Force, Jun. 2010. [Online]. Available: http://www.ietf.org/rfc/rfc5905.txt
[53] D. L. Mills, "Improved algorithms for synchronising computer network clocks," *IEEE/ACM Transactions on Networking*, vol. 3, no. 3, June 1995.
[54] S. Meier, H. Weibel, and K. Weber, "Ieee 1588 syntonization and synchronization functions completely realized in hardware," in *International IEEE Symposium on Precision Clock Synchronization for Measurement, Control and Communication (ISPCS 2008)*, 2008.
[55] D. L. Mills, "Precision synchronisation of computer network clocks," *ACM Computer Communication Review*, vol. 24, no. 2, April 1994.
[56] *GStreamer Documentation*, Freedesktop.org, Oct. 2010. [Online]. Available: http://gstreamer.freedesktop.org/documentation/
[57] W. J. Scheirer and T. E. Boult, "Cracking fuzzy vaults and biometric encryption," in *Proceedings of Biometrics Symposium, Baltimore, USA*, 2007.
[58] T. Ignatenko and F. M. J. Willems, "Information Leakage in Fuzzy Commitment Schemes," in *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 2, Jun. 2010, p. 337.
[59] K. Fenzi and D. Wreski, *Linux Security HOWTO*, Jan. 2004. [Online]. Available: http://www.ibiblio.org/pub/linux/docs/howto/other-formats/pdf/Security-HOWTO.pdf
[60] R. G. Brown, "Dieharder: A random number test suite," http://www.phy.duke.edu/∼rgb/General/dieharder.php, 2011.
[61] N. Kuiper, "Tests concerning random points on a circle," in *Proceedings of the Koinklijke Nederlandse Akademie van Wetenschappen*, vol. Series a 63, 1962, pp. 38–47.
[62] M. Stephens, "The goodness-of-fit statistic v_n: Distribution and significance points," *Biometrika*, vol. 52, 1965.

Dominik Schürmann received his bachelor of science from the TU Braunschweig, Germany in 2010. His research interests include unobtrusive security in distributed systems and cryptographic algorithms in general.

Stephan Sigg received his diploma in computer sciences from the University of Dortmund, Germany in 2004 and finished his PhD in 2008 at the chair for communication technology at the University of Kassel, Germany. He currently works in the Information Systems Architecture Science Research Division at the National Institute of Informatics (NII), Japan. His research interests include the analysis, development and optimisation of algorithms for Pervasive Computing Systems.