POSTER: Location Privacy Leaking from Spectrum Utilization Information in Database-driven Cognitive Radio Network

Zhaoyu Gao[†], Haojin Zhu[†], Yao Liu[‡], Muyuan Li[†] and Zhenfu Cao[†]
[†]Shanghai Jiao Tong University, Shanghai 200240, P. R. China
{gaozy1987, leilmyxwz}@gmail.com, {zhu-hj, zfcao}@sjtu.edu.cn
[‡]University of South Florida, yliu584@gmail.com

ABSTRACT

The Database-driven Cognitive Radio Network is regarded as a promising way for a better utilization of radio channels without introducing the interference to the primary user. However, it is also facing a series of security threats. In this study, we identify a new kind of location privacy related attack which could geo-locate a secondary user from the spectrum he used. We propose a Spectrum Utilization based Location Inference Algorithm, which is based on the intersection of the possible location sets revealed by each channel access or channel transition event under the presence of the primary user. We implement our algorithm on the data extracted from Google Earth Coverage Maps released by FCC. Our experiement results show that, 80% SUs could be located to 10 cells based on 25 or less channels.

Categories and Subject Descriptors

C.2.m [**Computer-Communication Networks**]: Miscellaneous

General Terms

Security

Keywords

DB-driven Cognitive Radio Network, Location Privacy, Spectrum Utilization

1. INTRODUCTION

Over the last decade, unlicensed channels have been used by prevalent wireless technologies like wireless LAN, mesh, BLUETOOTH. However, unlicensed channels only take a very small portion of the entire spectrum used by people today. As wireless technologies become more and more popular, unlicensed channels become more crowed. Cognitive Radio Networks (CRNs) thus have been proposed to address the increasing demand for wireless bandwidth and support emerging wireless technologies. CRNs have two types of users: Primary Users (PUs) and Secondary User (SUs). PUs are licensed users assigned with certain channels to operate, and SUs are unlicensed users allowed to use PUs' channels only when the channel are not occupied by the PU.

CCS'12, October 16–18, 2012, Raleigh, North Carolina, USA.
ACM 978-1-4503-1651-4/12/10.

Spectrum sensing and white space database are two typical ways to determine which channels are locally available for reuse by the SUs. However, the latest FCC's rule [1] in May 2012 eliminates spectrum sensing as a requisite for cognitive radio devices. Instead, it requires that all fixed or mobile cognitive radio devices (i.e., SUs) should query a database to obtain Spectrum Availability Information (SAI) at their location, and register their operations in this database. FCC has designated nine entities (e.g. Comsearch, Google Inc.) as TV bands device database administrators. Recently, two TV Bands database systems designed by Koos Technical Services, Inc. and Telecordia Technologies, Inc, have been approved by FCC for operation.

Though white space database is regarded as a promising implementation for CRN, privacy issues draw people's attention. A straightforward location privacy attack towards white space database is compromising SU's location privacy from the location-based spectrum availability query, in which the SU sends queries to the white space database to retrieve SAI at its location. The similar attacks also happen in conventional location-based service and are expected to be mitigated via cryptographic approach such as private information retrieval technique or k-anonymity technique.

In this study, however, we revealed a new attack which enables an attacker to learn the location of an SU. This attack arises from the fact that a secondary user can gain access to a channel if and only if the presence of the PU is not detected in this location (e.g., out of the coverage of PU). In other words, any event that an SU can or can not access to a channel with the presence of the PU will leak his location information partially. Such correlation between the spectrum utilization information of an SU and his physical location could be exploited to geo-locate an SU by intersecting the coverage of different channels that the SU has used.

2. BACKGROUNDS & ASSUMPTIONS

We assume that the adversary's goal is to get the location of an SU by inferring from the spectrum utilization information. The adversary could either be the curious-but-honest service provider, who is incentivized to collect the location information of the equipment of secondary user [1], or any external adversary who tries to infer the users' location information by analyzing users' spectrum utilization information leaking from database due to a variety of reasons including malware, security break-in, etc.

A location-based database driven CRN consists of four components: Primary Users (PUs), Secondary Users (SUs),

Base Station (BS) and a white space Database DB, who possesses all the SAI of the region. We denote a BS covered region as C which is divided into $n \times n$ square cells. The side length of a cell is assumed to be 600m $\sim$ 800m, which is determined by both the shadowing correlation[3] and the efficiency of spectrum utilization[4]. The radius of C varies from $30km$ to $100km$ as suggested in IEEE 802.22 standard. In the BS covered region, there are K PUs, which are denoted as $PU_k, k \in \{1, \ldots, K\}$. The channel of each primary user PU_k is referred to ch_k. The coverage of PU_k's signal is coined as a covered region C_k. In this case, ch_k is regarded as unavailable in the area C_k if PU_k's state is ON. Instead, the channel ch_k is regarded as available in the whole region C if PU_k's state is OFF.

In a typical SAI query process, an SU first query the database by submitting his location information. Then DB will return the SAI to the SU. Based on the SAI, the SU will choose a channel ch_k and register the operation on ch_k on the database. After this process, the link between SU and BS is established through the channel ch_k. When the state of PU_k varies from OFF to ON, DB will inform all the SUs who are using ch_k. Then the SUs will query the database following the same process as described above. In this case, the SUs in PU_k's coverage C_k have to choose another vacant channel. On the other hand, the SUs out of C_k will not change his channel, since ch_k is still available for them.

3. ATTACK METHODOLOGY

Algorithm 1: SULI Algorithm

Input: event sequence $E = \{e_1, e_2, \ldots, e_l, \ldots\}$, where e_l could be *Event I* (SU, t, ch_k) or *Event II* $(SU, t, ch_{k_1}, ch_{k_2})$.
Output: the SU's possible location set S.
Initialization: Let SU's possible location set $S = C$.
Run:
while an event occurs about SU **do**
  if the event is Event I **then**
    PUCC(E_t^k)
  else
    ECS($E_t^{k_1 \to k_2}$)
  end if
end while
function PUCC(E_t^k)
if PU_k's state is ON at time t **then**
  $S \leftarrow S \bigcap (C - C_k)$
end if
end function
function ECS($E_t^{k_1 \to k_2}$)
$S \leftarrow S \bigcap C_k$
PUCC($E_t^{k_2}$)
end function

3.1 The Basic Idea

The spectrum will leak the users' location privacy, since any PU_k has a certain coverage C_k, for a specific time slot, a specific channel is only available for a SU which is out of C_k. Therefore, given the knowledge of presence of PU_k in the channel ch_k, the possible location set S of an SU could narrow down to the complement of C_k, we take this as *Primary User Coverage Complement Attack*, or *PUCC attack*. Further, we denote the event that, at the time slot t, an SU accesses a channel ch_k in the presence of PU_k as *Event I*, $E_t^k = (SU, t, ch_k)$.

Once an SU makes a handover from one channel ch_{k_1} to another channel ch_{k_2} due to the unavailability of this channel (the state of PU_{k_1} changing from OFF to ON), an SU within the coverage C_{k_1} will be enforced to switch to another channel, thus the fact that the SU is in C_{k_1} will be known by DB. This attack is referred as *Enforced Channel Switch Attack* (or *ECS attack*), such an event that an SU has to switch to another channel in the case that the state of PU_{k_1} changing from OFF to ON is denoted as *Event II*, $E_t^{k_1 \to k_2} = (SU, t, ch_{k_1}, ch_{k_2})$.

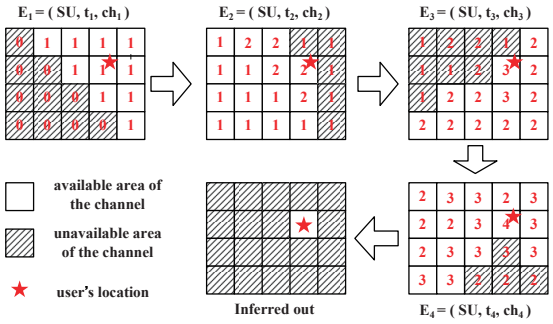


Figure 1: An example of PUCC attack

.

3.2 The Proposed Spectrum Utilization based Location Inferring Algorithm

We formalize the attack approach as below:

- **Case I: PUCC Attack** Given PU's state is ON, an event (SU, t, ch_k) that SU is using the channel ch_k at the moment t indicates that the possible location set of SU could be derived from the complement of the coverage of channel ch_k as

$$S \in C - C_k \tag{1}$$

- **Case II: ECS Attack** Event II $(SU, t, ch_{k_1}, ch_{k_2})$ indicates that SU switches from channel ch_{k_1} to channel ch_{k_2} due to the state transition (from OFF to ON) of PU_{k_1}. This will further introduce two cases. If the state of PU_{k_2} is ON, it can be derived that SU should be within the coverage of ch_{k_1} while in the complement of the coverage of ch_{k_2} as

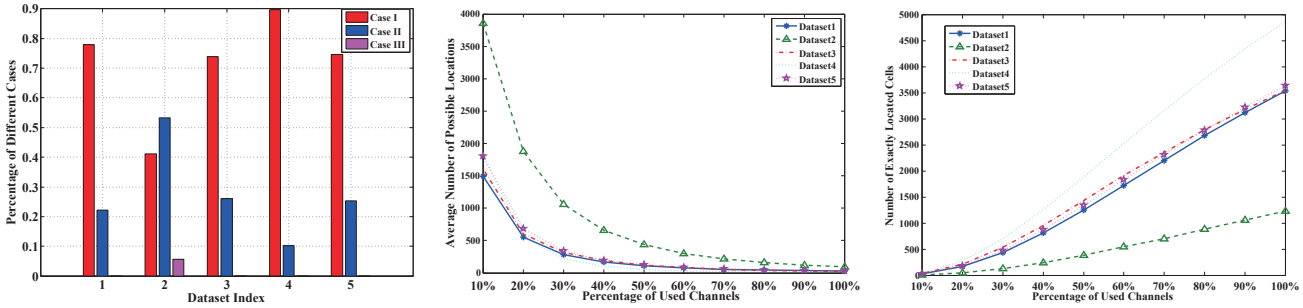
$$S \in C_{k_1} \bigcap (C - C_{k_2}) \tag{2}$$

Otherwise, the possible location set of SU is still in the coverage of ch_{k_1}

$$S \in C_{k_1} \tag{3}$$

Based on the above conclusion, we propose Spectrum Utilization based Location Inference Algorithm (SULI Algorithm) as shown in Algorithm 1. Let SU's possible location set S start from whole BS covered region C. Thereafter, for each inferring step, the possible location set of SU will shrink based on the two cases discussed above.

We show a simple example in Fig.1. It shows that the user's location could be inferred out after he accesses four



(a) Percentage of Each Case (b) Average Result of All Cases (c) Number of Exactly Located Region

Figure 3: Evaluation Result of the Location Privacy Leakage

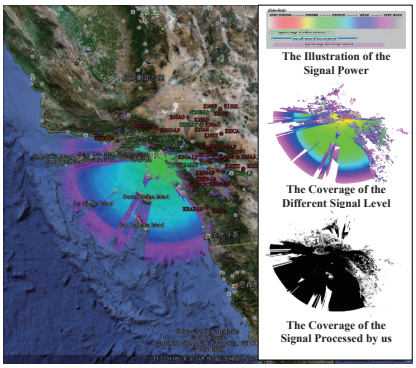


Figure 2: The coverage of TV tower KRCA-0-D

channels ch_1, ch_2, ch_3, ch_4. Based on the first event E_1, DB could derive the possible cells, which are labelled by "1", while assigning those impossible cells by the number "0". For the subsequent Event, we label those overlapping possible cells by increasing their numbers by "1" while keeping other non-overlapping cells unchanged. It is obvious that, after four events, the cell with the highest number is derived as the most likely position of the target user.

4. SYSTEM SETUP AND EVALUATIONS

We setup the white space database by adopting the spectrum availability information of Los Angles released on TV-Fool [2], and implementing all FCC restrictions on all TV towers. In LA area, there are 129 channels totally, one of which is shown in Fig.2. Then we extract the SAI from these data and choose 5 sample regions of with the scale of $75km \times 75km$. Each region is divided into 100×100 cells. We perform 20 Monte Carlo experiments by randomly choosing different percentage of channels accessed by the secondary users during the presence of PU or enforced channel switch.

We measure the privacy leaking inferred based on spectrum utilization. The results are classified into three categories: the Case I (a good case located to less than 25 cells), the Case II (located to $25 \sim 500$ cells), and the Case III (a bad case located to more than 500 cells).

Table.1 shows the inference results in the case that users have traversed all of the channels under the presence of PUs. The result shows that the SUs could be located to $1 \sim 2$ cells in the Case I while could achieve the localization accuracy of $1 \sim 5$ cells. Fig.3-(a) further gives the distribution of the

dataset	number of inferred locations		
	average case	Case I	Case III
1	2.1697	1.7098	-
2	5.3505	2.3107	568.9999
3	2.2292	1.6761	-
4	1.6661	1.5044	-
5	2.1580	1.6279	-

Table 1: Number of Inferred Possible Location Set

Case I, II and III cases for 5 data sets. It also shows that, only one out of total 5 data sets have the Case III, which means, given enough spectrum utilization information, the users could be located with a high accuracy.

We also investigate the situation when no enough spectrum information is provided. We evaluate the average localization performance under the different percentage of channels accessed by SUs in case of PU's state is ON. In Fig.3-(b), it implies that, along with the increasing percentage of used channels, the inference accuracy could be significantly improved. Specifically, with more than 50% channel information exploited, SUs could be located to less than 100 cells. Fig.3-(c) shows the number of exactly located SUs (located to only one cell) under different percentage of used channels. It shows that more than 10% regions will be exactly distinguished with only 40% channel information is used.

In our experiments, around 80% SUs could be located to less than 10 cells by using 25 or less channels. This further demonstrates the practicality of the proposed scheme.

ACKNOWLEDGEMENT

This research is supported by National Natural Science Foundation of China (Grant No.61003218, 70971086, 61161140320, 61033014), Doctoral Fund of Ministry of Education of China (Grant No.20100073120065).

5. REFERENCES

[1] FCC, Third Memorandum Opinion and Order, ET Docket No FCC 12-36A1, May 2012.

[2] http://www.tvfool.com/. March, 2012.

[3] H. Kim and K. Shin. In-band spectrum sensing in cognitive radio networks: energy detection or feature detection? In *Proc. of the Mobicom'08*. ACM, 2008.

[4] R. Murty, R. Chandra, T. Moscibroda, and P. Bahl. Senseless: A database-driven white spaces network. In *DySPAN'11*. IEEE, 2011.