Location Privacy Preservation in Collaborative Spectrum Sensing

Shuai Li[†], Haojin Zhu[†], Zhaoyu Gao[†], Xinping Guan[†], Kai Xing[‡] and Xuemin (Sherman) Shen[§]
[†] Shanghai Jiao Tong University, Shanghai, China
{shuailee, zhu-hj, zy-gao, xpguan}@sjtu.edu.cn
[‡] University of Science and Technology of China, P.R. China
kxing@ustc.edu.cn
[§] University of Waterloo, Waterloo, Ontario, Canada
xshen@bbcr.uwaterloo.ca

Abstract—**Collaborative spectrum sensing has been regarded as a promising approach to enable secondary users to detect primary users by exploiting spatial diversity. In this paper, we consider a converse question: could space diversity be exploited by a malicious entity, e.g., an external attacker or an untrusted Fusion Center (FC), to achieve involuntary geolocation of a secondary user by linking his location-dependent sensing report to his physical position. We answer this question by identifying a new security threat in collaborative sensing from testbed implementation, and it is shown that the attackers could geo-locate a secondary user from its sensing report with a successful rate of above 90% even in the presence of data aggregation. We then introduce a novel location privacy definition to quantify the location privacy leaking in collaborative sensing. We propose a Privacy Preserving collaborative Spectrum Sensing (PPSS) scheme, which includes two primitive protocols: Privacy Preserving Sensing Report Aggregation protocol (PPSRA) and Distributed Dummy Report Injection Protocol (DDRI). Specifically, PPSRA scheme utilizes applied cryptographic techniques to allow the FC to obtain the aggregated result from various secondary users without learning each individual's values while DDRI algorithm can provide differential location privacy for secondary users by introducing a novel sensing data randomization technique. We implement and evaluate the PPSS scheme in a real-world testbed. The evaluation results show that PPSS can significantly improve the secondary user's location privacy with a reasonable security overhead in collaborative sensing.**

Keywords – **Location Privacy, Cognitive Radio Security, Collaborative Sensing**

I. INTRODUCTION

The proliferation of smart phones and mobile Internet based applications requires a better utilization of radio channels. To address the ever increasing demand for wireless bandwidth, cognitive radio (CR) networks have been proposed to increase the efficiency of channel utilization under the current static channel allocation policy [1]. Unlike the conventional spectrum regulation paradigms in which the majority of the spectrum is allocated to fixed licensed users (or primary users) for exclusive usage, a CR system permits unlicensed users (or secondary users) to utilize the idle spectrum as long as it does not introduce the interference to the primary users.

One major technical challenge of designing dynamic spectrum access system is to detect the presence of the primary users' transmissions and thus determine the availability of a certain spectrum. It is recently discovered that collaboration among multiple secondary users can significantly improve the performance of spectrum sensing by exploiting the spatial diversity of them. Therefore, collaborative sensing has been widely adopted in the standard proposals for CR networks, e.g., IEEE 802.22 WRAN, CogNeA, IEEE 802.11af and WhiteFi [2]–[4].

However, collaborative sensing is also facing a series of security threats. Recently, security issues in collaborative sensing has been increasingly attracting researching attentions. So far, most of the existing research works mainly focus on incentive issues in collaboration [5], [6], or preventing the malicious nodes from reporting inaccurate or even fake messages [8]–[11]. In this paper, we consider a new type of threat, *Location Privacy Leaking in Collaborative Sensing*. Specifically, the existing works show that the sensing report on the signal propagation of primary users is highly dependent on the secondary user's physical location [8], which is also demonstrated in our experiments. Therefore, similar to geolocating the individuals via WiFi or Bluetooth signals, the correlation of CR sensing reports and their physical location can be exploited by malicious attackers to geo-locate a user and thus compromise the user's location privacy.

A potential approach to prevent location privacy leaking in collaborative sensing is Privacy Preserving Aggregation Techniques, with which the FC may aggregate spectrum availability data from various CR devices and, at the same time, to conceal the spectrum sensing data from leaking. However, there are several research challenges which make the considered privacy preserving spectrum sensing issue fundamentally different from any existing privacy preserving aggregation solutions [12], [13]. First, different from existing works which assume the aggregator trusted, we consider a honest-but-curious aggregator attack model, in which the FC may honestly perform sensing report aggregation while has a high interest in collecting users' location information. This attack model is justified by the recent researches, which show that the location privacy of mobile users might be compromised by the untrusted wireless service providers in illegitimate cases (e.g., worm based malware) or even legitimate cases (location based advertisement networks, theft locators, or Amber Alert services) [14], [15]. Therefore, if fusion centers are run by an untrusted service provider, it is possible for them to illegitimately track the individuals from the sensing report. Secondly, CR networks are characterized to be of a dynamic network topology, which makes the privacy preserving aggregation techniques for static networks unsuitable in CR networks. Further, the dynamic network topology and the assumption on untrusted FC introduce a new kind of attack towards the privacy preserving aggregation, named as *Differential Location Privacy* (DLP) Attack, in which the adversary could estimate a specific node's submitted sensing report and thus infer his location information by comparing the changes of the aggregation result if this node joins or leaves the network.

To address the above challenges in collaborative sensing, we first introduce a novel Location Privacy Model to quantify the privacy leakage in dynamic CR networks. Based on the proposed privacy model, we introduce a novel Privacy Preserving collaborative Spectrum Sensing (PPSS) scheme to achieve sensing report aggregation without location privacy leaking. PPSS is comprised of two primitive protocols: Privacy Preserving Sensing Report Aggregation protocol (PPSRA) and Distributed Dummy Report Injection protocol (DDRI). PPSRA enables the sensing devices to submit their encrypted sensing data to FC while FC could obtain the sum of all sensing reports without learning each individual's values. The PPSRA includes a novel self-organized key management scheme, which can well support the secondary users' dynamic join/leave in collaborative sensing. To further combat DLP attack, the proposed DDRI algorithm could prevent the changes of the aggregation data from leaking users' individual sensing report by adding some dummy report within a pre-defined time window.

We evaluate the effectiveness and efficiency of PPSS by implementing it in a real-world testbed. Our experiment results show that the malicious FC can geo-locate the secondary user by correlating its sensing reports and its physical location with an geographical accurate range of 10-50 meters and a successful rate of more than 90%. We also evaluate the performance of PPSS in terms of a series of performance metrics, such as privacy gain, computation overhead and impact on the performance of collaborative sensing. The extensive experiment results show that the proposed scheme is a practical approach to protect secondary users' location privacy in the collaborative sensing.

The contributions of this work are summarized as follows:

1) We identify and formulate a new security threat in collaborative sensing. Specifically, a malicious aggregator could compromise a secondary user's location privacy by correlating its submitted sensing report with its physical location.
2) We introduce a novel method, PPSS, to protect secondary users' location privacy in collaborative sensing. PPSS can work well in a dynamic CR network with untrusted FC. Furthermore, it can thwart DLP attack when the users join or leave the network.
3) We evaluate the effectiveness and efficiency of PPSS by implementating it in a real experiment. From the experiment, we demonstrate that a secondary user's physical location could be linked to its sensing results. It also shows that our PPSS scheme could successfully protect the user's location privacy in collaborative sensing.

The remainder of this paper is organized as follows. In section II, we present the problem formulation and system model. In section III, we introduce our scheme in details, which protects individual data privacy in report aggregation and injects Dummy Sensing report to eliminate the threats of the user's location privacy when he joins or leaves. We evaluate the performance of PPSS via experiment in Section IV, which is followed by the conclusion and future work.

II. Problem Formulation

A. Privacy Threats in Collaborative Spectrum Sensing

Although cooperative sensing can significantly improve the sensing accuracy compared to individual sensing, it raises a privacy concern: an attacker may try to geo-locate a CR user by his sensing reports, for instance, by the received signal strength (RSS) on TV band. So, like RSS based localization in WiFi or sensor networks, the attacker can infer the location of a CR user from his sensing report, coined as *Single CR Report Location Privacy* attacks (or SRLP attacks). Further, in addition to SRLP attacks for single report, sensing reports can also be used to compromise a user's location privacy in the aggregation mode. Specifically, when RSS measurements from multiple CR users have been aggregated to the sum, the adversary can still get a specific node's submitted reports and thus infer his location by comparing the aggregation variations if this node joins/leaves the network. We coin the second kind of attack as *Differential Location Privacy* Attacks (or DLP attacks). Then, we will use a series of experiments to show the practicality of SRLP and DLP attacks.

The experiment is taken in Building of Electronic Information and Electrical Engineering School at Shanghai Jiao Tong University. We use Universal Software Radio Peripheral (USRP) to detect the TV radio signal of 13 sampling regions within the building as shown in Fig. 1 (a), and these 13 regions basically cover the whole indoor area. We find significant location-dependent fluctuation in the RSS sensing of three Digital TV (DTV) channels (662-670MHz, 750-758MHz and 798-806MHz), which are enough to distinguish these 13 sampling places. The average signal strength of 3 DTV channel in 3 sampling regions is shown as follows:

region	662-670MHz	750-758MHz	798-806MHz
2	-25.3854	-19.8791	-29.3976
3	-26.7225	-26.5512	-27.6911
· · ·	· · ·	· · ·	· · ·
6	-19.6562	-17.0178	-22.6402

To localize a specific user with its sensing report, we adopt a machine learning method, K-Means to classify the data collected, and get the center $c_i, i = 1, \cdots, 13$ of each sampling place cluster [16]. Fig. 1 (b) shows the classification result of two channels with 3 sampling place clusters. In SRLP attack, we randomly choose a report r from the data pool of these 13 places and compare the distance between r and the training data, and if a report satisfies $\sum |r - c_i|^2 \leq \epsilon$, the location i

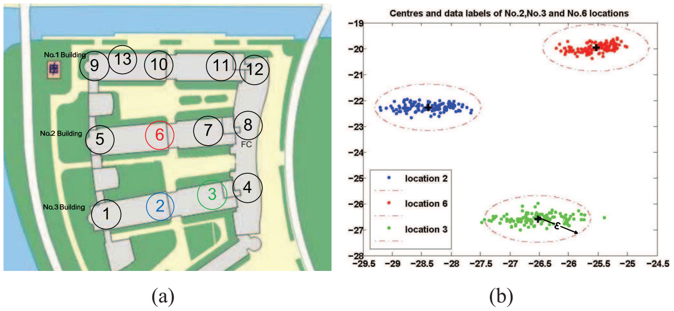


Fig. 1. (a) Sampling regions in the experiment, (b) Classification results of two channels with three sampling places:2,3,6.

is determined to be the possible position for the report. The result varies along with the parameter ϵ that a report may belong to none of these 13 places as well as several of them. We stipulate only when there is a single correct result which could be inferred out, then it is regarded as a successful attack. In DLP attack, we calculate the expectation of a leaving user's report in 10 rounds before and after the user's leave/join, then localize the user in the same way with SRLP attack.

Our experiment results show that if launching the SRLP attacks or DLP attacks, the attackers could localize a user within 10-50 meters accuracy with 90% confidence interval by choosing a proper parameter ϵ. We list the experiment results with different parameters in TABLE I. These results demonstrate that collaborative spectrum sensing may incur a serious privacy threat for secondary users without appropriate security guarantees.

Attack Type	ϵ	Max	Min	Average
SRLP	1.44	100%	76.92%	91.31%
	2.25	100%	92.31	99.15%
	4.00	61.54%	46.15%	56.77%
DLP	2.25	92.31%	46.15%	71.08%
	4.00	92.31%	53.85%	79.31%
	6.25	100%	69.23%	84.38%

TABLE I
THE ATTACKING SUCCESSFUL RATE UNDER DIFFERENT ϵ

B. System Model

We consider a centralized CR network, which consists of a FC and multiple secondary users in a range of 1 to $2km$ [4]. The set of secondary users is denoted by $\mathcal{U}_s = \{u_1, u_2, \ldots, u_n\}$. The primary users considered in this paper are mainly TV broadcasts, whose transmission power is nearly invariant. We assume each node is equipped with energy detectors. The choice of energy detection is due to its widespread acceptance and ease of implementation and analysis. During the collaborative sensing, each node reports its received signal powers to the FC, which aggregates these reports and makes the combined decision on the available spectrum. According to [10], the sensing reports have the following distribution:

$$r_i^k \sim \begin{cases} \mathcal{N}(N_0, \frac{N_0^2}{M}) & \mathcal{H}_0 \\ \mathcal{N}(P_i^k + N_0, \frac{(P_i^k+N_0)^2}{M}) & \mathcal{H}_1 \end{cases} \tag{1}$$

where $\mathcal{H}_0$ denotes the spectrum is occupied, and $\mathcal{H}_1$ represents the spectrum is idle. N_0 is the noise power, P_i^k is u_i's received signal power for spectrum k, and M is the signal sample number. The FC will combine the sensing reports as follows:

$$T_{\Sigma}^k = \sum_{i=1}^{n} w_i r_i^k, \tag{2}$$

where T_{Σ}^k is the decision statistic towards k-th spectrum, and w_i is the weight of the secondary user u_i in the aggregation. Without loss of generality, we adopt equal gain combination and set w_i to 1 in this study [10].

C. Attack Model and Assumptions

Consider an adversary $\mathcal{A}$ aiming to track the location of secondary users which are involved in CR networks. This adversary could be an external adversary, a compromised CR node or even the untrusted FC. Specifically, we consider the following three kinds of attacks:

- **Single Report Location Privacy (SRLP) Attack**: The adversary tries to compromise the location privacy of a CR user by correlating his sensing report and physical location.
- **Differential Location Privacy (DLP) Attack**: For the aggregated reports, the adversary aims to compromise a user's location privacy by estimating the aggregation difference before and after this node joins/leaves the network.
- **Collusion Attack**: The above two attacks could be further complicated by the collusion of two or more nodes. For example, the untrusted FC could collude with several compromised CR nodes to launch SRLP or DLP attack.

We assume that each node is pre-distributed with secret keys for mutually authenticating and securing the transmission links, which is different from the aggregation keys proposed in Section III. In this study, we only consider sensing report based location privacy related attacks. Falsely reporting attack [8]–[11], incentive issue [5]–[7], DoS attacks are not our focus. Measuring secondary beacon signal strength and employing RSS-based localization approach to localize the users are also out of the scope of this work [17]. We believe they deserve separate studies and there are existing works such as [5], [6], [8]–[11], [17], which have proposed a series of countermeasures to prevent these attacks.

D. Location Privacy Definition in Collaborative Sensing

To quantify the privacy leaking from the privacy preserving collaborative spectrum sensing under different attacks, we use the concept of entropy from Shannon's information theory [18] and have the definition on location privacy as follows:

DEFINITION 1 *Let $\mathcal{G} = \{g_1, g_2, \ldots, g_m\}$ be the set of spatial regions covered by the CR network and $\mathcal{U}_s = \{u_1, u_2, \ldots, u_n\}$ be the set of nodes in collaborative sensing. In privacy preserving collaborative sensing, the adversary observes the CR reports and predicts the possible matching between node $u_a \in \mathcal{U}_s$ and sub-region $g_b \in \mathcal{G}$. Let $p_{a|b} =$*

$Pr(u_a \in \mathcal{U}_s$ corresponds to $g_b \in \mathcal{G})$, which is the probability that user u_a is located in the sub-region g_b. We define the uncertainty of the adversary and thus the location privacy level of a node involved in a successful privacy preserving spectrum sensing to be

$$\mathcal{A}(a) = -\sum_{b=1}^{m} p_{a|b} log(p_{a|b}), \tag{3}$$

and the location privacy level for overall system as

$$\mathcal{A} = \sum_{a=1}^{n} \mathcal{A}(a). \tag{4}$$

It is easy to see that if there is no any privacy preserving techniques and the attacker can uniquely identify u_a's location from his sensing report, we can get $p_{a|b} = 1$, $\mathcal{A}(a) = 0$. On the other hand, the entropy is maximum for a uniform probability distribution $p_{a|b}$, which would provide node a with a location privacy level of $\log_2 m$. Note that, such a privacy definition can be applied to both SRLP and DLP attacks. The only difference is that the former predicts the matching between the node and its location based on a single report while the latter makes the prediction by estimating the aggregation differences before and after a user join/leave the networks.

III. The Proposed Scheme

In this section, we present *Privacy Preserving collaborative Spectrum Sensing* (or PPSS) scheme in details. The basic idea of PPSS is to exploit the diversity of sensing reports of different CR users to confuse the correlation between the report and users' location. In particular, to address SRLP and DLP attacks we propose two protocols, including PPSRA protocol, which utilizes data aggregation to hide sensing reports, and DDRI protocol, which prevents users' privacy from leaking during join/leave phase by injecting some dummy information. The detailed protocols are presented in the follows:

A. Privacy Preserving Sensing Report Aggregation protocol

PPSRA Protocol aims to hide the content of a specific sensing report by introducing a Privacy Preserving Aggregation (PPA) process. The proposed scheme is based on PPA algorithms proposed in [19], the basic idea of which is *secret sharing*. By sharing FC's secret among n participants, the aggregator cannot obtain the aggregation result unless he can collects all of the participants' reports, which enable FC to obtain the aggregation results without learning each individual's values.

However, the original PPA scheme is limited to the static environment, which may not be suitable for a dynamic CR network. Therefore, PPSRA further extends the PPA by sharing each user's private key among other users and the FC. This enables PPSRA to work well in a dynamic CR network in which some nodes may temporally join or leave the network.

1) System Parameter: Let $\{u_1, u_2, \ldots, u_{n-1}, u_n\}$ be the set of secondary nodes in CR networks and u_0 be the FC. We denote $\mathcal{U} = \{u_0, u_1, u_2, \ldots, u_{n-1}, u_n\}$ as all of participants in a spectrum sensing and $\mathcal{K} = \{sk_0, sk_1, sk_2, \ldots, sk_n\}$ is their corresponding secret keys. We represent the scanned spectrum as $\widetilde{C} = \{\widetilde{C}_1, \widetilde{C}_2, \cdots, \widetilde{C}_M\}$ and denote user u_i's sensing report on $\widetilde{C}_k$ as r_i^k. Let $\mathbb{G}$ denote a cyclic group of prime order p for which Decisional Diffie-Hellman is hard and $H : \mathbb{Z} \to \mathbb{G}$ denote a hash function modeled as a random oracle.

2) Key Generation: For any two nodes $u_i, u_j \in \mathcal{U}$, they will randomly generate their pairwise secret keys sk_{ij} and $sk_{ji} \in \mathbb{Z}_p$, such that $sk_{ij} + sk_{ji} = 0$. Therefore, the final secret key held by a node u_i could be represented by $sk_i = \sum_{j=0}^{n} sk_{ij}$. It is obvious that we could obtain the following equation:

$$\sum_{i=0}^{n} sk_i = \sum_{u_i, u_j \in \mathcal{U}} sk_{ij} + sk_{ji} = 0 \tag{5}$$

3) Sensing Report Encrypting: Each secondary user $u_i \in \mathcal{U}$ senses the spectrum $\widetilde{C}_k$ at the time slot t, and then encrypts his sensing report r_i^k with his secret key as follows:

$$c_i^k = g^{r_i^k} \cdot H(t)^{sk_i}. \tag{6}$$

Then u_i sends the encrypted sensing report c_i^k to the FC.

4) Aggregation Phase: After receiving the sensing reports from all CR users, the FC could obtain the final aggregated sensing results by computing:

$$V_k = H(t)^{sk_0} \prod_{u_i \in \mathcal{U}} c_i^k \tag{7}$$

Since $\prod_{u_i \in \mathcal{U}} c_i^k = g^{\sum_{i=1}^{n} r_i^k} \cdot H(t)^{\sum_{i=1}^{n} sk_i}$, with Equation (5), it is easy to derive $V_k = g^{\sum_{i=1}^{n} r_i^k}$. Therefore, to obtain the aggregated sensing result for time slot t, the FC needs to compute the discrete log of V_k base g and then obtain $\sum_{i=1}^{n} r_i^k$. Note that, the RSS value in collaborative sensing report is typically not large. In our experiment, RSS value varies in the range of [−30, 5], which makes the plaintext space quite small. As pointed out by [19], when the plaintext space is small, decryption can be accomplished via a brute-force search. To further speed up the decryption speed, Pollards lambda method is suggested for fast decryption, which requires decryption time roughly square root in the plaintext space.

5) Local Collaboration to Handle Users' Joins and Leaves: PPSRA can be adaptive to the dynamic CR networks where the users may temporarily join/leave. Specifically, whenever a user leaves/joins the network, to allow privacy preserving aggregation, it requires the key update so that the decryption could be successfully proceeded. In PPSRA, since the secret key of each CR user is shared by other $n-1$ nodes and the FC, the key update can be locally finished under the collaboration among the CR users and the FC when a node joins/leaves the networks. The detailed procedure is presented as follows.

- *User's Leave:* When a secondary user u_l leaves the CR networks, it will inform other participants by broadcasting a *LEAVE* message. After receiving the *LEAVE* message, each secondary user $u_i \in \mathcal{U}/u_l$ will update its secret key

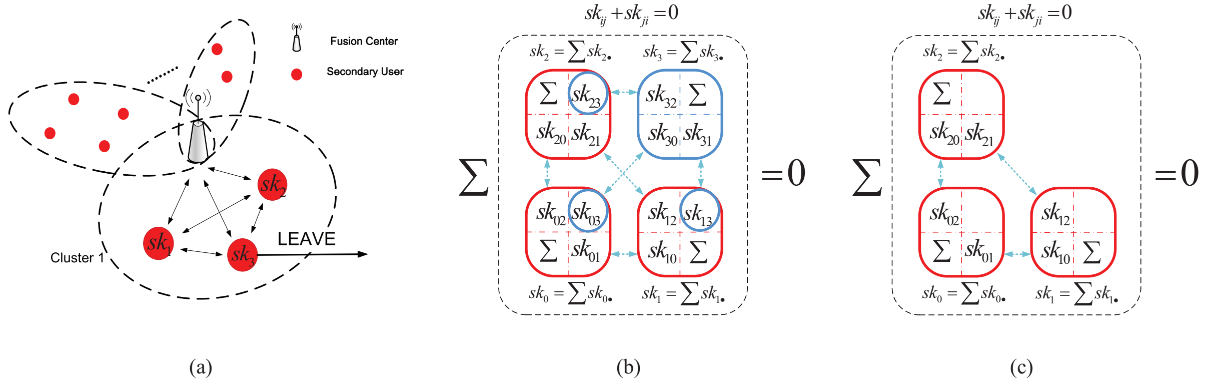


Fig. 2. **An illustration for key management in PPSRA:** (a) The CR users are divided into several clusters. (b) In each cluster, each pair of CR users and the FC cooperatively generates the keys. (3) When a CR user leaves the cluster, the remaining users would delete the keys shared with this user.

by removing its pairwise key sk_{il} with the user u_l. After that, u_i obtains his new secret key $sk_i' = \sum_{u_j \in \mathcal{U}/u_l} sk_{ij}$.

- *User's Join:* When a new secondary user u_{n+1} joins the CR networks, it informs other participants by broadcasting a *JOIN* message. After performing the mutual authentication with other nodes, u_{n+1} shares a new pairwise secret key with each node $u_i \in \mathcal{U}$. This new pairwise secret key is denoted as $sk_{(n+1)i}$ and $sk_{i(n+1)}$, such that $sk_{i(n+1)} + sk_{(n+1)i} = 0$. After this process, u_{n+1} obtains its overall secret $sk_{n+1} = \sum_{i=0}^{n} sk_{(n+1)i}$ while any other nodes' secret key is updated to $sk_i' = sk_i + sk_{i(n+1)}$.

In both cases, Equation (5) always holds, enabling PPSRA executed normally without a trusted third party's involvement.

6) Security Analysis: We will demonstrate the effectiveness of the PPSRA in concealing the sensing report r_i^k of the node $u_i \in U$ in a single collaborative sensing round. Firstly, FC has no idea about r_i^k. In PPSRA, the FC can only obtain the encrypted data c_i^k from u_i, and according to [19], FC cannot deduce the sensing report r_i^k if lacking the node's secret key sk_i. In addition, even when FC colludes with several nodes in the collaboration, he is still unable to get r_i^k. Because $sk_i = \sum_{j=0}^{n} sk_{ij}$, FC needs to collude with all the other nodes in the collaboration in order to get u_i's secret key. This means PPSRA can bear the collusion attack in most of cases.

7) Using Clustering to Reduce Key Management Complexity: By using the distributed key management, PPSRA requires any two nodes $u_{i,j} \in \mathcal{U}$ to communicate and have the key negotiation such that $sk_{ij} + sk_{ji} = 0$. Thus, PPSRA incurs a key management complexity of $\binom{n+1}{2} \cdot C_{tr}$, where C_{tr} refers to the transmission overhead incurred by the pairwise key negotiation. To reduce the key management complexity, we partition the whole networks into several virtual clusters, and in each cluster CR users and the FC can share their private key with each other. We show an example of CR network clustering as well as key management of PPSRA in Fig. 2. Suppose that cluster size is m. After clustering, the key management complexity C_{ost} of PPSRA could be reduced to

$$C_{ost}(m) \approx \lceil \frac{n}{m} \rceil \cdot \binom{m+1}{2} \cdot C_{tr} \tag{8}$$

It is clear that a larger cluster size leads to a higher key management complexity. But a larger cluster size is preferred for the sake of improved security level under collusion attacks. Next, we will discuss the tradeoff between Key Management Complexity and Security Level under Collusion Attack.

8) Discussions on Tradeoff between Key Management Complexity and Security Level under Collusion Attack: Generally, let m_c denote the minimum cluster size. We assume a subset of the participants may be compromised and collude with FC. If at least γ fraction of the participants are honest and not compromised, we should make sure that the possibility of recovering a specific node's secret key by collusion attack should be less than a certain possibility p_c. The following theorem discusses, given p_c, how to tune parameter m_c to reduce communication overhead in cluster formation phase.

Theorem 1: *Given n as the total number of secondary users in CR networks and γ as the fraction of the participants who are honest and not compromised, to ensure that the collusion attack successful rate is less than the threshold p_c, the minimum key establishment overhead for the network is*

$$C_{ost}(m_c) \approx \lceil \frac{n}{\lceil \log_{(1-\gamma)} p_c \rceil} \rceil \cdot \binom{\lceil \log_{(1-\gamma)} p_c \rceil + 1}{2} \cdot C_{tr} \tag{9}$$

Proof: To successfully launch a collusion attack towards a node u_i, the FC needs to collude with all of the nodes except u_i in the same cluster. Therefore, we obtain $(1-\gamma)^{m_c} \leq p_c$ and thus $m_c \geq \log_{(1-\gamma)} p_c$. By substituting m_c into m of Equation (8), we have the minimum key establishment overhead as above. Note that, the equation (9) only defines the lower bound of the key establishment overhead. In practice, to tolerate l nodes temporarily leaving the networks, we need to set m_c larger than $\log_{(1-\gamma)} p_c + l$ to ensure than the security level is always above the threshold.

B. Distributed Dummy Report Injection Protocol

In the previous section, we present PPSRA, which could effectively protect the collaborative sensing participants from leaking their location privacy via privacy preservation aggregation. However, as shown in section II, a user's join/leave will also leak its location privacy. In traditional differential privacy literature, standard procedure for ensuring differential privacy is for FC to add an appropriate magnitude of noise or each participant adds the noise in a distributed way before publishing the desired statistic [20]. However, adding noise to the sensing reports may seriously degrade the performance of collaborative sensing, which obviously deviates from the original goal of collaborative sensing. To address this problem, we introduce a *Distributed Dummy Report Injection protocol (or DDRI)* to protect the DLP of the secondary users.

1) The Proposed DDRI Algorithm: The basic idea of the DDRI is: during the user leaving/joining phase, other users could use dummy sensing reports r_0^k, which could be provided by FC's own sensing [4] (or any voluntary secondary user), to replace their real sensing report. Different from the traditional noise based differential privacy protection technique which may have a negative effect on collaborative sensing, such a dummy report based approach will not *pollute* the aggregation result. Instead, it only increases the weight of a real sensing report from the FC of the overall aggregation result and reduce the number of real participants involved in the collaborative sensing, which are two major metrics considered in the subsequent performance analysis part. Next, we present the detailed algorithm in Algorithm 1.

Algorithm 1: Distributed Dummy Report Injection

```
for A secondary user joins or leaves do
    for Each node $u_i \in \mathcal{U}/u_l$ do
        Randomly choose noise parameter $\delta_i$ from
        $\mathcal{N}(\mu, \sigma^2)$;
        Generate a random number $\tau \in [0, 1]$ ;
        for the subsequent $T^k$ time slots do
            if Sensing report $r_i^k$ doesn't change then
                if $\tau \le \delta_i$ then
                    Submit the sensing report $r_0^k$;
                else
                    Submit the sensing report $r_i^k$;
                end if
            else
                break;                //Remove the noise of $u_i$
            end if
        end for
    end for
end for
return valid;
```

Upon leaving or joining the CR networks, a user u_l broadcasts a *LEAVE* message to all the nodes in the network. A remaining node $u_i \in \mathcal{U}/u_l$ will randomly generate the parameter δ_i. During the time window T^k, u_i injects the dummy sensing report of r_0^k at the probability of δ_i while submitting his real sensing report at the probability of $1 - \delta_i$. Here, $\delta_i, u_i \in \mathcal{U}/u_l$ follows the following distribution:

$$\forall u_i \in U/u_l, \delta_i \sim \mathcal{N}(\mu, \sigma^2) \tag{10}$$

where, μ and σ are two predefined parameters.

This algorithm will be executed within time window T^k. However, it is not desirable for all the nodes to stop injecting the dummy message at the same time. This is because the adversaries could derive the value of noise by comparing the aggregation results before and after T^k and then obtain the real sensing report of u_l to launch DLP attack. Here, estimating the value of noise within and after the time window is similar to estimating a leaving node's sensing report in DLP attack. Therefore, to prevent injected dummy report from leaking, we require that each node $u_i \in \mathcal{U}/u_l$ stop injecting the dummy message in a distributed way, which is, before time window T^k, each node u_i could stop injecting if its sensing report changes, and such a change is large enough to conceal the user's noise. Therefore, at the end of this window time, all remaining dummy reports could be eliminated and the system will become normal. The selection of window time should ensure that in this period the users' sensing reports r_i^k have changed and should be more than a pre-defined threshold Φ.

2) Effectiveness Analysis of DDRI: In DLP attacks, the attacker violates the CR user's location privacy by executing differential operation on the aggregation results when this user leaves/joins the networks. The approach can be described as:

$$\widehat{r}_l^k = \widehat{\mu}(\sum_{u_i \in U} r_i^k) - \widehat{\mu}(\sum_{u_i \in U/u_l} r_i^k) \tag{11}$$

where $\widehat{\mu}$ is the estimator for expectation. When the samples used for estimation are large enough, the value $\widehat{r}_i^k$ should converge to the expected value of r_l^k. This is because:

$$E[r_l^k] = E[\sum_{u_i \in U} r_i^k] - E[\sum_{u_i \in U/u_l} r_i^k] \tag{12}$$

After obtaining the estimated $\widehat{r}_l^k$, the attacker could infer the user's leaving/joining location.

With DDRI, a leaving/joining secondary user's sensing report, which may be leaked from DLP attack, is protected by injecting the dummy reports. Specifically, after executing DDRI, a joint noise n^k generated in the distributed manner will be introduced into the estimation of r_l^k's mean, and such noise is described in the following theorem:

Theorem 2: *Given the value μ,σ, the noise n^k introduced to the estimation of r_l^k's expectation follows the distribution:*

$$n^k \sim \mathcal{N}(\sum_{u_i \in U/u_l} \mu E[r_0^k - r_i^k], \sum_{u_i \in U/u_l} \sigma^2 (E[r_0^k - r_i^k])^2) \tag{13}$$

Proof: See the Appendix VII-A.

If this noise is large enough and unknown by the attacker, then the DLP attack should be no longer effective. From the equation (13), it is observed that by selecting proper parameters μ and σ, the secondary users could successfully generate a large joint noise. And, at the same time, due to the uncertainty of the value δ_i as well as $E[r_0^k - r_i^k]$, the attacker is unable to derive such a noise. Thus, DLP attack is no longer effective when the our proposed DDRI protocol is in place.

3) The Impact on Collaborative Sensing: In this section, we discuss the impact of our scheme on collaborative sensing performance. In DDRI protocol, when a secondary user leaves or joins CR networks, every other secondary user submits a dummy report from FC to generate the noise with a certain probability. Therefore, DDRI could improve the location privacy of the leaving nodes at the cost of reducing the collaborative sensing performance. To measure the impact on collaborative sensing, we first define the concept of actual cooperator number A_n in collaborative sensing.

Definition 2: *The number of actual cooperators A_n is defined as the number of the cooperators who submit their authentic sensing reports in a single time slot.*

The number of actual cooperators is upper bounded by the value n, and $A_n = n$ when the system operates in traditional mode. After executing DDRI protocol, if the actual cooperator number is close to n, our protocol should have little impact on the collaborative sensing, and vise versa. To measure DDRI's impact on collaborative sensing, in the following theorem, we can give the actual cooperator number in our DDRI:

Theorem 3: *Given μ and σ, the expected value of the actual cooperative number follows the distribution:*

$$E[A_n] \sim |U/u_l| - \mathcal{N}(\mu \cdot |U/u_l|, |U/u_l| \cdot \sigma^2) \qquad (14)$$

where $|U/u_l|$ denotes the cardinality of the set U/u_l.

Proof: See the Appendix VII-B.

Besides, another performance metric to evaluate the impact on collaborative sensing is the weight w_o of FC's sensing report, which is increased along with the increased usage of the dummy reports. We could have following theorem:

Theorem 4: *Given the parameter μ and σ, after the secondary user's leaves or enters, the FC's expected weight w_0 follows the following distribution:*

$$E[w_0] \sim \mathcal{N}(\mu \cdot |U/u_l|, |U/u_l| \cdot \sigma^2) + 1 \qquad (15)$$

Proof: See the Appendix VII-C.

From the equation (15), we find when the system adopts a small μ and σ, the expectation of the number of actual cooperator is close to the upper bound while the weight of the FC report is close to 1, which means that little impact on sensing performance is introduced by DDRI. In the experiment, we would demonstrate a small μ and σ is enough to effectively protect the user's differential location privacy, and therefore our scheme has little impact on the collaborative sensing.

IV. Evaluations

In this section, we evaluate the effectiveness and efficiency of the proposed PPSS from following aspects: 1) Setup of our experiments; 2) Evaluation of the SRLP and DLP attack; 3) Evaluation of the computational overhead; 4) Effectiveness of PPSS and 5) Impact of PPSS on collaborative sensing.

A. Experiments Setup

The experiment is set up as follows. We use USRP with a TVRX daughterboard (50 MHz to 860 MHz Receiver) and a wide band antenna (70 MHz to 1000 MHz) to detect the TV broadcasts in the building. Then in order to build a spectrum sensing database of the sampling places, we scan the channel from 600 MHz to 860 MHz at these 13 places with each spectrum scanned for 10 seconds while every 8 MHz spectrum scan costs 33ms.

To evaluate the SRLP attack, we first place the CR user in a certain location, and execute spectrum sensing to get the corresponding sensing results. We use the database to geo-locate this user. The geo-location algorithm adopted has been mentioned in Section II-A and the obtained result is the user's possible location set, which may include the locations the CR user doesn't belong to, which is coined as the false locations.

After obtaining the possible location set, we could calculate the privacy level according to Location Privacy Definition I, which is shown in Section II.D. We execute the same experiment for 100 rounds to get the expected privacy level at that location, and do the same experiments in other locations to obtain the privacy level expectation over all locations. Note that, in the privacy entropy calculation, if the possible location set doesn't include the user's correct location, then we set the entropy to be $log(m)$, which is the maximum entropy value in our experiment. This is because, from the attacker point of view, he cannot obtain any useful information from the false location set, which makes SRLP attack invalid.

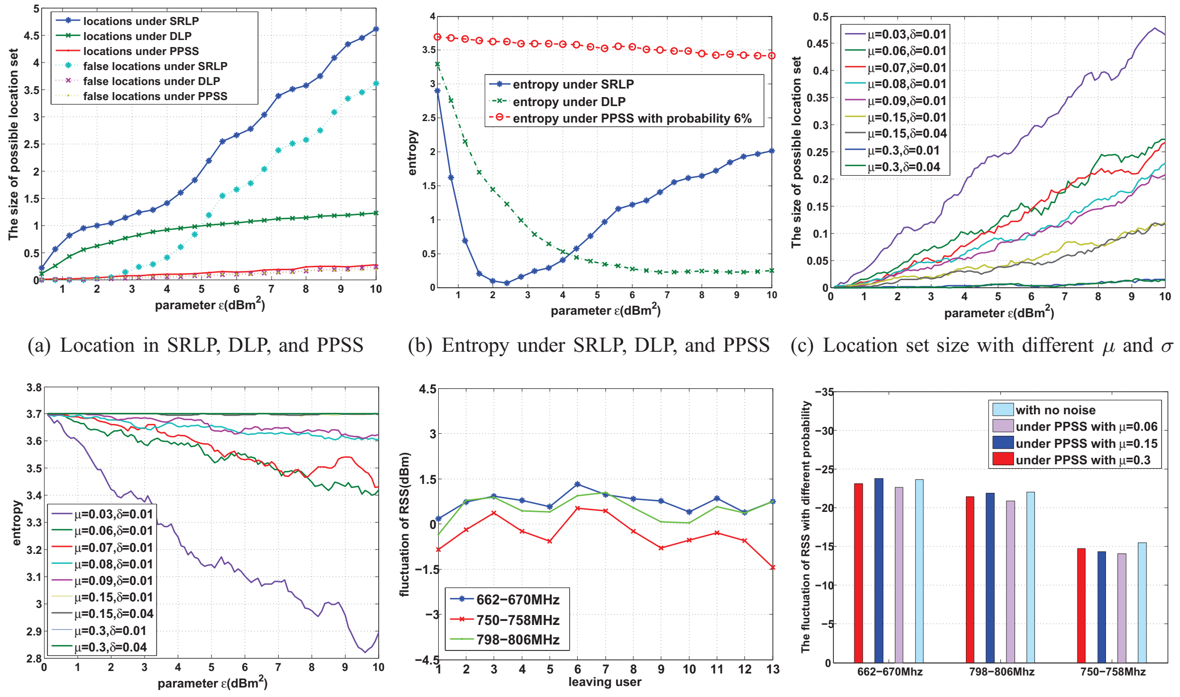
For the evaluation of DLP attack, we assume 13 secondary users are located in these 13 places separately, and select one of them to leave the collaboration. We utilize 10 samples respectively to estimate the expected aggregation result before and after a secondary user's leave, and obtain the secondary user's reports to geo-locate him. The experiment is also executed for 100 rounds to obtain the expected results. Furthermore, we execute the same experiment over all other secondary users to obtain the expected value over all users.

B. Evaluation of the SRLP and DLP Attacks

We start our evaluation on the practicality of SRLP attack and DLP attacks from number of possible locations. The experiment results are shown in Fig. 3 (a) and Fig. 3 (b).

In Fig. 3 (a), it shows the average number of possible locations under SRLP and DLP attacks, which is compared with the one protected with PPSS protocol. In Fig. 3 (a), it is observed that when ϵ remains lower than 1, the obtained number of possible locations and false locations are both less than 1, which means the attack can rarely bring useful information to the attacker. This result is mainly due to the fluctuation of the spectrum measurements, which indicates most of sensing reports keep at least ϵ distance from the cluster centroid. As ϵ raises into the range $1 - 4dBm^2$, which means the allowed error range of a channel is $1 - 2dBm$, the SRLP attack have the best performance with more than 90% correct locations and almost no false locations. When adopting a larger ϵ and therefore larger allowed error range, the false location number will increase dramatically compared with the correct location number, and the effectiveness of attack will degrade. From the Fig. 3, we could also see that a larger parameter ϵ is favorable in DLP attack, this is because the user's sensing reports obtained by DLP attack have more fluctuation.

In Fig. 3 (b), it shows the entropy of SRLP and DLP attack varies when the parameter ϵ changes from 0 to 10 at the



(a) Location in SRLP, DLP, and PPSS (b) Entropy under SRLP, DLP, and PPSS (c) Location set size with different μ and σ

(d) The entropy with different μ and σ (e) The fluctuation of RSS under PPSS (f) The fluctuation of RSS with different μ

Fig. 3. The evaluation results about the SRLP attack, DLP attack, the effectiveness of PPSS, and PPSS's imapct on collaborative sensing

interval of 0.4. We see when parameter ϵ takes the value $1-3dBm^2$ under SRLP attack and $7-10dBm^2$ under DLP attack, the entropy will have the small value close to 0, which means the attacker has more certainty about the user's location.

C. Evaluation of Computation Overhead

We evaluate the computation overhead incurred by PPSRA scheme. Suppose there are n secondary users and each of them utilizes v bits to represent his sensing results. Then in the encryption phase, the secondary users and the FC should execute a hash operation, two modular exponentiations, and one multiplication in Diffie-Hellman group, respectively. In the decryption phase, the fusion center should execute $n+1$ multiplication in Diffie-hellman group and $\frac{2^v n}{2}$ at average modular exponentiations. According to the benchmarking data in [21], the computation overhead are dominated by the modular exponentiation, and when we adopt prime order 1024 bit for group $\mathbb{G}$ and the curve "25519", one modular exponentiation needs roughly 0.3ms in desktop PC. Therefore, encryption could be executed in 0.6ms, and for decryption, if we adopt the parameter n=10, v=4, the total computation time is roughly 48ms for one aggregation. When utilizing the Pollard's lambda method, this computation time would be reduced to 6.93ms. Such a computational overhead can satisfy the real-time requirements of collaborative sensing, in which the time interval for two regular CR sensing is 2s [10].

D. Evaluation of the Effectiveness of PPSS

In this section, we evaluate the effectiveness of PPSS scheme when protecting the user's differential privacy. In the experiment, we set the parameter $\mu = 0.06$ and $\sigma = 0.1$, from the Fig. 3 (a), we could see that the obtained correct location number decreases dramatically no matter what value the parameter ϵ takes. In the Fig. 3 (b), after executing PPSS, the entropy maintains the value about 3.7, which indicates that the attacker has a high uncertainty about the user's location.

We also evaluate the impact of the parameter μ and σ when preserving the user's differential location privacy. In the Fig. 3 (c) and Fig. 3 (d), we can see when the system adopts a larger parameter μ, the obtained possible location number will decrease and the entropy will increase. This result shows our scheme can effectively protect the user's differential privacy. It also shows that the larger μ is, the larger noise is introduced to the user's sensing reports which is obtained by DLP attack.

E. PPSS's Impact on Collaborative Sensing

The PPSS's impact on the system performance is shown in Fig. 3 (e) and Fig. 3 (f). It can be seen that for the fluctuation of RSS in different channels, the maximum change is lower than $1.5dBm$, which means our scheme affects little on the system performance. Furthermore, with different probabilities μ, the variations fall in such a reasonable range that the greatest change is less than 2dBm, as showed in Fig. 3 (f). This demonstrates that PPSS has little negative effect on the performance of collaborative sensing.

V. Conclusion

In this paper, we identify the location privacy leakage problem in collaborative sensing and focus on two potential attacks, SRLP and DLP. To address these two new location privacy threats, we propose PPSS scheme consisting of a basic scheme, PPSR, and an advanced scheme, DDRI. PPSRA enables the CR user to conceal his reports in aggregation, while DDRI can protect the user's DLP. We evaluate the location privacy leakage in collaborative sensing and demonstrate the effectiveness of PPSS by implementing it in a realistic testbed. In the future work, we would investigate the location privacy issues in ad hoc CR networks.

ACKNOWLEDGEMENT

This research was supported by National Natural Science Foundation of China (Grant No. 61003218, 61033014, 61161140320, 70971086, 61174127, 60974123, 61170267, and 60934003), 973 Program (Grant No.2011CB302905), and Doctoral Fund of Ministry of Education of China (Grant No.20100073120065).

REFERENCES

[1] I. F. Akyildiz, W. Y. Lee, M. C. Vuran, and S. Mohanty, "Next eneration/dynamic spectrum access/cognitive radio wireless networks: a survey," *Computer Networks*, September 2006.

[2] CogNea: Cognitive Networking Alliance. http://www.cognea.org/.

[3] IEEE 802.22 WRAN WG on Broadband Wireless Access Standards. www.ieee802.org/22.

[4] P. Bahl, R. Chandra, T. Moscibroda, R. Murty, and M. Welsh,"White space networking with wi-fi like connectivity," in *SIGCOMM'09*.

[5] C. Song and Q. Zhang, "Achieving cooperative spectrum sensing in Wireless cognitive radio networks," in *ACM MC2R, Special Issue on Cognitive Radio Technologies and Systems*, Vol. 13, Issue 2, April 2009.

[6] B. Wang, K.J. Liu, and T.C. Clancy, "Evolutionary cooperative spectrum sensing game: how to collaborate?" *IEEE Trans. on Communications*, vol.58, no.3, pp.890-900, March 2010.

[7] H. Zhu, X. Lin, R. Lu, Y. Fan, and X. Shen, SMART: A Secure Multi-Layer Credit based Incentive Scheme for Delay-Tolerant Networks, *IEEE Trans. on Vehicular Technology*, vol.58, no.8, pp.4628-4639, 2009.

[8] O. Fatemieh, A. Farhadi, R. Chandra, and C.A. Gunter, "Using classification to protect the integrity of spectrum measurements in white space networks," in Proc. of *NDSS*, 2011.

[9] H. Li and Z. Han, "Catch me if you can: an abnormality detection approach for collaborative spectrum sensing in cognitive radio networks," *IEEE Trans. on Wireless Communications*, vol.9, no.11, pp.3554-3565, 2010.

[10] A. Min, K. Shin, X. Hu, "Secure cooperative sensing in IEEE 802.22 WRANs using shadow fading correlation," *IEEE Trans. on Mobile Computing*, vol.10, no.10, pp. 1434-1447, 2011.

[11] S. Li, H. Zhu, B. Yang, C. Chen, and X. Guan,"Believe yourself: A user-centric misbehavior detection scheme for secure collaborative spectrum sensing", in Proc. of *ICC*, 2011.

[12] W. He, X. Liu, H. Nguyen, and K. Nahrstedt, and T. Abdelzaher, "PDA: privacy-preserving data aggregation in wireless sensor networks," in Proc. of *INFOCOM'07*, 2007.

[13] J. Shi, R. Zhang, Y. Liu, and Y. Zhang, "Prisense: privacy-preserving data aggregation in people-centric urban sensing systems," in Proc. of *INFOCOM'10*, 2010.

[14] P. Wang, Z. Gao, X. Xu, Y. Zhou, H. Zhu and K.Q. Zhu,"Automatic inference of movements from contact histories," in *SIGCOMM'11*, 2011.

[15] N. Husted, S. Myers, "Mobile location tracking in metro areas: malnets and others,"in Proc. of *CCS'10*, 2010.

[16] C.M. Bishop,"Pattern recognition and machine learning,"Springer, 2006.

[17] T. Wang, and Y. Yang, "Location privacy protection from RSS localization system using antenna pattern synthesis," in *INFOCOM'11*, 2011.

[18] C.E. Shannon, and W. Weaver, "The mathematical theory of communication," Citeseer, 1959.

[19] E. Shi, T. Chan, E. Rieffel, R. Chow, and D. Song,"Privacy-preserving aggregation of time-series data," in Proc. of *NDSS'11*, 2011.

[20] C. Dwork "Differential privacy," Invited talk at *ICALP*, 2006.

[21] D. J. Bernstein and T. L. (editors). eBACS: ECRYPT benchmarking of cryptographic systems. http://bench.cr.yp.to, accessed 10th, July.

APPENDIX

A. Proof of Theorem 2

After the CR user u_l leaves/enters the network, the expected sensing reports $\widetilde{r}_i^k$ submitted by user $u_i \in U/u_l$ is:

$$\begin{aligned} E[\widetilde{r}_i^k] &= (1-\delta_i)E[r_i^k] + \delta_i E[r_0^k] \\ &= E[r_i^k] + \delta_i E[r_0^k - r_i^k] \end{aligned} \tag{16}$$

Then, the user i introduces the noise $\delta_i E[r_0^k - r_i^k]$ to the estimation of r_l^k's expectation. Since the noise parameter $\delta_i, i \in U/u_l$ follows the normal distribution $\mathcal{N}(\mu, \sigma^2)$, the generated noise by the individual follows:

$$\delta_i E[r_0^k - r_i^k] \sim \mathcal{N}(\mu E[r_0^k - r_i^k], \sigma^2(E[r_0^k - r_i^k])^2)$$

Since the noises are generated independently by the CR users, then the sum of these noises have the following distribution:

$$\begin{aligned} &\sum_{u_i \in U/u_l} \delta_i E[r_0^k - r_i^k] \\ &\sim \mathcal{N}(\sum_{u_i \in U/u_l} \mu E[r_0^k - r_i^k], \sum_{u_i \in U/u_l} \sigma^2(E[r_0^k - r_i^k])^2) \end{aligned} \tag{17}$$

B. Proof of Theorem 3

When given parameter $\delta_i, i \in U/u_l$, the actual cooperator number follows Poisson binomial distribution, and the probability mass function could be written as follows:

$$\mathbb{P}(A_n = k) = \sum_{A \in F_k} \prod_{i \in A} (1 - \delta_i) \prod_{j \in A^c} \delta_j \tag{18}$$

where F_k is the set of all subsets of k indexes that can be selected from U/u_l, and A^c is the complement of A. Further, we could obtain the expected value of the A_n:

$$E[A_n] = \sum_{u_i \in U/u_l} (1 - \delta_i) \tag{19}$$

Notice that the parameter δ_i follows the distribution $\mathcal{N}(\mu, \sigma^2)$, and these parameters are generated independently, we could obtain the distribution of the expected value of A_n:

$$E[A_n] \sim |U/u_l| - \mathcal{N}(\mu \cdot |U/u_l|, |U/u_l| \cdot \sigma^2) \tag{20}$$

C. Proof of Theorem 4

In each time slot, the relationship between A_n and the w_0 in aggregation could be shown as:

$$w_0 = |U/u_l| - A_n + 1 \tag{21}$$

Then, we could have the expected w_0:

$$E[w_0] = |U/u_l| - E[A_n] + 1 \tag{22}$$

Substituting $E[A_n]$ with the equation (19), we could obtain:

$$E[w_0] = \sum_{u_i \in U/u_l} \delta_i + 1 \tag{23}$$

Then given the parameter $\delta_i \sim \mathcal{N}(\mu, \sigma^2), u_i \in U/u_l$, the expected value of w_0 follows the distribution:

$$E[w_0] \sim \mathcal{N}(\mu \cdot |U/u_l|, |U/u_l| \cdot \sigma^2) + 1 \tag{24}$$