

RESEARCH

Open Access

Heterogeneous spectrum sensing: challenges and methodologies

Wei Liu^{1*}, Mikolaj Chwalisz², Carolina Fortuna¹, Eli De Poorter¹, Jan Hauer², Daan Pareit¹, Lieven Hollevoet³ and Ingrid Moerman¹

Abstract

Distributed sensing is commonly used to obtain accurate spectral information over a large area. More and more heterogeneous devices are being incorporated in distributed sensing with the aim of obtaining more flexible sensing performance at lower cost. Although the concept of combining the strengths of various sensing devices is promising, the question of how to compare and combine the heterogeneous sensing results in a meaningful way is still open. To this end, this paper proposes a set of methodologies that are derived from several spectrum sensing experiments using heterogeneous sensing solutions. Each of the solutions offers different radio frequency front-end flexibility, sensing speed and accuracy and varies in the way the samples are processed and stored. The proposed methodologies cover four fundamental aspects in heterogeneous sensing: (i) storing experiment descriptions and heterogeneous results in a common data format; (ii) coping with different measurement resolutions (in time or frequency domain); (iii) calibrating devices under strictly controlled conditions and (iv) processing techniques to efficiently analyse the obtained results. We believe that this paper provides an important first step towards a standardized and systematic approach of heterogeneous sensing solutions.

Keywords: Cognitive radio; Heterogeneous spectrum sensing; Methodology; Experimentation

1 Introduction

It has been shown that a licensed spectrum is often underused in time and space. As a concrete example, a 6-day measurement campaign in multiple cities of Europe indicates that the UHF TV bands are less than 62.2% utilized [1]. In order to improve spectrum usage, the research community proposed the concept of cognitive radio (CR) [2,3], which offers a mechanism for unlicensed users (i.e. secondary users) to use the licensed spectrum in the absence of the primary users. There are two fundamentally opposite approaches to determine whether the primary user is active. In the first approach, the secondary user senses the spectrum [4] to decide whether it is available or not; while in the second approach, a geo-location database built upon information of primary transmitters and radio propagation models is accessed in order to locate spectrum opportunities [5]. The first approach tends to be less accurate while the second has difficulties

adapting to dynamic scenarios. More recently, also hybrid geo-location databases have been considered to improve accuracy [6].

While initially the Federal Communications Commission (FCC) only allowed opportunistic access to licensed spectrum via the geo-location database approach, in a recent revision, requirements for devices relying solely on spectrum sensing have been included in the FCC regulation [7]. The new regulation states that sensing devices must demonstrate a very high degree of confidence to avoid interfering with the incumbents and they must meet the minimum sensitivity requirements for several types of primary signal (i.e. -114 dBm for analog and digital TV signal and -107 dBm for low power auxiliary).

Currently, mainly due to cost considerations, these constraints are hard to satisfy for commercial devices. Fortunately, the lack of individual sensitivity can be compensated by using cooperative sensing with multiple sensing devices [8]. These sensing devices can be installed as part of the infrastructure for the sole purpose of spectrum sensing [9], or existing user devices can be involved

*Correspondence: wei.liu@intec.ugent.be

¹University Ghent - iMinds, Gaston Crommenlaan 8, 9050 Ghent, Belgium
Full list of author information is available at the end of the article

through a crowdsourcing approach [10]. In both cases, the devices involved in the collaborative spectrum sensing are likely to be heterogeneous, thus performing *heterogeneous spectrum sensing*. Unlike spectrum sensing using homogeneous devices, heterogeneous sensing is more challenging due to the fact that it involves high- to low-end sensing devices and their corresponding approaches to process or store the results. More particularly, heterogeneous spectrum sensing experiments are faced by the following challenges:

1.1 Storage format

Spectrum sensing devices may record data in different formats. Each type of device has its own way of logging data, which could be plain text, or different types of binary formats. Apart from the format of actual measurements, the lack of common configuration and well-defined meta information makes it hard to initialize the measurements and interpret the results. Therefore, we need a uniform and well-structured storage mechanism for heterogeneous sensing devices.

1.2 Measurement resolution

In addition, measurements from heterogeneous devices may have different resolutions in time and frequency domains. Consider the following example: two devices (A and B) are both monitoring the same range of spectrum; device A measures the power spectrum with a 1-MHz resolution bandwidth updating each second, while device B measures with a resolution bandwidth of 10 KHz at the rate of two times per second. Obviously, it is not possible to directly compare or combine the results from A and B. When dealing with the output from heterogeneous devices, the situation in the above example often occurs. Thus, we need a method to obtain a common resolution before a meaningful comparison or combination of the heterogeneous data sources can be made.

1.3 Calibration

Some devices, particularly the low-end ones, may be uncalibrated. Simple wireless devices typically provide the spectrum information using the existing channel assessment module for the purpose of wireless medium access control (MAC). For MAC purposes, the output of the channel assessment module is used within the device, thus it only needs to be relatively correct, rather than absolutely. However, for the purpose of heterogeneous spectrum sensing, it is important that sensing results are produced against a common reference; hence, there is a need for calibration. Though most high-end devices are calibrated individually by their manufactures, a uniform calibration process is still desired in case there are noticeable differences in the factory calibration process or even individual differences among the same type of devices. For

instance, prior to experiments, the authors in [11] calibrate the internal noise level of multiple radio receivers of the same type in a shielded environment. However, this approach cannot be used to calibrate heterogeneous devices, as internal noise level is one of the primary heterogeneities among devices. In conclusion, there is a need for a uniform calibration mechanism for heterogeneous spectrum sensing.

1.4 Processing methods

Due to the use of different devices and complex experiment scenarios (i.e. distributed measurements, multiple iterations), heterogeneous spectrum sensing usually generates a large amount of data. Efficient processing methods are crucial to reach an objective conclusion in a reasonable amount of time. Although for certain performance metrics, well-accepted processing approaches exist; most of the time, it is not feasible to pursue a uniform processing mechanism for all experiments. A more pragmatic way is to simply make use of existing methods when applicable. Thus, we believe there is a need to share experiences related to processing heterogeneous data.

In this paper, we identify the challenges and discuss the methodologies of heterogeneous spectrum sensing regarding the following aspects: storage format, measurement resolution, calibration and processing methods. First, we propose a common data format for uniformly defining experiments and storing the results. Then we provide a set of methodologies regarding measurement resolution, device calibration and data processing, which are implemented, validated and evaluated on reference scenarios. Thus, the main contributions of this paper are the common data format, the methodologies, the validation and the evaluation.

The remainder of this paper is structured as follows: first, the related work is discussed in Section 2; after that, Section 3 describes the proposed methodologies regarding the aforementioned four challenges, while the performance of these methodologies are verified with concrete implementations and real-life experiments in Section 4. Finally, we conclude this paper in Section 5.

2 Related work

This section gives an overview of related work for heterogeneous spectrum sensing. The following aspects are discussed: (i) the storage format, (ii) the measurement resolution, (iii) calibration methods and (iv) processing mechanisms.

2.1 Storage format

The IEEE 1900.6 standard [12] defines spectrum sensing-related parameters and data structures, and as such may be considered as a guideline of data storage for spectrum sensing. Since this is also the main interest of our work,

the terminologies available in this standard have been considered and used whenever applicable. In comparison, the IETF PAWS standard [13] is limited to the communication among TV white space devices and databases, thus it is less relevant to the low level challenges in heterogeneous spectrum sensing.

Apart from existing standards, authors in [14,15] use heterogeneous spectrum sensing to construct a radio environment map (REM), which includes a dedicated spectrum data server (SDS) to collect heterogeneous sensing data. After the data collection phase, the server fusion (SF) interface is used to communicate with other processing units. It is seen that the architecture of splitting storage and processing units performs well in real experiments and can be easily adapted as part of existing LTE network [16]. Compared to [14,15], this work has a more general perspective for sensing experimentation. We aim to provide tools for common data storage that focuses more on low level spectrum data and not specifically tailored to any application purposes.

2.2 Measurement resolution

Measurement resolution is an important indication of the sensing capabilities of a device. Advanced sensing devices usually have more flexibilities for configuring which resolution to use for the measurements.

A straightforward way to achieve a common resolution is to use common configurations at the measurement time (e.g. choose a resolution bandwidth that is available on all devices) [17]. When no common setting is available among the considered devices, the authors in [17] determine the settings for each device by performance (e.g. choose a sweeping pattern or detector type after a number of initial trial measurements). While this approach is practical and reasonable, it is not clear how experimenters can achieve data with common resolutions for further processing when no common settings could be found.

The authors of [18] use experimental measurements to illustrate how the choices of measurement resolutions can influence the detection performance. However, the methodologies discussed in [18] do not take heterogeneous devices into account.

Therefore, this paper proposes a post-processing approach to derive data with common resolutions regardless of the settings chosen at the measurement time.

2.3 Calibration

As described in the previous section, the internal noise level used in [11] is not an ideal metric for calibrating heterogeneous sensing devices, because it is expected to be different. The authors of [8] propose to use satellite band signals for cooperative sensing devices to identify shadow effects. Since satellite signals have relatively constant strength over a wide area, devices that receive weaker

satellite signals have a higher probability to be shadowed by obstacles. Though this solution is not directly related to calibration, the idea of using satellite band signals as reference could be used to calibrate devices in outdoor measurements. In general, for the calibration of heterogeneous devices, a known and constant power level is needed as input reference [14,15]. This is also the main principle of calibration followed by our previous work [19]. Apart from a stable reference signal, calibration experiments need to be strictly controlled. In this paper, we share our experience of using coaxial-cable-based experiments to achieve accurate calibration solutions.

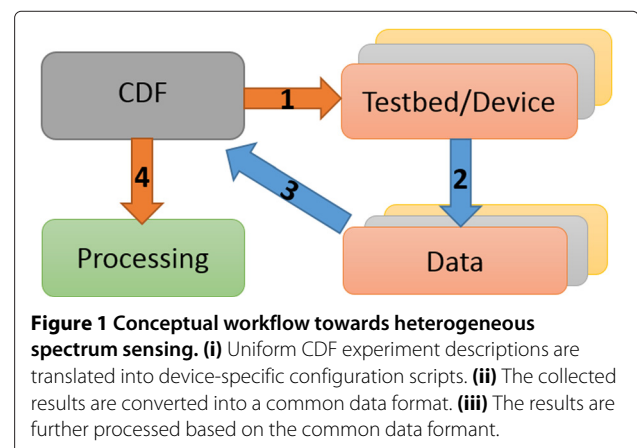
2.4 Processing methods

Unlike the previous challenges, the processing mechanisms are very experiment specific. For energy detection, the magnitude of the complex samples is calculated to represent the received signal strength. Alternatively, the magnitude of the output from the fast Fourier transform (FFT) could be used instead of time domain samples [14,15]. For the purpose of detection performance, a receiver operating characteristic (ROC) plot is often used to observe the device's sensitivity under various input conditions. The work in [14,15] goes one step further by using the inverse distance weighting (IDT) technique to interpolate discrete energy measurements into a REM. Though it is not feasible to have a uniform processing approach, we believe there is a need to share processing experience in order to improve the efficiency.

3 Methodologies for realizing heterogeneous sensing

This section describes the methodologies for realizing heterogeneous sensing regarding the previously identified challenges.

Conceptually, we propose the following workflow for heterogeneous spectrum sensing (see Figure 1). The ini-



tial phase for performing sensing experiments consists of configuring the heterogeneous devices, sending them the instructions for starting the sensing and collecting the data. This involves creating a series of device-specific scripts. As one of the contributions of this paper, we propose a uniform way of providing the configuration to the devices and storing the data from the devices. We use a common data format (CDF) for experiment description and data storage that is device independent and machine readable. Spectrum sensing descriptions and settings are defined using this common data format as depicted in Figure 1. In the first step, these uniform descriptions are then converted into device (or testbed/infrastructure)-specific configurations and control scripts. In the second step, the results of the experiments are transformed into a common representation format. As the third and final step, the resulting uniformly described data is further processed by a set of tools to align the resolution, achieve calibration and compute spectrum occupancy-related metrics.

It should be noted that, the information needed in the calibration phase comes from separate experiments, which will be denoted as *calibration experiments* in the remaining part of the paper. Calibration experiments are not necessarily part of every experiment iteration, but it should be at least performed once before the real sensing measurements start.

3.1 Common data format

The proposed common data format has been developed to ease spectrum sensing experimentation across devices and testbeds and contains three main parts. The first part

refers to the description of the experiment abstract, the second part refers to the spectrum sensing experiment, thus the so-called meta-data. The third part focuses on the actual traces resulted from the experiment.

The experiment description provides a detailed description of the experiment, such as how it was performed and what kind of data was collected. From the top level, the description contains the following fields: *experiment abstract*, *meta-information* and *experiment iteration(s)*. Below each field (except for experiment abstract), some sub fields are defined, as shown in Figure 2.

3.1.1 Experiment abstract

Experiment abstract is a high level description of the experiment, providing a basic idea of the experiment motivation, as well as the expected output. It is possible to relate to other experiments by adding relevant information. For instance, when experiment B is a scaled extension of experiment A, the following sentence ‘repetition of experiment A on a larger scale’ can be noted in the abstract of experiment B. In addition, we provide means to link to related documentations, such as publications that are based on a given experiment.

3.1.2 Meta-information

Meta-information is the information required for describing, understanding and evaluating the experiment. All experimental details except the data itself should be described in this field. The most important items are the description of involved devices, physical setup of the experiment, the selected signal type and frequency, as well as the description of the measurements.

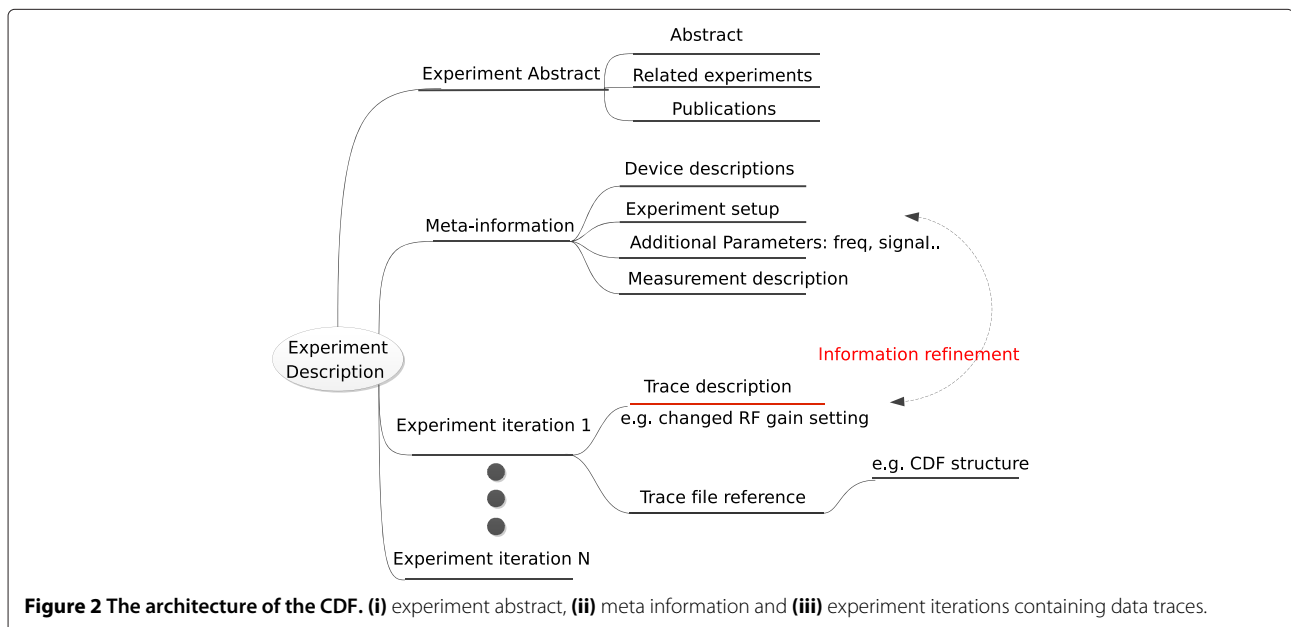


Figure 2 The architecture of the CDF. (i) experiment abstract, (ii) meta information and (iii) experiment iterations containing data traces.

The description of the involved devices is critical to reproduce the experiment. It should not only be limited to textual description but also provide references to the relevant data sheets. Moreover, we recommend to include information of related software and, if necessary, the operating system. The bottom line is that the collected information must suffice to repeat the experiment from scratch, starting from finding the same devices to setting up the identical software environment.

The physical experiment setup mainly refers to the description of how devices are positioned and connected. Ideally, there should be a location map to indicate the topology of the devices. Wireless experiments are sensitive to environmental factors, such as if an experiment is conducted indoor or outdoor, or if an experiment is conducted under a static or rather dynamic environment. Thus, we recommend to document this information in the meta-information as well.

Furthermore, the operating frequency and the characteristics of the used signals are noted as additional parameters. This creates a convenient way of indexing the existing experiments, e.g. one can easily find all sensing experiments in the TV white space. Thus, it allows experimenters to reuse past experiments more efficiently.

Finally, the measurement description contains a common description of the recorded data of all the *experiment iterations*, allowing experimenters to understand and process the data more smoothly. It specifies the configuration used by each device (e.g. gain settings, sample frequency) and the collected data types (e.g. frequency, signal power, time stamp). In addition, each data type is associated with a measurement unit (e.g. Hz, dBm, μ s). For more information related to defining measurement units, readers are referred to the IEEE 1900.6 [12] standard.

3.1.3 Experiment iteration(s)

Experiment iteration provides information that is related to the execution of a particular experiment round. There are two sub fields in each experiment iteration: the trace description and the trace file reference. The trace description is similar to the description in the *meta-information* but may extend or refine the *meta-information* partially if necessary, as shown by the red line in Figure 2. For instance, if a set of measurements is used to compare the influence of different radio frequency (RF) front-end gain settings, trace description is an ideal place to indicate what gain setting is used in each experiment iteration. This way, different settings among experiment iterations can be highlighted without the need of describing the entire experiment setup over and again.

The trace file reference is a 'pointer' towards the measurement data, which indicates where the measurement trace is physically stored.

A reference implementation of the CDF architecture is presented in Section 4.1.

3.2 Measurement resolution

Typically, one spectrum sensing trace cannot be directly compared to another, due to the differences in frequency and/or time domain. To overcome the heterogeneous *frequency resolution*, the easiest and most straightforward approach is to integrate the power spectral density (PSD) in a certain frequency interval and use the integrated power as the metric for comparison. This also implies that the selected interval for integration needs to be wider than the largest resolution bandwidth among all the sensing solutions.

There are different approaches to overcome the differences in *time resolution*. The easiest way is to apply averaging on the traces obtained in the same time duration. Alternatively, instead of using averaging, one can apply max-hold filtering, so the combined trace contains every transient signal that ever appeared in the observation period. By using integration in the frequency domain and averaging or max-holding in the time domain, a common metric is derived from various raw spectra. This is referred to as the common metric in the remainder of the paper. We provide a reference implementation of this processing scheme in the CDF toolbox (`pw_integration` function).

3.3 Calibration

Calibration of heterogeneous devices essentially means comparing the received power of each device to its corresponding input signal strength. The calibration process consists of four steps. First, a set of reference signals has to be selected. Second, the path loss between the signal source and the devices under calibration must be strictly controlled. Third, a suitable metric for performing the calibration has to be identified and fourth, the offset between the reference signal and the signals received by the devices has to be computed.

For the first step, it is generally advisable to use a set of diversified input signals (i.e. different bandwidth and signal strength) so that the calibration experiment is general enough to cope with different types of input. Also the generated signal needs to be continuous so that the recorded signal has a constant amplitude. This ensures that the sensing performance in terms of timing does not affect the performance in terms of power accuracy. The produced signal strength needs to be tuned within the dynamic range of all devices. If the signal is too strong, it may saturate the device under calibration; on the other hand, when it is too weak,

the signal might be buried by noise. Both situations should be avoided. Ideally, a high-end signal generator should be used as the signal source to meet the above constraints.

For the second step, the most at-hand method is to use a coaxial cable for controlling the path loss between the signal source and the sensing devices. An alternative way would be to use an anechoic chamber where the path loss is not affected by the multi-path effect.

In the third step, the received signal strength needs to be calculated from the power spectral density, which comes down to performing integration over the interval where the signal is transmitted in the frequency domain. If the integration interval is not the same as the signal bandwidth, the obtained metric will rely partially on the device's noise floor instead of solely on the input signal, thus, not qualified for power calibration.

Finally, in the fourth step, the power offset is then computed according to Equation 1, where the transmit power is denoted as P_{tx} , the received power is denoted as P_{rx} , and the total attenuation caused by coaxial cables and splitters is denoted as P_{atten} :

$$P_{offset} = P_{tx} - P_{atten} - P_{rx} \quad (1)$$

In Equation 1, P_{offset} accounts for the combined heterogeneity of the RF front-end, analog-to-digital converter (ADC) and the processing unit. However, it does not include the influence of the antenna, as the antenna is replaced by the coaxial cable connections. For devices using different types of antenna, the power offset needs to be readjusted with the antenna gain.

If the relative position of the transmitter and receiver is known, the influence of the radiation pattern should also be taken into account. For omnidirectional antenna, the radiation pattern changes with the elevation angle between the transmitter and receiver; while for directional antenna, the radiation pattern varies with both horizontal and vertical angles [20]. When the relative position of transmitter and receiver is unknown, it is necessary to rotate the directional antenna several times to cover the 360° [18].

Sometimes, P_{offset} varies with the input signal strength and the settings of the sensing device (i.e. gain settings). For instance, it is mentioned in [15] that the RFX2400 daughter-board of USRP does not have a linear input and output (IO) relationship. In this case, more measurements need to be performed to cope with different input signal strength and sensing configurations.

3.4 Processing

Sensitivity and accuracy are two important metrics to compare spectrum sensing devices. For heterogeneous

sensitivity analysis, experimenters tend to form a mainstream processing style, which is discussed in the first part of this section. As for power accuracy, generally, a high-end device (i.e. spectrum analyser) is used as benchmark in various measurements. However, when it comes to large scale heterogeneous measurements, this approach becomes very tedious. Thus, there is a need to process data in a more elegant approach, which is what we discuss in the second part of this section.

3.4.1 Heterogeneous sensitivity analysis

The sensitivity of a sensing device is reflected by its noise floor. Unlike power accuracy, sensitivity cannot be evaluated by the common metric derived in Section 3.2. This is because the noise floor is affected by the resolution bandwidth, thus the integrated power metric will always be higher than the original noise floor.

The most straightforward way is to observe the mean and variance value of the spectrum trace when no signal is present. Alternatively, we can also use the receiver operating characteristic. The ROC is obtained by expressing the probability of detection (P_d) as a function of the probability of false alarm (P_f). Some papers utilize the probability of missed detection (P_m) which is simply given by $1 - P_d$.

Despite of the heterogeneity in power spectra, ROC can be obtained via a common approach:

- Record spectrum traces when no signal is present.
- Vary P_f from 0% to 100% in small steps and determine a detection threshold for each P_f based on the previously recorded trace.
- Apply a signal at the input of the sensing device and record spectrum trace again.
- Compute P_d or P_m for all the detection thresholds determined in the second step from the trace recorded in the previous step

The advantage of ROC analysis is that it is device independent, as for a given false alarm, each device can have its own threshold. The only constraint is that the detection threshold should be calculated in a uniform approach for all devices. This is why it is commonly applied in the heterogeneous sensitivity studies [14,15,19].

As for the method to obtain detection threshold, there are many optimized variants [21-23]. As an example, the constant false alarm (CFA) approach [11] is described in Equation 2, where σ_n denotes the variance of the noise samples, N denotes the number of spectrum samples, P_f denotes the target false alarm and λ denotes the calculated detection threshold, respectively.

$$\lambda = \sigma_n^2 \left(1 + \frac{Q^{-1}(P_f)}{\sqrt{N/2}} \right) \quad (2)$$

3.4.2 Heterogeneous accuracy analysis

As stated previously, distributed heterogeneous measurements usually generate a large amount of data, which needs more efficient processing mechanisms. When processing a large dataset, the basic approach is to look at how the data is distributed, which can be achieved by computation of several statistics (i.e. mean, variance). However, to gain more insights of the data (i.e. discover a common behaviour, or a group of data that displays similarity within the entire set), then more advanced techniques, such as correlation, various linear regression algorithms need to be involved. Essentially, we recommend to use the basic techniques of data mining for analysing large scale heterogeneous sensing experiments, among which four most relevant techniques are exemplified as follows:

- Dependency modelling - the establishment of relationships between variables. This could be that the detection probability depends on the target signal strength or the distance between the transmitter and the sensing device.
- Outlier detection - the identification of the unusual spectrum records, which could be caused by malfunctioning devices or other unknown interferences.
- Regression - is a statistical way to explore the relationship among variables which models the data with the least error.
- Clustering - is the task of discovering groups and structures in the data that are in one way or another 'similar.' In case of spectrum sensing, this could be that a group of sensing devices are shadowed by a common obstacle.

The outlier detection is a rather basic step, which can be achieved by many statistical tools or simply manual observations. The procedure of 'Clustering' and 'Regression' are addressed with a concrete experiment in Section 4.4. For the dependency modelling, we find that the path loss model (the relationship between received signal strength and distance) is generally applicable for dependency modelling in the case of distributed sensing measurements. More particularly, the well-known log-distance path loss model can be expressed by two parameters - the path loss coefficient exponent α and the path loss offset β :

$$PL(d) = 20 \times \alpha \times \log_{10}(d) + \beta, \quad (3)$$

where d is the distance between the transmitter and the receiver. When using the logarithmic distance as the argument, the equation reduces to a simple linear expression.

Hence, various approaches, such as least square regression, can be used to estimate α and β .

The role of the path loss model is essentially a way to extract new parameters out of the raw data sets. It is a tool to correlate data from distributed locations. Although deriving the path loss model is not always easy or feasible, the basic idea of correlating data to extract new parameters is generally applicable and highly valuable in our experience.

4 Reference implementation and experimentation

This section first presents the implementation of the common data format in Section 4.1 and then illustrates how to apply the methodologies defined in the previous section with real-life experiments. More specifically, Section 4.2 describes a calibration experiment, which uses the method proposed in Section 3.2 to overcome the different measurement resolutions and then obtains the power offset following the mechanisms proposed in Section 3.3. Sections 4.3 and 4.4 present two experiments that use processing techniques discussed in Section 3.4 to evaluate two fundamental sensing performance metrics (sensitivity and accuracy), respectively.

4.1 Common data format implementation

The reference implementation of the CDF architecture consists of three parts: the CDF experiment description, the CDF data structure for common storage and the CDF toolbox for additional functionalities such as conversion between formats and result analysis.

4.1.1 The CDF experiment description

The CDF experiment described in Section 3.1 can be easily translated into modern markup languages such as XML and JSON. We made a design choice to use XML because (i) it can be read and processed by a large set of programming languages, and (ii) it is internally used by OMF - a testbed cOntrol and Management Framework [24] which is widely adopted by many modern wireless testbed facilities [25,26]. The goal here is to ensure that the CDF experiment description can be easily translated to testbed/device-specific implementations. Additionally, we provide an XML schema [27] to validate the semantic correctness of an experiment description.

4.1.2 The CDF data structure

The CDF data structure is one of the formats that the trace file field in the CDF experiment description can reference to, it is meant to be a starting point where users can easily load data from different devices. The content of the CDF data structure is illustrated below:

```

p           = common data format structure
p.Name     = Unique identifier of the sensing device
p.Location  = Location of the sensing device (m) e.g. [x,y,z]
p.CentreFreq = Array of centre frequencies corresponding to
              the columns of the power matrix (Hz)
p.BW       = Bandwidth around each centre frequency (Hz)
p.Tstart   = Start time of the measurement in datestr format
              e.g. '24-Jan-2003 11:58:15'
p.SampleTime = Array of timestamps relative to Tstart (s)
              corresponding to rows of the power matrix (Hz)
p.Power    = Matrix containing power measurements (dBm)
              row contains all frequencies for one timestamp

```

For FFT-based sensing devices, defining both 'BW' and 'CentreFreq' is unnecessary. However, for pure sweeping-based sensing devices, the resolution bandwidth depends on the width of the band pass filter at the RF front-end, which is not necessarily the same as the distance between the consecutive RF centre frequencies. Therefore, both fields are included in the data structure so that it is suitable to store results from all types of sensing devices.

The field 'Location' is included in both the *CDF experiment description* and the *CDF data structure*, because it not only is important as 'meta-information' but also needed in various of calculations. Note that the device location in 'meta-information' is a description of the general experiment topology, while in the CDF data structure, it has to be expressed in coordinates.

The remaining fields of the CDF data structure are self-explaining, hence, omitted from further explanations.

4.1.3 The CDF toolbox

The CDF toolbox is a set of functions implemented in Matlab to support the usage of the CDF data structure. We choose Matlab as the programming environment because it is powerful in matrix processing and widely used among research and academic institutions. The most often used functions are listed below:

- The `create_structure` function extracts information from the input spectrum trace, and store the data in the CDF data structure.
- The `pw_integration` computes the power of a certain frequency band by integrating the PSD over the corresponding interval. It takes three input parameters - the spectrum trace in the CDF data structure, the begin and the end of the frequency interval, and produces the integrated power as output.
- Several plotting and analysis tools are implemented based on the fields of the CDF data structure, which makes the CDF more attractive from practical point of view.

Sample scripts of the CDF toolbox are made available online [27]. The scripts are currently only developed for

devices used in the CREW^a project. However, they can serve as valuable examples for other devices, given the fact that the existing examples already cover large varieties of sensing devices and data formats. Although it is not yet a full-fledged toolbox for sensing analysis, we believe it is one step towards the support of heterogeneous devices.

4.2 Calibration and resolution

This section first gives an overview of the devices involved in the experimental evaluations of the entire Section 4 and then describes a calibration experiment using the common metric derived in Section 3.2 and the instructions given in Section 3.3.

4.2.1 Overview of sensing devices

TelosB [28] is a sensor node developed at UC Berkeley. It is widely used by wireless sensor network community. The platform uses the IEEE 802.15.4-compliant CC2420 transceiver [29], which operates in the 2.4 GHz ISM band. The sensing application is built above TinyOS [30]. In our experiments, the device sweeps over the target spectrum in steps of 2 MHz and measures RF energy in each step. A single received signal strength indication (RSSI) is collected at every RF centre frequency. The RSSI is transferred to the host computer via USB connection in real-time. It takes around 2 ms to sweep over the entire 2.4 GHz ISM band. The collected RSSI value and its timestamp are stored in a comma-separated value (CSV) file. *TelosB* has less flexibility towards spectrum sensing applications, both in processing algorithms and RF functionalities; however, it has the lowest price as well.

The *metaGeek Wi-Spy 2.4x* and [31] *Airmagnet* [32] are both low-cost spectrum analysing devices, and they both use USB dongles as the RF front-end. The sensing mechanism of *Wi-Spy* resembles *TelosB* in the sense that it also uses a narrow-band RF receiver to scan across the band of interest in tiny steps. The step width ranges from 50 KHz to over 600 KHz, depending on the width of the frequency span. This essentially determines the resolution bandwidth of the spectrum trace. In our experiments, the *Wi-Spy* is used jointly with the *Kismet Spectools* [33] in

Linux environment, instead of the standard ‘Chanalyzer’ software. By doing so, the power spectrum trace can be stored in a non-proprietary format, which is more convenient for further processing. Unlike Wi-Spy, Airmagnet relies on FFT-based sensing algorithm. The radio of Airmagnet has a 20-MHz instantaneous bandwidth. It performs sweeping in steps of 20 MHz to cover a bandwidth that is wider than the instantaneous span. For 2.4 GHz ISM band, it has a fixed span of 83 MHz with 156 KHz resolution bandwidth. The PSD of Airmagnet can be stored with either CSV format or other proprietary formats.

USRP [34] is a relatively low-cost SDR platform that consists of two parts - a fixed motherboard and a removable daughterboard. The motherboard contains ADC and digital-to-analog converter (DAC), a field programmable gate array (FPGA) for digital down sampling and an interface connected to a host computer. The daughterboard provides RF front-end functionalities. There are many third-party software platforms, such as GNU Radio [35] and Iris SDR platform [36], which can communicate with the USRP. Thus, spectrum sensing applications can be implemented in many ways.

The *imec sensing engine* [37] is an integrated sensing device based on a custom design that targets for low-power and hand-held devices. Hence, it is powered and configured over a single USB connection. Similar to USRP, it has a separate PCB for the RF front-end functionality. The imec sensing engine has a very wide RF frequency range (from 100 MHz up to 6 GHz) and a programmable instantaneous bandwidth between 1 MHz and 40 MHz. Additionally, it uses a dedicated integrated circuit (IC) for signal processing instead of using the host computer. There are several pre-defined modes in the IC, including sensing based on FFT and sensing based on fast sweeping over a set of consecutive RF frequencies. The host application of imec sensing engine is written in C; therefore, the storage format is also flexible.

The overview of the devices are summarized in Table 1.

4.2.2 Calibration experiment

For the first step, the reference signals are defined as continuous OFDM signals with two different bandwidths (22 MHz and 5 MHz), transmitted on three different channels

(Wi-Fi channel 1,6,11 for the 22 MHz signal and Zigbee channel 11,16,26 for the 5 MHz signal), with three input signal strength (−60 dBm, −70 dBm, −80 dBm). The 22 MHz and 5 MHz bandwidth are selected to emulate Wi-Fi and Zigbee signals, respectively. A Rohde & Schwarz signal generator is used as the signal source.

The signal generator is connected to the sensing devices with coaxial cables and splitters as shown in Figure 3. The idle terminals are properly connected to terminators with matching impedance (50Ω). The calibration experiment consists of two simultaneous operations: continuous RF signal is produced by the signal generator, and at the same time, all devices record the sensing data to cover the same frequency span (2.4 GHz ISM band) for the same amount of time (1 min). This process is repeated for all the predefined reference signals, which means 18 iterations in total (2 types of bandwidth, 3 input levels, 3 channels).

After the recording, the raw spectra traces are converted in to the CDF data structure. Then the received power within the transmitted signal bandwidth is calculated from the raw PSD. More particularly, the `create_structure` and the `pw_integration` functions within the CDF toolbox are used to perform these operations. As the input signal has a constant amplitude, and the devices are configured to sense for the same amount of time, averaging the P_{rx} over the entire sensing duration is the most logical way to compute the common metric. Finally the power offset of each device against each reference signal is calculated according to Equation 1.

The minimum, maximum and average value of the measured offsets are plotted in Figure 4. The results indicate that the USRP solution has the largest offset, most likely due to the fact that it is a general research platform and the output of the customized software is not strictly calibrated. Airmagnet and Wi-Spy are both commercial USB spectrum devices, but their offsets are the opposite of each other. Additionally, Airmagnet has much bigger variations during the measurements (seen by the difference between the maximum and minimum offset values). We believe that this is most likely caused by the difference in sensing approach - Airmagnet uses FFT-based processing approach while Wi-Spy relies on pure narrow band sweeping (see Table 1). As this paper does not focus on

Table 1 Overview of the sensing solutions

Device name	Resolution bandwidth	Frequency span	Processing	Sweeping	Cost
TelosB	2 MHz	2.4 GHz ISM band	RSSI	Narrow band	Medium
Wi-Spy2.4x	50 to 600 KHz	2.4 GHz ISM band	RSSI	Narrow band	Medium
Airmagnet	155 KHz	2.4 and 5 GHz ISM bands	FFT	In 20 MHz	Low
USRP	48 KHz*	RF board dependent	FFT	In 20 MHz*	Medium
imecse	78 KHz*	100 MHz to 6 GHz	FFT	In 20 MHz*	High

*Indicates that the entry is configurable and only a typical value is entered.

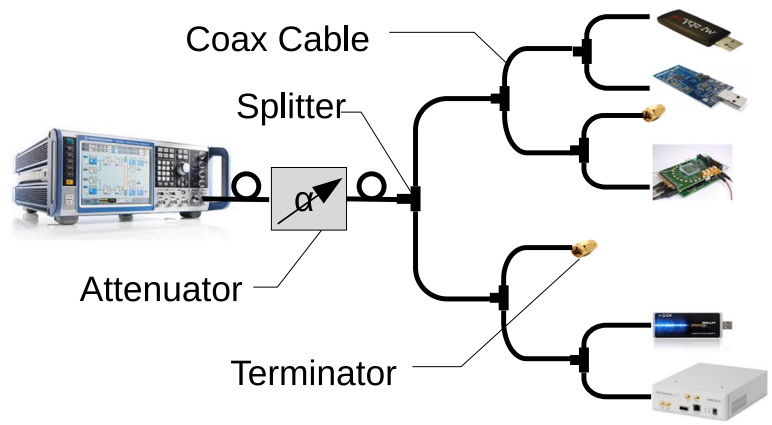


Figure 3 Replace wireless medium with coaxial cable and splitters.

the performance of sensing devices, the exact cause of the different offsets is not examined. However, the fact that heterogeneous devices have very different offsets confirms the need for calibration.

Finally, we would like to examine the calibration result by looking at the spectra of different devices with the same resolution bandwidth. To do so, we first subtract the mean P_{offset} from the collected raw spectra, and then divide the entire 2.4 MHz ISM band into a set of 2 MHz wide consecutive intervals (2 MHz is the largest resolution bandwidth among the considered devices, see Table 1) and perform integration over each of these intervals using the CDF toolbox. This operation essentially brings all spectra to the same frequency resolution. For comparison, the original raw spectra and the resulting spectra of the measurement obtained under one type of reference signal (22 MHz OFDM on Wi-Fi channel 1 with -60 dBm input power) are shown in plots (a) and (b) of Figure 5, respectively. Plot (a) shows that devices with larger resolution bandwidth have higher level of PSD than devices with finer resolution bandwidth. This behaviour can be best illustrated when comparing the raw spectra of TelosB and imec sensing

engine. Compared to plot (a), the spectra in plot (b) are much smoother due to the coarser resolution bandwidth and closer to each other within the interval where the signal is present, which is the expected behaviour after the calibration and resolution conversion. At the same time, we notice that there is still big differences between the spectra where the signal is not present. This difference is no longer linked to the influence of resolution bandwidth, but solely reflects the internal noise level of devices. Finally, the envelope of the 22 MHz OFDM signal is slightly shifted to the right for the case of TelosB, this is because the resolution bandwidth of TelosB is not fine enough to resolve the exact boundary of the OFDM signal in the frequency domain.

4.3 Reference experiment for heterogeneous sensitivity analysis

This section describes an experiment for sensitivity analysis following the instructions in Section 3.4.1. The experiment setup involves an MS3700A signal generator, placed 9 m away from the rest of the sensing devices. This experiment does not include the Airmagnet device due to

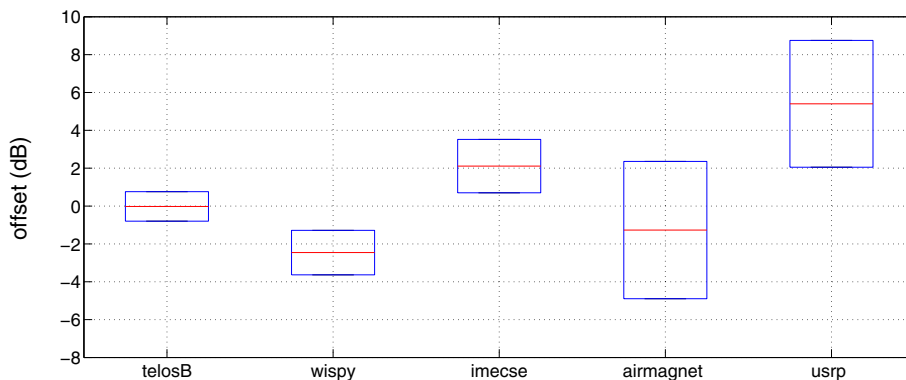


Figure 4 Offset vs device.

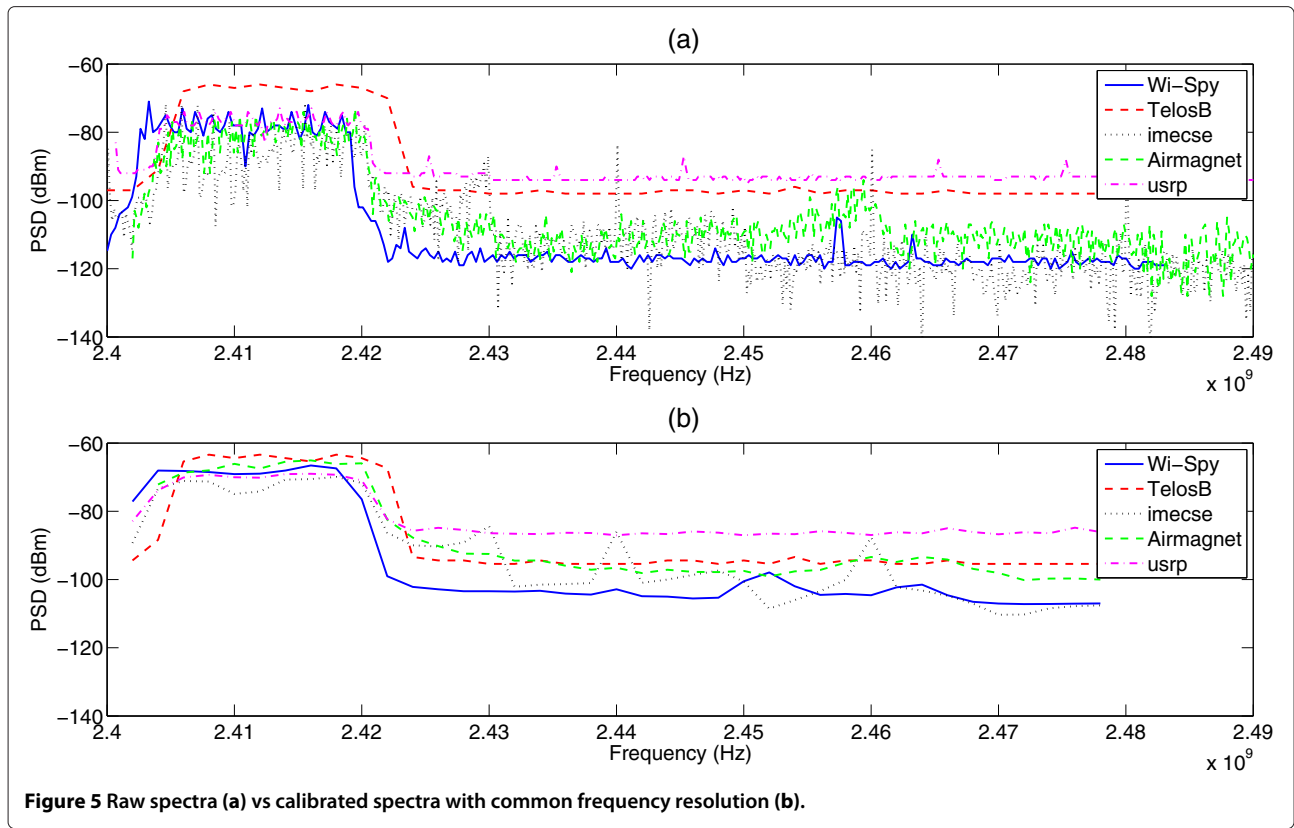


Figure 5 Raw spectra (a) vs calibrated spectra with common frequency resolution (b).

practical limitations at the time of this measurement. All devices are placed on the same horizontal level, with no obstacles in between.

First, samples are recorded when devices are shielded away from external signals. Then recordings are made when an 8 MHz wide OFDM signal is transmitted with various signal strength. The raw traces are converted to

the CDF data structure via the `create_structure` function in the CDF toolbox. In the next step, the detection thresholds for a set of P_f are calculated according to Equation 2, and finally the corresponding P_d values are computed.

The ROC plot obtained under -4 dBm signal strength is displayed in Figure 6. It shows that imec sensing

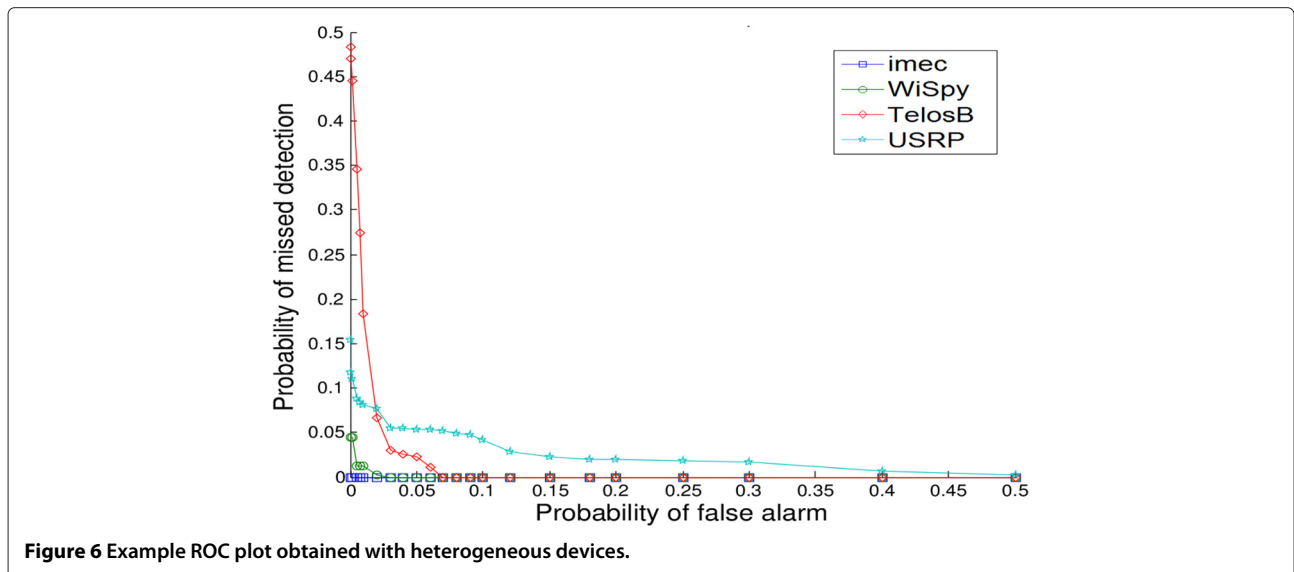


Figure 6 Example ROC plot obtained with heterogeneous devices.

engine and Wi-Spy have better sensitivity than USRP and TelosB. The lack of sensitivity for TelosB is clearly caused by the limitation of its large resolution bandwidth. While for USRP, the low sensitivity is due to insufficient amplification applied at the time of the experiment, which can be resolved in more recent sensing implementations [38].

4.4 Reference experiment for heterogeneous accuracy analysis

This section uses an experiment to illustrate how the processing mechanisms described in Section 3.4.2 can be used in power accuracy analysis. The experimental setup is shown in Figure 7, where 23 measurement locations and one transmitter’s location are chosen within an indoor cafeteria. During the measurement, the signal generator transmits a constant 20 MHz wide OFDM signal on Wi-Fi channel 8 (2.447 GHz), with 3 dBm transmit power. Each of the aforementioned sensing devices is configured to record the spectrum at all locations for a duration of minimum 30 s.

As stated in 3.4.2, the simple logarithmic path loss model is used for *dependency modelling*. Based on

Equation 2, path loss model is characterized by two parameters: the path loss exponent α and offset β . Therefore, the dependency modelling is realized by estimating the two parameters with the distributed power measurements.

Only one measurement from Wi-Spy device is identified as *outlier* due to the fact that its output was abnormal, the rest of the measurements are considered valid.

Two types of *regression* techniques are applied for the estimation: the least square regression and the robust regression. The least square regression attributes an equal weight to all input data so that the resulting α and β give the minimum mean squared error over the input dataset. On the other hand, the robust regression iteratively attributes weights to different ranges of the dataset, so that the impact of potential outliers is minimized.

Finally, the measurement locations are grouped into different *clusters* so that the locations which has a line-of-sight (LOS) topology, with respect to the transmitter, are separated from those that are in a none line-of-sight (NLOS) condition (shadowed by the coffee machine). The resulting path loss models estimated with the least

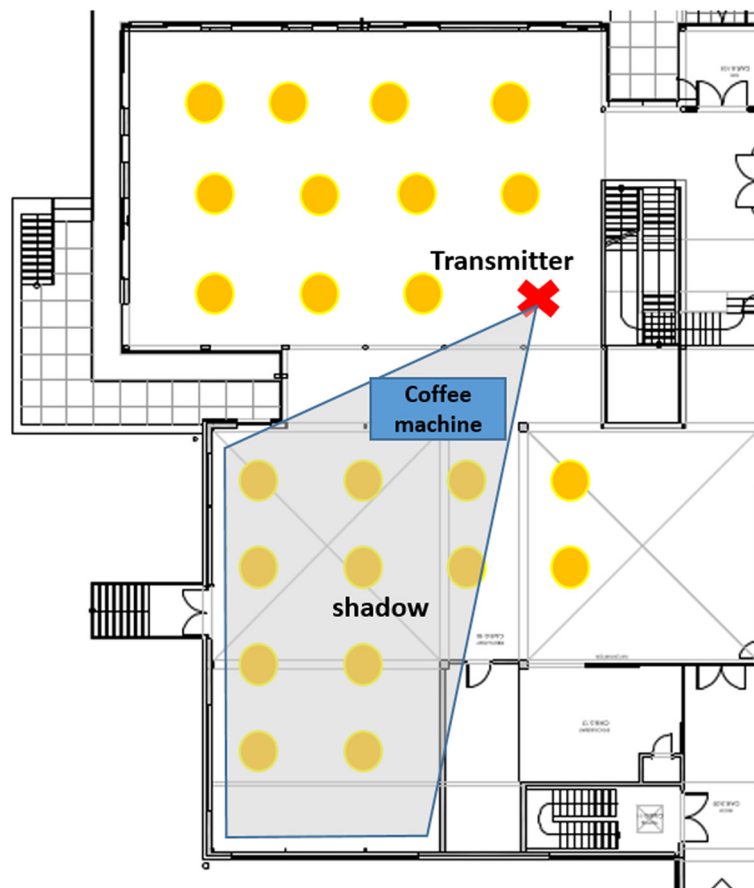


Figure 7 Experiment setup and ‘Cluster’ of locations adapted from [39].

square regression for both location clusters are shown in Figure 8. We observe that compared to LOS model, NLOS has a smaller slope but higher offset. The high offset in NLOS estimation is caused by the shadow effect. For the same reason, around shadow, the increment of path loss caused by distance can be compensated by the decreasing amount of shadowing; hence, the path loss exponent appears to be smaller than the LOS estimation. This analysis gives extra insight on the impact of shadowing in the power accuracy measurement, thanks to the ‘cluster’ of devices. More technical details about this experiment and its result analysis are presented in [39].

5 Conclusions

In this paper, we identify and address several challenges in heterogeneous spectrum sensing. First, we provide a *common data format (CDF)* (consisting of data structure and toolbox) to configure sensing devices and store measurement results in a uniform approach. We show that the use of CDF can effectively reduce the experiment overhead, however its implementation requires device-specific scripts.

Second, we apply aggregation techniques to raw spectra in both frequency and time domain (through CDF toolbox) to overcome heterogeneous measurement

resolution. We show that this technique can be used to compare heterogeneous spectra conveniently, however, it cannot be applied in sensitivity analysis, since sensitivity and resolution are related.

Third, we propose the use of a strict calibration process: replace the wireless medium by coaxial cables, use high-end signal generator as reference, derive the power offsets among devices and perform calibration. We validate experimentally that this approach is highly reliable and repeatable. The drawback is that the usage of coaxial cable leaves the antenna out of the calibration system. This can be resolved by adjusting the measured power offset with the offsets among antenna gains.

Finally, we share our experience of analysing two fundamental heterogeneous sensing metrics. More particularly, a well-accepted common procedure is applied to heterogeneous devices to achieve fair sensitivity analysis; basic data-mining techniques are used to extract new parameters concerning distributed power accuracy analysis. It can be seen that these techniques greatly improve the processing efficiency and trigger profound understanding of the measurements. Though processing mechanisms generally vary with experiment details, sharing these to the community will lead to a quicker harmonisation of approaches.

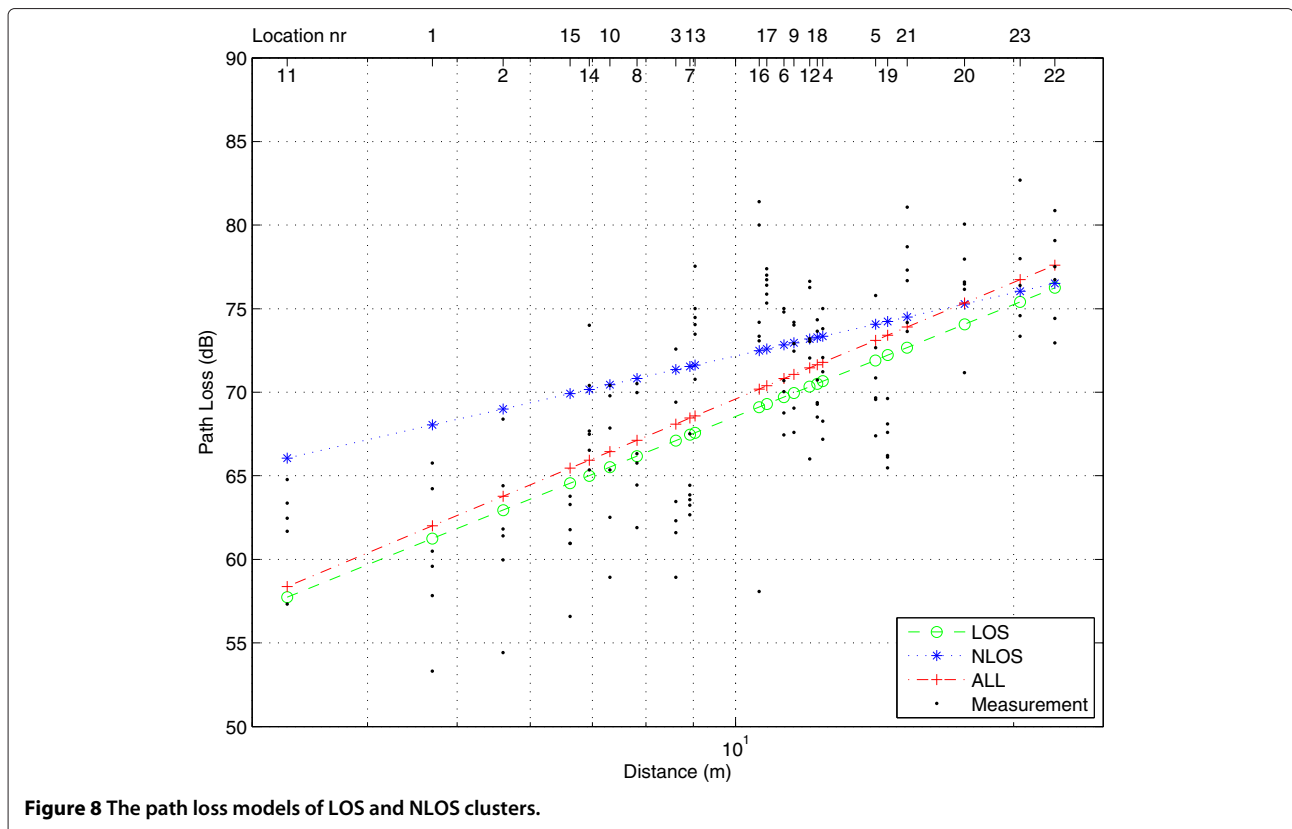


Figure 8 The path loss models of LOS and NLOS clusters.

In the future, before extending the CDF to support more devices, the device-specific implementation can be simplified and validated via standard procedures. Also more advanced aggregation techniques could be explored in order to meet the needs of different analysis and the influence of the antenna could be studied more in depth to improve the calibration accuracy.

Endnote

^aCognitive Radio Experimental World (CREW) is a project in European Unions Seventh Framework Programme FP7/2007-2013.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

The research leading to these results has received funding from the European Union's Seventh Framework Programme FP7/2007-2013 under grant agreement n 258301 (CREW project).

Author details

¹University Ghent - iMinds, Gaston Crommenlaan 8, 9050 Ghent, Belgium.

²Telecommunication Networks Group (TKN)/TU Berlin, Sekretariat FT 5, Einsteinufer 25, 10587 Berlin, Germany. ³imec, Kapeldreef 75, Leuven, 3001, Belgium, 3000 Leuven, Belgium.

Received: 28 February 2014 Accepted: 9 February 2015

Published online: 13 March 2015

References

1. V Valenta, R Marsalek, G Baudoin, M Villegas, M Suarez, F Robert, in *Cognitive Radio Oriented Wireless Networks & Communications (CROWNCOM), 2010 Proceedings of the Fifth International Conference On*. Survey on spectrum utilization in europe: measurements, analyses and observations (IEEE, Cannes, France, 2010), pp. 1–5
2. S Haykin, Cognitive radio: brain-empowered wireless communications. *Sel. Areas Commun. IEEE J.* **23**(2), 201–220 (2005)
3. J Walko, Cognitive radio. *IEEE Rev.* **51**(5), 34–37 (2005)
4. T Yucek, H Arslan, A survey of spectrum sensing algorithms for cognitive radio applications. *Commun. Surveys Tutorials, IEEE.* **11**(1), 116–130 (2009)
5. HR Karimi, in *New Frontiers in Dynamic Spectrum Access Networks (DySPAN), 2011 IEEE Symposium On*. Geolocation databases for white space devices in the uhf tv bands: Specification of maximum permitted emission levels (IEEE, Aachen, Germany, 2011), pp. 443–454
6. J Zander, LK Rasmussen, KW Sung, P Mahonen, M Petrova, R Jantti, J Kronander, On the scalability of cognitive radio: assessing the commercial viability of secondary spectrum access. *Wireless Commun. IEEE.* **20**(2), 28–36 (2013)
7. Federal Communications Commission — Rules and Regulations for Title 47. <http://www.fcc.gov/encyclopedia/rules-regulations-title-47>. Last accessed on 02/21/2015.
8. R Tandra, SM Mishra, A Sahai, What is a spectrum hole and what does it take to recognize one? *Proc. IEEE.* **97**(5), 824–848 (2009)
9. CF Tomaz Solc, M Mohorcic, in *Cognitive Radio and Networking in Heterogeneous Wireless Networks*. Low-cost testbed development and its applications in cognitive radio prototyping (Springer, 2015)
10. A Nika, Z Zhang, X Zhou, BY Zhao, H Zheng, in *Proceedings of the 1st ACM Workshop on Hot Topics in Wireless*. Towards commoditized real-time spectrum monitoring (ACM, Los Angeles, California, USA, 2014), pp. 25–30
11. D Cabric, A Tkachenko, RW Brodersen, in *Proceedings of the First International Workshop on Technology and Policy for Accessing Spectrum*. Experimental study of spectrum sensing based on energy detection and network cooperation (ACM, Boston, Massachusetts, USA, 2006), p. 12
12. M Murrioni, RV Prasad, P Marques, B Bochow, D Nogueta, C Sun, K Moessner, H Harada, IEEE 1900.6: spectrum sensing interfaces and data structures for dynamic spectrum access and other advanced radio communication systems standard: technical aspects and future outlook. *Commun. Mag. IEEE.* **49**(12), 118–127 (2011)
13. J Caufield, Protocol to Query a White Space Database. <http://tools.ietf.org/html/draft-caufield-paws-protocol-for-twws-01>
14. D Denkovski, V Rakovic, M Pavloski, K Chomu, V Atanasovski, L Gavrilovska, in *Wireless Communications and Networking Conference (WCNC), 2012 IEEE*. Integration of heterogeneous spectrum sensing devices towards accurate rem construction (IEEE, Paris, France, 2012), pp. 798–802
15. V Atanasovski, J van de Beek, A Dejonghe, D Denkovski, L Gavrilovska, S Grimoud, P Mahonen, M Pavloski, V Rakovic, J Riihijarvi, B Sayrac, in *New Frontiers in Dynamic Spectrum Access Networks (DySPAN), 2011 IEEE Symposium On*. Constructing radio environment maps with heterogeneous spectrum sensors (IEEE, Aachen, Germany, 2011), pp. 660–661
16. J Van De Beek, T Cai, S Grimoud, B Sayrac, P Mahonen, J Nasreddine, J Riihijarvi, How a layered rem architecture brings cognition to today's mobile networks. *Wireless Commun. IEEE.* **19**(4), 17–24 (2012)
17. TF TRL, FC TRL, SS TRL, W Workpackage, Flexible and spectrum aware radio access through measurements and modelling in cognitive radio systems. *Evaluation.* **5**(2) (2009)
18. M López-Benítez, F Casadevall, Methodological aspects of spectrum occupancy evaluation in the context of cognitive radio. *Eur. Trans. Telecommunications.* **21**(8), 680–693 (2010)
19. D Finn, JC Tallon, LA DaSilva, P Van Wesemael, S Pollin, W Liu, S Bouckaert, J Vanhie-Van Gerwen, N Michailow, J Hauer, D Wilkomm, C Heller, in *Proceedings of the 6th ACM International Workshop on Wireless Network Testbeds, Experimental Evaluation and Characterization*. Experimental assessment of tradeoffs among spectrumsensing platforms (ACM, New York, USA, 2011), pp. 67–74
20. WL Stutzman, WA Davis, *Antenna Theory*. (Wiley Online Library, 1998)
21. Y Liu, C Zeng, H Wang, G Wei, in *Advanced Computer Control (ICACC), 2010 2nd International Conference On*. Energy detection threshold optimization for cooperative spectrum sensing, vol. 4 (IEEE, Shenyang, Liaoning, China, 2010), pp. 566–570
22. W Zhang, RK Mallik, K Letaief, in *Communications, 2008. ICC'08. IEEE International Conference On*. Cooperative spectrum sensing optimization in cognitive radio networks (IEEE, Beijing, China, 2008), pp. 3411–3415
23. N Wang, Y Gao, X Zhang, Adaptive spectrum sensing algorithm under different primary user utilizations. **17**(9), 1838–1841 (2013)
24. T Rakotoarivelo, M Ott, G Jourjon, I Sesar, Omf: a control and management framework for networking testbeds. *ACM SIGOPS Operating Syst. Rev.* **43**(4), 54–59 (2010)
25. S Bouckaert, P Becue, B Vermeulen, B Jooris, I Moerman, P Demeester, in *Testbeds and Research Infrastructure. Development of Networks and Communities*. Federating wired and wireless test facilities through emulab and omf: the ilab. t use case (Springer, thessaloniki, Greece, 2012), pp. 305–320
26. V Handziski, A Köpke, A Willig, A Wolisz, in *Proceedings of the 2nd International Workshop on Multi-hop Ad Hoc Networks: from Theory to Reality*. Twist: a scalable and reconfigurable testbed for wireless indoor experiments with sensor networks (ACM, Florence, Italy, 2006), pp. 63–70
27. CREW Project. <http://www.crew-project.eu/repository/>. Last accessed on 02/16/2015.
28. Ultra low power IEEE 802.15.4 compliant wireless sensor module. <http://www.eecs.harvard.edu/~konrad/projects/shimmer/references/tmote-sky-datasheet.pdf>. Last accessed on 02/16/2015.
29. 2.4 GHz IEEE 802.15.4 ZigBee-ready RF Transceiver. <http://www.ti.com/lit/ds/symlink/cc2420.pdf>. Last accessed on 02/16/2015.
30. P Levis, S Madden, J Polastre, R Szewczyk, K Whitehouse, A Woo, D Gay, J Hill, M Welsh, E Brewer, D Culler, in *Ambient Intelligence*. Tinyos: An operating system for sensor networks (Springer, 2005), pp. 115–148
31. How Wi-Spy Works. <http://www.metageek.net/blog/2011/01/how-wi-spy-works>. Last accessed on 02/16/2015.
32. Airmagnet Survey. http://airmagnet.flukenetworks.com/assets/datasheets/AirMagnet_Survey_Datasheet.pdf. Last accessed on 02/16/2015.
33. J Bock, M Lynn, *Hacking Exposed Wireless*. (McGraw-Hill, Inc., 2007)
34. Ettus Research. <http://www.ettus.com/>. Last accessed on 02/16/2015.
35. E Blossom, Gnu radio: tools for exploring the radio frequency spectrum. *Linux J.* **2004**(122), 4 (2004)

36. PD Sutton, J Lotze, H Lahlou, SA Fahmy, KE Nolan, B Ozgul, TW Rondeau, J Noguera, LE Doyle, Iris: an architecture for cognitive radio networking testbeds. *Commun. Mag. IEEE*. **48**(9), 114–122 (2010)
37. S Pollin, L Hollevoet, P Van Wesemael, M Desmet, A Bourdoux, E Lopez, F Naessens, P Raghavan, V Derudder, S Dupont, A Dejonghe, in *New Frontiers in Dynamic Spectrum Access Networks (DySPAN), 2011 IEEE Symposium On*. An integrated reconfigurable engine for multi-purpose sensing up to 6 ghz (IEEE, Aachen, Germany, 2011), pp. 656–657
38. W Liu, D Pareit, E De Poorter, I Moerman, Advanced spectrum sensing with parallel processing based on software-defined radio. *EURASIP J Wireless Commun. Netw.* **2013**(1), 1–15 (2013)
39. P Van Wesemael, W Liu, M Chwalisz, J Tallon, D Finn, Z Padrah, S Pollin, S Bouckaert, I Moerman, D Willkomm, in *Future Network & Mobile Summit (FutureNetw), 2012*. Robust distributed sensing with heterogeneous devices (IEEE, 2012), pp. 1–9

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
