Ensuring Operational Privacy of Primary Users in Geolocation Database-Driven Spectrum Sharing

Summary Report for Task 1: Prior-Art Privacy Preserving Databases and Threat Models

Jung-Min "Jerry" Park and Jeffrey H. Reed

Wireless @ Virginia Tech Group
Bradley Department of Electrical and Computer Engineering
Virginia Tech

Technical Point of Contact: Jerry Park
540-231-8392, jungmin@vt.edu

June 30, 2013
Blacksburg, Virginia

Contents

List of Figures

List of Tables

Chapter 1

Introduction

When incumbent users (a.k.a. primary users (PUs)) and secondary users (SUs) share spectrum, the SUs must adopt technologies that enable them to avoid causing any interference to PUs. The FCC ruling on TV white spaces proposes relying on a database of the incumbents' spectrum usage information as the primary means of determining white space availability at any white space device (WSD) [2]. The database is required to house an up-to-date repository of incumbents including television stations, and in certain cases, wireless microphones, and use this information to determine white spaces availability at a WSD's location. It has been shown that sensing-only devices do not generally utilize spectrum as efficiently as geolocation enabled devices, due to the large margins in incumbent detection thresholds that must be built into sensing-only devices [3]. geolocation enabled devices have knowledge of the specific interference protection requirements of each licensed incumbent, which allows varying levels of protection to be applied, and thus maximizing utilization of the spectrum.

Although using geolocation databases for spectrum sharing has many advantages, it poses a potentially serious privacy problem. For instance, SUs, through seemingly innocuous queries to the database, can determine the *types* and *locations* of incumbent systems operating in a given region of interest—we refer to this as the *operational privacy* of the incumbents. When the incumbent systems are commercial systems, such as the case in TV spectrum, this is not an issue. However, when the incumbents are federal government systems, including military systems, then the information revealed by the databases can result in serious breach of privacy.

The primary incumbent contents in the SAS database that must be secured for primary user protection are the followings:

1. Transmitter Identity (i.e. the Call Sign of the transmitter in FCC CDBS)
2. Its geolocation (i.e. Latitude and Longitude)
3. Antenna Parameters (HAAT, etc)

4. Power (Max EIRP, Avg operation Power, etc)
5. Protection Contours of Operations (co-channel, Adjacent channel etc)
6. Periods of Operations

The prospects of spectrum sharing between federal government systems and non-government systems has been heightened by a recent *notice of proposed rule making (NPRM)* published by the Federal Communications Commission (FCC). In December 2012, the FCC published an NPRM to create a *Citizens Broadband Service* in the 3.5 GHz band that will promote two major technical advances that enable more efficient use of spectrum: small cells and spectrum sharing [4]. As part of the NPRM, the FCC is looking at whether it will be feasible to open up approximately 100 megahertz of spectrum in the 3550-3650 MHz bands for small cell technologies on an unlicensed basis. The 3.5 GHz band is currently used by the U.S. Department of Defense (DoD) for certain radar installations as well as by non-federal users. It is highly likely that spectrum sharing in the 3.5 GHz band will be enabled by geolocation databases aided by local spectrum sensing. In the 3.5 GHz band, the criticality of the incumbents' operational privacy is obvious. The operational privacy of the incumbents is one of the major hurdles holding back the federal government from opening up some of its spectrum to spectrum sharing with commercial systems. Research programs recently launched by federal funding agencies, including the NSF and DARPA, have stressed the importance of security and privacy in spectrum sharing [5], [6].

The problem of operational privacy of PUs cannot be addressed by tightly controlling access to the database, since all SUs need access to it to enable spectrum sharing. A more viable approach is to "obfuscate" the information revealed by the database in an intelligent manner such that a certain level of privacy is assured while supporting efficient use of the spectrum.

The rest of the report is organized as follows. We first discuss the Spectrum Access System (SAS) and some use case scenarios in chapters 2 and 3 respectively. Chapter 4 describes the database access protocol that is used by the geolocation database. Chapter 5 discusses the wireless standardization activities that involve geolocation databases. In chapter 6, we introduce some of the most well known privacy-preserving techniques that are widely used to ensure user's privacy in relational databases. We propose our threat model for operational privacy of incumbent users in chapter 7, and finally discuss how we can apply the existing privacy-preserving techniques to the geolocation databases in chapter 8.

Chapter 2

Spectrum Access System (SAS)

2.1 Introduction

The FCC released its National Broadband Plan [7] citing the exponentially growing demand for mobile data services and the critical need to utilize the radio spectrum efficiently. Moreover, in the President's Council of Advisors on Science and Technology (PCAST) report [1] emphasized the role of the spectrum as an important economic growth mechanism. The PCAST report proposed a shared spectrum access model, wherein a heterogeneous mix of wireless systems of differing access priorities, QoS requirements, and transmission characteristics need to coexist without causing harmful interference to each other. In this spectrum sharing model, secondary users identify unused spectrum by accessing a geolocation database that is constantly updated with the primary user's spectrum utilization information. The PCAST report defines a three-tier hierarchy for access to federal spectrum bands. The federal primary users have an exclusive right to use the spectrum when they deploy their networks or systems. The secondary users are allocated short term rights for operation in a specific geographical area. They are assured of interference protection with priority over opportunistic users. Any spectrum resources that are left over from the first and second tier users are made available to the general authorized access users. They opportunistically use these spectrum resources with an obligation to clear the spectrum in case of a federal primary or secondary user appears in that spectrum band.

For shared access to federal bands, a Spectrum Access System (SAS) with a geolocation database will be used [1] . Access to spectrum is authorized after successful communication and registration with the database. Unlike the TV whitespace database, the federal SAS [1] has more functional and operational requirements that need to be satisfied. Table 2.1 compares the functional requirements of a TV White space database with those of SAS.

Table 2.1: Comparison of Spectrum Access System Capability [1]

Functionality	White Space in TV Bands	White Space Federal Addition	Spectrum Access System (SAS)
Accept specific interference contours for federal primary access users and specific secondary uses	Yes	Yes	Yes
Automatically determine interference possibilities for any secondary technology	No	No	Yes
Register the location of secondary devices authorized to operate	No	Yes	Yes
Provide deconfliction of secondary spectrum users	No	No	Yes
Provide real time input of primary user operating locations and periods	No	Yes	Yes
Provide marketplace for leasing of spectrum and revenue to treasury	No	No	Yes
Provide the Spectrum Management Team (SMT) metrics [1] and advanced features like time to live (TTL)	No	No	Yes

2.2 Requirements

For an effective operational system [8], the SAS must fulfill the requirements of both the primary users and secondary users in the shared spectrum bands. The requirements regarding the protection of primary users include the following.

1. No interference to existent primary users of the spectrum band.
2. Secondary user must have the ability to reconfigure for accommodating changes in primary use like waveform types, occupancy, and locations, etc.
3. Secondary users must have backup bands to allow the primary licensed users to reclaim their spectrum at any time.
4. Systems must have mechanisms for enforcement of spectrum rules to track down interference events quickly and reliably.
5. System must be protected against any unauthorized/accidental use, and security must be provided against hackers.
6. Efficient system management for operating complex secondary to secondary system to ensure that agreed parameters are not violated.
7. Security of government agencies and protection of their classified information.

The requirements regarding secondary users spectrum access in the context of the SAS system [8] must fulfill are the following.

1. Interference requirements for the secondary systems must be reasonable for practical system deployment.

2. Existing broadband system architecture must be supported with minimal changes to existing standards and with low software integration costs, etc.

3. Secondary QoS requirements, low power operation must be achievable with no harmful interference to the primary systems.

4. A fair use policy with reliability and assured access for all secondary users must be enforced.

5. System must be secured against any unauthorized/accidental use and security must be provided against hackers.

2.3 Features

To fulfill the above-mentioned requirements, the SAS includes many features that are intended for interference protection of incumbent primary and secondary systems [9]. Along with the ability to accommodate changes in the federal user's operational parameters or response to unforeseen interference scenarios. Central entity of the SAS model is the database but sensing, and dynamic frequency selection must be used to enhance the spectrum utilization efficiency of the model. These capabilities are incorporated through the operational features defined by the PCAST report and summarized below.

1. *Channel selection* decisions for devices based upon their location, QoS requirements of the application and spectrum access rules specified by the database for that geographic location.

2. *Enforcement* of the spectrum reclaims from the primary user by switching off secondary communications on certain frequency channels through commands from the database or signal beacons.

3. Database must *validate the equipment* used for secondary access at the time of a spectrum request. Equipment validation can be done through the use of FCC certification identities.

4. The SAS system implemented either as centralized or distributed system must have defined generic terms of use in all the available bands in the SAS database. In other

words, from a user perspective, the database must provide a single consistent interface for accessing all geolocation database information and channel allocation methods like the Internet Domain Naming System (DNS) to facilitate a greater opportunistic use of the spectrum.

5. All the registrations and reservations for use must be time limited and renewed as appropriate. The database must have Time To Live (TTL) mechanisms, that can also be used for enforcement by revoking the secondary user spectrum authorizing.

6. Security must be provided by the SAS for both the database request and the response using a public key cryptographic system. Experience from the Digital Rights Management (DRM) systems can be used in securing the whole operational mechanism.

2.4 Data Model

In addition to the above-mentioned requirements, the SAS and the access protocol [1], [8], [10] must have algorithms and data types to support implementation specific details. Some of these requirements are mentioned below.

1. *Radio and Spectrum Regulatory Concepts.* They include channel, co-channel, adjacent channel, modulation, waveform, data rate, type of filtering, block size, device characteristics and attributes (e.g., FCC ID, serial number, transmitter, receiver, detector), Equivalent Isotropically Radiated Power (EIRP), mean EIRP, peak receiver (RX) power, frequency, center frequency, power masks, bandwidth, duty cycle, signal detector [including detection threshold, frequency range, sample rate, precision, Signal to Noise Ratio (SNR), and Received Signal Strength Indicator (RSSI)], and signal type.

2. Support for *signal evidence* like detected signal, sensed frequency intervals, peak sensed power, detected time, scan time/duration, count along with location evidence and time evidence.

3. Support *Scalar constraints* like restrictions on frequencies, time and dates.

4. *Security considerations* like the type of encryption, keys, key exchange, credential, security mechanism (e.g., integrity, confidentiality, authentication, authorization), and security/classification level.

5. *Networking concepts* like node identities, network membership, and types of networks i.e., peer 2 peer etc.

6. *Policy authority* for primary and secondary spectrum markets with primary and non-primary users.

2.5 Functional Components

The Spectrum Access System (SAS) is a dynamic database system and consists of many logical and physical components that will allow a number of unique capabilities like real time channel availability from dynamic calculations of the protection contour zones of stationary and mobile primary users. Interference protection and coexistence capability to calculate the secondary network interference power spectral density (IPSD) at primary incumbent location for allowing operations inside the protection zones as well.

The spectrum access system consist of the following components.

1. Spectrum Manager
2. Database Management System
3. Primary Incumbent Update Mechanism
4. Access Mechanism to SAS

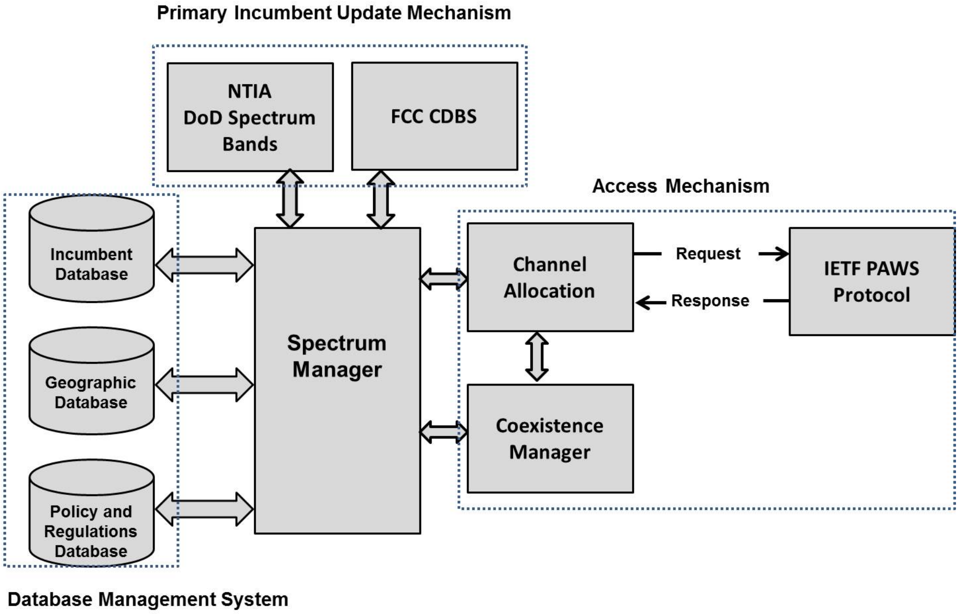

Figure 2.1: Block diagram of Spectrum Access System (SAS) with functional components.

2.5.1 Spectrum Manager

Spectrum manager is responsible for ensuring the protection of incumbents and efficient spectrum utilization while complying with regulatory policies [11]. Following are some of the functions of the spectrum manager.

1. Maintain spectrum availability information
2. Calculations for available channels, spectrum quality ranking and prioritization
3. Association control mechanisms for devices with the SAS
4. Channel set management for prioritized access
5. Scheduling for spectrum sensing
6. Enforcing regulatory domain policies
7. Managing spectrum mobility
8. Facilitating the coexistence of multiple wireless services

Spectrum manager has the database management system that it uses for the spectrum availability information through available white-space channel calculations, ranking the spectrum quality, developing the spectrum prioritization and maintaining a channel set for prioritized primary access, secondary licensed access, and general authorized access users.

2.5.2 Database Management System

The DBMS system will reside inside the spectrum manager and include a number of database entities having information about primary incumbents in different spectrum bands like TV stations (54-72 MHz, 76-88 MHz, 174-216 MHz and 470-806 MHz) [2], Satellite systems (1675-1710 MHz) [9], RADAR systems (3500-3650 MHz) [9], Federal communication networks (1755-1850 MHz) [9], Radio Altimeters(4200-4400 MHz) [9], etc. Each incumbent has their own database with primary entity parameters, geographical parameters, regulatory protection constraints, and available white-space resources.

Entity Parameters

The entity parameters of each wireless device will depend on application for the device like the application parameters defined in the FCC CDBS [12] [13]. The entity parameters will include transmitter, receiver, antenna, mobility, and network parameters [10], [11], [14].

The transmitter parameters will include but not limited to the device ID, its modulation, bandwidth, transmission power, power spectral density (PSD), spurious emissions, access methodology (periodic/continuous) .

Receiver/sensing parameters will include but not limited to the type of detector, noise estimate, sampling rate, bandwidth, time stamp, location stamp, sensing values, etc.

The antenna parameters will define the maximum gain, antenna pattern, elevation angle, azimuth angle, Effective Isotropically Radiated Power (EIRP), height of antenna above terrain (HAAT), polarization, beam width, number of sectors, maximum sweep angle, number of elements, and system type.

Mobility parameters will specify the direction of motion and speed of the device with constructs like speed, velocity, and acceleration, etc.

The entity parameters will also include the network parameters to which the device is associated with i-e Network ID, device role, context parameters, etc.

Geographical Parameters

The geographic parameters includes the primary incumbent‘s locations, the terrain features of the environment, and the signal propagation conditions. Primary incumbent locations are used to characterize the geographical features around the incumbents. These parameters are in the form of latitude, longitude and height above the terrain.

Terrain features specify the administrative/political boundaries and other features like plain flat areas, hilly terrain, mountains, and bodies of water like lakes, rivers and oceans, etc. These terrain features will be generated from publicly available National Elevation Database [15]

The signal propagation conditions [16] defines all the parameters that affect the propagation of radio frequency electromagnetic waves. These parameters include but not limited to type of terrain, terrain irregularities, type of built environment, electrical ground constants, radio climatic conditions, and surface refractivity, etc.

Regulatory Protection Requirements

The SAS system maintains a database of regulatory protection requirements [1], [9], [17]. The regulatory requirements will specify a framework for sharing the spectrum bands by providing aggregate Interference Power Spectral density (IPSD) limits, along with spectrum masks, underlay masks, co-channel and adjacent channel interference limits, etc. The spectrum manager using the regulatory protection requirements [17] establish the following .

1. Utilize the regulatory-approved interference prediction model, associated input parameters and aggregate IPSD distributions for authorizing access to commercial within and out of incumbent protection zones.

2. Access the IPSD limits of the secondary network on the primary incumbent user to

facilitate coordination among primary and secondary users for authorizing the use of spectrum bands within the protection areas.

3. Use the spectrum sensing functionality for collecting real time spectrum use data to enforce the IPSD limits in the geographical area of its operation.

The regulatory protection database has constructs for implementing the above-mentioned capabilities along with the ability for sharing and acquisition with the regulatory update mechanisms of the FCC CDBS and NTIA databases.

Available White-Spaces

The spectrum manager will use the entity parameters from the databases, the geographical parameters and calculate the available white-space channels through calculation‘s methods [18], [19], [20], [21], [22] by taking in order to account the regulatory protection requirements. Most of these channel calculation algorithms use the entity parameters, antenna patterns, geographic locations, terrain databases for calculating the protection contours around the primary incumbents like TV transmitters, CMRS/PLMRS, Wireless Microphones, Radio Astronomy sites, etc. through propagation models [23], [16], [24] defined by these standards.

The protection contours are overlaid in the form of channel power plots [18] to calculate the available channels for white space devices (WSD) in a geographical area. The WSD channel availability is specified through minimum allowed transmit power level for the cognitive radio link. The transmit power levels for WSD varies across a region, due to high spatial variability of the primary incumbent protected service areas. This leads to fewer high powered channels and more lower powered channels making the spectrum availability highly dynamic across a region.

2.5.3 Primary Incumbent Update Mechanism

The primary incumbents‘ information is updated in the geolocation database from regulatory databases of FCC Consolidated Database Systems (CDBS), and the NTIA database of DoD incumbent systems. This mechanism ensures reliability of information in the geolocation databases for proper operations.

The database administrators must ensure the accuracy of their primary incumbent information by regular updates of their information about the primary incumbents locations, access patterns (i.e., periodic or continuous), regulatory interference protection levels, spectrum availability for open spectrum markets, etc from the FCC Consolidated Database Systems (CDBS) [12], and NTIA Federal Spectrum Management System (FSMS) [25].

These regulatory databases are designed as license databases not as spectrum management databases and cannot be used for new applications of wireless spectrum [26]. The SAS will fulfill this gap and provide all the necessary functionality to promote technological innovations.

The database administrators provide web-based interfaces for registration of primary incumbent users (DTV stations, Wireless microphones, etc.) and the recent FCC rules require these incumbents to update their database registrations daily.

2.5.4 Access Mechanism to SAS

The access mechanism for the SAS includes a generalized database access protocol with a data model that can support the implementation of the objectives mentioned in the Table 2.1. The white space allocation of channels from the database is analogous to the Internet Domain Naming System (DNS), which maps the symbolic names to the IP addresses and is transparent to the Internet users. The SAS access protocol must map the user spectrum requirements to a database query and perform all the necessary protocol related operations like secure database discovery, determining essential query parameters, and exchanging messages with the database for desired frequency channels of operation. This ease of access would enable greater opportunistic use of the RF spectrum.

The Internet Engineering Task Force (IETF) is currently working on development of a Protocol for Access to Whitespace Spectrum (PAWS), that will be used to request resources from the geolocation database. The protocol support spectrum queries agnostic of the spectrum bands, and takes in user device description (type, ID, capabilities), location, antenna characteristics, etc. These parameters are sent in the query message to the geolocation database asking for available spectrum. The database responds back with a set of available channels, time schedule for use of the spectrum, rule set for that area, and maximum allowable location change after which the spectrum lease needs to be renewed.

The current implementation of the IETF PAWS does not support enforcement of spectrum rules, coexistence mechanisms, dynamic interference protection. But provides flexible and extensible data structures to implement the SAS functionalities with extended constructs for enforcement, coexistence and interference assessment.

Chapter 3

Use Case Scenarios

There are a number of scenarios in which the White Space Database (WSD) will be useful in establishing and maintaining a communication network. In this section, we discuss a few use case scenarios as case studies in describing our spectrum access system (SAS). In all these scenarios, the white space database supported networks have the following objectives.

1. Sensing over a wide frequency band and identifying primary users in these bands.
2. Characterizing the available spectrum opportunities and populating them in a local database
3. Accessing the geolocation database for available spectrum and updating the database with field measurements and statistics from the local database.
4. Communication between the devices to coordinate the use of identified opportunities, i.e., resource allocation
5. Expressing and applying interference-limiting policies, i.e., regulatory enforcement
6. Enforcing behavior consistent with applicable policies while using identified opportunities.

These objectives are helpful in implementing the spectrum access system (SAS) as mentioned in the section 2.2. For implementation of the above-mentioned objectives, unique functionalities are needed in the wireless networks. These functionalities include spectrum sensing, spectrum decision, spectrum sharing, and spectrum mobility.

Spectrum sensing is an important requirement for white space operations, as radios need to be aware of their environment and must respond to changes in this environment. Spectrum sensing enables the radios to exploit the unused spectrum portions adaptively. Spectrum sensing involves three main processes of PU detection, cooperation model and sensing control.

PU detection is necessary to distinguish between used and unused spectrum bands and is achieved by utilizing matched filter detection, energy detection, or cyclostationarity feature detection of local RF observations. Sensing process is based on hypothesis testing.

$$R(t) = \begin{cases} n(t) & H_0(t) \\ h \times s(t) + n(t) & H_1(t) \end{cases} \tag{3.1}$$

Where $r(t)$ is the signal received by the radio, $s(t)$ is the transmitted signal of the PU, $n(t)$ is a noise component and h is the amplitude gain of the channel. H_0 is a null hypothesis, which states that there is no licensed user signal in a certain spectrum band. H_1 is the alternative hypothesis, which indicates that there exists some PU signal.

The main objective of spectrum sensing is to find more spectrum access opportunities and reduce the interference to primary networks by reliable PU detection. For this purpose, the sensing operation of each radio is controlled and coordinated by a sensing controller, which determines how long and frequently radios should sense the spectrum for achieving sufficient sensing accuracy for in-band sensing and how quickly radios can find the available spectrum band in out-of-band sensing.

Spectrum sensing requires an efficient cooperation scheme in order to prevent interference to PU outside the observation range of each radio. This cooperation scheme can be through an infrastructure based cellular system or an ad-hoc network (AHN) architecture.

Spectrum decision requires capabilities to choose on the best spectrum band among the available bands according to the QoS requirements of the application. It depends on three fundamental requirements of spectrum characterization, spectrum selection, and reconfiguration.

Spectrum characterization is based on the radio environment observation. Radios determine the characteristics of each available spectrum and the primary user activity model.

In *spectrum selection*, the radios determine the finest spectrum band to satisfy the end-to-end QoS requirements.

The radios *reconfigure* communication protocol as well as the PHY and MAC layer protocols to determine Tx Power, with optimum combination of modulation and FEC codes according to the radio environment and the user QoS requirements.

The radio environment for ad-hoc networks contains information about interference, path loss, wireless link errors and link layer delay. It also contains the metrics for the primary user activity parameters.

Spectrum Sharing coordinates the transmission attempts between multiple users. Spectrum sharing provides the capability to maintain the QoS of each user without causing interference to the primary user by coordinating the multiple access of CR users as well as allocating communication resources adaptively to the changes in the radio environment. Spectrum sharing has some functional requirements similar to the spectrum sensing.

In *Resource Allocations*, each user is assigned a channel and transmit power to achieve the QoS requirements as well as resource fairness. Power control is important, so as not to violate the interference constraints.

Spectrum Access enables multiple users to share the spectrum resource by determining who will access the channel or when a user may access a channel. Spectrum access depends on the MAC layer protocols. They are classified into random access the slotted access or hybrid methods, etc.

Spectrum Mobility is the switching of communication from one channel to the next one in case the PU appears in the channel. Users need spectrum hand-off due to the following reasons.

1. PU is detected in the current operating band
2. Radio loses its connection due to the mobility of users involved in the ongoing communications
3. Current spectrum band cannot provide the QoS requirements

Spectrum hand-offs are classified as proactive or reactive spectrum hand-offs. Spectrum mobility is closely related to the routing protocol. It involves the recovery of link failures on the end-to-end route. For uninterrupted communications, a backup list of channels is maintained for high probability of finding channels for use in the shortest period of time.

3.1 Centralized Infrastructure Based Cellular System

A centralized system consists of a number of radios that are geographically dispersed in a region served by a base station. The base station manages the allocation of resources for the processes of spectrum sensing and communications. In the spectrum sensing process, the base station selects a band of frequency channels for spectrum sensing, selects a set of radios for sensing each frequency channel, and when the radios report sensing results, it combines them using a fusion rule to achieve the desired level of spatial and spectral diversity to mitigate fading and shadowing effects in sensing process.

The desired band for spectrum sensing is divided into frequency channels based on the knowledge of existing incumbents in the band. In case of a TV Channel, the subdividing is performed using the knowledge of signal features, such as pilot signals and other characteristic signal levels in the channel. In case of radar bands, the existence of primary radar can be confirmed by measuring the main beam frequency or the side lobe frequency channels to avoid high transmit powers.

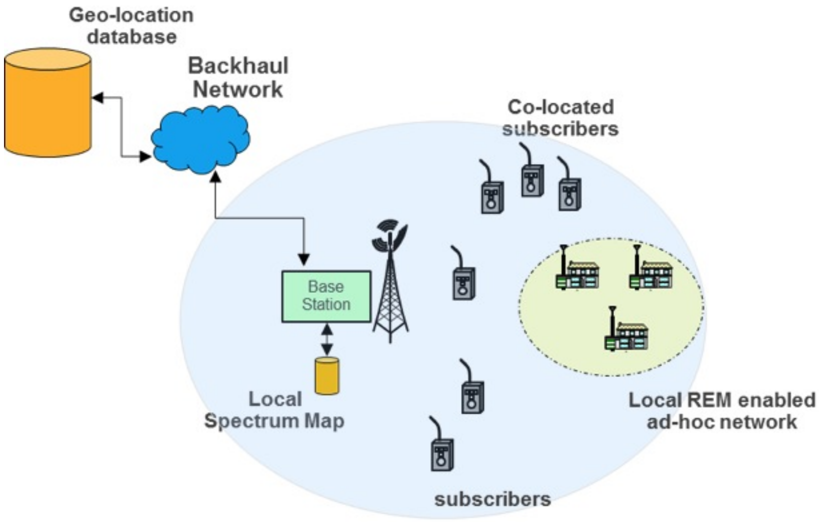

Figure 3.1: Centralized Infrastructure Based Spectrum Sharing with geolocation database.

Table 3.1: measurement plan

Radio	Channel	Measurement	Threshold
1	Ch1	FFT based RSSI	-90 dBm
2	Ch2	Cyclostationary detector	-121 dBm
3	Ch1	FFT based RSSI	-90 dBm
⋮	⋮	⋮	⋮
n	Ch10	Correlation based detector	-70 dBm

Once the desired band is channelized into frequency channels, sufficient number of subscriber radios are selected based on their geographical locations. The base station formulates a spectrum sensing scheme and allocates frequency channels to different sets of radios for sensing. For example, in Figure 3.1, radios that are co-located near each other are assigned different frequencies for sensing to achieve frequency diversity and radios that are further apart are assigned similar frequencies to achieve spatial diversity. Spatial diversity allows better utilization of available radios in the environment for sensing a channel to mitigate shadowing and fading. Frequency diversity is achieved through sensing multiple frequencies by co-located radios.

One way of formulating the measurement plan for each subscriber radio is formulated as in Table 3.1

The radios may respond with raw measurement data or use the thresholds to report presence or absence of the incumbent primary users. The base station combines these results using data fusion rules to achieve a desired level of diversity in frequency, time, space and subscriber radios. The base station maintains a pool of spectrum in its radio environment map database and contents of an example database tables are mentioned below.

Table 3.2: A Database Model for a Single Entity Parameters

S. No	Parameter	Description	Data Structure
Emitter Parameters			
1	Device ID	Unique ID for each radio device	Scalar
2	Modulation	GMSK, BPSK, QPSK, QAM etc	Enumeration
3	Occupied bandwidth	operational frequency band size	Scalar
4	Power (average or peak)	transmit power of the radio	Scalar
5	Power spectral density	Shape of output waveform	Associative array (value, frequency offset)
6	Spurious emission	Out of band spectral emissions	Associative array (value, frequency offset)
7	Data rate	Different data rates supported	Scalar
8	Time access	Continuous vs. pulsed or slotted	String
9	Error correction	Supported FEC mechanisms	Enumeration
Receiver/Sensing Parameters			
1	Bandwidth	frequency band size	Scalar
2	Duration	Sensing and communications duration	Scalar
3	Time Stamp	Time when observation is made	Scalar
4	Type of Detector	Energy/cyclostationarity Correlation based detectors	Enumeration
5	Type of Signal	Signal detection with type of signal	Enumeration
6	Sensed values	Raw or detection output results	Associative array (value, frequency offset)
7	Noise estimate	Noise floor of the environment	Scalar
8	Sample rate	Sampling rate of the receiver	Scalar or
9	Precision	Mathematical precision of the processor	Scalar
10	Position stamp	Location coordinates in Latitude and Longitude units	Enumeration

S. No	Parameter	Description	Data Structure
Antenna Parameters			
1	Maximum gain	Gain of the antenna in dBi units	Scalar
2	Antenna pattern	Isotropic or Directional	Multidimensional associative array (value, frequency offset, polarization)
3	Occupied bandwidth	frequency band size	Scalar
4	Elevation angle	Vertical height covered by the antenna beam measured in degrees	Scalar
5	Azimuth angle	Horizontal area covered by the antenna in degrees	Scalar
6	EIRP	Equivalent Isotropically Radiated Power at the antenna	Scalar
7	Height of antenna above ground	Depends on the device type	Scalar
8	Polarization	Vertical, Horizontal left-hand circular, circular	Enumeration
9	Beam width	Width of the antenna beam	Scalar
10	Number of sectors	Number of sectors in a cell served by the antenna	Scalar
11	Maximum sweep angle	angular area covered by the antenna beams	Scalar
12	Number of elements	total antenna array elements	Enumeration
13	Type of system	Satellite/radar/cellular public safety	Enumeration
Mobility Parameters			
1	Speed	Distance covered by radio in $\frac{m}{sec}$ or $\frac{km}{sec}$	Single value or range
2	Velocity	Distance covered by radio in a particular direction	Scalar or range of directions
3	Acceleration	Velocity change of radio	Value and direction or range
Network Parameters			
1	Device ID	Unique ID for each radio device	Scalar
2	Network ID	Unique ID for each network	Scalar
3	User role	Primary incumbent type i-e subscriber radio, base station, satellite transponder, TV base station	Enumeration
4	Context parameters	stationary, moving indoor, outdoor	Enumeration

3.2 Ad-hoc Networks

Operations in the white space spectrum need network support from the infrastructure of a communication system network. There is no infrastructure support in ad-hoc networks (AHN). So they need to develop their own network support architecture based on inter-node collaboration and sharing of spectrum sensing information. For this purpose, ad-hoc networks need unique functionalities of spectrum sensing, spectrum decision, spectrum sharing and spectrum mobility. These functionalities augmented with access to the geolocation database allow ad-hoc networks to achieve the above-mentioned objectives.

Ad-hoc network architectures are useful in a number of scenarios. Consider a disaster scenario in which the infrastructure of wireless communication networks is severely affected. To setup a replacement network, spectrum bands are required. In such a scenario, spectrum needs to be quickly cleared and reallocated to the disaster response organizations. White space databases are a preferred option for spectrum allocation in such conditions.

In the ad-hoc networks, there is a group of radios distributed in a geographical area. The network topologies can be classified into ad-hoc centralized network and ad-hoc distributed network. In the *ad-hoc centralized network*, there is a master node which coordinates the spectrum management and communications among the radios. While in *distributed ad-hoc networks*, each radio acts independently in its communications and spectrum selection.

For white space operations, network topologies of ad-hoc networks require wireless links to the Internet through a public/private network or a satellite link. In both network topologies, handling of RF channel allocation for such radios, a master node with a back-haul link relays or proxies the database query for the network. So nodes without a back-haul link can query the geolocation database through a master node with a back-haul link. In such cases, the client node is a slave node. The ad-hoc network radios utilize the allocated RF channels from the geolocation database for operations with the required access rules mentioned by the database.

For development of network support for ad-hoc networks a cooperation scheme is needed among the radios. A common cooperative scheme is forming clusters to share sensing information locally. In this case, the cluster head collects sensing measurements from other peers and makes the final sensing decision. For AHN without a central entity, a distributed cooperation method is implemented. In which, when a user detects the PU activity, it should notify its observations promptly to its neighbors to evacuate the busy spectrum. Cooperative detection is more accurate since the uncertainty in a single user's detection can be minimized through collaboration. Moreover, multi-path fading and shadowing effects can be mitigated so that detection probability is improved in a heavily shadowed environment. However cooperative approaches cause adverse effects due to cooperation overhead.

Consider the ad-hoc network Figure 3.2 with a set of wireless nodes that will sense a number of channels and exchange their sensing information with each other. The sensing nodes

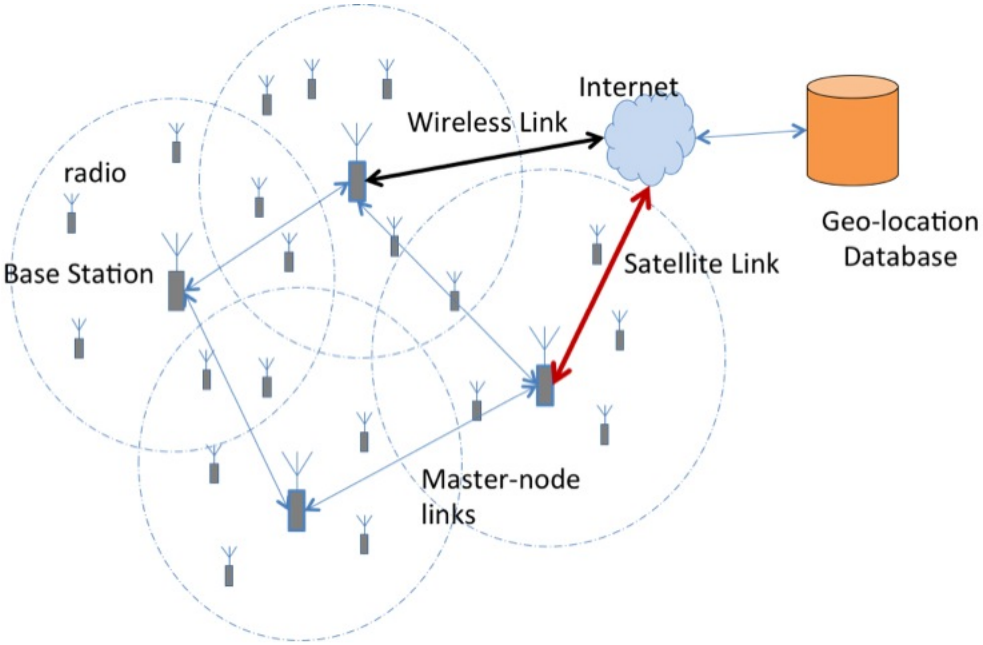

Figure 3.2: Emergency ad-hoc network with partly connected nodes supported by geolocation database.

together will develop a wide-band signal detector, in which nodes will sense a number of channels simultaneously and collaboratively detect PU transmitters in the wide-band. Each node will maintain an embedded radio environment map (REM) database that will store sensing related statistics and parameters. A database enabled sensing scheduling scheme will ensure reliable transmitter detection and will maximize the scanning speed of the whole spectrum band. The goal for the database enabled spectrum sensing system is to reliably and quickly detect existing PU in a wide frequency band. The primary user detection will help the secondary network to characterize the available white space in the desired spectrum bands.

Tables below provide example REM database tables for AHN.

Table 3.3: A Database Model for Ad-hoc Networks

S. No	Parameter	Description	Data Structure
Detected Signals			
1	Signal ID	Numeric ID for each signal detected	Scalar
2	Sensor ID	Numeric ID for each sensor	Scalar
3	Emitter ID	Numeric ID for each Transmitter	Scalar
4	Protocol Name	Name of standard like LTE, WiMAX, GSM, etc	String
5	Time Stamp	Time when measurement was taken	Scalar
6	Frequency	operational frequency	scalar
7	Bandwidth	size of operational band	Scalar
8	RSSI	Received signal strength of the signal	String
Emitters			
1	Emitter ID	Numeric ID for each transmitter	Scalar
2	Protocol Name	Name of standard like LTE, WiMAX, GSM, etc	String
3	Time Stamp	Time when measurement was taken	Scalar
4	Lat	latitude of the Transmitter location	Scalar
5	Lon	Longitude of the Transmitter location	Scalar
6	Frequency	operational frequency	scalar
7	Bandwidth	size of operational band	Scalar
Protocols			
1	Protocol ID	Numeric ID for each signal detected of a protocol	Scalar
4	Protocol Name	Name of standard like LTE, WiMAX, GSM, etc	String

S. No	Parameter	Description	Data Structure
Scan Results			
1	Sensor ID	Numeric ID for each sensor	Scalar
2	Time Stamp	Time when measurement was taken	Scalar
4	FrequencyLo	lower tuning frequency of the sensor	Scalar
5	FrequencyHi	higher tuning frequency of the sensor	Scalar
6	Number signals	Number of signals found during scanning	scalar
Spectrum Opportunity Map			
1	Channel Index	Numeric ID for all the channels	Scalar
2	State	Channel Idle/busy indicator	String
3	Availability	Number of available channels	Enumerations
4	Avg PU Utilization	Average amount of time channel used by PU	Scalar
5	Avg SU Utilization	Average amount of time channel used by SU	Scalar
6	RSSI	Channel power measured in the channel	scalar

The REM database will provide network support to the radios by efficiently representing and storing environmental and operational information. The embedded REM database will include provisions for its real time and historical databases. The level of content at the embedded REM database will include information accessed from the geolocation database and realtime sensing statistics derived from spectrum sensing. This information can be used at the MAC layer for timing requirements of coexistence with other systems. Knowledge needed for non-real time demands such as channel recommendation or spectrum handover by higher layers. Historical spectrum information for initiating a new application or information about spectrum bands from other sources like dedicated sensor network or regulatory authority.

Chapter 4

Database Access Protocol

Recently, a lot of research and regulatory activities [10], [11], [20], [27] are going on in many countries around the world in the use of the white-space spectrum. It's an understanding among the community that a common interface between white space devices and database must be defined for utilizing the unused spectrum. The Internet Engineering Task Force (IETF) is currently working on development of a Protocol for Access to Whitespace Spectrum (PAWS) [28], for accessing the geolocation database. The standard protocol defines the database interface to have the following attributes.

4.1 IETF PAWS Attributes

Global Applicability: This common interface must be conceptually similar to the Internet Domain Naming System (DNS) and would allow users to find the suitable spectrum for their application, either on a secondary or an unlicensed basis. This ease of access would enable greater opportunistic use of the RF spectrum.

Spectrum Agnostic: The protocol should be spectrum independent and able to be used in any spectrum band where white space sharing is permitted. Devices can operate in any location where such spectrum is available, and a common interface ensures uniformity in implementations and deployment.

*Regulatory Support:*To allow the global use of white space devices in different countries (whatever the regulatory domain), the protocol should support the database communicating applicable regulatory rule set information to the white space device.

Flexible and extensible data structures: Different databases are likely to have distinct requirements for all kinds of data needed for registration (different regulatory rule sets that apply to the registration of devices), and other messages sent by the device to the database. For instance, different regulators might require distinct device-characteristic information to be passed to the database.

4.1.1 IETF PAWS Primitives with Geolocation Database

Database discovery

The master device must obtain the address of a trusted database [29], which it will query for available white-space spectrum. If the master device uses a discovery service to locate a trusted database, it may perform the following steps (this description is intended to be descriptive, not prescriptive):

1. The master device constructs and sends a request (e.g., over the Internet) to a trusted discovery service.
2. If no acceptable response is received within a pre-configured time limit, the master device concludes that no trusted database is available. If at least one response is received, the master device evaluates the response(s) to determine if a trusted database can be identified where the master device is able to receive service from the database. If so, it establishes contact with the trusted database.

Optionally, and in place of steps 1-2 above, the master device can be pre-configured with the address (e.g., URI) of one or more trusted databases. The master device can establish contact with one of these trusted databases.

Device Registration

The master device must register with the database before it queries the database for available spectrum. A registration process may consist of the following steps:

1. The master device sends registration information to the database. This information may include the device ID, serial number assigned by the manufacturer, device location, device antenna height above ground, name of the individual or business that owns the device, and the name, street and email address, and telephone number of a contact person responsible for the device's operation.
2. The database responds to the registration request with an acknowledgement to indicate the success of the registration request or with an error if the registration was unsuccessful. Additional information may be provided by the database in its response to the master device.

Protocol

A protocol that enables a white space device to query a database to obtain information about available spectrum is needed. A device may be required to register with the database with some credentials prior to being allowed to query. The requirements for such a protocol are specified in this document.

Data Model Definition

A data model is required which enables the white space device to query the database while including all the relevant information such as geolocation, radio technology, power characteristics, etc., which may be country and spectrum and regulatory dependent. All databases are able to interpret the data model and respond to the queries using the same data model that is understood by all devices.

4.2 Database Query and Response Protocol

Following are the primitives of the PAWS protocol for accessing the geolocation database.

1. Master device discovers the geolocation database
2. Master device establishes HTTPS connection with database
3. Master device initialization message to database to exchange capabilities
4. Database responds
5. Device registration
6. Device sends available spectrum request to database
7. Master on behalf of a slave. So slave also has to register itself through master device
8. Database responds with available spectrum response
9. Master device spectrum usage info to the database
10. Database responds by spectrum usage acknowledgement response message

Details of the protocol operations and messages exchanged between the white space device (WSD) and the geolocation database are available in the IETF PAWS Documents [28].

Content of a database query

In this section, the parameters in the spectrum request are briefly summarized in this table.

Table 4.1: Database Query Contents

Parameter	Content	Data Structure
database query		
Device descriptor		
	Manufacturer serial number	string
	ruleset ID	list
	FCC ID	string
	Device type	string
	Radio Access Technology (RAT)	string
geolocation	Point, Region, Center	
	Latitude, Longitude	
Antenna characteristics	antenna height,antenna type	list
	antenna direction, radiation pattern	
	antenna gain, antenna polarization	
Device owner		vcard
Device capabilities		
	authority	string
	Max location change	float
	Max polling sec	int
	Rule set IDs	list

Content of a database response

After receiving a spectrum request, the database responds with a spectrum response. The contents of the spectrum response are in the table below.

Table 4.2: Database Response Contents

Parameter	Content	Data Structure
database query		
Time stamp		
	start time	string
	stop time ID	list
Spectrum schedule		
	Bandwidth	
	frequency range	
	Bandwidth	
Spectrum report	antenna height	list
	antenna type	
	antenna direction	
	radiation pattern	
	antenna gain	
	antenna polarization	
rule set ID		vcard
Device capabilities		
	authority	string
	Max location change	float
	Max polling sec	int
	Rule set IDs	list
Location	geolocation and	
	selected from the spectrum request	
maxLocationChange	renew the spectrum request	

Details of the primitives exchanged between the WSD and the geolocation database are available in the IETF PAWS Documents [28].

Chapter 5

Wireless Standardization Activities Using Geolocation Database

5.1 FCC 3.5 GHz NPRM and Citizens Broadband Services Database

FCC released a Notice of Proposed Rule Making (FCC NPRM- 12-148) [30] to create a new Citizens Broadband Service in the 3550-3650 MHz band (3.5 GHz Band) currently used for military and satellite operations. It will promote two major advances that enable more efficient use of the radio spectrum: small cells and spectrum sharing.

The 3.5 GHz Band was identified by the National Telecommunications and Information Administration (NTIA) for shared federal and non-federal use in the 2010 Fast Track Report [9]. Current FCC's proposal builds on experience with spectrum sharing in the television white spaces (TVWS), and prepares ideas for the new notice of Inquiry on Dynamic Spectrum Access technologies, and broadly reflects recommendations made in a recent report by the Presidents Council of Advisors on Science and Technology (PCAST) [1].

Moreover, these proposed new and flexible rules can be extended to the neighboring 3650-3700 MHz band, which is already used for commercial broadband services. Together, these proposals would make up to 150 megahertz of a contiguous spectrum available for innovative mobile and fixed wireless broadband services without displacing mission-critical incumbent systems.

5.2 European Communication Commission (ECC) and Ofcom Efforts

Efficient use of the radio spectrum is a global regulatory goal from Europe, to Canada or Singapore with a primary focus on what spectrum sharing can add to existing spectrum management options. European Commissions recent release [31] on promoting the shared use of radio spectrum and the interest from regulators such as the UKs Ofcom [32] and Singapore‘s IDA [33] in trials and commercial pilots of the White Spaces technology, ensures that a significant portion of spectrum is available for license-exempt sharing on a nationwide basis, increasing the amount of such spectrum in the urbanized areas within these countries.

The reason is simple: the number of wireless devices is growing exponentially. According to the European Commission statistics [34], by 2015 there will be 7.1 billion phones, tablets and other mobile devices connected to the Internet globally. Five years further down the line, the number of smart devices connected to the Internet is going to be staggeringly larger particularly when wireless sensors and other machine-to-machine communication devices are counted. Unfortunately, the way spectrum is managed today, does not readily allow the flexibility to adapt to meet that projected demand. Gaps in coverage and network overload in busy areas are already resulting in poor service for end-users.

These problems can be alleviated through spectrum sharing, as proposed in the European Commissions Communication [31]. Fortunately, the technology is now ready to make this happen. A Commission funded project called COGEU [35], which brings together research institutes and private companies from Portugal, France, Ireland, Germany, Poland, Slovakia, Greece and Cyprus, has lately set out to quantify the white spaces sharing opportunities in key central European states. In August, Europe‘s another commercially-authorized TV White Space geolocation database established in Finland [36]. Ireland is on the cusp of launching its own White Spaces initiative [37], and the French regulator has recently granted a TV white space test license.

With this growing interest and investment in White Spaces technology, it has become increasingly clear that its benefits can be much broader than just ensuring good wireless broadband services. Trials and pilots have explored a range of solutions from increasing the affordability of the Internet and enabling a machine to machine (M2M) communication, to turn a highly populated metropolitan area into a leading Smart City with substantial long-term environmental benefits. A popular use case is looking at how spectrum resources can help overcome inadequate Internet access in remote rural areas.

5.3 IEEE Standardization Activities

IEEE standards are key drivers for standardization and implementation of novel communications system concepts. A number of IEEE standards have incorporated the use of geolocation database as a source of available white space channels for operation. Some of these standards are summarized below.

IEEE 802.11 af Standard

IEEE 802.11af working group [38] has been set up to define a standard to implement WiFi technology within the TV unused spectrum, or TV white space. Research and standardization groups around the world are taking a more flexible approach to spectrum allocations, the idea of low power systems that are able to work within portions of RF spectrum that may need to be kept clear of high-power transmitters to ensure coverage areas do not overlap are being seriously investigated. IEEE 802.11af that use TV white space, the overall system must not cause interference to the primary users. There are many benefits from a system such as IEEE 802.11af from using TV white space.

Propagation characteristics: In view of the fact that the 802.11af WiFi system operating the TV white spaces would use frequencies below one GHz. This would allow for greater distances to be achieved. Current Wi-Fi systems use frequencies in the ISM bands - the lowest band is 2.4 GHz and here signals are easily absorbed.

Additional bandwidth: One of the advantages of using TV white space is that unused frequencies can be accessed. However, it will be necessary to aggregate several TV channels to provide the bandwidths that Wi-Fi uses on 2.4 and 5.6 GHz, to achieve the required data throughput rates. IEEE 802.11 af uses many new technological concepts of cognitive radio and dynamic spectrum access with sensing and database access technologies.

Cognitive Radio: To utilize the available spectrum as efficiently as possible, there is a need to utilize radio technology that can sense the environment and configure itself accordingly - Cognitive Radio. The technology is heavily dependent upon Software Defined Radio technology as the radio needs to be configurable according to the prevailing radio environment.

Database Access: The IEEE 802.11af systems have provisions for accessing the geolocation database to get available spectrum for use with the available channel powers mentioned in the spectrum response.

IEEE P1900 Standard

The IEEE P1900 standard gives a complete overview of a cognitive radio network (CRN). The standard provides basic definitions and interconnection between cognitive radio and cognitive radio network concepts. For example, informative tables and diagrams explain the relationships of CRs, software-controlled radios, intelligent radios, and adaptive radios. The P1900.1 definitions and terminology are categorized into (1) definitions of advanced radio system concepts, (2) definitions of radio system functional capabilities, (3) definitions of network technologies that support advanced radio system technology (4) spectrum management definitions, and (5) a glossary of ancillary definitions.

The P1900 committee objective is to develop supporting standards dealing with new technologies and techniques being developed for next-generation radio and advanced spectrum management The different tasks [39] for the working groups are defined below.

1. IEEE P1900.1: Terminology and Concepts for Next-Generation Radio Systems and Spectrum Management,

   http://grouper.ieee.org/groups/dyspan/1/index.htm

2. IEEE P1900.2: Recommended Practice for Interference and Coexistence Analysis,

   http://grouper.ieee.org/groups/dyspan/2/index.htm

3. IEEE P1900.3: Recommended Practice for Conformance Evaluation of Radio (SDR) Software Modules,

   http://grouper.ieee.org/groups/dyspan/3/index.htm

4. IEEE P1900.4: Architectural Building Blocks Enabling Network-Device Distributed Decision Making for Optimized Radio Resource Usage in Heterogeneous Wireless Access Networks,

   http://grouper.ieee.org/groups/dyspan/4/index.htm

5. IEEE P1900.5: Policy Language and Policy Architectures for Managing Cognitive Radio for Dynamic Spectrum Access Applications,

   http://grouper.ieee.org/groups/dyspan/5/index.htm

6. IEEE 1900.6 Working Group on Spectrum Sensing Interfaces and Data Structures for Dynamic Spectrum Access and other Advanced Radio Communication Systems,

   http://grouper.ieee.org/groups/dyspan/6/index.htm

7. IEEE 1900.7 White Space Radio Working Group,

   http://grouper.ieee.org/groups/dyspan/7/index.htm

IEEE 802.22 Standard

The IEEE 802.22 WG was formed to use the frequencies in the TV-band between 54 MHz and 862 MHz at 6, 7, or 8 MHz bandwidths. The standard defines a cognitive radio based wide-area regional network (WRAN) that contains cognitive radio devices that can sense the immediate spectrum. The 802.22 WRAN is the first IEEE standard [40] to define how cognition in radios can be used in the base station and user terminals in a regional area network.

Communications between fixed point-to-multipoint environment with specific use of television channels and guard bands was considered in the WG specifications. Specifically, the primary goal was to develop a standard for a CR-based PHY/MAC air interface for use in license-exempt wireless communication devices on a non-interfering basis with a TV broadcast spectrum.

Moreover, deployment in different geographic areas, including sparsely populated rural areas, while preventing harmful interference to incumbent licensed services in the TV broadcast bands. A secondary objective for the 802.22 standard is to serve dense population areas where spectrum is available.

The 802.22 standard have the capability for accessing the geolocation database for available specrum in its operational area. The specification defines protocol for accessing the database from the base station perspective as well as the individual CPE device.

Chapter 6

Existing Privacy-Preserving Models and Methods

In this chapter we introduce some of the most well known privacy-preserving techniques that are used widely in different types of databases.

6.1 k-anonymity

The concept of k-anonymity was originally introduced in the context of relational data privacy [41]. It addresses the question of ”How can a data holder release its private data with guarantees that the individual subjects of the data cannot be identified while the data remain practically useful” [42]. For instance, a medical institution may want to release a table of medical records with the names of the individuals replaced with dummy identifiers. However, some set of attributes can still lead to identity breaches. These attributes are referred to as the $quasi-identifier$. For instance, the combination of zip code, age and the nationality attributes in the disclosed table 6.1 can uniquely determine an individual. By joining such a medical record table with some publicly available information source, like a voters list table, the medical information can be easily linked to individuals. k-anonymity prevents such privacy breach by ensuring that each individual record can only be released if there are at least $k-1$ distinct individuals whose associated records are indistinguishable from the former in terms of their quasi-identifier values. Table 6.2 illustrates an example of 4-anonymous impatient microdata for the data in table 6.1. It has been shown that the problem of optimal k-anonymization is NP-hard [89], nevertheless, the problem can be solved quite effectively by the use of a number of heuristic methods.

In Location-Based Services (LBSs) and mobile clients, location k-anonymity refers to k-anonymous usage of location information. A subject is considered location k-anonymous if and only if the location information sent from a mobile client to an LBS is indistinguishable

Table 6.1: Impatient Microdata

ID	Zip Code	Age	Nationality	Condition
1	24060	22	American	Heart Disease
2	24061	25	Indian	Heart Disease
3	24061	29	American	Viral Infection
4	24060	23	American	Viral Infection
5	28255	42	Chinese	Cancer
6	28231	51	Ameican	Heart Disease
7	28255	45	Indian	Viral Infection
8	28231	47	Chinese	Viral Infection
9	24054	31	American	Cancer
10	24054	33	Chinese	Cancer
11	24021	33	American	Cancer
12	24021	39	Indian	Cancer

Table 6.2: 4-anonymous Impatient Microdata

ID	Zip Code	Age	Nationality	Condition
1	240**	< 30	*	Heart Disease
2	240**	< 30	*	Heart Disease
3	240**	< 30	*	Viral Infection
4	240**	< 30	*	Viral Infection
5	282**	≥ 40	*	Cancer
6	282**	≥ 40	*	Heart Disease
7	282**	≥ 40	*	Viral Infection
8	282**	≥ 40	*	Viral Infection
9	240**	$3*$	*	Cancer
10	240**	$3*$	*	Cancer
11	240**	$3*$	*	Cancer
12	240**	$3*$	*	Cancer

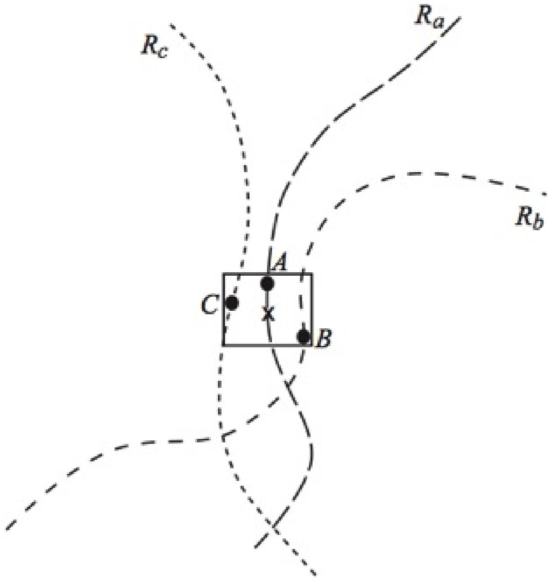

Figure 6.1: k-anonymity for location privacy.

from the location information of at least $k-1$ other mobile clients. In order to do this, each user needs to find regions that are called cloak region that at least $k-1$ other users exist in this region, and instead of its exact location, send this region to the LBS [43].

Following the terminology introduced in [44], location privacy is defined in two levels: microscopic and macroscopic. Microscopic location privacy is defined as the users' location privacy on a small scale, i.e., corresponding to a single query, and reflects how accurately the adversary can infer the users' locations after observing their individual queries, given his a priori knowledge. Macroscopic location privacy represents users' privacy on a large scale, e.g., given multiple (possibly correlated) queries from users as they move.

Even using k-anonymity for location privacy of mobile clients (e.g. smartphones) of LBSs has already been proven ineffective at the macroscopic level [45], [46]. Also note that in the context of geolocation databases in cognitive radio network, since instead of revealing the exact location of PUs, the database only responds to SUs with a maximum transmission power level, it is providing some kind of spatial cloacking for primary users. In other words, finding the exact location of the primary user, using the maximum transmission power value, is not possible, and this value only reveals a possible region (usually modeled as a circle around the SU) that the primary user is possibly located in this region.

Table 6.3: 3-diverse impatient microdata

ID	Zip Code	Age	Nationality	Condition
1	240**	< 40	*	Heart Disease
9	240**	< 40	*	Cancer
10	240**	< 40	*	Cancer
4	240**	< 40	*	Viral Infection
5	282**	≥ 40	*	Cancer
6	282**	≥ 40	*	Heart Disease
7	282**	≥ 40	*	Viral Infection
8	282**	≥ 40	*	Viral Infection
2	240**	< 40	*	Heart Disease
3	240**	< 40	*	Viral Infection
11	240**	< 40	*	Cancer
12	240**	< 40	*	Cancer

6.2 l-diversity

The l-diversity model was designed to handle some weaknesses in the k-anonymity model since protecting identities to the level of k individuals is not the same as protecting the corresponding sensitive values, especially when there is homogeneity of sensitive values within a group [47]. For example in the 4-anonymous medical database of table 6.2, we can see a user with zipcode $606 * **$ and age $3*$, has cancer (sensitive data) with probability 1. In other words although the identity of user is not revealed, but the sensitive data is inferred from non-sensitive data. To protect users' sensitive data against this privacy problem, the concept of intra-group diversity of sensitive values is promoted within the anonymization scheme. Table 6.3 shows 3-diversity for the medical database in table 6.1 while preserving 4-anonymity.

l-diversity has also been used for preserving privacy in location-based services [48], in order to mitigate query homogeneity attacks [49].

6.3 t-closeness

The t-closeness model is a further enhancement on the concept of l-diversity. One characteristic of the l-diversity model is that it treats all values of a given attribute in a similar way irrespective of its distribution in the data. This is rarely the case for real data sets, since the attribute values may be very skewed. This may make it more difficult to create feasible l-diverse representations. Often, an adversary may use background knowledge of the global distribution in order to make inferences about sensitive values in the data. Furthermore, not all values of an attribute are equally sensitive. For example, an attribute corresponding to

Table 6.4: Confidence Bounding

Job	Gender	Age	Condition
Professional	Male	[35-40)	Hepatitis
Professional	Male	[35-40)	Hepatitis
Professional	Male	[35-40)	HIV
Artist	Female	[30-35)	Flu
Artist	Female	[30-35)	HIV
Artist	Female	[30-35)	HIV
Artist	Female	[30-35)	HIV

a disease may be more sensitive when the value is positive, rather than when it is negative. In [79], a t-closeness model was proposed which uses the property that the distance between the distribution of the sensitive attribute within an anonymized group should not be different from the global distribution by more than a threshold t. The Earth Mover distance (a.k.a Wasserstein metric) is used in order to quantify the distance between the two distributions [50]. Furthermore, the t-closeness approach tends to be more effective than many other privacy-preserving data mining methods for the case of numeric attributes.

6.4 Confidence Bounding

Wang et al. [51] considered bounding the confidence of inferring a sensitive value from a quasi-identifier (QID) group by specifying one or more privacy templates of the form, $(QID \rightarrow s, h)$; s is a sensitive value, QID is a quasi-identifier, and h is a threshold. Let $Conf(QID \rightarrow s)$ be $\max\{conf(qid \rightarrow s)\}$ over all qid groups on QID, where $conf(qid \rightarrow s)$ denotes the percentage of records containing s in the qid group.

A table satisfies $(QID \rightarrow s, h)$ if $Conf(QID \rightarrow s) \leq h$. In other words, $(QID \rightarrow s, h)$bounds the attacker's confidence of inferring the sensitive value s in any group on QID to the maximum h. For example in the 3-anonymous and 2-diverse table 6.4, with $QID = \{$Job, Sex, Age$\}$, $(QID \rightarrow \text{HIV}, 10\%)$ states that the confidence of inferring HIV from any group on QID is no more than 10%. But for the data in the table, this privacy template is violated because the confidence of inferring HIV is 75% in the group {Artist, Female, [30-35)}.

6.5 Personalized Privacy

Not all individuals or entities are equally concerned about their privacy. For example, a corporation may have very different constraints on the privacy of its records as compared to an individual. This leads to the natural problem that we may wish to treat the records in a given data set very differently for anonymization purposes. The notion of personalized privacy is proposed [52] to allow each record owner to specify her own privacy level. This model assumes that each sensitive attribute has a taxonomy tree and that each record owner specifies a guarding node in this tree. The record owner's privacy is violated if an attacker is able to infer any domain sensitive value within the subtree of her guarding node with a probability, called breach probability, greater than a certain threshold.

Although both confidence bounding and personalized privacy take an approach to bound the confidence or probability of inferring a sensitive value from a QID group, they have differences. In the confidence bounding approach, the data publisher imposes a universal privacy requirement on the entire data set, so the minimum level of privacy protection is the same for every record owner. In the personalized privacy approach, a guarding node is specified for each record by its owner. The advantage is that each record owner may specify a guarding node according to her own tolerance on sensitivity. Experiments show that this personalized privacy requirement could result in lower information loss than the universal privacy requirement

6.6 (X, Y)-Privacy

As we described earlier, k-anonymity in states that each group of insensitive attributes X, has at least k distinct values on sensitive values Y (e.g., diseases). However, if some Y values occur more frequently than others, the probability of inferring a particular Y value can be higher than $\frac{1}{k}$. To address this issue, Wang and Fung [53] proposed a general privacy model, called (X, Y)-Privacy, which combines both k-anonymity and confidence bounding. The general idea is to require each group x on X to contain at least k records and $conf(x \rightarrow y) \leq h$ for any $y \in Y$, where Y is a set of selected sensitive values and h is a maximum confidence threshold.

6.7 Perturbation

The perturbative masking method (a.k.a randomization method) is a technique for privacy-preserving databases that uses data distortion in order to mask the attribute values of records. In this method, we add sufficiently large noise to individual record values to prevent recovery of these values by an adversary. One key advantage of the randomization method is that it is relatively simple, and does not require knowledge of the distribution of other records in the data, and can be implemented at data collection time. This is not true of other methods such as k-anonymity which require the knowledge of other records in the data. We describe some of the perturbative methods that can be used for our problem in this section. In [54], it is shown that perturbation-based mechanisms can outperform privacy-preserving techniques that use cloaking regions for location based services. We need

Table 6.5: Students grades relational database

Student Number	Name	Age	Grade
101111	Lewis	18	78
101112	Venus	19	89
101201	Carl	18	92
101205	Mary	20	82
101206	Alice	20	80

to compare various perturbation-based mechanisms with other existing privacy-preserving techniques in the context of our problem to see which one has better performance.

6.7.1 Additive Noise

Additive noise is the most basic perturbative method that can be used for privacy-preserving databases. There are four noise addition algorithms in the literature:

1. Masking by uncorrelated noise addition: The vector of observations x_j for the j-th attribute of the original dataset X_j is replaced by a vector $z_j = x_j + \epsilon_j$, where ϵ_j is a vector of normally distributed errors and $Cov(\epsilon_t, \epsilon_l) = 0$ for all $t \neq l$.

2. Masking by correlated noise addition: Similar to the previous masking method with the difference that $Cov(\epsilon_t, \epsilon_l)$ could be nonzero.

3. Masking by noise addition and linear transformation. In this method after adding noise to the data, we perform an additional linear transformation to ensure the sample covariance matrix of the masked attributes is an unbiased estimator for the covariance matrix of the original attributes.

4. Masking by noise addition and nonlinear transformation. In this method we combine simple additive noise to the data with a non-linear transformation. This method preserves more information but is time-consuming and requires expert knowledge on the data set.

It is obvious that additive noise is not suitable to protect categorical data. As an example os applying additive noise consider the relational table 6.5. Perturbing the grades (sensitive data) in this database using the simple uniform noise function $N = (\text{Add}, 5)$ results in the perturbed database in table 6.6.

Table 6.6: Students grades relational database

Student Number	Name	Age	Grade
101111	Lewis	18	83
101112	Venus	19	94
101201	Carl	18	97
101205	Mary	20	87
101206	Alice	20	85

Table 6.7: Students grades relational database

Student Number	Name	Age	Grade
101111	Lewis	18	92
101112	Venus	19	80
101201	Carl	18	82
101205	Mary	20	78
101206	Alice	20	89

6.7.2 Data Swapping

Data swapping is another type of perturbative method, in which the values across different records are swapped in order to perform the privacy preservation. Note that since this technique does not allow the value of a record to be perturbed independently of the other records, it cannot be used for preserving operational privacy of primary users in geolocation database-driven cognitive radio networks.

Table 6.7 illustrates applying this technique to the relational database in table 6.5.

6.7.3 Rounding

Rounding methods replace original values of attributes with rounded values. For a given attribute X_i, rounded values are chosen among a set of rounding points defining a rounding set (often the multiples of a given base value). Table 6.8 illustrates applying this technique to the relational database in table 6.5.

6.8 Differential Privacy

Instead of comparing the prior probability and the posterior probability before and after accessing the published data, Dwork [55] proposed differential privacy to compare the risk with and without the record owner's data in the published data. In other words differential privacy is a condition on the release mechanism and not on the dataset.

Table 6.8: Students grades relational database

Student Number	Name	Age	Grade
101111	Lewis	18	80
101112	Venus	19	90
101201	Carl	18	90
101205	Mary	20	80
101206	Alice	20	80

Assume that the responses of the database are modeled via a randomized algorithm A. The randomized algorithm A is ϵ-differentially private if for all datasets D_1 and D_2 that differ on a single element (i.e., data of one person), and all $S \subset \mathit{Range}(A)$, $\Pr\{A(D_1) \in S\} \leq e^{\epsilon} \times \Pr\{A(D_2) \in S\}$, where the probability is taken over the coins of the algorithm and Range(A) denotes the output range of the algorithm A. This means that for any two datasets which are close to one another (that is, which differ on a single element) a given differentially private algorithm A will behave approximately the same on both data sets.

Chapter 7

A General Threat Model

Depending on characteristics of the secondary users' network, primary users' network and the services provided for secondary users by the geolocation database, the operational privacy of primary users can be threatened by the adversary in different ways. The first question that we need to answer for creating a threat model is: "who is the adversary?". The adversary can be a secondary user (querier) or a group of colluding secondary users that are located throughout the region that the geolocation database covers, or alternatively an entity who eavesdrops on communications between SUs and the geolocation database (GDB).

In our framework, the adversary is actually the entity who observes the output of privacy-preserving mechanisms and hence has access to a subset of observable events O. The subset of O that is accessible to a given adversary is called the set of *observed event* by that adversary and is denoted by $\hat{O} \subseteq O$. The properties of this subset and the implication of this observation on the PUs, operational privacy depend on the characteristic of the adversary. Note that the adversary might have multiple observation points (e.g. sending queries from different locations and observing the GDB's replies), from each of which she can observe a different set of events. At each observation point, the adversary observes different transformations of the same actual events, nevertheless the structure of transformations (not their settings) is the same. Thus in this section, we focus on the set of adversary's observed events at a single observation point, which is shown by $\hat{O}$.

In this framework, we model an adversary based on the following three factors: her *Means*, *Actions*, and *Goals*.

7.1 Means

The means of the adversary are the technologies available to er for capturing events, her access credentials in the system, and her a priori knowledge about the system.

7.1.1 Access

The adversary might eavesdrop on the communication channel between SUs and the GDB or she might be a secondary user herself. In both cases, the adversary has access to the content of the SU's queries and the GDB's corresponding replies to these queries. However, in the former case, based on the level of sophistication of the adversary's eavesdropping devices, the accuracy of observed events changes. We also assume that the adversary has access to spectrum sensing technology and may combine local spectrum sensing results with the information that she obtains from the GDB's replies to the SUs' queries to pose a more serious privacy threat to PUs.

7.1.2 Knowledge

The a priori knowledge of the adversary is composed of multiple pieces. Here, we categorize the adversary's knowledge into multiple classes. The precision and confidence of the adversary's knowledge about each class determines her a priori knowledge. Her knowledge in each class can be deterministic or probabilistic and this should be clarified in each threat model.

Users. The adversary might know the (exact or estimated) number of primary users at any time, or more precisely the set of primary users U, that implies knowing the real identity of active primary users. This knowledge can evolve over time, or she may remain oblivious about the dynamics of the set of primary users and their joining/leaving. The adversary might also have some background knowledge about the capabilities of primary users that operate on a specific channel. Her knowledge about PUs may include their transmission power range, protected contour, antenna characteristics, employed wireless technology (e.g. LTE), required SINR, etc.

Identities. This class specifies to what extent the adversary knows about the primary users' identities and the pseudonyms used by them. The adversary might know the relation between pseudonyms of a user, and also the constraints on the set of pseudonyms (e.g., how many pseudonyms a user can have). The extent to which each pseudonym is linkable to its holder's real-identity is also part of the knowledge of the adversary in this class.

Space. The knowledge of the adversary on the space in which primary users move falls into this class. This knowledge consists of three layers. The first layer models the geographical space in which users can move, in a discrete way, using a grid of cells that represents the two dimensional space. Each cell is called a *location instance* and has a unique identity. The second layer models the places or location sites such as roads, houses, etc. Each location site consists of a non-empty set of location instances and has a unique name and address. The third layer captures the type of location sites and their similarity. A location type may represent the usage of the location, e.g., military, residential, or landscape type, e.g. lake, mountain, forest, etc. The connection between users and places must also be specified here. For example, does the adversary know the address of the origin and destination of the moving primary users.

Events. The adversary might have access to some actual events that are performed before the observation time. Moreover, in many cases the adversary has some statistics about the typical behavior of primary users. For example, she knows the (im)possibility or the probability that one specific actual event can be performed by a user, or that two specific events belong to the same user. Knowledge of the adversary about mobility profile of users (which represents how probable/possible it is for a specific user or a mass of users to move from one location to another location in a specific time period) falls into this class.

In addition to these factors, we assume that the adversary knows the application, employed privacy tools, and also the privacy metric that the geolocation database uses.

7.2 Actions

The action scope of an adversary is determined by the size of the location areas and the duration of time periods in which the adversary observes the system. Considering these factors, adversaries can consequently be divided into different categories. An adversary is *global* if she observes the observable events occurred at any location in the space, i.e. she can query the geolocation database from any location instance. Whereas, she is called *local* if during the observation period she cannot observe the transformation of some events that are generated in specific location areas. Similarly, based on the observation time, an adversary is referred to as a *short-term* attacker if the transformation of events performed at some time periods are not observable by the adversary. In the case there is no such time restrictions, she is called a *long-term* attacker. In the case an attack is global and long-term, we have $\hat{O} = O$.

7.3 Goals

7.3.1 Presence vs. Absence Disclosure

An adversary's goals of observing users' activities in a mobile network can be divided into two main categories: presence disclosure or absence disclosure. In the former category, the adversary's goal is to find out if a set of primary users are present at some place(s). Whereas, in the latter category, the adversary wants to know whether a specific set of primary users are not present at some place(s). Usually the adversary is interested in presence disclosure, thus most of the attacks presented in the literature of location privacy fall into the first category. However, there are some reports about the consequences of absence disclosure attacks on some primary users such as ordinary people. As an obvious example, by misusing her access to a geolocation database, the adversary can find out the best time to break into a person's house. Therefore we need to obfuscate the location of primary users to protect their location privacy against both presence and absence disclosure attacks.

7.3.2 Individual vs. Mass Target

The inference attacks against a geolocation database can disclose the private information of a specific primary user in an *individual target* attack, or it might be targeting a set of primary users, collectively, in a *mass target* attack where the adversary does not distinguish primary users in the set, for example when they belong to a community.

7.3.3 Tracking vs. Identification

The two main known attacks on primary users' location privacy, which are used usually to disclose primary users' presence, are *tracking* and *identification*. These two attacks are tightly related to each other, although they have different ways of obtaining primary users' private location information.

In tracking attacks, the adversary's goal is to reconstruct the primary users' actual trajectories (which might have been distorted by privacy preserving mechanisms) and subsequently identify the locations that the primary users have visited. This information can also be used to predict the future locations of primary users. The tracking can be done in various manners depending on the adversary's goal. The adversary might want to know the trace of location instances (i.e., coordinates) visited by the primary users in a given time period along with his average speed, direction etc., or the location sites (e.g., specific military base) where they have been to, or only the type of places that the primary users are used to periodically visit in order to detect the type of primary users.

In *identification* attacks, the adversary wants to discover the real identity of her targets. This can be done on a small scale where the adversary is interested in de-anonymizing a specific observed event, or on a large scale where the adversary is interested in finding the identity of users from whom the adversary has observed some anonymous traces of events. The identification is done using some inference attacks based on the adversary's knowledge on the linkability of the primary users to sensitive areas such as their homes or work places. Identification can also leverage on the mobility pattern of users, because primary users tend to visit certain places regularly.

It is clear that the success of each of the two above-mentioned attacks also paves the way for the other. In the case the adversary manages to discover the actual trajectory of a user, the identification of the user is not a difficult task. Especially if the adversary has access to the information about location sites such as homes or work places of the users, which contain a lot of information about their identities. In the case that the adversary has already de-anonymized some events of a user, the recovery of the user's actual trajectory (i.e., tracking him) can be done more easily, if the adversary has access to the mobility profile of the users.

7.4 Using the threat model for evaluating the privacy-preserving techniques

We can use the described threat model in this chapter to evaluate different privacy-preserving techniques that are designed for the GDB. To evaluate a privacy-preserving technique, we need to apply different attacks against it and find the attacker's cost and its impact.

To find the attacker's cost in an attack, one needs to know the required means (e.g. amount of knowledge and access level), and type of actions (for instance the cost of a short-term attack should be defined differently from the cost of a long-term attack). This won't be possible without applying the threat model to the attack precisely.

On the other hand in order to measure the attacker's impact or its success rate, we need to use the threat model to learn about the attacker's goals and proposing a well-defined metric for the attacker's impact.

Defining attacker's cost and impact using the threat model will help us in evaluating privacy-preserving techniques against different types of attacks which in tern results in designing more robust techniques.

Chapter 8

Applying Existing Privacy-Preserving Models and Methods to Our Problem

8.1 *k*-anonymity

In the context of location privacy of PUs in database-driven spectrum sharing, we can achieve location k-anonymity by combining protected contours of k primary users that are closest together and creating a larger protected contour that works like the cloak box for LBSs. The secondary users are not allowed to transmit in the area covered by this larger protected contour. This increases the privacy of primary users but will decrease the performance of the secondary network. The main problem of this approach is that unlike mobile clients of LBSs in metropolitan areas, PUs that operate on the channels that may contain spectrum holes, usually are not close enough to create a small cloak box that covers k PUs, such that k is large enough for preserving privacy and the cloak box is small enough to avoid causing significant degradation in spectrum utilization efficiency. Figure 8.1 illustrates this idea for $k = 3$. Further research and experimental results on this method is required.

Consider the following simplified scenario. Suppose table 8.1 shows the location and the radius of the protected contour of the 6 primary users in figure 8.1. Also assume that all PUs are operating in the same band. To answer a SU's query, the database checks if the SU is located inside the protected contour of a PU or not.

To implement 3-anonymity, we replace the following table with table 8.1, where PU1' and PU2' are replaces with the set of PUs {PU1,PU2,PI3} and {PU4,PU5,PU6} respectively. The location of the virtual PU that is replaced with 3 actual PUs is the circumcenter of the triangle that consists of the location of the 3 PUs (the circumcenter is located at the same distance from all three points). And the radius of the protected contour is the distance of this point from one of the PUs plus the largest radius of protected contour of the three PUs.

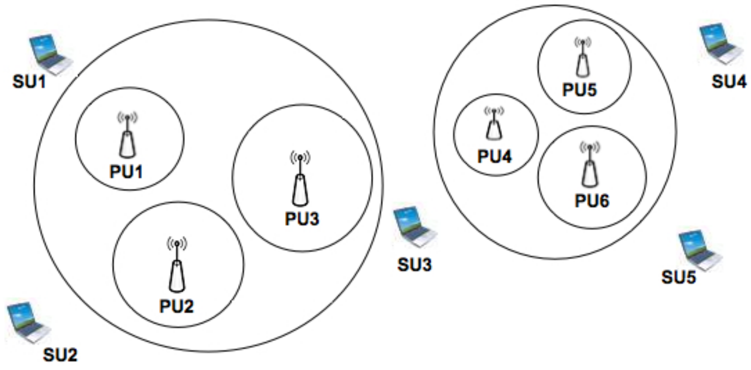

Figure 8.1: 3-anonymity for primary users' location privacy.

Table 8.1: Simplified Geolocation Database.

ID	2D Location	Radius of Protected Contour (km)
PU1	(1,2)	0.2
PU2	(3,-3)	0.25
PU3	(5,1)	0.3
PU4	(9,2)	0.1
PU5	(12,3)	0.15
PU6	(12,1)	0.2

Table 8.2: Geolocation Database with 3-anonymity.

ID	2D Location	Radius of Protected Contour (km)
PU1'	(2.56,-0.28)	3.06
PU2	(10.67,2)	1.86

Note that although 3-anonymity by spatial cloaking preserves the privacy of PUs, it results in poor spectral efficiency for secondary network, because the SUs cannot transmit in some regions that are not part of the protected contour of real PUs.

8.2 l-diversity

In the context of privacy of PUs in database-driven spectrum sharing, we can achieve l-diversity by providing l different answers (for example l different combinations of power values and associated availability time) to the same query, i.e. queries that are sent by a specific SU from a specific location. This seems to degrade the performance of secondary users' network significantly. Further research is required to find out if this method will perform better in preserving the privacy level of primary users.

8.3 Confidence Bounding

In the context of geolocation database, we can implement confidence bounding, by bounding the number of repetitions of different answers that database gives to the same query. For example if the database replies to the SU at location (x,y) with 3 different power values 40mW, 50mW and 80 mW, and 80mW is the true value for the current deployment of PUs, by replying with 80mW twice for 10 identical queries, the database bounds the confidence of the attacker to 20% for the correct value.

8.4 (X, Y)-Privacy

In the context of geolocation databases, this technique can be implemented similar to what we described for confidence-bounding technique.

8.5 Perturbation

8.5.1 Additive Noise

In the context of PUs location privacy, we can use different perturbative methods. For example we can add noise to the radius of the protected contour of PUs and then compute maximum transmission power and report it to the querier (masking by noise addition and nonlinear transformation) or we can use false positives such as reducing maximum allowed transmission power (masking by uncorrelated noise addition) in the reply to SUs queries.

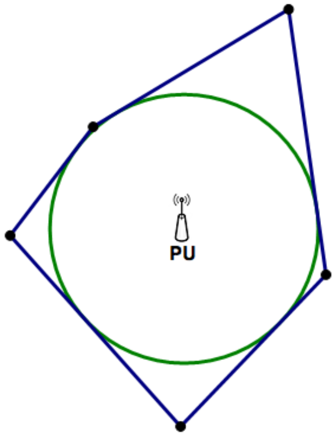

Figure 8.2: Perturbation using transfiguration.

8.5.2 Rounding

In the context of location privacy of PUs in geolocation databases, the database may round the maximum transmission power value which it obtains from the propagation model that it uses. Therefore instead of the exact distance between the querier (SU) and PU, the SU can only obtain an interval for the possible distances, i.e. less information about the location of the PU.

8.5.3 Transfiguration

Another form of perturbation that might be used to preserve the privacy of primary users, is changing the shape of protected contour of PUs. Replacing the circular protected contours with random convex shapes that inscribe the actual circular protect contour will increase the location privacy of PUs. Figure 8.2 illustrates using a pentagon as the protected contour of a PU, that inscribes the real circular shape protected contour of the PU.

8.6 Differential Privacy

In the context of geolocation databases, we can define ϵ-differential privacy as follows: Assume that two queries that are sent by a SU, only differ in the location of the SU. In other words, the identity of the SU, the closest PU to the SU, and all other parameters are fixed and the only change that has happened is the distance between the SU and the PU that is changed from d_1 to d_2. We can say that the database has differential privacy if for $|d_1 - d_2| < \delta$: $\Pr\{f(d_1) - f(d_2) \neq 0\} < \epsilon$, where $f(d)$ is the database's response (e.g.

maximum allowed transmission power) for the distance d.

Chapter 9

Conclusions

This report drafts the deficiencies of the existing geolocation database implementations for spectrum sharing to achieve the recommendations made in a number of studies [1], [7], [9], conducted for identifying the 500MHz of spectrum as mentioned in the National Broadband Plan (NBP) and tiered approach to spectrum access proposed by the President‘s Council of Advisors on Science and Technology (PCAST) Report.

To fulfill this gap, the requirements for such a database system along with some new functionality are developed. A generalized dynamic database model is proposed in Chapter 2. It consists of many components that will allow a number of unique capabilities like real time channel availability from dynamic calculations of the protection contour zones of stationary and mobile primary users. Interference protection and coexistence capability to calculate the secondary network interference power spectral density (IPSD) at primary incumbent's location for allowing operations inside the protection zones as well. It can provide deconfliction among the secondary users for achieving coexistence of multiple wireless standards.

In Chapter 3, some use case scenarios for centralized infrastructure based networks, and Adhoc networks are outlined. Example database implementation tables are provided that can accommodate locally identified spectrum opportunities along with the resources requested from the geolocation database.

The IETF PAWS protocol for accessing the geolocation database is summarized in chapter 4. The protocol provides a common interface between white space devices and the geolocation database for utilizing the unused spectrum. The attributes of the protocol and primitives exchanged with the database are outlined.

Wireless standardization activities around the world in different regulatory bodies are summarized in chapter 5. IEEE standards like IEEE 802.11af, IEEE P1900, IEEE802.22, etc. that use the geolocation database for requesting spectrum resources are also summarized in this chapter.

This report also provides an overview to existing privacy-preserving techniques for protecting sensitive data in a relational database, and how these techniques can be utilized for preserving operational privacy of incumbents in a geolocation database.

Privacy-preserving data mining finds numerous applications in surveillance which are naturally supposed to be privacy-violating applications. The key is to design methods for databases, which continue to be effective, without compromising privacy. Privacy-preserving techniques have been used for many applications such as bio-surveillance, facial de-identification, identity theft, etc. In chapter 6, we studied an overview of the state-of-the-art in privacy-preserving data mining techniques.

In order to evaluate different privacy-preserving techniques that are designed for the geolocation database, we introduced a threat model in chapter 7. This threat model helps us to find the attacker's cost and its impact, for all privacy-preserving techniques.

Finally in chapter 8, we investigated the problem of preserving the operational privacy of incumbents in a geolocation database-driven cognitive radio network. We showed how we can use the existing privacy-preserving techniques for this application in a high level discussion.

Bibliography

[1] PCAST, "Realizing the full potential of government-held spectrum to spur economic growth," tech. rep., Report to the President, 2012.

[2] F. C. Commission, "Third order and memorandum opinion and order, in the matter of unlicensed operation in the tv broadcast bands, additional spectrum for unlicensed devices below 900 mhz and in the 3 ghz band," April 2012.

[3] D. Gurney, G. Buchwald, L. Ecklund, S. Kuffner, and J. Grosspietsch, "Geo-location database techniques for incumbent protection in the tv white space," in *IEEE Dyspan*, 2008.

[4] F. C. Commission, "Enabling innovative small cell use in 3.5 ghz band nprm & order (fcc 12-148)," December 2012.

[5] NSF, "Enhancing access to the radio spectrum (ears)." Program Solicitation NSF 13-539.

[6] DARPA, "Baa: Shared spectrum access for radar and communications (ssparc)." DARPA-BAA-13-24, February 2013.

[7] FCC, "Connecting america: The national broadband plan," tech. rep., White House, 2010.

[8] NTIA, "Spectrum sharing working group (sswg) working notes." `csmac_spectrum_sharing_workinggroup_notes_03feb2012_v3.pdf`, 2012. Commerce Spectrum Management Advisory Committee (CSMAC).

[9] "An assessment of the near-term viability of accommodating wireless broadband systems in the 1675-1710 mhz, 1755-1780 mhz, 3500-3650 mhz, 4200-4220 mhz, and 4380-4400 mhz bands," tech. rep., NTIA:Fast Track Report, 2010.

[10] IEEE, "P1900.6 standard for spectrum sensing interfaces and data structures for dynamic spectrum access and other advanced radio communication systems," tech. rep., 2011.

[11] "Ieee 802.22-2011(tm):standard for cognitive wireless regional area networks (ran) for operation in tv bands," tech. rep., 2011.

[12] FCC, "Fcc media bureau databases for am, fm, tv, lptv, and dtv broadcast stations." `http://transition.fcc.gov/mb/databases/cdbs/`. Parameters of Devices in Different Applications.

[13] "Fcc tv database query." `http://www.fcc.gov/encyclopedia/tv-query-broadcast-station-search`.

[14] J. Stine, "Model-based spectrum management: Loose coupling spectrum management and spectrum access," in *New Frontiers in Dynamic Spectrum Access Networks (DySPAN), 2011 IEEE Symposium on*, pp. 628–631, 2011.

[15] USGS, "National map." `http://nationalmap.gov/viewer.html`.

[16] "Ntia-irregular terrain model (itm) (longley-rice)." `http://www.ntia.doc.gov/report/1982/guide-use-its-irregular-terrain-model-area-prediction-mode`.

[17] NTIA, "Commerce spectrum management advisory committee final report working group 1 –1695-1710 mhz meteorological-satellite." `www.ntia.doc.gov/files/ntia/publications/wg-1_report_v2.pdf`.

[18] D. Gurney, G. Buchwald, L. Ecklund, S. Kuffner, and J. Grosspietsch, "Geo-location database techniques for incumbent protection in the tv white space," in *New Frontiers in Dynamic Spectrum Access Networks, 2008. DySPAN 2008. 3rd IEEE Symposium on*, pp. 1–9, IEEE, 2008.

[19] "Fcc guidelines for channel calculations for white spaces guidelines." `www.ic.gc.ca/eic/site/smt-gst.../SMSE-012-11-WSDBA-annex1.pdf`.

[20] "Further definition of technical and operational requirements for the operation of white space devices in the band 470-790 mhz," tech. rep., Complementary Report to ECC Report 159, 2013.

[21] "Audit of methods for calculating white space spectrum availability." `stakeholders.ofcom.org.uk/binaries/.../uhf.../Real_Wireless_audit.pdf`.

[22] K. Harrison, "Cognitive radios in the tv whitespaces: challenges and opportunities," Master's thesis, EECS Department, University of California, Berkeley, Dec 2011.

[23] I. T. U. (ITU), "P.1546 : Method for point-to-area predictions for terrestrial services in the frequency range 30 mhz to 3 000 mhz." `http://www.itu.int/rec/R-REC-P.1546-4-200910-I/en`. Online, accessed 06-18-2013.

[24] FCC, “Fm and tv propagation curves graphs.” http://transition.fcc.gov/Bureaus/MB/Databases/FM_TV_DTV_Propagation_Curves_Graphs/FM_TV_DTV_propagation_curves_graphs.html. Online, accessed 25-01-2013.

[25] NTIA, “Ntia planning and processes need strengthening to promote the efficient use of spectrum by federal agencies.” http://www.gao.gov/products/GAO-11-352. Online, accessed 25-01-2013.

[26] NTIA, “Report of the spectrum management improvements working group.” http://www.ntia.doc.gov/files/ntia/meetings/spectrum_management_improvements_report_10nov2011.pdf. Online, accessed 06-25-2013.

[27] FARAMIR, “Faramir a european fp7 project for radio environment mapping.” http://www.ict-faramir.eu/. Online, accessed 06-25-2013.

[28] IETF, “Protocol to access ws database (paws).” http://datatracker.ietf.org/wg/paws/. Online, accessed 06-25-2013.

[29] “Ietf paws database discovery.” http://datatracker.ietf.org/doc/draft-wei-paws-database-discovery/. Online, accessed 06-25-2013.

[30] “Enabling innovative small cell use in 3.5 ghz band nprm and order.” http://www.fcc.gov/document/enabling-innovative-small-cell-use-35-ghz-band-nprm-order. Online, accessed 06-25-2013.

[31] “On technical conditions regarding spectrum harmonisation for terrestrial wireless systems in the 3400-3800 mhz frequency band.” www.cept.org/.../ecc/.../ECC(12)052-Annex15_Draft-CEPT-Report-49. Online, accessed 06-25-2013.

[32] “Ofcom tv whitespace databases.” http://stakeholders.ofcom.org.uk/spectrum/tv-white-spaces/white-spaces-pilot/TV-White-Space-Database/. Online, accessed 06-25-2013.

[33] “Information development authority of singapore.” http://www.ida.gov.sg/. Online, accessed 06-25-2013.

[34] “Delivering the digital dividend.” http://ec.europa.eu/digital-agenda/en/delivering-digital-dividend. Online, accessed 06-25-2013.

[35] “Cognitive radio system for efficient sharing of tv white spaces in european context.” http://www.ict-cogeu.eu/. Online, accessed 06-25-2013.

[36] “Fairspectrum.” http://www.fairspectrum.com/propagating-thoughts. Online, accessed 06-25-2013.

[37] “White spaces ireland.” http://whitespacesireland.wordpress.com/2012/10/22/first-tv-white-space-database-for-ireland/. Online, accessed 06-25-2013.

[38] "Ieee 802.11af standard." `http://grouper.ieee.org/groups/802/11/`. Online, accessed 06-25-2013.

[39] B. Fette, *Cognitive Radio Technology.* Elsevier Science, 2009.

[40] "Ieee std 802.22-2011tm, standard for wireless regional area networks part 22: Cognitive wireless ran medium access control (mac) and physical layer (phy) specifications: Policies and procedures for operation in the tv bands." `http://www.ieee802.org/22/`. Online, accessed 06-25-2013.

[41] P. Samarati and L. Sweeney, "Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression," in *IEEE Symposium on Research in Security and Privacy*, 1998.

[42] L. Sweeney, "k-anonymity: A model for protecting privacy," *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, vol. 10, no. 5, pp. 557–570, 2002.

[43] B. Gedik and L. Liu, "Protecting location privacy with personalized k-anonymity: Architecture and algorithms," *IEEE transaction on mobile computing*, vol. 7, no. 1, 2008.

[44] R. Shokri, J. Freudiger, and J.-P. Hubaux, "A unified framework for location privacy," Tech. Rep. EPFL-REPORT-148708, EPFL, 2010.

[45] B. Hoh, M. Gruteser, H. Xiong, and A. Alrabady, "Preserving privacy in gps traces via uncertainty-aware path cloaking," in *ACM Symposium on Information, Computer and Communication Security*, 2007.

[46] R. Shokri, J. Freudiger, M. Jadliwala, and J.-P. Hubaux, "A distortion-based metric for location privacy," in *Workshop on Privacy in the Electronic Society*, 2009.

[47] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "l-diversity: Privacy beyond k-anonymity," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 1, no. 1, 2007.

[48] F. Liu, K. Hua, and Y. Cai, "Query l-diversity in location-based services," in *International Conference on Mobile Data Management: Systems, Services and Middleware*, pp. 436–442, 2009.

[49] Z. Xiao, J. Xu, and X. Meng, "p-sensitivity: A semantic privacy-protection model for location-based services," in *International Workshop on Privacy-Aware Location-Based Mobile Services*, 2008.

[50] Y. Rubner, C. Tomasi, and L. Guibas, "A metric for distributions with applications to image databases," in *ICCV*, pp. 59–66, 1998.

[51] K. Wang, B. Fung, and P.Yu, "Handicapping attacker's confidence: An alternative to k-anonymization," *Knowledge and Information Systems (KAIS)*, vol. 11, no. 3, pp. 345–368, 2007.

[52] X. Xiao and Y. Tao, "Personalized privacy preservation," in *ACM SIGMOD international conference on Management of data*, pp. 229–240, 2006.

[53] K. Wang and B. Fung, "Anonymizing sequential releases," in *12th ACM SIGKDD Conference*, 2006.

[54] R. Dewri, "Local differential perturbations: Location privacy under approximate knowledge attacker," *IEEE transaction on mobile computing*, 2012.

[55] C. Dwork, "Differential privacy," *Automata, languages and programming*, pp. 1–12, 2006.