Some aspects of physical prototyping in Pervasive Computing

Distributed adaptive beamforming, Device-free recognition of activities from RF, Secure keys from ambient audio and calculation of mathematical functions on the wireless channel

Habilitationsschrift zur Erlangung der
Lehrbefugnis im Fach Informatik an der
Karl-Friedrich-Gauss-Fakultät der
Technischen Universität Carolo
Wilhelmina zu Braunschweig

Vorgelegt von: Dr. rer. nat. Stephan Sigg
sigg@ibr.cs.tu-bs.de
Institut für Betriebssysteme und Rechnerverbund
Technische Universität Braunschweig

Braunschweig im Februar 2015

Habilitation: 17.03.2017

Contents

1 Introduction

This document summarises the results of several research campaigns over the past seven years. The main connecting theme is the physical layer of widely deployed sensors in Pervasive Computing domains. In particular, we have focused on the RF-channel or on ambient audio. Instead of plugging together existing technologies to solve a particular task, we have been re-prototyping the use and interaction via these interfaces for a particular purpose. In particular, the initial problem from which we started this work was that of distributed adaptive transmit beamforming. We have been looking for a simple method to align the phases of jointly transmitting nodes (e.g. sensor or IoT nodes). The algorithmic solution to this problem was to implement a distributed random optimisation method on the participating nodes in which the transmitters and the receiver follow an iterative question-and-answer scheme. In this scheme, phases of transmitters are randomly altered. The algorithm works on the physical layer, not utilising existing protocols. We have been able to derive sharp asymptotic bounds on the expected optimisation time of an evolutionary random optimiser and presented an asymptotically optimal approach (cf. section 2.1)[1]. The latter approach, however, requires richer feedback from the transmit devices which restricts its application. Given the strong unimodality of the underlying search space, we then derived improved sharp bounds on a local random search approach (cf. section 2.2)[2].

One thing that we have learned from the work on these physical layer algorithms was that the signals we work on are fragile and perceptive to physical environmental changes. These could be obstacles such as furniture, opened or closed windows or doors as well as movement of individuals.

This observation motivated us to view the wireless interface as a sensor for environmental changes in Pervasive Computing environments. Pioneering this field of device-free recognition of activities and situations, we have demonstrated the feasibility of this sensing paradigm with software radios and also sensor nodes (cf. section 2.3)[3]. The essential novelty of this sensing paradigm enabled by looking at the physical layer directly is that monitored entities do not need to be equipped with any hardware or with any part of the sensing system. By reflecting and blocking RF-signals, the monitored entities are implicitly integral parts of the sensing system. Improving the recognition accuracy of these systems, we could show that an accurate recognition of activities is also possible utilising only ambient signals (i.e. not controlling the transmitter) (cf. section 2.4)[4]. In particular, ambient FM radio signals have been utilised. Finally, we could also demonstrate that a (lower accuracy) recognition is also possible on consumer devices, such as smartphones and that software radios are not mandatory for this recognition scheme. In this work, gestures and activities have been distinguished by analysing the fluctuation in the received signal

strength indicator (RSSI) of received IEEE 802.11 packets (cf. section 2.5)[5].

Another use of physical layer RF-signals is for security applications (e.g. [6]). The essential idea is that the signal fluctuations at two distinct physical locations are uncorrelated given that the devices are separated by at least half the wavelength of the signal. Then, close devices can use their correlated signal for the generation of common secure keys whereas devices which are farther apart are not able to generate identical keys following the same protocol since their input to generate the keys is uncorrelated. The security of this scheme relies on the difficulty to predict a channel at a particular remote physical location.

However, due to the high frequency of RF-signals, devices can be separated by few centimeters at most in order to generate common secure keys by this approach.

Instead, we exploit ambient audio which shares a number of properties with RF but operates at a lower frequency, so that higher physical separation of devices is acceptable. In particular, we presented a scheme for the generation of secure cryptographic keys from ambient audio (cf. section 2.6) [7]. The approach has been exploited in various environmental conditions and we have not been able to find BIAS using statistical tests. Later, we ported this approach towards common smartphones (cf. section 2.7) [8]. In this, hardware inconsistencies and insufficient synchronisation on common smartphone platforms had to be solved algorithmically.

This collection of applications demonstrates the potential of physical prototyping of Pervasive Computing applications. Existing protocols simplify communication among and with devices and the interaction with common interfaces. However, such protocols also introduce overhead and abstract from available information. While this enables a wide and easy application of the protocols, some information is only available on the physical level and some level of efficiency only possible by re-prototyping the physical layer.

We are currently working to further push these mentioned directions and novel fields of physical prototyping as detailed, for instance, in [9, 10, 11]. In particular, the calculation of mathematical operations on the wireless channel at the time of transmission appears to contain good potential for gains in efficiency for communication and computation in Pervasive Computing domains.

1.1 Contribution

This thesis presents the work of seven publications at international Journals or Conferences between 2010 and 2014. In particular, these are

[1] Stephan Sigg, Rayan Merched El Masri and Michael Beigl: Feedback based closed-loop carrier synchronisation: A sharp asymptotic bound, an asymptotically optimal approach, simulations and experiments, in IEEE Transactions on Mobile Computing (TMC), 2011
(DOI: http://dx.doi.org/10.1109%2FTMC.2011.21)

[2] Stephan Sigg: A fast binary feedback-based distributed adaptive carrier synchronisation for transmission among clusters of disconnected IoT nodes in smart spaces, Elsevier Journal on Ad Hoc Networks, vol. 16, May 2014, pp. 120-130
(DOI: http://10.1016/j.adhoc.2013.12.006)

[3] Stephan Sigg, Markus Scholz, Shuyu Shi, Yusheng Ji and Michael Beigl: RF-sensing of activities from non-cooperative subjects in device-free recognition systems using ambient and local signals, in IEEE Transactions on Mobile Computing (TMC), Feb. 2013, vol. 13, no. 4
(DOI: http://doi.ieeecomputersociety.org/10.1109/TMC.2013.28)

[4] Shuyu Shi, Stephan Sigg, Wei Zhao, and Yusheng Ji: Monitoring of Attention from Ambient FM-radio Signals, IEEE Pervasive Computing, Los Alamitos, CA, USA, IEEE Computer Society, Jan-Mar 2014, vol. 13, no. 1, pp. 30-36, 2014
(DOI: http://dx.doi.org/10.1109/MPRV.2014.13)

[5] Stephan Sigg, Ulf Blanke and Gerhard Troester: The Telepathic Phone: Frictionless Activity Recognition from WiFi-RSSI, IEEE International Conference on Pervasive Computing and Communications (PerCom), Budapest, Hungary, March 24-28, 2014
(DOI: http://dx.doi.org/10.1109/PerCom.2014.6813955)

[7] Dominik Schuermann and Stephan Sigg: Secure communication based on ambient audio, in IEEE Transactions on Mobile Computing (TMC), Feb. 2013, vol. 12 no. 2
(DOI: http://doi.ieeecomputersociety.org/10.1109/TMC.2011.271)

[8] Ngu Nguyen, Stephan Sigg, An Huynh and Yusheng Ji: Pattern-based Alignment of Audio Data for Ad-hoc Secure Device Pairing, in 2012 16th International Symposium on Wearable Computers (ISWC), pp.88-91, 18-22 June 2012
(DOI: http://dx.doi.org/10.1109/ISWC.2012.14)

2 Original work

2.1 Feedback based closed-loop carrier synchronisation: A sharp asymptotic bound, an asymptotically optimal approach, simulations and experiments [1]

We derive an asymptotically sharp bound on the synchronisation speed of a randomised black box optimisation technique for closed-loop feedback based distributed adaptive beamforming in wireless sensor networks. We also show that the feedback function that guides this synchronisation process is weak multimodal. Given this knowledge that no local optimum exists, we consider an approach to locally compute the phase offset of each individual carrier signal. With this design objective an asymptotically optimal algorithm is derived. Additionally, we discuss the concept to reduce the optimisation time and energy consumption by hierarchically clustering the network into subsets of nodes that achieve beamforming successively over all clusters. For the approaches discussed we demonstrate their practical feasibility in simulations and experiments.

2.1.1 Introduction

In recent years, sensor nodes of extreme tiny size have been envisioned [12, 13, 14]. In [15], for example, applications for square-millimetre sized nodes that seamlessly integrate into an environment are detailed. At these small form-factors transmission power of wireless nodes is restricted to several microwatts. Communication between a single node and a remote receiver is then only feasible at short distances. It is possible, however, to increase the maximum transmission range by cooperatively transmitting information from distinct nodes of a network [16, 17]. Cooperation can increase the capacity and robustness of a network of transmitters [18, 19] and decreases the average energy consumption per node [20, 21, 22].

Related research branches are cooperative transmission [23], collaborative transmission [24, 25], distributed adaptive beamforming [26, 27, 28, 29], collaborative beamforming [30] or cooperative/virtual MIMO for wireless sensor networks [31, 32, 33, 34]. One approach is

---

to utilise neighbouring nodes as relays [35, 36, 37] as proposed by Cover and El Gamal in [38]. Cooperative transmission is then achieved by multi-hop [39, 40, 41] or data flooding [42, 43, 44, 45] approaches. The general idea of multi-hop relaying based on the physical channel is to retransmit received messages by a relay node so that the destination will receive not only the message from the source node but also from the relay. In data flooding approaches, a node will retransmit a received message at its reception. It has been shown that the approach outperforms non-cooperative multi-hop schemes significantly. In particular, the transmission time is reduced compared to traditional transmission protocols [46].

In these approaches, nodes are not tightly synchronised and transmission may be asynchronous. Synchronous transmission, however, is achieved by virtual MIMO techniques. In these implementations, identical RF carrier signal components from various transmitters that function as a distributed beamformer are superimposed. When the relative phase offset of these carrier signal components at a remote receiver is small, the signal strength of the received sum signal is improved. In virtual MIMO for wireless sensor networks, single antenna nodes are cooperating to establish a multiple antenna wireless sensor network [32, 31, 33]. Virtual MIMO has capabilities to adjust to different frequencies and is highly energy efficient [34, 22]. However, the implementation of MIMO capabilities in WSNs requires accurate time synchronisation, complex transceiver circuits and signal processing that might surpass the power consumption and processing capabilities of simple sensor nodes.

Other solutions proposed are open-loop synchronisation methods such as round-trip synchronisation [47, 48, 49]. In this scheme, the destination transmits beacons in opposed directions along a multi-hop circle in which each of the nodes appends its part of the overall message to the beacons. Beamforming is achieved when the processing time along the multi-hop chain is identical in both directions. This approach, however, does not scale with the size of a network.

Closed loop feedback based approaches include full-feedback techniques, in which carrier synchronisation is achieved in a master-slave manner. The phase-offset among the carrier signals of destination nodes is corrected by a receiver node. Diversity between RF-transmit signal components is achieved over CDMA channels [50]. This approach is applicable only to small network sizes and requires sophisticated processing capabilities at the source nodes.

A more simple and less resource demanding implementation is the one-bit feedback based closed-loop synchronisation considered in [50, 51]. The authors describe an iterative process in which n source nodes $i \in [1, \ldots, n]$ randomly adapt the phases γ_i of their carrier signal $\Re\left(m(t)e^{j(2\pi(f_c+f_i)t+\gamma_i)}\right)$. Here, $m(t)$ is the transmit message and f_i denotes the frequency offset of node i to a common carrier frequency f_c. Initially, i.i.d. phase offsets γ_i of carrier signals are assumed. When a receiver requests a transmission from the network, carrier phases are synchronised in an iterative process.

1. Each source node i adjusts its carrier phase offset γ_i and frequency offset f_i randomly.

2. The source nodes transmit to the destination simultaneously as a distributed beamformer.

10
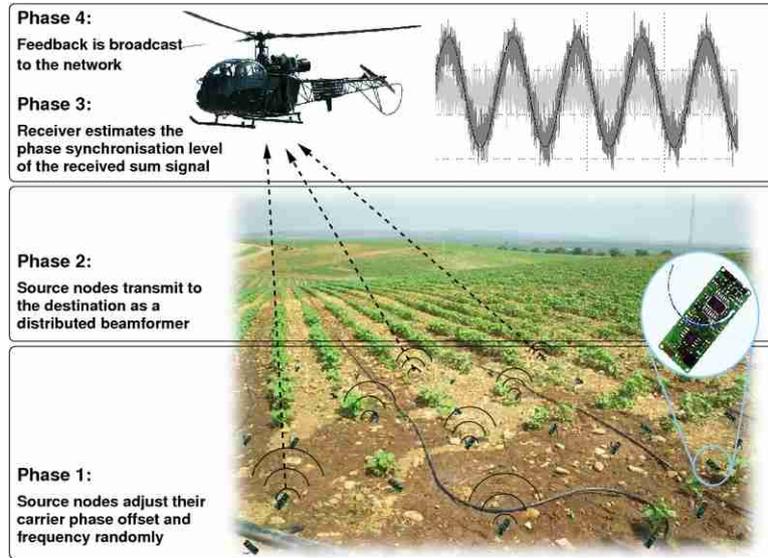
Figure 2.1: Schematic illustration of feedback based distributed adaptive beamforming in wireless sensor networks (©2011 IEEE Published by the IEEE CS, CASS, ComSoc, IES, SPS)

3. The receiver estimates the level of phase synchronisation of the received sum signal (for instance by the SNR).

4. This value is broadcast as a feedback to the network. Nodes interpret this feedback and adapt the phase of their carrier signal accordingly.

These four steps are iterated repeatedly until a stop criterion is met (e.g. maximum iteration count or sufficient synchronisation). Fig. 2.1 illustrates this procedure. It has been studied by different authors [52, 53, 54, 24]. The distinct approaches proposed differ in the implementation of the first and the fourth step specified above. The authors of [54] show that it is possible to reduce the count of transmitters in a random process and still achieve sufficient synchronisation among all nodes.

In [52, 53, 54] a process is described in which each node alters its carrier phase offset γ_i according to a normal distribution with small variance in step one. In [24] a uniform distribution is utilised instead but the probability for one node to alter the phase offest of its carrier signal is low. We show in section 2.1.5 that both approaches achieve a similar performance. Only in [53] not only the phase but also frequency is adapted.

Significant differences among these approaches also apply to the feedback and the reactions of nodes in step four. In [26, 53, 54] a one-bit feedback is utilised. Nodes sustain their phase modifications when the feedback has improved and otherwise reverse them. In [54] it was shown that the optimisation time is improved by a factor of two when a node as response to a negative feedback from the receiver applies a complementary phase offset instead of simply reversing its modification. In [24], authors suppose to utilise more than one bit as feedback so that parameters of the optimisation can be adapted with regard to the optimisation progress.

The strength of feedback based closed-loop distributed adaptive beamforming in wireless sensor networks is its simplicity and low processing requirements that make it feasible for the application in networks of tiny sized, low power and computationally restricted sensor nodes.

We study aspects of this transmission scheme and derive sharp asymptotic lower and upper bounds on the expected optimisation time of a common implementation in section 2.1.2. Together with these bounds we show that the feedback function is weak multimodal so that no local optimum exists. By small modifications of the common algorithm, however, further improvements in the synchronisation time can be achieved. In section 2.1.3 we discuss a hierarchical clustering scheme that exploits that the superimposed signal strength of a set of nodes is increased at a slower pace than the synchronisation time with increasing node count. When additional information available at a receiver node is utilised, further improvements are possible. In section 2.1.4 we show that by providing more than one bit as feedback to the transmitters, knowledge about the feedback function can be derived from measurements of a single node altering the phase offset of its carrier signal. We present an asymptotically optimal algorithm that utilises this knowledge and significantly improves the synchronisation process. In section 2.1.5, algorithms are compared for their synchronisation performance in numerical simulations. In these simulations, the impact of various environmental settings and algorithmic configurations can be approximated. Finally, in section 2.1.6 we demonstrate the feasibility of distributed adaptive beamforming in wireless sensor networks in a near-realistic instrumentation with software radios. Section 2.1.7 draws our conclusion.

2.1.2 Synchronisation time analysis

We analyse the process of distributed adaptive beamforming in wireless sensor networks as described in section 2.1.1 and assume that each one of the n nodes decides with probability $\frac{1}{n}$ to change the phase of its carrier signal uniformly at random in the interval $[0, 2\pi]$. On obtaining the feedback of the receiver, nodes that recently updated the phase of their carrier signal either sustain this decision or reverse it, depending on whether the feedback has improved or not. A feedback function $\mathcal{F} : \zeta^*_{\text{sum}} \to \mathbb{R}$ maps the superimposed received carrier signal

$$\zeta_{\text{sum}} = \Re\left(m(t)e^{j2\pi f_c t} \sum_{i=1}^{n} \text{RSS}_i e^{j(\gamma_i + \phi_i + \psi_i)} \right) \tag{2.1}$$

to a real-valued feedback score. In equation (2.1) the RSS_i denotes the received signal strength of the i-th signal out of n received signal components. As local oscillators are not synchronised and nodes are spatially distributed, ϕ_i and ψ_i account for the phase offset in the received signal components due to the offset in the local oscillators and due to distinct signal propagation times. A possible feedback function that is proportional to the distance between an observed superimposed carrier ζ_{sum} and an optimum sum carrier signal

$$\zeta_{\text{opt}} = \Re\left(m(t)\text{RSS}_{\text{opt}} e^{j(2\pi f_c t + \gamma_{\text{opt}})} \right) \tag{2.2}$$

is $\mathcal{F}(\zeta_{\text{sum}}) = \int_{t=0}^{2\pi} |\zeta_{\text{sum}} - \zeta_{\text{opt}}|$. Since this function can be mapped onto other feedback measures as, for instance, the signal to noise ratio (SNR) or the received signal strength (RSS), the following discussion remains valid for these feedback measures. While the multimodality of this feedback function is straightforward, we derive in Appendix A that it is also weak multimodal so that no local optima exist.

Distributed adaptive beamforming in wireless sensor networks is a search problem. The search space $\mathcal{S}$ is given by the set of possible combinations of phase and frequency offsets γ_i and f_i for all n carrier signals. A global optimum is a configuration of individual carrier phases that result in identical phase and frequency offset of all received direct signal components. For the analysis, we assume that the optimisation aim is to achieve for an arbitrary k a maximum relative phase offset of $\frac{4\pi}{k}$ between any two carrier signals. This means that we can control the quality of the synchronisation achieved by the variable k. An optimum is then reached when the phases of all carrier signal components of a receiver are within an interval of $\frac{4\pi}{k}$ in the phase space. When k is increased this directly translates to an improved phase synchronisation among signal components. Naturally, we can expect that the accuracy of the synchronisation also impacts the synchronisation time.

For our analysis we logically divide the phase space for a single carrier signal into k intervals of width $\frac{2\pi}{k}$. Observe that this is half of the interval that was used to define the optimum synchronisation. Consequently, when the achieved phase offset of each received signal component is within a maximum distance of $\frac{2\pi}{k}$ to the optimum phase offset, all received carrier phase signals are within an interval of $\frac{4\pi}{k}$ in the phase space and the optimum is reached.

For a specific superimposed carrier signal ζ at a receiver we represent the corresponding search point $s_\zeta = (\Gamma_t, F_t)_\zeta \in \mathcal{S}$ at iteration t by a specific combination of phase and frequency offsets with $\Gamma_t = (\gamma_{t,1}, \dots, \gamma_{t,n})$ and $F_t = (f_{t,1}, \dots, f_{t,n})$. In order to respect neighbourhood similarities we represent search points as Gray encoded binary strings $s_\zeta \in \mathbb{B}^{n \cdot \log(k)}$ so that similar points have a small Hamming distance [55]. A search point is then composed from n sections of $\log(k)$ bits each. Every block of length $\log(k)$ describes one of the k intervals for the phase offset of one carrier signal. For the analysis, we assume that the frequency offset f_i is zero for all carriers. Observe, however, that the discussion can be easily adapted to also cover a simultaneous carrier frequency synchronisation. In [53] the authors demonstrated, that the same random synchronisation approach can be utilised to synchronise carrier frequencies when in each iteration not only the carrier phase but also the frequency of the transmit signals is altered. By this generalisation, the search space of the algorithm is increased. For each node, not only k distinct possibilities exist, but $k \cdot \rho$ where ρ denotes the count of distinct frequencies that can possibly be applied for each carrier signal. The optimisation time then increases by a factor of ρ. However, the analytical discussion becomes more complicated as the common period of the received sum signal might be increased considerably.

The optimisation problem is denoted as $\mathcal{P}$ and $T_{\mathcal{P}}$ describes the count of iterations required to reach one optimum for the problem $\mathcal{P}$.

An upper bound on the expected synchronisation time

The value of the feedback function increases with the number of carrier signals ζ_i that share the same interval for their phase offset γ_i at the receiver. Assume that $\kappa \in [1, k]$ is the interval that contains most of the carrier phase offsets. As worse feedback values are not accepted, we count the iterations required for all carrier signals to change to interval κ. We can roughly divide the values of the feedback function into n partitions $L_1, \ldots, L_n$ depending on the number of carrier signals with their phase in the interval κ. For each one transmitter, the probability to adapt its phase to one specific interval is $\frac{1}{k}$. The probability to increase the feedback value so that at least the next partition is reached is then

$$\frac{1}{k} \cdot (n - L_i) \cdot \frac{1}{n} \tag{2.3}$$

since one carrier signal ζ_i is altered with probability $\frac{1}{n}$ and the probability to reach any particular of the $(n - L_i)$ partitions that would increase the feedback value is $\frac{1}{k}$. In partition i, a total of

$$\binom{n - i}{1} = n - i \tag{2.4}$$

carrier signals each suffice to improve the feedback value with probability $\frac{1}{n} \cdot \frac{1}{k}$. We therefore require that at least one of the not synchronised carrier signals is correctly altered in phase while all other $n - 1$ signals remain unchanged. This happens with probability

$$\begin{aligned}
&\binom{n - i}{1} \cdot \frac{1}{n} \cdot \frac{1}{k} \cdot \left(1 - \frac{1}{n}\right)^{n-1} \\
= \ &\left(\frac{n - i}{n \cdot k}\right) \cdot \left(1 - \frac{1}{n}\right)^{n-1}.
\end{aligned} \tag{2.5}$$

Since

$$\left(1 - \frac{1}{n}\right)^n < \frac{1}{e} < \left(1 - \frac{1}{n}\right)^{n-1} \tag{2.6}$$

We obtain the probability $P[L_i]$ that L_i is left and a partition j with $j > i$ is reached as

$$P[L_i] \geq \frac{n - i}{n \cdot e \cdot k}. \tag{2.7}$$

The expected number of iterations to change the layer is bounded from above by $P[L_i]^{-1}$. We consequently obtain the overall expected synchronisation time as

$$\begin{aligned}
E[T_{\mathcal{P}}] \ &\leq \ \sum_{i=0}^{n-1} \frac{e \cdot n \cdot k}{n - i} \\
&= \ e \cdot n \cdot k \cdot \sum_{i=1}^{n} \frac{1}{i} \\
&< \ e \cdot n \cdot k \cdot (\ln(n) + 1) \\
&= \ \mathcal{O}\left(n \cdot k \cdot \log n\right).
\end{aligned} \tag{2.8}$$

A lower bound on the expected synchronisation time

After the initialisation, the phases of the carrier signals are identically and independently distributed. Consequently for a superimposed received sum signal ζ, each bit in the binary string s_ζ that represents the corresponding search point has an equal probability to be 1 or 0. The probability to start from a search point s_ζ with Hamming distance $h(s_{\mathrm{opt}}, s_\zeta)$ not larger than $l \in \mathbb{N}$; $l \ll n \cdot \log(k)$ to one of the global optima s_{opt} directly after the random initialisation is at most

$$
\begin{aligned}
P[h(s_{\mathrm{opt}}, s_\zeta) \leq l] &= \sum_{i=0}^{l} \binom{n \cdot \log(k)}{n \cdot \log(k) - i} \cdot \frac{k}{2^{n \cdot \log(k) - i}} \\
&\leq \frac{(n \cdot \log(k))^{l+2}}{2^{n \cdot \log(k) - l}}
\end{aligned}
$$

In this formula,

$$
\binom{n \cdot \log(k)}{n \cdot \log(k) - i} \tag{2.9}
$$

is the count of possible configurations with i bit errors to a given global optimum, $\frac{1}{2^{n \cdot \log(k) - i}}$ represents the probability for all these bits to be correct and k is the count of global optima. This means that with high probability (w.h.p.) the Hamming distance to the nearest global optimum is at least l. We use the method of the expected progress to calculate a lower bound on the optimisation time.

Let (s_ζ, t) denote the situation that search point s_ζ is achieved after t iterations of the algorithm. We assume a progress measure $\Lambda : \mathbb{B}^{n \cdot \log(k)} \to \mathbb{R}_0^+$ such that $\Lambda(s_\zeta, t) < \Delta$ represents the case that a global optimum was not found in the first t iterations. For every $t \in \mathbb{N}$ we have

$$
\begin{aligned}
E[T_\mathcal{P}] &\geq t \cdot P[T_\mathcal{P} > t] \\
&= t \cdot P[\Lambda(s_\zeta, t) < \Delta] \\
&= t \cdot (1 - P[\Lambda(s_\zeta, t) \geq \Delta]). \tag{2.10}
\end{aligned}
$$

With the help of the Markov-inequality we obtain

$$
P[\Lambda(s_\zeta, t) \geq \Delta] \leq \frac{E[\Lambda(s_\zeta, t)]}{\Delta} \tag{2.11}
$$

and therefore

$$
E[T_\mathcal{P}] \geq t \cdot \left(1 - \frac{E[\Lambda(s_\zeta, t)]}{\Delta}\right). \tag{2.12}
$$

This means that we can obtain a lower bound on the optimisation time by providing the expected progress after t iterations. The probability for l bits to correctly flip is at most

$$
\begin{aligned}
&\left(1 - \frac{1}{n \cdot \log(k)}\right)^{n \cdot \log(k) - l} \cdot \left(\frac{1}{n \cdot \log(k)}\right)^l \\
&\leq \frac{1}{(n \cdot \log(k))^l}. \tag{2.13}
\end{aligned}
$$

In this formula, $\left(1 - \frac{1}{n \cdot \log(k)}\right)^{n \cdot \log(k) - l}$ describes the probability that all 'correct' remain unchanged while the remaining l bits flip with probability $\left(\frac{1}{n \cdot \log(k)}\right)^{l}$. The expected progress in one iteration is therefore

$$
\begin{aligned}
E[\Lambda(s_\zeta, t), \Lambda(s_{\zeta'}, t+1)] &\leq \sum_{i=1}^{l} \frac{i}{(n \cdot \log(k))^i} \\
&< \frac{2}{n \cdot \log(k)}
\end{aligned}
\tag{2.14}
$$

and the expected progress in t iterations is consequently not greater than $\frac{2t}{n \cdot \log(k)}$. When we choose $t = \frac{n \cdot \log(k) \cdot \Delta}{4} - 1$, the double of the expected progress is still smaller than Δ. With the Markov inequality we can show that this progress is not achieved with probability $\frac{1}{2}$. Altogether we conclude that the expected synchronisation time is bounded from below by

$$
\begin{aligned}
E[T_\mathcal{P}] &\geq t \cdot \left(1 - \frac{E[\Lambda(s_\zeta, t)]}{\Delta}\right) \\
&\geq \frac{n \cdot \log(k) \cdot \Delta}{4} \cdot \left(1 - \frac{\frac{2 \cdot n \cdot \log(k)}{4 \cdot n \cdot \log(k)} \cdot \Delta}{\Delta}\right) \\
&= \Omega(n \cdot \log(k) \cdot \Delta)
\end{aligned}
\tag{2.15}
$$

With $\Delta = k \cdot \frac{\log(n)}{\log(k)}$ we obtain a lower bound in the same order as the upper bound derived in section 2.1.2 and consequently an asymptotically sharp bound of

$$
E[T_\mathcal{P}] = \Theta\left(n \cdot k \cdot \log(n)\right).
\tag{2.16}
$$

Note that in [52] an upper bound on the expected asymptotic synchronisation time was derived that scales linearly in the number of nodes n when the probability distribution is optimally altered repeatedly during the synchronisation. However, simulation results derived for a fixed uniform distribution in this study also indicate a logarithmic factor in the synchronisation time of one-bit feedback based synchronisation.

2.1.3 Hierarchical clustering

A further improvement of the synchronisation time can be achieved by synchronising smaller clusters of nodes separately. Since this bound on the synchronisation time grows faster than linearly with the network size n but the received signal strength RSS_{sum} of the received superimposed signal grows linearly with n, the overall energy consumption and synchronisation time might be reduced when fewer nodes transmit for a shorter time but with an increased transmission power. Note that currently most low cost radios are not capable of altering their transmission power and therefore are not able to exploit this property. More sophisticated radios could, however, achieve carrier phase synchronisation more efficiently when this fact is utilised. We propose the following hierarchical clustering scheme that synchronises all transmit nodes iteratively in clusters of reduced size.

1. Determine clusters (e.g. by a random process initialised by the receiver node)

2. Synchronise clusters successively as described above with possibly increased transmit power. When cluster ι is sufficiently synchronised, nodes in this cluster sustain their carrier signal and stop transmitting until all clusters are synchronised.

3. At this stage, carrier signals in all clusters are in phase but carrier phases of distinct clusters might differ. Determine representative nodes from all clusters and synchronise these.

4. Nodes in all clusters alter their carrier phase by the phase offset experienced and broadcast by the corresponding representative node (broadcast).
   Let $\zeta_i = \Re\left(m(t)\mathrm{RSS}_i e^{j2\pi f_c t(\gamma_i + \phi_i + \psi_i)}\right)$ and $\zeta_i' = \Re\left(m(t)\mathrm{RSS}_i e^{j2\pi f_c t(\gamma_i' + \phi_i + \psi_i)}\right)$ be the carrier signals of representative node i from cluster ι before and after synchronisation between representative nodes was achieved. A node h from cluster ι alters its carrier signal $\zeta_h = \Re\left(m(t)\mathrm{RSS}_h e^{j2\pi f_c t(\gamma_h + \phi_h + \psi_h)}\right)$ to $\zeta_h' = \Re\left(m(t)\mathrm{RSS}_h e^{j2\pi f_c t(\gamma_h + \phi_h + \psi_h + \gamma_i - \gamma_i')}\right)$. Under ideal conditions, all nodes are now in phase.

5. To account for synchronisation errors a final synchronisation phase in which all nodes participate concludes the overall synchronisation process.

Fig. 2.2 illustrates this procedure. The crucial idea of this approach is applied in step 4. Since nodes inside a cluster have already been synchronised, they are still in phase after all apply an identical phase offset. Because this offset is the phase alteration the representative nodes experienced to during their synchronisation, all nodes should be synchronised after this step.

A potential problem for this approach is phase noise. Since only one cluster is synchronised at a time, phases of nodes in the inactive clusters experience phase noise and start drifting out of phase due to practical properties of oscillators. However, we show in section 2.1.5 that sufficient synchronisation is possible in the order of milliseconds. Therefore, we do not consider phase noise an important issue. Observe that all coordination is initiated by the receiver node so that no inter-node communication is required for coordination.

Depending on the network size, more than one hierarchy stage might be optimal for the synchronisation time and the energy consumption. To estimate the optimal hierarchy depth and the optimum cluster size, the count of nodes participating in the synchronisation must be computed. We assume that the nodes themselves do not know the network size. This means that the remote receiver derives the network size, calculates optimal cluster sizes and hierarchy depths and broadcasts this information. In [56] it was demonstrated that the superimposed sum signal from arbitrarily synchronised nodes is sufficient to estimate the number of transmitters. We derive the optimum hierarchy depth and cluster size by integer programming in time $\mathcal{O}(n^2)$ (cf. Appendix B). The expected synchronisation time is dependent on the cluster count, cluster size and hierarchy depth. Since for each cluster a small instance of the original problem is solved, the synchronisation time can be composed from the synchronisation times of individual clusters.
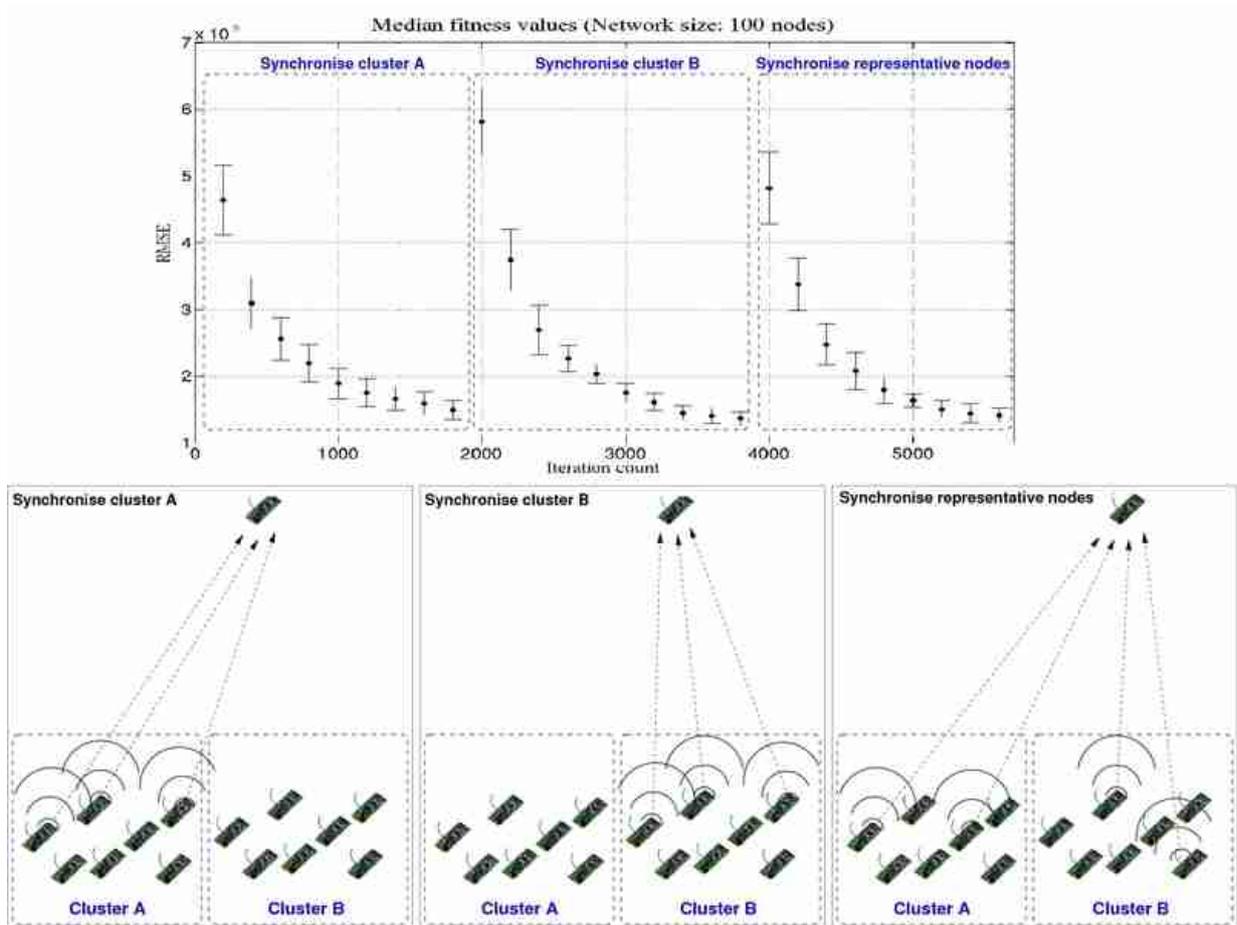
Figure 2.2: Illustration of the approach to cluster the network of nodes in order to improve the synchronisation time of feedback based closed-loop distributed adaptive beamforming. (©2011 IEEE Published by the IEEE CS, CASS, ComSoc, IES, SPS)

18

2.1.4 An asymptotically optimal algorithm

Since no local optimum exists in the search space due to its weak multimodality (cf. Appendix A), the performance of $\Theta(n \cdot k \cdot \log(n))$ of the random search seems weak. Previously, only one feedback bit was utilised. Due to this reduced information, some of the information available at the receiver is not available by the nodes that actually react on the feedback. When more information is included in the feedback of the receive node we are able to design an asymptotically optimal synchronisation algorithm.

In every iteration the receiver provides additional information over a feedback value so that a node i can learn the optimum phase offset of its own carrier

$$\zeta_i = \Re\left(m(t)\text{RSS}_i e^{j2\pi f_c t(\gamma_i + \phi_i + \psi_i)}\right)$$

relative to the superimposed sum signal

$$\zeta_{\text{sum} \smallsetminus i} = \Re\left(m(t)e^{j2\pi f_c t} \sum_{o \in [1,n]; o \neq i} \text{RSS}_o e^{j(\gamma_o + \phi_o + \psi_o)}\right)$$

of all other nodes, provided that the latter does not change significantly. $\zeta_{\text{sum} \smallsetminus i}$ is a sinusoidal signal. The feedback is maximal when ζ_i and $\zeta_{\text{sum} \smallsetminus i}$ have identical phase offset at a receiver. With increasing phase offset

$$|(\gamma_i + \phi_i + \psi_i) - (\gamma_{\text{sum} \smallsetminus i} + \phi_{\text{sum} \smallsetminus i} + \psi_{\text{sum} \smallsetminus i})|$$

the feedback value decreases symmetrically. Consequently, the feedback function has the form $\mathcal{F}(\gamma_i) = A\sin(\gamma_i + \Phi) + c$. This is an equation with the three unknowns A (amplitude), Φ (phase offset of $\mathcal{F}$) and the additive term c so that a node i can calculate it with three distinct measurements. Fig. 2.3 illustrates the accuracy of this procedure for 100 transmitters. The root of the mean square error (RMSE) is calculated as

$$RMSE = \sqrt{\sum_{t=0}^{\tau} \frac{\left(\zeta_{\text{sum}} + \zeta_{\text{noise}} - \zeta_{\text{opt}}\right)^2}{n}}. \tag{2.17}$$

Here, τ is chosen to cover several signal periods.

For the optimisation process, a node will during each of four subsequent iterations either alter the phase offset of its carrier signal or sustain it for all four iterations. The probability to alter the phase offset should be low as, for instance, $\frac{1}{n}$. A node that decides to alter its phase offset, will do this three times to measure feedback values for distinct phase offsets, derive with these measurements the feedback function, alter its phase offset accordingly and finally transmit a fourth time to obtain the amount by which the achieved feedback value deviates from the expected value. If the deviation is small, the node will not alter its phase further since the current phase offset is considered optimal. All other nodes then adapt the probability to alter their carrier phase so that one node alters the phase of its carrier signal on average per iteration (for instance from $\frac{1}{n}$ to $\frac{1}{n-1}$).
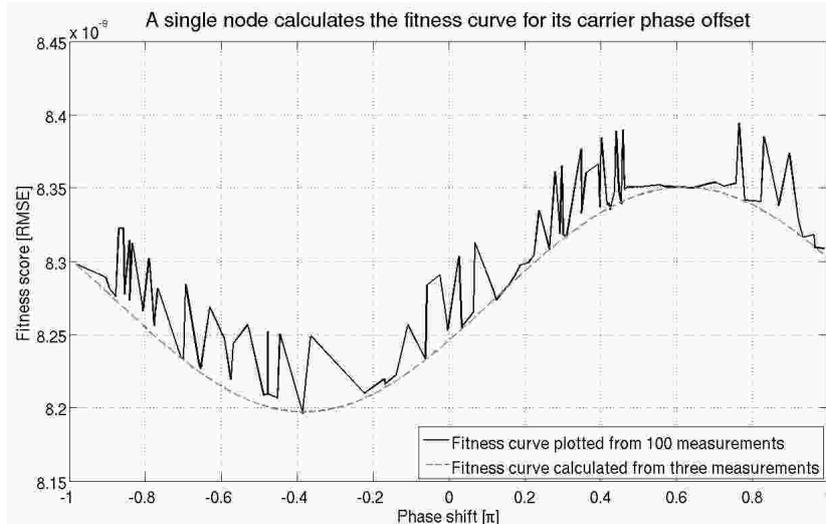
Figure 2.3: Deviation of the feedback curve calculated from three measurements to the feedback curve plotted from 100 measurements. (©2011 IEEE Published by the IEEE CS, CASS, ComSoc, IES, SPS)

As nodes are chosen according to a random process, it is possible that more than one node simultaneously alters its phase offset. In this case, the node's conclusions on the impact of their phase-alteration on the feedback value are biased. Therefore, in the fourth measurement, when the measured value deviates significantly from the expected feedback a node concludes that it was not the only one to alter its phase and reverses its decision.

In our measurements, the deviation of the calculated feedback curve did not exceed 0.6% when only one node adapts its phase offset. With two nodes simultaneously adapting their phase offset we already experienced a deviation of approximately 1.5%. As this procedure is guided purely by the feedback broadcast by the receiver, inter-node communication is not required.

Asymptotically, the synchronisation time of this algorithm is $\Theta(n)$ since on average the count of carrier signals that are in phase increases by 1 in each iteration. Further performance improvements can be achieved when nodes utilise only three subsequent iterations and acquire the first measurement from the last transmission of the preceding three subsequent iterations.

The asymptotic synchronisation time derived for this approach is optimal when we assume that individual nodes have to compute their optimal carrier phase offset independently since n carrier signals have to be adapted. When, however, a synchronisation scheme is utilised in which information about the optimum relative carrier phase offsets of all nodes is provided, as e.g. in typical open-loop carrier synchronisation schemes (cf. [29]), the asymptotic synchronisation time can be further reduced.

This improved carrier synchronisation scheme can be applied in any scenario in which a rich feedback as, for instance, the SNR can be provided. It is, however, not applicable when only binary feedback is provided by the receiver. When, for example, high noise

Table 2.1: Configuration of the simulations. P_{rx} is the the received signal power, d is the distance between transmitter and receiver and λ is the wavelength of the signal

Property	Value
Node distribution area	$30m \times 30m$
Location of the receiver	$(15m, 15m, 30m)$
Mobility	stationary nodes
Base band frequency	$f_{base} = 2.4$ GHz
Transmission power of nodes	$P_{tx} = 1$ mW
Gain of the transmit antenna	$G_{tx} = 0$ dB
Gain of the receive antenna	$G_{rx} = 0$ dB
Iterations per simulations	6000
Identical simulation runs	10
Random noise power [57]	-103 dBm
Pathloss calculation (P_{rx})	$P_{tx} \left(\frac{\lambda}{4\pi d}\right)^2 G_{tx} G_{rx}$

and interference would force an impractically complex error correction scheme, it might be beneficial to utilise the one bit feeback based carrier synchronisation instead.

2.1.5 Simulation studies

We have implemented the scenario of distributed adaptive beamforming in Matlab to obtain a better understanding of the impact of environmental parameters and algorithmic configurations. In particular, the effect of distinct probability distributions as well as the count of transmitters and the transmission distance are considered. In these simulations, 100 transmit nodes are placed uniformly at random on a $30m \times 30m$ square area. The receiver is located $30m$ $(100m, 200m, 300m)$ above the centre of this area. Receiver and transmit nodes are stationary. Simulation parameters are summarised in Table 2.1.

Frequency and phase stability are considered perfect. We derived the median and standard deviation from 10 simulation runs. One iteration consists of the nodes transmitting, feedback computation, feedback transmission and feedback interpretation by transmitters. It is possible to perform these steps within few signal periods so that the time consumed for 6000 iterations is in the order of milliseconds for a base band signal frequency of 2.4 GHz. Signal quality is measured by the RMSE of the received signal to an expected optimum signal as detailed in equation (2.17). The optimum signal is calculated as a perfectly aligned and properly phase shifted received sum signal from all transmit sources. For the optimum signal, noise is disregarded.

Fig. 2.4a depicts the optimum carrier signal, the initial received sum signal and the synchronised carrier after 6000 iterations when carrier phases are altered with probability $\frac{1}{n}$ in each iteration according to a uniform distribution. In Fig. 2.4b, the phase offset of received signal components for an exemplary simulation run with the same parameters are illustrated. We observe that after 6000 iterations about 98% of all carrier signals converge to a relative phase offset of about +/- 0.1π. The median of all variances of the

(a) Received sum signal from 100 transmit nodes without synchronisation and after 6000 iterations

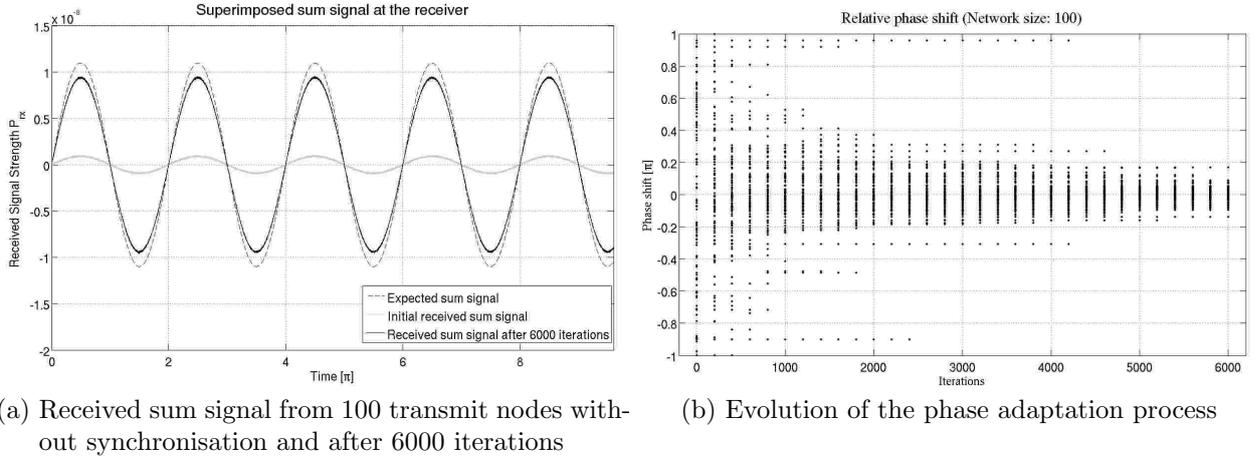(b) Evolution of the phase adaptation process

Figure 2.4: Simulation results for a simulation with 100 transmit over 6000 iterations of the random optimisation approach to distributed adaptive beamforming in wireless sensor networks. (©2011 IEEE Published by the IEEE CS, CASS, ComSoc, IES, SPS)

phase offsets for simulations with this configuration is 0.2301 after 6000 iterations. The actual synchronisation time is dependent on the time to complete a single iteration. In each iteration, a synchronisation signal is transmitted, the received sum signal is analysed, feedback is calculated, broadcast to the network and interpreted by transmit nodes. While the processing speed might be improved with improved hardware, the round trip time of the signal poses a definite lower bound for the time a single iteration lasts. At a distance of 30 meters, for instance, we can not hope to complete a single iteration in less than $0.2\mu s$.

Uniform vs. normal distribution

Distributed adaptive beamforming in wireless sensor networks has been studied in the literature according to various random phase alteration processes. The authors in [26, 27, 28] report good results when the probability p_γ to alter the phase of a single carrier signal in one iteration is 1 for all nodes and the phase offset is chosen according to a normal distribution. The variance σ_γ^2 applied is not reported. In [24, 25] p_γ was set to $\frac{1}{n}$ for each one of the n nodes while the phase is altered according to a uniform distribution.

For both, uniform and normal distributed processes, we consider several values for p_γ and σ_γ^2. Generally, we achieved good performance when modifications in one iteration were small. For the uniform distribution this translates to $p_\gamma = \frac{1}{n}$. For the normal distribution, good results are achieved when σ_γ^2 and p_γ are balanced so that the modification to the overall sum signal is small. With increasing p_γ good results are achieved with decreasing σ_γ^2. Fig. 2.5 depicts the results for $p_\gamma = \frac{1}{n}$ and $\sigma_\gamma^2 = 0.5\pi$

The figure shows the median RMSE value achieved in 10 simulations by normal and uniform distributed processes over the course of 6000 iterations. For ease of presentation, error bars are omitted in this figure. However, the standard deviation is low for both processes
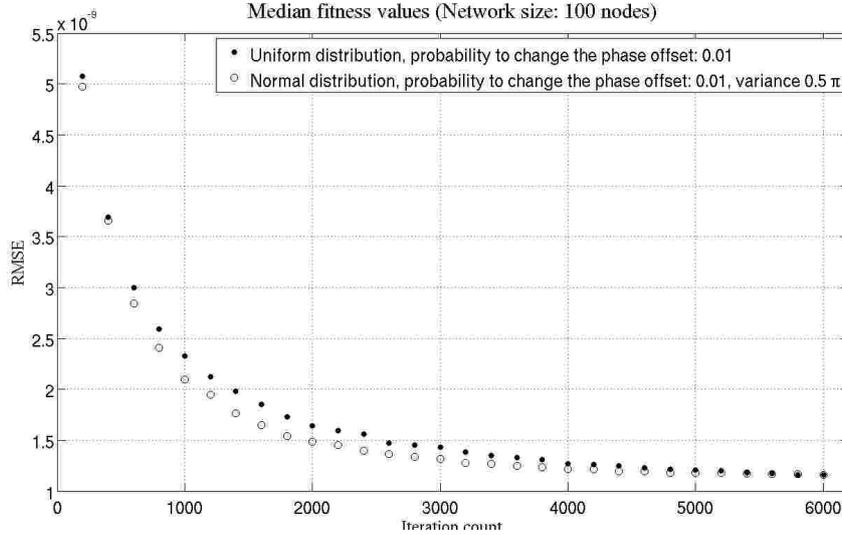
Figure 2.5: Performance of normal and uniform distributions for a network size of 100 nodes and $p_\gamma = 0.01, \sigma_\gamma^2 = 0.5\pi$. (©2011 IEEE Published by the IEEE CS, CASS, ComSoc, IES, SPS)

(the standard deviation of this normal distributed process is depicted in Fig. 2.9b).

The normal distributed process has a slightly improved synchronisation performance. The optimum feedback value reached is, however, identical.

Impact of the network size

When the count of nodes that participate in the synchronisation is altered, this also impacts the performance of this process (cf. section 2.1.2). We conducted several simulations with network sizes ranging from 20 to 100 nodes. Fig. 2.6 depicts the performance of several synchronisation processes with varying network sizes.

In these simulations, we set $p_\gamma = 0.05$ and utilised a uniformly distributed phase alteration process. We see that the maximum feedback value achieved is lower for smaller network sizes. This is due to the RMSE measure that compares the achieved sum signal to an expected optimum superimposed signal. As the count of participating nodes diminishes, also the amplitude of the optimum signal decreases. As expected, the optimum value is reached earlier for smaller network sizes.

Impact of the transmission distance

We are also interested in the performance of distributed adaptive beamforming when the distance between the network and a receiver is increased. For a uniformly distributed phase alteration process with $p_\gamma = \frac{1}{n}$ we increase the transmission distance successively. Fig. 2.7 depicts the phase coherency achieved and the received sum signal for various transmission distances.
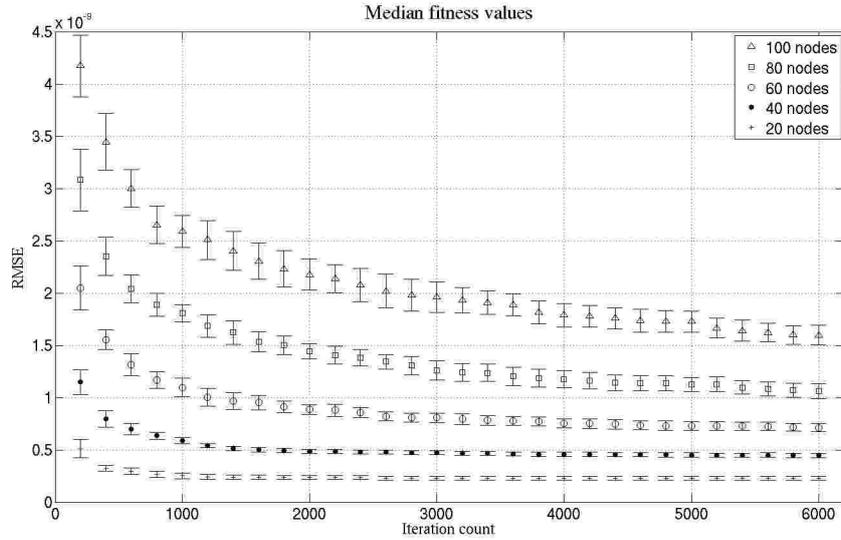
Figure 2.6: The synchronisation performance for various network sizes in a uniformly distributed process with $p_\gamma = 0.05$. (©2011 IEEE Published by the IEEE CS, CASS, ComSoc, IES, SPS)

Although the noise power relative to the sum signal increases, synchronisation is possible at about 200 meters distance. Observe that in our model with $P_{tx} = 1mW$ we expect a signal strength at the receiver of $0.1\mu W$ or $-40dBm$ for each single carrier at this distance.

When the distance is further increased to 300 meters, however, synchronisation is not possible with this configuration, due to the high impact of the noise fluctuation on the received signal. This has a higher impact on the signal than the alteration of single carrier signals.

However, when more carrier signals are altered simultaneously, a weak synchronisation is still possible. Fig. 2.8 depicts the received carrier signal after 100 iterations for the uniformly distributed process with $p_\gamma = 0.2$ and $p_\gamma = 0.6$. We see that the synchronisation quality is improved with increasing p_γ. While the superimposed signal is indistinguishable for $p_\gamma = 0.2$, the synchronisation quality increases with $p_\gamma = 0.6$. Although the signal is heavily distorted, the carrier can be extracted.

Utilisation of additional feedback information

We also conducted simulations in which our implementation of the asymptotically optimal algorithm described in section 2.1.4 is compared to the classical process with normal distributed phase alterations. When optimum phase offsets are calculated by solving multivariable equations at the transmit nodes, the synchronisation performance can be greatly improved as detailed in section 2.1.4. Fig. 2.9 depicts the performance improvement achieved by solving multivariable equations to determine the feedback function compared to a global random search approach. We observe that the global random search heuristic is outperformed already after about 1000 iterations and the feedback value reached is greatly
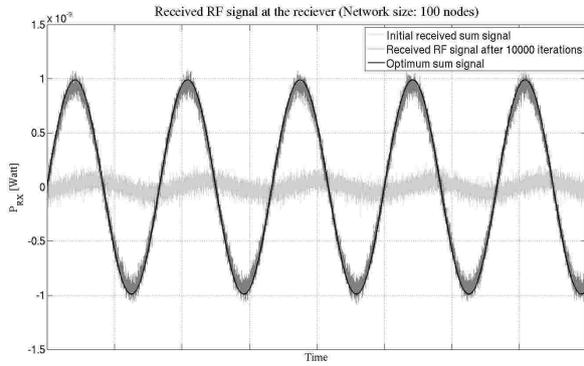
(a) Receiver distance: 100 meters – Received RF signal
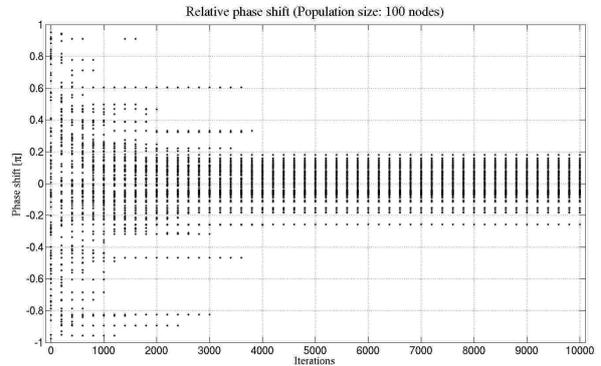
(b) Receiver distance: 100 meters – Relative phase shift of signal components
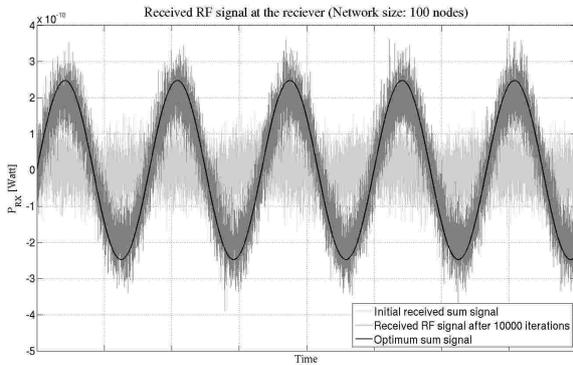
(c) Receiver distance: 200 meters – Received RF signal
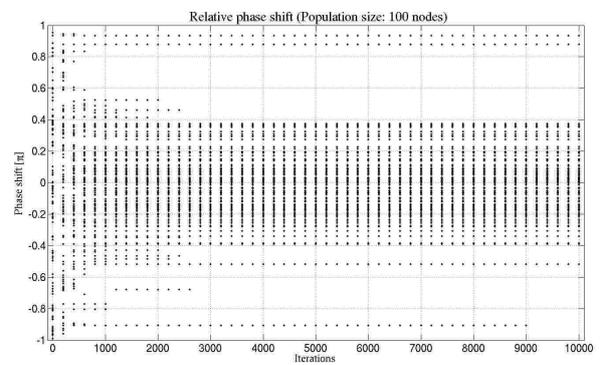
(d) Receiver distance: 200 meters – Relative phase shift of signal components
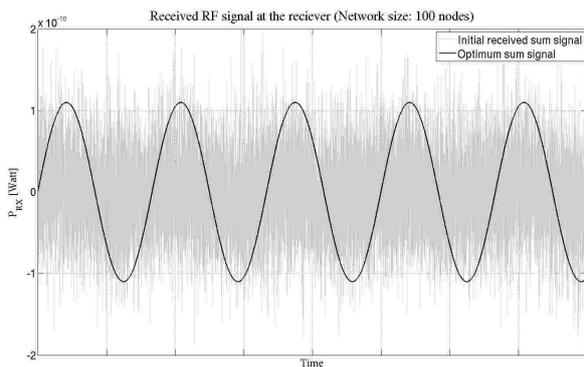
(e) Receiver distance: 300 meters – Received RF signal
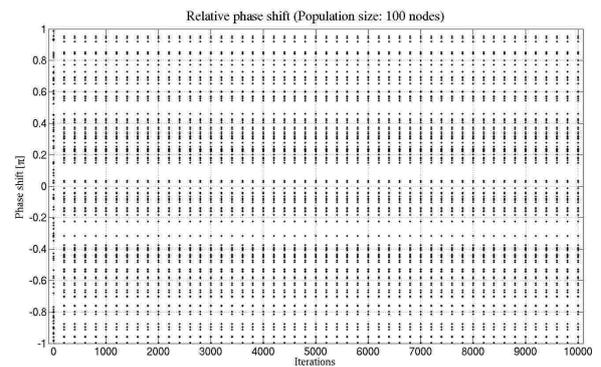
(f) Receiver distance: 300 meters – Relative phase shift of signal components

Figure 2.7: RF signal strength and relative phase shift of received signal components for a network size of 100 nodes after 10000 iterations. Nodes are distributed uniformly at random on a $30m \times 30m$ square area and transmit at $P_{TX} = 1mW$ with $p_\gamma = \frac{1}{n}$. (©2011 IEEE Published by the IEEE CS, CASS, ComSoc, IES, SPS)

Probability to alter phase offset: 0.2
Superimposed sum signal at the receiver (Network size: 100 nodes)

Probability to alter phase offset: 0.6
Superimposed sum signal at the receiver (network size: 100 nodes)

(a) Receiver distance: 300 meters, $p_\gamma = 0.2$ (b) Receiver distance: 300 meters, $p_\gamma = 0.6$
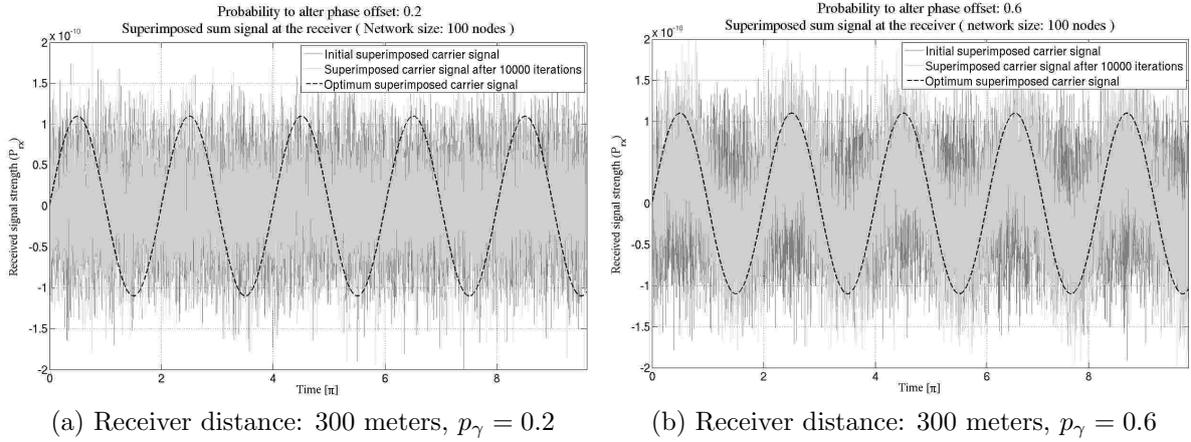
Figure 2.8: RF signal strength and relative phase shift of received signal components for a network size of 100 nodes after 10000 iterations. Nodes are distributed uniformly at random on a $30m \times 30m$ square area and transmit at $P_{TX} = 1mW$.

(©2011 IEEE Published by the IEEE CS, CASS, ComSoc, IES, SPS)

improved.

The phase offset of distinct nodes is within $+/-0.05\pi$ for up to 99% of all nodes.

2.1.6 Near realistic instrumentation

We have utilised USRP software radios (http://www.ettus.com) to model a sensor network capable of distributed adaptive transmit beamforming. The software radios are controlled via the GNU radio framework (http://gnuradio.org). The transmitter and receiver modules implement the feedback based distributed adaptive beamforming[2]. For the superimposed transmit channel and the feedback channel we utilised widely separated frequencies so that the feedback could not impact the synchronisation performance. We conducted experiments with several transmit frequencies of nodes. In these experiments we repeatedly synchronised the carrier phases of the three transmit devices with the help of the 1-bit feedback based algorithm described in [27, 29] with uniform or normal probability distribution on the phase modulation. Table 2.2 summarises the configuration and results of two experiments with low and high transmit frequencies of 27MHz and 2.4GHz, respectively. After 10 experiments at an RF transmit frequency of 27MHz we achieved a median gain in the received signal strength of 3.72dB for three independent transmit nodes after 200 iterations. In 14 experiments with 4 independent nodes that transmit at 2.4GHz the achieved median gain of the received RF sum signal was 2.19dB after 500 iterations.

---

[2]The software for our feedback based closed loop implementation is constantly further improved and extended in student projects. It is currently not recommended for productive environments. If you are interested to receive a copy of the code in order to participate in the development and testing of the implementation please contact sigg@ibr.cs.tu-bs.de.

Table 2.2: Experimental results of software radio instrumentations (©2011 IEEE Published by the IEEE CS, CASS, ComSoc, IES, SPS)

Experimental setting	Experiment 1	Experiment 2
Separation of TX antennas [m]	≈ 0.21	≈ 0.3
Distance to receiver [m]	≈ 0.75	≈ 4
Transmit RF Frequency [MHz]	$f_{TX} = 2400$	$f_{TX} = 27$
Receive RF Frequency [MHz]	$f_{RX} = 902$	$f_{RX} = 902$
Iterations per experiment	500	200
Mobility	stationary	stationary
Identical experiments	14	10
Transmit devices	4	3
Receive devices	1	1
Algorithmic configuration		
Random distribution	uniform	uniform
Phase alteration probability	0.25	0.33
Hardware		
Transmit board	RFX2400	LFTX
Receive board	RFX900	RFX900
Gain of receive antenna [dBi]	$G_{RX} = 3$	$G_{RX} = 3$
Gain of transmit antenna [dBi]	$G_{TX} = 3$	$G_{TX} = 1.5$
Median gain (P_{RX}) [dB]	2.19	3.72
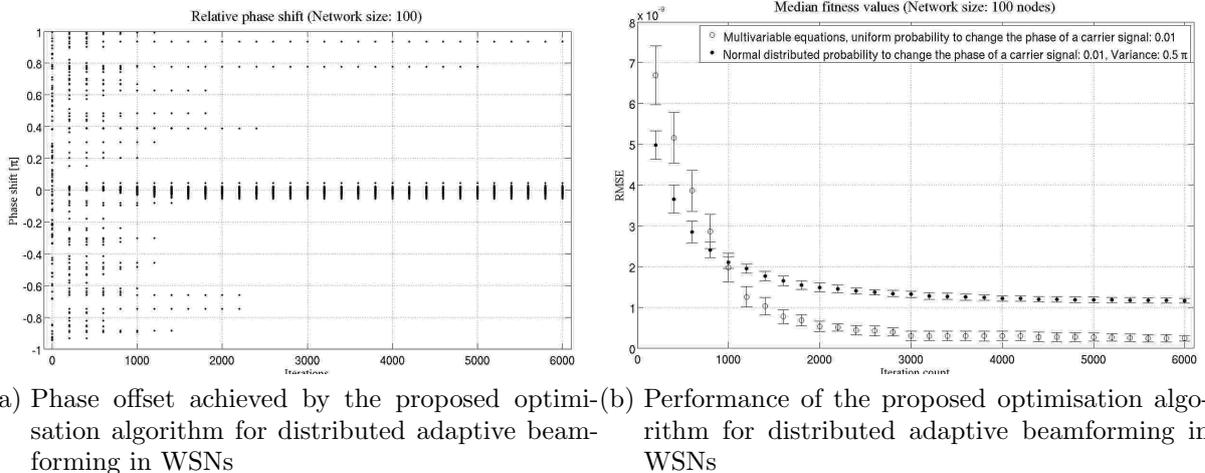
(a) Phase offset achieved by the proposed optimi-(b) Performance of the proposed optimisation algo-
sation algorithm for distributed adaptive beam- rithm for distributed adaptive beamforming in
forming in WSNs   WSNs

Figure 2.9: Distributed adaptive beamforming with a network size of 100 nodes where
phase alterations are drawn uniformly at random. Each node adapts its carrier
phase offset with probability 0.01 in one iteration. In this case, multivariable
equation are solved to determine the optimum phase offset of the carrier signal.

For the transmitters we utilised the clock of the first device for all transmit nodes. The
receive node utilises its own clock and is therefore not synchronised to any of the transmit
nodes. Apart from this clock synchronisation no other communication or synchronisation
between transmitters was applied. In future implementations it is possible to utilise GPS
for the clock synchronisation.

In a third experiment we altered the transmission distance and the phase alteration
variance for a normal distributed random process. Fig. 2.10 depicts our experimental
setting. Table 2.3 summarises our experimental configuration.

Carrier phases have been adapted for each transmit device independently following a
normal distributed random process. We modified the probability to alter the phase offset
of one device and the variance for its normal distributed random process as well as the
distance between transmit and receive devices.

Some results derived are depicted in Fig. 2.11. In the figure, the mean gain of the
received sum signal over all 12 experiments to the initially received sum signal is depicted.
As expected, we observe that the synchronisation process differs for different parameter
settings. Again, best results are achieved when small changes are applied in each iteration.
Therefore, the experiments in which the phase alteration probability and the variance are
small achieve superior results.
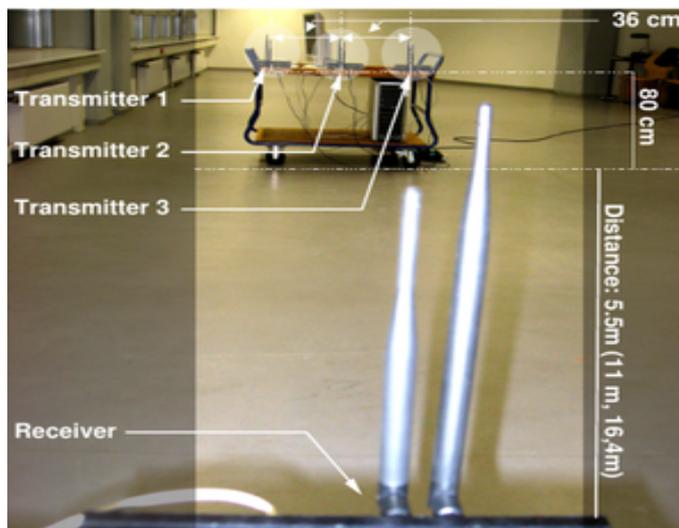
Figure 2.10: Experimental instrumentation of distributed adaptive beamforming among three transmit USRP devices and one receive USRP device. (©2011 IEEE Published by the IEEE CS, CASS, ComSoc, IES, SPS)
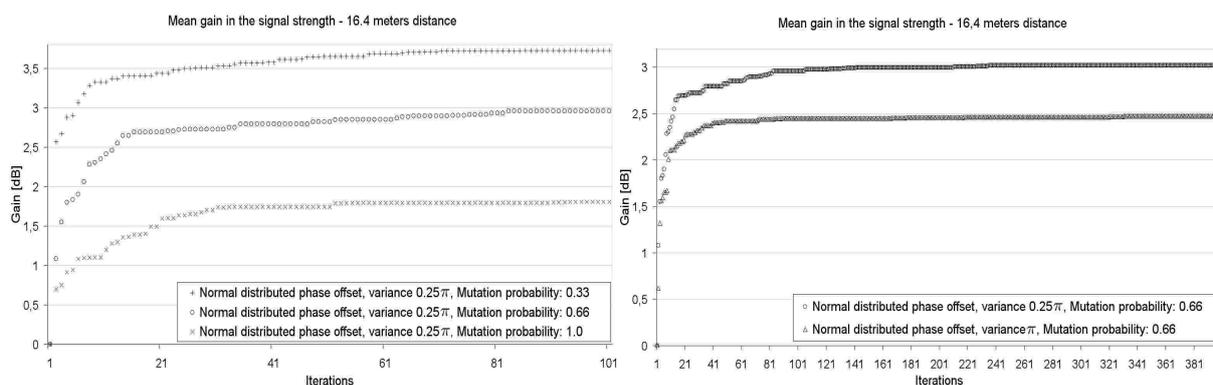
(a) Mean gain in the signal strength at a transmission distance of 16.4 meters and a variance of the random process of 0.25π

(b) Mean gain in the signal strength at a transmission distance of 5.5 meters and a variance of the random process of 0.25π and π

Figure 2.11: Mean gain in the signal strength of three collaboratively transmitting devices

(©2011 IEEE Published by the IEEE CS, CASS, ComSoc, IES, SPS)

Table 2.3: Configuration of the experiment (©2011 IEEE Published by the IEEE CS, CASS, ComSoc, IES, SPS)

Experimental setting	Experiment 3
Separation of transmit antennas [m]	0.36
Distance to receive antenna [m]	5.5 / 11 / 16.4
Transmit frequency [MHz]	$f_{TX} = 2400$
Receive frequency [MHz]	$f_{RX} = 902$
Iterations per experiment	400
Mobility	stationary
Identical experiments	12
Transmit devices	3
Receive devices	1
Algorithmic configuration	
Random distribution of the phase alteration	normal
Phase alteration probability	0.33 / 0.66 / 1.00
Variance of normal distribution [π]	0.25 / 1
Hardware	
Transmit board	RFX2400
Receive board	RFX900
Transmit antenna	VERT2450
Receive antenna	VERT900
Gain of receive antenna [dBi]	$G_{RX} = 3$
Gain of transmit antenna [dBi]	$G_{TX} = 3$

2.1.7 Conclusion

We have considered randomised search approaches to solve the problem of distributed adaptive transmit beamforming. In an analytic consideration an asymptotically tight bound on the expected optimisation time of $\Theta(n \cdot k \cdot \log(n))$ was derived.

Additionally, a protocol to further reduce the optimisation time and energy consumption of distributed adaptive beamforming was introduced. In this protocol, the problem was divided into sub-problems that were solved iteratively. Since the decrease in the synchronisation time is greater than the increase in transmission power in smaller clusters, this approach can improve the optimisation time and reduce energy consumption.

Furthermore, an asymptotically optimal algorithm was derived. For this approach we considered the possibility to estimate the unknown feedback function by an individual node so that an optimisation approach is possible that scales linearly with the network size n. This approach is asymptotically optimal since each carrier signal has to be considered at least once individually in order to find its optimum phase offset.

In mathematical simulations we demonstrated the effect of several configurations for distributed adaptive transmit beamforming with uniform and normal distributed phase alteration methods. Generally, a low mutation probability translates to a better performance in the phase synchronisation process. An adaptive probability over the course of the optimisation might further improve the optimisation speed. While a moderate mutation probability is beneficial at the beginning of the simulation, a smaller mutation probability shows an improved optimisation speed later in the process. Also, our implementation of the asymptotically optimal method greatly outperforms the global random search approach in the synchronisation achieved and the optimisation speed.

Finally, in an instrumentation with USRP software radios we demonstrated the feasibility of distributed adaptive transmit beamforming in a concrete implementation with up to four transmitters.

APPENDIX A: On the multimodality of the feedback function

We easily see that the feedback function is multimodal. The reason is that, given the search point corresponding to an optimum sum signal ζ_{opt} we can state another optimum by adding the same phase offset γ' to all carrier signals. In particular, the feedback function is weak multimodal so that no local optimum exists.

Identical transmit frequencies

When carrier frequencies among nodes are identical a local optimum exists if we can identify at least one search point $s_{\overline{\zeta}}$ for which all small phase modifications decrease the feedback value, while some larger modifications increase it. The smallest possible modification is realised when the transmit phase is altered for exactly one carrier signal ζ_i. Fig. 2.12 illustrates that the feedback of a signal is given by the distance between the rotation angles φ_{opt} and φ_i of an optimal configuration $s_{\zeta_{\mathrm{opt}}}$ and s_{ζ_i} as $\left|\cos(\varphi_{\mathrm{opt}}) - \cos(\varphi_i)\right|$.
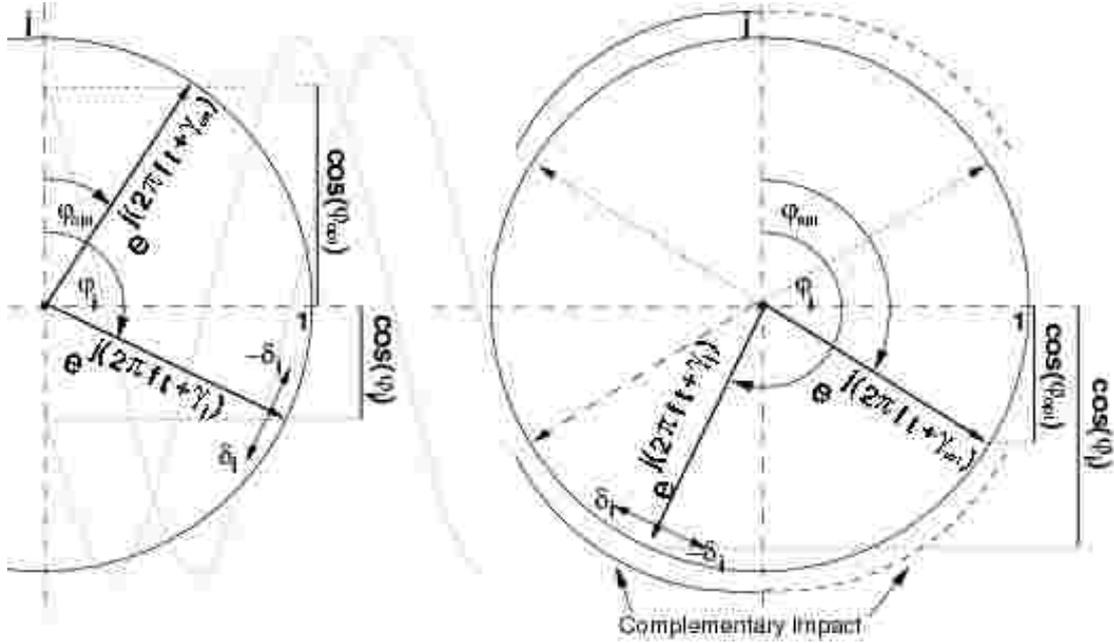
Figure 2.12: Fitness calculation of signal components. The feedback of the superimposed sum signal is impacted by the relative phase offset of an optimally aligned signal and a carrier signal i. (©2011 IEEE Published by the IEEE CS, CASS, ComSoc, IES, SPS)

Compared to s_{opt} no configuration short of the optimum configuration $s_i = s_{\text{opt}}$ exists for which the phase offset between signal components is increased for phase offset δ_i regardless of the sign of δ_i (cf. Fig. 2.12).

Distinct transmit frequencies

When signal frequencies differ, the feedback function is not affected by phase modifications only. The reason for this is that we can for every positive contribution to the feedback function also find a negative contribution of the same amount in the common period of ζ_{opt} and ζ_i.

APPENDIX B: Calculation of optimal hierarchy depth and cluster size

We estimate the expected optimisation time for a network of size n by $E[T_{\mathcal{P}n}] = c{\cdot}k{\cdot}n{\cdot}\log(n)$ for a suitable constant c and the expected energy consumption by $E[\mathcal{E}_{\mathcal{P}n}] = c \cdot k \cdot n \cdot \log(n) \cdot \overline{\mathcal{E}_{\mathcal{P}n}}$ where $\overline{\mathcal{E}_{\mathcal{P}n}}$ is the energy consumption of all n nodes in one iteration [25]. A hierarchy and cluster structure that minimises these formulae when summed over all hierarchy stages is optimal in our sense. We derive the optimum cluster sizes and hierarchy depths by integer

programming. For a cluster size of m the above formulae have the property

$$E[T_{\mathcal{P}n}] = E[T_{\mathcal{P}\frac{n}{m}}] \cdot \frac{n}{m} \cdot E[T_{\mathcal{P}m}]$$

$$E[\mathcal{E}_{\mathcal{P}n}] = E[\mathcal{E}_{\mathcal{P}\frac{n}{m}}] \cdot \frac{n}{m} \cdot E[\mathcal{E}_{\mathcal{P}m}].$$

We define the recursion by

$$E_{\text{opt}}[T_{\mathcal{P}n}] = min_m \left[E_{\text{opt}}[T_{\mathcal{P}\frac{n}{m}}] \cdot \frac{n}{m} \cdot E_{\text{opt}}[T_{\mathcal{P}m}] \right]$$

$$E_{\text{opt}}[\mathcal{E}_{\mathcal{P}n}] = min_m \left[E_{\text{opt}}[\mathcal{E}_{\mathcal{P}\frac{n}{m}}] \cdot \frac{n}{m} \cdot E_{\text{opt}}[\mathcal{E}_{\mathcal{P}m}] \right]$$

and the start of the recursion by $E_{\text{opt}}[T_{\mathcal{P}\eta}]$ and $E_{\text{opt}}[\mathcal{E}_{\mathcal{P}\eta}]$ with η being the minimum feasible cluster size when the maximum transmission power and distance are given. Since η is dependent on the distance to the receiver, it can be calculated over the round trip time between the sensor network and the receiver.

The time required for the calculation of the optimum hierarchy depth and cluster sizes is quadratic. With a network of n nodes, at most n^2 distinct terms $E_{\text{opt}}[T_{\mathcal{P}i}]$ and $E_{\text{opt}}[\mathcal{E}_{\mathcal{P}i}]$ with $i \in \{1, \ldots, n\}$ are of relevance. We can start by calculating $E_{\text{opt}}[\mathcal{E}_{\mathcal{P}\eta}]$ and $E_{\text{opt}}[T_{\mathcal{P}\eta}]$ and obtain all other values by table look-up according to $E_{\text{opt}}[T_{\mathcal{P}n}]$ and $E_{\text{opt}}[\mathcal{E}_{\mathcal{P}n}]$ in time $\mathcal{O}(n^2)$ since every one of the (at most) n entries has not more than n possible predecessors.

2.2 A fast binary feedback-based distributed adaptive carrier synchronisation for transmission among clusters of disconnected IoT nodes in smart spaces [3]

We propose a transmission scheme among groups of disconnected IoT devices in a smart space. In particular, we propose the use of a local random search implementation to speed up the synchronisation of carriers for distributed adaptive transmit beamforming. We achieve a sharp bound on the asymptotic carrier synchronisation time which is significantly lower than for previously proposed carrier synchronisation processes. Also, we consider the impact of environmental conditions in smart spaces on this synchronisation process in simulations and a case study.

2.2.1 Introduction

The advancing miniaturisation of electronics and its integration into everyday objects fosters smart spaces as an antecedent to an Internet of Things (IoT). In such environments, arbitrarily distributed, sharply resource restricted devices share data acquired by their sensors and cooperate in their data processing in order to establish an intelligent and responsive smart space. Instead of being equally distributed among an environment, the processing and communicating devices are likely clustered in distant physical spaces. Consequently, for the sharply resource restricted devices it might be difficult to establish a connection among the spread clusters of devices. Natural clusters are given, for instance, by the set of devices worn or carried by a person or also by a working place constituted of a high density of electronically enhanced tools. In a smart space, clusters should be connected to share information and provide additional value to an individual in this space.

Since IoT devices (possibly featuring RFID or Organic Electronics [58]) will have a sharply restricted transmission range for their low energy budget available, these clusters, however, might be frequently disconnected. From a communication perspective, the signal strength of such resource restricted devices in one cluster might be too weak to reach a remote cluster at sufficient Signal-to-Noise Ratio (SNR). Therefore, although nodes in a remote cluster might sense some activity on the channel, the signal strength is too weak for them to decode information.

One solution to increase the transmission range of nodes in a cluster and thereby to establish a connection is to combine their transmit signals during simultaneous, phase-aligned transmission on the wireless channel.

By superimposing signals on the wireless channel in-phase, they are accumulated and therefore strengthened so that the transmission range can be extended.

In the literature, several approaches for such beamforming among distributed nodes are proposed [53, 42, 52, 59]. The most common ones require either code divisioning techniques

---

[3]Reprinted from 'Elsevier Journal on Ad Hoc Networks, vol. 16, Stephan Sigg, A fast binary feedback-based distributed adaptive carrier synchronisation for transmission among clusters of disconnected IoT nodes in smart spaces, pp. 120-130, May 2014, with permission from Elsevier.'

or for the receiving node to conduct significant computation [28, 60]. Such process, however, is exhaustive for sharply resource restricted IoT nodes in a smart space. A simpler, less resource consuming, method was proposed by Mudumbai and others in [52, 26, 54]. The authors employ an iterative random search mechanism in which nodes in each iteration may randomly change the phase of their carrier signal conditioned on a binary feedback from the receiver. This approach is better suited for IoT nodes for its low computational complexity to randomly draw alternative signal phases at nodes. Since the binary feedback can be encoded as an energy efficient on/off (burst transmission = 1, no transmission = 0) scheme, the required SNR can be low. This carrier synchronisation scheme is applicable also with inexpensive crystal oscillators with high frequency derivations [61, 62] such as we can expect for IoT devices.

For this scheme we derived a sharp asymptotic bound on the expected optimisation time for n transmit devices and k possible transmit carrier phases each, in the order of $\Theta(n \cdot k \cdot \log n)$ iterations in [1]. This performance is the main drawback of the beamforming scheme for smart spaces. Significant count of iterations and therefore a high number of transmissions are required which slows down the synchronisation [63].

In [64] an alternative asymptotically optimal, iterative optimisation approach was presented. Although its optimisation performance was as low as $\mathcal{O}(n)$, this improved performance was achieved at the cost of a more descriptive receiver feedback so that it can not be implemented as a simple on/off scheme and is therefore less well suited for resource restricted IoT nodes.

Another possibility to improve the synchronisation performance is to modify the random search for synchronised transmit phases at nodes. The original approach employed an evolutionary random search [27]. However, as indicated in [59], the search space of the problem is rather simple and does not contain any local optima.

Therefore, we propose in this paper to utilise a fast local random search to establish carrier synchronisation among nodes. In particular, we derive an asymptotic upper and lower bound for the expected optimisation time and compare the approach to the one presented in [1, 27] in simulations and a case study.

The contributions of this paper are

1. an improved and more exact upper bound for iterative feedback-based closed-loop carrier synchronisation with local random search,

2. a lower bound in the same asymptotic order,

3. a discussion of environmental impacts on the performance of iterative feedback-based carrier synchronisation,

4. simulations and

5. a case study with software-defined radio devices.

This consideration of a local random search method for feedback-based iterative carrier synchronisation improves and extends the discussion on a simple bound in [59]. Our

analysis provides an improved and more exact upper bound and in addition derives a lower bound in the same asymptotic order.

After introducing the related work and discussing iterative random carrier synchronisation in section 2.2.2 we propose a local random search mechanism and study its expected synchronisation time in section 2.2.3. In section 2.2.4 we show that the synchronisation quality of iterative feedback-based carrier synchronisation among IoT devices in a smart space is impacted by environmental stimuli. In section 2.2.5, the impact of environmental stimuli on the optimum choice of optimisation parameters is demonstrated in mathematical simulations and experimental case studies. Section 2.2.6 draws our conclusion.

2.2.2 Distributed adaptive carrier synchronisation

For distributed IoT devices in a smart space to establish a transmission beam to a remote receiver, carrier phases of transmit signals have to be synchronised with respect to the receiver location and the phase and frequency offset of the distributed local oscillators. After synchronisation, a message $m(t)$ is transmitted simultaneously by all transmit devices $i \in [1..n]$ as

$$\zeta_i(t) = \Re\left(m(t)e^{j(2\pi(f_c+f_i)t+\gamma_i)}\right) \tag{2.18}$$

so that the receiver observes the superimposed signal

$$\zeta_{\text{sum}}(t) + \zeta_{\text{noise}}(t) =$$
$$\Re\left(m(t)\sum_{i=1}^{n}\text{RSS}_i e^{j2\pi(f_c+f_i)t+(\gamma_i+\phi_i+\psi_i)}\right) + \zeta_{\text{noise}}(t) \tag{2.19}$$

with minimum phase offset between carrier signals:

$$\min\left(|(\gamma_i + \phi_i + \psi_i) - (\gamma_j + \phi_j + \psi_j)|\right) \tag{2.20}$$
$$\forall i, j \in [1..n], i \neq j.$$

In equation (2.18) and equation (2.19), f_i denotes the frequency offset of device i to a common carrier frequency f_c. The values γ_i, ϕ_i and ψ_i represent the carrier phase offset of node i as well as the phase offset in the received signal component due to the offset in the local oscillators of nodes (ϕ_i) and due to distinct signal propagation times (ψ_i). $\zeta_{\text{noise}}(t)$ denotes the noise and interference in the received sum signal. We assume additive white Gaussian noise (AWGN) here. With RSS_i we describe the received signal strength of IoT device i.

Algorithms for distributed adaptive carrier synchronisation are distinguished by closed-loop carrier synchronisation and open-loop carrier synchronisation techniques [26]. Closed-loop synchronisation can be achieved by a master-slave approach as detailed in [50]. The central idea is that transmit IoT devices send a synchronisation sequence simultaneously on code-divisioned channels to a destination device in the smart space. The destination calculates the relative phase offset of the received signals and broadcasts this information to all transmitters which then adapt their carrier signals accordingly.
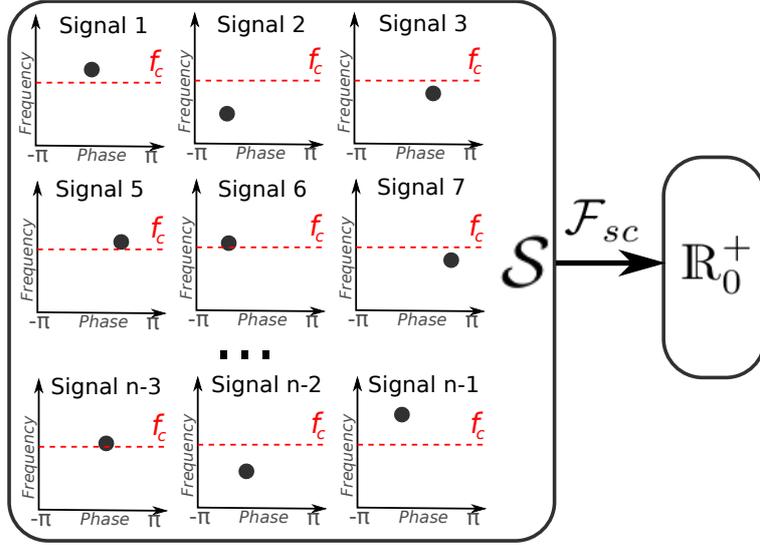
Figure 2.13: Search space and score function for binary feedback-based distributed adaptive carrier synchronisation

Due to the high computational complexity burden for the destination node to derive the relative phase offset of all received signals and for all nodes due to the utilisation of code divisioning techniques, this implementation is not suggestive for the application in smart spaces with a high count of strictly resource limited devices.

Alternatively, in a master-slave open-loop synchronisation [29], the relative phase offset among nodes is determined by the transmit nodes with a method similar to [50] but among transmit IoT devices only. The receiver then broadcasts a carrier signal once so that the transmit nodes are able to correct their frequency offsets. In this method, however, the high complexity for the nodes is shifted from the receiver node to one of the transmit nodes. Therefore, this approach also suffers from its high computational complexity.

A simpler and less resource demanding distributed carrier synchronisation scheme was proposed in [27]. This closed-loop approach is computationally cheap at the cost of increasing the time required for carrier synchronisation. It utilises a binary feedback on the achieved synchronisation quality that is transmitted in each iteration from a remote receiver [26, 51]. In particular, such binary feedback can be implemented by a simple on/off burst scheme also for sharply resource restricted IoT devices.

The central optimisation procedure consists of n devices $i \in [1, \ldots, n]$ randomly altering the phases γ_i of their carrier signal $\zeta_i(t)$ in each iteration. Implicitly, with this process a global random search is applied. The search space $\mathcal{S}$ is spanned by all possible combinations of carrier frequencies and carrier phase offsets for all transmit nodes (cf. figure 2.13).

The figure illustrates the global search space constituted from all possible combinations of phase and frequency configurations of transmit signals at local IoT nodes.

Each specific phase-frequency combination $s \in \mathcal{S}$ is associated with a score $\mathcal{F}_{sc} : \mathcal{S} \to \mathbb{R}_0^+$ that denotes its synchronisation quality. Without loss of generality we assume that the

---

Algorithm 1: Feedback-based distributed adaptive carrier synchronisation

---
1: **repeat**
2:    With probability $P_{\mathrm{mut},\gamma_i}$ and P_{mut,f_i}, each transmit node i adjusts its carrier phase offset γ_i and frequency offset f_i following a probability distribution $P_{\mathrm{dist},\gamma_i}$ (P_{dist,f_i}) with variance V_{γ_i} (V_{f_i}).
3:    Nodes transmit to the destination simultaneously as a distributed beamformer.
4:    Receiver estimates the level of phase synchronisation of the received sum signal $\zeta_{\mathrm{sum}}(t) + \zeta_{\mathrm{noise}}(t)$ (for instance by the SNR).
5:    A binary feedback (e.g. burst/no burst) indicating whether this value has improved is broadcast.
6:    When the feedback is worse than in the previous iteration, each transmit node that altered its phase offset in this iteration reverses this decision, by re-setting its carrier phase offset γ_i to the previous value.
7: **until** Sufficient synchronisation achieved

---

optimisation aim is to maximise $\mathcal{F}_{sc}$. A natural choice to compute such a score value is, for instance, the Signal-to-Noise-Ratio (SNR) of the received sum signal detailed in equation (2.19).

Feedback-based distributed carrier synchronisation approaches are characterised by the parameters

$P_{\mathbf{mut},\gamma_i}$ Probability to alter the phase-offset of device i ($P_{\mathrm{mut},\gamma_i} \in [0,1]$)

$P_{\mathbf{mut},f_i}$ Probability to alter the frequency-offset of device i ($P_{\mathrm{mut},f_i} \in [0,1]$)

$P_{\mathbf{dist},\gamma_i}$ Probability distribution (phase) for the random process at device i ($P_{\mathrm{dist},\gamma_i} \in \{\text{normal, uniform,} \dots \}$)

$P_{\mathbf{dist},f_i}$ Probability distribution (frequency) for the random process at device i ($P_{\mathrm{dist},f_i} \in \{\text{normal, uniform,} \dots \}$)

V_{γ_i} Variance for the random phase alteration process at device i ($V_{\gamma_i} \in [0,\pi]$)

V_{f_i} Variance for the random frequency alteration process at device i ($V_{f_i} \in [-f_\Delta, f_\Delta]$ for frequency range f_Δ)

The carrier synchronisation process is described by algorithm 1. Intuitively, each IoT node may in one iteration alter its transmit carrier phase offset (step 2), superimpose a synchronisation signal simultaneously with all other smart devices (step 3) and receive a binary feedback on the quality of the synchronisation (better/worse; step 5). These iterations are repeated until a random distribution of carrier phases is achieved that scores a sufficient synchronisation quality [53, 52, 65]. Initially, independent and identically distributed (i.i.d.) phase offsets γ_i of carrier signals are assumed. Since a decreasing signal quality is not accepted (cf. step 6 in algorithm 1), and since a global random search is implemented

by this approach (every possible combination of carrier phase offsets of nodes has a positive probability in each iteration) the method eventually converges to the optimum with probability 1 [52]. For this result an idealised environment without noise and interference was considered. In a realistic environment, the impact of the noise figure determines the accuracy that can be achieved.

In [66], an implementation of this carrier synchronisation approach was presented for software defined radio (SDR) devices which does not rely on any wired connections between devices (for instance, for clock synchronisation of the SDR nodes).

The authors of [53] then demonstrated in a case study that the method is feasible to synchronise frequency as well as phase of carrier signal components. Without loss of generality we will in our discussion only consider phase synchronisation and assume the frequency synchronisation as perfect. Our discussion can be easily extended to cover frequency synchronisation also by adding additional dimensions for the frequency of carrier signals to the search space $\mathcal{S}$ [53]. As an alternative, sufficiently accurate separate frequency synchronisation schemes have been discussed for this approach in [67].

The distinct implementations in the literature differ in the 2nd and the 6th step of algorithm 1. For instance, in [53, 54], devices alter their carrier phase γ_i according to a normal distribution with small variance. In [24], a uniform distribution with a small probability to alter the phase offset of one individual device is utilised instead. For a fixed uniform distribution over the whole optimisation process, a sharp asymptotic bound of $\Theta(n \cdot k \cdot \log n)$ on the expected optimisation time was derived [1]. Here, k denotes the maximum number of distinct phase offsets a physical transmitter can generate.

In all previous studies, a global random search is considered, in which nodes choose their next carrier phase and frequency offset uniformly at random from all possible values. Since the search space does not contain local optima [59] we restrict the search neighbourhood to reduce the number of possible next configurations in one iteration that would worsen the synchronisation quality. We propose to modify step 2 of algorithm 1 to follow a local random search instead of the previously applied global random search mechanism. In particular, an IoT node i will, when it changes its phase and frequency offset, draw the new values from a restricted neighbourhood of size $\mathcal{N}$ that is centred around the current values of γ_i (and f_i). This addresses a recent critique expressed in [63] regarding the convergence speed for this binary feedback-based iterative adaptive carrier synchronisation. In section 2.2.3 we derive upper and lower bounds on the expected synchronisation performance of a local random search mechanism for feedback-based distributed adaptive carrier synchronisation. These bounds improve the existing bounds known for the global random search method. Section 2.2.4 shows that the optimum values for the parameters $P_{\mathrm{mut},\gamma_i}$, P_{mut,f_i}, $P_{\mathrm{dist},\gamma_i}$, P_{dist,f_i}, V_{γ_i} and V_{f_i} are conditioned on environmental situations in a smart space.

2.2.3 Local random search

Recent approaches to 1-bit feedback-based distributed carrier synchronisation utilise a global random search that reaches any search point $s \in \mathcal{S}$ with a positive probability in
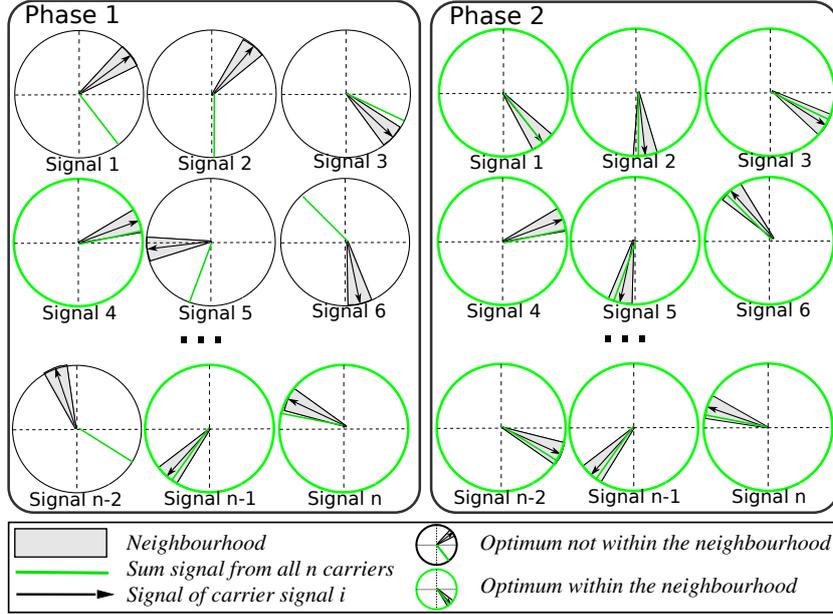
Figure 2.14: Phase 1 and phase 2 of the synchronisation process

each iteration [53, 52, 59, 54]. The probability to achieve by these random phase and frequency perturbations of s a search point s' with $\mathcal{F}_{sc}(s') \geq \mathcal{F}_{sc}(s)$ decreases with increasing synchronisation quality ($\mathcal{F}_{sc}(s)$ score). A restricted search neighbourhood can, however, ensure a constant steady progress since the search space does not contain local optima as derived in [1]. Any local search heuristic that manages to follow a path with increasing $\mathcal{F}_{sc}$ score will find a global optimum with probability 1. We assume that each transmit node is able to apply k distinct phase-offsets and define a global optimum as superimposition of transmit signals in which all phases are within $\frac{2\pi}{k}$ of a superimposition with perfect phase coherency.

Theorem 2.2.1. *Let $s \in \mathcal{S}$ be a current search point of a local random search algorithm $\mathcal{A}$ for feedback-based distributed adaptive carrier synchronisation with neighbourhood size $\mathcal{N} \leq |\mathcal{S}|$ and $\forall i : P_{dist,\gamma_i} = uniform$. For each phase and frequency perturbation (step 2 in algorithm 1) of one transmit carrier signal, the probability to arrive at a search point (superimposition of transmit signals) s' with $\mathcal{F}_{sc}(s') \geq \mathcal{F}_{sc}(s)$ is at least $\frac{1}{2}$ for each transmit signal that has the optimum search point s^* not within its neighbourhood $\mathcal{N}$.*

(Refer to the Appendix for a proof of the results)

Consequently, we divide the following analysis into two phases. In the first, the optimum is not within the neighbourhood of at least one node so that at least one node can improve the fitness with probability $\frac{1}{2}$ or more (cf. figure 2.14).

In the second phase, all nodes have the optimum within their neighbourhood. The probability to decrease the distance to the optimum might then be worse than $\frac{1}{2}$. An optimisation process with restricted neighbourhood-size therefore has a probability of $\frac{1}{2}$
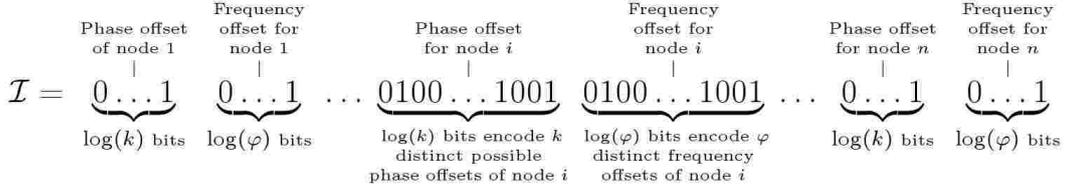
Figure 2.15: Binary representation of search points as a concatenation of grey encoded phase and frequency offsets

to increase the fitness-value for a long time until the optimum point is within the neighbourhood of each single carrier signal. The price for this high probability to improve the fitness-value in each iteration is that the chance to achieve great progress in one step (as possible with an unrestricted neighbourhood-size) is lost. Since this event is significantly less probable, we are prepared to pay this price. An individual node i then alters the phase-offset γ_i of its carrier-signal uniformly at random within a range of $\left[\gamma_i - \frac{\mathcal{N}}{2}, \gamma_i + \frac{\mathcal{N}}{2}\right]$ for suitable $\mathcal{N}$. To simplify the analysis we represent search points (superimpositions of transmit signals) in a binary encoding. Figure 2.15 sketches this encoding.

We assume that each transmit node is able to apply k distinct phase-offsets. We encode a search point $s \in \mathcal{S}$ represented by $n \cdot k$ distinct phase-offsets as binary string of length $n \cdot \log(k)$ [4]. We assume that configurations are encoded so that their Hamming-distance increases with increasing difference in phase-offsets [68, 69]. We analyse the count of bit mutations of this bit-string until an encoding of a global optimum is found. We choose the probability to alter a bit in the binary sequence as $\frac{1}{n \cdot \mathcal{N}}$ for n nodes with neighbourhood size $\mathcal{N}$. For the binary representation this is analogue to having a probability of $\frac{1}{n}$ for each signal to alter its phase uniformly at random within the $\mathcal{N}$ possible values in $\left[\gamma_i - \frac{\mathcal{N}}{2}, \gamma_i + \frac{\mathcal{N}}{2}\right]$. Then, one node on average alters its phase offset within the neighbourhood boundaries in each iteration. With Chernoff bounds we can show that with high probability the Hamming-distance to an optimum configuration of these offsets for all carrier signals is not much smaller than $\frac{n \cdot \log(k)}{2}$.

Theorem 2.2.2. *For a network of n transmit and one receive node, let $\mathcal{N}$ be the neighbourhood size of a local random search method $\mathcal{A}$ for feedback-based distributed adaptive carrier synchronisation with $\forall i: P_{dist,\gamma_i} = uniform, P_{mut,\gamma_i} = \frac{1}{n}$. Further assume that each node is capable of transmitting signals at up to k distinct carrier phases and that new carrier phases are drawn uniformly at random from the neighbourhood. The expected number of iterations for distributed adaptive carrier synchronisation is bounded by*

$$\mathcal{O}\left(n \cdot \mathcal{N} \cdot \log(n) + \frac{\log(k)}{\mathcal{N}}\right). \tag{2.21}$$

(Refer to the Appendix for a proof of the results)

---

[4]When distinct frequency offsets are also considered, a search point $s \in \mathcal{S}$ would be represented by $n \cdot k \cdot f$ distinct phase- and frequency-offsets as binary string of length $n \cdot \log(k) \cdot \log f$

Theorem 2.2.3. *For a network of n transmit and one receive node, let $\mathcal{N}$ be the neighbourhood size of a local random search method $\mathcal{A}$ for feedback-based distributed adaptive carrier synchronisation with $\forall i : P_{dist,\gamma_i} = uniform, P_{mut,\gamma_i} = \frac{1}{n}$. Further assume that each node is capable of transmitting signals at up to k distinct carrier phases and that new carrier phases are drawn uniformly at random from the neighbourhood of size $\mathcal{N}$. For a suitable Δ the expected number of iterations for distributed adaptive carrier synchronisation is bounded by*

$$\Omega(n \cdot \mathcal{N} \cdot \Delta)$$

(Refer to the Appendix for a proof of the results)

Theorem 2.2.4. *For a network of n transmit and one receive node, let $\mathcal{N}$ be the neighbourhood size of a local random search method $\mathcal{A}$ for feedback-based distributed adaptive carrier synchronisation with $\forall i : P_{dist,\gamma_i} = uniform, P_{mut,\gamma_i} = \frac{1}{n}$. Further assume that each node is capable of transmitting signals at up to k distinct carrier phases and that new carrier phases are drawn uniformly at random from the neighbourhood of size $\mathcal{N}$. The expected number of iterations for distributed adaptive carrier synchronisation is bounded by*

$$E[T_{\mathcal{P}}] = \Theta\left(n \cdot \mathcal{N} \cdot \log(n) + \frac{\log(k)}{\mathcal{N}}\right). \tag{2.22}$$

(Refer to the Appendix for a proof of the results)

Observe that equation (2.22) evolves to

$$\Theta\left(n \cdot \log(n) + \log(k)\right) \tag{2.23}$$

for $\mathcal{N} \to 1$ and to

$$\Theta\left(n \cdot k \cdot \log(n)\right) \tag{2.24}$$

for $\mathcal{N} \to k$. Equation (2.24) is identical to the bound derived on the expected optimisation time of the global random search method [59], where in fact the neighbourhood size is $\mathcal{N} = k$. Observe that it is more beneficial to have a smaller local search neighbourhood than to utilise a global random search method with $\mathcal{N} = k$.

The minimum value for $E[T_{\mathcal{P}}]$ is achieved for $\mathcal{N} \to 1$. Since this is an asymptotic consideration, the optimum absolute value for $\mathcal{N}$ depends on the choice of n and k.

2.2.4 Environmental impacts

The performance of the local random search guided carrier synchronisation is impacted by P_{mut,γ_i}, P_{mut,f_i}, P_{dist,γ_i}, P_{dist,f_i}, V_{γ_i} and V_{f_i}. In [1] we observed that a good synchronisation quality is achieved when either P_{mut,γ_i}, P_{mut,f_i}, V_{f_i} or V_{γ_i} are small, so that the search space is propagated in rather small steps, eventually approaching the optimum. This observation also agrees with our discussion in the last section. However, the environment may impact the optimum value for these parameters. We discuss three possible impacts, namely the number of participating nodes, the noise figure and movement of devices.
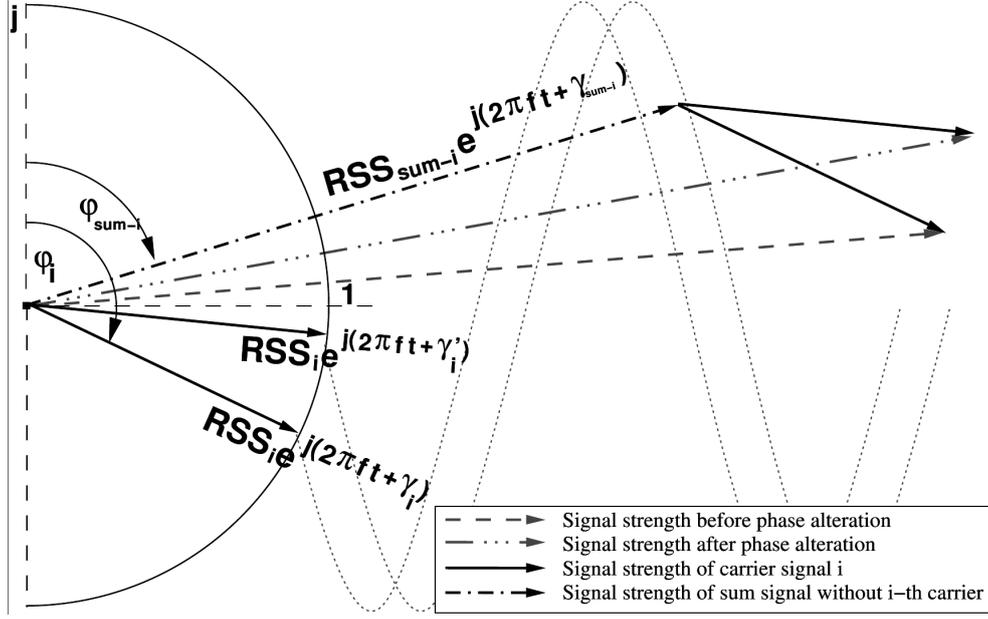
Figure 2.16: Illustration of the impact of carrier phase alteration on the overall received signal strength

Impact of noise and interference

The signal observed by a receiver is composed of the signal $\zeta_{\text{sum}}(t)$ and noise $\zeta_{\text{noise}}(t)$ (cf. equation (2.19)). Noise and interference might differ due to opened windows or doors, people moving or other nearby electronic devices [70, 71, 3]. The impact of the phase alteration of a single link $i \in [1..n]$ on the SNR of $\zeta_{\text{sum}}(t) + \zeta_{\text{noise}}(t)$ is not greater than $2 \cdot \text{RSS}_i$. This is the case when the phase of signal $\zeta_i(t)$ and the sum signal $\zeta_{\text{sum-i}}(t)$ without the signal of i have been separated in phase by π before γ_i is then shifted by π. With the cosine rule we can calculate the change in the received signal strength of the received superimposed signal at the event of a change of the carrier phase from γ_i to γ_i' as

$$
\begin{aligned}
\overline{\text{RSS}}(\gamma_i, \gamma_i') = & \\
& \sqrt{\text{RSS}_{\text{sum-i}}^2 + \text{RSS}_i^2 - 2\text{RSS}_{\text{sum-i}}\text{RSS}_i \cos(\gamma_i + \gamma_i')} \\
& - \sqrt{\text{RSS}_{\text{sum-i}}^2 + \text{RSS}_i^2 - 2\text{RSS}_{\text{sum-i}}\text{RSS}_i \cos(\gamma_i)}
\end{aligned} \tag{2.25}
$$

as illustrated in figure 2.16. In this equation we denote the received signal strength achieved by the superimposition of all signals short of i by

$$
\text{RSS}_{\text{sum-i}} = \sum_{\bar{i} \neq i} \text{RSS}_{\bar{i}} e^{j(2\pi(f_c + f_{\bar{i}})t + \gamma_{\bar{i}} + \phi_{\bar{i}} + \psi_{\bar{i}})}; \bar{i} \in [1..n]. \tag{2.26}
$$

Since the phase alteration is a random process, the actual gain of a single phase modification is typically smaller than the maximum possible value. When we assume that a

single node can establish up to k equally probable carrier phases, the average gain by the alteration of one carrier signal is then

$$\frac{\sum_{i=1}^{k} \overline{\text{RSS}}\left(\gamma_i, \gamma_i + \frac{2\pi}{k} \cdot i\right)}{k}. \tag{2.27}$$

Consequently, when the noise figure is in the same order or greater, alterations of individual carriers have little effect. In such a situation it is beneficial to increase the average distance of consecutive search points in the search space in a single iteration. This can be achieved by increasing the variance V_{γ_i}, the neighbourhood size $\mathcal{N}$ or the probability P_{mut,γ_i} (cf. section 2.2.5 and section 2.2.5).

Impact of the network size

The number of nodes that participate also impacts the performance. Since the synchronisation is achieved by a random process over all possible combinations of phase and frequency offsets, the synchronisation time is increased with the count of nodes [29]. The optimum performance is achieved with small P_{mut,γ_i}, P_{mut,f_i}, V_{f_i} and V_{γ_i} [59]. On the other hand, the relative impact of an individual node on $\zeta_{\text{sum}}(t)$ decreases with increasing node count. We can see this again from equation (2.25). The value $\overline{\text{RSS}}(\gamma_i, \gamma_i')$ decreases with increasing $\text{RSS}_{\text{sum-i}}$. With increasing node count n, it is therefore beneficial to chose P_{mut,γ_i}, P_{mut,f_i}, V_{γ_i} and V_{f_i} slightly higher than $\frac{1}{n}$ in order to increase the impact of modifications in one iteration (cf. section 2.2.5).

Impact of node mobility

Movement impacts the synchronisation of nodes since phases drift apart when the receiver or transmit nodes move [64]. Synchronisation has to be significantly faster than the velocity experienced. An increased value for P_{mut,γ_i}, P_{mut,f_i}, V_{γ_i} or V_{f_i} might therefore be beneficial in the presence of node mobility (cf. section 2.2.5).

2.2.5 Simulation and case studies

In a Matlab-based simulation, up to 100 IoT devices are distributed uniformly at random on a 4 m×6 m square area (e.g. spread across a wall in a factory building) with a remote IoT receiver located up to 11 m in orthogonal direction from the centre of this area. Frequency and phase stability are considered perfect. We calculate the phase offset of the received dominant signal component from each transmitter according to the transmission distance in a direct line of sight. Path loss was calculated by the Friis free space equation $P_{tx}\left(\frac{\lambda}{2\pi d}\right)^2 G_{tx} G_{rx}$ with antenna gain for transmitter and receiver as $G_{rx} = G_{tx} = 0$ dB. Signals are transmitted at 2.4 GHz with transmit power $P_{tx} = 1$ mW. All received signal components calculated in this manner are then summed up in order to achieve the

superimposed sum signal

$$\zeta_{\text{sum}}(t) = \sum_{i=1}^{n} \left(\Re \left(m(t)\text{RSS}_i e^{j(2\pi(f_c+f_i)t+\gamma_i+\phi_i+\psi_i)} \right) \right). \tag{2.28}$$

Finally, a noise signal $\zeta_{\text{noise}}(t)$ is added to $\zeta_{\text{sum}}(t)$ in order to estimate the signal at the receiver. We utilise AWGN at -103 dBm as proposed in [57]. For a given configuration we repeated each simulation 10 times with identical parameters.

Each simulation lasts for 6000 iterations. Signal quality of a signal during the synchronisation phase is measured by the Root of the Mean Square Error (RMSE) of the received signal $\zeta_{\text{sum}}(t)$ to an expected optimum signal $\zeta_{\text{opt}}(t)$:

$$RMSE = \sqrt{\sum_{t=0}^{\varrho} \frac{\left(\zeta_{\text{sum}}(t) + \zeta_{\text{noise}}(t) - \zeta_{\text{opt}}(t) \right)^2}{n}} \tag{2.29}$$

In equation (2.29), ϱ is chosen to cover several signal periods.

The optimum signal $\zeta_{\text{opt}}(t)$ is calculated as perfectly aligned and properly phase shifted received sum signal from all transmit sources. For the optimum signal, noise is disregarded.

Performance of a local random search

We implement a local random search with neighbourhood radius $\frac{\mathcal{N}}{2} \in [0, \pi]$ where each node $i \in [1, n]$ alters the phase offset γ_i of its carrier signal $\zeta_i(t)$ with probability $\frac{1}{n}$ to $\gamma_i' \in [\gamma_i - \frac{\mathcal{N}}{2}, \gamma_i + \frac{\mathcal{N}}{2}]$. Figure 2.17 depicts the performance of the algorithm with a neighbourhood size of $\mathcal{N} = 0.6\pi$ compared to a global random search approach ($\mathcal{N} = 2\pi$). We observe that, although the local random search method naturally has a slower start than the global random search method, it then reaches lower RMSE values faster.

In particular, in the critical part ot the optimisation, the RMSE values reached by the local random search approach are reached only about 400-500 iterations later by the global random search method. Due to noise and therefore a general saturation of the optimisation process, the synchronisation quality is not much improved afterwards so that the global random search eventually catches up.

Case study with software defined radio devices

We approximated realistic conditions in an experimental setting with Universal Software Radio Peripheral (USRP) software radios[5] to represent distributed devices.

Three USRP devices have been utilised as transmitters with one device as receiver. In order to achieve identical transmit frequencies among devices, the clock of the first transmit device was utilised as reference for the other two transmitters. The clock of the receiver node was, however, independent. Alternatively, clocks might be synchronised via GPS or by the iterative frequency synchronisation approach described in [53]. Table 2.4 summarises our experimental configuration.

---

[5]http://www.ettus.com
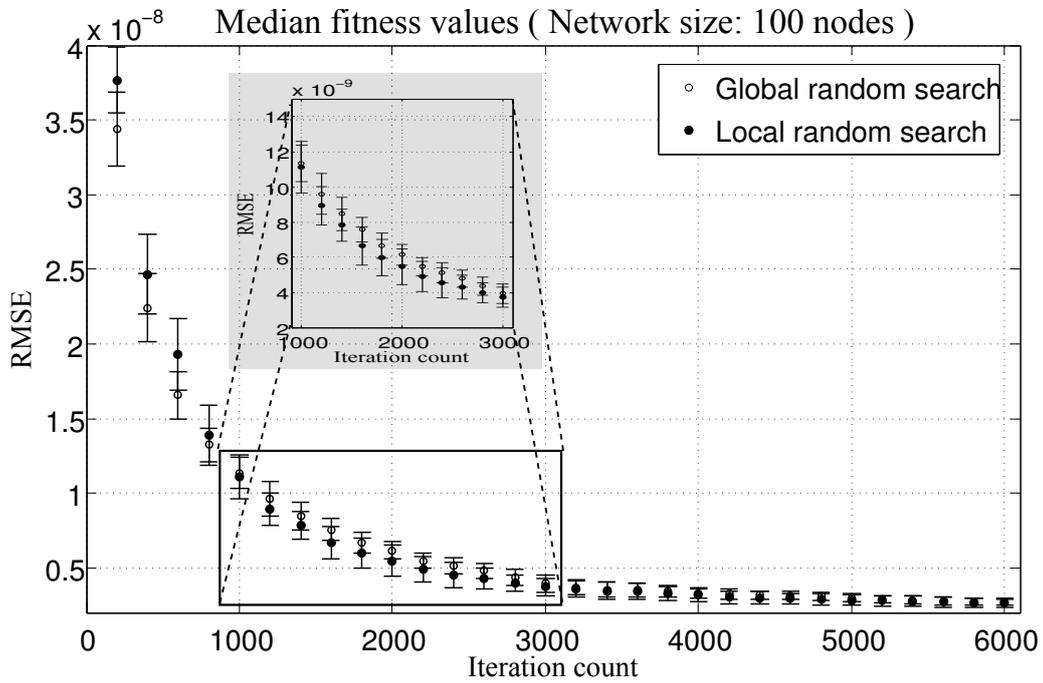
Figure 2.17: Comparison of the performance achieved by global and local random search

Table 2.4: Configuration of the experimental case study

Experimental setting	
Separation of antennas [m]	0.44
Distance to receive antenna [m]	5.5 / 11 / 16.4
Transmit frequency [MHz]	$f_{TX} = 2400$
Receive frequency [MHz]	$f_{RX} = 902$
Iterations per experiment	400
Mobility	stationary
Identical experiments	12
Transmit devices	3
Receive devices	1
Algorithmic configuration	
Random distribution of the phase alteration	normal
Phase alteration probability $P_{\mathrm{mut},\gamma_i}$	0.33/0.66/1.0
Variance V_{γ_i} [π]	0.25 / 1

Figure 2.18: Illustration of the experimental setting utilised. Devices are placed on tables with a height of 72 cm.

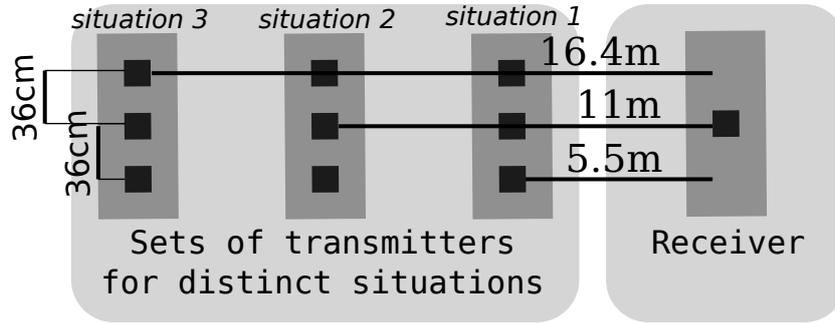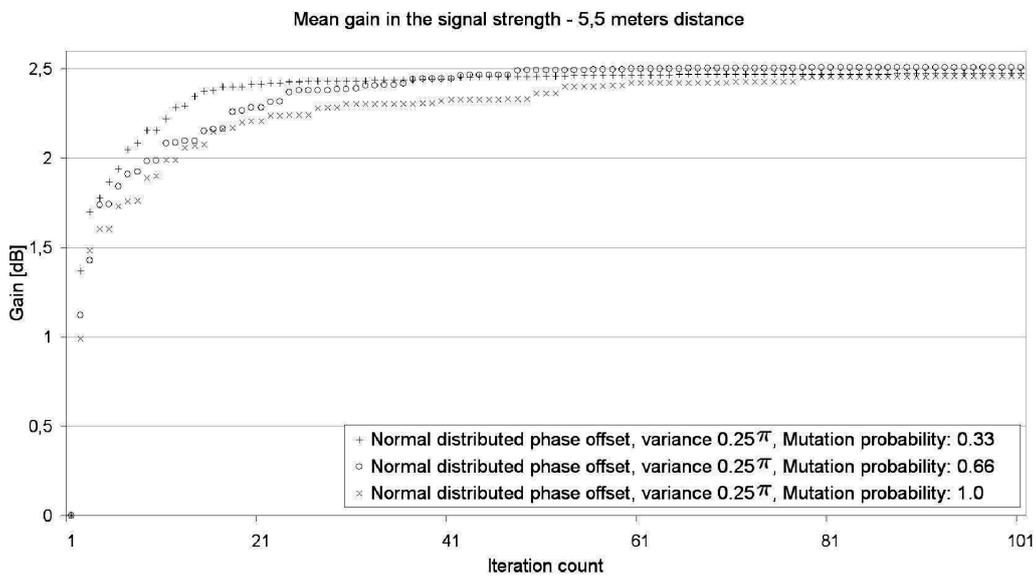Figure 2.19: Mean gain in the signal strength at a transmission distance of 5.5 meters and a variance of the random process of 0.25π

The experimental setting is sketched in figure 2.18. For the three different situations, the transmit nodes were moved to various distances accordingly. We modified the probability to alter the phase offset of one device and the variance for its normal distributed random phase perturbation process as well as the distance between transmit and receive devices to account for distinct environmental situations. The three transmit devices were controlled by a single computer running three independent and non-communicating processes. The receiver device was controlled by a second computer. During the experiments the room was vacated so that no movement or presence of individuals could impact the synchronisation process.

Results derived in these experiments are depicted in figure 2.19 and figure 2.20.

The figure displays the mean gain in signal strength compared to an unsynchronised transmission at the beginning of the synchronisation. The synchronisation performance
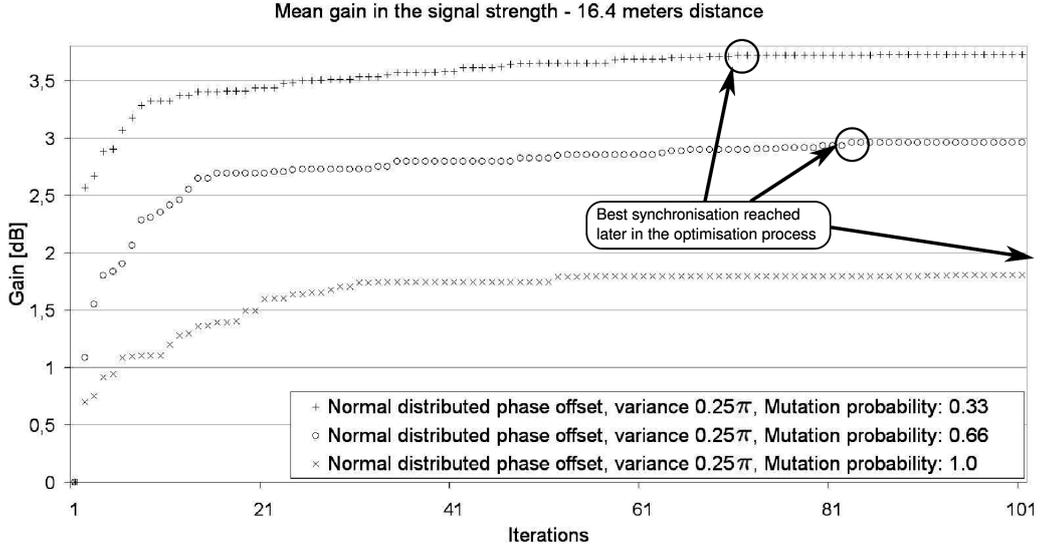
Figure 2.20: Mean gain in the signal strength at a transmission distance of 16.4 meters and a variance of the random process of 0.25π

differs for different environmental situations. When the transmission distance increases, the relative noise figure also increases. The best synchronisation is then generally reached later in the synchronisation process. For instance, in figure 2.19, at a transmission distance of 5.5 meters, the best value is reached after about 40 to 50 iterations. In figure 2.20 (16.4 meters) we observe that the optimum synchronisation is reached after about 60 to 80 iterations. Also, the choice of the optimum configuration differs dependent on the scenario. In figure 2.20 at a distance of 16.4 meters, a variance of 0.25π and a probability to alter the phase offset of $P_{\mathrm{mut},\gamma_i} = 0.33$ achieves the best results. At shorter distances, the configuration with $P_{\mathrm{mut},\gamma_i} = 0.66$ results in a slightly better synchronisation performance. For the variance, a similar effect was not observed.

2.2.6 Conclusion

We analysed and evaluated a local random search-based approach to distributed adaptive carrier synchronisation for IoT nodes in a smart space with an iterative feedback-based carrier synchronisation method. We derived a sharp asymptotic bound of

$$E[T_{\mathcal{P}}] = \Theta\left(n \cdot \mathcal{N} \cdot \log(n) + \frac{\log(k)}{\mathcal{N}}\right)$$

on the expected synchronisation performance. This bound is significantly lower than the expected synchronisation performance derived recently for a global random search heuristic for this carrier synchronisation method. Intuitively, although the global random search approach has, unlike the local random search, a positive (but very small) probability to reach a global optimum in each iteration, its probability to generally reach any point

that would improve the synchronisation quality decreases with increasing synchronisation quality. For the local random search, however, we could show that there is at least one node that would improve the synchronisation with probability not smaller than $\frac{1}{2n}$ for a long time during the synchronisation process.

Also, we discussed the impact of environmental effects on the synchronisation performance. In particular, the relative noise figure, the count of participating devices and the mobility of nodes have been identified to impact the synchronisation performance. However, by changing the probabilities $P_{\mathrm{mut},\gamma_i}$, P_{mut,f_i} to alter the phase offset or the variance V_{γ_i}, V_{f_i} for each node i, the synchronisation approach can be adapted to these environmental impacts.

We presented simulations and case studies with software defined radios on the iterative feedback-based carrier phase synchronisation by a local random search approach that also showed an improved performance compared to the global random search approach and an effect of the distance between nodes on the synchronisation performance.

Appendix − Proofs

Proof of Theorem 2.2.1. We see this from figure 2.21. In the figure, a vector of all carrier signals short of signal i is denoted by

$$\zeta_{\mathrm{sum}-i}(t) = \mathrm{RSS}_{\mathrm{sum}-i}e^{j(2\pi ft+\gamma_{\mathrm{sum}-i})}. \tag{2.30}$$

The vector associated with carrier i is identified by

$$\zeta_i(t) = \mathrm{RSS}_i e^{j(2\pi ft+\gamma_i)}. \tag{2.31}$$

When a single carrier signal $\zeta_i(t)$ is modified within the neighbourhood of $\mathcal{N}$, this means that $\zeta_i(t)$ is rotated by ν or $-\nu$ with $\nu \in \left[0, \frac{\mathcal{N}}{2}\right]$. As long as $\zeta_{\mathrm{sum}-i}(t)$ is not within the neighbourhood, a rotation of ν will increase (decrease) the amplitude of $\zeta_{\mathrm{sum}-i}(t) + \zeta_i(t)$ while a rotation by $-\nu$ will decrease (increase) it. The probability to improve the fitness value is $\frac{1}{2}$ when carrier phase offsets are chosen uniformly at random. $\square$

Proof of Theorem 2.2.2. We divide the analysis into two phases (cf. theorem 2.2.1). In the first phase, at least one node does not have the optimum phase offset within its neighbourhood. Then, there is always at least one node that will by altering its carrier phase improve the synchronisation with probability at least $\frac{1}{2}$. The probability that in one iteration one such node alters its phase offset while all other $n-1$ nodes do not change it is at least

$$\frac{1}{n} \cdot \left(1 - \frac{1}{n}\right)^{n-1} \geq \frac{1}{e \cdot n}. \tag{2.32}$$

We define the expected progress as the expected count of bits in the binary representation of search points that are altered in one iteration. Since a new search point is drawn uniformly at random from all possible values in the neighbourhood of size $\mathcal{N}$, the expected

progress when a node that has not the optimum within its neighbourhood randomly alters its carrier phase and improves the overall synchronisation is therefore at most

$$\frac{1}{2} \cdot \frac{\mathcal{N}}{2} \cdot e \cdot n = \frac{e \cdot n \cdot \mathcal{N}}{4}. \tag{2.33}$$

The expected upper bound on the iterations required to reach a global optimum is then determined by the maximum distance to an optimum. The Hamming-distance to a binary representation that describes a global optimum is $n \cdot \log(k)$ at most. Consequently, the expected number of these iterations until a binary representation is found for which all nodes have the optimum within their neighbourhood is at most

$$\frac{n \cdot \log(k) \cdot 4}{\mathcal{N} \cdot e \cdot n} = \mathcal{O}\left(\frac{\log(k)}{\mathcal{N}}\right). \tag{2.34}$$

In the second phase, each node has the optimum carrier phase offset within the neighbourhood around its current carrier phase. Assume that a set of i nodes has already reached an optimum synchronisation of their carrier phases. In this case, the probability that one of the $n - i$ nodes which have not yet found the optimum phase offset applies a correct mutation which would alter the carrier phase to the optimum value with respect to all other carrier phases is

$$\binom{n - i}{1} \cdot \frac{1}{n} \cdot \frac{1}{\mathcal{N}} \cdot \left(1 - \frac{1}{n}\right)^{n-1} \geq \frac{n - i}{n \cdot \mathcal{N} \cdot e}. \tag{2.35}$$

In equation (2.35), the term $\binom{n - i}{1} \cdot \frac{1}{n}$ describes the number of possible cases that one node out of $n - i$ nodes which are not yet perfectly synchronised alters its phase offset with probability $\frac{1}{n}$. Since all phases are with equal probability drawn from the Neighbourhood of size $\mathcal{N}$, this alteration leads to the one optimum phase offset within the neighbourhood with probability $\frac{1}{\mathcal{N}}$. The term $\left(1 - \frac{1}{n}\right)^{n-1}$ describes the probability that all other $n - 1$ nodes do not alter their phase offset in this iteration. When this event happens $n - 1$ times for each possible number of already synchronised nodes ($n - i$ with $i \in [1..n]$), the carrier phase offsets of all nodes are finally synchronised. Therefore, an upper bound on the synchronisation time in the second phase is given by

$$
\begin{aligned}
&\sum_{i=0}^{n-1} \frac{n \cdot \mathcal{N} \cdot e}{n - i} \\
= &\sum_{i=1}^{n} \frac{n \cdot \mathcal{N} \cdot e}{i} \\
= &\mathcal{O}\left(n \cdot \mathcal{N} \cdot \log(n)\right).
\end{aligned} \tag{2.36}
$$

Overall, the expected asymptotic synchronisation time is then

$$\mathcal{O}\left(n \cdot \mathcal{N} \cdot \log(n) + \frac{\log(k)}{\mathcal{N}}\right). \tag{2.37}$$

$\square$

Proof or Theorem 2.2.3. After initialisation, the phases of the carrier-signals are identically and independently distributed. Consequently for a superimposed received sum-signal $\zeta_{\text{sum}}(t)$, each of the $n \cdot \log(k)$ bits in the binary string $s_{\zeta_{\text{sum}}}$ that represents the corresponding search-point has an equal probability to be 1 or 0. The probability to start from a search-point $s_{\zeta_{\text{sum}}}$ with Hamming-distance $h(s_{\text{opt}}, s_{\zeta_{\text{sum}}})$ not larger than $l \in \mathbb{N}$; $l \ll n \cdot \log(k)$ to one of the global optima s_{opt} is at most

$$
\begin{aligned}
P[h(s_{\text{opt}}, s_{\zeta_{\text{sum}}}) \leq l] \quad &= \quad \sum_{i=0}^{l} \binom{n \cdot \log(k)}{n \cdot \log(k) - i} \cdot \frac{k}{2^{n \cdot \log(k) - i}} \\
&\leq \quad \frac{k \cdot (n \cdot \log(k))^{l+2}}{2^{n \cdot \log(k) - l}}
\end{aligned}
\tag{2.38}
$$

In this formula,

$$
\binom{n \cdot \log(k)}{n \cdot \log(k) - i}
\tag{2.39}
$$

is the count of possible configurations with i bit-errors to a given global optimum, $\frac{1}{2^{n \cdot \log(k) - i}}$ represents the probability for all these bits to be identical to the respective bits in one of the k global optima. Observe that we have a global optimum for each possible k phase offsets since from one global optimum we reach an arbitrary other global optimum by shifting the carrier phases of all nodes by an equal amount.

This means that with high probability the Hamming-distance to the nearest global optimum is at least l. We will use the method of the expected progress to calculate a lower bound on the optimisation-time required to flip these l bits. The general idea is the following.

Let $(s_{\zeta_{\text{sum}}}, \tau)$ denote the situation that search-point $s_{\zeta_{\text{sum}}}$ was achieved after τ iterations of the algorithm. We define a progress measure $\Lambda : \left(\mathbb{B}^{n \cdot \log(k)}, t \right) \to \mathbb{R}_0^+, t \in \mathbb{N}$ such that $\Lambda(s_{\zeta_{\text{sum}}}, \tau) < \Delta$ represents the case that a global optimum was not found in the first τ iterations. For every $\tau \in \mathbb{N}$ we have

$$
\begin{aligned}
E[T_{\mathcal{P}}] \quad &\geq \quad \tau \cdot P[T_{\mathcal{P}} > \tau] \\
&= \quad \tau \cdot P[\Lambda(s_{\zeta_{\text{sum}}}, \tau) < \Delta] \\
&= \quad \tau \cdot (1 - P[\Lambda(s_{\zeta_{\text{sum}}}, \tau) \geq \Delta]).
\end{aligned}
\tag{2.40}
$$

With the help of the Markov-inequality we obtain

$$
P[\Lambda(s_{\zeta_{\text{sum}}}, \tau) \geq \Delta] \leq \frac{E[\Lambda(s_{\zeta_{\text{sum}}}, \tau)]}{\Delta}
\tag{2.41}
$$

and therefore

$$
E[T_{\mathcal{P}}] \geq \tau \cdot \left(1 - \frac{E[\Lambda(s_{\zeta_{\text{sum}}}, \tau)]}{\Delta} \right).
\tag{2.42}
$$

This means that we can obtain a lower bound on the optimisation-time by providing the expected progress after τ iterations. The probability for l bits to correctly flip is at most

$$
\left(1 - \frac{1}{n \cdot \mathcal{N}} \right)^{n \cdot \mathcal{N} - l} \cdot \left(\frac{1}{n \cdot \mathcal{N}} \right)^l \leq \frac{1}{(n \cdot \mathcal{N})^l}.
\tag{2.43}
$$

In this formula, $\left(1 - \frac{1}{n \cdot \mathcal{N}}\right)^{n \cdot \mathcal{N} - l}$ describes the probability that all 'correct' bits do not flip while the remaining l bits mutate with probability $\left(\frac{1}{n \cdot \mathcal{N}}\right)^l$.

This means that for all n nodes in one iteration on average 1 bit flips inside their neighbourhood of size $\mathcal{N}$. The expected progress in one iteration is therefore

$$E[\Lambda(s_{\zeta_{\mathrm{sum}}}, \tau), \Lambda(s_{\zeta'_{\mathrm{sum}}}, \tau + 1)] \;\leq\; \sum_{i=1}^{l} \frac{i}{(n \cdot \mathcal{N})^i} < \frac{2}{n \cdot \mathcal{N}} \tag{2.44}$$

and the expected progress in τ iterations is consequently not greater than $\frac{2\tau}{n \cdot \mathcal{N}}$.

When we choose $\tau = \frac{n \cdot N \cdot \Delta}{4} - 1$, the double of the expected progress is still smaller than Δ. With the Markov inequality we can show that this progress is not achieved with probability $\frac{1}{2}$. Altogether we conclude that the expected optimisation-time is bound from below by

$$
\begin{aligned}
E[T_{\mathcal{P}}] \;&\geq\; \tau \cdot \left(1 - \frac{E[\Lambda(s_{\zeta_{\mathrm{sum}}}, \tau)]}{\Delta}\right) \\
&\geq\; \frac{n \cdot \mathcal{N} \cdot \Delta}{4} \cdot \left(1 - \frac{\frac{2 \cdot n \cdot \mathcal{N}}{4 \cdot n \cdot \mathcal{N}} \cdot \Delta}{\Delta}\right) \\
&=\; \Omega(n \cdot \mathcal{N} \cdot \Delta)
\end{aligned} \tag{2.45}
$$

$\square$

Proof of Theorem 2.2.4. The asymptotically sharp bound is a result of theorem 2.2.2 and theorem 2.2.3 with

$$\Delta = \log(n) + \frac{\log(k)}{n \cdot \mathcal{N}^2}. \tag{2.46}$$

$\square$

2.3 RF-sensing of activities from non-cooperative subjects in device-free recognition systems using ambient and local signals [6]

We consider the detection of activities from non-cooperating individuals with features obtained on the Radio Frequency channel. Since environmental changes impact the transmission channel between devices, the detection of this alteration can be used to classify environmental situations. We identify relevant features to detect activities of non-actively transmitting subjects. In particular, we distinguish with high accuracy an empty environment or a walking, lying, crawling or standing person, in case-studies of an active, device-free activity recognition system with software defined radios. We distinguish between two cases in which the transmitter is either under the control of the system or ambient. For activity detection the application of one-stage and two-stage classifiers is considered. Apart from the discrimination of the above activities, we can show that a detected activity can also be localised simultaneously within an area of less than 1 meter radius.

2.3.1 Introduction

In the approaching Internet of Things (IoT), virtually all entities in our environment will be enhanced by sensing, communication and computational capabilities [72, 73]. These entities will provide information on environmental situations, interact in the computation and processing of data [9] and store information. In order to sense environmental situations, common sensors in current applications are light, movement, pressure, audio or temperature [74]. Clearly, for reasons of cost and sensor size it is desired to minimise the count of distinct sensors in IoT entities. The one sensor class that defines the minimum set naturally available in virtually all IoT devices is the Radio Frequency (RF)-transceiver to communicate with other wireless entities [75]. It is also shipped with nearly every contemporary electronic device like mobile phones, notebooks, media players, printers as well as keyboards, mouses, watches, shoes and rumour has spread about even media cups. Therefore, the RF transceiver is a ubiquitously available sensor class. It is capable of sensing changes or fluctuation in a received RF-signal. Radio waves are blocked, reflected or scattered at objects. At a receiver, the signal components from distinct signal paths add up to form a superimposition. When objects that block or reflect the signal path of some of these signal components are moved, this is reflected in the superimposition of signal waves at the receiver. We assert that specific activities in the proximity of a receiver generate characteristic patterns in the received superimposed RF-signal. By identifying and interpreting these patterns, it is possible to detect activities of non-cooperating subjects in an

RF-receiver's proximity.

Although the wireless channel is occasionally utilised for location detection of other RF devices [76, 77] or passive entities [78, 79], it is seldom used to detect other contexts like activities from entities which are not equipped with a RF-transceiver.

We consider the detection of activities of device-free entities from the analysis of RF-channel fluctuations induced by these very activities. In analogy to the definition of device-free radio-based localisation systems (DFL) [78] we define device-free radio-based activity recognition systems (DFAR) as *systems which recognise the activity of a person using analysis of radio signals while the person itself is not required to carry a wireless device* (cf. [80]). In addition to the sensor type employed, we further categorise radio-based activity recognition systems by the parameters enlisted in table 2.5. In particular, we distinguish between passive and active systems depending on whether a transmitter is part of and under control of the radio-based recognition system. Also, an ad-hoc system can be installed in a new environment without re-training the classifier, while a non-ad-hoc system requires initial training or configuration.

In this work, we focus on the detection of static and dynamic activities of single individuals by active and passive, non-ad-hoc DFAR systems. The active system employs a dedicated transmitter as part of the recognition hardware while the passive system utilises solely ambient FM radio from a transmitter not under the control of the system. Compared to preliminary work on RF-based activity recognition [81, 82, 70, 71, 75], the novel contributions are

1. A comprehensive discussion of research campaigns utilising RF-channel based features for the detection of location or activities (section 2.3.2)

2. A concise investigation on possible features for RF-based activity recognition (section 2.3.4)

3. A case study on activity classification of a single individual from RF-channel based features for

   3a) an active DFAR system utilising 900 MHz software defined radio nodes (section 2.3.5), and

   3b) a passive DFAR system utilising ambient FM radio signals at 82.5 MHz (section 2.3.5) considering in both cases

4. the classification accuracy with respect to activity and location.

The majority of the features we consider are amplitude-based. Since with the Received Signal Strength Indicator (RSSI), a related value is commonly provided by contemporary transceiver hardware, the features utilised in this study can be implemented similarly for most current mobile devices.

Our discussion is structured as follows. In section 2.3.2 we review the related work on activity and location recognition with a particular focus on radio frequency based or related environmental features. Section 2.3.3 then discusses use-cases and application

Table 2.5: Classification-parameters for radio-based context recognition systems (1536–1233 ©2013 IEEE)

Parameter	Values
Sensor type	Device-bound; Device-free
Sensing modality	Passive; Active
Setup	Ad-hoc; Non-ad-hoc (requires training)

scenarios for RF-based activity recognition. The features utilised in our case-studies are introduced, analysed and discussed in section 2.3.4. Based on some of these features, we report from the experiments in section 2.3.5. In particular, we demonstrate the detection of five activities with active and passive DFAR systems. We can also show that a localisation of these activities is feasible simultaneously from the same set of features. Section 2.3.6 summarises the results and closes our discussion.

2.3.2 Related work

Activity recognition comprises the challenge to recognise human activities from the input of sensor data. A broad range of sensors can be applied for this task. Traditionally, accelerometer devices have evolved as the standard equipment for activity recognition both for their high diffusion and convincing recognition rates [83, 84]. General research challenges for activity recognition regard the accurate classification of noisy data captured under real world conditions [85] or the automation of recognition systems [86]. Another problem that is addressed in depth only recently is the creation of classification systems that scale to a large user base. With increasing penetration of sensor enriched environments and devices, the diversity in user population poses new challenges to activity recognition. Abdullah et al. for instance address this challenge by maintaining several groups of similar users during training to identify inter-user differences without the need for individual classifiers [87].

Even more fundamental and aligned to this scaling problem is the required cost for accurately equipping subjects, training them to the system, equipping the environment or the users and most importantly, having them to actually wear the sensing hardware. The classification accuracy is highly dependent on the accurate sensor location. The integration of sensors in clothing as well as the recent remarkable progress in the robustness to rotation or displacement have improved this situation greatly [88]. However, a subject is still required to cooperate and at least wear the sensors [89]. This requirement can not be assured generally in real-world applications. In particular, even devices as private as mobile phones, which are frequently assumed to be constantly in the same context as its owner [90, 91, 92], can not serve as a sensor platform suitable to accurately capture the context of an individual. Dey et al. investigated in [93] that users have their mobile phone within arms reach only 54% of the time. This confirms a similar investigation of Patel et al. in

2006 [94] which reported a share of 58% for the same measure.

These general challenges of activity recognition can be overcome be using an environmental sensing modality. Naturally, vision-based approaches, such as video [95] and recently also the Kinect and wii concepts have been employed by scientists to classify gestures and activities [96]. However, the burden of installation and cost make such approaches hard to deploy at scale [89]. Recently, researchers therefore explore alternative sensing modalities that are pre-installed and readily available in environments and therefore minimise installation cost.

Patel et. al. coined the term infrastructure-mediated sensing and demonstrated in 2007 that alterations in resistance and inductive electrical load in a residential power supply system due to human interaction can be automatically identified [97]. They leveraged transients generated by mechanically switched motor loads to detect and classify such human interaction from electrical events. In a related work from 2010, Gupta et al. analysed electromagnetic interference (EMI) from switch mode power supplies [98]. In [99] they showed that it is even possible to detect simple gestures near compact fluorescent light by analysing the EMI-structures, effectively turning common light bulbs in a house into sensors. Environmental sensing with atypical sensing devices is also considered by Campbell et al. and Thomaz et al. who present an activity detection method utilising residential water pipes [100, 101]. In [102], Cohn et al. form a residual Power-line system into a large distributed antenna to sense low power signals from parasitic or distant devices. These approaches all require explicit interaction between a cooperating individual and a specific sensing entity. These approaches are bound to specific environments or installations and typically are only feasible indoors. An infrastructure mediated sensing medium with greater range is the RF-channel. Signal strength, amplitude fluctuation or noise level provide information that can be utilised to classify environmental situations.

Several authors considered the localisation of individuals based on measurements from the RF-sensor. Results are typically achieved by analysing the RF-signal amplitude, namely the RSSI of a received signal. Classical approaches are device-bound and utilise the RF-sensor for location estimation of an active entity equipped with a RF transceiver. In these approaches, the impact of multi-path fading and shadowing on the transmission channel and therefore the strength of an RF signal is exploited. These approaches were driven by the attempt to provide capabilities of indoor localisation. The first promising work was the RADAR system presented by Bahl et al. [103]. The authors took advantage of existing communications infrastructure, WiFi access points, and employed RSSI fingerprints to identify locations off-line. With location, this approach was then applied also with GSM networks [104, 105], FM radio signals [106, 107] and domestic powerline [94, 108]. Recently, automations have been proposed for such fingerprinting approaches [109, 110].

While these systems rely on a two staged approach in which first a map of fingerprints is created off-line, recent work achieves on-line real-time localisation of entities equipped with a wireless transceiver based on WiFi or FM radio [111, 112, 113]. These studies were initiated in 2006 by Woyach et al. who detail various environmental changes and their effect on a transmit signal [76]. The authors utilise MICAz nodes to show that motion detection based on RSSI measurements can be more accurate than accelerometer

data when changes are below the sensitivity of the accelerometer. They experimentally employ several indoor-settings in which a receive node analyses a signal obtained from a transmitter. In this study they focused on an increased fluctuation in the RSSI signal level. Additionally, the authors showed that velocity of an entity can be estimated by analysing the RSSI pattern of continuously transmitted packets of a moving node. This work was advanced by Muthukrishnan et al. who study in 2007 the feasibility of motion sensing in a WiFi network [77]. They analyse fluctuation in the 1 byte 802.11 RSSI indicator to sense whether a device is moving. The authors consider only the two cases of motion and no motion and achieve a classification accuracy of up to 0.94. A more fine grained distinction was made by Anderson et al. and Sohn et al. based on fluctuations in GSM signal strength [114, 115]. The authors of [114] implement a neural network to detect the travel mode of a mobile phone. They monitor the signal strength fluctuation from cells in the active set to distinguish between walking, driving and stationary with an accuracy between 0.8 and 0.9.

Sohn et. al describe a system that extracts seven features from GSM signal strength measurements to distinguish six velocity levels with an accuracy of 0.85 [115]. The features mainly build on distinct measures of variation in signal strength and the frequency of cell-tower changes in the active set.

While all previously mentioned results considered special installations of the wireless transmitters, Sen et al presented a system that allows the localisation of a wireless device with an accuracy of about 1 meter from WiFi physical layer information even when the receiver is carried by a person that might induce additional noise to the captured features [116].

Summarising, these studies are examples of device-bound and active velocity and location estimation approaches since they require that the located entity is equipped with a RF transceiver.

Recently, some authors also consider RF-sensing to detect the presence or location of passive entities. Since these systems require at least one active transmitter, they can be classified as active, device-free systems.

Youssef defines this approach as Device-Free Localisation (DFL) in [78] to localise or track a person using RF-Signals while the entity monitored is not required to carry an active transmitter or receiver. They localised individuals by exchanging packets between 802.11b nodes in corners of a room and analysed the moving average and its variance of the RSSI [78]. Classification accuracy reached up to 1.0 for some configurations. Additionally, they presented a fingerprint-based localisation system with an accuracy of 0.9. Later, they improved their approach using less nodes [79]. A passive radio map was constructed offline before a Bayesian-based inference algorithm estimated the most probable location. These experiments have been conducted under Line-of-Sight (LoS) conditions. Also, Wilson and Patwari showed in conformance with the findings of Kosba et al. [117] that the variance of the RSSI can be used as an indicator of motion of non-actively transmitting individuals regardless of the average path loss that occurs due to dense walls and stationary objects [118]. The area in which environmental changes impact signal characteristics was then considered by Zhang et al. They used 870 MHz nodes arranged in a grid to show that for each link an

elliptical area of about 0.5 to 1 meters diameter exists for which RSSI fluctuation caused by an object traversing this area exceeds measurements in a static environment [119]. They identified a valid region for detecting the impact (i.e. the RSSI fluctuations exceeding the measured threshold in a static environment) for transceiver distances from 2 m to 5 m for the considered 870 MHz frequency range [120]. By dividing a room into hexagonal cell-clusters with measurements following a TDMA scheduling, an object position could be derived with an accuracy of around 1 meter. This accuracy was further improved by Wilson and Patwari in 2011 [118]. They utilised a dense node array to locate individuals within a room with an average error of about 0.5 meters. This was possible by instrumenting a tomographic image over the 2-way RSSI fluctuations of nodes [121]. All these studies consider a single experimental setting.

In a related work, Lee et al. sense the presence of an individual in five distinct environments [122]. They showed that the RSSI peak is concentrated in a restricted frequency band in a vacant environment while it is spread and reduced in intensity in the presence of an individual. In 2011, Kosba et al. presented a new system for the detection of human movement in a monitored area [123]. Using anomaly detection methods they achieved 6% miss detection and a 9% false alarm rate when utilising the mean and standard deviation of the RSSI in two environments. They further implemented techniques to counteract effects of dispersion. This was accomplished by continuously adding newly measured data which did not trigger the detection. The previous results all considered the localisation of a single individual.

The simultaneous localisation of multiple individuals at the same time was first mentioned and studied by Patwari and Wilson in [124]. The authors derive a statistical model to approximate the position of a person based on RSSI variance which can be extended to multiple persons. This aspect together with the previously untackled problem that environmental changes over time might necessitate frequent calibration of the location system was approached by Zhang and others in [125]. The authors isolate the LoS path by extracting phase information from the differences in the RSS on various frequency spectrums at distributed nodes. Their experimental system is with this approach able to simultaneously and continuously localise up to 5 persons in a changing environment with an accuracy of 1 meter.

We summarise that most work conducted in the area of RF-based classification with passive participants is related to the localisation of individuals. The feasibility of this approach was verified in various environmental settings and at various frequencies. The features utilised are mostly the RSSI, its moving average, mean or RSSI fingerprint. Also, 2-way RSSI variance was employed. With these features a localisation accuracy of about 0.5 meters was possible or the simultaneous localisation of up to 5 persons in a changing environment with an accuracy of 1 meter.

While the localisation of individuals based on features from the radio channel can therefore generally be considered as solved, recently, some authors considered active DFAR approaches to also detect activities.

Patwari et al. monitor breathing based on RSS analysis [126]. The monitored area was surrounded by twenty 2.4 GHz nodes and the two-way RSSI was measured. Using a

maximum likelihood estimator they approximated the breathing rate within 0.1 to 0.4 beats accuracy.

Recently, we also conducted preliminary studies regarding the use of features from a RF-transceiver to classify static environmental changes such as opened or closed doors, presence, location and count of persons with an accuracy of 0.6 to 0.7 [70, 71, 75, 127]. We utilised USRP Software defined radio devices (SDR)[7] from which one constantly transmits a signal that is read and analysed by other nodes. Devices were equipped with 900 MHz transceiver boards. With the software radios a higher sampling frequency than in previous studies is possible and we can also sample the actual channel instead of only tracking the RSSI. In these studies we concentrated on features related to the signal amplitude and derivation of the instantaneous amplitude from its mean. Furthermore, we conducted preliminary studies on passive device free situation awareness by utilising ambient signals from a FM radio station not under the control of the recognition system. In these studies, static environmental changes such as opened doors have been detected with an accuracy of about 0.9 [81] and a first study on suitable features to detect human activities could achieve an accuracy of about 0.8 with a two stage recognition approach [82].

DFAR is still a mostly unexplored research field. Open research questions regard the optimum frequencies and the impact of the frequency on the classification accuracy, the optimum sampling rate of the signal, the detection range and the impact of this distance on the classification accuracy as well as the minimum Signal-to-Noise Ratio (SNR). Furthermore, a set of activities that can be recognised by RF-based classification is yet to be identified as well as a suitable design of the detection system. In particular, the impact of the count and height of transmitting and receiving nodes has not yet been considered comprehensively as well as even the actual necessity of a transmit node as part of the recognition system since potentially the system might utilise ambient radio. Also, it is not clear whether and how activities of multiple persons can be identified simultaneously and if features exist that enable ad-hoc DFAR systems. A more detailed discussion of most of these aspects is given by Scholz et al. in [80].

In the present study, we identify and evaluate features for the classification of activities from RF-signals in two frequency bands (900 MHz and 82.5 MHz) with systems utilising ambient radio as well as a system-generated signal. Four activities, two dynamic and two static, together with the empty environment are considered.

2.3.3 Application scenarios for DFAR

We believe that DFAR research can provide a foundation for the realisation of an IoT and for Ubicomp in general. The RF-Sensor has a high penetration in common equipment and will be available in virtually all IoT devices. To reduce cost and complexity, hardware designers and application developers might then rather investigate and utilise the common RF-transceiver to sense environmental stimuli than integrating additional sensing hardware. Currently, the information provided by the RF-channel is, although available

---

[7]http://www.ettus.com

virtually for free, mostly disregarded and discarded unused.

Apart from modulated data, the signal strength, amplitude fluctuation or noise level provide additional information about environmental situations. In the following sections we exemplify two applications for DFAR in emergency situations and elderly monitoring.

Monitoring in disaster stricken areas

Despite tremendous efforts, careful preparation and training for a 'worst case', increased security precautions and costly installations of early warning systems, disaster situations either caused by nature or human intervention frequently strike also highly developed countries. Recent cautionary tales are the flooding in Thailand or also the Tohoku earthquake near Sendai, Japan that let to a devastating tsunami and was the cause of the atomic crisis around the Fukushima-Daichi power plant.

In the time since this event, research efforts have been taken in the search of systems that can assist auxiliary forces in areas where most of the infrastructure is destroyed. One important and urgent issue in such situations is the search for survivors and injured persons that might reside, for instance, in partly destroyed buildings [128, 129].

When the existing infrastructure is destroyed, RF-sensing might provide a cheap and wide-ranging alternative to assist rescue forces. With a single RF-transmitter such as an RF-radio tower or a base station, a large area can not only be supplied with voice and data communication but the fluctuation in RF-channel characteristics might be employed to detect individuals and identify their status from activities such as lying, crawling, standing or walking. Auxiliary forces might bring out a network of RF-transceiver devices in order to monitor an area via RF-channel fluctuation as part of their professional routine while at the same time establishing communication means via this RF-transceiver infrastructure [130]. The range, optimum installation height and features for ad-hoc operation are still open research questions for DFAR but the results presented in this work show that assistance in such scenarios can be provided by RF-sensing (although due to the lack of prior training the set of activities recognised might be reduced, for instance, to 'some movement' and 'some static alteration'). These additional sensing capabilities come virtually for free on top of the installation of wireless communication.

Supporting well-being in domestic areas

Most accidents happen at home. The primary reason for these accidents are falls which make up about 40% of the total number of accidents [131]. Most of these accidents leave the affected person in an unusual posture such as lying at an unusual location. While the automatic detection of fall and fall prevention has gained large interest in the research community and various approaches have been proposed, these alarm system either need body-attached sensors, require the installation of a complex infrastructure or have strong privacy related implications as, for instance, video based systems [132, 133, 134]. By utilising the RF-sensor for this kind of detection we would reduce privacy issues, avoid the need of having to carry sensors and ideally reduce installation requirements to a minimum.

The sensor could further become a crucial component of (Health) Smart Home systems [135, 136] relieving users from the necessity to wear a device. In fact, for Smart Home systems, the sensor needs to provide a rough localisation capability as well as the recognition of at least a basic set of activities of daily living. Among such activities are walking, standing and sleeping [137].

Considering the demographic change in developing and developed countries, the application of the RF-sensor for alarm systems or Smart Homes could further play an important role towards the extension of self-sustained living of the elderly. The present study illustrates the potential of the ubiquitously available RF-sensor for the detection of relevant activities in Smart Home environments.

2.3.4 Features for DFAR

In the following we discuss the RF-based features we considered and their achieved classification accuracy. We identify a set of three most relevant features for active and passive DFAR systems.

For our active DFAR system, we deploy a USRP SDR transmit node constantly broadcasting a signal $m(t)$ at a frequency of $f_c = 900$ MHz. In the passive DFAR system a FM radio signal $m(t)$ from a local radio station at $f_c = 82.5$ MHz is utilised. In both cases, the received signal

$$\zeta_{\text{rec}}(t) = \Re\left(m(t)e^{j2\pi f_c t}\text{RSS}e^{j(\psi+\phi)}\right) \tag{2.47}$$

is read by one USRP SDR node and is analysed for signal distortion and its fluctuation due to channel characteristics. In equation (2.47) the RSS denotes the Received Signal Strength. The value ϕ accounts for the phase offset in the received signal due to the signal propagation time. This continuous received signal is sampled from the USRP devices $64 \cdot 10^9$ times per second at distinct time intervals $t = 1, 2, \ldots$ in a resolution of 12 bits.

We considered the following features for activity classification. For all features we employed a window $\mathcal{W}$ of $|\mathcal{W}|$ samples to calculate their value. The blocking or damping of signal components by subjects or other entities impacts the amplitude of the received signal. A feature to measure this property is the maximum peak of the signal amplitude. We calculate it by the difference between the maximum and minimum amplitude within one sample window

$$\mathcal{P}_{\text{eak}} = \max_{t \in \mathcal{W}}\left(\zeta_{\text{rec}}(t)\right) - \min_{t \in \mathcal{W}}\left(\zeta_{\text{rec}}(t)\right) \tag{2.48}$$

We utilise the Mean amplitude μ of the received signal frequently as a reference value to compare the current amplitude of a signal $\zeta_{\text{rec}}(t)$ to the average amplitude in a training situation:

$$\mu = \frac{\sum_{t=1}^{|\mathcal{W}|} \zeta_{\text{rec}}(t)}{|\mathcal{W}|} \tag{2.49}$$

The Root of the Mean Square (RMS) deviation of the signal amplitude $|\zeta_{\text{rec}}(t)|$ to the

mean μ is also utilised. With lower RMS we expect fewer alterations in an environment.

$$\text{RMS} = \sqrt{\frac{\sum_{t=1}^{|\mathcal{W}|} \left(\zeta_{\text{rec}}(t) - \mu\right)^2}{|\mathcal{W}|}} \tag{2.50}$$

Furthermore, we investigate the second and third central moment that express the shape of a cloud of measured points. The second central moment describes the variance σ^2 of a set of points. It can be used to measure how far a set of points deviates from its mean.

$$\sigma^2 = \frac{\sum_{t=1}^{|\mathcal{W}|} \left(\zeta_{\text{rec}}(t) - \mu\right)^2}{|\mathcal{W}|} \tag{2.51}$$

Additionally, we consider the third central moment.

$$\gamma = \text{E}[\zeta_{\text{rec}}(t) - \mu)^3] \tag{2.52}$$

In equation (2.52), $\text{E}[x]$ defines the expectation of a value x.

All above features are taken from the time domain of the received signal. In the frequency domain, we consider the DC component a_0, the spectral energy $\mathcal{E}$ and the entropy H of the signal.

The feature a_0 represents the average of all samples a Fast Fourier Transform (FFT) was applied to. It describes the vertical offset of an observed signal.

We calculate its ith frequency component as

$$\text{FFT}(i) = \sum_{t=1}^{|\mathcal{W}|} \zeta_{\text{rec}}(t) e^{-j\frac{2\pi}{N}it}. \tag{2.53}$$

In equation (2.53) we choose the window size $|\mathcal{W}|$ as the quantity of the samples in the FFT.

The DC component is defined by the first Fourier coefficient $\text{FFT}(i)$ and is separately calculated as

$$a_0 = \frac{\int_{-\frac{|\mathcal{W}|}{2}}^{\frac{|\mathcal{W}|}{2}} \left(\zeta_{\text{rec}}(t)\right) dt}{|\mathcal{W}|} \tag{2.54}$$

The signal energy $\mathcal{E}$ can be computed as the squared sum of its probability density of spectrum in each frame. The probability of each spectral $\text{FFT}(i)$ band is

$$\text{P}(i) = \frac{\text{FFT}(i)^2}{\sum_{j=1}^{|\mathcal{W}|/2} \text{FFT}(j)^2}. \tag{2.55}$$

Consequently, we calculate the spectral energy as

$$\mathcal{E} = \sum_{i=1}^{|\mathcal{W}|/2} \text{P}(i)^2. \tag{2.56}$$

Table 2.6: Best feature combinations for the passive DFAR system (1536–1233 ©2013 IEEE)

acc	$\mathcal{P}_{\text{eak}}$	μ	a_0	$\mathcal{E}$	H	σ^2	γ	RMS
.866	x		x			x		
.863	x					x	x	
.861	x		x				x	
.861	x		x					
.861	x						x	

(a) Distinction between dynamic and static activities

acc	$\mathcal{P}_{\text{eak}}$	μ	a_0	$\mathcal{E}$	H	σ^2	γ	RMS
.902	x		x					
.898	x	x	x					
.898	x		x				x	
.896	x	x						x
.894	x	x					x	

(b) Distinction between standing, lying and empty

acc	$\mathcal{P}_{\text{eak}}$	μ	a_0	$\mathcal{E}$	H	σ^2	γ	RMS
.817			x	x		x		
.706		x	x					x
.701	x		x		x			
.701	x			x	x			
.701	x					x	x	

(c) Distinction between walking and crawling

acc	$\mathcal{P}_{\text{eak}}$	μ	a_0	$\mathcal{E}$	H	σ^2	γ	RMS
.694	x					x	x	
.686	x		x				x	
.683	x		x					
.679	x	x	x					
.679	x		x					x

(d) Distinction between all five activities: standing, walking, crawling, lying and empty

We compute the entropy of a set of points as

$$H = \sum_{i=1}^{|\mathcal{W}|/2} P(i) \cdot \ln\left(P(i)\right). \tag{2.57}$$

For all possible combinations of up to three of these features we exploited their classification accuracy of the five activities considered in section 2.3.5 and in section 2.3.5. Table 2.6 details the accuracy for the best five feature combinations[8] of the passive DFAR system with $|\mathcal{W}| = 32$.

The table distinguishes between one-stage and two-stage classification. For the one-stage classification, all five activities are distinguished in one single classification step. For two-stage classification, first the classifier distinguishes between dynamic (walking or crawling) and static (lying, standing or empty) activities. Then, the final classification is done in one of these classes. We observe that in particular $\mathcal{P}_{\text{eak}}$ and a_0 are well suited to achieve a high classification accuracy.

A high $\mathcal{P}_{\text{eak}}$ value indicates a dynamic activity. The feature is therefore well suited to distinguish between dynamic and static activities. Consequently, for the distinction between the two dynamic activities in table 2.7c, this feature is less prominent. The DC-component a_0 mostly represents the vertical offset of the signal. In can therefore serve as an indicator to distinguish whether a person is standing or walking, lying or standing or whether the room is empty.

For the active DFAR system, the most significant features are the variance σ^2, the third central moment γ when applied twice and the minimum over a window of maximum values.

---

[8]The complete table with all results is available at http://klab.nii.ac.jp/˜sigg/TMC-2012-01-0047_PassiveDFARAcc.pdf

We achieved good results for a window size of $|\mathcal{W}| = 20$ which translates to $|\mathcal{W}| = 400$ for features applied on preprocessed data. By the adding further combinations of features, the overall classification accuracy can be further improved slightly.

Generally, for the activities considered, the dynamic activities have a greater number of significant features as they also have characteristic alterations over time. Static activities are therefore in principle harder to distinguish from each other and it will likely not be possible to re-use a trained classifier for static activities without re-training in another scenario.

2.3.5 RF-based DFAR

In order to explore the activity recognition capabilities and limits of the RF-sensor, we conducted case studies for active and passive DFAR implementations. In particular, three subjects have conducted the four activities lying, standing, crawling and walking in a corridor of our institute. Additionally, the empty corridor was considered as a baseline activity. All experiments have been conducted in after-hours to ensure a controlled environment in which all important external parameters are kept stable. In particular, no additional subjects have been present in the corridor or in adjacent rooms that could have interfered with the experimental conditions. Figure 2.22 depicts the setting employed for the case study.

The experimental space was divided into five areas with respect to their distance to the receiver. For active and passive systems, the receiver was placed at the same location in the center of the detection area. For the active DFAR system, the transmitter was positioned in two meters distance from the receiver.

The activities were conducted at the five locations which are labelled A, B, C, D, and E. Locations A and E are in a distance of 2.20 meters from the receiver, locations B and D are separated by 1.35 meters and location C is 0.5 meters apart. All locations are arranged in a circle around the receiver in their center.

Each of the three subjects repeated all activities at every location for about 60 seconds. We took arbitrary patterns from these sample sequences for classification.

For the active DFAR system the transmitter constantly modulated a signal to a 900 MHz carrier which was then sampled at the receiver at 70 Hz. USRP 1 devices[9] were utilised as transmitter and receiver with RFX900 daughterboards[10] and VERT900 Antennas[11] with 3dB antenna gain.

The receiver of the passive DFAR system sampled a signal from an ambient FM radio station at 82.5 MHz with a sample rate of 255 kHz. We employed a USRP N200[12] device with a WBX daughterboard[13] together with a VERT900 Antenna[14] with 3dB antenna

---

[9]https://www.ettus.com/product/details/USRP-PKG

[10]https://www.ettus.com/product/details/RFX900

[11]https://www.ettus.com/product/details/VERT900

[12]https://www.ettus.com/product/details/UN200-KIT

[13]https://www.ettus.com/product/details/WBX
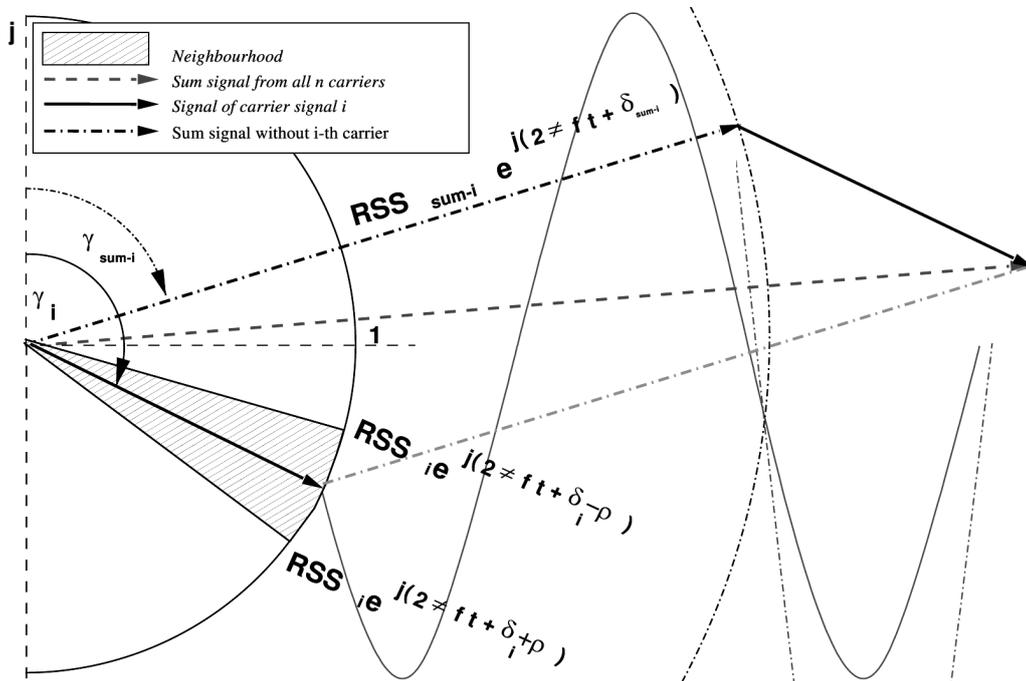
[14]https://www.ettus.com/product/details/VERT900

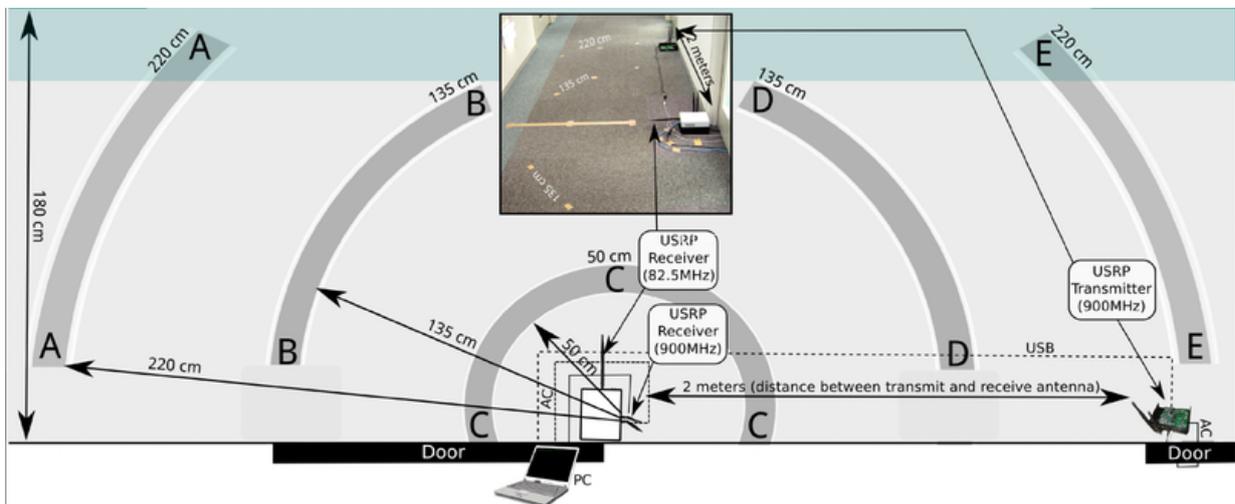Figure 2.21: Illustration of the neighbourhood of the local random search approach

Figure 2.22: Schematic illustration of the corridor in which the case-study was performed. Locations at which activities were conducted are marked (A,B,C,D,E). Both receive nodes are located in the center of the recognition area on top of each other. (1536–1233 ©2013 IEEE)
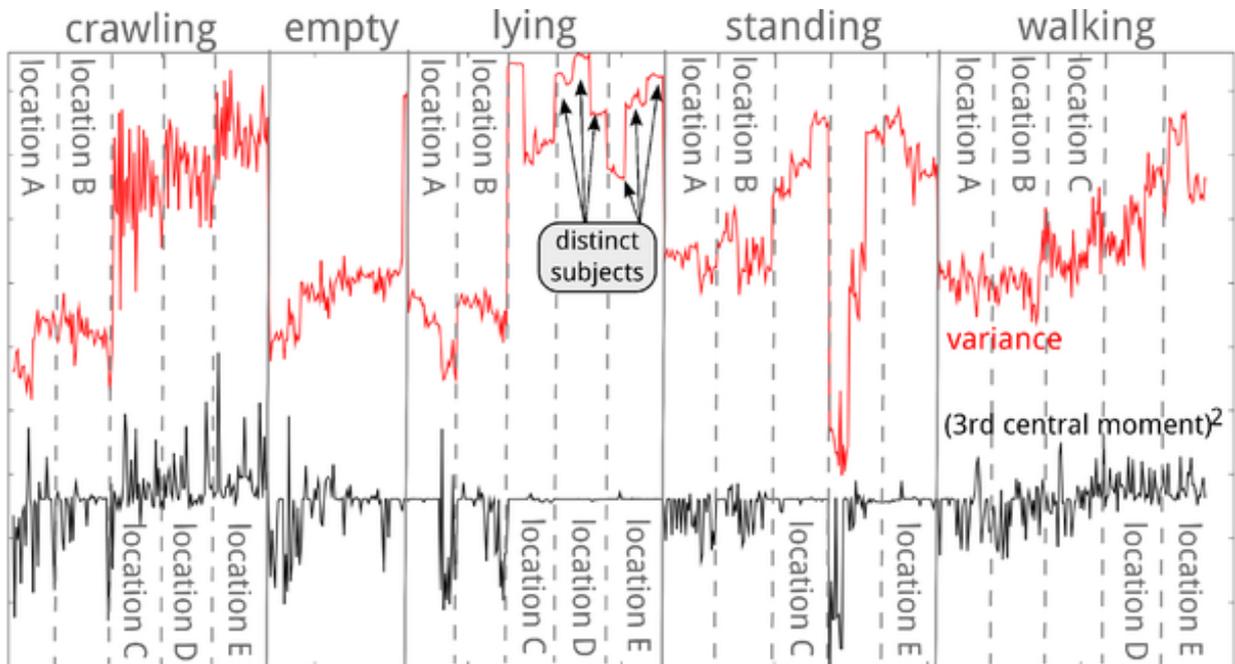
Figure 2.23: Exemplary feature samples (variance and twice applied 3rd central moment; over 400 samples each) from all activities, locations and subjects for active DFAR (1536–1233 ©2013 IEEE)

gain [82].

Active device-free activity recognition

For the detection of the described activities with our active DFAR system we utilise a one-stage classification approach. In particular we use as features the mean μ, the variance σ^2, the third central moment γ, the RMS, the count of amplitude peaks within 90% of the maximum, the distance of zero crossings, the Energy $\mathcal{E}$ and the entropy H, over a window of 400 samples.

For classification we utilise a k-nearest neighbour (k-NN) classifier with $k = 10$ and a decision tree (DT). Figure 2.23 depicts values for the variance σ^2 and the third central moment γ applied twice for part of the sample data. Distinct activities are clearly distinguishable in this plot already.

From this data we observe that activities conducted at locations A and B are seemingly harder to distinguish from the empty case. The reason is that activities at these locations are conducted relative to the transmitter behind the receiver and therefore have less impact on the received signals.

Classification results after 10-fold cross validation are depicted in table 2.7. Table fields with very low values (i.e. 0.0) are left blank. The table depicts the classification accuracy when classifiers for the five activities empty, walking, standing, lying, crawling have been trained on features obtained for all five locations and subjects. Due to the challenging

Table 2.7: Classification of activities conducted by three subjects at Locations A to E by a k-nearest neighbour and a decision tree classifier in an active DFAR system

	Classification				
	cr	em	ly	st	wa
crawling	**.713**	.024	.06		.204
empty	.022	**.593**	.187	.121	.077
lying	.042	.048	**.743**	.144	.024
standing	.011	.067	.078	**.777**	.067
walking	.166	.029	.034	.051	**.72**

(a) Confusion matrix for the k-NN classifier over samples from all locations and subjects

	Classification				
	cr	em	ly	st	wa
crawling	**.659**	.006	.054	.024	.257
empty	.055	**.582**	.154	.121	.088
lying	.054	.042	**.784**	.102	.018
standing	.022	.056	.095	**.771**	.056
walking	.189	.023	.011	.057	**.72**

(b) Confusion matrix for the classification tree classifier over samples from all locations and subjects

Table 2.8: Accuracy, Information score and Brier score for the classification algorithms

	Accuracy	Information score	Brier score
Classification tree	0.716	1.529	0.567
k-NN classifier	0.722	1.518	0.4

feature value fluctuations for locations A and B we have not been able to achieve a higher accuracy in this case. In particular, we notice that the distinction of the empty class is hard for the classifiers since other activities conducted at locations A and B have a similar feature value footprint. The overall classification accuracies are 0.714 and 0.722 for the classification tree and the k-NN classifier as depicted in table 2.8. The table also shows the Brier score and the Information score as defined by Kononenko and Bratko [138]. These basic accuracies can be improved when classifiers are trained at specific locations and when the classification of activities is segmented for distinct locations as derived in the next sections.

Spatial impact on accuracy   To improve accuracy we spatially restricted the classification area. In particular, we utilised feature values only from activities conducted at one distinct location (A,B,C,D or E). Table 2.9 shows the classification results for location C. The classification accuracy is increased in this case compared to the previous general setting. This is also due to the subjects conducting activities in only about 50 cm distance from the receive antenna. The impact on the signal is therefore significant.

With increasing distance to the receiver, the classification accuracy slowly deteriorates as visible in table 2.10. The table depicts the classification accuracy of the k-NN classifier. Classification accuracies for the decision tree are comparable as shown in table 2.11. We observe that indeed locations E, D and C achieve best classification results. The impact of reflected signals from an action conducted behind the receiver quickly diminishes, so that

Table 2.9: Classification of activities conducted by three subjects at Location C by a k-nearest neighbour and a decision tree classifier in an active DFAR system (1536–1233 ©2013 IEEE)

	Classification				
	cr	em	ly	st	wa
crawling	**.848**			.03	.121
empty		**.978**			.022
lying			**.839**	.161	
standing			.108	**.892**	
walking	.143				**.857**

(a) Confusion matrix for the k-NN classifier over samples from all subjects at location C

	Classification				
	cr	em	ly	st	wa
crawling	**.727**	.03	.03	.03	.182
empty		**.978**	.022		
lying			.032	**.677**	.29
standing	.054	.027	.135	**.784**	
walking	.2				**.8**

(b) Confusion matrix for the classification tree classifier over samples from all subjects at location C

Table 2.10: Classification of activities conducted by three subjects at Locations A, B, D or E by a k-nearest neighbour classifier in an active DFAR system (1536–1233 ©2013 IEEE)

	Classification				
	cr	em	ly	st	wa
crawling	**.556**	.167	.194		.083
empty	.044	**.769**	.055	.066	.066
lying	.125	.25	**.625**		
standing		.286		**.686**	.029
walking	.03	.394			**.576**

(a) Confusion matrix for the k-NN classifier over samples from all subjects at location A

	Classification				
	cr	em	ly	st	wa
crawling	**.8**	.133			.067
empty		**.78**	.132	.077	.011
lying		.242	**.727**		.03
standing		.278		**.722**	0
walking	.03	.242	.091	.061	**.576**

(b) Confusion matrix for the k-NN classifier over samples from all subjects location B

	Classification				
	cr	em	ly	st	wa
crawling	**.765**			.029	.206
empty		**.967**	.033		
lying			**.971**	.029	
standing		.194	.056	**.722**	.028
walking	.111				**.889**

(c) Confusion matrix for the k-NN classifier over samples from all subjects at location D

	Classification				
	cr	em	ly	st	wa
crawling	**.676**		.029		.294
empty		**.967**	.033		
lying			**.865**	.135	
standing	.029		.114	**.829**	.029
walking	.263				**.737**

(d) Confusion matrix for the k-NN classifier over samples from all subjects location E

Table 2.11: Accuracy, Information score and Brier score for the classification algorithms in an active DFAR system (1536–1233 ©2013 IEEE)

	Accuracy	Information score	Brier score
Location A			
Classification tree	0.674	1.284	0.652
k-NN classifier	0.674	1.309	0.449
Location B			
Classification tree	0.825	1.696	0.35
k-NN classifier	0.735	1.385	0.383
Location C			
Classification tree	0.841	1.702	0.317
k-NN classifier	0.907	1.892	0.129
Location D			
Classification tree	0.832	1.709	0.337
k-NN classifier	0.887	1.846	0.168
Location E			
Classification tree	0.779	1.549	0.442
k-NN classifier	0.851	1.767	0.214

the classification accuracy of activities conducted at these locations (A and B) quickly worsens with distance.

Localising an action In the previous case, the classifiers were trained for actions of a specific location without considering actions taking place at other locations. We now train the classifiers on all five activities at all five locations, respectively. The action 'empty' is identical regardless of the location. Overall, we then distinguish between 21 classes. Table 2.12 depicts our results.

We observe that the right action is classified for the right location most often. Moreover, we see a locality in the classifications. Misclassifications are seldom in different activities but most often regarding the correct activity in a neighbouring location. With increasing distance to the place where the action was trained, the misclassification error increases. The distance between locations was 85 cm. We therefore conclude that a localisation of activities is possible alongside classification with an error of less than 1 meter. We further observe that the static activities standing and lying as well as the dynamic activities walking and crawling are harder to distinguish for the classifier since their features are not so well separated.

Summary on active DFAR studies All five activities are classified with varying accuracy depending on the setting considered. Higher accuracy can be achieved when the activities are conducted near the receiver node or between the transmitter and receiver. With increasing distance to the receiver the classification accuracy deteriorates. When the activities are trained at various locations, a localisation of the classified activity within less than 1 meter radius is possible.

Table 2.12: Accuracy when training is accomplished including activities at all locations in an active DFAR system (1536–1233)

	crawling					empty	lying					standing					walking				
	A	B	C	D	E		A	B	C	D	E	A	B	C	D	E	A	B	C	D	E
crawling at A	**.528**	.028				.111	.167	.028													
crawling at B	.033	**.867**				.067															
crawling at C	.03		**.455**	.152	.091				.029										.03	.03	
crawling at D			.176	**.176**	.147															.324	.182
crawling at E			.088	.176	**.529**																.206
empty	.022					**.681**	.011	.088		.011	.022	.044	.077								
lying at A	.094					.188	**.594**	.125													
lying at B	.091					.152	.152	**.545**							.03	.03					
lying at C									**.677**	.059		.097	.065	.129							
lying at D									.054	**.765**	.088										
lying at E										.162	**.595**		.135			.108					
standing at A						.2						**.686**	.057				.057	.033			
standing at B						.194						.306	**.444**				.056	.111	.028		
standing at C									.054	.027	.054			**.676**	.028				.03	.029	
standing at D		.028				.028			.111					.114	**.472**	.111			.028	.118	
standing at E									.057	.086	.114			.114	.057	**.514**					.029
walking at A	.061					.152						.057					**.364**	.242	.061		.061
walking at B		.03				.03						.056					.303	**.364**	.061		
walking at C			.086													.029		.061	**.657**	.139	.026
walking at D			.056	.028		.028														**.75**	.028
walking at E			.158	.184	.158														.114	.028	**.421**

Table 2.13: Classification of activities of all subjects conducted at Location C by a k-NN and a classification tree classifier in a passive DFAR system

	Classification				
	empty	lying	standing	walking	crawling
empty	**.942** (.152)		.058 (.008)		(.001)
lying		**.773** (.268)	.027 (.007)	.093 (.021)	
standing	.093 (.024)	.027 (.005)	**.853** (.236)	.027 (.005)	
walking	.0 (.001)	.026 (.008)	.077 (.014)	**.795** (.261)	.103 (.032)
crawling	.0 (.001)	.121 (.038)	.014 (.004)	.176 (.071)	**.689** (.214)

(a) Confusion matrix for the k-NN classifier

	Classification				
	empty	lying	standing	walking	crawling
empty	**.986** (.008)		.014 (.003)		
lying		**.787** (.252)	.013 (.004)	.133 (.006)	.067
standing	.027 (.006)	.013 (.004)	**.933** (.009)		.027 (.005)
walking		.026 (.005)	.077 (.014)	**.769** (.210)	.128 (.033)
crawling		.108 (.031)	.027 (.006)	.135 (.045)	**.730** (.245)

(b) Confusion matrix for the classification tree classifier

Passive device-free activity recognition

In the previous section we considered an active transmitter as one part of the classification system. The disadvantage in such a system is that in a practical situation, a separate transmitter has to be brought out and positioned in the proximity of the receiver that is constantly transmitting. However, since the freely available frequency spectrum is sparse, we can assume to be exposed to some kind of radio signals continuously. The highest coverage is probably reached by FM radio. We attempt to utilise ambient FM radio signals from a nearby FM radio station in order to detect the five activities described above in the same setting. In our case study, the FM-receiver was placed on top of the 900 MHz USRP receiver to sample ambient signals. Samples have been taken simultaneously to the active DFAR studies described above. We utilise 10 fold cross validation with k-NN and decision tree classifiers. A two-stage classification approach with the feature sets that reached best classification accuracy was shown in table 2.6).

Table 2.13 details the classification accuracy when activities are conducted at location C only and classifiers are trained only on these feature values. The table depicts the median classification accuracy and the variance over 10 separate classifications. For ease of presentation, table entries with very low values (0.0 (0.0)) are left empty.

With increasing distance to the receiver, the classification accuracy deteriorates. Naturally, since we utilise ambient signals, the direction in which activities are moved away from the receiver is of minor importance (cf. table 2.14). These tables show classification results of the k-NN classifier. Classification accuracies of the classification tree have been comparable. However, we observe that especially the detection of static activities, in particular lying and standing suffer from the increased distance to the receiver. We explain

Table 2.14: Classification of activities of all subjects at Locations A,B,D and E by a k-NN classifier (1536–1233 ©2013 IEEE)

	Classification				
	empty	lying	standing	walking	crawling
empty	**.812** (.231)	.145 (.074)		.014 (.006)	.029 (.010)
lying	.155 (.091)	**.239** (.177)	.155 (.066)	.169 (.073)	.282 (.139)
standing		.27 (.098)	**.893** (.158)		.080 (.004)
walking	.054 (.003)	.162 (.055)		**.649** (.294)	.135 (.042)
crawling	.053 (.009)	.027 (.004)	.093 (.007)	.120 (.007)	**.707** (.302)

(a) Confusion matrix for the classification at location A

	Classification				
	empty	lying	standing	walking	crawling
empty	**.783** (.243)	.072 (.009)	.116 (.009)	.0 (.002)	.029 (.006)
lying	.0 (.001)	**.507** (.419)	.440 (.338)	.053 (.006)	
standing		.214 (.110)	**.786** (.208)		
walking	.035 (.005)	.047 (.006)	.012 (.003)	**.659** (.361)	.247 (.157)
crawling	.013 (.003)	.041 (.009)		.230 (.108)	**.716** (.293)

(b) Confusion matrix for the classification at location B

	Classification				
	empty	lying	standing	walking	crawling
empty	**.875** (.186)		.125 (.045)		
lying		**.758** (.308)	.242 (.112)		
standing	.222 (.097)	.25 (.113)	**.500** (.301)		.027 (.012)
walking	.0 (.001)	.120 (.034)	.173 (.044)	**.680** (.143)	.027 (.005)
crawling		.214 (.085)	.071 (.026)	.262 (.063)	**.452** (.296)

(c) Confusion matrix for the classification at location D

	Classification				
	empty	lying	standing	walking	crawling
empty	**.725** (.262)	.087 (.021)	.159 (.034)	.029 (.004)	
lying	.289 (.103)	**.461** (.186)	.145 (.032)	.105 (.028)	.0 (.001)
standing	.130 (.041)	.273 (.125)	**.558** (.371)	.039 (.004)	
walking	.025 (.007)	.120 (.058)	.025 (.004)	**.667** (.286)	.160 (.071)
crawling	.026 (.007)	.013 (.004)		.171 (.068)	**.789** (.228)

(d) Confusion matrix for the classification at location E

Table 2.15: Accuracy for the k-NN classifier when training is accomplished with activities from all subjects conducted at Location C only in a passive DFAR system.

	Classification				
	empty	lying	standing	walking	crawling
empty at location A	**1.0**				
lying at location A	.355	**.156**	.298	.085	.106
standing at location A		.691	**.215**	.054	.04
walking at location A	.068	.102	.245	**.251**	.333
crawling at location A	.094	.168	.309	.342	**.087**
empty at location B	**1.0**				
lying at location B	.207		.427	.327	.04
standing at location B	.446		**.331**	.223	
walking at location B	.053	.059	.112	**.647**	.129
crawling at location B	.034	.047	135	.608	**.176**
empty at location C	**.986**		.014		
lying at location C		**.787**	.013	.133	.067
standing at location C	.027	.013	**.933**		.027
walking at location C		.026	.077	**.769**	.128
crawling at location C		.108	.027	.135	**.730**
empty at location D	**1.0**				
lying at location D	.113	**.035**	.423	.373	.056
standing at location D	.165	.152	**.468**	.171	.044
walking at location D	.04	.228	.201	**.262**	.268
crawling at location D	.054	.118	.344	.269	**.215**
empty at location E	**1.0**				
lying at location E	.408		.355	.224	.013
standing at location E	.442	.013	**.325**	.201	.019
walking at location E	.204	.043	.228	**.42**	.105
crawling at location E	.099	.132	.152	.503	**.113**

this with the missing LoS signal component between transmitter and receiver. Without a dominant signal component which could have high impact on the received signal when, for instance, blocked, all incoming signal components have equal or similar impact on the signal at the receiver.

For the utilisation of an ambient signal, the distance to a receiver is more critical than in the active DFAR case. In particular, short of the empty corridor, all classification accuracies drop significantly when activities are conducted at locations remote to the location at which the classifier was trained. Table 2.15 shows this property for a case in which the classifier is trained from feature values of activities conducted at location C but applied to activities conducted at all five locations.

When training the classifier with feature values from activities conducted at various locations the accuracy decreases (cf. table 2.16).

Furthermore, a localisation of the conducted activities as it was feasible for the active DFAR case is hardly possible with the passive DFAR system. The classifier can at most give a hint on the possible location as it can be observed from table 2.17. The table details

Table 2.16: Classification of activities at all locations by a k-NN and a classification tree classifier trained on activities conducted by all subjects on location C only in a passive DFAR system (1536–1233 ©2013 IEEE)

	Classification				
	em	ly	st	wa	cr
empty	**.949**			.043	.007
lying	.154	**.473**	.241	.065	.067
standing	.192	.156	**.577**	.049	.025
walking	.054	.128	.126	**.476**	.216
crawling	.032	.153	.126	.199	**.490**

(a) Confusion matrix for the k-NN classifier

	Classification				
	em	ly	st	wa	cr
empty	**.978**		.022		
lying		**.500**	.198	.301	.001
standing			**.787**	.212	.001
walking		.001		**.824**	.130
crawling			.051	.433	**.516**

(b) Confusion matrix for the classification tree

the classification accuracy for all 21 classes considering the activities and their respective locations.

Summary on passive DFAR studies   Summarising, we conclude that activity classification is also feasible with a passive DFAR system utilising ambient FM radio signals. We achieved best classification accuracies when the activity was conducted within 0.5 to 1 meters from the receiver. At higher distances, however, the classification accuracy quickly deteriorated and is hardly usable. A passive DFAR system must therefore employ a higher count of receive devices but can omit a dedicated transmitter. In short distance, classification accuracy is comparable to active DFAR systems.

2.3.6 Conclusion

We have proposed a classification scheme for device-free radio-based activity recognition systems. Following this scheme, we considered non-ad-hoc, active and passive, device-free activity recognition systems.

Classification was achieved by k-NN and decision tree classifiers with similar classification accuracy. For one-stage and two-stage active and passive DFAR systems we derived a set of most significant features with respect to their classification accuracy in our case studies. The presented work is the first to detect the considered activities from RF-channel measurements and also the first to do this with active and passive DFAR systems. Despite some recent advances on device-free radio-based localisation systems, this is also the first study to combine an activity recognition and localisation in one classification algorithm on a common set of features. For the activities lying, crawling, standing and walking we were able to localise them within less than 1 meter in USRP-SDR-based case-studies with the active DFAR system.

The results of this study effectively enable the use of arbitrary wireless devices as sensing equipment.

Still, open challenges remain and present future research questions for radio-based activity recognition systems. Among them are the development of algorithms and features

Table 2.17: Passive DFAR classification accuracy of the k-NN classifier and localisation of activities when training is accomplished including activities at all locations (1536–1233 ©2013 IEEE)

	crawling					empty	lying					standing					walking				
	A	B	C	D	E		A	B	C	D	E	A	B	C	D	E	A	B	C	D	E
crawling at A	.134	.027	.027	.067	.013	.014	.027	.04	.027	.027	.02	.06	.047	.013	.067	.04	.08	.034	.06	.081	.094
crawling at B	.007	.304	.041	.014	.155	.014	.027	.007		.02				.027	.007	.007	.007	.135	.101	.054	.074
crawling at C	.02	.048	.476	.02	.014		.007	.027	.109					.007	.007		.041	.061	.068	.082	.014
crawling at D	.086	.022	.043	.097	.054	.011	.022	.075		.065	.011	.108	.022	.022	.043	.043	.032	.065	.075	.065	.043
crawling at E	.007	.106	.013	.046	.285	.02		.007	.04	.106					.007			.172	.079	.02	.093
empty						1.0															
lying at A	.035	.014	.007	.021	.007	.141	.206	.092	.014	.043	.043	.014	.028	.071	.035	.05	.071	.021	.014	.050	.021
lying at B	.08	.013	.02	.047	.007	.027	.1	.147		.033	.087		.073	.047	.06	.133	.007		.02	.027	.073
lying at C	.04		.107		.047		.013		.638			.054		.007	.02		.013	.013	.007	.04	
lying at D	.014	.035	.007	.056	.134	.007	.014	.035		.282	.035	.028	.028	.042	.049	.028	.035	.014	.063	.063	.028
lying at E	.026			.007		.125	.013	.072		.033	.191		.118	.086	.066	.145	.02		.02	.013	.066
standing at A	.06			.067			.013		.054	.013		.544		.007	.121	.013	.04			.067	
standing at B	.036			.014		.115	.014	.101		.036	.144		.223	.029	.072	.173	.007		.007		.029
standing at C	.007	.027	.007	.013		0.18	.08	.053	.033	.027	.093	.007	.033	.293	.067	.007	.04	.007	.007	.007	.013
standing at D	.063	.006		.025	.013	.013	.044	.063	.013	.025	.063	.114	.063	.07	.228	.07	.013		.025	.082	.006
standing at E	.039			.032	.006	.102	.071	.123		.013	.117		.156	.026	.065	.13	.013	.013	.013	.026	.052
walking at A	.061	.007	.048	.027	.007	.021	.048	.007	.027	.034	.034	.034	.014	.054	.02	.02	.395	.027	.007	.088	.02
walking at B	.006	.153	.071	.018	.147	.006	.029	.012	.012	.029				.012		.012	.024	.224	.147	.024	.076
walking at C	.032	.097	.045	.039	.084		.013	.013	.013	.045	.013			.006	.006	.013	.006	.155	.258	.058	.103
walking at D	.06	.04	.067	.04	.027		.074	.04	.04	.054	.007	.087		.027	.067	.02	.067	.020	.054	.154	.054
walking at E	.093	.074	.012	.024	.148	.037	.006	.056		.031	.062		.031	.024	.006	.043	.019	.056	.08	.062	.136

Classification

which reduce the amount of training effort or the amount of additional required knowledge in order to use the RF-sensor in a different setting. We further need to investigate the required coverage, height and relative location of the sensor in order to deduce how the number of sensor entities affect the resolution of the system. Other questions include the activity detection of multiple persons or the inclusion of mobile nodes within a DFAR system.

Nevertheless, with the presented investigation results it could be shown that the RF-sensor can support applications such as monitoring of emergency situations or the creation of Smart Home systems. In both applications the sensor could not only provide a classification accuracy comparable to the currently used technologies but also provide novel services, such as detecting non-cooperating persons, and increases the level of convenience, for instance, by not having to wear an actual monitoring device. Based on these findings and the truly pervasive character of the underlying physical entity we believe that RF-based sensing can be essential in the pervasive systems of the upcoming Internet of Things.

2.4 Monitoring of Attention Using Ambient FM-radio Signals [15]

We investigate the classification of FM-radio signal fluctuation for the monitoring of attention by individuals in motion towards a static object. In particular, we distinguish in a corridor, whether it is empty or populated by moving or standing individuals as well as the attention of these subjects towards poster frames in that corridor. We consider the distinction in front of which poster these subjects are walking or standing as well as their walking speeds or changes therein. This information can provide some hint whether a person is paying attention to a specific poster in this corridor as well as the location of the particular poster.

2.4.1 Introduction

Attention determines for a system the potential to impact the actions and decisions taken by an individual [139]. The management of attention covers the activation of attention as well as its detection and timely exploitation. The same action of the same system might be considered either as annoyance or be appreciated as helpful depending on whether the individual was focusing part or all of her attention towards the system or not. In the literature, we find various definitions that classify attention as well as its determining characteristics [140]. A straightforward measure of attention might be the tracking of gaze [141]. In general, aspects such as Saliency, Effort, Expectancy and Value are important indicators of attention [142]. Alois Ferscha and others extended this model and put a greater stress on the effort a person takes towards an object [143].

We consider the following scenario. In a corridor, a series of electronic poster frames are installed while people are walking by these frames. From the perspective of a specific poster, a significant part of its message shall be recognised by passers-by. Therefore, the poster should draw the attention of people passing by and, when this is achieved, it might possibly transport additional information. Consequently, the poster frame should be aware of people passing by, know where people are in order to attract attention at the right moment and detect whether attention is attracted. In this work we assume that the monitored individuals are not cooperating with the system and hence are not equipped with any part of the sensing hardware.

Such detection and management of attention may require elaborate installations and very specific sensors in order to accurately sense quantities such as Saliency, Effort, Expectancy and Value [139]. However, we believe that for many commercial installations, cost and ease of installation and not primarily the highest achievable accuracy are most important.

Also more general, environmental sensors can provide sufficient information to estimate the attention state of individuals.

We propose to utilise ambient FM-radio signals for the detection of attention since it has a nearly perfect coverage in populated areas and features cheap receiver hardware [144].

... but which aspects of attention can actually be captured by an FM-receiver?

Ferscha and others [143] discuss various aspects of attention and identify as most distinguishing factors changes in walking speed, direction or orientation. From FM-radio signals it is hard to detect the orientation of a person. However, it is feasible to classify walking speeds, walking direction or location of individuals. We show that with a straightforward installation, we can distinguish

- an empty corridor, a person walking by and a person standing in front of a poster frame

- the specific poster the person is observing

- the location where a person is walking

- the walking speed of a person

We therefore argue that attention levels can be inferred upon interpretation of the changes in walking speed or direction as derived from our system. This information can provide some indication on the attention of persons towards a poster frame. Also, it can enable a frame to take action in order to catch the attention of a person just in the right moment.

2.4.2 Sensing passive entities from the RF-channel

The RF-channel has been recently utilised by various authors for the detection of location or activities of passive entities, not equipped with a transmitter or receiver (See [145] and references therein). These studies exploit fluctuation of received signal strength or signal amplitude conditioned on changes (for instance, movement, altered location of objects) in the physical proximity of a receiver [76]. This work is related to the passive radar literature. In an early publication, Kaipin Tan et. al presented the potential of a GSM-based passive radar prototype for detecting and tracking different types of ground-moving targets [146] in an outdoor environment. For the simple binary detection of presence in an indoor environment, Masahiro Nishi et. al proposed an indoor human detection system leveraging the multi-path radio propagations of VHF-FM and UHF-TV broadcasting signals [147]. In their work, they leveraged incoming signal waves. Moustafa Youssef and others then demonstrated the localisation of individuals by analysing the Received-Signal-Strength indicator (RSSI) in received packets at 802.11b nodes [78]. Neal Patwari and Joey Wilson introduced a statistical, empirically verified model to approximate the position of a person based on the variance in the Received Signal Strength Indicator [126]. The aspect of localising up to five individuals at a time, together with the previously untackled problem

that environmental changes over time might necessitate frequent calibration of the location system was approached by Dian Zhang and others using a grid of wireless sensor nodes in [125]. They isolated the Line-of-Sight path by extracting phase information from the differences in the received signal strength on various frequency spectrums at distributed nodes.

These studies on localisation of individuals assume a transmitter as part of and under the control of the recognition system. However, we recently demonstrated that a recognition is also possible when an ambient signal source is utilised. In particular, we analysed fluctuation in ambient signals from an FM-radio station not under the control of the recognition system. Static environmental changes such as opened doors have been detected with an accuracy of about 0.9 and a first study on suitable features to detect human activities could achieve an accuracy of about 0.8 with a two-staged recognition approach [82]. In this article we extend this work towards the monitoring of attention of a single subject.

2.4.3 Monitoring attention from FM-radio

For the monitoring of attention of non-cooperating individuals, an environmental signal source is required. We propose the utilisation of RF-signals for their ease of deployment and nearly perfect coverage in indoor and outdoor locations through existing, pre-installed systems in populated areas [144], argue why we favour FM-radio above other RF-technologies and detail the features we utilise for the detection of attention.

Radio-based monitoring of attention

Radio waves are electromagnetic waves, defined by their amplitude, phase and frequency. During signal propagation from the transmitter to the receiver, the radio waves are impacted by physical phenomena, for instance, damping, reflection and scattering. Assume a signal observed at a receiver, at some frequency f_c [Hz]. Naturally, as we are considering an indoor environment in which no direct-line-of-sight exists, incoming signals arrive over mutliple paths at roughtly equal strenth from all directions. In the event that a signal wave encounters any structure such as an object or individual, the main signal component will be damped (continue its path with reduced energy) or even completely blocked. Additionally, the signal is typically reflected or scattered at this occasion. Reflection describes the event that the signal wave bounces away from an object in a modified direction. Typically, the signal will also experience scattering, which is the splitting of a signal wave due to the not perfectly even structure of an encountered object and the propagation of these signal components into diverse directions.

Therefore, the composition of incoming signal components at a receiver is conditioned on the movement and position of objects in its proximity [145]. A change in position of objects or static activities like standing will generally affect the mean amplitude, while movement, such as walking, induces a characteristic pattern on the signal over time [82]. Figure 2.24 illustrates the received signal's strength and some extracted feature sequences from an ambient FM-radio station over 1 minute for 3 different situations of a single subject
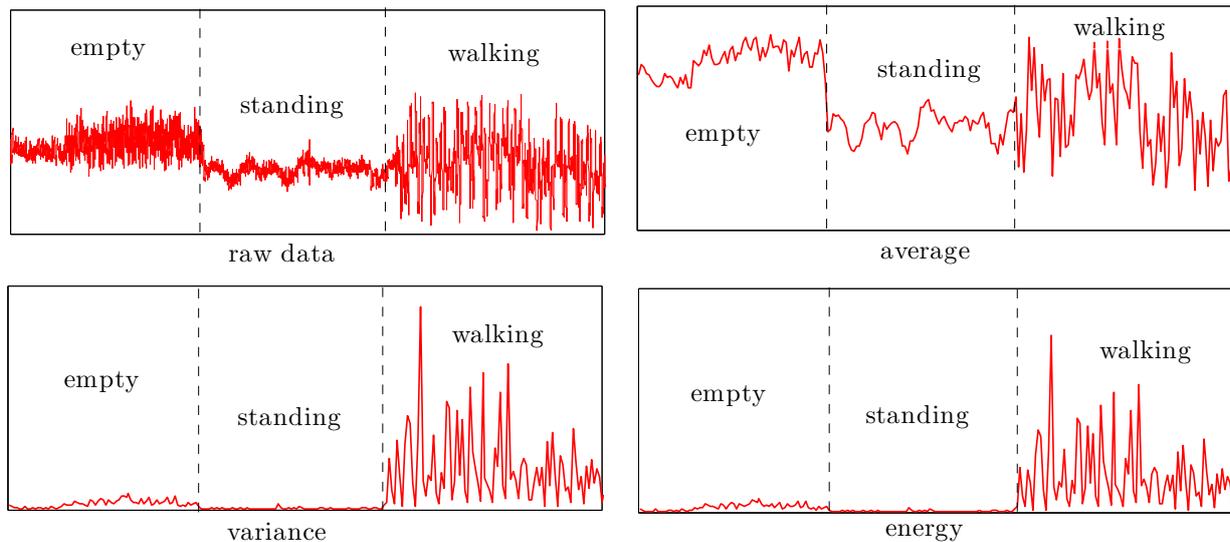
Figure 2.24: Evolution of signal strength for empty corridor, standing and walking at the area B, performed by a single subject

in a corridor (the setting is illustrated in Figure 2.25). Figure 2.24 indicates a correlation between the characteristics of an RF signal and the activities conducted.

Consideration of various RF-technologies

In the literature, RF-sensing is applied on various signal frequencies and technologies such as WiFi, GSM or FM-radio [119, 78]. We believe that FM-radio is best suited for attention monitoring for the following reasons.

Since FM-radio features a low operating frequency, a simple modulation mechanism and a wide area of coverage, it is possible to design more robust and discriminative signatures for RF-fingerprinting than for WiFi and GSM [144].

FM-radio signals experience, when compared with WiFi, 3G or 4G signals, lower variation in signal strength over time [144]. Consequently, for attention monitoring, FM-radio signals induce a lower process noise than signals from WiFi, 3G or 4G systems. Also, FM-radio is, compared to the other named systems which operate at higher frequencies, less susceptible to weather conditions, such as rain and fog [144].

Additionally, in order to increase spectrum efficiency, spread spectrum techniques such as frequency hopping or code divisioning are employed in WiFi, 3G and 4G access points. CDMA interleaves the transmissions to multiple devices including additional potential noise and for hopping schemes, a passive recognition system would need to follow this RF-signal activity variations in order to extract the signal fluctuations on-top which are caused by activities in proximity. This would be a more difficult task. Therefore, the use of FM-radio signals which are not conditioned on the use of such data transfer schemes appears to be a more appropriate choice.
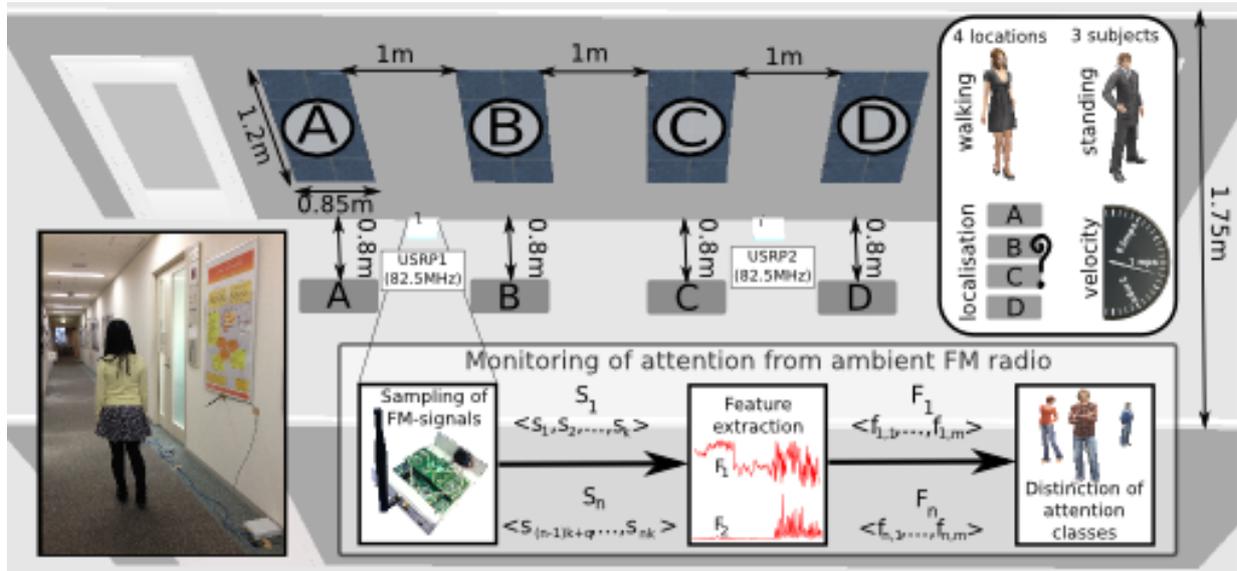
Figure 2.25: Sketch of the evaluation setting. The attention-monitoring system extracted features combining the data acquired by both USRP devices.

Furthermore, FM-radio stations are widely implemented and continuously broadcast signals with higher coverage than WiFi, 3G or 4G systems. Finally, an FM-radio receiver is less costly than receivers for the other mentioned systems.

For these reasons, we believe that FM-radio is best suited for the utilisation in a passive attention monitoring system.

Features for FM-based attention monitoring

We extract features for attention monitoring from FM-signals continuously broadcast by an FM-radio station (cf. figure 2.25). The features are obtained from a series of continuous measurements $s_1, s_2, \ldots, s_t$ which are samples of the amplitude of ambient FM signals and grouped in windows $S_1, \ldots, S_n$ of k consecutive samples each $\langle s_{(i-1)k+1}, s_{(i-1)k+2}, \ldots, s_{ik} \rangle$. From these S_i, sets of features $F_i = \langle f_{i,1}, f_{i,2}, \ldots, f_{i,m} \rangle$ are extracted and used for the monitoring of attention. The features we utilized are the mean (Avg_i), the variance (Var_i) and the energy (E_i) of S_i as detailed in figure 2.26. These features have been derived among a greater set of features as well suited to achieve high accuracy for activity recognition conditioned on passive RF-based recognition systems in [145]. After the extraction of features, we randomly divided the collection F of all feature sets F_i into a training set Tr and a classification set Cl that met the conditions $\mathrm{Tr} \cup \mathrm{Cl} = \mathrm{F}$ and $\mathrm{Tr} \cap \mathrm{Cl} = \emptyset$. The set Tr is used to train the classifiers. After the training, classifiers will process Cl.

For a set of k activities $\mathcal{A} = \{a_1, \ldots, a_k\}$ let $\mathcal{I}(a_i)$, with $\mathcal{I}(a_1) \cup \mathcal{I}(a_2) \cdots \cup \mathcal{I}(a_k) = \mathrm{Cl}$, be the total number of instances for activity a_i and $\mathcal{I}_{\mathrm{cor}}(a_i)$ be the number of correctly classified instances for this activity in which the classification matches the ground truth.

<table>
<tr><td>

Average signal strength

The average signal strength over a window of measurements represents static characteristic changes in the received signal strength.

It provides means to distinguish a standing person as well as her approximate location.

$$\text{Avg}_i = \left(\frac{\sum_{t=(i-1)n+1}^{in} \text{s}(t)}{n} \right)^2$$

</td><td>

Normalised spectral energy

The normalised spectral energy is a measure in the frequency domain of the received signal.

It has been used to capture periodic patterns such as walking, running or cycling.

$$\text{E}_i = \sum_{k=1}^{n} \text{P}_i(k)^2$$

Here, $\text{P}_i(k)$ denotes the probability or dominance of a spectral band k:

$$\text{P}_i(k) = \frac{\text{FFT}_i(k)^2}{\sum_{j=1}^{n} \text{FFT}_i(j)^2}$$

As usual, we calculate the k-th frequency component as

$$\text{FFT}_i(k) = \sum_{t=(i-1)n+1}^{in} \text{s}(t)e^{-j\frac{2\pi}{N}kt}$$

</td></tr>
<tr><td>

Variance of the signal's strength

The variance of the signal strength represents the volatility of the received signal.

It can provide some estimation on changes in a receiver's proximity such as movement of individuals.

$$\text{Var}_i = \sqrt{\frac{\sum_{t=(i-1)n+1}^{in} (s(t) - \text{Avg}_i)^2}{n}}$$

</td></tr>
</table>

Figure 2.26: Features utilised for the monitoring of attention via received FM-signals (Published by the IEEE CS n 1536-1268/14/\$31.00 ©2014 IEEE

We define the accuracy by which an activity a_i can be detected as

$$\mathcal{ACC}(a_i) = \frac{\mathcal{I}_{\text{cor}}(a_i)}{\mathcal{I}(a_i)}$$

2.4.4 Evaluation

In this section, we discuss case studies to demonstrate the viability of monitoring attention of people passing by several poster frames towards these frames. In all cases, we consider a corridor with posters attached along one side (cf. figure 2.25).

Four posters of 0.85 m × 1.2 m which are separated by 1 m are attached alongside one wall of the corridor. We place the USRP devices between the two leftmost and rightmost poster frames on the floor. These N210[16] USRP devices are equipped with WBX[17] daughter boards and VERT900[18] antennas with 3 dBi antenna gain. Both devices continuously recorded the signal strength with a sample rate of 64 Hz, emitted by an ambient FM-radio station at 82.5 MHz while the attention of subjects towards the poster frames is monitored. We distinguish between four locations, 0.8 m in front of the posters (labelled A, B, C, and D) and the rest of the corridor. During the case studies, the subjects were walking along the corridor and through the marked areas or standing in front of one of the posters at the

---

[16]https://www.ettus.com/content/files/2987_Ettus_N200-210_DS_FINAL_1.27.12.pdf
[17]https://www.ettus.com/product/details/WBX
[18]https://www.ettus.com/product/details/VERT900

Table 2.18: Mean accuracy for the distinction of the corridor states 'empty', (person)'standing' and (person)'walking' (Published by the IEEE CS n 1536-1268/14/$31.00 ©2014 IEEE

	Classification		
	empty	standing	walking
Truth — empty	**.906**	.034	.06
standing	.136	**.765**	.099
walking	.021	.100	**.879**

(a) Classification accuracy achieved by a k-NN classifier

	Classification		
	empty	standing	walking
Truth — empty	**.877**	.064	.059
standing	.041	**.852**	.107
walking		.071	**.929**

(b) Classification accuracy achieved by a DT classifier

marked locations. As a baseline, the received signal from the empty corridor was recorded. For each action and all three subjects about two minutes of sample data each have been collected.

For all features detailed in figure 2.26, we utilise a window of 128 signal measurements, spanning a total of 2 seconds. Features are extracted from the data sets collected by USRP1 and USRP2 (cf. figure 2.25) and are merged for the distinction of attention classes. For this, we utilise a decision tree (DT) and a k-nearest-neighbour (k-NN) classifier from the Orange data mining Toolkit[19]. The k-NN classifier utilised 5 neighbours and weights their distance by the Euclidean distance. The decision tree utilises at minimum 10 instances in its leaves for pre-pruning and a recursive merge of leaves of the same major class with an m-estimate of 2 for post pruning. We apply a 10-fold cross validation.

State of the corridor

In our scenario, a corridor is equipped with electronic poster frames which shall detect the attention of passers-by and act accordingly. The most basic case to distinguish for the frames is the state of the corridor. In particular, we consider whether the corridor is empty or occupied by a person and, when it is occupied, whether this person is walking or standing. In the case of electronic poster frames, the devices might change into an energy saving mode when the corridor is empty or also display more or less complex information conditioned on whether the person in the corridor is walking or standing.

Table 2.18 depicts the classification accuracy for these classes. For all classes, the mean classification accuracy over the sample windows of 2 seconds is near or above 0.8. In a second stage, we can now obtain information related to the attention of passers-by.

Focused attention towards specific frames

While walking by poster frames, brief snippets of the content can be grasped by individuals. However, an intense engagement with the more complex content of a poster requires a

---

[19]http://orange.biolab.si/

Table 2.19: Mean accuracy for the distinction in front of which poster a person is standing

	Classification (standing at)						Classification (standing at)			
(Truth)	Loc.A	Loc.B	Loc.C	Loc.D		(Truth)	Loc.A	Loc.B	Loc.C	Loc.D
Loc.A	**.876**	.011	.022	.090		Loc.A	**.933**		.011	.056
Loc.B		**.788**	.203	.008		Loc.B		**.735**	.257	.009
Loc.C		.063	**.929**	.009		Loc.C		.080	**.911**	.009
Loc.D	.009	.027		**.964**		Loc.D	.054		.027	**.919**

(a) Classification accuracy achieved by a k-NN classifier

(b) Classification accuracy achieved by a DT classifier

person to slow down her walking speed [143] and possibly come to a stand in front of the poster. We demonstrate the distinction in front of which poster a person is standing in the scenario depicted above. All parameters of the recognition system remain identical to section 2.4.4.

All subjects have been standing and observing a poster at one of the locations labelled A, B, C or D in figure 2.25.

The most characteristic feature to distinguish these cases is the mean of the signal strength. The average classification accuracy after 10-fold cross validation is depicted in table 2.19 We observe that the classification accuracy is in most cases above 0.9, in all cases it is near or above 0.8.

Tracking individuals in motion

While people are passing by poster frames in a corridor, a specific poster frame might have the intention to actively attract the attention of a passer-by. This attempt is most successful when the person is in the proximity of the poster, facing towards it. In order to optimally schedule such action, the location of the walking person has to be available at the system. We show that the location of a single person walking along a corridor can be traced by analysing fluctuation of an incoming FM-radio signal. Similar to the case study detailed in section 2.4.4 we detect in front of which poster a person walking in the corridor is located.

Table 2.20 depicts our results. We observe that the classification of the location where a person is walking is harder than the classification of the location where a person is standing. However, the classification accuracy reached is still near or above 0.8.

Changes in the walking speed

As detailed in [143], an important indicator of the attention state of a person are changes in the walking speed. When a person is interested in a specific content of a poster, she might likely slow down to better perceive the content.

Table 2.20: Mean accuracy for the distinction of walking at location A, B, C or D in the environment depicted in figure 2.25

	Classification (walking at)			
	Loc.A	Loc.B	Loc.C	Loc.D
Loc.A	**.779**	.118	.015	.088
Loc.B	.107	**.804**	.071	.018
Loc.C		.017	**.933**	.05
Loc.D	.139			**.785**

(a) Classification accuracy achieved by a k-NN classifier

	Classification (walking at)			
	Loc.A	Loc.B	Loc.C	Loc.D
Loc.A	**.754**	.228	.018	
Loc.B	.175	**.772**	.035	.018
Loc.C	.029	.017	**.953**	
Loc.D	.125		.071	**.804**

(b) Classification accuracy achieved by a DT classifier

Table 2.21: Confusion matrices for the discrimination between walking speeds (0.5 m/s, 1 m/s, 2 m/s) achieved by k-NN and Decision Tree classifiers

	Classification		
	0.5m/s	1m/s	2m/s
0.5m/s	**.641**	.219	.141
1m/s	.109	**.453**	.438
2m/s	.134	.284	**.582**

(a) Classification accuracy achieved by a k-NN classifier

	Classification		
	0.5m/s	1m/s	2m/s
0.5m/s	**.844**	.094	.063
1m/s	.094	**.578**	.328
2m/s	.164	.343	**.493**

(b) Classification accuracy achieved by a DT classifier

We obtain the walking speed of a passer-by from the fluctuation in ambient FM-radio signals. We collected for all three subjects and for three different velocities (0.5 m/s, 1 m/s, 2 m/s) samples of a duration of 2 minutes each. Again, k-NN and DT classifiers are utilised for training and classification.

Table 2.21 illustrates our results. We observe that, although there is an indication towards the correct velocity in all cases, the accuracy greatly drops compared to the previous considerations. The confusion of these velocity levels especially for higher walking speeds is owing to the reduced duration an individual is located in front of a single poster during her walk.

We can, however, achieve a higher recognition accuracy without increasing the distance between posters by abstracting from the 1 m/s walking speed (cf. table 2.22), distinguishing only between a slow walk and a running person. Although we are then not able to distinguish the medium walking speed, note that the attraction of attention of a person in a hurry is not the intention of the considered system. Rather, we are focusing towards individuals in a relaxed, open state of mind to receive external stimuli and information.

Since the change in walking speed at a particular location might correspond to the

Table 2.22: Confusion matrices for the discrimination between walking speeds (0.5 m/s, 2 m/s) achieved by k-NN and Decision Tree classifiers

	Classification	
	0.5m/s	2m/s
Truth 0.5m/s	**.896**	.104
Truth 2m/s	.219	**.782**

(a) Classification accuracy achieved by a k-NN classifier

	Classification	
	0.5m/s	2m/s
Truth 0.5m/s	**.925**	.075
Truth 2m/s	.025	**.750**

(b) Classification accuracy achieved by a DT classifier

attention level of passer-by, the information on the walking speed, monitored over time, can be utilised to grasp her attention level.

Altering the count of receive devices

In the above considerations, we have utilised two USRP devices since the experimental setting spans over five meters and the classification accuracy deteriorates with increasing distance to the receive antenna [82]. However, for economic reasons, a simple installation might be designed in favour of only one receive device at the cost of a slightly reduced recognition accuracy for greater distances. In order to evaluate this impact for the monitoring of attention of passers-by within a corridor, we also consider the classification accuracy when the data from only one of the receive devices is utilised. The classification system and location of receive devices was not changed. Figure 2.27 depicts our results. We observe that the classification accuracy benefits from the addition of the second device in all cases. With only one device, the overall classification accuracy drops by about 0.05 to 0.1 since the classification accuracy for individuals in greater distance deteriorates.

2.4.5 Discussion

The monitoring of the attention state of passers-by towards interactive poster frames can provide additional information to the display system when to display which information.

We demonstrated the distinction of attention classes from features extracted from ambient FM-radio signals. In particular, we utilised the mean, variance and energy of a signal received at 82.5 MHz in order to distinguish occupancy states in a corridor as well as locations at which persons are walking or standing and finally walking speeds. The attention level of persons is, among other factors, related to walking speeds or changes in velocity or acceleration. Therefore, we can use the information extracted from the fluctuation in the received FM-signal as an indicator towards various attention states. This information might control the information provided by a poster frame, conditioned on the attention of passers-by.

Due to the low cost of FM-receiver hardware and the high coverage of FM-radio, the
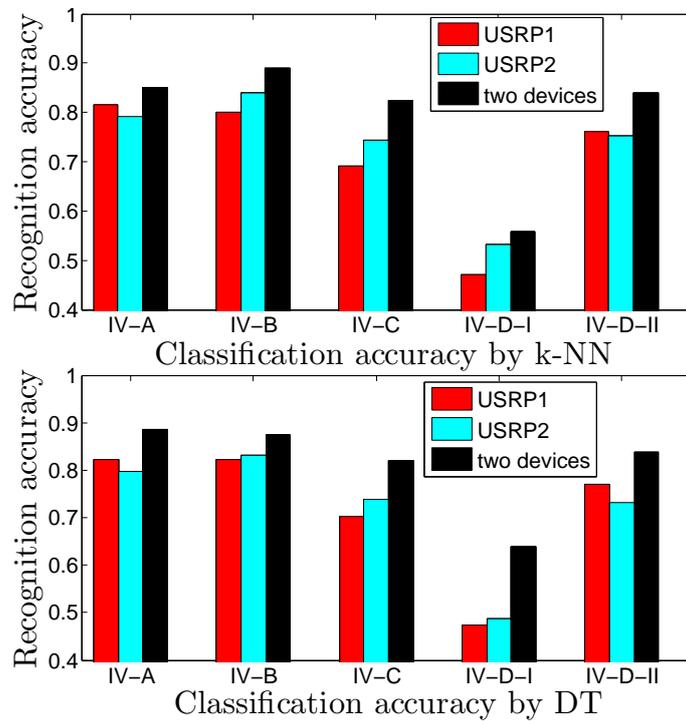
Figure 2.27: Comparison of the classification accuracy for the 4 cases described in section 2.4.4 when only one receive device is utlised

described attention-monitoring system has the potential to be widely deployed in systems that benefit from knowing the attention levels of people in proximity. Future directions cover the simultaneous detection of attention levels of multiple persons as well as the implementation of the system using off-the-shelf receiver hardware.

2.5 The Telepathic Phone: Frictionless Activity Recognition from WiFi-RSSI [20]

We investigate the use of WiFi Received Signal Strength Information (RSSI) at a mobile phone for the recognition of situations, activities and gestures. In particular, we propose a device-free and passive activity recognition system that does not require any device carried by the user and uses ambient signals. We discuss challenges and lessons learned for the design of such a system on a mobile phone and propose appropriate features to extract activity characteristics from RSSI. We demonstrate the feasibility of recognising activities, gestures and environmental situations from RSSI obtained by a mobile phone. The case studies were conducted over a period of about two months in which about 12 hours of continuous RSSI data was sampled, in two countries and with 11 participants in total. Results demonstrate the potential to utilise RSSI for the extension of the environmental perception of a mobile device as well as for the interaction with touch-free gestures. The system achieves an accuracy of 0.51 while distinguishing as many as 11 gestures and can reach 0.72 on average for four more disparate ones.

2.5.1 Introduction

Mobile phones are a popular sensing platform for the multitude of sensors they incorporate and for their status as personal device kept close to or on the body [148, 149]. However, these mobile sensing platforms focus on inertial motion to recognize physical activity. When a device is no longer worn on the body, its sensing capabilities are greatly reduced. Indeed, although people are in the same room with their mobile device almost 90% of the time, their device is within arms reach less than 55% of a day [93, 94]. Therefore, the mobile phone can hardly serve as a continuous sensing platform with sensors such as accelerometers or gyroscopes.

To still obtain information about situations or activities, we need to exploit sensors that react on ambient stimuli. Possible choices are video [150], or audio for the classification of device-locations based on audio signatures [151] as well as localisation via audio-based fingerprinting [7]. However, video is restricted by the sensor's field of vision while audio is limited to general locations or situations [152].

We propose the use of another environmental sensor: the wireless interface to the Radio Frequency (RF) channel. By monitoring the fluctuation in the received signal strength indicator (RSSI) that is calculated at a receiver for each incoming packet, we attempt to classify the situation (e.g. crowd size), activities or gestures performed in proximity of a mobile phone (See figure 2.28). This approach allows operation even when the device is not carried by the user but near to her – a scenario where most activity recognition systems

---
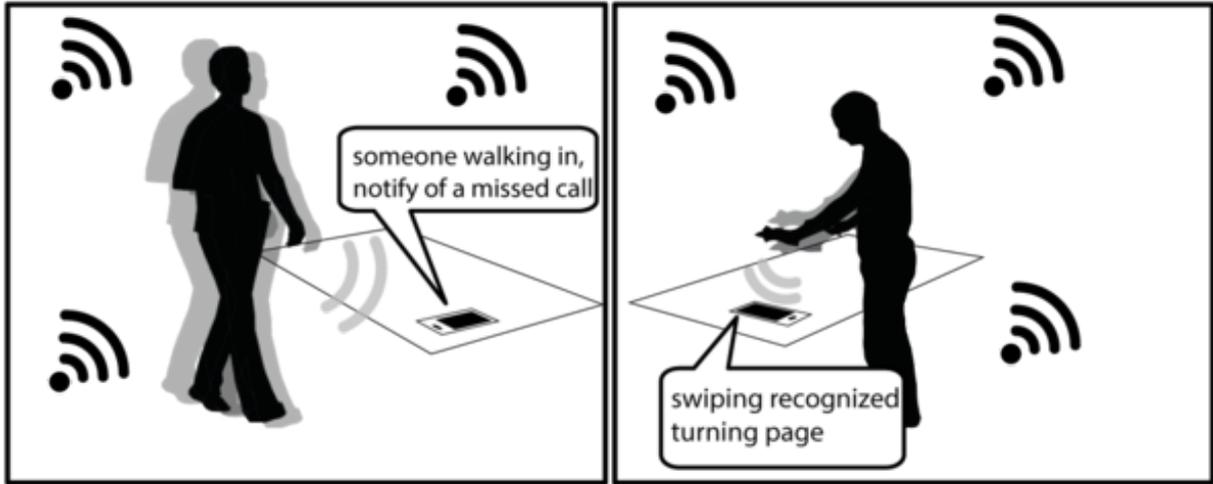
Figure 2.28: Activity obtained from RSSI-signatures. Two example use-cases: user walking in with the smartphone implicitly reacting (left) and a no-touch explicit interaction (right). (978-1-4799-3445-4/14/$31.00 ©2014 IEEE)

fail.

We can utilise RSSI also in dark or quiet environments when audio or video might not provide sufficient information. In urban spaces, WiFi connectivity can be presumed (cf. section 2.5.5). In addition, RF might be perceived as less privacy intrusive when compared to audio or video.

While there is some work on the device-free recognition of activities from RF-channel fluctuation [153, 154, 3], these systems require sophisticated Software Defined Radio (SDR) devices in order to obtain frequency domain features. In contrast, we attempt to utilise signal strength fluctuation on off-the shelf mobile phone hardware and from ambient WiFi traffic. On such devices, already the capturing of RSSI data in sufficient frequency is challenging. In addition, the data captured is less accurate and bursty. We discuss necessary pre-processing as well as the design of features suitable for highly bursty and low-resolution environmental RSSI data together with the final recognition step. In case studies we demonstrate the potential and limitations of using RSSI for recognition. The contributions of this work are:

1. System design and definition of feature space for RSSI-based 'frictionless' recognition

2. Analysis of RSSI-influencing factors in a controlled setting (e.g. direction, distance)

3. Feasibility study of situation, activity, and gesture recognition with off-the-shelf mobile phones.

Our results indicate that RF-based sensing of environmental *situations*, *crowd* and *individual activity* provides additional information for activity or context classification tools.

2.5.2 Related Work

Device-free RF-based recognition was introduced by Youssef and others [78] as the localisation of an entity not equipped with any transmitter or receiver. In recent years, some groups work in this direction using hardware that ranges from SDR devices [155], laptop-class computers [79] over sensor nodes [125] or RFID tags [156] and achieve high accuracies of about 1 meter. This work is also related to a considerable body of practical and theoretical results on passive radar (cf. [146, 157] and references therein) where vehicles and individuals are detected and tracked from signals such as HF radio, UHF television broadcasts or DAB, DVB and GSM.

Recognition utilising signals on the wireless channel has been generalised in [3] to activities and we can further imagine also situations [75], gestures [154] or attention [4] to be identified by RF-based device-free implementations. These systems can be grouped into active and passive approaches conditioned on the presence of an active transmitter. Most previous work in this direction uses SDR devices.

Kassem et al. sense traffic situations by tracking frequency and speed of passing cars that intercept the direct line of sight between a pair of nodes [158]. The authors of [3] classify simple activities in an SDR-based active device-free system by extracting and interpreting features from a continuous signal between two nodes. Their approach explores also the multipath effects induced by persons that are not intercepting the direct path between nodes. It was later demonstrated that also simultaneously conducted activities from multiple persons can be distinguished by leveraging purely signal-strength based features [159]. Furthermore, it was shown by Pu and others that simultaneous detection of gestures from multiple individuals is possible utilising multi-antenna nodes and micro Doppler fluctuations [154, 160]. In a related system, Adib and Katabi employ MIMO interference nulling and combine samples taken over time to achieve the same result while compensating for the missing spatial diversity in a single-antenna system [153].

While the above are active approaches that require a dedicated transmitter, Ding and others have presented a passive system leveraging RF noise from engines of vehicles [161]. In addition, Shi et al. recognised activities and locations from fluctuation in the signal strength of broadcast FM radio [82].

Also, active systems utilising non-SDR nodes have been studied. Most notably, Patwari and others estimated the breathing frequency of an individual surrounded by nodes from the RSSI of exchanged packets [124]. Following other directions, Xu et al. have counted crowd [162] from RSSI within a field of sensor nodes. Their unsupervised learning approach is able to predict the count of up to 10 stationary or moving individuals. Recently, the recognition of general activities from RSSI in a sensor network has been considered [163]. In particular, the activities standing, sitting, lying, walking and empty have been distinguished with an accuracy of 0.8-0.9.

For these studies, either a sophisticated SDR device or transmit-receive pairs of nodes were required. Both cases are hard to establish with end-user equipment in spontaneous use. We propose a usable RF-based device-free recognition approach on phones by leveraging received RSSI from packets of WiFi access points (APs). We are not aware of previous

work on such RSSI-based passive device-free recognition system.

2.5.3 Capturing RSSI on Phones

In IEEE 802.11, data is exchanged in packets on 11 partly overlapping frequency channels. In normal communication, a WiFi receiver discards all packets not addressed to itself. However, we can force the interface into monitor mode to log all traffic. For each packet, the receiver calculates the signal strength from the 8 bit preamble. Due to the lower data rates, control packets differ in their estimated RSSI significantly.

While the APIs of contemporary mobile phone operating systems (OSs) provide means to access the RSSI, this information is averaged and refreshed at about 1 Hz only. Another access to the RSSI is possible via the interface directly with tools such as `airodump-ng` or `tcpdump`. This requires root permissions to access the interface in monitor mode.[21] WiFi-firmware with sufficient access to relevant parameters is sparse. More severe even for mobile phones, most handsets implement a similar chipset family (e.g. Broadcom bcm4329, bcm4330(B1/B2), bcm4334, bcm4335) for which the default firmware does not provide access to the desired information (even as root). The only solution to avoid root access and which abstracts from this chipset family is via an external antenna[22]. However, this considerably extends the dimensions and complexity of the hardware, so that we decided against it. Instead, we used a modified firmware for the above mentioned chipset family [164] on a Nexus One phone running Cyanogen mod 7.2 and executed tcpdump on the interface in monitor mode to capture RSSI of packets. In monitor mode, no data can be transmitted and consequently no impact can be taken on the frequency in which packets are received. We can, however, adjust the channel we listen on and might utilise data from multiple APs transmitting on the same channel. In summary, while it is practically possible to monitor RSSI, the support of manufacturers for the operating systems to perform this out of the box is limited. However, we can track RSSI fluctuation with a modified OS, but without hardware modifications.

Figure 2.29 shows an exemplary snippet of sampled RSSI. In the experiments conducted, the RSSI usually ranged from -98dBm to -47dBm. Since the RSSI calculated for control packets differs, we disregarded them for the generation of this data. At the time of this recording, the phone was lying on a table within approximately 0.5 meters distance of a person sitting at that table. We observe that the data is very bursty. While there might be only one packet within 0.1 seconds at times, we can also observe five or more packets in the same interval. Clearly, when compared to SDR-based recognition systems that have direct access to the physical channel, the amount of information available from RSSI is severely reduced. Even compared to active RSSI-based systems that contain a transmitter omitting packets at high rate, our passive approach has to deal with more bursty traffic and a lower packet arrival rate. In addition, the granularity of RSSI is low. In our case, the 1dB granularity observed in the figure could not be improved for the WiFi interface.

---

[21]Monitor mode is obligatory in our case since otherwise the tool is executed in Ethernet emulation which does not provide RSSI information

[22]github.com/brycethomas/liber80211/blob/master/README.md
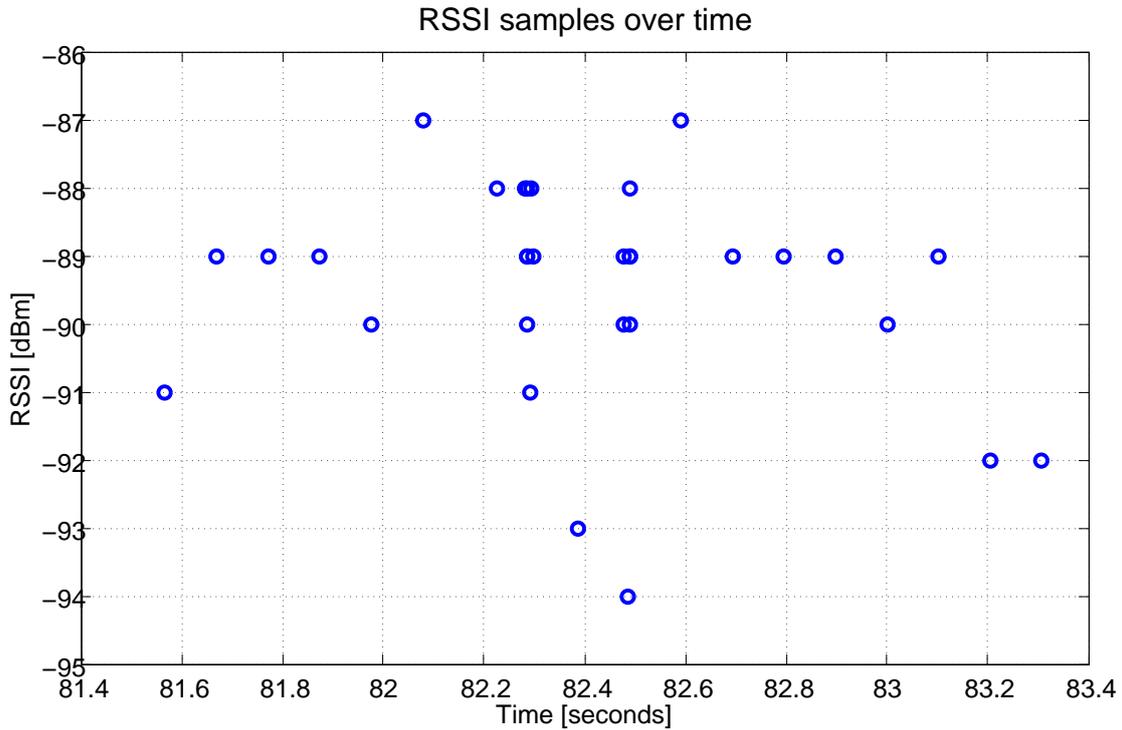
Figure 2.29: RSSI from packets of a single AP (978-1-4799-3445-4/14/$31.00 ©2014 IEEE)

We conclude that it would be hard to apply any curve fitting that could successfully predict the RSSI evolution at a higher sample rate.

2.5.4 Features for RSSI-based Recognition

Considering this structure of the data, we used simple features that express general properties such as the overall weight or mass as well as their spread.[23] As a tribute to the bursty traffic, the low granularity (cf. figure 2.29) and a fluctuating packet arrival rate, we simply fixed non-overlapping windows of two seconds and then utilised all RSSI values that would arrive during this period for feature calculation. The window length was set to 2s since we aim to design a system that would be practically usable with a good response time. A higher accuracy can be achieved with increased window size or via majority votes over successively calculated features (cf. section 2.5.5). In total, 18 different features have been considered. On a data set with the three basic cases

1. A phone lying on a table in an empty room

2. A phone lying on a table with a person moving

---

[23]No frequency domain features could be used; Features as zero crossings or direction changes were not meaningful on the undersampled signal.

3. A person holding and handling the phone

we applied a feature selection from the orange data mining toolkit [24]. From the remaining 9 features, we manually tweaked a combination that achieves good accuracy. Several combinations of *mean, median, variance, maximum* and the *difference between minimum and maximum* could achieve best and comparable classification results. For the case studies (section 2.5.5), we decided for a combination of *mean, variance, maximum* and *difference between maximum and minimum.* For the gesture recognition, also the slope was considered.

2.5.5 Case Studies

We conducted case studies in indoor environments at ETH Zurich and TU Braunschweig (cf. figure 2.30). Occasionally the phone was connected to a computer via `adb shell` as an alternative to the slow on-screen keyboard which made no difference for the recorded data. All recordings were conducted multiple times and over several days. We intentionally altered the environments between recordings (e.g. moving furniture, placing the device slightly different).

Data was processed off-line. However, we have developed a toolchain for the processing and classification that is sufficiently lightweight to be executed on the phone in realtime.[25] The tool groups packets for their source address (since the mean RSSI differs among senders) and disregards control packets (since also their RSSI level differs).

We now consider general RSSI properties and then investigate limits of RSSI-based recognition. The studies were conducted over two months in Braunschweig, Germany and Zurich, Switzerland. A total of 11 persons (9 male and 2 female; 26 to 37 years) have participated and overall about twelve hours of continuously sampled data has been produced.

First, we investigate properties of urban WiFi with respect to traffic and sampling rate (section 2.5.5). Then, we study coarse characteristics with respect to the presence of a user. Finally, we provide experiments on fine-grained gesture recognition.

WiFi Traffic in Urban Spaces

For the recognition of activities and gestures from RSSI, the rate of incoming packets is essential since this is the rate of fluctuation induced by environmental stimuli. We sampled packets over some days at various locations in Zurich to estimate a typical rate of packets in urban places. Figure 2.31 shows the number of packets per second from the most active AP at various locations on all channels. Short packets, such as acknowledgements, were removed (cf. section 2.5.4).

---

[24]http://orange.biolab.si/

[25]The python tools to extract and process RSSI information from pcap files and to classify situations are available at http://www.stephansigg.de/DeviceFree/pcapTools.tar.gz
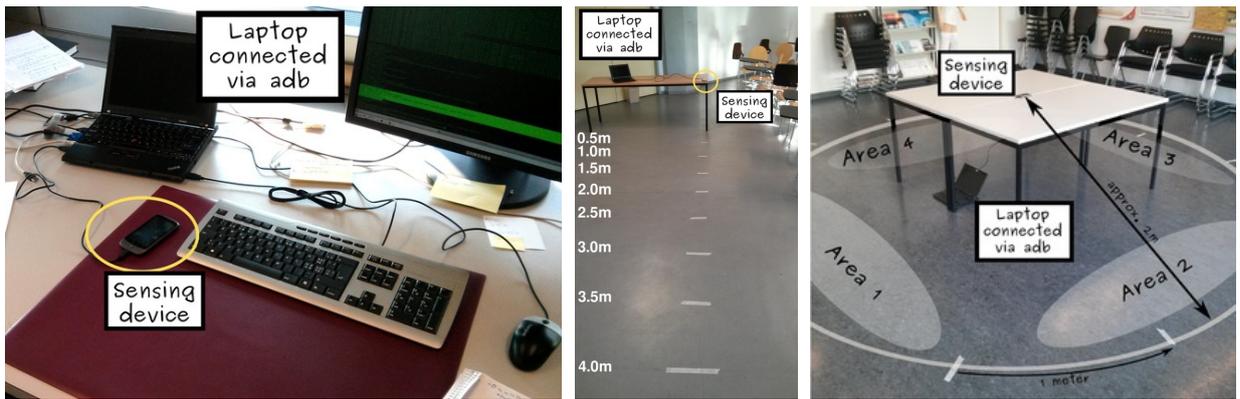
(a) Office environment at ETH

(b) Lecture room at TU-BS

(c) Scenario for the distinction of walking speed
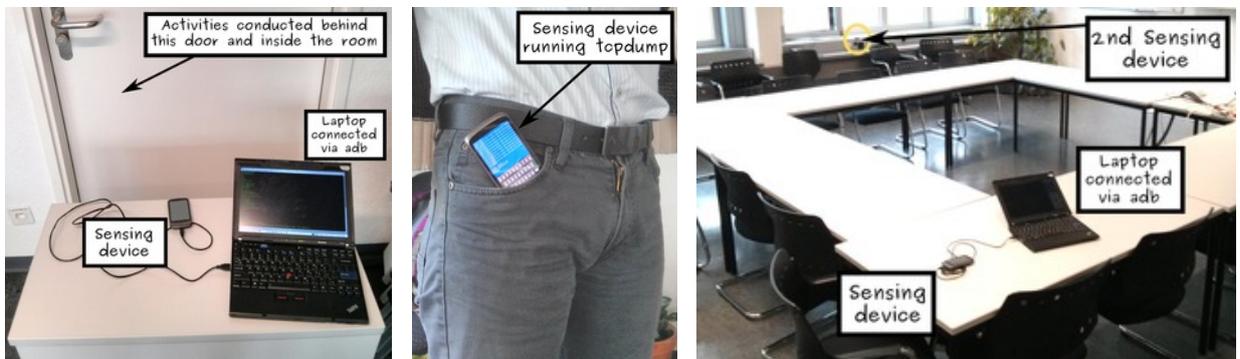
(d) Activities conducted behind a closed door

(e) Sensing device inside pocket

(f) Meeting room at ETH

Figure 2.30: Environments for our case studies. Surrounding furniture and objects were intentionally altered in all cases.

Train station

channel	Packets/sec
1	10.45
2	9.18
3	9.01
4	21.91
5	23.70
6	22.31
7	21.34
8	21.58
9	0.55
10	0.62
11	14.51

Café in center

channel	Packets/sec
1	15.29
2	8.86
3	11.06
4	1.41
5	2.15
6	10.99
7	4.45
8	1.23
9	11.09
10	10.79
11	23.30

University (ETH)

channel	Packets/sec	
	ground	3rd fl.
1	34.06	14.88
2	0.61	5.02
3	4.45	0.19
4	0.10	0.04
5	53.97	192.29
6	0.14	0.04
7	0.28	0.03
8	0.11	0.07
9	26.46	2.88
10	0.06	0.09
11	0.05	0.02

Dormitory

channel	Packets/sec
1	10.28
2	10.03
3	12.13
4	9.92
5	9.30
6	1.77
7	0.09
8	0.19
9	6.92
10	0.47
11	0.36

Suburban flat

channel	Packets/sec
1	0.85
2	0.35
3	0.32
4	0.20
5	0.85
6	3.10
7	2.59
8	11.85
9	4.46
10	2.05
11	9.61

Figure 2.31: Packets per second from the most active AP at various locations and over all 11 WiFi channels

The locations span a University building at two distinct floors, a dormitory, a café in the city center, the main train station and a flat in a suburb of Zurich. Only the University locations share APs. All other locations are well separated over the city. All locations have characteristic properties. While the café has the most equally distributed traffic over all channels, in the dormitory, traffic is clustered in few channels. University locations feature few, heavily trafficked channels while at the suburban flat only few channels are frequented. In all cases, we find at least one channel with 10 or more packets per second from a single AP. While this most frequented channel might differ spatially, a brief scan easily reveals most suited channels.

Since the receiver has no impact on the packet arrival rate, it relies on traffic from other devices. We considered the impact of the RSSI samples per second on the classification accuracy. In the case study (cf. figure 2.30a), we distinguish an empty office with the mobile phone lying on a table, the same room with a person walking next to the table and a person holding and handling the phone. Recordings were taken over four days at different times of day. Each activity is sampled for five minutes in a row. This was repeated on each day twice for all activities.

Table 2.23a shows the classification accuracy (CA), information score (IS), Brier score and area under the ROC[26] curve (AUC) [165, 166]. The IS measures how well a classifier learned a data set. It is higher when the correct class is predicted more often. Brier score

---

[26]Receiver Operating Characteristic

	CA	IS	Brier	AUC
5 samples/s	.593	.594	.512	.813
7 samples/s	.607	.622	.502	.814
10 samples/s	.652	.703	.446	.831
15 samples/s	.671	.806	.408	.856
20 samples/s	.836	1.127	.229	.957

(a) Performance of a k-NN classifier with distinct sample rates

		Classification			
		activity	empty	holding	recall
Gr. truth	activity	**.829**	.014	.157	.829
	empty	.021	**.921**	.057	.921
	holding	.207	.036	**.757**	.757
precision		.784	.949	.779	

(b) Confusion matrix for the k-NN classifier with 20 RSSI samples/sec

Table 2.23: Impact of the sample rate on the classification (978-1-4799-3445-4/14/$31.00 ©2014 IEEE)
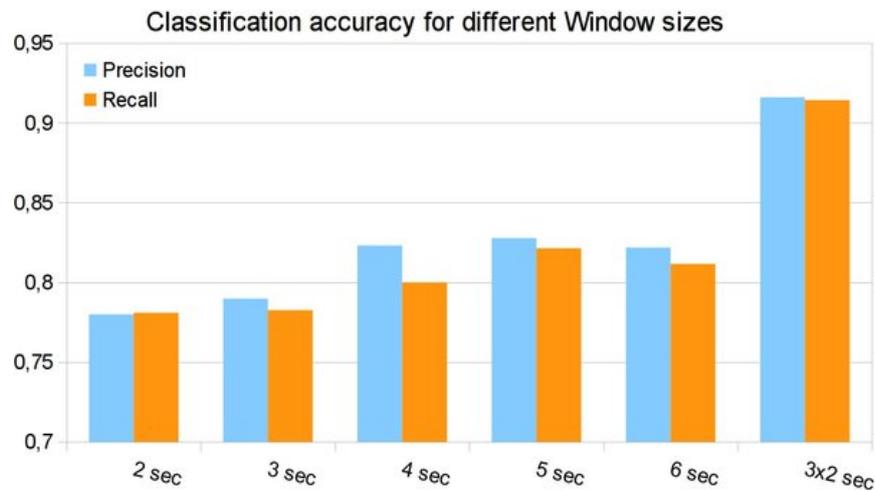
Figure 2.32: Accuracy for the distinction between three basic cases with varying feature window size. A majority vote over three windows of 2 seconds outperforms greater windows (978-1-4799-3445-4/14/$31.00 ©2014 IEEE)

measures the mean squared difference between a predicted probability for an outcome and the actual class. AUC is the probability that a classifier ranks a random positive instance higher than a random negative one.

For these results, we used a k-NN with $k = 20$ (best results reached with $k \in [10..20]$), and a 10-fold cross validation. While higher sample rates improve accuracy, also 10 to 15 samples per second allow an indication about a class. The Confusion matrix for 20 samples per second is depicted in table 2.23b. Observe that activity and holding suffer from slight confusion. In the empty room almost no confusion is seen. Then the signal is stable and not influenced by movement.

The classification accuracy is impacted by the sampling window size (cf. figure 2.32). A majority vote over three successive windows of two seconds can reach higher accuracy than a greater window size. However, since the system is more responsive with shorter windows, we choose 2s windows.

	Distance [meters]									CA	IS	Brier	AUC
	.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	em				
1 AP	x							x	x	.809	1.115	.258	.939
		x						x	x	.730	.796	.434	.866
							x	x	x	.528	.472	.599	.743
	x	x	x					x	x	.483	.933	.644	.831
	x	x	x	x	x	x	x	x	x	.379	1.19	.762	.823
2AP	O	O	O	O	O	O	O	O	O	.427	1.329	.722	.857

(a) Performance using 1 (x) and 2 (O) APs

		Classification			
		.5m	4.0m	empty	recall
Gr. truth	.5m	**.981**	.019		.981
	4.0m	.026	**.768**	.206	.768
	empty	.013	.310	**.677**	.677
	precision	.962	.700	.766	

(b) Classification accuracy with fairly separated locations
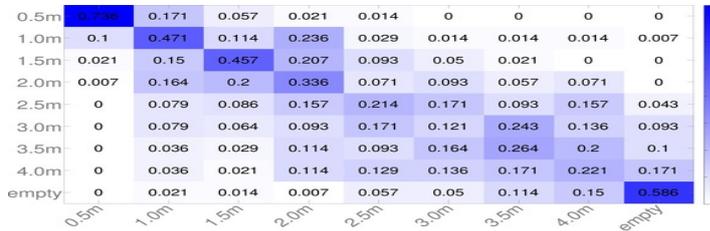
(c) Confusion matrix over all distances

		Classification			
		0.5m	4m	empty	recall
Gr. truth	0.5m	**1.0**			1.0
	4m	.027	**.846**	.128	0.846
	empty	.007	.134	**.859**	0.859
	precision	.968	.863	.871	

(d) Accuracy with two APs

Table 2.24: Classification of activity in various distances. (978-1-4799-3445-4/14/$31.00 ©2014 IEEE)

Distance to the phone

How does the distance to the sensing hardware impact the capability to detect an activity. The case studies depicted in figure 2.30b were conducted at TU-Braunschweig over two consecutive days with repetitions of experiments on both days. On the floor, locations were marked in increasing distance of 0.5m up to 4.0m. At these locations, an individual walked around or move for at least 5min for each distance and day.

We investigated the distinction between an empty environment, a person moving in 4 meters distance and a person moving closer to the mobile phone (cf. table 2.24b). The classification accuracy deteriorates when the locations are closer together (cf. table 2.24a). However, when we tolerate an error of about 0.5-1m, reasonable accuracy can be achieved (cf. table 2.24c). Furthermore, distance to an activity can be estimated from RSSI. In conclusion, there is good potential to classify activities also in this distance so that for indoor environments a mobile phone can cover a typical room sufficiently.

In addition, we employed another equally active AP operating at the same frequency. Although the signal strength between both differed by about 10 dB, classification accuracy was comparable using packets from either AP. In addition, when features are created from RSSI information of both APs, the accuracy can be further improved (cf. table 2.24d). We used the same features for both access points, effectively doubling the number of features for one time window.

Gr. truth	Classification				
	s 1	s 2	s 3	s 4	recall
side 1	**.486**	.193	.121	.2	.486
side 2	.3	**.321**	.086	.293	.321
side 3	.286	.136	**.179**	.4	.179
side 4	.221	.121	.214	**.443**	.443
precision	.376	.417	.298	.332	

(a) Naive Bayes

Gr. truth	Classification				
	s 1	s 2	s 3	s 4	recall
side 1	**.471**	.243	.15	.136	.471
side 2	.314	**.279**	.236	.171	.279
side 3	.3	.314	**.214**	.171	.214
side 4	.293	.257	.171	**.279**	.279
precision	.342	.255	.278	.368	

(b) Classification tree

Gr. truth	Classification				
	s 1	s 2	s 3	s 4	recall
side 1	**.421**	.207	.221	.15	.421
side 2	.271	**.336**	.214	.179	.335
side 3	.271	.186	**.279**	.264	.279
side 4	.221	.143	.264	**.371**	.371
precision	.355	.385	.285	.385	

(c) k-NN classifier

Table 2.25: Confusion matrices for the distinction of the direction in which a person was performing activities (978-1-4799-3445-4/14/$31.00 ©2014 IEEE)

Direction of Movement or Activity

To identify locations of performed activities, in addition to distance also relative direction must be distinguished. We conducted a study in the environment depicted in figure 2.30c in which the mobile was placed in the center of a 2m×2m table. In parallel to the four borders of the table a subject conducted activity (walking up and down) in approximately 1m distance. In figure 2.30c, the regions are marked as 'Area 1–4'. The experiment was repeated multiple times for each side and each time for at least five minutes continuously.

We then attempted to distinguish at which side the activity was performed. However, it turned out that it is hardly possible to tell this from the RSSI. We were not able to find a subset of features that would achieve reasonable accuracy with three distinct classifiers[27] (cf. table 2.25 for exemplary results).

Detection of activity behind a door/wall

WiFi signals can traverse obstacles such as walls or doors but the signal will be damped at this occasion so that the recognition of activity based on this data might be more challenging. We distinguished activity inside or outside a room. As depicted in figure 2.30d, we placed the phone inside a room next to the door. Then, a person was present and moving either inside or on the other side of the door. In the third case, nobody was present in the room or outside on the corridor. For each case, RSSI samples have been recorded for at least five minutes. Table 2.26 depicts the results. While all three cases can be distinguished, the activity conducted outside the room is indeed most confused. This is, because although there is increased fluctuation, signals are weak so that classes are more likely confused for one of the other classes which represent either stronger activity or weakly fluctuating signals.

---

[27]For the results depicted in this table, we utilise a Naive Bayes classifier with 100 sample points and a Loess window of .5, a classification tree with two or more instances at its leaves and a k-NN classifier with $k = 20$

	CA	IS	Brier	AUC
Naive Bayes	.710	.784	.423	.880
Classification tree	.669	.855	.663	.795
k-NN	.724	.843	.393	.903

(a) Performance of various classifiers

		Classification			
		empty	inside	outside	recall
Gr. truth	empty	**.814**	.036	.150	.814
	inside	.064	**.743**	.193	.743
	outside	.2	.186	**.614**	.614
precision		.755	.770	.642	

(b) Confusion matrix

Table 2.26: Classification of activity inside and outside a room (978-1-4799-3445-4/14/\$31.00 ©2014 IEEE)

	CA	IS	Brier	AUC
5 samples/s	.681	.777	.409	.881
10 samples/s	.717	.823	.388	.894
15 samples/s	.767	.910	.355	.905

(a) Performance for different sampling rates

		Classification			
		$.5\frac{m}{\sec}$	$1.0\frac{m}{\sec}$	$2.0\frac{m}{\sec}$	recall
Gr. truth	$.5\frac{m}{\sec}$	**.864**	.071	.064	.864
	$1\frac{m}{\sec}$	.121	**.657**	.221	.657
	$2\frac{m}{\sec}$	.050	.171	**.779**	.779
precision		.834	.730	.732	

(b) Confusion of walking speeds

Table 2.27: Classification of walking speed ($k = 18$; $10\frac{\text{samp.}}{\sec}$) (978-1-4799-3445-4/14/\$31.00 ©2014 IEEE)

Detection of Walking Speed

Walking speed can be derived from signal strength with an SDR-based active device-free system [4]. We investigate the performance of a passive RSSI-based system. In the setting shown in figure 2.30c, a person walked around the table with the mobile phone in its center in a distance of about 2m. The phone sampled the RSSI while the person was moving at $0.5\frac{m}{\sec}$, $1\frac{m}{\sec}$ and $2\frac{m}{\sec}$. This experiment was conducted for at least 5 minutes at each recording and repeated for each velocity twice and also clockwise and counter-clockwise. The speed was controlled autonomously by the subject. For this purpose we marked the circle with an interleaving of 1m and equipped the subject with a stopwatch so that she could adjust her speed. Best accuracy was achieved considering median, mean, minimum and standard deviation. Results are depicted in table 2.27. All velocities can be well distinguished. The confusion is greater for velocity pairs that are closer to each other.

Sensing Crowd

An important ingredient for context-recognition is the size of the surrounding crowd. Different sizes can indicate different situations. For instance, having a conversation between few people or listening to or giving a talk in a meeting with multiple people. We attempted to distinguish between the empty room depicted in figure 2.30f and the same room occupied by 1, 5 or 10 persons. In the room, two phones where placed to record the RSSI. Phone 1

	CA	IS	Brier	AUC
Phone 1	.759	1.309	.354	.946
Phone 2	.805	1.397	.304	.937

(a) Performance of a k-NN classifier with data from various phones

	Classification				
	0P	1P	5P	10P	recall
0 Persons	1.0				1.0
1 Persons		.857	.129	.014	.857
5 Persons		.129	.671	.2	.671
10 Persons		.114	.193	.693	.693
precision	1.0	.779	.676	.764	

Gr. truth

(b) Confusion matrix (Phone 2)

Table 2.28: Classification of crowd (k-NN; 20 samples/s)

is located near the entrance on a table and the second one is placed beside a window across the room. The latter was farther away from the nearest AP which is located right next to the door outside the room. For the case study, 10, 5, 1 or no person would be present for at least five minutes. Participants were instructed not to stand still for longer periods of time but otherwise should move or act freely. They have then, for instance, moved around, stood in front of a poster and discussed it or leaned over a map to plan a weekend trip. Table 2.28 shows that this broad distinction of the number of persons present is possible with reasonable accuracy. Empty room is perfectly recognised with 100% of accuracy. While different crowd sizes are confused in particular for the 5 and 10 persons, the performance is still far above random guess.

Observe that with the data captured by the phone placed near the window (Phone 2) the recognition accuracy is higher. We account this to the fact that individuals in the room continuously have resided in the area between the AP outside the room and the window. Therefore, the impact on the WiFi packets due to blocking or damping was greater for this phone.

Detect Activity while the Device is Carried

When the phone is carried, we expect significant noise for the recognition of situations from packets blocked by the user carrying the phone. We investigated whether RSSI can still be utilised to classify simple situations. For instance, it might be possible to derive whether a person is alone or in company. For this study, the phone was carried in the pocket of a person (cf. figure 2.30e). Then, the person was standing or walking alone and while another person was walking in proximity. For each class, data has been recorded for at least five minutes. Table 2.29 depicts the results. We are well able to detect whether the person wearing the device is stationary and alone or if there is movement either by the device holder or by someone else. However, when the device holder is herself moving, the distinction of other activity is more confused.

	Classification				
	Stat.–Empty	Stat.–pres.	Walking–Empty	Walking–pres.	recall
Stationary–Empty	**.921**	.079			**.921**
Stationary–presence	.15	**.693**	.079	.079	**.693**
Walking–Empty		.171	**.457**	.371	**.457**
Walking–presence		.107	.364	**.529**	**.529**
precision	.860	.660	.508	.540	

Table 2.29: Classification of presence when device is carried <small>(978-1-4799-3445-4/14/$31.00 ©2014 IEEE)</small>
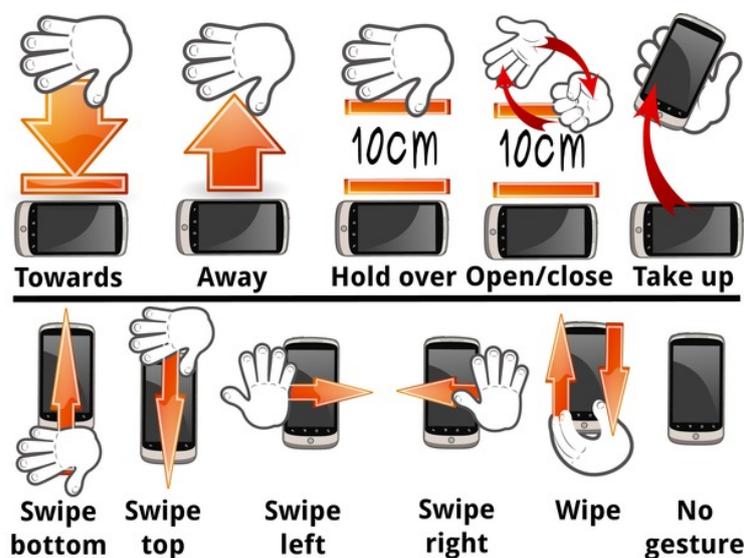
Figure 2.33: Gestures performed <small>(978-1-4799-3445-4/14/$31.00 ©2014 IEEE)</small>

Recognition of gestures

Finally, we considered the recognition of gestures. For this study, we placed the phone on the table as depicted in figure 2.30a and performed 10 single-handed gestures (figure 2.33). Each gesture lasted for approximately 0.4 to 0.7 seconds and was performed in a distance of about 1cm to 20cm. Only for two gestures, *Take up* and *Wipe*, the hand was in actual physical contact with it. Each gesture was repeated 100 times for a total of 1100 recordings of the distinct cases.

Best results have been achieved with the features *mean, variance, signal peaks within 10% of the maximum* and the *fraction between the mean of the first and second half of a feature window*. Table 2.30 shows the classification results. We observe that, while some gestures have a reasonable accuracy and average results are far above guess, a high confusion for other classes inhibits correct classification. In particular, the gesture *Hold over* can hardly be be distinguished. Furthermore, some of the swiping gestures are confused. Therefore, we merged cases into a single gesture. Table 2.31 shows two levels of merging gestures. When

Ground truth		Classification											recall
		Away	Hold over	Towards	No gesture	Open/close	Take up	S. bottom	S. left	Wipe	S. right	S. top	
	Away	**.54**				.06	.03	.15	.12		.05	.05	.540
	Hold over		**.26**	.16	.05	.16	.08	.03	.06	.04	.14	.02	.260
	Towards		.09	**.71**	.07	.04	.01	.01		.06		.01	.710
	No gesture		.04	.06	**.67**	.05	.01		.01	.15		.01	.670
	Open/close		.1	.07	.09	**.47**	.07	.01		.14	.03	.02	.470
	Take up	.01	.08	.03	.02	.09	**.46**	.06	.03	.09	.06	.07	.460
	S. bottom	.13	.06		.01	.04	.09	**.36**	.06		.2	.05	.360
	S. left	.12	.01	.01	.01		.08	.07	**.49**		.07	.14	.490
	Wipe		.04	.1	.08	.16	.09	.01		**.51**		.01	.510
	S. right	.03	.03		.01	.03	.01	.1	.01	.01	**.68**	.09	.680
	S. top	.07	.02	.01		.03	.08		.21		.11	**.47**	.470
	precision	.600	.356	.617	.663	.416	.455	.450	.495	.510	.507	.500	

Table 2.30: Confusion matrices for the distinction of gestures (978-1-4799-3445-4/14/$31.00 ©2014 IEEE)

merging to 7 distinct gestures[28] we achieve a mean accuracy of about 0.56. In the table, labels are shortened to the first two letters for space limitations. While most gestures are well recognized, especially the swipe gestures still achieve mediocre performance. When further reducing to the four gestures *away, towards, no gesture* and *swipe* by merging all swipe gestures, an average accuracy of 0.66 is achieved.

This suggests that some gestures can indeed be utilised to interact with phones or other WiFi capable devices. Possible applications cover a touch-free, frictionless interface to control mobile devices also through clothes, an extended interface for wearable devices or interface-free devices in an Internet of Things.

Discussion

We have investigated a passive, device-free RSSI-based activity recognition system considering several situations captured by a mobile phone. Figure 2.34 summarises shows the accuracies achieved in our case studies relative to a random classifier[29] as a baseline. In general, the overall accuracy falls with increasing number of classes to distinguish. However, short of the recognition of direction, the results are far above random guess in all cases. The simple distinction of distance and three well separated situations reached best results and could be further improved by considering multiple APs or majority votes over several windows of features.

---

[28] *Hold over, Open/close, Take up* and *Wipe* were labelled as *No gesture*

[29] The random Classifier takes each possible choice with equal probability

	Classification							recall
	Aw	No	To	Sb	Sl	Sr	St	
Aw	**.58**	.09		.13	.11	.05	.04	.58
No		**.872**	.05	.014	.012	.034	.018	.872
To		.4	**.59**				.01	.59
Sb	.15	.22		**.32**	.04	.22	.05	.32
Sl	.12	.11	.01	.06	**.48**	.08	.14	.48
Sr	.04	.15		.06	.01	**.67**	.07	.67
St	.03	.18	.01	.01	.24	.1	**.43**	.43
prec	.630	.791	.686	.492	.511	.519	.518	

(a) Confusion of 7 distinct gestures – all remaining gestures shifted to 'no gesture'

	Classification				recall
	Away	Towards	No gesture	Swipe	
Away	**.45**	.06		.49	.45
Towards		**.834**	.052	.114	.834
No gesture		.41	**.56**	.03	.56
Swipe	.063	.128	.005	**.805**	.805
precision	.643	.810	.667	.747	

(b) Confusion of 4 gestures

Table 2.31: Performance with fewer gestures (978-1-4799-3445-4/14/$31.00 ©2014 IEEE)
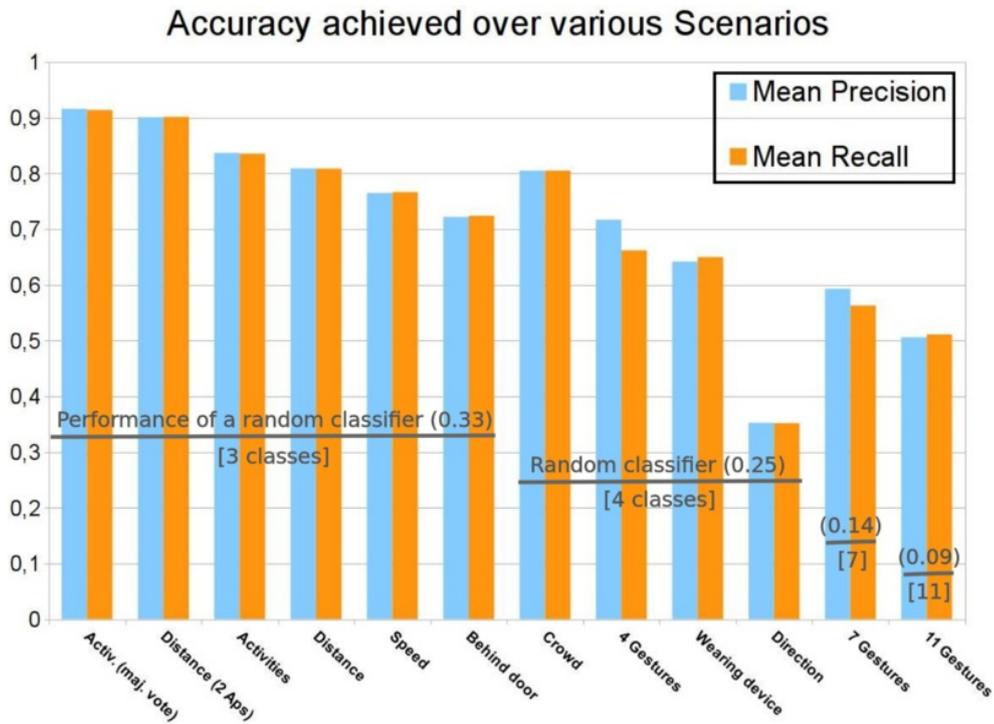
Figure 2.34: Accuracies achieved for various scenarios considered (978-1-4799-3445-4/14/$31.00 ©2014 IEEE)

The failure in the distinction in which direction activity was performed indicates limitations of passive device-free RSSI-based recognition. Since the system has to rely on data transmitted from an AP which can be located in an arbitrary direction and the device might be in arbitrary orientation, it is hardly possible to obtain fine grained information on environmental situation. The sequence of received RSSI samples is highly bursty and of low granularity and rate. Consequently, the classes that can be distinguished are limited too. Additional studies we conducted on the recognition of further activities (sitting, standing, walking, reading, typing on a computer) could not yield a useful recognition accuracy.

However, our results show that an RSSI-based passive device-free recognition system can provide basic environmental awareness when classical phone-based recognition systems fail (e.g. when the phone is not carried on the body). In addition, for special cases such as the distinction of gestures where movement is conducted in close proximity to the device, RSSI-based passive recognition might provide an innovative ad-hoc alternative to more complex solutions.

Unfortunately, our solution requires a modified WiFi firmware, root access and is currently limited to a small set of phones. Much work is still required in order to allow operating-system supported non-root access to RSSI information in sufficient frequency.

2.5.6 Conclusion

We have proposed and discussed the utilisation of RSSI information from mobile phones for the characterisation of situations, activities and gestures. We reported problems to be solved for the acquisition of RSSI from received packets on mobile phones and discussed the structure of the data as well as features suited for the recognition of activities and gestures.

In case studies we investigated the feasibility of RSSI-based recognition on mobile phones for multiple scenarios. Summarising, these results show that it is possible to distinguish simple activities and to some extent also gestures from RSSI fluctuation captured by a mobile phone. However, it also shows the limitations of this device-free recognition approach for instance, regarding a localisation of activities. Furthermore, the accuracies achieved stay below what would be possible with classical sensors such as accelerometers.

However, we could demonstrate, that there is a good potential to extend the perception of a phone beyond its boundaries into the environment. A recognition in distances of 4 meters is still feasible. RSSI-based recognition can cover cases where classical sensors can not provide meaningful results.

Regarding the recognition of gestures, we see a good potential to extend the interface of body-worn devices with RSSI-based gesture recognition.

2.6 Secure communication based on ambient audio [30]

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.

2.6.1 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 [167]. Nevertheless, the range of a communication network most likely bridges devices in various contexts.

As context, proximity and trust are related [167], 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 2.6.2 we introduce related work on context-based security mechanisms and security with noisy input data. Section 2.6.3 discusses the algorithmic background required for ambient audio-based key generation and implementation details. In section 2.6.4 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 2.6.5 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 2.6.5 in a controlled office environment. Section 2.6.5 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 2.6.5 and in a canteen environment in section 2.6.5. The entropy of the ambient audio-based characteristic binary sequences generated by our method is discussed in section 2.6.6. In section 2.6.7 we draw our conclusion.

2.6.2 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 [168, 169, 170, 171]. 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. [170]. The authors propose to utilise the accelerometer of the Smart-It [172] device to extract characteristic features

from simultaneous shaking processes of two devices. Later, Mayrhofer et al. presented an authentication mechanism based on this principle [173]. 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 [174] 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 [175, 169]. 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 [171]. Hershey et al. utilise physical layer features to derive secret keys for a pair of devices [176]. In the absence of interference and non-linear components, transmitter and receiver experience identical channel response [177]. 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 [178], an eavesdropper is not capable of guessing information about the secret. This scheme was validated in an indoor environment in [179]. 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 [151]. 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 [180] is possible. Li et al. analyse the usage of biometric or multimedia data as part of an authentication process and propose a protocol [181]. 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 [182] 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. [183]. The authors establish a key distribution based on a fuzzy vault [184] using data measured by devices worn on the human body. The fuzzy vault scheme, also utilised in [185], enables the decryption of a secret with any key that is substantially similar to the key used for encryption.

2.6.3 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 2.6.3). This binary sequence is designed to fall onto a code-space of an error correcting code (cf. section 2.6.3). 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 2.6.3). 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 2.6.3). 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 2.6.5, 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.

Audio-fingerprinting

Audio-fingerprinting is an approach to derive a characteristic pattern from an audio sequence [186]. 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 [187, 188, 189]. Common applications include the retrieval of a specific music file in an audio database [190], duplicate detection in such a database [191] as well as identification of music based on short samples [192]. The capabilities of detecting similar audio sequences in the presence of heavy signal distortion are prominently demonstrated by applications such as query by humming [193]. The authors utilise autocorrelation, maximum likelihood and Cepstrum analysis to describe the pitch of an audio sequence as a Parsons encoded music contour [194]. Similar audio sequences are detected by approximate string matching [195]. McNab et al. added rhythm information by analysing note duration to match the beginning of a song [196]. A similar approach is presented by Prechelt et al. [197]. 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 [198]. Notes are detected by amplitude-based note segmentation. Later, Shiffrin et al. showed that songs can be described by Markov-chains [199] where states represent note transitions. Retrieval of songs is then achieved by the HMM Forward algorithm [200] 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 [201].

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 [202] 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 [190]. Due to its reported robustness, several authors employ slightly modified versions of this approach [192]. 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 [203]. Alternatively, Burges et al. enhance the former approach by utilising a distortion discriminant analysis [204]. 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. [205].

Also, Yang presented a method to utilise characteristic energy peaks in the signal spectrum in order to extract a unique pattern [206]. A general framework that supports this scheme was later presented by Yang et al. [207]. Building on these ideas, a similar algorithm was then successfully applied commercially by Avery Wang on a huge data base of audio sequences [208, 209].

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 $F_1, \ldots, F_n$ of identical length $d = |F_i| = r \cdot \frac{l}{n}$. On each frame a discrete Fourier transformation (DFT) weighted by a Hanning window (HW) is applied:

$$\forall i \in \{0, \ldots, n-1\},$$
$$S_i = DFT\left(HW(F_i)\right)$$

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

$$b = \frac{\text{maxfreq}(S_i) - \text{minfreq}(S_i)}{m}. \tag{2.58}$$

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.

$$\forall j \in \{0, \ldots, m-1\},$$
$$S_{ij} = \text{bandfilter}_{b \cdot j, b \cdot (j+1)}(S_i) \tag{2.59}$$
$$E_{ij} = \sum_k S_{ij}[k] \tag{2.60}$$

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

$$f(i,j) = \begin{cases} 1, & \begin{aligned} &(E(i,j) - E(i,j+1)) - \\ &(E(i-1,j) - E(i-1,j+1)) > 0 \end{aligned} \\ 0, & \text{otherwise.} \end{cases} \tag{2.61}$$

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 2.6.4 and section 2.6.5). 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. [190]. 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 [210, 211].

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 [212]. Since we experienced a substantial difference in the audio-fingerprints created (cf. section 2.6.4) 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 [213]. 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:

$$\mathcal{A} = \mathbb{F}_q^m, \tag{2.62}$$

$$\mathcal{C} = \mathbb{F}_q^n, \tag{2.63}$$

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:

$$a \xrightarrow{Encode} c, \tag{2.64}$$

This step adds redundancy to the original words with $n > m$, based on polynomials over Galois fields [213].

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

$$\tilde{c} \xrightarrow{Decode} a \in \mathcal{A}. \tag{2.65}$$

The value

$$t = \left\lfloor \frac{n - m}{2} \right\rfloor \tag{2.66}$$

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 [214]. 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.

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

$$\mathrm{Ham}(f, f') \leq t \tag{2.67}$$

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

$$\delta = f \ominus c. \tag{2.68}$$

Then, a SHA-512 hash [215] $\mathtt{h}(a)$ 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 $\mathbb{F}_{2^k}^n$.

$$c' = f' \ominus \delta. \tag{2.69}$$

Afterwards c' is decoded to a' as

$$a' \in \mathcal{A} \xleftarrow{\mathrm{Decode}} c' \in \mathcal{C}. \tag{2.70}$$

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 (2.66)) 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) [216] 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:

Let $\kappa = (\kappa_0, \ldots, \kappa_{31})$, whereas

$$\kappa_i = \left(\sum_{j=0}^{7} c_{(i*8)+j} \right) \quad \mathrm{mod}\ 2^8 - 1$$

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 (2.66)) between fingerprints we can then set m flexibly to define the minimum required fraction u of identical bits in fingerprints as

$$t^* = \lceil (1 - u) \cdot n \rceil , \tag{2.71}$$
$$m = n - 2 \cdot t^*. \tag{2.72}$$

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 2.6.4).

$$m = 512 - 2 \cdot \lceil (1 - 0.7) \cdot 512 \rceil \tag{2.73}$$
$$= 204$$

We therefore use Reed-Solomon codes with

$$RS(2^{10}, 204, 512). \tag{2.74}$$

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

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) [217, 218], the Precision Time Protocol (PTP) [219] 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 [220, 217], some error in the synchronisation of audio samples remains. For instance, the usage of sound subsystems, like GStreamer [221], to record ambient audio introduces new delays.
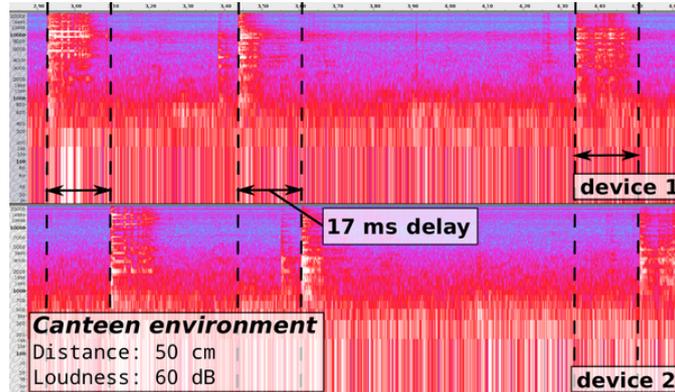
Figure 2.35: Synchronisation offset of NTP synchronised audio recordings

Figure 2.35 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.

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. [222]. In particular, fuzzy commitment is evaluated regarding information leakage by Ignatenko et al. [223]. 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 [224]. For generating `h(a)` a one-way-function has to be chosen to make sure that no assumptions on a can be made based on `h(a)`. We utilise SHA-512 which is certified by the NIST and was extensively evaluated [215].

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 [225] set of statistical tests. Generally, we could not find any bias in the fingerprints created from ambient audio. Section 2.6.6 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.

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}.

Denial-of-service (DoS) An attacker could stress the communication while Alice and Bob are using the fuzzy pairing. The pairing would fail if $(\delta, \text{h}(a))$ 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 $(\delta, \text{h}(a))$ or δ is transmitted. As mentioned in section 2.6.3, with a careful choice of the fingerprint mechanism the exchange of data can be avoided.

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 $(\delta, \text{h}(a))$, she must generate an audio-fingerprint $\bar{f}$ 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 $(\delta, \text{h}(a))$ or δ.

Table 2.32: 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

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 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 2.6.5 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.

2.6.4 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 2.32.

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 2.36 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.

(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

Figure 2.36: Hamming distance observed for fingerprints created for recorded audio samples at distinct loudness levels and distances between microphones and the audio source (1536-1233/13/$31.00 ©2013 IEEE Published by the IEEE CS, CASS, ComSoc, IES, SPS)

Table 2.33: 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

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 2.6.3). 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 2.33 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.

Table 2.34: 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

2.6.5 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 2.6.5, we observe how the proposed method can establish an ad-hoc secure communication 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 2.6.5 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 2.6.5 and section 2.6.5 we describe experiments in common environments where ambient audio was utilised exclusively. In section 2.6.5 we placed devices at distinct locations in a canteen and studied the success probability based on the distance between devices. In section 2.6.5 we study the feasibility of establishing a secure communication channel with road-traffic as background noise. Figure 2.37 summarises all settings considered. To capture audio we utilised the build-in microphones of the computers. The only exception is the reference scenario 2.37a in which simple off-the-shelf external microphones have been utilised. For both devices, the manufacturer and audio device types differed. Table 2.34 details further configuration of the scenarios conducted and the hardware utilised.

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 2.37a. We distinguish four distinct scenarios

2.37a$_1$ Both devices inside the office at locations a and b. 1-2 Individuals speaking at
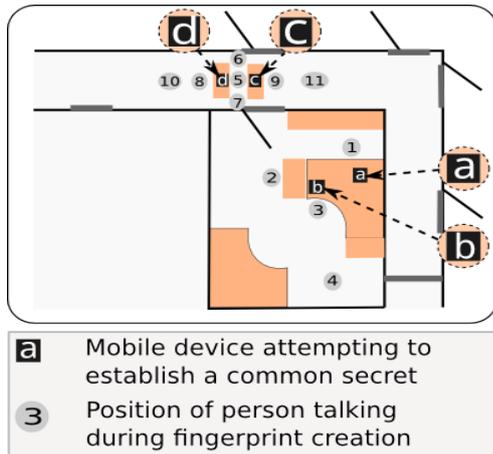
(a) Office setting. Devices and speakers located at distinct positions.
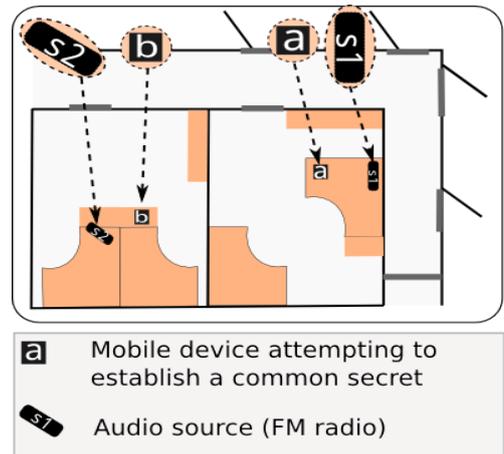
(b) Office setting. Synchronised dominant audio source established in both rooms via FM radios.
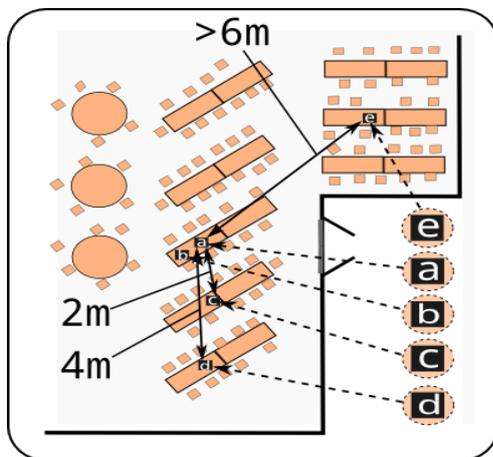
(c) Canteen setting. Devices located at various distances. No dominant noise source.
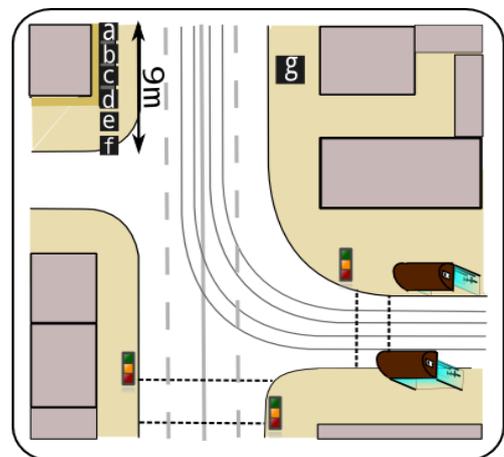
(d) Roadtraffic setting. Devices arranged alongside a road. No dominant noise source.

Figure 2.37: Environmental settings of the case-studies conducted.

locations 1 to 4.

2.37a$_2$ 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.

2.37a$_3$ Both devices in the corridor in front of the office at locations c and d. 1-2 Individuals speaking at locations 5 to 11.

2.37a$_4$ 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 2.6.3. 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 2.37a$_1$ and 2.37a$_3$ represent the situation of two friendly devices willing to establish a secure communication channel. The setting 2.37a$_2$ could constitute the situation in which a person passing by is accidentally witnessing the communication and part of the audio context. In setting 2.37a$_4$, the communication partners might have closed the office door intentionally in order to keep information secure from persons outside the office.

In scenario 2.37a$_1$, where both devices share the same audio context a fraction of 0.9 of all synchronisation attempts have been successful. Also, for scenario 2.37a$_3$, 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 2.37a$_2$, 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 2.37a$_4$, 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.

Finally, we attempted to establish a synchronisation in the scenarios 2.37a$_1$, 2.37a$_2$ and 2.37a$_3$ 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.
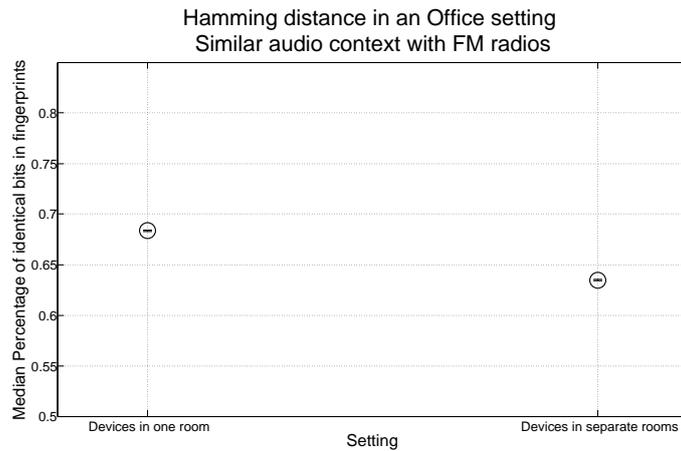
Figure 2.38: Median percentage of bit errors in fingerprints generated by two mobile devices in an office setting. The audio context was dominated by an FM radio tuned to the same channel.

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 2.37b).

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 2.38 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
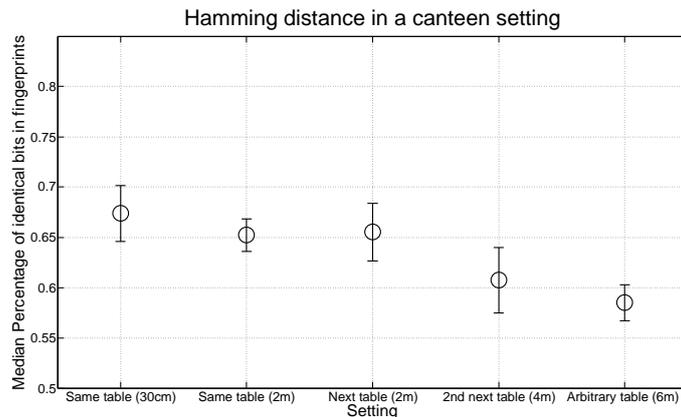
Figure 2.39: Median percentage of bit errors in fingerprints generated by two mobile devices in a canteen environment. (1536-1233/13/$31.00 ©2013 IEEE Published by the IEEE CS, CASS, ComSoc, IES, SPS)

even though information on the audio context might be leaking.

Canteen environment

We studied the accuracy of the approach in the canteen of the TU Braunschweig (cf. figure 2.37c). 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 2.39 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.

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
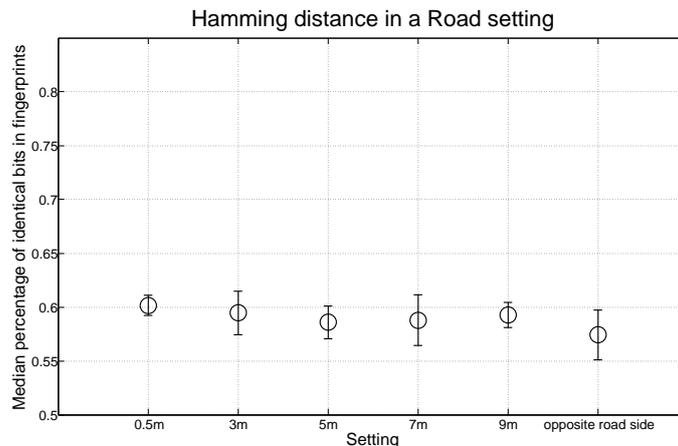
Figure 2.40: Median percentage of bit errors in fingerprints from two mobile devices beside a heavily trafficked road.

running motors in front of the measurements. The loudness level was about 60 dB for both devices. The setting is depicted in figure 2.37d. 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 2.40. 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.

2.6.6 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 [225] 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 [226]. All tests are applied to a set of fingerprints of 480 bits length. We utilised all samples obtained in section 2.6.4 and section 2.6.5.

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[31]. Each specific test was repeated at least 70 times. The p-values are calculated according to the statistical test of Kuiper [226, 227].

Figure 2.41 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 [226] 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

$$1 - \alpha \pm 3 \cdot \sqrt{\frac{(1-\alpha) \cdot \alpha}{m}}. \tag{2.75}$$

Alternatively, we could not observe any distinction between indoor and outdoor settings (cf. figure 2.41a and figure 2.41b) and conclude that also the increasing noise figure and different hardware utilised [32] 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 other samples. We could not identify a significant impact of music on the outcome of the test results (cf. figure 2.41c).

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 2.41d).

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.

---

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

[32]Overall, the microphones utilised (2 internal, 2 external) were produced by three distinct manufacturers
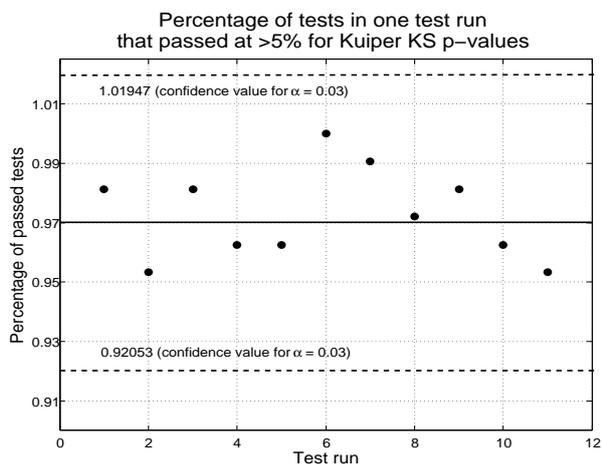
(a) Proportion of sequences from an indoor labora-
tory environment passing a test
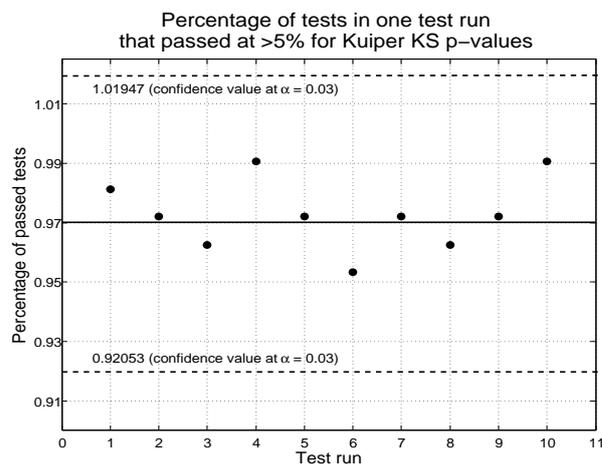
(b) Proportion of sequences from various outdoor
environments passing a test
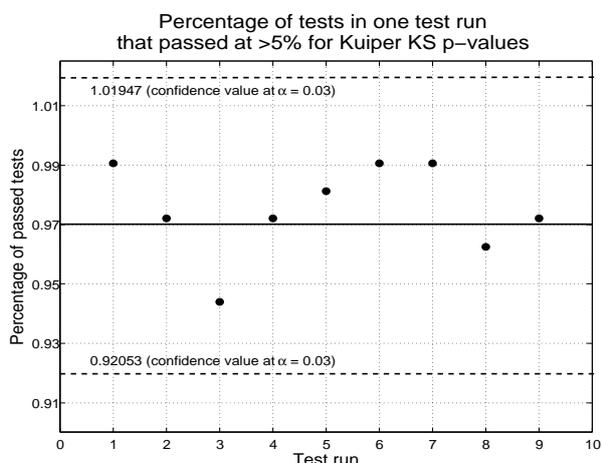
(c) Proportion of sequences from all but music sam-
ples passing a test
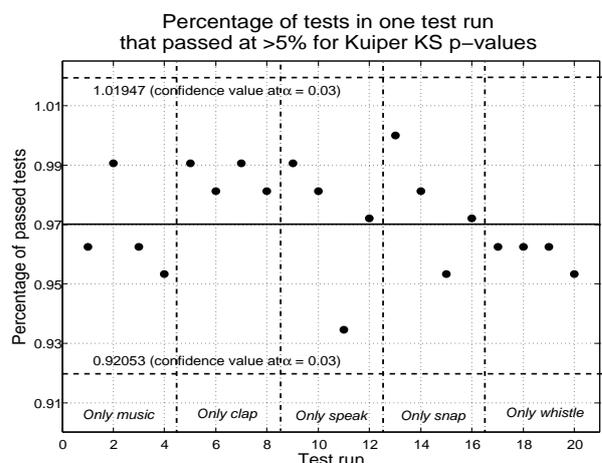
(d) Proportion of sequences belonging to a specific
audio class passing a test

Figure 2.41: Illustration of P-Values obtained for audio-fingerprints by applying the DieHarder battery of statistical tests. (1536-1233/13/$31.00 ©2013 IEEE Published by the IEEE CS, CASS, ComSoc, IES, SPS)

2.6.7 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.

APPENDIX A: Fingerprint creation

The fingerprinting method implemented is detailed in algorithm 2.

After initialisation, the audio sequence is split into frames F_i. For each frame, a Hanning window weighted absolute Fourier transform is then applied. Afterwards, the energy difference between successive frequency bands is calculated and concatenated to a fingerprint. When the energy was increased, the corresponding position in the binary fingerprint is associated with the value 1, and else with 0.

APPENDIX B: Commit

Algorithm 3 details the commit function we utilised to create a public pair $(\delta, \mathtt{h}(a))$ describing the difference between a fingerprint and a randomly chosen code value. Generally, after randomly choosing a from $\mathcal{A}$, a hash $\mathtt{h}(a)$ is generated and a is stored privately. Following a suitable error correcting code (in our case Reed-Solomon codes with $RS(2^{10}, 204, 512)$), a corresponding codeword c is derived from a. After calculating the distance δ between c and f, pair $(\delta, \mathtt{h}(a))$ is then published as a verification that sufficiently similar fingerprints have been created by both devices.

APPENDIX C: Decommit

The decommitment algorithm, utilised to verify the similarity of fingerprints created by remote devices, is detailed in algorithm 4. The public pair $(\delta, \mathtt{h}(a))$ provided by algorithm 3 is used in order to verify the similarity between f and f'. Given the fingerprint f', the codeword c' is calculated as the codeword with distance δ to f'. This value is then decoded to a word $a' \in \mathcal{A}$. Due to the properties of the error correction methods, up to $t = \lfloor \frac{n-m}{2} \rfloor$ bits difference between the fingerprints can be corrected. Since c' and c have the distance δ to the fingerprints f' and f in common, they also share the same Hamming distance. Consequently, from the Hamming distance $\mathrm{Ham}(f, f')$ between the two fingerprints we obtain

$$\mathrm{Ham}(f, f') \leq t \tag{2.76}$$
$$\Rightarrow \quad \mathtt{h}(a') = \mathtt{h}(a). \tag{2.77}$$
$$\Rightarrow \quad a' = a \tag{2.78}$$

Therefore, when the hash values are observed to be identical, a can be used as the common secret among devices. Otherwise, the pairing failed.

Algorithm 2: Fingerprint(S)

Input: Audio sequence S with sample rate r
Data: l: length of S in seconds, n: number of frames, m: number of frequency bands
Output: fingerprint f as bit sequence

begin
    $d \leftarrow r \cdot \frac{l}{n}$; // length of one frame
    $\mathcal{F} \longleftarrow \{F_0, \ldots, F_{n-1}\}$;

    for $i \leftarrow 0$ **to** $n-1$ **do**
        $s \leftarrow i \cdot d$;
        $F_i \leftarrow S[s : s+d]$; // split into frames
    end

    foreach F_i *in* $\mathcal{F}$ **do**
        $F_i \leftarrow$ `HW`(F_i);
        $F_i \leftarrow$ `Abs(FFT`$(F_i))$;
    end
    // calculate energy per frequency band on every frame
    for $i \leftarrow 0$ **to** $n-1$ **do**
        divide into frequency bands $B_0, \ldots, B_{m-1}$;
        for $j \leftarrow 0$ **to** $m-1$ **do**
            $E[i,j] \leftarrow \sum_k B_j[k]$;
        end
    end

    for $i \leftarrow 1$ **to** $n-1$ **do**
        for $j \leftarrow 0$ **to** $m-2$ **do**
            if $E[i,j] - E[i,j+1] - (E[i-1,j] - E[i-1,j+1]) > 0$ **then**
                $f \leftarrow f || 1$;
            end
            else
                $f \leftarrow f || 0$;
            end
        end
    end

    return f
end

Algorithm 3: Commit(f)

Input: fingerprint f
Data: k, m, n for initialising $RS(2^k, m, n)$
Output: $(\delta, \mathtt{h}(a))$

begin

    $\mathcal{A} = \mathbb{F}_{2^k}^m$;
    $\mathcal{C} = \mathbb{F}_{2^k}^n$;

    randomly choose $a = (a_0, \ldots, a_{m-1}) \in \mathcal{A}$;
    generate $\mathtt{h}(a)$;
    $\mathtt{SavePrivate}(a)$;

    $c \in \mathcal{C} \xleftarrow{\text{Encode}} a \in \mathcal{A}$;
    $\delta \leftarrow f \ominus c$; // $\ominus$: subtraction in $\mathcal{C} = \mathbb{F}_{2^k}^n$
    return $(\delta, \mathtt{h}(a))$

end

Algorithm 4: Decommit($f', (\delta, \mathtt{h}(a))$)

Input: fingerprint f', received $(\delta, \mathtt{h}(a))$
Data: k, m, n for initialising $RS(2^k, m, n)$

begin

    $\mathcal{A} = \mathbb{F}_{2^k}^m$;
    $\mathcal{C} = \mathbb{F}_{2^k}^n$;

    $c' \leftarrow f' \ominus \delta$; // $\ominus$: subtraction in $\mathcal{C} = \mathbb{F}_{2^k}^n$
    $a' \in \mathcal{A} \xleftarrow{\text{Decode}} c' \in \mathcal{C}$;
    generate $\mathtt{h}(a')$;

    if $\mathtt{h}(a') == \mathtt{h}(a)$ **then**
        return *"decommitment successful"*;
        $\mathtt{SavePrivate}(a')$;
    end
    else
        return *"decommitment failed"*;
    end

end

2.7 Pattern-based Alignment of Audio Data for Ad-hoc Secure Device Pairing [33]

When studying the use of ambient audio to generate a secure cryptographic shared key among mobile phones, we encounter a misalignment problem for recorded audio data. The diversity in software and hardware causes mobile phones to produce badly-aligned audio chunks. It decreases the identical fraction in audio samples recorded in nearby mobile phones and consequently the common information available to create a secure key. Unless the mobile devices are real-time capable, this problem can not be solved with standard distributed time synchronisation approaches. We propose a pattern-based approximative matching process to achieve synchronisation without communication independently on each device. Our experimental results show that this method can help to improve the similarity of the audio fingerprints, which are the source to create the communication key.

2.7.1 Introduction

With recent advances in smart-phone dissemination and their computational capabilities, smart-phones can be seen as a kind of wearable device for the masses. These general-purpose devices are capable of solving several wearable computing tasks. Due to their high penetration, security in communication among devices becomes a relevant issue. Common security schemes for mobile devices require explicit user input to provide a shared piece of information. A wearable device, however, should not distract its holder from other tasks. How can we provide security among possibly unacquainted devices without any user interaction?

We consider an interaction-free common key generation scheme for proximate devices conditioned on ambient audio. Since the seed to the key is implicit with the context, no information that could be used to reconstruct the key is transmitted on a wireless channel during key generation. Each device computes a binary characteristic sequence for a synchronised recording: An audio-fingerprint [190, 209, 228, 229, 230]. This binary sequence is designed to fall onto a code-space of an error correcting code [213]. Devices then exploit the error correction capabilities of the error correcting code to map fingerprints to codewords as described in [214, 212]. For fingerprints with a Hamming-distance within the error correction threshold of the error correcting code the resulting codewords are identical and then utilized as secure keys. The Hamming distance in fingerprints rises with increasing distance of devices so that distant devices are unlikely to guess the correct key. Our fingerprint extraction scheme is adapted from [190] to extract fingerprints from synchronized ambient audio recordings in a noisy environment without exchanging any information about the resources among devices. However, when audio sequences utilized

---

are not well aligned, similarity in fingerprints decreases. This is due to the fingerprint generation which exploits the relative fluctuation of energy over time. Small shifts in the audio data will likely produce completely differing fingerprints. This is a relevant problem since simple time synchronisation approaches, such as for instance the network time protocol (NTP), are not suitable to sufficiently synchronise audio recordings due to the delays in the recording hardware.

This paper addresses the accurate alignment of recorded audio sequences from remote devices. The challenging point here is to achieve an alignment between audio samples taken from distinct devices interaction free and without any inter-device communication other than an initial plain pairing request.

We will in section 2.7.2 discuss related work on secure ad-hoc pairing of mobile devices. Problems that prevent accurate audio sequence alignment and our pattern-based approximative matching method to reduce the mismatching are detailed in in section 2.7.4. Section 2.7.5 describes a case study conducted with smart-phone devices to investigate the accuracy of our approach in a realistic setting and results of our alignment scheme. Section 2.7.7 draws our conclusion.

2.7.2 Related Work

Contextual or sensor information of mobile devices can be incorporated as a solution for authentication [170]. When the seed to the key is implicit with the context, no information that could be used to reconstruct the key is transmitted during key generation. For instance, McCune et al. [231] introduced *Seeing-Is-Believing*, utilizing the camera of a mobile device to capture a 2D barcode which is displayed on the screen of another device. *Loud and Clear* of Goodrich et al. [232] implements a similar scheme but exploits spoken audio. A user reads aloud a text message displayed on one device and a second device recognizes the speech for authentication. A further example mechanism by Mayrhofer et al. [173] uses accelerometer readings when devices are shaken simultaneously by a single person. Also, Mayrhofer derived in [174] that the sharing of secret keys is possible with a similar protocol by repeatedly exchanging hashes of key-sub-sequences until a common secret is found. Bichler et al. generalise this approach to noisy acceleration readings [175, 169]. They utilize a hash function that maps similar acceleration patterns to identical key sequences. These approaches require explicit user interaction.

By utilising a context source that provides a sufficient amount of unique, context-related information, such as audio or radio frequency (RF), it is possible to get the user out of the loop. Mathur et al. introduced ProxiMate that enables wireless devices in proximity to pair automatically and securely using their shared ambient RF-signals [6]. They generate fingerprints from RF-channel fluctuations and map these onto a codespace of an error-correcting code. By correcting potential errors in the fingerprints, they are mapped onto the closest regular codeword in the codespace. When the similarity between fingerprints is high, codewords are identical. Sigg et. al proposed to use audio instead of RF in a similar implementation [233]. They study the entropy of audio fingerprints and identify time synchronisation as a main hindrance to practically apply the method for mobile

devices. Their instrumentation requires idealized conditions regarding the synchronisation of devices and to account for this a high number of fingerprints must be created (201 in their experiments) in order to find one matching fingerprint. For extensive computational load, this is feasible only in an offline approach. The high number of fingerprints created, however, was necessary since the utilized NTP synchronisation is not sufficiently accurate.

In this paper, we present an alignment mechanism that enables a synchronisation accuracy of recorded audio in the order of less than 10 milliseconds among unsynchronised mobile devices in the same context without transmitting information about the audio sequence over the wireless channel. The synchronisation is achieved by processing a weakly NTP-synchronised recording without additional communication among the devices.

2.7.3 Extracting fingerprints from ambient audio

Most of the previous studies applied audio fingerprints to a large audio database of musical information. In this type of data, there exists a dominant sound, which is the song itself. In our research, we need to extract a fingerprint that can represent all of the characteristics of ambient sounds around the devices. For instance, when the user is in a lobby room, there are various audio sources such as human voice, music, noise of opening and closing doors, etc. All of the recorded sounds are potentially equally important to describe a context and should contribute to form an audio fingerprint.

Our approach is adapted from [190] to extract the most similar fingerprint from synchronized ambient audio recordings in a noisy environment without exchanging any information about the resources among devices. In our research, each bit in the binary audio-fingerprint expresses the difference between energy on frequency bands.

An ambient audio chunk is split into non-overlapping equal-length frames. Then we perform a Discrete Fast Fourier Transform on these frames. After that, each frame is divided into a set of frequency bands with identical width.

An energy matrix E is created as follows. Its size is *the number of frames $\times$ the number of frequency bands per frame*. Each value of the matrix represents the total energy of a frequency band in the corresponding frame.

From the matrix E, we generate the binary fingerprint f whose bits contain the information about the energy change of frequency bands on two successive frames.

$$f(i,j) = \begin{cases} 1 & \text{, if } \begin{array}{l} (E[i,j] - E[i,j+1]) \\ -(E[i-1,j] - E[i-1,j+1]) > 0 \end{array} \\ 0 & \text{, otherwise.} \end{cases}$$

In the above formula, if the energy gain between two consecutive frames is positive, the corresponding element in the binary sequence is assigned with the value of 1; otherwise, it has the value of 0. Each audio fingerprint contains 512 bits. Because the fingerprint extraction is performed based on the relative fluctuation of energy, there is no effect caused by the difference of raw intensity values in time domain for distinct devices. However, since the audio signal is of consecutive time windows, small shifts in the audio data will likely

produce completely differing fingerprints. When this method is utilized for the generation of identical binary strings on remote mobile devices, the timing difference likely prevents the generation of sufficiently similar information on both devices.

Also, if there are more audio sources around the device, the audio fingerprints are more sensitive with inter-device distance. The closer the mobile devices to each other, the more similar their fingerprints are.

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 [212]. Since we experienced a substantial difference in the audio-fingerprints created 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 utilize possibly different audio hardware.

The proposed cryptographic protocol shall work completely ad-hoc with devices previously not known to each other. 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. Apart from these requirements we want to choose the threshold for a working encrypted communication based on contextual conditions of different physical locations.

With fuzzy encryption schemes, we are able to overcome these challenges. Generally, 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 are similar for all communicating devices in the same context are constituted by fingerprints generated from ambient audio.

We implemented a Fuzzy Commitment scheme with Reed-Solomon codes [213]. The following discussion provides a short introduction into 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 initialized as:

$$\mathcal{A} = \mathcal{F}_q^m, \tag{2.79}$$

$$\mathcal{C} = \mathcal{F}_q^n, \tag{2.80}$$

with $q = p^k, p$ prime, $k \in \mathcal{N}$. 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:

$$a \xrightarrow{Encode} c, \tag{2.81}$$

This step adds redundancy to the original words with $n > m$, based on polynomials over Galois fields [213].

Decoding utilizes 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

$$\tilde{c} \xrightarrow{Decode} a \in \mathcal{A}. \tag{2.82}$$
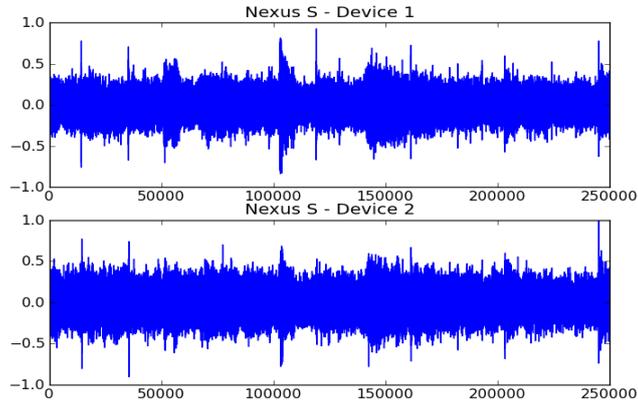
Figure 2.42: Audio files recorded on two Samsung Nexus S mobile phones. (1550-4816/12 $26.00 ©2012 IEEE)

The value

$$t = \left\lfloor \frac{n-m}{2} \right\rfloor \qquad (2.83)$$

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 [214]. In the algorithm implemented, 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.

Our current fingerprint extraction scheme, which is based on energy in frequency domain, can deal with the difference in amplitude values. However, when the audio chunks are shifted in time, for example, since the ambient sounds are not recorded at the same time, the audio fingerprints misalign so that we can not create the common key. In our experiments, we observed this effect. In particular, although the NTP-synchronised devices intend to start their recordings at the same time, we observed significant differences that exceed the time offset due to NTP in accuracy by several milliseconds. This effect can be observed in figure 2.43.

2.7.4 Pattern-based alignment of audio data

When developing the scheme of unobtrusive secure device pairing with audio fingerprints for Android-based mobile devices, we encountered practical issues not evident when considering the problem theoretically. One issue in practical implementation is differing audio hardware. For instance, the Samsung Google Nexus S[34] and HTC Google Nexus One[35]

---

[34]Nexus S: `http://www.google.com/nexus/tech-specs.html`
[35]Nexus One: `http://www.google.com/phone/detail/nexus-one`
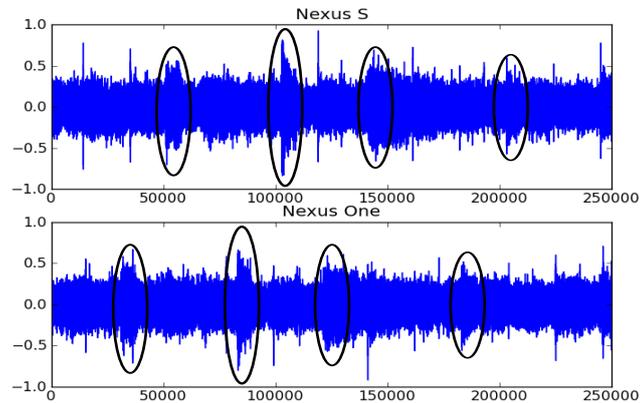
Figure 2.43: Misalignment of audio files from difference mobile phones. The above is a Samsung Nexus S and the below is a HTC Nexus One.

devices we utilized apply different audio-pre-processing routines that render the unprocessed audio outputs on these devices unusable for the generation of identical fingerprints. Furthermore, time synchronisation is a serious problem for the approach. In particular, not only the clocks on remote devices have to be synchronised as usual for distributed devices, but also the generally unknown and possibly non-constant hardware specific delays on both devices need to be taken into account.

Misalignment of audio data

Due to the differences in software and hardware of the devices, the recorded audio sequences are not exactly the same. An example of this phenomenon is shown in figure 2.44. All audio files are recorded with the same settings and at the same time (clocks synchronised by a NTP service[36]). Basing on some similar visual appearances in the waveform format of the signals in figure 2.44, the recording start time of the Nexus One was heavily delayed when compared to the Nexus S. In our recorded audio files, time difference values can be from 0.3 second to more than 1 second. The time difference in some pairs of audio files is shown in Table 2.35.

We observe that the offset when both recordings start is fluctuating and in all cases much higher than the accuracy expected from NTP synchronisation. Additionally, we observed from the figure that clearly, the higher frequency bands of the signal available at the Nexus One device completely differ from that of Nexus S readings because the Nexus One employs a hardware noise cancellation. There is no way to bridge the noise cancellation on that device to obtain the unmodified signal. Both these effects are unfortunate for our fingerprinting method.

---

[36]Navy Clock II application: `https://market.android.com/details?id=com.cognition.navyclock`

Table 2.35: Time Difference before Pattern-based Alignment.

Audio file pair	Time difference [seconds]
1	0.320
2	0.849
3	0.740
4	0.459
5	1.341
6	0.450
7	0.765
8	0.562
9	0.744
10	0.375
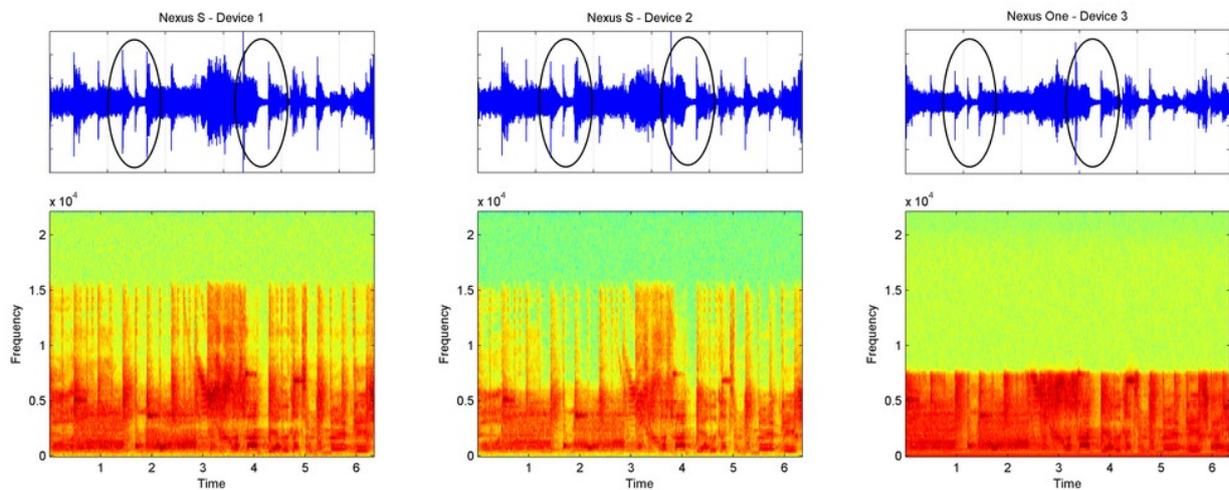
Figure 2.44: Waveform format (upper) and spectrogram (lower) representation of audio recordings from three devices. For the Nexus One, the hardware noise cancellation cuts higher parts of the frequency spectrum. The recording of the Nexus One is delayed comparing to the two Nexus S devices which are tightly synchronised.
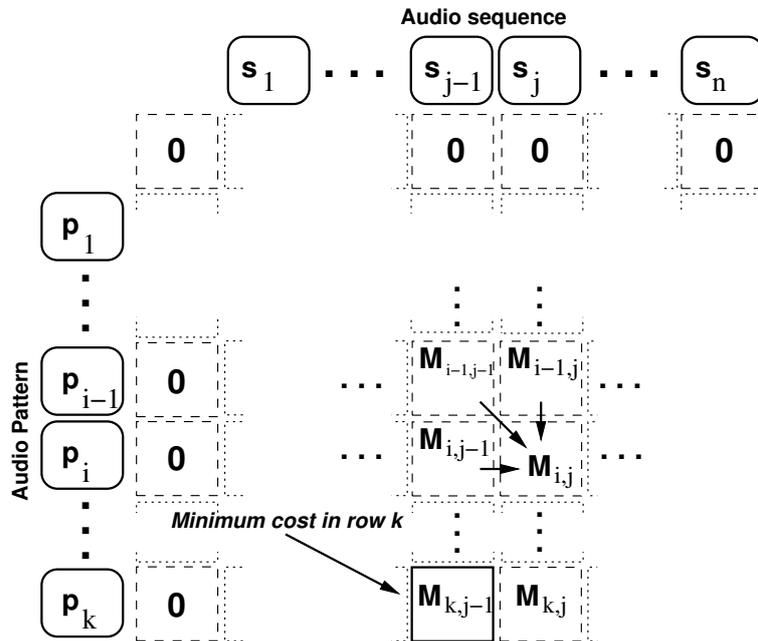
Figure 2.45: Schematic illustration of the Smith-Waterman pattern matching (1550-4816/12 $26.00 ©2012 IEEE)

Aligning recorded audio data

One important condition for our secure device pairing scheme is that no information regarding the recorded audio shall be exchanged between the devices. Otherwise, the security of the key might be impaired by information leaking from these transmissions. We propose a synchronization scheme to reduce the mismatch between audio data from neighbouring devices. Our solution is based on the Smith-Waterman algorithm [234], an approximative pattern matching technique. In the algorithm, two strings are compared for the best approximative pattern among them, as depicted in figure 2.45. The figure illustrates the operation to find a sub-sequence similar to a pattern $p = p_1 \ldots p_k$ in a longer sequence $s = s_1 \ldots s_n$. The matrix is filled from left to right and from top to bottom. Each matrix entry M_{ij} contains the minimum cost to align $p_1 \ldots p_i$ with $s_1 \ldots s_j$. It is obtained iteratively from the cost of the cells $M_{i-1,j-1}, M_{i-1,j}, M_{i,j-1}$ by choosing the minimum cost for an extension of any of these three alignments (extension of both strings, extension of only the first string or extension of only the second string). The first row is initialised with 0 to allow the pattern to start at any position within s. The careful choice of the gap-penalty g is crucial for the approach to produce good matching results. The best matching string is obtained by backtracking the smallest entry in the last row.

Since the matching is approximative, we will always find a best matching position even though the absolute similarity of this very position might not be high. We exploit this property to be able to resign from any information transmitted among devices on the actual recorded audio. In particular, we utilize a predefined, characteristic pattern on both

devices. since the pattern is known in advance, no information need to be transmitted. The downside to this implementation is, of course, that the pattern utilised might be very different from the actual audio recorded. However, since the Smith-Waterman algorithm always computes a best matching, we can speculate that this best matching is found at similar positions in the audio recordings, provided that the data has significant similarity. The best matching, although it might not be a good matching in absolute terms, is likely to be found at a similar position in the recorded audio of the remote devices. The specific pattern used for matching is extracted randomly from consecutive samples of an arbitrary audio sequence. In our experiments, its length is 100 samples (longer patterns increase running time of the algorithm). The pattern p is matched in the first 100000-sample part of each audio file. The matching score of p and a local part l of s is the difference between amplitude values in p and l. The less the score, the more similar p and l are. According to our experiments, the gap penalty that can yield an acceptable matching is 150. Then, we eliminate all samples preceding the matching positions and generate the audio fingerprints from the remaining ones.

It turned out that we achieve only seldom a perfect synchronizing result among two best matching points on both devices. We therefore calculated a set of k best matching parts. We assume that one device chooses the best matching locations, encodes a data sequence with the resulting keys and transmits this to the second device. This receiver device then attempts to decode the data with the keys generated from its best matching alignment points. This process might require the transmit device to transmit a data sequence several times, encrypted with different keys each time. Although this increases the transmit load, this process could be implemented in an iterative fashion and is required only for the first encounter of the devices to derive the secret key. Since all data blocks are encrypted, only marginal additional information is provided to a potential adversary.

2.7.5 Experimental results

We utilize the Android-based mobile phones Samsung Google Nexus S and HTC Google Nexus One. The Nexus One smart-phone has a secondary microphone dedicated for dynamic noise suppression, while the Nexus S devices utilise software noise cancellation. The ambient audio data are recorded for 6375 milliseconds at a sampling rate of 44100 Hz. In our experiments, we attempt to generate a secure key among a Nexus S and a Nexus One device after applying a bandpass filter to the recorded data and aligning audio sequences with the pattern matching approach. The smart-phones are put at the same distance d from an audio source and we increase the distance between two devices from $10cm$ to $100cm$. With each distance, we record 10 audio files on each phone. Before matching with the arbitrary pattern, we perform a bandpass filtering which only retains signals whose frequency is between $4000Hz$ and $4500Hz$.
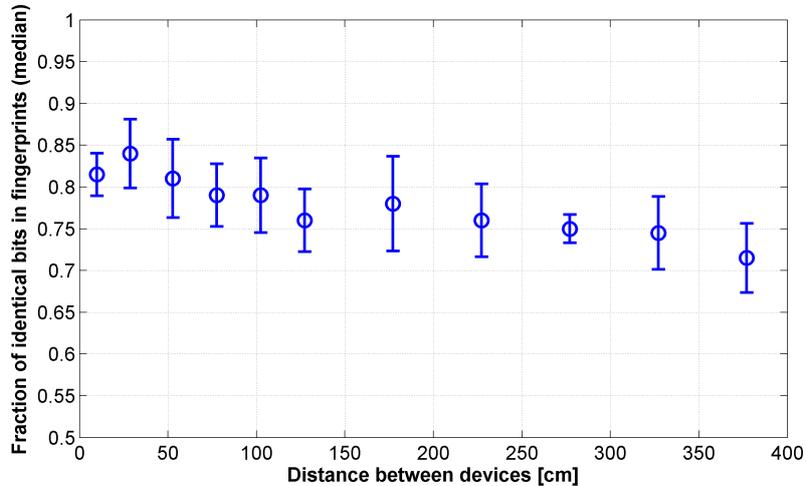
Figure 2.46: Median fraction of identical bits in fingerprints created from synchronised audio recordings at two Nexus S devices (1550-4816/12 $26.00 ©2012 IEEE)

2.7.6 Key generation without sequence alignment

We conducted a case study with two Nexus S and one Nexus One device to study the alignment of audio data at various inter-device distances. The bit-errors in fingerprints were not related to the distance to an audio source or the loudness of the sound. Therefore, we consider the deterioration in the similarity of fingerprints while the distance among recording devices is gradually increased. We are interested in the Hamming distance between the generated fingerprints.

As detailed in section 2.7.3, we can use error correcting codes (ECC) to account for the Hamming distance in fingerprints and generate an identical cryptographic key at devices independently without communication, provided that the Hamming distance is significantly smaller than 50% of the sequence length (which would be the Hamming distance to a random binary sequence). We implemented the fingerprinting method based on energy differences as described above.

Devices were placed in an angle of 22.5° and −22.5° to an audio source and the distance between devices was altered gradually. In several sets of experiments, we increased the distance among devices and to the audio source while keeping the angle to the audio source fixed. For each distance the experiment was repeated at least 10 times.

The figure shows the median fraction of identical bits in fingerprints created by the devices. Each point was created from 10-12 separate experiments with identical environmental conditions. The method achieves an accuracy of about 0.75 to 0.85 in all tests. In particular, the Hamming distance in fingerprints only slightly decreases with increasing distance. This demonstrates that a good synchronisation is feasible on mobile devices. Also, devices in about 2 m distance can be considered to be in the same security aura for this environment. Devices farther away can be excluded by properly configuring the error correction threshold of the ECC utilized.

In particular, when the ECC is configured to correct up to about 17% of the bits in a fingerprint, devices in more than 60 cm distance would only occasionally be able to successfully generate the same key. For devices in more than 2 meters distance, it is highly unlikely that they generate a matching key. Observe that, for a 512 bit key, although a Hamming distance of 1% of all bits translates to only about 32 different possibilities, the positions of the respective 5 bits in the 512 bits are not known. This leaves a total of

$$\left(\begin{array}{c} 5 \\ 512 \end{array} \right) \cdot 32 > 9 \cdot 10^{12} \tag{2.84}$$

different possibilities for an attacker to reduce the Hamming distance further by only 1%.

For the Nexus One, only about 50% of the bits were identical in all cases, regardless of the distance among devices and the specific Nexus S device used for pairing. Since this is the similarity, we would expect for a random guess of the fingerprint, we conclude that a secure key can not be derived with this approach from these devices which differ in hardware and software.

Figure 2.47 depicts the median time difference after alignment when increasing the number of trials in the transmitter. The receiver attempts to decode the data with its sequences from the top 3 and 10 matching positions.

We can achieve a synchronisation in the order of 10 milliseconds already with 10 trials at the transmitter and only 3 trials at the receiver (cf. figure 2.47).

Moreover, when the receiver tries 10 times, we observe that, already with the best matching trial at the transmitter, the matching is greatly improved compared to time difference between misaligned audio sequences in table 2.35. With 3 trials of the transmitter, the synchronisation time is accurate enough for our secret key generation approach. Since the alignment approach calculates all possible alignments at once, the cost for additional trials at the receiver is low. Additional trials at the transmitter, however, directly impact the communication load.

We conclude, that the alignment approach greatly improves the synchronisation of audio recordings on remote devices. This can be seen from figure 2.51. The figure depicts the fraction of identical bits among fingerprints before and after the alignment matching was applied. While the fingerprint similarity only marginally deviates from the similarity to a random sequence, it is greatly improved after the alignment matching was applied. The fraction of identical bits in the audio fingerprints decreases when inter-device distance increases. In particular, also the variance in the data is reduced. These characteristics allow for a sharper threshold of the error correcting code.

We also extracted the pattern from one file in each misaligned audio file pair and found it in the other. The quality of the audio fingerprints is not much higher than with the arbitrary patterns, even with the best results of matching. The comparison is shown in figure 2.51. Moreover, in this approach, there is data transmission between the devices.
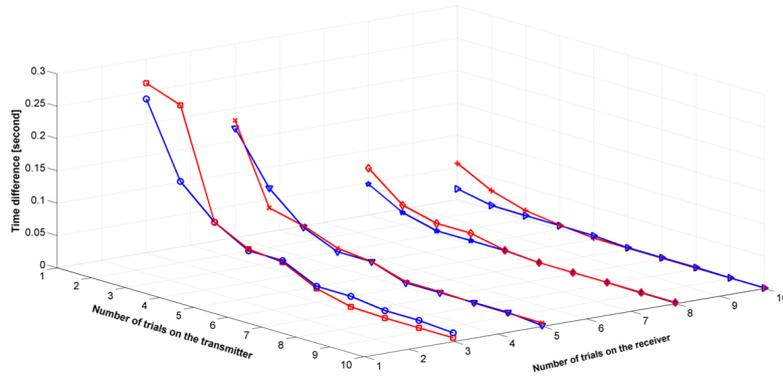
Figure 2.47: Time difference of the audio after alignment with increasing number of trials on the transmitter. The number of trials on the receiver is respectively 3 and 10. (1550-4816/12 $26.00 ©2012 IEEE)
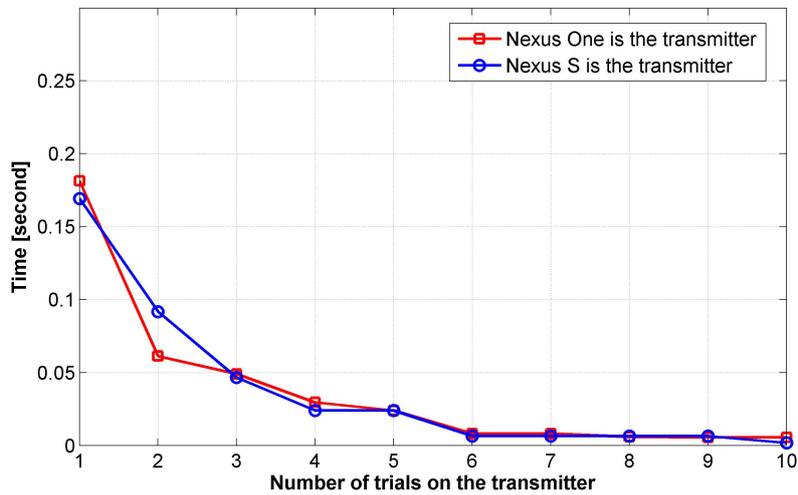
Figure 2.48: Time offset of the audio after synchronisation with increasing number of trials on the transmitter. The number of trials on the receiver is always 5. (1550-4816/12 $26.00 ©2012 IEEE)
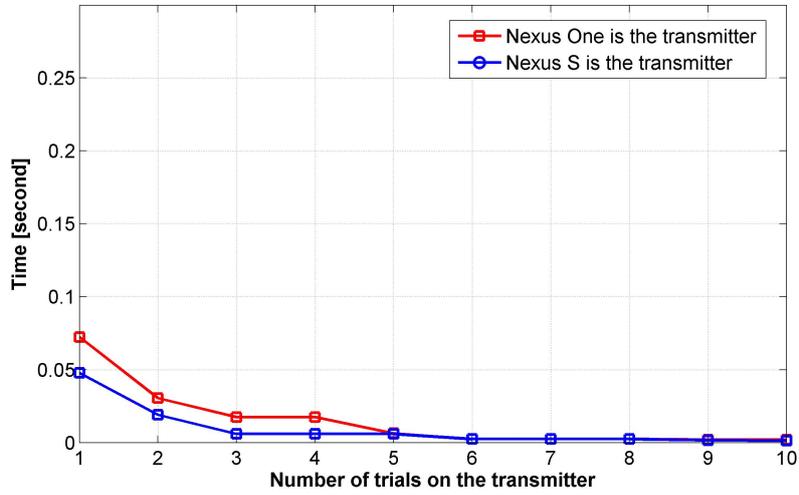
Figure 2.49: Time offset of the audio after synchronisation with increasing number of trials on the transmitter. The number of trials on the receiver is always 8.
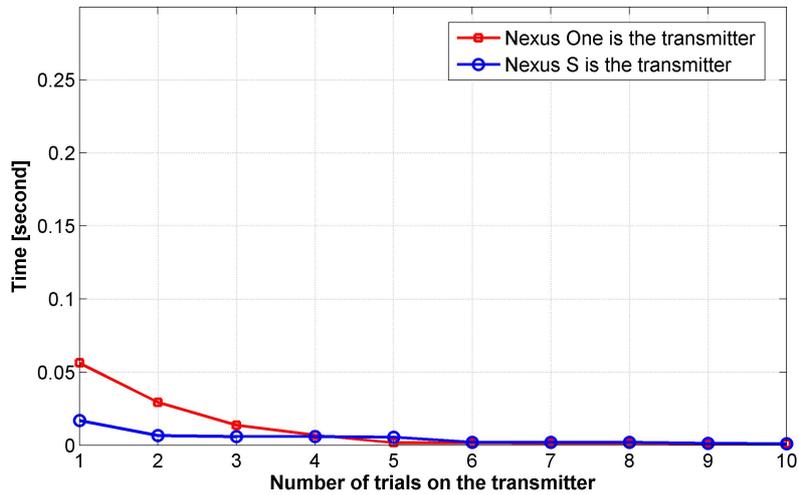
Figure 2.50: Time offset of the audio after synchronisation with increasing number of trials on the transmitter. The number of trials on the receiver is always 10.
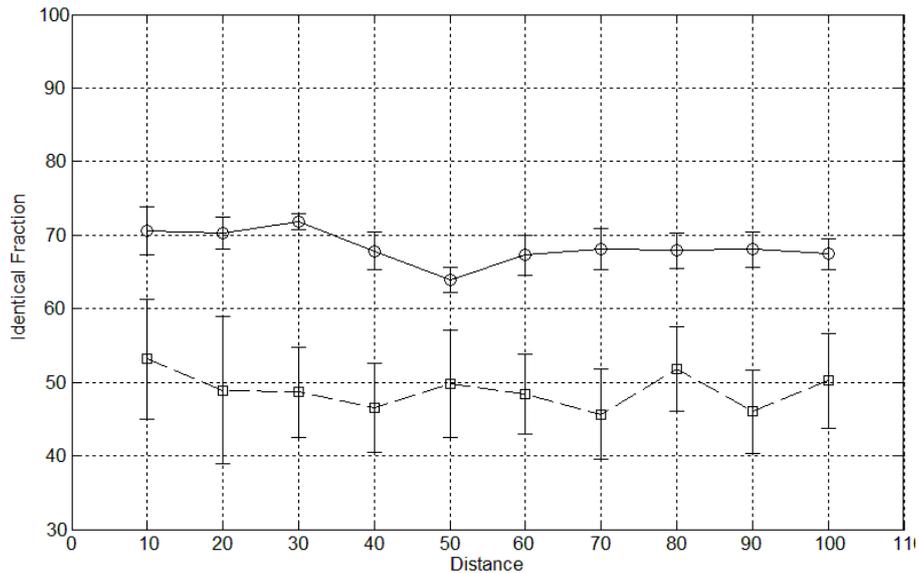
Figure 2.51: Identical fraction in bits of audio fingerprints created on Nexus S and Nexus One devices when the inter-device distance increases. (1550-4816/12 $26.00 ©2012 IEEE)

2.7.7 Conclusion

When implementing an audio-based ad-hoc secure device pairing mechanism for previously unacquainted mobile devices, the diversity of hardware and software can affect the offset in audio recordings of even clock-synchronised mobile devices. We propose an approximative pattern matching algorithm to align the corresponding audio without communication between the devices. The devices synchronise their audio sequences without any knowledge about the recorded audio on the remote device other than their own recorded contextual information. Hence, no information about the audio utilized as a seed for the secure key generation, can leak. To improve the alignment quality, we can choose more than one matching position on each device at the cost of increasing the communication load for one device. We can obtained a synchronization among devices of less than 2 milliseconds when both devices utilize up to 10 trials. With moderate additional communication load, a synchronisation in the order of 10 milliseconds is reasonable.

3 Discussion

3.1 Feedback based closed-loop carrier synchronisation: A sharp asymptotic bound, an asymptotically optimal approach, simulations and experiments

Solutions proposed for carrier synchronisation comprise open-loop synchronisation methods such as round-trip synchronisation [47, 48, 49]. In this scheme, the destination transmits beacons in opposed directions along a multi-hop circle in which each of the nodes appends its part of the overall message to the beacons. Beamforming is achieved when the processing time along the multi-hop chain is identical in both directions. This approach, however, does not scale with the size of a network.

Closed loop feedback based approaches include full-feedback techniques, in which carrier synchronisation is achieved in a master-slave manner. The phase-offset among the carrier signals of destination nodes is corrected by a receiver node. Diversity between RF-transmit signal components is achieved over CDMA channels [50]. This approach is applicable only to small network sizes and requires sophisticated processing capabilities at the source nodes.

A more simple and less resource demanding implementation is the one-bit feedback based closed-loop synchronisation considered in [50, 51]. The authors describe an iterative process in which n source nodes $i \in [1, \ldots, n]$ randomly adapt the phases γ_i of their carrier signal $\Re\left(m(t)e^{j(2\pi(f_c+f_i)t+\gamma_i)}\right)$. Here, $m(t)$ is the transmit message and f_i denotes the frequency offset of node i to a common carrier frequency f_c. Initially, i.i.d. phase offsets γ_i of carrier signals are assumed. When a receiver requests a transmission from the network, carrier phases are synchronised in an iterative process.

Our study of binary-feedback-based distributed adaptive transmit beamforming derived for the first time a concise theoretical understanding of this carrier synchronisation approach. The potential to achieve carrier synchronisation by this iterative method had been reported in [27] and brief estimates on its performance had been done based on empirical studies, for instance in [235].

The main breakthrough towards the estimation of the optimisation time has been done in preliminary work [59, 65, 236] where the model was already described as evolutionary random optimisation. In previous empirical studies, the implemented algorithm frequently employed a normal random distribution for the mutation of transmit phases and a mutation probability of 1. The upper and lower bounds have then been derived with standard tools for the analysis of simple evolutionary algorithms (assuming standard bit mutation and no crossover). Namely, we utilised the method of the expected progress in order to derive the

lower bound and the method of fitness-based partitions for the upper bound.

Based on the understanding of the problem domain, it was then also possible to derive further algorithmic improvements on the optimisation time. Two examples for this are included in the publication, namely the hierarchical clustering of transmit nodes and the asymptotically optimal optimisation method. The former exploits the fact that the synchronisation time grows more than linearly with the number of nodes that participate in the synchronisation. Therefore, hierarchically clustering node sets and finding synchronisation for these smaller sets can reduce the overall synchronisation time.

The asymptotically optimal approach exploits that, after random initialisation, with high probability, already more than half of all nodes are well synchronised to each other. It is then sufficient to alter the phase of each node in the corresponding direction. This is done with a minimum number of transmissions by solving an equation with unknowns of the underlying fitness function.

Summarising, this publication solves the main open issues regarding binary-feedback-based distributed carrier synchronisation and, with the concise mathematical description of the underlying problem space, provides tools to design optimal beamformers based on randomly distributed phase perturbations.

3.2 A fast binary feedback-based distributed adaptive carrier synchronisation for transmission among clusters of disconnected IoT nodes in smart spaces

The above studied closed-loop approach is computationally cheap at the cost of increasing the time required for carrier synchronisation. It utilises a binary feedback on the achieved synchronisation quality that is transmitted in each iteration from a remote receiver [26, 51]. In particular, such binary feedback can be implemented by a simple on/off burst scheme also for sharply resource restricted devices.

The central optimisation procedure consists of n devices $i \in [1, \ldots, n]$ randomly altering the phases γ_i of their carrier signal $\zeta_i(t)$ in each iteration. Implicitly, with this process a global random search is implemented. The search space $\mathcal{S}$ is spanned by all possible combinations of carrier frequencies and carrier phase offsets for all transmit nodes.

Each specific phase-frequency combination $s \in \mathcal{S}$ is associated with a score $\mathcal{F}_{sc} : \mathcal{S} \to \mathbb{R}_0^+$ that denotes its synchronisation quality. Without loss of generality we assume that the optimisation aim is to maximise $\mathcal{F}_{sc}$. A natural choice to compute such a score value is, for instance, the Signal-to-Noise-Ratio (SNR) of the received sum signal.

Intuitively, each node may in one iteration alter its transmit carrier phase offset, super-impose a synchronisation signal simultaneously with all other nodes and receive a binary feedback on the quality of the synchronisation. These iterations are repeated until a random distribution of carrier phases is achieved that scores a sufficient synchronisation quality [53, 52, 65]. Initially, independent and identically distributed (i.i.d.) phase offsets γ_i of carrier signals are assumed. Since a decreasing signal quality is not accepted, and

since a global random search is implemented by this approach (every possible combination of carrier phase offsets of nodes has a positive probability in each iteration) the method eventually converges to the optimum with probability 1 [52]. For this result an idealised environment without noise and interference was considered. In a realistic environment, the impact of the noise figure determines the accuracy that can be achieved.

In [66], an implementation of this carrier synchronisation approach was presented for software defined radio (SDR) devices which does not rely on any wired connections between devices (for instance, for clock synchronisation of the SDR nodes).

The authors of [53] then demonstrated in a case study that the method is feasible to synchronise frequency as well as phase of carrier signal components.

In all previous studies, a global random search is considered, in which nodes choose their next carrier phase and frequency offset uniformly at random from all possible values. However, this global random optimisation approach comparably slow in the optimisation process.

While an asymptotically optimal algorithm has been derived in [1], this requires more than binary feedback. In particular, function values of the feedback function have to be transmitted over the wireless channel. Assuming that the nodes seek for a common phase for transmission in order to establish a minimum Signal-To-Noise Ratio (SNR) at the receiver, this might be critical. Binary feedback can be encoded as simple burst-protocol (transmission vs. no transmission) which can be read out also at very low SNR. More advanced protocols capable of transmitting additional information, however, also require an increased SNR.

Since the search space does not contain local optima [59] we restrict the search neighbourhood to reduce the number of possible next configurations in one iteration that would worsen the synchronisation quality. We propose to modify the algorithm to follow a local random search instead of the previously applied global random search mechanism. In particular, a node i will, when it changes its phase and frequency offset, draw the new values from a restricted neighbourhood of size $\mathcal{N}$ that is centred around the current values of γ_i (and f_i). This addresses a recent critique expressed in [63] regarding the convergence speed for this binary feedback-based iterative adaptive carrier synchronisation.

The derived sharp asymptotic bounds on the expected optimisation time are again improved compared to the earlier studied global random search method.

Another question which was motivated by these studies on distributed adaptive transmit beamforming was on the impact of environmental effects. In our case studies, we had observed that the performance of the binary feedback-based carrier synchronisation approach was perceptive to presence and movement of individuals. In a first step, this motivated a series of studies on an adaptive learning approach for binary-feedback-based distributed adaptive beamforming in which the parameters of the optimisation algorithm – namely the mutation probability and the variance of the normal distribution on the neighbourhood function – were adapted according to the environment [127].

These considerations let to our work on RF-based device-free recognition systems (cf. section 2.3, section 2.4 and section 2.5) in which we considered the inmpact of different environmental conditions on the evolution of received RF-signals.

3.3 RF-sensing of activities from non-cooperative subjects in device-free recognition systems using ambient and local signals

In the approaching Internet of Things (IoT), virtually all entities in our environment will be enhanced by sensing, communication and computational capabilities [72, 73]. These entities will provide information on environmental situations, interact in the computation and processing of data [9] and store information. In order to sense environmental situations, common sensors in current applications are light, movement, pressure, audio or temperature [74]. Clearly, for reasons of cost and sensor size it is desired to minimise the count of distinct sensors in IoT entities. The one sensor class that defines the minimum set naturally available in virtually all IoT devices is the Radio Frequency (RF)-transceiver to communicate with other wireless entities. It is also shipped with nearly every contemporary electronic device like mobile phones, notebooks, media players, printers as well as keyboards, mouses, watches, shoes and rumour has spread about even media cups. Therefore, the RF transceiver is a ubiquitously available sensor class. It is capable of sensing changes or fluctuation in a received RF-signal. Radio waves are blocked, reflected or scattered at objects. At a receiver, the signal components from distinct signal paths add up to form a superimposition. When objects that block or reflect the signal path of some of these signal components are moved, this is reflected in the superimposition of signal waves at the receiver. We assert that specific activities in the proximity of a receiver generate characteristic patterns in the received superimposed RF-signal. By identifying and interpreting these patterns, it is possible to detect activities of non-cooperating subjects in an RF-receiver's proximity.

In the context of indoor localisation, Youssef defines this approach as Device-Free Localisation (DFL) in [78] to localise or track a person using RF-Signals while the entity monitored is not required to carry an active transmitter or receiver. They localised individuals by exchanging packets between 802.11b nodes in corners of a room and analysed the moving average and its variance of the RSSI [78]. A passive radio map was constructed offline before a Bayesian-based inference algorithm estimated the most probable location. These experiments have been conducted under Line-of-Sight (LoS) conditions. Also, Wilson and Patwari showed in conformance with the findings of Kosba et al. [117] that the variance of the RSSI can be used as an indicator of motion of non-actively transmitting individuals regardless of the average path loss that occurs due to dense walls and stationary objects [118]. The area in which environmental changes impact signal characteristics was then considered by Zhang et al. They used 870 MHz nodes arranged in a grid to show that for each link an elliptical area of about 0.5 to 1 meters diameter exists for which RSSI fluctuation caused by an object traversing this area exceeds measurements in a static environment [119]. They identified a valid region for detecting the impact (i.e. the RSSI fluctuations exceeding the measured threshold in a static environment) for transceiver distances from 2 m to 5 m for the considered 870 MHz frequency range [120]. By dividing a room into hexagonal cell-clusters with measurements following a TDMA scheduling, an

object position could be derived with an accuracy of around 1 meter. This accuracy was further improved by Wilson and Patwari in 2011 [118]. They utilised a dense node array to locate individuals within a room with an average error of about 0.5 meters. This was possible by instrumenting a tomographic image over the 2-way RSSI fluctuations of nodes [121]. All these studies consider a single experimental setting.

The simultaneous localisation of multiple individuals at the same time was first mentioned and studied by Patwari and Wilson in [124]. The authors derive a statistical model to approximate the position of a person based on RSSI variance which can be extended to multiple persons. This aspect together with the previously untackled problem that environmental changes over time might necessitate frequent calibration of the location system was approached by Zhang and others in [125]. The authors isolate the LoS path by extracting phase information from the differences in the RSS on various frequency spectra at distributed nodes. Their experimental system is with this approach able to simultaneously and continuously localise up to 5 persons in a changing environment with an accuracy of 1 meter.

While the localisation of individuals based on features from the radio channel was well elaborated, activity recognition from RF was still in its infants.

Patwari et al. monitored breathing based on RSS analysis [126]. The monitored area was surrounded by twenty 2.4 GHz nodes and the two-way RSSI was measured. Using a maximum likelihood estimator they approximated the breathing rate within 0.1 to 0.4 beats accuracy.

Our presented work was preceded by several smaller studies in which we were considering a simple set of classes to distinguish [75] and also open issues in device-free recognition of situations and activities [80].

Our study mainly focused on the recognition of the activities 'standing, lying, crawling, walking and empty' which we believe to be most expressive in emergency or elderly care situations.

The recognition of these activities has been considered by three device-free systems which we distinguish by the transmit and receive systems utilised. In particular, we utilised an active Software Defined Radio (SDR)-based Device-Free Activity Recognition (DFAR) system utilising Universal Software Radio Peripheral (USRP) transmit and receive device, an active RSSI-based DFAR system utilising sensor nodes as transmitters and recievers and a passive DFAR system exploiting ambient FM-radio. In all cases, unlike in most related work, we have restricted the system to a single receive device. While the accuracy can be improved by the utilisation of multiple receive nodes, we envision that in practical applications the number of available nodes will also be limited.

In comparison with a body-worn accelerometer, we could show that the recognition accuracy of the device-free systems on the recognition of the above mentioned classes in comparable. Consequently, this shows that, for the recognition of simple classes in, for instance, emergency or elderly care situations, device-free systems can replace body-worn sensing systems.

This opens also a range of new applications in which the monitored entity is not required to wear any device. Examples are cases in which the monitored entity is not cooperating

or where the requirement to wear a device induces additional cost (e.g. monetary, comfort, freedom of movement).

The optimum features identified for all systems differ slightly but mainly cover the mean, variance or maximum of the received signal strength.

3.4 Monitoring of Attention Using Ambient FM-radio Signals

After our initial studies which demonstrated the general feasibility of RF-based device-free recognition of activities via various recognition systems, we have considered various aspects of these systems regarding the recognition performance. Examples are the recognition of mulitple individuals [159] and also studies considering the distance of the monitored entity to a receive antenna [145, 237]. One result from these studies was that static classes such as lying or standing, although they can in principle be detected from RF-based DFAR systems, require extensive training in a specific environment and it is therefore unrealistic to sense such classes in practical applications. Dynamic classes, however, such as walking, crawling or also walking speed can be well recognised also across different environments and slight modifications of parameters of the sensing system such as topology or transmission power.

Likewise, other groups have been approaching RF-based passive recognition utilising different sensing technology like, for instance, RFID nodes [156] or also more sophisticated recognition algorithms [238]. In this sense, RF-based device-free recognition received a broad recognition from various directions and we were also trying to push the limits of the recognition algorithms further towards the detection of non-directly observable properties such as attention.

In the featured study, we were considering attention of individuals towards poster frames in a corridor. Attention determines for a system the potential to impact the actions and decisions taken by an individual [139]. The management of attention covers the activation of attention as well as its detection and timely exploitation. In the literature, we find various definitions that classify attention as well as its determining characteristics [140, 239]. A straightforward measure of attention might be the tracking of gaze [141]. In general, aspects such as Saliency, Effort, Expectancy and Value are important indicators of attention [142, 239]. Others extended this model and put a greater stress on the effort a person takes towards an object [143]. Such detection and management of attention may require elaborate installations and very specific sensors in order to accurately sense quantities such as Saliency, Effort, Expectancy and Value [139, 240]. However, we believe that for many commercial installations, cost and ease of installation and not primarily the highest achievable accuracy are most important. Also more general, environmental sensors can provide sufficient information to estimate the attention state of individuals.

We propose to utilise ambient FM-radio signals for the detection of attention since it has a nearly perfect coverage in populated areas and features cheap receiver hardware [144].

In particular, authors in [143] discuss various aspects of attention and identify as most distinguishing factors changes in walking speed, direction or orientation. We therefore argue that attention levels can be inferred upon interpretation of the changes in walking speed or direction as derived from our system.

Our passive, FM-radio based DFAR system then detects the location and walking speed of persons in that corridor. From this information, for instance the location of the person and change in her walking speed, it is then possible to estimate attention levels towards specific poster frames in the corridor.

This work also raises questions whether further complex classes, such as, for instance, sentiment can be reliably detected from device free RF-based sensing system.

3.5 The Telepathic Phone: Frictionless Activity Recognition from WiFi-RSSI

Throughout 2013, research campaigns in particular at the University of Washington and at the MIT have further pushed the limits of which classes can be accurately sensed from RF-based device-free recognition systems. In particular, the Doppler shift was exploited by these groups as a powerful feature to accurately sense fine-grained movement.

When an object reflecting a signal wave is in motion, this causes Doppler Shift. The direction and speed of the movement conditions the strength and nature of this frequency shift. Pu and others showed that simultaneous detection of gestures from multiple individuals is possible by utilising multi-antenna nodes and micro Doppler fluctuations [154, 160]. They utilise a USRP SDR multi antenna receiver and one or more single antenna transmitters distributed in the environment to distinguish between a set of 9 gestures with an average accuracy of 0.94. Their active device-free system exploits a MIMO receiver in order to recognise gestures from different persons present at the same time. By leveraging a preamble gesture pattern, the receiver estimates the MIMO channel that maximises the reflections of the desired user.

A main challenge was for them that the Doppler shift from human movement was several magnitudes smaller than the bandwidth of the signal employed. The authors therefore proposed to transform the received signal into several narrowband pulses which are then analysed for possible Doppler fluctuation. The group discussed application possibilities of their system in [241].

In a related system, Adib and Katabi employ MIMO interference nulling and combine samples taken over time to achieve a similar result while compensating for the missing spatial diversity in a single-antenna receiver system [153]. In their system, they leverage standard WiFi hardware at 2.4 GHz.

Later, this work was extended to 3D motion tracking by utlising three or more directional receive antennas in exactly defined relative orientation [242]. In particular, the system is able to track the center of a human body with an error below 21cm in any direction and can also detect movement of body parts and directions of a pointing body part, such as

a hand. This localisation is possible through time-of-flight estimation and triangulisation. Higher accuracy of this estimation is granted by utilising frequency modulated carrier waves (sending a signal that changes linearly in frequency with time) over a bandwidth of 1.69 GHz. Impact of static objects could be mitigated by subtracting successive sample means whereas noise was filtered by its speed of changes in energy over frequency bands.

For these systems, however, sophisticated SDR devices are are necessary as well as extensive training of the system.

Towards another dimension, we were therefore considering the RF-based device-free recognition of activities, situations and gestures on consumer devices. On such device classes, the high accuracy and also the direct access to signals on the wireless channel are not provided. Instead, utilising IEEE 802.11g signals, we are restricted to the RSSI information calculated for each received packet. It turned out that this information is particularly challenging since the packet reception is sparse, bursty and RSSI calculation is inaccurate and discrete. We therefore limited our study to very simple time-domain features, mainly the mean RSSI, variance as well as minimum and maximum RSSI. In this first study on passive RSSI-based DFAR, we investigated the distinction of distance, movement speed, crowd size as well as simple presence and interaction with the phone. With about 20 packets per second, these classes can be recognised with an accuracy between 0.9 and 0.7. It was, however, not possible to distinguish the direction in which an activity was performed. Furthermore, gestures showed only a low recognition accuracy.

With this study, the first to utilise passive RSSI on a smartphone device, we presented an alternative to the frequent use of accelerometers for the recognition of activities. When the smartphone is not worn on the body, which occurs more than 50% of the time according to [93, 94], accelerometer-based sensing of activities is infeasible. In such cases, i.e. when the smartphone is lying on the table, RSSI-based device-free sensing paradigms still enable the detection of simple activities or situations. Furthermore, this sensing paradigm enables a range of novel application cases in which environmental conditions might trigger actions on a smartphone.

Rencent developments further advanced these ideas to also cover the recognition of emotional states from gestures captured from RSSI or channel state information (CSI) fluctuation [243, 244].

3.6 Secure communication based on ambient audio

Secure authentication and communication in wireless settings is still a topic with a number of unsolved issues. In mobile computing, when two devices meet for the first time, how is it possible to authenticate devices and to establish a secure key without the need of a trusted third party.

In the absence of a shared key, the first communication over the channel is necessarily unencrypted. Hence, this opens the door for an eavesdropper within the communication range. Furthermore, how can the communication partner be authenticated. It is easy in such wireless scenario for an attacker to forge the identity of the legal communication

partner.

Context as an implicit input might help to limit the amount of potential communication partners. This concept was presented first 2001 by Holmquist et al. [170]. The authors propose to utilise the accelerometer of the Smart-It [172] device to extract characteristic features from simultaneous shaking processes of two devices. Later, Mayrhofer et al. presented an authentication mechanism based on this principle [173]. 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 [174] that the sharing of secret keys is possible with a similar protocol. The proposed protocol that can be utilised with arbitrary contextual 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 [175, 169]. 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.

In our publication 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 [171]. Hershey et al. utilise physical layer features to derive secret keys for a pair of devices [176]. In the absence of interference and non-linear components, transmitter and receiver experience identical channel response [177]. 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 [178], an eavesdropper is not capable of guessing information about the secret. This scheme was validated in an indoor environment in [179]. 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.

In our work we present an approach to establish a common cryptographic key for co-located devices from ambient audio, without leaking information to eavesdroppers in a different audio context. 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 [180] is possible. The strength of the key established is conditioned on the length of the audio sequences considered but sequence of less than 6 seconds are already sufficient to obtain 512 bit keys.

The main novelty of our approach compared to other context-based device pairing mechanisms is the use of fuzzy cryptography to mitigate differences in the audio fingerprints of co-located audio recordings. Through the strong error correcting codes employed (we utilise Reed-Solomon error correcting codes), it is possible to exactly define how many differences are tolerated in fingerprints of co-located devices so as to establish identical cryptographic keys.

We implemented the approach using python on a pair of desktop computers. A critical problem we faced was the time synchronisation among devices. Standard NTP-based synchronisation showed to be too inaccurate. In particular, since the fingerprints are created from chunks of audio recordings in the time domain of length 0.375 seconds, a time difference of more than 20 milliseconds already results in significant difference of the generated fingerprint.

Our study demonstrated that ambient audio is a rich source of randomness with good Entropy. Indeed, employing sets of statistical tests we could not identify a bias in the generated fingerprints.

Our case studies also showed that even with partial information on the audio context, an eavesdropper can be prevented from obtaining a valid fingerprint. In this case, however, the parameters of the Reed-Solomon algorithm have to be stricter so that also more of the legitimate pairing attempts fail. An automatic and adaptive choice of parameters in a given environment is yet an open issue to solve.

The presented approach raises the bar for attacks in wireless environments since attackers would have to establish physical proximity in order to be able to authenticate. Furthermore, it enables simpler, easier to use security schemes since part of the key can be derived automatically via ambient audio.

Li et al. analyse the usage of biometric or multimedia data as part of an authentication process and propose a protocol [181]. 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 [182] 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. [183]. The authors establish a key distribution based on a fuzzy vault [184] using data measured by devices worn on the human body. The fuzzy vault scheme, also utilised in [185], enables the decryption of a secret with any key that is substantially similar to the key used for encryption.

Recently, the scheme developed by us in this work has been further advanced towards other domains, such as e.g. pairing conditioned on gait or co-presence on the same

body [245].

3.7 Pattern-based Alignment of Audio Data for Ad-hoc Secure Device Pairing

Our instrumentation described above requires idealised conditions regarding the synchronisation of devices and to account for this a high number of fingerprints must be created (201 in our experiments) in order to find one matching fingerprint. For extensive computational load, this is feasible only in an offline approach. The high number of fingerprints created, however, was necessary since the utilised NTP synchronisation is not sufficiently accurate.

In a later paper, we present an alignment mechanism which enables a synchronisation accuracy of recorded audio in the order of less than 10 milliseconds among unsynchronised mobile devices in the same context without transmitting information about the audio sequence over the wireless channel. The synchronisation is achieved by processing a weakly NTP-synchronised recording without additional communication among the devices.

This implementation solved several practical issues and constitutes a prototype implementation for smart and pervasive environments. The android-implementation is available online[1] and the method was also included in the Open-UAT API for secure applications on android [2].

While the main pairing approach is identical, challenges arised through hardware peculiarities of the phones utilised.

In particular, the hardware-based audio-pre-processing on one of the phones and the non-real-time capability of the operating system posed greatest difficulties. While the hardware-pre-processing issue could be easily solved by restricting the frequency range from which audio-fingerprints are created, the non-real-time capability of the phone posed greater difficulties. In particular, for the start of the audio recording, we experienced unpredictable delays of up to two seconds. We solved the challenge to find an identical starting point in these sequences without inter-device-communication by utilising approximate pattern matching between a common short audio sequence and the recording on each of the devices. While the common sequence was uncorrelated to the recording, a best matching would be found with high probability approximately at the same location within the audio recording provided that the recordings are similar.

With this approach we have been able to identify an identical starting time for the generation of the fingerprints with an offset in the order of ten milliseconds.

---

[1]https://github.com/stephansigg/AdhocPairing.git
[2]http://www.openuat.org/

4 Acknowledgements

A number of people have contributed to this thesis in various ways. My work and studies towards this thesis have profited much from these positive contributions which have shaped and impacted the direction of my work.

First of all, I would like to thank my three girls, Nelja, Freyja and Aniko. Thank you for your patience, understanding and support!

I would like to thank Professor Dr. Michael Beigl for giving me the opportunity to work with much freedom on these topics. Without the financial support he mobilised to acquire the hardware for the experimental studies, many findings and experiences could not have been made. Likewise, Professor Dr. Yusheng Ji has greatly supported this work with suggestions and advises and also financial support for equipment. Professor Dr. Lars Wolf helped me significantly with many issues after my return from japan and shared an office in his group which greatly helped to improve my working conditions at that time. Professor Dr. Xiaoming Fu greatly supported and re-integrated me into the German academic system and thereby essentially enabled the writing on this document.

Over the last seven years, I have had the opportunity to cooperate with many great scientists which have had a strong impact on my work and the results presented in this document. In particular, I would like to thank the team at TecO for stirring discussions and helpful advice on very diverse topics. I have experienced a very competitive research environment in this group. Dawud Gordon, in particular, has always been available for discussions and practical advice. In Behnam Banitalebi I have found an electrical engineer with whom I have had the good fortune of working with. Many of the studies on device-free recognition have been conducted in close cooperation with Markus Scholz. I am thankful for constructive discussions on this topic and for his patience. Furthermore, Matthias Budde, Till Riedel, Hedda Schmidtke, Predrag Jakimovski and Martin Berchtold have contributed to parts of the work presented here and I am very thankful for having had the chance to cooperate with them.

I would further like to thank the k-lab team at the National Institute for Informatics, in particular Shuyu Shi. I am honoured to have worked closely together with Shuyu during 2011 and 2012. Your passion and clear focus on research has inspired me greatly. Furthermore, I have had the chance to cooperate with passionate advanced scientists at NII, in particular, Olga Streibel, Sven Wohlgemuth and Christoph Lofi.

From the IBR at TU-Braunschweig, in particular Dominik Schuermann, Felix Buesching and Sebastian Schildt had a great impact on parts of the work presented here. Dominik

has helped me a lot with security and cryptography-related issues and has significantly contributed to the implementation of the audio-based device pairing approach. Felix has introduced the RF-based recognition to the inga sensor nodes and has significantly contributed to our studies on the active RSSI-based device-free activity recognition system.

I am very thankful towards the team of the wearable computing laboratory of Professor Dr. Gerhard Troester at ETH Zurich. I was impressed by the high scientific standard each team member maintained. In particular, Ulf Blanke greatly contributed to our study on passive device-free activity recognition from received 802.11g packets. I greatly enjoyed the constructive and positive discussions with Sinziana Mazilu, Alberto Calatroni, Sebastian Feese, Franz Gravenhorst, Giovanni Antonio Salvatore, Zhu Zack and with my roommate Simon Christen.

I would also like to mention the very warm welcome in the Nodes group of the University of Helsinki and good and fruitful discussions, in particular with Sasu Tarkoma, N. Asokan, Eemil Lagerspetz, Ella Peltronen, Petteri Nurmi, Hien Truong and Markus Miettinen. I have learned a lot during this stay, in particular, about big data management and algorithmic tools.

Also, I am grateful for discussions and support with all kinds of academic issues by the ComTec team at Kassel University, in particular, Professor, Dr. Klaus David, Niklas Klein, Rico Kusber and Andreas Jahn.

Finally, the members of the ComNet group at University of Goettingen provide me with a new academic home and have inspired me with new ideas and research aspects towards future networking. In particular, David Koll, Mayutan Arumaithurai, Konglin Zhu and Jiachen Chen.

During these years I am grateful to have had the opportunity to closely work together with very talented and passionate students. The interesting and deep discussions have often led to refinements or clarifications of my work. In particular, I would like to mention Jialin Wang, Weiwei Liang, Rayan Merched El Masri, Julian Ristau, Aaron Israel, Georg von Zengen, Gerrit Bagschik, Toni Günther, Johannes Starosta, Sebastian Schwarzl, Markus Reschke, Timo Schulz, Sascha Lity, Stephen Roettger, Sergei Dechand, Ngu Nguyen, David Rieger, Mario Hock, An Huynh, Stephan Mueller, Philipp Specht, Christoph Rauterberg, Marko Becker, Matthias Velten and Chuong Thach Nguyen.

List of Tables

List of Figures

Bibliography

[1] Sigg, S., Masri, R.M.E., Beigl, M.: Feedback based closed-loop carrier synchronisation: A sharp asymptotic bound, an asymptotically optimal approach, simulations and experiments. Transactions on mobile computing **10**(11) (2011) 1605–1617

[2] Sigg, S.: A fast binary feedback-based distributed adaptive carrier synchronisation for transmission among clusters of disconnected iot nodes in smart spaces. Ad Hoc Netw. (2014) http://dx.doi.org/10.1016/j.adhoc.2013.12.006.

[3] Sigg, S., Scholz, M., Shi, S., Ji, Y., Beigl, M.: Rf-sensing of activities from non-cooperative subjects in device-free recognition systems using ambient and local signals. IEEE Transactions on Mobile Computing **13**(4) (2013)

[4] Shi, S., Sigg, S., Zhao, W., Ji, Y.: Monitoring of attention from ambient fm-radio signals. IEEE Pervasive Computing, Special Issue - Managing Attention in Pervasive Environments (2014)

[5] Sigg, S., Blanke, U., Troester, G.: The telepathic phone: Frictionless activity recognition from wifi-rssi. In: IEEE International Conference on Pervasive Computing and Communications (PerCom). PerCom '14 (2014)

[6] Mathur, S., Miller, R., Varshavsky, A., Trappe, W., Mandayam, N.: Proximate: Proximity-based secure pairing using ambient wireless signals. In: Proceedings of the ninth International Conference on Mobile Systems, Applications and Services (MobiSys 2011). (2011)

[7] Schuermann, D., Sigg, S.: Secure communication based on ambient audio. IEEE Transactions on mobile computing **12**(2) (2013)

[8] Nguyen, N., Sigg, S., Huynh, A., Ji, Y.: Pattern-based alignment of audio data for ad-hoc secure device pairing. In: Proceedings of the 16th annual International Symposium on Wearable Computers (ISWC). (2012)

[9] Sigg, S., Jakimovski, P., Beigl, M.: Calculation of functions on the rf-channel for iot. In: 3rd International Conference on the Internet of Things (IOT). (2012) 107–113

[10] Jakimovski, P., Schmidtke, H.R., Sigg, S., Weiss, L., Chaves, F., Beigl, M.: Collective communication for dense sensing environments. Journal of Ambient Intelligence and Smart Environments (JAISE) **4**(2) (2012)

[11] Sigg, S., Jakimovski, P., Ji, Y., Beigl, M.: Utilising an algebra of random functions to realise function calculation via a physical channel. In: 14th IEEE workshop on Signal Processing Advances in Wireless Communications (SPAWC). (2013)

[12] Culler, D., Estrin, D., Srivastava, M.: Overview of sensor networks. IEEE Computer **37**(8) (2004) 41–49

[13] Zhao, F., Guibas, L.: Wireless Sensor Networks: An Information Processing Approach. Morgan Kaufmann, Los Altos, CA (2004)

[14] Norman, D.: The invisible computer. MIT press (1999)

[15] Butera, W.J.: Programming a paintable computer. PhD thesis, Massachusetts Institute of Technology (2002)

[16] Pillutla, L., Krishnamurthy, V.: Joint rate and cluster optimisation in cooperative mimo sensor networks. In: Proceedings of the 6th IEEE Workshop on signal Processing Advances in Wireless Communications. (2005) 265–269

[17] Scaglione, A., Hong, Y.W.: Opportunistic large arrays: Cooperative transmission in wireless multihop ad hoc networks to reach far distances. IEEE Transactions on Signal Processing **51**(8) (2003) 2082–2092

[18] Sendonaris, A., Erkop, E., Aazhang, B.: Increasing uplink capacity via user cooperation diversity. In: IEEE Proceedins of the International Symposium on Information Theory (ISIT). (2001) 156

[19] Laneman, J., Wornell, G., Tse, D.: An efficient protocol for realising cooperative diversity in wireless networks. In: Proceedings of the IEEE International Symposium on Information Theory. (2001) 294

[20] Hong, Y.W., Scaglione, A.: Critical power for connectivity with cooperative transmission in wireless ad hoc sensor networks. In: IEEE Workshop on Statistical Signal Processing. (2003)

[21] Hong, Y.W., Scaglione, A.: Energy-efficient broadcasting with cooperative transmission in wireless sensor networks. IEEE Transactions on Wireless communications (2005)

[22] Jayaweera, S.K.: Energy analysis of mimo techniques in wireless sensor networks. In: 38th conference on information sciences and systems. (2004)

[23] del Coso, A., Sagnolini, U., Ibars, C.: Cooperative distributed mimo channels in wireless sensor networks. IEEE Journal on Selected Areas in Communications **25**(2) (2007) 402–414

[24] Sigg, S., Beigl, M.: Collaborative transmission in wsns by a (1+1)-ea. In: Proceedings of the 8th International Workshop on Applications and Services in Wireless Networks (ASWN'08). (2008)

[25] Sigg, S., Beigl, M.: Randomised collaborative transmission of smart objects. In: 2nd International Workshop on Design and Integration principles for smart objects (DIPSO2008) in conjunction with Ubicomp 2008. (2008)

[26] Mudumbai, R., Brown, D.R., Madhow, U., Poor, H.V.: Distributed transmit beamforming: Challenges and recent progress. IEEE Communications Magazine (2009) 102–110

[27] Mudumbai, R., Wild, B., Madhow, U., Ramchandran, K.: Distributed beamforming using 1 bit feedback: from concept to realization. In: Proceedings of the 44th Allerton conference on communication, control and computation. (2006) 1020–1027

[28] Barriac, G., Mudumbai, R., Madhow, U.: Distributed beamforming for information transfer in sensor networks. In: Proceedings of the third International Workshop on Information Processing in Sensor Networks. (2004)

[29] Mudumbai, R., Barriac, G., Madhow, U.: On the feasibility of distributed beamforming in wireless networks. IEEE Transactions on Wireless communications **6** (2007) 1754–1763

[30] Ochiai, H., Mitran, P., Poor, H.V., Tarokh, V.: Collaborative beamforming for distributed wireless ad hoc sensor networks. IEEE Transactions on Signal Processing **53**(11) (2005) 4110 – 4124

[31] Chen, W., Yuan, Y., Xu, C., Liu, K., Yang, Z.: Virtual mimo protocol based on clustering for wireless sensor networks. In: Proceedings of the 10th IEEE Symposium on Computers and Commmunications. (2005)

[32] Youssef, M., Yousif, A., El-Sheimy, N., Noureldin, A.: A novel earthquake warning system based on virtual mimo wireless sensor netwroks. In: Canadian conference on electrical and computer engineering. (2007) 932–935

[33] del Coso, A., Savazzi, S., Spagnolini, U., Ibars, C.: Virtual mimo channels in cooperative multi-hop wireless sensor networks. In: 40th annual conference on information sciences and systems. (2006) 75–80

[34] Jayaweera, S.K.: Energy efficient virtual mimo based cooperative communications for wireless sensor networks. IEEE Transactions on Wireless communications **5**(5) (2006) 984–989

[35] Laneman, J., Wornell, G.: Distributed space-time coded protocols for exploiting cooperative diversity in wireless networks. IEEE Transactions on Information theory **49**(10) (2003) 2415–2425

[36] Sendonaris, A., Erkip, E., Aazhang, B.: User cooperation diversity – part i: System description. IEEE Transactions on Communications **51**(11) (2003) 1927–1938

[37] Zimmermann, E., Herhold, P., Fettweis, G.: On the performance of cooperative relaying protocols in wireless networks. European Transactions on Telecommunications **16**(1) (2005) 5–16

[38] Cover, T.M., Gamal, A.A.E.: Capacity theorems for the relay channel. IEEE Transactions on Information Theory **525**(5) (1979) 572–584

[39] Kramer, G., Gastpar, M., Gupta, P.: Cooperative strategies and capacity theorems for relay networks. IEEE Transactions on Information Theory **51**(9) (2005) 3037–3063

[40] Scaglione, A., Hong, Y.W.: Cooperative models for synchronization, scheduling and transmission in large scale sensor networks: An overview. In: 1st IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing. (2005) 60–63

[41] Gupta, P., Kumar, R.P.: The capacity of wireless networks. IEEE Transactions on Information Theory **46**(2) (2000) 388–404

[42] Ochiai, H., Mitran, P., Poor, H.V., Tarokh, V.: Collaborative beamforming for distributed wireless ad hoc sensor networks. IEEE Transactions on Signal Processing **53**(11) (2005) 4110 – 4124

[43] Simeone, O., Spagnolini, U.: Capacity region of wireless ad hoc networks using opportunistic collaborative communications. In: Proceedings of the International Conference on Communications (ICC). (2006)

[44] Krohn, A., Beigl, M., Decker, C., Varona, D.G.: Increasing connectivity in wireless sensor network using cooperative transmission. In: 3rd International Conference on Networked Sensing Systems (INSS). (2006)

[45] Krohn, A.: Optimal non-coherent m-ary energy shift keying for cooperative transmission in sensor networks. In: 31st IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). (2006)

[46] Hong, Y.W., Scaglione, A.: Cooperative transmission in wireless multi-hop ad hoc networks using opportunistic large arrays. In: 4th IEEE workshop on Signal Processing Advances in Wireless Communications (SPAWC). (2003)

[47] Brown, D.R., Prince, G., McNeill, J.: A method for carrier frequency and phase synchronization of two autonomous cooperative transmitters. In: sixth IEEE workshop on signal processing advances in wireless communications. (2005)

[48] Brown, D.R., Poor, H.V.: Time-slotted round-trip carrier synchronisation for distributed beamforming. IEEE Transactions on Signal Processing **56** (2008) 5630–5643

[49] Ozil, I., Brown, D.R.: Time-slotted round-trip carrier synchronisation. In: Proceedings of the 41st Asilomar conference on signals, signals and computers. (2007) 1781–1785

[50] Tu, Y., Pottie, G.: Coherent cooperative transmission from multiple adjacent antennas to a distant stationary antenna through awgn channels. In: Proceedings of the IEEE Vehicular Technology Conference. (2002) 130–134

[51] Mudumbai, R., Hespanha, J., Madhow, U., Barriac, G.: Scalable feedback control for distributed beamforming in sensor networks. In: Proceedings of the IEEE International Symposium on Information Theory. (2005) 137–141

[52] Mudumbai, R., Hespanha, J., Madhow, U., Barriac, G.: Distributed transmit beamforming using feedback control. IEEE Transactions on Information Theory **56**(1) (2010)

[53] Seo, M., Rodwell, M., Madhow, U.: A feedback-based distributed phased array technique and its application to 60-ghz wireless sensor network. In: IEEE MTT-S International Microwave Symposium Digest. (2008) 683–686

[54] Bucklew, J.A., Sethares, W.A.: Convergence of a class of decentralised beamforming algorithms. IEEE Transactions on Signal Processing **56**(6) (2008) 2280–2288

[55] Bennett, W.: Introduction to signal transmission. McGraw-Hill (1971)

[56] Krohn, A.: Superimposed Radio Signals for Wireless Sensor Networks. PhD thesis, Technical University of Braunschweig (2007)

[57] 3GPP: 3rd generation partnership project; technical specification group radio access networks; 3g home nodeb study item technical report (release 8). Technical Report 3GPP TR 25.820 V8.0.0 (2008-03) (March)

[58] Jakimovski, P., Becker, F., Sigg, S., Schmidtke, H.R., Beigl, M.: Collective communication for dense sensing environments. In: 7th IEEE International Conference on Intelligent Environments (IE). (2011) (**Best paper**).

[59] Sigg, S., Beigl, M.: Algorithms for closed-loop feedback based distributed adaptive beamforming in wireless sensor networks. In: Proceedings of the fifth International Conference on Intelligent Sensors, Sensor Networks and Information Processing - Symposium on Adaptive Sensing, Control, and Optimization in Sensor Networks. (2009)

[60] Hagmann, W.: Network synchronisation techniques for satellite communication systems. PhD thesis, USC, Los Angeles (1981)

[61] Mudumbai, R., Madhow, U., Brown, R., Bidigare, P.: Dsp-centric algorithms for distributed transmit beamforming. In: 2011 Conference Record of the 45th Asilomar Conference on Signals, Systems and Computers (ASILOMAR). (2011) 93–98

[62] Quitin, F., Rahman, M.M.U., Mudumbai, R., Madhow, U.: A scalable architecture for distributed transmit beamforming with commodity radios: Design and proof of concept. IEEE Transactions on Wireless Communications (**12**(3)) 1418–1428

[63] Mudumbai, R., Bidigare, P., Pruessing, S., Dasgupta, S., Oyarzun, M., Raeman, D.: Scalable feedback algorithms for distributed transmit beamforming in wireless networks. In: IEEE Internatinal Conference on Acoustics, Speech and Signal Processing (ICASSP). (2012) 5213 –5216

[64] Masri, R.M.E., Sigg, S., Beigl, M.: An asymptotically optimal approach to the distributed adaptive transmit beamforming in wireless sensor networks. In: Proceedings of the 16th European Wireless Conference. (2010)

[65] Sigg, S., Beigl, M.: Algorithmic approaches to distributed adaptive transmit beamforming. In: Fifth International Conference on Intelligent Sensors, Sensor Networks and Information Processing - Symposium on Theoretical and Practical Aspects of Large-scale Wireless Sensor Networks. (2009)

[66] Rahman, M.M., Baidoo-Williams, H.E., Mudumbai, R., Dasgupta, S.: Fully wireless implementation of distributed beamforming on a software-defined radio platform. In: Proceedings of the 11th international conference on Information Processing in Sensor Networks. IPSN '12 (2012) 305–316

[67] Quitin, F., Rahman, M.M.U., Mudumbai, R., Madhow, U.: Distributed beamforming with software-defined radios: frequency synchronisation and digital feedback. In: Proceedings of the 55th International Global Communications Conference (Globecom). (2012)

[68] Savage, C., Winkler, P.: Monotone gray codes and the middle levels problem. Journal of Combinatorial Theory **70**(2) (1995) 230–248

[69] Knuth, D.E.: The Art of Computer Programming – Combinatarial Algorithms, Part 1. Addison-Wesley (2011)

[70] Reschke, M., Starosta, J., Schwarzl, S., Sigg, S.: Situation awareness based on channel measurements. In: Vehicular Technology Conference (VTC Spring), 2011 IEEE 73rd. (2011)

[71] Reschke, M., Schwarzl, S., Starosta, J., Sigg, S., Beigl, M.: Context awareness through the rf-channel. In: Proceedings of the 2nd workshop on Context-Systems Design, Evaluation and Optimisation. (2011)

[72] Li, T., Chen, L.: Internet of things: Priinciples, frameworks and applications. In: Proceedings of the Future Wireless Networks and Information Systems. Number 144 in LNEE (2008) 477–842

[73] Haller, S.: The things in the internet of things. In: Proceedings of the Internet of Things Conference 2010. (2010)

[74] Beigl, M., Krohn, A., Zimmer, T., Decker, C.: Typical sensors needed in ubiquitous and pervasive computing. In: First International Workshop on Networked Sensing Systems. Volume 4 of Society of Instrument and Control Engineers. (2004) 153–158

[75] Scholz, M., Sigg, S., Shihskova, D., von Zengen, G., Bagshik, G., Guenther, T., Beigl, M., Ji, Y.: Sensewaves: Radiowaves for context recognition. In: Video Proceedings of the 9th International Conference on Pervasive Computing (Pervasive 2011). (2011)

[76] Woyach, K., Puccinelli, D., Haenggi, M.: Sensorless sensing in wireless networks: implementation and measurements. In: Proceedings of the Second International Workshop on Wireless Network Measurement (WiNMee). (2006)

[77] Muthukrishnan, K., Lijding, M., Meratnia, N., Havinga, P.: Sensing motion using spectral and spatial analysis of wlan rssi. In: Proceedings of Smart Sensing and Context. (2007)

[78] Youssef, M., Mah, M., Agrawala, A.: Challenges: Device-free passive localisation for wireless environments. In: Proceedings of the 13th annual ACM international Conference on Mobile Computing and Networking (MobiCom 2007). (2007) 222–229

[79] Seifeldin, M., Saeed, A., Kosba, A., El-Keyi, A., Youssef, M.: Nuzzer: A large-scale device-free passive localization system for wireless environments. IEEE Transactions on Mobile Computing (TMC) **12**(7) (2013)

[80] Scholz, M., Sigg, S., Schmidtke, H.R., Beigl, M.: Challenges for device-free radio-based activity recognition. In: Proceedings of the 3rd workshop on Context Systems, Design, Evaluation and Optimisation, in conjunction with MobiQuitous 2011. (2011)

[81] Shi, S., Sigg, S., Ji, Y.: Passive detection of situations from ambient fm-radio signals. In: Proceedings of the 2012 ACM Conference on Ubiquitous Computing. UbiComp '12 (2012)

[82] Shi, S., Sigg, S., Ji, Y.: Activity recognition from radio frequency data: Multi-stage recognition and features. In: IEEE Vehicular Technology Conference (VTC Fall). (2012)

[83] Ravi, N., Dandekar, N., Mysore, P., Littman, M.L.: Activity recognition from accelerometer data. In: Proceedings of the 17th conference on Innovative applications of artificial intelligence - Volume 3. IAAI'05 (2005) 1541–1546

[84] Cao, H., Nguyen, M.N., Phua, C.C.W., Krishnaswamy, S., Li, X.: An integrated framework for human activity classification. In: Proceedings of the 14th ACM International Conference on Ubiquitous Computing (UbiComp 2012). (2012)

[85] Bao, L., Intille, S.S.: Activity recognition from user-annotated acceleration data. In: Proceedings of PERVASIVE 2004. Volume LNCS 3001. (2004)

[86] Ploetz, T., Hammerla, N.Y., Rozga, A., Reavis, A., Call, N., Abowd, G.D.: Automatic assessment of problem behavior in individuals with developmental disabilities. In: Proceedings of the 14th ACM International Conference on Ubiquitous Computing (Ubicomp 2012). (2012)

[87] Abdullah, S., Lane, N.D., Choudhury, T.: Towards population scale activity recognition: A scalable framework for handling data diversity. In: Proceedings of the 26th conference on artificial intelligence (AAAI 2012). (2012)

[88] Chavarriaga, R., Bayati, H., del R. Millan, J.: Unsupervised adaptation for acceleration-based activity recognition: robustness to sensor displacement and rotation. Personal and Ubiquitous Computing (2011)

[89] Cohn, G., Morris, D., Patel, S.N., Tan, D.S.: Humantenna: Using the body as an antenna for real-time whole-body interaction. In: Proceedings of ACM CHI 2012. (2012)

[90] Patridge, K., Golle, P.: On using existing time-use study data for ubiquitous computing applications. In: Proceedings of the 10th International Conference on Ubiquitous Computing. (2008) 144–153

[91] Varshavsky, A., Patel, S.N. Ubiquitous Computing Fundamentals. In: Location in Ubiquitous Computing. Taylor and Francis Group (2008) 285–319

[92] Lane, N.D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T., Campbell, A.T.: A survey of mobile phone sensing. IEEE Communications magazine **48**(9) (2010) 140–150

[93] Dey, A.K., Wac, K., Ferreira, D., Tassini, K., Hong, J.H., Ramos, J.: Getting closer: An empirical investigation of the proximity of user to their smart phones. In: Proceedings of the 13th international conference on Ubiquitous computing. (2011)

[94] Patel, S.N., Kientz, J.A., Hayes, G.R., Bhat, S., Abowd, G.D.: Farther than you may think: An empirical investigation of the proximity of users to their mobile phones. In: Proceedings of the 8th international conference on Ubiquitous computing. (2006) 123–140

[95] Hongeng, S., Nevatia, R., Bremond, F.: Video based event recognition: Activity representation and probabilistic recognition methods. Computer Vision and Image Understanding **96** (2004) 129–162

[96] Bannach, D., Lukowicz, P., Amft, O.: Rapid prototyping of activity recognition applications. Pervasive Computing, IEEE **7**(2) (2008) 22 –31

[97] Patel, S.N., Robertson, T., Kientz, J.A., Reynolds, M.S., Abowd, G.D.: At the flick of a switch: Detecting and classifying unique electrical events on the residential power line. In: Proceedings of the 9th International Conference on Ubiquitous Computing (UbiComp 2007). (2007) 271–288

[98] Gupta, S., Reynolds, M.S., Patel, S.N.: Electrisense: Single-point sensing using emi for electrical event detection and classificaiton in the home. In: Proceedings of the 13th international conference on Ubiquitous computing. (2010)

[99] Gupta, S., Chen, K.Y., Reynolds, M.S., Patel, S.N.: Lightwave: Using compact fluorescent lights as sensors. In: Proceedings of the 13th international conference on Ubiquitous computing. (2011)

[100] Campbell, A.T., Larson, E., Cohn, G., Froehlich, J., Alcaide, R., Patel, S.N.: Wattr: A method for self-prowered wireless sensing of water activity in the home. In: Proceedings of the 12th international conference on Ubiquitous computing. (2010)

[101] Thomaz, E., Bettadapura, V., Reyes, G., Sandesh, M., Schindler, G., Ploetz, T., Abowd, G.D., Essa, I.: Recognizing water-based activities in the home through infrastructure-mediated sensing. In: Proceedings of the 14th ACM International Conference on Ubiquitous Computing (Ubicomp 2012). (2012)

[102] Cohn, G., Stuntebeck, E., Pandey, J., Otis, B., Abowd, G.D., Patel, S.N.: Snupi: Sensor nodes utilizing powerline infrastructure. In: Proceedings of the 12th international conference on Ubiquitous computing (UbiComp 2010). (2010)

[103] Bahl, P., Padmanabhan, V.: Radar: an in-building rf-based user location and tracking system. In: Proceedings of the 19th IEEE International Conference on Computer Communications (Infocom). (2000)

[104] Otsason, V., Varshavsky, A., LaMarca, A., de Lara, E.: Accurate gsm indoor localisation. In: Proceedings of the 7th ACM International Conference on Ubiquitous Computing (Ubicomp 2005). (2005)

[105] Varshavsky, A., de Lara, E., Hightower, J., LaMarca, A., Otsason, V.: Gsm indoor localization. Pervasive and Mobile Computing **3** (2007)

[106] Krumm, J., Cermak, G.: Rightspot: A novel sense of location for a smart personal object. In: Proceedings of the 5th ACM International Conference on Ubiquitous Computing (Ubicomp 2003). (2003)

[107] Youssef, A., Krumm, J., Miller, E., Cermak, G., Horvitz, E.: Computing location from ambient fm radio signals. In: Proceedings of the IEEE Wireless Communications and Networking Conference. (2005)

[108] E.P. Stuntebeck, S.P., Robertson, T., Reynolds, M., Abowd, G.: Wideband powerline positioning for indoor localization. In: Proceedings of the 10th ACM International Conference on Ubiquitous Computing (Ubicomp 2008). (2008)

[109] Jiang, Y., Pan, X., Li, K., Lv, Y., Dick, R.P., Hannigan, M., Shang, L.: Ariel: Automatic wi-fi based room fingerprinting for indoor localization. In: Proceedings of the 14th ACM International Conference on Ubiquitous Computing (Ubicomp 2012). (2012)

[110] Pulkkinen, T., Nurmi, P.: Awesom: Automatic discrete partitioning of indoor spaces for wifi fingerprinting. In: Proceedings of the 10th International Conference on Pervasive Computing. (2012)

[111] Schougaard, K.R., Gronbaek, K., Scharling, T.: Indoor pedestrian navigation based on hybrid route planning and location modelling. In: Proceedings of the 10th International Conference on Pervasive Computing (Pervasive2012). (2012)

[112] Wang, H., Sen, S., Elgohary, A., Farid, M., Youssef, M., Choudhury, R.R.: No need to war-drive – unsupervised indoor localization. In: Proceedings of the 10th International Conference on Mobile Systems, Applications and Services (Mobisys2012). (2012)

[113] Chen, L., Nugent, C., Wang, H.: A knowledge-driven approach to activity recognition in smart homes. Knowledge and Data Engineering, IEEE Transactions on **24**(6) (2012) 961–974

[114] Anderson, I., Muller, H.: Context awareness via gsm signal strength fluctuation. In: 4th international conference on pervasive computing, late breaking results. (2006)

[115] Sohn, T., Varshavsky, A., LaMarca, A., Chen, M.Y., Choudhury, T., Smith, I., Consolvo, S., Hightower, J., Grisworld, W.G., de Lara, E.: Mobility detection using everyday gsm traces. In: Proceedings of the 8th international conference on Ubiquitous computing. (2006)

[116] Sen, S., Radunovic, B., Choudhury, R.R., Minka, T.: You are facing the mona lisa – spot localization using phy layer information. In: Proceedings of the 10th International Conference on Mobile Systems, Applications and Services (Mobisys2012). (2012)

[117] Kosba, A., Youssef, M.: Rasid demo: A robust wlan device-free passive motion detection system. In: 2012 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops). (2012) 531–533

[118] Wilson, J., Patwari, N.: See-through walls: Motion tracking using variance-based radio tomography. IEEE Transactions on Mobile Computing **10**(5) (2011) 612–621

[119] Zhang, D., Ni, L.: Dynamic clustering for tracking multiple transceiver-free objects. In: Proceedings of the 7th IEEE International Conference on Pervasive Computing and Communications. (2009)

[120] Zhang, D., Liu, Y., Ni, L.: Rass: A real-time, accurate and scalable system for tracking transceiver-free objects. In: Proceedings of the 9th IEEE International Conference on Pervasive Computing and Communications (PerCom2011). (2011)

[121] Wilson, J., Patwari, N.: Radio tomographic imaging with wireless networks. IEEE Transactions on Mobile Computing **9** (2010) 621–632

[122] Lee, P.W.Q., Seah, W.K.G., Tan, H.P., Yao, Z.: Wireless sensing without sensors - an experimental study of motion/intrusion detection using rf irregularity. Measurement science and technology **21** (2010)

[123] Kosba, A.E., Saeed, A., Youssef, M.: Rasid: A robust wlan device-free passive motion detection system. In: IEEE International Conference on Pervasive Computing and Communications (PerCom). (2012)

[124] Patwari, N., Wilson, J., Ananthanarayanan, S., Kasera, S.K., Westenskow, D.: Monitoring breathing via signal strength in wireless networks (2011) submitted to IEEE Transactions on Mobile Computing, 18 Sept., 2011, available: arXiv:1109.3898v1.

[125] Zhang, D., Liu, Y., Guo, X., Gao, M., Ni, L.M.: On distinguishing the multiple radio paths in rss-based ranging. In: Proceedings of the 31st IEEE International Conference on Computer Communications. (2012)

[126] Patwari, N., Wilson, J.: Spatial models for human motion-induced signal strength variance on static links. IEEE Transactions on Information Forensics and Security **6**(3) (2011) 791–802

[127] Sigg, S., Beigl, M., Banitalebi, B.: 5.4. Organic Computing - A Paradigm Shift for Complex Systems, Autonomic Systems Series. In: Efficient adaptive communication from multiple resource restricted transmitters. Springer (2011)

[128] Song, W.Z., Huang, R., Xu, M., Shirazi, B., LaHusen, R.: Design and deployment of sensor network for real-time high-fidelity volcano monitoring. IEEE Transactions on Parallel and Distributed Systems **21**(11) (2010) 1658–1674

[129] Palmer, N., Kemp, R., Kielmann, T., Bal, H.: The case for smartphones as an urgent computing client platform. In: Proceedings of the International Conference on Computational Science, ICCS 2012. Volume 9. (2012) 1667–1676

[130] Ramirez, L., Dyrks, T., Gerwinski, J., Betz, M., Scholz, M., Wulf, V.: Landmarke: an ad hoc deployable ubicomp infrastructure to support indoor navigation of firefighters. Personal and Ubiquitous Computing **16**(8) (2012) 1025–1038

[131] for the Prevention of Accidents (RoSPA), T.R.S.: Home and leisure accident surveillance system. http://www.hassandlass.org.uk/query/index.htm (2002)

[132] Noury, N., Fleury, A., Rumeau, P., Bourke, A., Laighin, G., Rialle, V., Lundy, J.: Fall detection - principles and methods. In: 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. (2007) 1663–1666

[133] Cucchiara, R., Prati, A., Vezzani, R.: A multi-camera vision system for fall detection and alarm generation. Expert Systems **24**(5) (2007) 334–345

[134] Lin, C.W., Ling, Z.H.: Automatic fall incident detection in compressed video for intelligent homecare. In: Proceedings of 16th International Conference on Computer Communications and Networks (ICCCN). (2007) 1172–1177

[135] Noury, N., Poujaud, J., Fleury, A., Nocua, R., Haddidi, T., Rumeau, P.: Smart sweet home... a pervasive environment for sensing our daily activity? In Chen, L., Nugent, C.D., Biswas, J., Hoey, J., Khalil, I., eds.: Activity Recognition in Pervasive Intelligent Environments. Volume 4 of Atlantis Ambient and Pervasive Intelligence. Atlantis Press (2011) 187–208

[136] Vacher, M., Istrate, D., Portet, F., Joubert, T., Chevalier, T., Smidtas, S., Meillon, B., Lecouteux, B., Sehili, M., Chahuara, P., Meniard, S.: The sweet-home project: Audio technology in smart homes to improve well-being and reliance. In: Proceedings of the 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2011). (2011)

[137] Lukowicz, P., Pentland, S., Ferscha, A.: From context awareness to socially aware computing. Pervasive Computing, IEEE **11**(1) (2012) 32 –41

[138] Kononenko, I., Bratko, I.: Information-based evaluation criterion for classifier's performance. Machine Learning **6** (1991) 67–80

[139] Xu, Y., Stojanovic, N., Stojanovic, L., Schuchert, T.: Efficient human attention detection based on intelligent complex event processing. In: Proceedings of the 6th ACM International Conference on Distributed Event-Based Systems. DEBS '12 (2012) 379–380

[140] Wu, F., Hubermann, B.: Novelty and collective attention. In: Proceedings of the National Academics of Sciences. Volume 104. (2007) 17599–17601

[141] Yonezawa, T., Yamazoe, H., Utsumi, A., Abe, S.: Gaze-communicative behavior of stuffed-toy robot with joint attention and eye contact based on ambient gaze-tracking. In: ICMI. (2007) 140–145

[142] Wickens, C., McCarley, J.: Applied attention theory. CRC Press (2008)

[143] Ferscha, A., Zia, K., Gollan, B.: Collective attention through public displays. In: 2012 IEEE Sixth International Conference on Self-Adaptive and Self-Organizing Systems (SASO). (2012) 211 –216

[144] Popleteev, A., Osmani, V., Mayora, O.: Investigation of indoor localization with ambient FM radio stations. In: Pervasive Computing and Communications (PerCom), 2012 IEEE International Conference on. (2012) 171 –179

[145] Sigg, S., Shi, S., Buesching, F., Ji, Y., Wolf, L.: Leveraging rf-channel fluctuation for activity recognition. In: Proceedings of the 11th International Conference on Advances in Mobile Computing and Multimedia (MoMM2013). (2013)

[146] Tan, D., Sun, H., Lu, Y., Lesturgie, M., Chan, H.: Passive radar using global system for mobile communication signal: theory, implementation and measurements. IEE Proceedings - Radar, Sonar and Navigation **152**(3) (2005) 116–123

[147] Nishi, M., Takahashi, S., Yoshida, T.: Indoor human detection systems using vhf-fm and uhf-tv broadcasting waves. In: Personal, Indoor and Mobile Radio Communications, 2006 IEEE 17th International Symposium on. (2006) 1–5

[148] Yang, D., Xue, G., Fang, X., Tang, J.: Crowdsourcing to smartphones: incentive mechanism design for mobile phone sensing. In: Proceedings of the 18th annual international conference on Mobile computing and networking. Mobicom '12 (2012) 173–184

[149] Lukowicz, P., Pentland, A., Ferscha, A.: From context awareness to socially aware computing. IEEE Pervasive Computing **11**(1) (2012) 32–41

[150] Aggarwal, J., Ryoo, M.: Human activity analysis: A review. ACM Computing Surveys **43**(3) (2011) 16:1–16:43

[151] Kunze, K., Lukowicz, P.: Symbolic object localization through active sampling of acceleration and sound signatures. In: Proceedings of the 9th International Conference on Ubiquitous Computing. (2007)

[152] Chaquet, J.M., Carmona, E.J., FernáNdez-Caballero, A.: A survey of video datasets for human action and activity recognition. Comput. Vis. Image Underst. **117**(6) (2013) 633–659

[153] Adib, F., Katabi, D.: See through walls with wi-fi. In: ACM SIGCOMM'13. (2013)

[154] Pu, Q., Gupta, S., Gollakota, S., Patel, S.: Whole-home gesture recognition using wireless signals. In: The 19th Annual International Conference on Mobile Computing and Networking (Mobicom'13). (2013)

[155] Popleteev, A.: Device-free indoor localization using ambient radio systems. In: Adjunct Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2013). UbiComp '13 (2013)

[156] Wagner, B., Timmermann, D.: Adaptive clustering for device-free user positioning utilizing passive rfid. In: Adjunct Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2013). UbiComp '13 (2013)

[157] Colone, F., Falcone, P., Bongioanni, C., Lombardo, P.: Wifi-based passive bistatic radar: Data processing schemes and experimental results. IEEE Transactions on Aerospace and Electronic Systems **48**(2) (2012) 1061–1079

[158] Kassem, N., Kosba, A., Youssef, M.: Rf-based vehicle detection and speed estimation. In: 75th IEEE Vehicular Technology Conference (VTC Spring). (2012) 1–5

[159] Sigg, S., Shi, S., Ji, Y.: Rf-based device-free recognition of simultaneously conducted activities. In: Adjunct Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2013). UbiComp '13 (2013)

[160] Kim, Y., Ling, H.: Human activity classification based on micro-doppler signatures using a support vector machine. IEEE Transactions on Geoscience and Remote Sensing **47**(5) (2009) 1328–1337

[161] Ding, Y., Banitalebi, B., Miyaki, T., Beigl, M.: Rftraffic: Passive traffic awareness based on emitted rf noise from the vehicles. In: ITS Telecommunications (ITST), 2011 11th International Conference on. (2011) 393 –398

[162] Xu, C., Firner, B., Moore, R.S., Zhang, Y., Trappe, W., Howard, R., Zhang, F., An, N.: Scpl: Indoor device-free multi-subject counting and localization using radio signal strength. In: The 12th ACM/IEEE Conference on Information Processing in Sensor Networks (ACM/IEEE IPSN). (2013)

[163] Scholz, M., Riedel, T., Hock, M., Beigl, M.: Device-free and device-bound activity recognition using radio signal strength. In: Proceedings of the 4th Augmented Human International Conference in cooperation with ACM SIGCHI. (2013)

[164] Ildis, O., Ofir, Y., Feinstein, R.: Wardriving from your pocket (2013)

[165] Kononenko, I., Bratko, I.: Information-based evaluation criterion for classifier's performance. Machine Learning **6**(1) (1991) 67–80

[166] Spackman, K.A.: Signal detection theory: valuable tools for evaluating inductive learning. In: 6th international workshop on Machine learning. (1989) 160–163

[167] Dupuy, C., Torre, A. In: Local Clusters, trust, confidence and proximity, Clusters and Globalisation: The development of urban and regional economies. (2006) pp. 175–195

[168] Mayrhofer, R., Gellersen, H.: Spontaneous mobile device authentication based on sensor data. information security technical report **13**(3) (2008) 136–150

[169] Bichler, D., Stromberg, G., Huemer, M., Loew, M.: Key generation based on acceleration data of shaking processes. In Krumm, J., ed.: Proceedings of the 9th International Conference on Ubiquitous Computing. (2007)

[170] Holmquist, L.E., Mattern, F., Schiele, B., , Alahuhta, P., Beigl, M., Gellersen, H.W.: 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)

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

[172] Gellersen, H.W., Kortuem, G., Schmidt, A., Beigl, M.: Physical prototyping with smart-its. IEEE Pervasive computing **4**(1536-1268) (2004) 10–18

[173] Mayrhofer, R., Gellersen, H.: Shake well before use: Authentication based on accelerometer data. Pervasive Computing (2007) 144–161

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

[175] Bichler, D., Stromberg, G., Huemer, M.: Innovative key generation approach to encrypt wireless communication in personal area networks. In: Proceedings of the 50th International Global Communications Conference. (2007)

[176] Hershey, J., Hassan, A., Yarlagadda, R.: Unconventional cryptographic keying variable management. IEEE Transactions on Communications **43** (1995) 3–6

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

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

[179] Hamida, S.T.B., Pierrot, J.B., Castelluccia, C.: 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)

[180] Tuyls, P., Skoric, B., Kevenaar, T.: Security with Noisy Data. Springer-Verlag (2007)

[181] Li, Q., Chang, E.C.: Robust, short and sensitive authentication tags using secure sketch. In: Proceedings of the 8th workshop on Multimedia and security, ACM (2006) 56–61

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

[183] Miao, F., Jiang, L., Li, Y., Zhang, Y.T.: 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) 2458–2461

[184] Juels, A., Sudan, M.: A Fuzzy Vault Scheme. Proceedings of IEEE Internation Symposium on Information Theory (2002) 408

[185] Dodis, Y., Katz, J., Reyzin, L., Smith, A.: Robust fuzzy extractors and authenticated key agreement from close secrets. Advances in Cryptology-CRYPTO 2006 (2006) 232–250

[186] Cano, P., Batlle, E., Kalker, T., Haitsma, J.: A Review of Algorithms for Audio Fingerprinting. The Journal of VLSI Signal Processing **41**(3) (2005) 271–284

[187] Baluja, S., Covell, M.: Waveprint: Efficient wavelet-based audio fingerprinting. Pattern Recognition **41**(11) (2008)

[188] Ghouti, L., Bouridane, A.: A robust perceptual audio hashing using balanced multiwavelets. In: Proceedings of the 5th IEEE International Conference on Acoustics, Speech, and Signal Processing. (2006)

[189] Sukittanon, S., Atlas, L.: Modulation frequency features for audio fingerprinting. In: Proceedings of the 2nd IEEE International Conference on Acoustics, Speech, and Signal Processing. (2002)

[190] Haitsma, J., Kalker, T.: A Highly Robust Audio Fingerprinting System. Journal of New Music Research **32**(2) (2003) 211–221

[191] Burges, C., Plastina, D., Platt, J., Renshaw, E., Malvar, H.: Using audio fingerprinting for duplicate detection and thumbnail generation. In: Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on. Volume 3. (2005) iii/9–iii12 Vol. 3

[192] Bellettini, C., Mazzini, G.: A framework for robust audio fingerprinting. Journal of Communications **5**(5) (2010)

[193] Ghias, A., Logan, J., Chamberlin, D., Smith, B.C.: Query by humming. In: Proceedings of the ACM Multimedia. (1995)

[194] Parsons, D.: The directory of tunes and musical themes. Cambridge University press (1975)

[195] Baeza-Yates, R.A., perleberg perleberg perleberg perleberg, C.H.: Fast and practical approximate string matching. Third annual symposium on combinatorial pattern matching (1992)

[196] McNab, R.J., Smith, L.A., Witten, I.H., Henderson, C.L., Cunningham, S.J.: Towards the digital music library: tune retrieval from acoustic iinput. Proceedings of the ACM (1996)

[197] Prechelt, L., Typke, R.: An interface for melody input. ACM Transactions on Computer Human Interactions **8** (2001)

[198] Chai, W., Vercoe, B.: Melody retrieval on the web. In: Proceedings of the ACM/SPIE conference on Multimedia Computing and Networking. (2002)

[199] Shifrin, L., Pardo, B., Birmingham, W.: Hmm-based musical query retrieval. In: Proceedings of the joint conference on digital libraries. (2002)

[200] Rabiner, L.: A tutorial on hidden markov models and selected accplications in speech recognition. Proceedings of the IEEE **77**(2) (1989)

[201] Zhu, Y., Shasha, D.: Warping indexes with envelope transforms for query by humming. In: Proceedings of the ACM SIGMOD International Conference on Management of Data. (2003)

[202] Haitsma, J., Kalker, T.: Robust audio hashing for content identification. In: In Content-Based Multimedia Indexing (CBMI. (2001)

[203] Lebossé, J., Brun, L., Pailles, J.C.: A robust audio fingerprint's based identification method. In: Proceedings of the 3rd Iberian conference on Pattern Recognition and Image Analysis, Part I. IbPRIA '07, Berlin, Heidelberg, Springer-Verlag (2007) 185–192

[204] Burges, C., Platt, J., Jana, S.: Distortion discriminant analysis for audio fingerprinting. Speech and Audio Processing, IEEE Transactions on **11**(3) (2003) 165–174

[205] Herre, J., Allamanche, E., Hellmuth, O.: Robust matching of audio signals using spectral flatness features. In: Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the. (2001) 127–130

[206] Yang, C.: Macs: music audio characteristic sequence indexing for similarity retrieval. In: Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the. (2001) 123–126

[207] Yang, C.: Efficient acoustic index for music retrieval with various degrees of similarity. In: Proceedings of the tenth ACM international conference on Multimedia. MULTIMEDIA '02, New York, NY, USA, ACM (2002) 584–591

[208] Wang, A.: The shazam music recognition service. Communications of the ACM **49**(8) (2006)

[209] Wang, A.: An industrial strength audio search algorithm. In: Proceedings of the International Conference on Music Information Retrieval (ISMIR). (2003)

[210] Hao, F.: On using fuzzy data in security mechanisms. PhD thesis, Queens College, Cambridge (2007)

[211] Ibarrola, A.C., Chavez, E.: A robust entropy-based audio-fingerprint. In: Proceedings of the 2006 International Conference on Multimedia and Expo (ICME 2006). (2006)

[212] Schneider, B.: Applied Cryptography: Protocols, Algorithms, and Source Code in C. 2 edn. John Wiley and Sons, Inc. (1996)

[213] Reed, I., Solomon, G.: Polynomial codes over certain finite fields. Journal of the Society for Industrial and Applied Mathematics (1960) 300–304

[214] Juels, A., Wattenberg, M.: A Fuzzy Commitment Scheme. Sixth ACM Conference on Computer and Communications Security (1999) 28–36

[215] National Institute of Standards and Technology: 180-3, Secure Hash Standard (SHS). Federal Information Processing Standards Publications (FIPS PUBS) (2008)

[216] National Institute of Standards and Technology: 197, Advanced Encryption Standard (AES). Federal Information Processing Standards Publications (FIPS PUBS) (2001)

[217] Mills, D., Martin, J., Burbank, J., Kasch, W.: Network Time Protocol Version 4: Protocol and Algorithms Specification. RFC 5905 (Proposed Standard) (2010)

[218] Mills, D.L.: Improved algorithms for synchronising computer network clocks. IEEE/ACM Transactions on Networking **3**(3) (1995)

[219] Meier, S., Weibel, H., Weber, K.: 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)

[220] Mills, D.L.: Precision synchronisation of computer network clocks. ACM Computer Communication Review **24**(2) (1994)

[221] Freedesktop.org: GStreamer Documentation. (2010)

[222] Scheirer, W.J., Boult, T.E.: Cracking fuzzy vaults and biometric encryption. In: Proceedings of Biometrics Symposium, Baltimore, USA. (2007)

[223] Ignatenko, T., Willems, F.M.J.: Information Leakage in Fuzzy Commitment Schemes. In: IEEE Transactions on Information Forensics and Security. Volume 5. (2010) 337

[224] Fenzi, K., Wreski, D.: Linux Security HOWTO. (2004)

[225] Brown, R.G.: Dieharder: A random number test suite. http://www.phy.duke.edu/~rgb/General/dieharder.php (2011)

[226] Kuiper, N.: Tests concerning random points on a circle. In: Proceedings of the Koinklijke Nederlandse Akademie van Wetenschappen. Volume Series a 63. (1962) 38–47

[227] Stephens, M.: The goodness-of-fit statistic v_n: Distribution and significance points. Biometrika **52** (1965)

[228] Baluja, S., Covell, M.: Content fingerprinting using wavelets. In: Proceedings of the Conference of Visual Media Production, London, UK (2006)

[229] Cano, P., Batlle, E., Kalker, T., Haitsma, J.: A review of audio fingerprinting. Journal of VLSI Signal Processing Systems **41 Issue 3** (2005)

[230] Chandrasekhar, V., Sharifi, M., Ross, D.: Survey and evaluation of audio fingerprinting schemes for mobile audio search. In: International Symposium on Music and Information Retrieval (ISMIR), Miami, Florida (2011)

[231] McCune, J.M., Perrig, A., Reiter, M.K.: Seeing-is-believing: Using camera phones for human-verifiable authentication. In: Proceedings of the 2005 IEEE Symposium on Security and Privacy. (2005)

[232] Goodrich, M.T., Sirivianos, M., Solis, J., Tsudik, G., Uzun, E.: Loud and clear: Human-verifiable authentication based on audio. In: Proceedings of the 26th IEEE International Conference on Distributed Computing Systems. (2006)

[233] Sigg, S., Schuermann, D., Ji, Y.: 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)

[234] Smith, T.F., Waterman, M.S.: Identification of common molecular subsequences. Journal of molecular biology **147**(1) (1981) 195–197

[235] Mudumbai, R., Brown, D.R., Madhow, U., Poor, H.V.: Distributed transmit beamforming: Challenges and recent progress. IEEE Communications Magazine (2009) 102–110

[236] Sigg, S., Masri, R., Ristau, J., Beigl, M.: Limitations, performance and instrumentation of closed-loop feedback based distributed adaptive transmit beamforming in wsns. In: Fifth International Conference on Intelligent Sensors, Sensor Networks and Information Processing - Symposium on Theoretical and Practical Aspects of Large-scale Wireless Sensor Networks. (2009)

[237] Shi, S., Sigg, S., Ji, Y.: Joint localisation and activity recognition from ambient fm broadcast signals. In: Adjunct Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2013). UbiComp '13 (2013)

[238] Hong, J., Ohtsuki, T.: Ambient intelligence sensing using array sensor: Device-free radio based approach. In: Adjunct Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2013). UbiComp '13 (2013)

[239] Wickens, C.: Processing resources in attention. Academic Press (1984)

[240] Gollan, B., Wally, B., Ferscha, A.: Automatic attention estimation in an interactive system based on behaviour analysis. In: Proceedings of the 15th Portuguese Conference on Artificial Intelligence (EPIA2011). (2011)

[241] : Bringing gesture recognition to all devices. In: Presented as part of the 11th USENIX Symposium on Networked Systems Design and Implementation, Berkeley, CA, USENIX (2014)

[242] Adib, F., Kabelac, Z., Katabi, D., Miller, R.C.: 3d tracking via body radio reflections. In: Usenix NSDI. Volume 14. (2014)

[243] Raja, M., Exler, A., Hemminki, S., Konomi, S., Sigg, S., Inoue, S.: Towards pervasive geospatial affect perception. Springer GeoInformatica (2017)

[244] Raja, M., Sigg, S.: Applicability of rf-based methods for emotion recognition: A survey. In: 2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops). (2016) 1–6

[245] Schürmann, D., Brüsch, A., Sigg, S., Wolf, L.: Bandana—body area network device-to-device authentication using natural gait. In: Pervasive Computing and Communications (PerCom), 2017 IEEE International Conference on, IEEE (2017) 190–196