Primary Users' Operational Privacy Preservation via Data-Driven Optimization

Jingyi Wang*, Yanmin Gong†, Lijun Qian‡, Riku Jäntti§, Miao Pan*, and Zhu Han*
*Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77204
†School of Electrical and Computer Engineering, Oklahoma State University, Stillwater, OK 74078
‡Department of Electrical and Computer Engineering, Prairie View A&M University, Prairie View, TX 77446
§Department of Communications and Networking, Aalto University, Espoo, Finland

***Abstract*—Recently opened spectrum within 3550-3700 MHz provides more accessing opportunities to secondary users (SUs), while it also raises concerns on the operational privacy of primary users (PUs), especially for military and government. In this paper, we propose to study the tradeoff between PUs' temporal privacy and SUs' network performance using the data-driven approach. To preserve PUs' temporal operational privacy, we develop an obfuscation strategy for PUs, which allows PUs to intentionally add dummy signals to change the distribution of temporal spectrum availability, and confuse the adversary. While generating the dummy signals for privacy, the PUs have to consider the utility of SUs and try their best to satisfy SUs' uncertain traffic demands. Based on the historical data, we employ a data-driven risk-averse model to characterize the uncertainty of SUs' demands. With joint consideration of PUs' privacy and uncertain SUs' demands, we formulate the data-driven risk-averse stochastic optimization, and provide corresponding solutions. Through numerical simulations, we show that the proposed scheme is effective in preserving PUs' temporal operational privacy while offering good enough spectrum resources to satisfy SUs' traffic demands.**

I. Introduction

In recent years, the exploding increase of mobile wireless devices and the proliferation of wireless services have accelerated the growth in demand for radio spectrum. With limited unlicensed spectrum, regulators are turning to dynamic spectrum sharing and looking for advanced techniques to improve spectrum utilization. As one promising technology, cognitive radio (CR) allows secondary users (SUs) to access the idle spectrum in temporal and spatial domain opportunistically, when primary users (PUs) are not active. To further meet the ever-increasing demand for spectrum, Federal Communication Commission (FCC) and National Telecommunications and Information Administration (NTIA) have agreed to open up the 3550-3700 MHz band for unlicensed communications [1], [2]. Note that most frequencies within 3550-3700 MHz are traditionally used by government agencies, e.g., Department of Defense [2], [3], and the operational information (such as time of use, geographical locations, anti-jamming capability, and so on) of government facilities, e.g., military radars, are very sensitive or even classified. Therefore, how to maintain the PUs' operational privacy while increasing SUs' spectrum accessing opportunities poses great challenges.

This work was supported by the U.S. National Science Foundation under grants CNS-1343361, CNS-1350230 (CAREER), and CNS-1702850.

There are several pioneering works about PUs' privacy preservation in existing literature. For example, Clark et al. in [3] discussed several attack models and PUs' obfuscation strategies, based on the assumption that all the information of PUs and SUs are stored in a database. The adversary might hack into the database or compromise SUs' devices to infer PUs' location information. Robertson et al. in [4] proposed to add false spectrum allocation entries into the database to prevent the adversary from learning the operational privacy of PUs. Bahrak et al. in [5] employed obfuscation methods to preserve PUs' privacy by constructing forbidden contours. A perturbation strategy was used to develop a pentagon-shaped contour, which envelops the PU's actual contour to preserve PU's location privacy. Dwork et al. [6] added noises in the sensing reports before publishing the desired sensing statistics, which achieved differential privacy of PUs. Since simply adding noise signals may degrade the performance of collaborative sensing results, Gao et al. in [7] further proposed a distributed dummy report injection protocol, which jointly prevents the pollution of the aggregation results and preserves location privacy of PUs. Based on attributed-based encryption techniques, Liu et al. in [2] developed the query policy for PUs' spectrum usage database to protect PUs' location privacy. In military communications, Fu et al. in [8] proposed a method that hides traffic characteristics from eavesdroppers by padding the traffic with constant/variable interarrival times, to mitigate the traffic analysis attacks. In addition, there are some previous works related to the time-based traffic model. For instance, Bonal et al. in [9]–[11] show that if the underlaying scheduler is fair, the flow-level (TCP) [12] throughput and delay admit simple time based form, which is independent of the actual inter-arrival distribution between MAC layer packets. However, most existing schemes are limited about the privacy preservation of PUs' temporal operational privacy (i.e., the time of usage). The temporal operations of PUs might include highly confidential or even classified information (e.g., the operational time of military radars). If such information is obtained by a malicious party, it may jeopardize national security and people's safety [13]. In addition, most existing PUs' privacy preserving designs have very limited consideration about how to create more accessing opportunities to satisfy SUs' traffic demands and improve spectrum utilization, which is the original purpose for opening up 3550-3700 MHz

978-1-5090-5019-2/17/$31.00 ©2017 IEEE

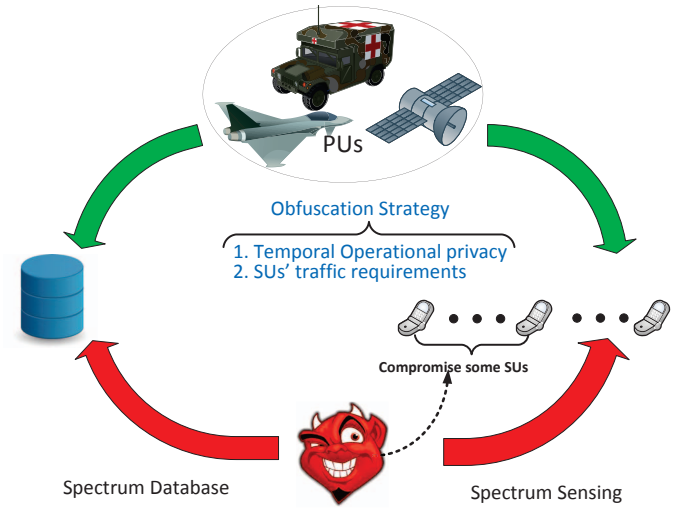

Fig. 1. System architecture and temporal operational attacks.

band for CR communications.

From the aspect of PUs' obfuscation strategy designing, it will be good for the PUs to precisely predict SUs' traffic demands. In this way, the PUs can intentionally add dummy signals to obfuscate the attackers[1] while trying their best to satisfy SUs' traffic demands. However, it is a challenging problem to characterize the uncertainty of SUs' traffic demands. Some previous efforts tried to employ robust optimization to address this issue. For instance, Lunden et al. in [14] proposed a robust computationally nonparametric cyclic correlation estimator, which does not require the distribution information of users' traffic. Gong et al. in [15] designed an algorithm to search optimal detection bound considering signal uncertainty. However, the robust optimization approach can be very conservative, since its objective is to minimize the worst case cost or worst case effectiveness. If PUs add too many dummy signals, according to the overly conservative analysis for privacy preservation, it would reduce the utility of SUs.

To address the issues above, we propose a novel PUs' obfuscation strategy design, formulate the PUs' operational privacy preservation problem into a data-driven risk-averse optimization, and provide robust solutions. Our salient contributions are summarized as follows:

- We introduce a new privacy preserving framework for PUs' obfuscation strategy design, which jointly considers PUs' operational privacy in the temporal domain, the obfuscation cost of PUs, the uncertainty of SUs' demands, and SUs' traffic demand satisfaction. Under such a framework, when PUs add dummy signals to obfuscate the adversary, they also need to consider the trade-off between preserving PUs' temporal privacy and satisfying SUs' traffic requirements, and thus cannot arbitrarily generate dummy signals for privacy preserving purposes.
- Under the proposed framework, with abundant historical data of SUs' traffic demands, we propose to let the PUs employ data-driven modeling to characterize the uncertainty of SUs' traffic demands. The PUs can build up a reference SUs' demand distribution from the historical data, and generate the predicted SUs' demand distribution, which is close enough to the reference distribution at a certain confidence level.
- Based on the modeling of SUs' uncertain traffic demands and temporal operational privacy metrics, we formulate the PUs' temporal privacy preservation problem into a risk-averse two-stage stochastic optimization, develop algorithms for robust solutions, and conduct simulations to verify our theoretical analysis.

The rest of paper is organized as follows. In Sec. II, we introduce the network model and formulate the PUs' operational privacy preservation problem. In Sec. IV, we develop the solutions to the proposed problem. Simulation results and discussions are presented in Sec. V, and the conclusion remarks are drawn in Sec. VI.

II. System Description

A. Network Configuration

As shown in Fig. 1, we consider a CR network consisting of $\mathcal{N} = \{1, 2, \cdots, i, \cdots, N\}$ SUs and $\mathcal{M} = \{1, 2, \cdots, j, \cdots, M\}$ PUs sharing 3550-3700 MHz bands. Following the principles of CR communications, SUs can opportunistically use the band when the PU owning that band is not active, and SUs must evacuate if the PU comes back. Here, we assume each PU is licensed to use a dedicated band, and each SU can only opportunistically access one band at a time. For an available spectrum band, different SUs are allowed to dynamically share the band in time-division multiple access (TDMA) manners. To preserve temporal operational privacy, PUs will send obfuscating dummy signals periodically, where the fixed period is denoted by $\mathcal{T}$. Let T_j represent the actual temporal spectrum availability for band j (i.e., available time for SUs' opportunistic spectrum accessing before PU j adds dummy signals), and y_j ($y_j \leq T_j$) be the transformed temporal spectrum availability for band j (i.e., the available time for SUs' opportunistic spectrum accessing after PU j adds dummy signals). Given the transmission rate, let a random variable ξ_i denote the required time to deliver the uncertain traffic demands of SU i within $\mathcal{T}$. For simplicity, we call ξ_i the demand of SU i in the rest of this paper, and let $\mathbb{P}_i$ be the distribution of ξ_i. For instance, $\mathcal{T} = 60$ mins, and PU j is actively using band j for 20 mins, so that T_j is equal to 40 mins. After PU j executes obfuscation strategy, $y_j = 30$ mins, and the demand of SU i is $\xi_i = 25$ mins.

B. Attack Model

In this work, we consider passive adversaries, who may learn the operational time of PUs from either spectrum database or collective spectrum sensing results by compromised SUs. The compromised SUs do not intercept or modify the messages sent by PUs. Specifically, the adversaries can either eavesdrop the communication between the spectrum

[1]It refers to the attackers either hacking into the spectrum usage database or employing multiple SUs to sense in order to learn the PUs' operational parameters [3].

database server and SUs or send queries to the database to learn spectrum availability in the database-driven approach [3], [5], or compromise some SUs' devices and collect spectrum sensing[2] results to infer PUs' operational characteristics in the spectrum sensing approach [3], as shown in Fig. 1.

III. Obfuscation Strategy and Problem Formulation

A. Utility Functions of PUs and SUs

From the PU's perspective, to preserve the temporal operational privacy from passive attackers, the PU executes obfuscation strategy by generating dummy signals for a certain time period when it actually has no traffic. In that way, the adversary cannot distinguish them from true signals, when the adversary detects those dummy signals, either through database or collective spectrum sensing. Thereafter, the adversary would obtain transformed temporal spectrum occupation of the PUs based on detected signals, which is a combination of the dummy and true signals. As long as the dummy signals are sent out frequently enough, the PUs' true operations can be hidden in those signals and the operational privacy of PUs can be preserved. Thus, the utility function of PUs' operational privacy preservation can be written as

$$U_{PU_j}(y_j) = c(T_j - y_j), \tag{1}$$

where c is a temporal privacy coefficient, T_j is the actual spectrum availability, and y_j is the transformed spectrum availability after the PU j's obfuscation strategy is executed. We can see that if $(T_j - y_j)$ is sufficiently large, PUs' temporal operational privacy is preserved effectively.

From the SUs' perspective, they attempt to transmit on available spectrum to satisfy their own demand. Since SUs can only observe the transformed spectrum availability of PUs, i.e., the spectrum availability after PUs execute obfuscation strategy, we denote the transformed spectrum availability for SU i over spectrum band j as $x_{i,j}$. Assuming the SU's traffic can be perfectly split, we let $\sum_{j=1}^{M} x_{i,j}$ denote the total available time that SU i can transmit over all spectrum bands. We define $d(\xi_i)$ as the actual time needed to satisfy the traffic demand of SU i. Then, $\min\Big(\sum_{j=1}^{M} x_{i,j}, d(\xi_i)\Big)$ represents the traffic delivery time of SU i. Specifically, when $d(\xi_i) < \sum_{j=1}^{M} x_{i,j}$, which indicates that the traffic demand is less than the transformed available spectrum supply, SU i will only transmit $d(\xi_i)$ to meet its service demands. On the other hand, if transformed available spectrum supply for SU i is less than its real demand, i.e., $\sum_{j=1}^{M} x_{i,j} < d(\xi_i)$, then SU$_i$ will deliver $\sum_{j=1}^{M} x_{i,j}$.

So, the utility function of SU i is

$$U_{SU_i}(\xi_i) = bE_{\mathbb{P}_i}\Big(\min\Big(\sum_{j=1}^{M} x_{i,j}, d(\xi_i)\Big)\Big), \tag{2}$$

where b is SUs' traffic delivery coefficient, and traffic demand ξ_i follows the distribution $\mathbb{P}_i$.

[2]Here, we assume SUs use energy detection for spectrum sensing.

B. PUs' Operational Privacy Preserving Optimization

Based on the utility functions of PUs and SUs, we expect an obfuscation strategy jointly considering PUs' operational privacy preservation and the satisfaction of SUs' uncertain traffic demands. Obviously, regardless of the PU's power consumption, generating more dummy signals can better protect the PU's operational privacy while reducing the available opportunistic accessing time of the SUs and thus diminishing SUs' traffic delivery. Taking the trade-off between PUs' privacy and SUs' utility into account, we can formulate the PUs' obfuscation strategy design into an optimization, a classic two-stage stochastic programming (SP) problem, described as follows:

$$\begin{aligned}
\text{Max} \quad & \sum_{j=1}^{M} c(T_j - y_j) \\
& + b\sum_{i=1}^{N} \mathbb{E}_{\mathbb{P}_i}\Big(\min\Big(\sum_{j=1}^{M} x_{i,j}, d(\xi_i)\Big)\Big), && (3)\\
\text{s.t.} \quad & T_j - y_j \geq \lambda, && (4)\\
& \sum_{i=1}^{N} x_{i,j} \leq y_j. && (5)
\end{aligned}$$

The function $min(\cdot,\cdot)$ in (3) returns the smaller value of $\sum_{j=1}^{M} x_{i,j}$ and $d(\xi_i)$. The constraint (5) indicates that total transmission time for SUs over PU j's spectrum should be less than the total transformed available spectrum supply of PU j. Besides, to preserve PU j's operational privacy, the time period of the sent dummy signals, i.e., $T_j - y_j$, is then required to be larger than a certain predefined privacy threshold λ, which is a constant, as shown in (4).

As we know, it is difficult to know the actual probability distribution of SUs' demands in practice. In this paper, we employ a data-driven approach, i.e., the risk-averse stochastic optimization approach (RA-SP) allowing distribution ambiguity [16], to characterize the uncertainty of SUs' demands. Instead of deriving a true distribution for the unknown parameter ξ, this optimization approach constructs a confident set D, which allows the distribution ambiguity to be within D under a certain confidence level (e.g., 99%). With RA-SP, considering the worst-case distribution, we can reformulate the problem as follows.

$$\begin{aligned}
\text{Max} \quad & \sum_{j=1}^{M} c(T_j - y_j) \\
& + \text{Min}_{\mathbb{P}_i \in D} \sum_{i=1}^{N} b\, \mathbb{E}_{\mathbb{P}_i} \min\Big(\sum_{j=1}^{M} x_{i,j}, d(\xi_i)\Big), && (6)\\
\text{s.t.} \quad & T_j - y_j \geq \lambda, \\
& \sum_{i=1}^{N} x_{i,j} \leq y_j.
\end{aligned}$$

We use a distance measurement proposed in [17], [18] to quantify the distance between two distributions. Specifically, a predefined distance measure $d(\mathbb{P}_i^0, \mathbb{P}_i)$ is constructed on confident set D, where $\mathbb{P}_i^0$ is the reference distribution estimated from historical data, and $\mathbb{P}_i$ is the ambiguous distribution of SU i. The distance d and confident set D can be defined as follows:

$$D = \{\mathbb{P}_i : d_\zeta(\mathbb{P}_i^0, \mathbb{P}_i) \leq \theta\}, \tag{7}$$

$$d_\zeta(\mathbb{P}_i^0, \mathbb{P}_i) = \sup_{h \in \mathcal{H}} \left| \int_\Omega h d\mathbb{P}_i^0 - \int_\Omega h d\mathbb{P}_i \right|, \tag{8}$$

where the distance under ζ-structure probability metric is denoted by $d_\zeta(\cdot, \cdot)$, the tolerance is denoted by θ, and $\mathcal{H}$ is a family of real-valued bounded measurable functions on Ω (the sample space on ξ). Tolerance θ is correlated to historical data size. It can be easily inferred that the more historical data that the PU can observe, the tighter D would be, and the closer the ambiguous distribution $\mathbb{P}_i$ would be to $\mathbb{P}_i^0$. More details of ζ-structure probability metric will be illustrated in the following section.

IV. Risk-Averse Stochastic Programming for Preserving Temporal Operational Privacy

This section is organized as follows. First, we describe how to determine the reference distribution $\mathbb{P}_i^0$ for SU i. Second, we represent how to determine tolerance θ on the amount of historical data under ζ-structure. Lastly but not least, we develop algorithms to solve the problem with respect to different probability distance metrics.

A. Reference Distribution

$$\mathbb{P}_i^0(x \leq X) = \frac{1}{Q} \sum_{q=1}^{Q} \delta_{\xi_q^0}(x) \tag{9}$$

First, we introduce how we determine the reference distribution $\mathbb{P}_0$. Suppose we use a set of historical data $\{\xi_1^0, \xi_2^0, \xi_3^0, \cdots, \xi_q^0\}$ to estimate the reference distribution $\mathbb{P}_0$. We utilize the empirical distribution of the historical data samples to construct $\mathbb{P}_0$. To be specific, the distribution of empirical data is defined as (9), the indicator variable $\delta_{\xi_q^0}(x)$ is equal to 1 when $\xi_q^0 \leq x$ and 0 otherwise. After that, the reference distribution data can be represented by its mass probability p_q^0 which equals to the ratio of the number of historical data samples matching ξ_q and Q since the supporting space is discrete.

B. Converge Rate under ζ-Probability Metrics

As described in Sec. IV, we employ three metrics and solve our problem under these constraints correspondingly. We define $\rho(x, y)$ as the distance between two variables x and y, and n is the dimension of Ω. $\mathbb{P} = \mathcal{L}(x)$ represents random variables x follows distribution $\mathbb{P}$. The metrics are derived as follows.

- **Kantorovich metric**: denoted as $d_K(\mathbb{P}_i^0, \mathbb{P}_i)$, $\mathcal{H} = \{h : ||h||_L \leq 1\}$, where $||h||_L := \sup\{h(x) - h(y)/\rho(x, y) : x \neq y \text{ in } \Omega\}$.
- **Fortet-Mourier metric**: denoted as $d_{FM}(\mathbb{P}_i^0, \mathbb{P}_i)$, $\mathcal{H} = \{h : ||h||_C \leq 1\}$, where $||h||_C := \sup\{h(x) - h(y)/c(x, y) : x \neq y \text{ in } \Omega\}$ and $c(x, y) = \rho(x, y) max\{1, \rho(x, a)^{p-1}, \rho^(y, a)^{p-1}\}$ for some $p \geq 1$ and $a \in \Omega$.
- **Uniform metric**: denoted as $d_U(\mathbb{P}_i^0, \mathbb{P}_i)$, $\mathcal{H} = \{I_{(-\infty, t]}, t \in R^n\}$.

The relationships among metrics are represented as follows [16], [18]:

$$d_{FM}(\mathbb{P}_i^0, \mathbb{P}_i) \leq \Lambda \cdot d_K(\mathbb{P}_i^0, \mathbb{P}_i), \tag{10}$$

where $\Lambda = \max\{1, \varnothing^{p-1}\}$, $\varnothing$ is the diameter of Ω.

From the definition of metrics and relationships between metrics under ζ-structure, we can derive the convergence property and convergence rate accordingly.

For the Uniform metric, the convergence rate is obtained by utilizing the Dvoretzky-Kiefer-Wolfowitz inequality [19]:

Proposition 1 For a single dimension case (i.e., $n = 1$),

$$\mathbb{P}(d_U(\mathbb{P}_i^0, \mathbb{P}_i) \leq \theta) \geq 1 - \exp\left(-\frac{\theta^2 Q}{2}\right). \tag{11}$$

In [16], the converge rate of Kantorovich metric is shown below:

Proposition 2 For a general dimension case (i.e., $n \geq 1$).

$$\mathbb{P}(d_K(\mathbb{P}_i^0, \mathbb{P}_i) \leq \theta) \geq 1 - \exp\left(-\frac{\theta^2 Q}{2\varnothing^2}\right). \tag{12}$$

From the relation between the Fortet-Mourier metric and Kantorovich metric (10), with Proposition 2, we can easily derive the convergence rate of other metrics.

Corollary 1 For a general dimension (i.e., $n \geq 1$), we have

$$\mathbb{P}(d_{FM}(\mathbb{P}_i^0, \mathbb{P}_i) \leq \theta) \geq 1 - \exp\left(-\frac{\theta^2 Q}{2\varnothing^2 \Lambda^2}\right). \tag{13}$$

With the convergence rate in (11)-(13), we can calculate the tolerance θ accordingly. For instance, in the Kantorovich metric, we assume the confidence level is η. Therefore, $\mathbb{P}(d_u(\mathbb{P}_i^0, \mathbb{P}_i \leq \theta)) \geq 1 - \exp(-\frac{\theta^2}{2\varnothing^2} Q) = \eta$ according to (11), and $\theta = \varnothing\sqrt{2log(1/(1-\eta)/Q)}$. After that, we explore how to solve the problem in (6). The sample space is $\Omega = \{\xi^1, \xi^2, \cdots, \xi^V\}$. Then the formulation can be simplified as :

$$\text{Max} \quad \sum_{j=1}^{M} c(T_j - y_j) + \text{Min}_{p_i^k} \sum_{i=1}^{N} \sum_{k=1}^{V} b p_i^k \min\left(\sum_{j=1}^{M} x_{i,j}, d(\xi_i)\right), \tag{14}$$

$$\text{s.t.} \quad T_j - y_j \geq \lambda, \sum_{i=1}^{N} x_{i,j} \leq y_j, \tag{15}$$

$$\sum_{k=1}^{V} p_i^k = 1, \forall i = 1, \cdots, N, \tag{16}$$

$$\max \sum_{k=1}^{V} h_k p_i^{k0} - \sum_{k=1}^{V} h_k p_i^k \leq \theta, \forall h_k : ||h||_\zeta \leq 1, \tag{17}$$

where $|h||_\zeta$ defined according to different metrics. In the Kantorovich metric and Bounded-Lipschits metric, $|h_x - h_y| \leq \rho(\zeta^x, \zeta^y)$. In the Fortet-Mourier, $|h_x - h_y| \leq \rho(\zeta^x, \zeta^y) \max\{1, \rho(\zeta^x, a)^{p-1}, \rho(\zeta^y, a)^{p-1}\}$. The constraints in (16)-(17) can be summarized as $\sum_k a_{kl} h_k \leq b_{kl}, l = 1, \cdots, L$. To reformulate the constraints, we consider the problem:

$$\text{Min}_{h_k} \quad \sum_{k=1}^{V} h_k p_i^{k0} - \sum_{k=1}^{V} h_k p_i^k, \tag{18}$$

$$\text{s.t.} \quad \sum_{k=1}^{V} a_{kl} h_k \leq b_{kl}, l = 1, \cdots, L. \tag{19}$$

Its dual problem is represented as:

$$\text{Min} \quad \sum_{l=1}^{L} b_l u_l, \tag{20}$$

$$\text{s.t.} \quad \sum_{l=1}^{l} a_{kl} u_l \geq p_i^{k0} - p_i^k, \forall k = 1, \cdots, V, \tag{21}$$

where u is the dual variable. Accordingly, the formulation can be reformulated as follows:

$$\text{Max} \quad \sum_{j=1}^{M} c(T_j - y_j) + \text{Min}_{p_i^k} \sum_{i=1}^{N} \sum_{k=1}^{V} b p_i^k \min\left(\sum_{j=1}^{M} x_{i,j}, d(\xi_i)\right), \tag{22}$$

$$\text{(SP-M)} \quad \text{s.t.} \quad T_j - y_j \geq \lambda, \tag{23}$$

$$\sum_{i=1}^{N} x_{i,j} \leq y_j \tag{24}$$

$$\sum_{k=1}^{V} p_i^k = 1, \sum_{l=1}^{L} b_l u_l \leq \theta, \tag{25}$$

$$\sum_{l=1}^{L} a_{il} u_l \geq p_i^{k0} - p_i^k, \forall i = 1, \cdots, N. \tag{26}$$

For the Uniform metric, we can have the reformulation from the Uniform metric definition:

$$\text{Max} \quad \sum_{j=1}^{M} c(T_j - y_j) + \text{Min}_{p_i^k} \sum_{i=1}^{N} \sum_{k=1}^{V} b p_i^k \min\left(\sum_{j=1}^{M} x_{i,j}, d(\xi_i)\right), \tag{27}$$

$$\text{(SP-U)} \quad \text{s.t.} \quad T_j - y_j \geq \lambda, \tag{28}$$

$$\sum_{i=1}^{N} x_{i,j} \leq y_j, \tag{29}$$

$$\sum_{k=1}^{V} p_i^k = 1, \forall i = 1, \cdots, N, \tag{30}$$

$$\left| \sum_{k=1}^{l} \left(p_i^{k0} - {p_i}^k\right) \right| \leq \theta, \forall l = 1, \cdots, L. \tag{31}$$

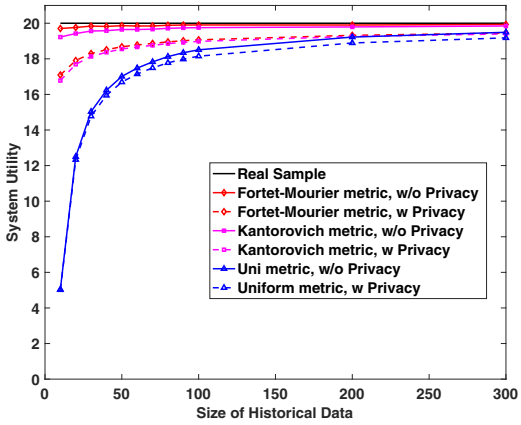

Fig. 2. Impact of historical data.

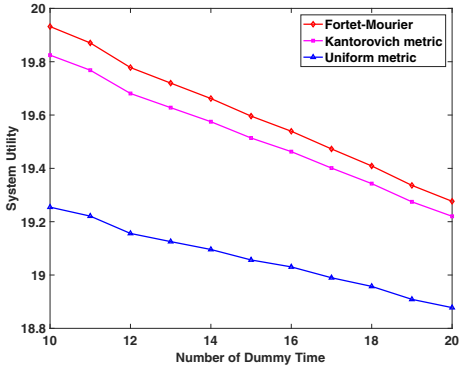

Fig. 3. Temporal operational privacy and system utility tradeoff.

The formulation SP-M and SP-U can be solved by CPLEX, etc. We also summarize the algorithm for the problem in Algorithm 1.

Algorithm 1 Algorithm for Obfuscation Strategy

1: **Input:** Historical data $\xi_i^1, \xi_i^2, \cdots, \xi_i^V$ from true distribution. Set η as the confident level of D.
2: **Out:** Objective value of the added time period of dummy signals.
3: Obtain the reference distribution $\mathbb{P}_0^i(x)$ and tolerance θ based on the historical data.
4: Use the reformulation (SP-M) or (SP-U) to solve the problem.
5: Output the solution.

V. Performance Evaluation

For illustrative purposes, in the simulations, we consider a CR network of 1 PU and 5 SUs. The actual available time of the PU's spectrum is $T = 30$ mins in a particular period $\mathcal{T} = 60$ mins. Moreover, we set the utility parameter for measuring operational privacy level c to be 3, and the utility parameter for SUs' traffic delivery b to be 5. We assume the SUs' traffic demand follows a discrete distribution with two scenarios: 10

mins and 20 mins with probabilities 0.4 and 0.6, respectively. We use this distribution to generate the historical data set for simulations.

First, we set the confidence level η to be 98% and the size of historical data from 100 to 300, to study the impact of the size of historical data. We also consider two strategies in the performance: with privacy obfuscating strategy ($\lambda = 15min$) and without obfuscating strategy ($\lambda = 0$). The results are reported in Fig. 2. From the figure, we can observe that the utility of network increases as the size of historical data increases, no matter under what kinds of metrics. The intuition behind the results is that the value θ decreases as the size of historical data increases. Therefore, the optimized problem (6) becomes less conservative. We can also see that when the number of samples exceeds 200, the gaps between results under the Fortet-Mourier metric and optimal value is very small. Furthermore, when sample size is 300, the gaps between system utility values are small under all metrics. Moreover, we study the performance under preserving privacy scheme. We set $\lambda = 15$ mins, which indicates that there is at least 15-minute gap between the transformed PUs' spectrum available time and the actual unoccupied period of the PU's spectrum. It can be observed that in Fig. 2, the total utility decreases after employing preservation privacy strategy since the PU's operational privacy preservation is at the cost of reducing accessing opportunities for SUs.

In addition, we explore the impacts of dummy signals time period on the system utility. The total number of historical sample is 300, and λ is set from 10 mins to 20 mins, and the results are shown in Fig. 3. We find that as the dummy signal time period increases, the overall system utility under all metrics decreases for chosen PU's privacy coefficient c, SUs' utility coefficient b, and confidence level. The reason is that the contributions of PU's privacy preservation is less important than the deductions of the denied SUs' traffic demands to current system. However, for a more PU's privacy oriented system (e.g., $c \gg b$), the system utility may increase while adding more dummy signals. For given PUs' and SUs' utility parameters, the proposed scheme can provide a design guideline for such a CR network considering the tradeoff between PUs' temporal operational privacy and SUs' performance.

VI. Conclusion

In this paper, we have proposed a novel obfuscation strategy for PUs within 3550-3700 MHz, which has a joint consideration of preserving PUs' temporal operational privacy and satisfying SUs' uncertain traffic demands. We have employed the data-driven risk-averse model in our scheme to characterize SUs' uncertain demand based on the historical data. With such a model, we have formulated the PUs' temporal operational privacy preservation problem into a risk-averse two-stage stochastic optimization, developed a robust algorithm for solutions, and presented simulation results to show the effectiveness of the proposed scheme in terms of preserving PUs' temporal operational privacy and satisfying SUs' traffic demands.

References

[1] FCC, "Spectrum policy task force report," Report of Federal Communications Commission, Et docket No. 02-135, November 2002.

[2] J. Liu, C. Zhang, H. Ding, H. Yue, and Y. Fang, "Policy-based privacy-preserving scheme for primary users in database-driven cognitive radio networks," in *Proceeding of the IEEE Global Communications Conference (GLOBECOM)*, Washington, DC, December 2016.

[3] M. Clark and K. Psounis, "Can the privacy of primary networks in shared spectrum be protected?" in *Proceeding of the IEEE International Conference on Computer Communications (INFOCOM)*, San Fransisco, CA, April 2016.

[4] A. Robertson, J. Molnar, and J. Boksiner, "Spectrum database poisoning for operational security in policy-based spectrum operations," in *IEEE Military Communications Conference*, San Diego, CA, November 2013.

[5] B. Bahrak, S. Bhattarai, A. Ullah, J.-M. Park, J. Reed, and D. Gurney, "Protecting the primary users operational privacy in spectrum sharing," in *IEEE International Symposium on Dynamic Spectrum Access Networks*, Mclean, VA, April 2014.

[6] C. Dwork, "Differential privacy: A survey of results," in *International Conference on Theory and Applications of Models of Computation*. Springer, 2008, pp. 1–19.

[7] Z. Gao, H. Zhu, S. Li, S. Du, and X. Li, "Security and privacy of collaborative spectrum sensing in cognitive radio networks," *IEEE Wireless Communications*, vol. 19, no. 6, pp. 106–112, December 2012.

[8] X. Fu, B. Graham, R. Bettati, and W. Zhao, "On effectiveness of link padding for statistical traffic analysis attacks," in *23rd International Conference on Distributed Computing Systems*, May 2003.

[9] T. Bonald, L. Massoulié, A. Proutiere, and J. Virtamo, "A queueing analysis of max-min fairness, proportional fairness and balanced fairness," *Queueing systems*, vol. 53, no. 1, pp. 65–84, 2006.

[10] T. Bonald, "Throughput performance in networks with linear capacity contraints," in *Information Sciences and Systems, 2006 40th Annual Conference on*, March 2006, pp. 644–649.

[11] T. Bonald and A. Proutiere, "Insensitive bandwidth sharing in data networks," *Queueing systems*, vol. 44, no. 1, pp. 69–100, 2003.

[12] X. Du, M. Rozenblit, and M. Shayman, "Implementation and performance analysis of snmp on a tls/tcp base," in *The Seventh IFIP/IEEE International Symposium on Integrated Network Management (IM 2001)*, May 2001.

[13] X. Yao, X. Han, X. Du, and X. Zhou, "A lightweight multicast authentication mechanism for small scale iot applications," *IEEE Sensors Journal*, vol. 13, no. 10, pp. 3693–3701, 2013.

[14] J. Lundén, S. A. Kassam, and V. Koivunen, "Robust nonparametric cyclic correlation-based spectrum sensing for cognitive radio," *IEEE Transactions on Signal Processing*, vol. 58, no. 1, pp. 38–52, January 2010.

[15] S. Gong, P. Wang, W. Liu, and W. Zhuang, "Performance bounds of energy detection with signal uncertainty in cognitive radio networks," in *INFOCOM, Proceedings IEEE*, 2013, pp. 2238–2246.

[16] C. Zhao and Y. Guan, "Data-driven risk-averse two-stage stochastic program with ζ-structure probability metrics," *Available on Optimization Online*, 2015.

[17] G. C. Calafiore, "Ambiguous risk measures and optimal robust portfolios," *SIAM Journal on Optimization*, vol. 18, no. 3, pp. 853–877, October 2007.

[18] D. Klabjan, D. Simchi-Levi, and M. Song, "Robust stochastic lot-sizing by means of histograms," *Production and Operations Management*, vol. 22, no. 3, pp. 691–710, February 2013.

[19] A. Dvoretzky, J. Kiefer, and J. Wolfowitz, "Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator," *The Annals of Mathematical Statistics*, pp. 642–669, 1956.