Leveraging P2P Interactions for Efficient Location Privacy in Database-driven Dynamic Spectrum Access

Erald Troja[1], Spiridon Bakiras[2]
(Corresponding author: Erald Troja)

The Graduate Center, City University of New York[1]
365 Fifth Avenue, New York, NY, 10016, USA
John Jay College, City University of New York[2]
524 West 59th Street, New York, New York 10019, USA
(Email: etroja@gc.cuny.edu, sbakiras@jjay.cuny.edu)
(Received Sept. 11, 2014; revised and accepted Jan. 16 & May 15, 2015)

Abstract

Dynamic spectrum access (DSA) is a novel communication paradigm that enables wireless clients to utilize statically allocated radio channels that are currently idle. Specifically, in the database-driven DSA model, clients learn their geographic location through a GPS device and use this location to retrieve a list of available channels from a centralized white-space database. To mitigate the potential privacy threats associated with location-based queries, existing work has proposed the use of private information retrieval (PIR) protocols when querying the database. Nevertheless, PIR protocols are very expensive and may lead to significant costs for highly mobile clients. In this paper, we propose a novel method that allows wireless users to collaborate in a peer-to-peer (P2P) manner, in order to share their *cached* channel availability information that is obtained from previous queries. To preserve location privacy against other users, we leverage an *anonymous veto* protocol that anonymizes the exchange of information among a group of users. Our experimental results with a real-life dataset show that our methods reduce the number of PIR queries by 50% to 60%, while incurring low computational and communication costs.

Keywords: Computer-communications networks, database applications, database management, distributed systems, GIS, spatial databases

1 Introduction

Radio spectrum is governed by federal agencies, mainly through a static sharing strategy. However, with the exponential increase of mobile devices capable of Internet connectivity, static spectrum sharing has led to a spectrum shortage. To this end, dynamic spectrum access (DSA) is a novel communication paradigm that enables wireless clients to utilize statically allocated radio channels when not in use by their licensed owners. DSA is accomplished through cognitive radio (CR), an intelligent wireless communication system that is aware of its operating spectral environment [33].

Currently, there exist two main approaches by which CR nodes acquire their spectral knowledge. On one end lies the distributed and cooperative sensing method, where nodes perform sheer power detection and coordination [18, 45, 47, 48]. On the other end lies the database-driven model, where nodes learn their spectral surroundings by querying a centralized white-space database (WSDB). In this latter model, mobile clients no longer bear the task of fusing spectral knowledge regarding the surrounding geographical area. Instead, they simply send their GPS coordinates to the database and receive the centrally fused repository report for that area.

On May 2012, the FCC issued a ruling [7] that obsoletes the distributed and cooperative sensing methods for white-space TV bands, thus requiring all CR nodes to implement the database-driven approach. Nevertheless, this approach suffers from severe location privacy leakage. According to the FCC specifications [8], a mobile TV band device (TVBD) must access the WSDB for a list of available channels, every time it is activated from a power-off state. Furthermore, a mobile TVBD must issue a new query to the WSDB whenever it moves further than 100m from its previous location. Given that the GPS coordinates are part of the query, the WSDB operator can easily build a detailed history of a mobile TVBD's trajectories. These trajectories would allow the WSDB to infer sensitive information about the mobile user, such as home location, health condition, lifestyle habits, political

and religious affiliations, etc.

To mitigate the potential privacy threats associated with location-based queries, existing work has proposed the use of private information retrieval (PIR) protocols when querying the database [10]. A PIR protocol enables a user to retrieve a record from a database server, while maintaining the identity of the record secret from the server. However, PIR protocols are very expensive and may lead to significant costs in the case of highly mobile clients that issue numerous queries throughout their trajectories. For example, the trivial PIR case is to download the entire database (e.g., once per day), which clearly preserves privacy but incurs an overwhelming communication cost.

Typical PIR protocols offer a trade-off between computational and communication complexity. Computational complexity has an adverse impact mostly at the server side, whereas communication complexity affects the end-user as well (especially in the case of wireless devices). For instance, the scheme by Trostle and Parrish [40] that is applied in previous work [10] is computationally efficient, but its communication cost is equal to a large percentage of the database size. On the other extreme, Gentry and Ramzan's protocol [11] is considered to attain the best communication complexity, but incurs a high computational cost due to its heavy use of cryptographic operations [35].

In this paper, we argue that any location privacy method for the database-driven DSA model is bounded by the limitations of the underlying PIR protocol. As such, it is desirable to identify new mechanisms for users to acquire the necessary spectral knowledge. Our intuition is that, in a white-space TV band network, mobile TVBD users will gradually develop a trajectory-specific spectrum knowledge *cache*, through a series of PIR requests. In the extreme case, some users might opt to download the entire WSDB (trivial PIR case) before initiating their travel[1]. Therefore, we propose that mobile users that are within communication range interact in a peer-to-peer (P2P) manner, in order to exchange their cached spectrum knowledge for the surrounding area.

However, a user's spectrum knowledge cache is a summary of his/her recent trajectory, and some users may be unwilling to share that information due to privacy concerns. To this end, we leverage the *anonymous veto network* (AV-net) protocol of Hao and Zieliński [17] that anonymizes the exchange of information among a group of users. Our experimental results with Microsoft's GeoLife trajectory dataset [49] show that our methods reduce the number of PIR queries by 50% to 60%, while incurring low computational and communication costs for the mobile clients.

The rest of this paper is organized as follows. Section 2 presents a literature review on location privacy and Section 3 provides the necessary background on the cryptographic primitives utilized in our methods. Section 4 describes the details of our P2P protocol and Section 5 presents the results of the experimental evaluation. Finally, Section 6 concludes our work.

[1]However, due to its overwhelming communication cost, the trivial PIR case may be infeasible for most users.

2 Related Work

Most existing protocols on location privacy build upon the notion of k-anonymity [38] or l-diversity [31]. A spatial query is considered to be k-anonymous if it is indistinguishable from at least $k-1$ other queries spawning from the same area, usually called the spatial cloaking region (SCR). The SCR is chosen such that it encapsulates the querying user as well as at least $k-1$ other users. To compute the SCR, existing k-anonymity algorithms typically extend the SCR around the query point until it encapsulates $k-1$ other users [16, 34, 44].

On the other hand, l-diversity based methods, such as the ones proposed in [42, 43], extend the SCR until $l-1$ different locations are included in the query. Both k-anonymity and l-diversity offer some degree of location privacy, but they are susceptible to semantic location information leakage. For example, if the SCR only contains casinos, the server can infer (to a certain degree) that the mobile user is interested in gambling. To this end, the work of Lee et al. [27] attempts to provide location privacy using location semantics.

The k-anonymity and l-diversity based approaches, as well as collaborative location privacy protection methods [9, 41], often rely on third party trusted *anonymizers*, an expensive and scarcely available option. Ghinita et al. [12] propose the first privacy-preserving protocol that does not require an anonymizer. They focus on nearest neighbor queries and introduce a method that achieves perfect location privacy via the PIR protocol of Kushilevitz and Ostrovsky [26].

Other protocols on location privacy revolve around the notion of data perturbation, location hiding, and the introduction of data point dummies. Meyerowitz et al. [32] introduce a data perturbation technique to protect personal location data against untrusted location based service (LBS) servers. In their work, they develop CacheCloak, a protocol that enables real time anonymization of location data. CacheCloak relies on a trusted anonymizing server to generate mobility prediction from historical data, and then submit intersecting predicted paths simultaneously to the LBS. Reliance on a trusted server is a very expensive and strong assumption that we would like to avoid in our proposed methods. Also, the intuition behind CacheCloak is to obscure the user's path by surrounding parts of it with other user's paths, effectively creating a k-anonymous region.

Huang at al. [20] study the problem of location privacy preservation with respect to an LBS that threatens a user's location privacy by tracking transmitting frames. The authors argue that correlation attacks between a node's old and new address are not sufficient. They suggest the concept of a *silent period*, defined as the transi-

tion period between the use of new and old pseudonyms, during which a node is not allowed to disclose neither the old nor the new address.

Furthermore, Huang at al. [21] extend their previous work [20] and study the problem of location privacy with respect to a user's communication with network access points. They mainly focus on the issue of how location privacy enhancements affect the perceived Quality of Service (QoS). The authors propose a silent cascade method to enhance a user's location privacy by trading end-to-end delay for anonymity. They abstract silent cascade as a mix-network model and evaluate its performance. In our setting, however, we are concerned with the effectiveness (computation and communication cost) of the location privacy-preserving protocol itself.

Kido et al. [25] suggest a location privacy-preserving method that uses the notion of *dummy* data (false positives), in order to hide the user's true location from the LBS. The authors argue that, after sending their GPS coordinates to the LBS, users can not delete or modify their disclosed location. In other words, users cannot prevent the service providers from analyzing motion patterns using stored location data. In their proposed method, users send their true location data along with several false ones (dummies) to the service provider, who subsequently creates a reply message for each received data point. Users then simply extract the correct information from the reply messages. However, it is clear that this scheme is essentially a k-anonymity based approach.

Similarly to Kido et al. [25], Lu at al. [30] introduce PAD, a method that injects dummy locations in the query, which are generated according to either a virtual grid or a circle. The virtual grid or circle cover the user's actual location, and their spatial extents are controlled by appropriate generating algorithms. However, PAD relies on a server-side front-end, in order to be integrated into existing client/server mobile service systems. Even though PAD takes into account the number of location points in the query, as well as the area of the region covered by those points, it can be effectively reduced to a pure k-anonymity based technique.

Other techniques such as routing anonymization and privacy-preserving wireless broadcast networks have been suggested. The authors in [2] suggest wireless anonymous routing (WAR) as the main approach of achieving anonymity in a wireless broadcast network. Ref. [50] and [28] propose lightweight ad hoc routing protocols in order to preserve location privacy of the mobile nodes. Lastly, the authors in [22] provide evidence that such anonymization and location privacy-preservation techniques can be applied even in radio frequency identification networks (RFID). Such techniques are orthogonal to our proposed methods and can be applied in an optional and complementary fashion in order to provide local network addressing anonymity as well as geo-location privacy.

Until recently, location privacy work in the dynamic spectrum access domain has mainly focused on the collaborative spectrum sensing model. In particular, most existing solutions attempt to protect the location privacy of mobile users that submit sensing reports to a fusion center [19, 29, 37]. The collaborative sensing and reporting approach was embraced as a superior method compared to the centralized database approach. This is no longer the case, though, at least in the white-space TV band realm.

Due to the recency of the FCC's ruling (May 2012), location privacy research in database-driven DSA networks is still in its infancy. The state-of-the-art protocol is due to Gao et al. [10], which builds upon a modified version of Trostle and Parrish's PIR scheme [40]. They assume a fixed grid of $n \times n$ cells, where each cell contains a bitmap that represents the channel availability information (typically 32 bits). Nevertheless, their scheme incurs a high communication cost of $(2n + 3) \cdot \log p$ bits, where p is a 2048-bit modulus. For example, if $n = 5000$, the amount of data transmitted to retrieve the bitmap of a single cell is 2.5 MB, which is approximately 2.6% of the whole database size. For highly mobile clients, the cost of this approach can exceed the cost of the trivial PIR case.

3 Preliminaries

In this section, we give a brief description of the cryptographic primitives incorporated in our methods. Section 3.1 provides some background on anonymous communication and Section 3.2 introduces the 2-round anonymous veto network (AV-net) of Hao and Zieliński [17]. Section 3.3 presents the threat model of our approach.

3.1 Anonymous Communication

Research on anonymous communication has evolved due to the *dining cryptographers* problem, introduced by Chaum in 1988 [5]. Essentially, a dining cryptographers network (DC-net) allows groups of $n > 2$ participating users to contribute their boolean bits towards a boolean-OR calculation of some statement, while preserving the privacy of the individual inputs. DC-nets have many weaknesses and are considered impractical due to complex key setup, message collisions, and vulnerability to disruptions. Alternatively, circuit evaluation techniques, such as the ones proposed in [13, 46], can also be used towards the secure computation of a boolean-OR function. However, as pointed out by Brandt [4], the circuit evaluation technique is expensive and impractical.

A similar problem is the anonymous veto network (AV-net), which allows groups of $n > 2$ participating users to vote against a given statement. In the setting of a white-space TV bands database, where channel availability can be represented via a boolean bit, a sample statement might be: "none of the group members knows that the channel is free." If any of the users in the group anonymously vetoes the statement, it means that "at least one of the users in the group knows that the channel is free."

Unlike DC-nets, AV-net protocols do not require secret channels in order to exchange messages. Furthermore, they have no message collisions and are very resistant to disruptions. Nevertheless, all existing AV-net protocols assume the existence of an *authenticated broadcast channel*, which is easily implemented using digital signatures [5]. Several such anonymous veto protocol designs exist in the literature [4, 15, 17, 24]. In our work, we leverage the 2-round AV-net protocol of Hao and Zieliński [17], because it is more efficient in terms of number of rounds, computation, and communication cost.

3.2 The 2-Round AV-net Protocol

Setup. All users participating in the protocol agree on two public parameters, namely G and g. G is a finite cyclic group of prime order q in which the Decision Diffie-Hellman (DDH) problem is hard [3], and g is a generator of G. These values are fixed and used in all protocol invocations. Subsequently, each participant P_i, $i \in \{1, 2, \ldots, k\}$, selects a random secret value $x_i \in_R \mathbb{Z}_q$.

Round 1. In the first round, every participant P_i broadcasts g^{x_i}. When the first round completes, each participant P_i computes

$$g^{y_i} = \prod_{j=1}^{i-1} g^{x_j} \Big/ \prod_{j=i+1}^{k} g^{x_j}$$

Round 2. In round 2, every participant P_i broadcasts a value $g^{c_i y_i}$ where c_i is either x_i or a random value $r_i \in_R \mathbb{Z}_q$, depending on whether participant P_i vetoes the statement or not.

$$g^{c_i y_i} = \begin{cases} g^{r_i y_i} & \text{if } P_i \text{ sends '1' (veto),} \\ g^{x_i y_i} & \text{if } P_i \text{ sends '0' (no veto).} \end{cases}$$

In order to test the final result, all participants compute $\prod_i g^{c_i y_i}$. If nobody vetoed the statement, then $\prod_i g^{c_i y_i} = \prod_i g^{x_i y_i} = 1$, since $\sum_i x_i y_i = 0$. If, however, one or more participants vetoes the statement by sending a '1', we have $\prod_i g^{c_i y_i} \neq 1$.

3.3 Threat Model and Security

In this work, we are not concerned with privacy against the WSDB operator. We assume that mobile users, when needed, query the WSDB through a standard PIR protocol. Instead, in our methods, the adversary is one or more users in the group that executes the AV-net protocol, or any eavesdropper that monitors the exchange of messages over the wireless channel. The adversary runs in polynomial time, and its goal is to identify a user that vetoes a certain statement.

Note that, in the case of malicious adversaries, the protocol described above necessitates zero-knowledge proof (ZKP) schemes, such as Schnorr's signature [36]. In particular, during each round, users must demonstrate knowledge of their own secret values, such as x_i and c_i. Nevertheless, in our work, we assume the *honest-but-curious* adversarial model, i.e., all users follow the protocol correctly but try to gain an advantage by examining the communication transcript. As a result, we do not implement zero-knowledge proofs.

Our methods inherit the security of the underlying AV-net protocol [17]. As such, they are *semantically secure* [14], i.e., it is infeasible to derive any information about a mobile client's input, given its published values and the public parameters. The security is based on the DDH assumption. Therefore, an eavesdropper is unable to determine whether a user has vetoed a statement. Our methods are also secure against *partial* collusions, i.e., when some participants collude (by revealing their secret values) to determine the input of a certain user. As explained in [17], only a *full* collusion against a single user can compromise security, i.e., when $k-1$ users reveal their values to identify whether the k-th user vetoed a statement.

4 P2P Protocol

In this section, we present the P2P protocol that allows a group of users to share anonymously their cached spectrum information. Section 4.1 describes the system architecture, and Section 4.2 explains the protocol initiation process. Section 4.3 presents the criteria for mobile nodes to participate in this protocol, and Section 4.4 describes the group formation mechanism. Finally, Section 4.5 introduces the details of the AV-net protocol invocation.

4.1 System Architecture

Similar to previous work [10], we assume a fixed grid of $n \times n$ cells, where mobile users can communicate through white-space TV bands, while maintaining their location privacy. According to the FCC specifications [8], each cell is 100m×100m in size, and users may need to query the WSDB whenever they move into a cell with no prior spectrum availability knowledge. The dimensions of the grid (i.e., n) can be made arbitrarily large, which has a direct effect on the database size.

Note that, mobile TVBDs are allowed to communicate only in the frequency ranges 512-608 MHz (TV channels 21-36) and 614-698 MHz (TV channels 38-51), i.e., there are a total of 31 possible white-space TV band channels that can be accessed in a DSA manner. Therefore, we represent the daily channel availability as 32 bits (per cell), where bit 0 represents a busy channel and bit 1 represents an idle channel. As an example, when $n = 5000$, the WSDB is 100 MB in size. The trivial PIR case is impractical in this setting, since it involves downloading the entire WSDB. This would take approximately 35 mins on 3G networks[2] [39].

[2]Furthermore, communication over a cellular network is a priced resource that should be avoided.

In our model, we assume an out-of-band common control channel (CCC) through a dedicated transceiver. This enables mobile users to exchange concurrently both control and data messages. Out-of-band CCC coordination can be realized over the 802.11 protocol in *ad-hoc* mode or through any of the methods proposed in [1, 6]. We emphasize that 802.11 is not a viable protocol for long range communications, hence it is only used to implement the out-of-band CCC for communications within a 100m×100m cell.

The FCC's white-space TV band DSA specifications state that "A mode II personal/portable device may load channel availability information for multiple locations around, i.e., in the vicinity of, its current location and use that information in its operation." Accordingly, in our methods, we assume a PIR protocol that retrieves channel information for multiple cells with a single query[3]. As a proof of concept, we consider a fixed grouping of the available cells into 4×4 blocks. Therefore, we assume that each PIR query retrieves the 16-cell block that contains the user's current cell.

Figure 1a shows an example of this approach. The black colored cells signify the locations where a new PIR query is issued, due to lack of spectrum availability knowledge. The alternating white and grey colored cells identify the different blocks, with the block *id* shown in the lower-left corner of the block. Note that, even though we assume a specific method for querying the WSDB, our protocol is *orthogonal* to the underlying PIR query/reply structure. Any WSDB indexing method is a viable candidate for our protocol, but for the PIR reply to be of some utility to the client, the retrieved cells should be spatially close to the user's location.

As illustrated in Figure 1a, each of the three mobile TVBDs gradually builds a spectrum knowledge *cache* containing channel availability information from their respective trajectories. When the users eventually meet at the diagonally striped cell, it may be beneficial to all of them to exchange their cached information. To maximize utility for all participating users, the sharing of spectrum information involves the area surrounding the current location (as users may continue their trajectories towards any direction). In particular, the TVBD nodes agree on the number of surrounding rings (AR) that they wish to explore during the protocol invocation. (Table 1 summarizes the symbols used in the remainder of this paper, along with the values tested in the experimental evaluation.) In the example of Figure 1b, $AR = 3$, and the explored region is shown in a darker shade.

To illustrate the location privacy leakage from a *plaintext* exchange of spectrum availability information (i.e., without the invocation of the AV-net protocol), consider the example of Figure 1b. We can infer that u_1 arrived at the current cell through block 9, while u_3 visited blocks 13 and 14. On the other hand, u_2's trajectory contains some uncertainty, as u_2 may have arrived at the current cell through blocks 2, 7, or 12. Furthermore, if two users participate in the same group multiple times (at different locations), they can derive more information about each other's movement patterns.

[3]All existing PIR schemes can retrieve multiple records from a database.

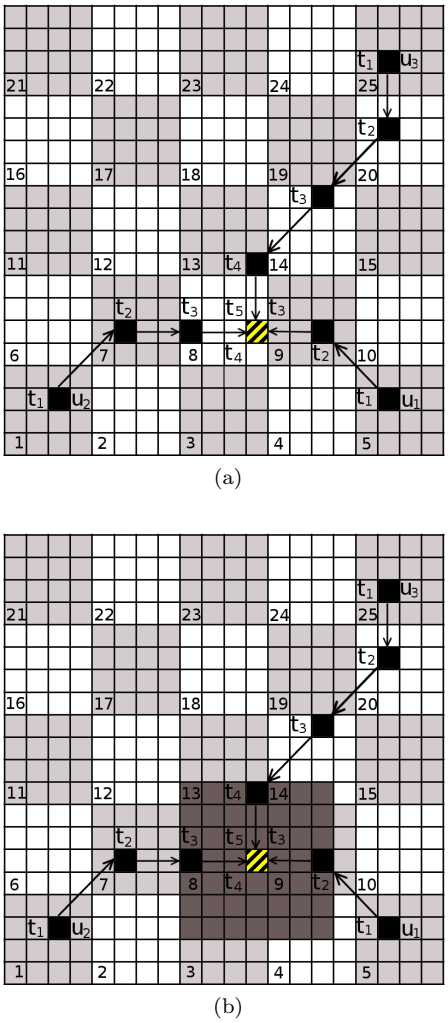


Figure 1: (a) Three mobile users querying a WSDB via PIR, and intersecting at the diagonally striped cell (b) Three mobile users invoking the AV-net protocol for the region identified by the darker shaded cells

We assume that intersecting users remain within communication range for ample periods of time (e.g., 1-2 minutes). However, they do not need to reside in the same cell continuously. The three conditions that control a successful invocation of our protocol are (i) protocol *initiation*, (ii) protocol *participation*, and (iii) successful *group formation*. Group formation is dependent on at least three users willing to engage in the P2P protocol, such that at least one of the engaging users is an *initiator*. We examine each condition separately in the following sections.

4.2 Protocol Initiation

Ideally, a mobile TVBD would like to maintain DSA connectivity throughout its trajectory, without any disrup-

Table 1: Summary of symbols

Symbol	Description	Range
GS	Group size	3-10
BS	Number of cells in a PIR block	16
AR	Ring(s) explored through AV-net invocations	1-3
AP	AV-net participation probability (fixed)	0-1
PI	AV-net participation probability increment (TCP)	0.05-0.2
K	AV-net initiation threshold	0.2-0.8
AK	Actual knowledge of the AR area	0-1

tions. As such, whenever the TVBD moves into a new cell, it measures the ratio of knowledge (AK) in the surrounding area. If that ratio falls under a system-defined threshold K, it initiates the protocol that triggers the group formation algorithm (described later). Algorithm 1 shows the detailed protocol initiation procedure. If there is no channel availability information for the current cell, the user always initiates the protocol (Lines 6–8), because it needs to identify free channels. On the other hand, if the current cell does exist in its cache, it computes the ratio AK for the present position (Lines 9–15). Specifically, each surrounding ring is assigned an identical aggregate weight (equal to $1/AR$), which is split equally among the individual cells. As a result, cells in the inner rings carry more weight than those in the outer rings, and lack of knowledge in the inner rings is more likely to initiate the protocol.

Algorithm 1 Protocol initiation Algorithm

```
1: procedure INITIATE-PROTOCOL
2:
3:     bool initiate = false;
4:     double AK = 0.0;
5:
6:     if no spectrum information for current cell then
7:         initiate = true;
8:     else
9:         for i = 1 to AR do
10:            for all cells c_j in ring i do
11:                if no spectrum information for c_j then
12:                    AK = AK + 1.0/(AR · 8 · i);
13:                end if
14:            end for
15:        end for
16:        if AK ≤ K then
17:            initiate = true;
18:        end if
19:    end if
20:    return initiate;
21: end procedure
```

4.3 Protocol Participation

Participation is defined as the selfless event, where one or more users in the group decide to participate in the AV-net protocol for the purpose of disseminating (and also collecting) channel information about the surrounding area. In order to avoid meaningless (due to repetition) AV-net protocol invocations that could lead to battery drainage, we propose the following three probabilistic AV-net participation methods.

Fixed Probability. This is the simplest approach where, whenever a protocol is initiated, a nearby TVBD always chooses to participate with probability AP. Larger AP values produce a greedy behavior that is optimal in terms of PIR query savings. On the other hand, this may also lead to numerous AV-net invocations in close (spatial) proximity, which are redundant in terms of gained knowledge.

TCP-like Approach. In the second method, we borrow from TCP Reno's congestion control mechanism [23]. In particular, a mobile user starts with a participation probability $AP = 1.0$. At each successful AV-net participation, AP is cut by half. Otherwise, if there is a protocol initiation but the TVBD does not participate, AP is incremented by PI units. This technique is expected to be the most conservative one, due to its aggressive back-off behavior.

Weighted Sliding Window. The final method is based on the weighted sliding window (SW) projection. We experimented with different window sizes, and decided to utilize a model with five entries, such that $W_1 = 0.5$, $W_2 = 0.25$, $W_3 = 0.15$, $W_4 = 0.07$, $W_5 = 0.03$, and $\sum_i W_i = 1$. (W_1 corresponds to the most recent entry.) The current window snapshot is stored as a 5-bit array, where '0' represents participation and '1' represents non-participation. In order to determine the probability of participation, a mobile user first checks its window and sums up past events for which it did not participate. For example, if the current SW snapshot is $(0, 1, 0, 1, 0)$, the user will participate with probability $0.25 + 0.07 = 0.32$. The weighted SW allows us to weight recent historical data more heavily than older ones, when determining the projected probability. This fits well with the intended participation model, in which more recent participation should lead to lower participation probability in the near future.

4.4 Group Formation

When Algorithm 1 (protocol initiation) returns *true*, the underlying TVBD initiates an invocation of the AV-net protocol. This is done by broadcasting its interest in the lowest out-of-band CCC channel. Assuming 801.11 as our out-of-band CCC implementation, any potential initiators will broadcast their unique MAC addresses, their current cell id, their *rendezvous* channel id[4], and an initiation flag over 802.11 channel 1. Users that are already engaged in an AV-net invocation/transmission will not hear such broadcast. We assume standard 802.11 MAC contention mechanisms are in place. We coin as "root" the first mobile user that successfully broadcasts the AV-net initiation control packet, regarding a specific cell id. Any other users (including other potential initiators situated in the same cell) that receive the first successful broadcast from a root node, and whose cell id *matches* the broadcast cell id, will use a simple three-way handshake group formation protocol.

[4]Assuming there is a free channel in the out-of-band CCC range.

Mobile users that decide to participate (based on the methods described earlier) or had attempted to initiate an AV-net invocation themselves, will first switch to the rendezvous channel. They will announce to the root user, through broadcast communication, their willingness to engage. We coin as "children" any of the users that have successfully rendezvoused in the channel id specified by the root user. The three-way handshake broadcast MAC protocol is summarized in Algorithm 2.

Algorithm 2 Group formation algorithm

```
1: procedure THREE-WAY-HANDSHAKE
2:
3:     [all children broadcast] Send Request To Join Group
4:
5:     while group size < GS do
6:         [root] randomly pick a child
7:         [root] send Clear to Join Group
8:         [root] increment group size counter
9:     end while
10:
11:    for all other children who sent a Request to Join do
12:        [root] send Reject to Join Group
13:    end for
14:    [all who received Clear to Join Group] Send Confirm To
   Join Group
15:
16:    [root] send ABORT if group size counter ≤ 3
17: end procedure
```

4.5 AV-net Protocol

When a group is formed, the nodes therein execute the AV-net protocol (as described in Section 3.2) for each bit of information that they want to share. However, to avoid excessive network delays due to the 2-round nature of the AV-net protocol, we group all individual invocations into two aggregate rounds, as shown in Algorithm 3. Specifically, the users first agree on the the specific order in which the cell information is transmitted, and then each user broadcasts its aggregate data to the rest of the group. The broadcast order can be arranged based on the unique MAC addresses of the TVBDs.

Algorithm 3 AV-net protocol

```
1: procedure AV-NET(G, g)
2:
3:     for all users i in the group do
4:         for all bits b in the explored area do
5:             compute g^{y_{i_b}};
6:         end for
7:     end for
8:
9:     for all users i in the group do
10:        for all bits b in the explored area do
11:            compute g^{c_{i_b} y_{i_b}};
12:        end for
13:        broadcast all exponentiations for user i;
14:    end for
15:
16:    for all users i in the group do
17:        for all bits b in the explored area do
18:            compute r_b = Π_i g^{c_{i_b} y_{i_b}};
19:            if r_b ≠ 1 then
20:                mark the corresponding channel as free;
21:            end if
22:        end for
23:    end for
24: end procedure
```

Lines 3–8 (Algorithm 3) correspond the the first round of the AV-net protocol, i.e., each node broadcasts a unique key for every bit of information in the surrounding AR rings. In the example of Figure 1b, where $AR = 3$, each node computes and broadcasts $32 \cdot (1 + \sum_{i=1}^{3} 8 \cdot i) = 1568$ modular exponentiation results. In Round 2 of the protocol (Lines 9–23), users publish their spectrum knowledge by choosing the appropriate values for c_{i_b} (as explained in Section 3.2). Specifically, if the underlying channel if free, the user vetoes that particular statement. Note that, in our running example, this step also involves the computation and broadcast of 1568 modular exponentiations. The result extraction phase of the algorithm (Lines 16–23) necessitates only GS modular multiplications per bit, and it is optional, i.e., it can be computed only when the user moves into the corresponding cell.

5 Experimental Evaluation

In this section we evaluate experimentally the performance of our methods. Section 5.1 describes the experimental setup and Section 5.2 presents our results.

5.1 Experimental Setup

We developed our experiments in Java SDK, running on a Ubuntu 10.4 LTS machine. To simulate the mobile TVBD users, we utilized Microsoft's GeoLife GPS Trajectories[5], which is an excellent dataset containing real-life trajectories from users traveling around Beijing, China. The GeoLife dataset [49] was collected as part of the Microsoft Research Asia GeoLife project, by monitoring numerous users for a period of over five years (from Apr. 2007 to Aug. 2012). A GPS trajectory from this dataset is represented as a sequence of time-stamped points, each containing information regarding the user's latitude, longitude, and altitude.

The dataset includes 17,621 trajectories, with a total distance of 1,292,951 kilometers, and a total duration of 50,176 hours. These trajectories were recorded by different GPS loggers and GPS-enabled phones, and have a variety of sampling rates. More specifically, 91.5 percent of the trajectories are logged in a dense representation, e.g., every 1–5 seconds or every 5–10 meters per point. We randomly selected 2774 intersecting trajectories, each simulating a unique user. For each trajectory, we measure (i) the average number of PIR queries issued by the user, and (ii) the average number of AV-net invocations that the user participates into.

In addition to the simulation results, we also implemented the basic cryptographic operations of the AV-net protocol on an iPhone 5, running iOS 7.1. Specifically, we cross compiled the GMP[6] multiple precision arithmetic library for the ARM architecture, and built a benchmark app to measure the cost of these operations on a handheld

[5] http://research.microsoft.com/en-us/projects/GeoLife/
[6] http://gmplib.org

device. We generated a cyclic group G of prime order q, where q is a 160-bit number. The group modulus was chosen as a 64-byte prime. Table 2 shows the cost of these operations.

Table 2: Cost of cryptographic primitives

Operation	Cost
Modular multiplication	0.004 ms
Modular exponentiation	0.518 ms

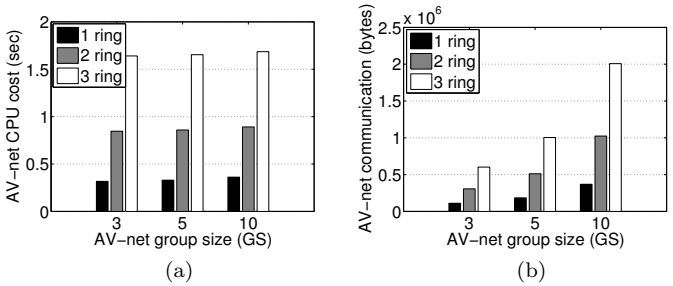
5.2 Results

Figure 2a illustrates the projected CPU time needed to run the AV-net protocol (Algorithm 3) on a handheld device. This cost is dominated by the expensive modular exponentiation operations and is, thus, unaffected by the group size GS. The major factor that determines this cost is the number of surrounding rings (AR) that are explored during a protocol invocation, since each cell contributes 32 modular exponentiations. Nevertheless, even for a value of $AR = 3$, the total CPU time is around 1.65 sec, which is an acceptable cost.

Furthermore, this cost can easily be reduced by 50%, using offline computations. Observe that, during the first round of the AV-net protocol, each node computes and publishes a large number of modular exponentiations. These values do not require any input from the other participating nodes and may, thus, be pre-computed offline. Specifically, a large pool of values (e.g., several hundred thousands) can be computed either at the mobile device during night time (when charging), or at a desktop machine for faster computations. The storage space required to maintain these values is insignificant compared to the storage capabilities of modern handheld devices.

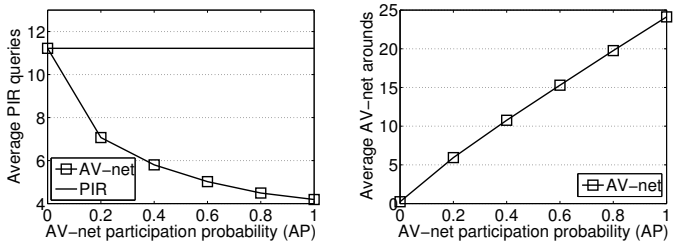
Figure 2b shows the total number of bytes that are broadcast during an AV-net protocol invocation. Clearly, the communication cost is linear in GS, as each group member needs to broadcast its own input to the protocol. We believe that $GS = 5$ is a very reasonable value for anonymity purposes, in which case the communication cost remains below 1 MB. While this cost might appear significant, we stress that, AV-net broadcasts occur over the 802.11 CCC band and do not involve the cellular network infrastructure.

Figure 3 investigates the effect of the fixed AV-net participation probability (AP) on the performance of our methods. For this experiment, we set $AR = 2$, $GS = 5$, and $K = 0.5$. The curve labeled "PIR" (Figure 3a) corresponds to the PIR-only approach, i.e., when users do not leverage our P2P protocol. When $AP = 0.5$, we observe a 50% reduction in the amount of PIR queries that are sent to the WSDB provider. Larger values naturally lead to better performance (over 60% reduction), but they increase considerably the number of AV-net invocations per user (Figure 3b). Nevertheless, as we have explained previously, PIR queries are much more expensive compared to the AV-net protocol.



(a) (b)

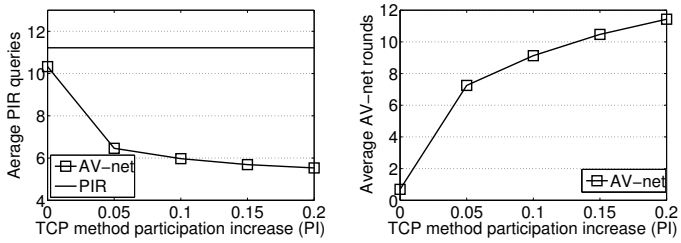
Figure 2: Cost of AV-net protocol on handheld devices (a) CPU cost (b) Communication cost



(a) (b)

Figure 3: Effect of varying the AV-net participation probability (a) Average number of PIR queries (b) Average number of AV-net invocations

Figure 4 shows the effect of the participation probability increment (PI) for the TCP-like approach ($AR = 2$, $GS = 5$, $K = 0.5$). Lower values of PI discourage users from participating in AV-net protocols and, thus, incur less cost compared to the fixed probability method (Figure 4b). However, as evident in Figure 4a, the TCP-like approach can still reduce the number of PIR queries by up to 50%.

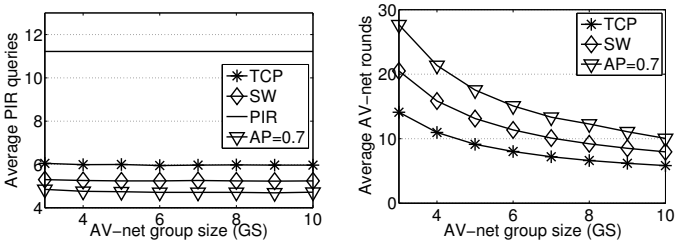


(a) (b)

Figure 4: Effect of varying the AV-net participation probability increment (TCP) (a) Average number of PIR queries (b) Average number of AV-net invocations

Figure 5 demonstrates the effect of the group size (GS) on the different methods ($AR = 2$, $K = 0.5$, $PI = 0.1$). As Figure 5a implies, larger groups do not contribute more information during the P2P data exchange. Therefore, the average number of PIR queries remains fairly constant. Nevertheless, users may still opt for larger

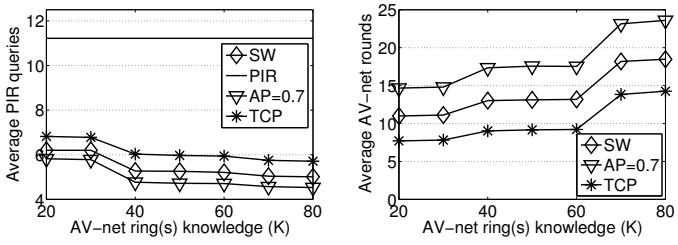
groups, in order to gain more privacy. On the other hand, a larger group size reduces the number of AV-net invocations (Figure 5b), because some groups may fail to form due to insufficient number of members. Among the three participation algorithms, the sliding window (SW) approach strikes a good balance between PIR savings (53%) and AV-net overhead (13 rounds, for $GS = 5$).



(a) (b)

Figure 5: Effect of varying the AV-net group size (a) Average number of PIR queries (b) Average number of AV-net invocations

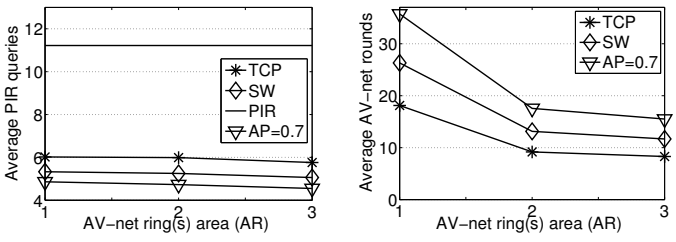
Figure 6 depicts the effect of the protocol initiation threshold (K) on the different methods ($AR = 2$, $GS = 5$, $PI = 0.1$). Recall that, this threshold represents a lower bound on the amount of spectrum knowledge that a mobile user must possess (regarding the surrounding area), in order to defer an AV-net protocol initiation. As evident in this figure, a knowledge of around 40%-50% is sufficient in terms of overall performance. Larger values to not offer much in terms of PIR reduction, but instead lead to unnecessary AV-net rounds. Similar to Figure 5, the SW participation method has the best performance.



(a) (b)

Figure 6: Effect of varying the AV-net initiation threshold (a) Average number of PIR queries (b) Average number of AV-net invocations

Finally, Figure 7 illustrates the effect of the number of surrounding rings (AR) that are explored during an AV-net protocol invocation ($K = 0.5$, $GS = 5$, $PI = 0.1$). The first observation, is that the number of PIR queries remains almost constant (Figure 7a). The reason is that, as shown in Figure 7b, exploring one ring at a time merely results in more AV-net rounds, since users invoke a new AV-net protocol once they move further away from their current position. However, the overall PIR reduction is not affected, because users still get most of their spectrum knowledge from the P2P protocol. A value of $AR = 2$ seems like the best choice, given that the number of AV-net rounds does not decrease significantly from 2 to 3 rings.



(a) (b)

Figure 7: Effect of varying the AV-net exploration area (a) Average number of PIR queries (b) Average number of AV-net invocations

6 Conclusions

Database-driven dynamic spectrum access is the standard mode of operation for cognitive radios in the white-space TV bands. This method requires mobile devices to periodically send their location to a centralized white-space database, in order to receive channel availability information in their surrounding area. Nevertheless, location-dependent queries pose a serious privacy threat, as they may reveal sensitive information about an individual. To mitigate this threat, previous work has proposed the use of private information retrieval (PIR) protocols when querying the database. In this work, we argue that PIR queries are very expensive and should be avoided, to the extent possible. To this end, we propose a novel approach that allows mobile users to share anonymously their cached channel availability information that is obtained from previous queries. Our experiments with a real-life dataset, indicate that our methods reduce the number of PIR queries by 50% to 60%. Furthermore, they are efficient in terms of both computational and communication cost.

Acknowledgments

This research has been funded by the NSF CAREER Award IIS-0845262.

References

[1] I. F. Akyildiz, W. Y. Lee, and K. R. Chowdhury, "CRAHNs: Cognitive radio ad hoc networks," *Ad Hoc Networks*, vol. 7, no. 5, pp. 810–836, 2009.

[2] M. Blaze, J. Ioannidis, A. D. Keromytis, T. G. Malkin, and A. Rubin, "Anonymity in wireless

broadcast networks," *International Journal of Network Security*, vol. 8, no. 1, pp. 37–51, 2009.

[3] D. Boneh, "The decision Diffie-Hellman problem," in *Algorithmic Number Theory*, pp. 48–63. 1998.

[4] F. Brandt, "Efficient cryptographic protocol design based on distributed ElGamal encryption," in *Information Security and Cryptology (ICISC'06)*, pp. 32–47. 2006.

[5] D. Chaum, "The dining cryptographers problem: Unconditional sender and recipient untraceability," *Journal of Cryptology*, vol. 1, no. 1, pp. 65–75, 1988.

[6] C. Cormio and K. R. Chowdhury, "A survey on MAC protocols for cognitive radio networks," *Ad Hoc Networks*, vol. 7, no. 7, pp. 1315–1329, 2009.

[7] FCC, "Third memorandum opinion and order," pp. 12–36, 2012.

[8] FCC, "Television band devices," p. 11, 2013.

[9] J. Freudiger, M. H. Manshaei, J. P. Hubaux, and D. C. Parkes, "On non-cooperative location privacy: A game-theoretic analysis," in *Proceedings of the 16th ACM conference on Computer and communications security (CCS'09)*, pp. 324–337, 2009.

[10] Z. Gao, H. Zhu, Y. Liu, M. Li, and Z. Cao, "Location privacy in database-driven cognitive radio networks: Attacks and countermeasures," in *IEEE INFOCOM'13*, pp. 2751–2759, 2013.

[11] C. Gentry and Z. Ramzan, "Single-database private information retrieval with constant communication rate," in *Proceedings of 32nd International Colloquium Automata, Languages and Programming (ICALP'05)*, pp. 803–815. 2005.

[12] G. Ghinita, P. Kalnis, A. Khoshgozaran, C. Shahabi, and K. L. Tan, "Private queries in location based services: Anonymizers are not necessary," in *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, pp. 121–132, 2008.

[13] O. Goldreich, S. Micali, and A. Wigderson, "How to play any mental game," in *Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing*, pp. 218–229, 1987.

[14] S. Goldwasser and S. Micali, "Probabilistic encryption & how to play mental poker keeping secret all partial information," in *Proceedings of the Fourth Annual ACM Symposium on Theory of Computing*, pp. 365–377, 1982.

[15] J. Groth, "Efficient maximal privacy in boardroom voting and anonymous broadcast," in *Financial Cryptography*, pp. 90–104, 2004.

[16] M. Gruteser and D. Grunwald, "Anonymous usage of location-based services through spatial and temporal cloaking," in *Proceedings of the 1st International Conference on Mobile Systems, Applications and Services*, pp. 31–42, 2003.

[17] F. Hao and P. Zieliński, "A 2-round anonymous veto protocol," in *Security Protocols*, pp. 202–211, 2009.

[18] S. Haykin, D. J. Thomson, and J. H. Reed, "Spectrum sensing for cognitive radio," *Proceedings of the IEEE*, vol. 97, no. 5, pp. 849–877, 2009.

[19] W. He, X. Liu, H. Nguyen, K. Nahrstedt, and T. T. Abdelzaher, "PDA: Privacy-preserving data aggregation in wireless sensor networks," in *IEEE INFOCOM'07*, pp. 2045–2053, 2007.

[20] L. Huang, K. Matsuura, H. Yamane, and K. Sezaki, "Enhancing wireless location privacy using silent period," in *IEEE Wireless Communications and Networking Conference*, vol. 2, pp. 1187–1192, 2005.

[21] L. Huang, H. Yamane, K. Matsuura, and K. Sezaki, "Silent cascade: Enhancing location privacy without communication qos degradation," in *Security in Pervasive Computing*, pp. 165–180, 2006.

[22] M. S. Hwang, C. H. Wei, and C. Y. Lee, "Privacy and security requirements for rfid applications," *Journal of Computers*, vol. 20, no. 3, pp. 55–60, 2009.

[23] IETF, "Tcp congestion control," RFC 2581, 2001. (https://tools.ietf.org/html/rfc2581)

[24] A. Kiayias and M. Yung, "Non-interactive zero-sharing with applications to private distributed decision making," in *Financial Cryptography*, pp. 303–320, 2003.

[25] H. Kido, Y. Yanagisawa, and T. Satoh, "An anonymous communication technique using dummies for location-based services," in *Proceedings of IEEE International Conference on Pervasive Services (ICPS'05)*, pp. 88–97, 2005.

[26] E. Kushilevitz and R. Ostrovsky, "Replication is not needed: Single database, computationally-private information retrieval," in *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pp. 364–373, 1997.

[27] B. Lee, J. Oh, H. Yu, and J. Kim, "Protecting location privacy using location semantics," in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1289–1297, 2011.

[28] C. T. Li and M. S. Hwang, "A lightweight anonymous routing protocol without public key en/decryptions for wireless ad hoc networks," *Information Sciences*, vol. 181, no. 23, pp. 5333–5347, 2011.

[29] S. Li, H. Zhu, Z. Gao, X. Guan, K. Xing, and X. Shen, "Location privacy preservation in collaborative spectrum sensing," in *IEEE INFOCOM*, pp. 729–737, 2012.

[30] H. Lu, C. S. Jensen, and M. L. Yiu, "PAD: privacy-area aware, dummy-based location privacy in mobile services," in *Proceedings of the Seventh ACM International Workshop on Data Engineering for Wireless and Mobile Access*, pp. 16–23, 2008.

[31] 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.

[32] J. Meyerowitz and R. Choudhury, "Hiding stars with fireworks: location privacy through camouflage," in *Proceedings of the 15th Annual ACM International Conference on Mobile Computing and Networking*, pp. 345–356, 2009.

[33] J. Mitola III, "Cognitive radio: An integrated agent architecture for software defined radio," *Doctoral Dissertation, KTH, Stockholm, Sweden*, May 2000.

[34] M. F. Mokbel, C. Y. Chow, and W. G. Aref, "The new Casper: query processing for location services without compromising privacy," in *Proceedings of the 32nd International Conference on Very Large Data Bases (VLDB'06)*, pp. 763–774, 2006.

[35] S. Papadopoulos, S. Bakiras, and D. Papadias, "pCloud: A distributed system for practical PIR," *IEEE Transactions on Dependable and Secure Computing*, vol. 9, no. 1, pp. 115–127, 2012.

[36] C. P. Schnorr, "Efficient signature generation by smart cards," *Journal of Cryptology*, vol. 4, no. 3, pp. 161–174, 1991.

[37] J. Shi, R. Zhang, Y. Liu, and Y. Zhang, "Prisense: privacy-preserving data aggregation in people-centric urban sensing systems," in *IEEE INFOCOM*, pp. 1–9, 2010.

[38] L. Sweeney, "k-anonymity: A model for protecting privacy," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 05, pp. 557–570, 2002.

[39] W. L. Tan, F. Lam, and W. C. Lau, "An empirical study on the capacity and performance of 3G networks," *IEEE Transactions on Mobile Computing*, vol. 7, no. 6, pp. 737–750, 2008.

[40] J. Trostle and A. Parrish, "Efficient computationally private information retrieval from anonymity or trapdoor groups," in *Information Security*, pp. 114–128. 2011.

[41] K. Vu, R. Zheng, and J. Gao, "Efficient algorithms for k-anonymous location privacy in participatory sensing," in *IEEE INFOCOM*, pp. 2399–2407, 2012.

[42] T. Wang and L. Liu, "Privacy-aware mobile services over road networks," *Proceedings of the VLDB Endowment*, vol. 2, no. 1, pp. 1042–1053, 2009.

[43] T. Xu and Y. Cai, "Exploring historical location data for anonymity preservation in location-based services," in *IEEE INFOCOM*, pp. 547–555, 2008.

[44] T. Xu and Y. Cai, "Feeling-based location privacy protection for location-based services," in *Proceedings of the 16th ACM Conference on Computer and Communications Security*, pp. 348–357, 2009.

[45] Z. Yang, G. Cheng, W. Liu, W. Yuan, and W. Cheng, "Local coordination based routing and spectrum assignment in multi-hop cognitive radio networks," *Mobile Networks and Applications*, vol. 13, pp. 67–81, 2008.

[46] A. C. C. Yao, "How to generate and exchange secrets," in *IEEE 27th Annual Symposium on Foundations of Computer Science*, pp. 162–167, 1986.

[47] W. Zhang, R. K. Mallik, and K. Letaief, "Optimization of cooperative spectrum sensing with energy detection in cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 8, no. 12, pp. 5761–5766, 2009.

[48] Q. Zhao, S. Geirhofer, L. Tong, and B. M. Sadler, "Optimal dynamic spectrum access via periodic channel sensing," in *IEEE Wireless Communications and Networking Conference (WCNC'07)*, pp. 33–37, 2007.

[49] Y. Zheng, L. Wang, R. Zhang, X. Xie, and W. Y. Ma, "GeoLife: Managing and understanding your past life over maps," in *IEEE 9th International Conference on Mobile Data Management (MDM'08)*, pp. 211–212, 2008.

[50] Z. Zhi and Y. K. Choong, "Anonymizing geographic ad hoc routing for preserving location privacy," in *25th IEEE International Conference on Distributed Computing Systems Workshops*, pp. 646–651, 2005.

Erald Troja is currently a Ph.D. student at the Graduate Center of the City University of New York. He received his bachelor degree in Computer Science from Brooklyn College, NY, in 1999, his M.S in Information Systems and Business Administration from Brooklyn College, NY, in 2009 and M.S in Computer Science from the Graduate Center, CUNY in 2012. His major research areas include dynamic spectrum access, cognitive radio networks and security and privacy in location based services.

Spiridon Bakiras received the BS degree in Electrical and Computer Engineering from the National Technical University of Athens in 1993, the MS degree in Telematics from the University of Surrey in 1994, and the PhD degree in Electrical Engineering from the University of Southern California in 2000. Currently, he is an associate professor in the Department of Mathematics and Computer Science at John Jay College, City University of New York. Before that, he held teaching and research positions at the University of Hong Kong and the Hong Kong University of Science and Technology. His current research interests include database security and privacy, mobile computing, and spatiotemporal databases. He is a member of the ACM and a recipient of the US National Science Foundation (NSF)CAREER award.