RF-based Network Inference: Theoretical Foundations

Shorya Consul, Hasan Burhan Beytur, Gustavo de Veciana, Haris Vikalo
Electrical and Computer Engineering Department, The University of Texas at Austin
shoryaconsul@utexas.edu, hbbeytur@utexas.edu, deveciana@utexas.edu, hvikalo@ece.utexas.edu

***Abstract*—We consider RF-based network inference based on channel usage. The proposed approaches rely on distributed spectrum sensing and are agnostic to the content and communication protocols. We consider inference based solely on observing nodes' channel usage and show it is equivalent to a Boolean matrix decomposition problem, which in general does not have a unique solution and is an NP-hard problem. We provide necessary and sufficient conditions for the Boolean matrix decomposition problem has a unique solution, i.e., for the network to be recoverable. We also propose a low-complexity network recovery algorithm that finds the unique solution under the recoverability conditions. In addition to that we provide an analysis of the required observation time to collect necessary channel usage data needed for the network recovery algorithm.**

***Index Terms*—collaborative spectrum sensing, wireless systems, network inference**

I. Introduction

With the burgeoning number of wireless applications and services, enabling spectrum sharing and monitoring its usage has became a growing area of interest [1], [2]. In this paper, we go beyond simply monitoring the spectrum usage and study inference of the network structure based on the observed RF activities of its nodes. This has natural applications in the policing of spectrum usage which aims to ensure compliance with the industry/public policies and standards. The inferred network structure could also pinpoint misbehaving or misconfigured nodes, thereby informing interventional procedures. Our focus is on ad-hoc networks consisting of spatially distributed nodes organized in possibly overlapping groups. The nodes in a group may hop in frequency over a set of channels; the groups' channel sets may overlap. We consider the setting where one is unaware of the network protocol and develop a protocol-agnostic methodology for learning the network structure. In particular, we develop and analyze methods for learning the network configuration from the RF activity data collected by low-cost sensors distributed in the monitored region.

A. Related Work

Depending on the system model and capabilities of the sensors, different approaches for network topology inference are proposed in literature. Some studies assume access to the packet's content, which is not always readily obtainable [3], [4]. In [5], a path inference approach for wireless sensor network is proposed; the method exploits the inter-packet correlations of packets generated from different nodes. Several inference techniques rely on the notions of causality and correlation. For instance, in [6], the authors infer network topology using spectral coherence as a measure of causality. Network topology inference can also be formulated as the problem of learning temporal causal structures among multiple time series. Along these lines, Granger causality is used in [7], [8] to model the acknowledgement mechanism of common communication protocols and learn the topological structure of time-multiplexed communication networks. The authors of [9] investigate the wireless network inference problem while considering the effect of imperfections such as packet collisions, shadowing and nodes' mobility. A method for reconstructing dynamic network topology from limited amounts of noisy data, under the assumption of sparse connectivity, is introduced in [10]; under these conditions, the superiority of the $L1$ optimizer is demonstrated.

In all of the above methods, network inference is performed based on data collected from multiple sources. In [11], an optimal activation policy for sensors collaboratively working to minimize the number of missed events is proposed. We take inspiration from [12], [13], where the authors proposed techniques to utilize inexpensive hardware such as USRP Radios for monitoring wideband spectrum.

B. Contributions

We study protocol and content-agnostic RF based inference of wireless networks with multiple groups, where users in each group share fixed channel sets. Moreover, the users in each group may perform frequency-hopping on their allocated, possibly overlapping, channel sets. We identify the conditions for which the network can be recovered with a low-complexity algorithm using only channel usage data. In addition to that we provide an analysis on the required observation time to recover the network. Our specific contributions are the following:

- We formulate network recovery from channel usage data as a Boolean matrix decomposition problem. This decomposition may not always admit a unique solution; even when a unique solution exits, the problem is NP-hard (Section III).
- To provide formal performance guarantees, we identify necessary and sufficient conditions for perfect network recovery; moreover, we develop an algorithm for network recovery under these, so called *distinguishability*, conditions (Theorem 1).
- We also identify more relaxed conditions for the partial recovery of nodes' group memberships or channel sets, which is possible even if channel or node sets are not distinguishable (Theorem 2).
- To provide insight on the required time to collect channel usage data, we provide a rough estimate on the required

time to collect channel activity data inside a group based on *uniform activity and observation model* (Section IV).

II. System Model and Notation

In this section we introduce the network model and associated notation. In particular:

- N denotes the set of all nodes participating in a network;
- G denotes the set of groups, where each group consists of nodes communicating with each other;
- N^g denotes the set of nodes participating in group $g \in G$, and $\mathcal{N} = \{N^g : g \in G\}$ denotes the set of all groups in the network;
- C denotes the set of channels over which frequency-hopping based communication takes place;
- C^g is the set of channels supporting group $g \in G$, i.e., the channels over which the nodes in g communicate; moreover, $\mathcal{C} = \{C^g : g \in G\}$ denotes the sets of possibly overlapping channels supporting various groups in the network;
- $C_n = \cup_{g:n\in N^g} C^g$ denotes the set of channels used by node n, i.e., the union over channels associated with the groups to which node n belongs;
- The tuple $(G, \mathcal{N}, \mathcal{C})$ denotes the network.

Note that $|B|$ denotes the cardinality of set B.

We assume that network nodes belong to one or more groups and communicate with each other using a fixed frequency or deploy frequency-hopping over a subset of channels. In the sequel, we assume that the sets of nodes and channels associated with groups are non-empty. In general, a node may participate in several groups and the sets of channels over which each group hops may overlap; such overlaps need to be avoided or be very infrequent among nearby nodes.

The proposed model is quite general and describes many types of networks. An example is the military adhoc system where frequency hopping spread spectrum (FHSS) schemes are favored due to their resilience to jamming and protection from unwanted detection. In such networks, multiple operational groups might be using FHSS over different (possibly overlapping) channels. Another system captured by our model is the Bluetooth network. To avoid collisions, Bluetooth uses a form of FHSS called adaptive frequency hopping. In this case, the groups can be thought of as pairs of master and slave Bluetooth devices which are communicating over fully overlapping (basically, identical) channels. A third example of a system described by our model is the WiFi network where an access point and the devices connected to it form a group, and groups communicate over non-overlapping WiFi subchannels. In general, any communication system using random access channels can be captured by our model.

III. Inferring Network Groups Based on Channel Usage Data

Based on the observed channel activities, we can in principle determine if a node is active on the channel set supporting a group and thus infer if the node is a member of that group. In this section, we will discuss approaches to network inference that rely only on the nodes' channel usage information.

To highlight some of the challenges and illustrate our approach to solving the network inference problem, let us consider the simple setting with two groups N^1 and N^2 communicating over channels sets C^1 and C^2, respectively. If $C^1 = C^2$, it is not possible to distinguish whether the network consists of one group of nodes $N^1 \cup N^2$ or two groups with nodes N^1 and N^2. Likewise, if $N^1 = N^2$, then it is not possible to distinguish between the setting where there are two groups with channel supports C^1 and C^2 and one group communicating over the set of channels $C^1 \cup C^2$. Clearly, identifying nodes in various groups using solely channel usage information will be possible only under certain conditions.

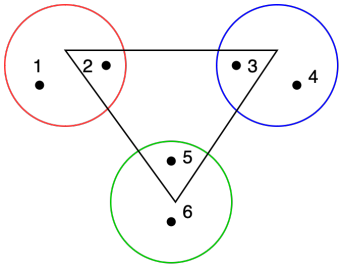

Fig. 1: Example of a network configuration

Next, we provide a few definitions which will be useful subsequently.

Definition 1. *A sensed data set $\mathcal{D}$ is said to be complete if it reveals the full set of subchannels used by each node, i.e., provides C_n for all $n \in N$.*

Definition 2. *For any group $g \in G$ we denote the set of channels used exclusively by the nodes in group g as $\bar{C}^g := C^g \setminus \cup_{f:f\neq g} C^f$, and the set of nodes which participate exclusively in group g as $\bar{N}^g := N^g \setminus \cup_{f:f\neq g} N^f$; these are referred to as the sets of **distinguishable** channels and nodes of group g, respectively. If $\bar{C}^g \neq \emptyset$ for all $g \in G$, the set of channels $\mathcal{C}$ is said to be distinguishable; likewise, if $\bar{N}^g \neq \emptyset$ for all $g \in G$, the set of nodes $\mathcal{N}$ is said to be distinguishable. If both channel and node sets are distinguishable, the network $(G, \mathcal{N}, \mathcal{C})$ is said to be **distinguishable**.*

In simple terms, distinguishability means that for each group g there exists a subset of channels $\bar{C}^g$ which are used only by the nodes in that group, and there is a subset of nodes $\bar{N}^g$ which are participating only in group g. For example, if the groups hop over disjoint sets of channels, the hopping sets are distinguishable.

A. Binary matrix representation of network $(G, \mathcal{N}, \mathcal{C})$

For convenience, we use a Boolean matrix $A \in \{0,1\}^{|C|\times|N|}$ to summarize information about channel usage by different network nodes. The (c, n) entry in A is defined as

$$A_{c,n} = \begin{cases} 1 & \text{if } c \in C_n, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

We refer to A as the *activity matrix*.

Similarly, we define the *channel assignment matrix* $S \in \{0,1\}^{|C|\times|G|}$ and the *node assignment matrix* $M \in \{0,1\}^{|G|\times|N|}$ with entries

$$S_{c,g} = \begin{cases} 1 & \text{if } c \in C^g \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad M_{g,n} = \begin{cases} 1 & \text{if } n \in N^g \\ 0 & \text{otherwise} \end{cases}, \tag{2}$$

respectively. Note that S and M together specify the network $(G,\mathcal{N},\mathcal{C})$. Further note that the activity matrix A can be expressed as

$$SM = A, \tag{3}$$

where the matrix multiplication is under the Boolean algebra (i.e., summation is the logical *or* operation, multiplication is the logical *and* operation). For any binary matrix K, we let $\text{rank}_B(K)$ denote the Boolean rank of K.

Remark: Boolean matrix representations have been used in a variety of settings including the Role Mining Problem [14], Set Basis Problem [15], Minimum Tiling Problem [16] and determining the minimum number of complete bipartite subgraphs [17]. Unfortunately, calculating the Boolean rank of a matrix, as well as finding the minimal product decomposition of a Boolean matrix, are NP-complete problems [15]–[17].

Note that a complete activity matrix A (i.e., with no missing entries) represents a complete data set. If A can be decomposed into S and M uniquely, then the network $(G,\mathcal{N},\mathcal{C})$ is **recoverable** based on the activity matrix.

Figure 1 shows a Venn diagram representing a network with 4 groups and 6 nodes. The color-coded sets represent groups, while dots represent the nodes in each group. The node assignment matrix M for this network is

$$M = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 \end{bmatrix},$$

where the color-coded rows indicate different groups and columns indicate nodes (ordered according to the node labels, 1–6). It is easy to see that M is a full Boolean rank matrix. However, the node set does not satisfy the distinguishability condition since the "black group" does not have a distinguishable node, i.e., one that belongs solely to that group. Continuing with this example, the associated channel assignment matrix S and the corresponding activity matrix A are

$$S = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \text{ and } A = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Note that although S is a full Boolean rank matrix, since the 2^{nd}, 3^{rd}, and 5^{th} columns of A are linear combinations of the 1^{st}, 4^{th} and 6^{th} columns of A, the Boolean rank of A is not 4. By the minimal product decomposition [14], A can be decomposed into two matrices, each of smaller dimension than S and M. Additionally, although Node 3 and Node 5 do not belong to the same group, the columns of A corresponding to Node 3 and Node 5 are equal – therefore, determining group memberships of these two nodes is not possible and the network is not uniquely recoverable from A. This example motivates study of conditions on node and channel assignments that should be satisfied for the network to be recoverable from A. Below, we state the necessary and sufficient conditions for network recovery from activity matrix.

Lemma 1. *A necessary condition for network inference from activity matrix A is that the Boolean rank of the activity matrix equals the number of groups, and channel and node assignment matrices are full Boolean rank, i.e., $\text{rank}_B(S) = \text{rank}_B(M) = \text{rank}_B(A) = |G|$.*

Proof of Lemma 1. If network $(G,\mathcal{N},\mathcal{C})$ is recoverable, the corresponding activity matrix A can be uniquely decomposed via matrices S and M. Recall that the dimensions of S and M are $|C|\times|G|$ and $|G|\times|N|$, respectively. Thus, A is of dimension $|C|\times|N|$.

Given the dimension of A, its Boolean rank satisfies

$$\text{rank}_B(A) \leq \min(|C|,|N|). \tag{4}$$

We shall consider two cases: (1) $|G| > \text{rank}_B(A)$, and (2) $|G| \leq \text{rank}_B(A)$.

First, suppose $|G| > \text{rank}_B(A) = K$. Then either $|C| < |G|$ or $|N| < |G|$, and one can find another pair of matrices S' and M' of dimension $|C|\times K$ and $K\times|N|$ such that $S'M' = A$. These can be obtained via the minimal product decomposition [14]. Note that the network represented by S' and M' is different than the one given by S and M, which contradicts the network recoverability assumption.

Now, suppose $|G| \leq \text{rank}_B(A)$. Due to (4), both $|G| \leq |C|$ and $|G| \leq |N|$ hold. Thus, $\text{rank}_B(S) \leq |G|$ and hence $\text{rank}_B(M) \leq |G|$. Consequently, the following Boolean rank inequality [18] holds,

$$\begin{aligned} \text{rank}_B(A) &= \text{rank}_B(SM) && (5)\\ &\leq \min\{\text{rank}_B(S), \text{rank}_B(M)\} && (6)\\ &\leq |G|. && (7) \end{aligned}$$

Note that, assuming $|G| \leq \text{rank}_B(A)$, (6) holds only if $|G| = \text{rank}_B(A)$. By the same inequality (6), $|G| = \text{rank}_B(S) = \text{rank}_B(M)$, which completes the proof. □

Theorem 1 provides a sufficient condition for recoverability.

Theorem 1. *Consider network $(G,\mathcal{N},\mathcal{C})$ with channel and node assignment matrices S and M, and activity matrix A. A sufficient condition for the network to be recoverable from A is that the channels and nodes satisfy the distinguishability condition.*

Proof of Theorem 1. Assume that $(G,\mathcal{N},\mathcal{C})$ is distinguishable, i.e., the channel and node sets satisfy the distinguishability condition. Then for each group $g \in G$ there is at least one node which does not participate in any other group, and at least one channel which is not used by any other group. Therefore, through proper reordering of groups, nodes and channels, the channel and node assignment matrices can be rewritten as

$$S' = \left[\begin{array}{c} I \\ \hline \widetilde{S'} \end{array}\right], \quad M' = \left[\begin{array}{c|c} I & \widetilde{M'} \end{array}\right]. \tag{8}$$

This special form with the identity submatrices of dimension $|G|\times|G|$ is possible because each group has at least one channel/node which is solely used by/belongs to that group. It follows that the activity matrix of $(G,\mathcal{N},\mathcal{C})$ can also be rewritten in the special form,

$$S'M' = A' = \left[\begin{array}{c|c} I & \widetilde{M'} \\ \hline \widetilde{S'} & \widetilde{S'}\widetilde{M'} \end{array}\right]. \tag{9}$$

Note that S' and M' are not known a priori – only A is known and it is not necessarily in the from given in (9). However, since the network $(G, \mathcal{N}, \mathcal{C})$ is distinguishable, one can identify distinguishable node-channel pairs in each group. Thus by properly reordering nodes and channels, the activity matrix can be rewritten as in (9), where the entries of the identity submatrix correspond to the distinguishable node-channel pairs of groups.

As indicated by (9) and (8), the node and channel assignment matrices are readily obtained from submatrices of appropriately reordered activity matrix. □

There are cases in which the node sets are recoverable while the channel sets are not, or vice versa. Theorem 2 below provides sufficient conditions for these cases where partial recoverability is possible.

Theorem 2. *Consider a network $(G, \mathcal{N}, \mathcal{C})$ with channel and node assignment matrices S and M, and activity matrix A. If the channel sets satisfy the distinguishability condition and the node assignment matrix M is of full Boolean rank with $\text{rank}_B(M) = |G|$, then the node sets are recoverable. Similarly, if the node sets satisfy the distinguishability condition and the channel assignment matrix S is of full Boolean rank, then sets of channels are recoverable.*

Proof of Theorem 2. If the channel sets are distinguishable, by the distinguishability condition each group $g \in G$ has at least one channel which is not used by any other group. Let e^g be a unit row vector with 1 for its g^{th} coordinate. The distinguishability of channel sets also can be interpreted as indicating that the corresponding channel assignment matrix S has at least one row vector e^g for each $g = 1 \dots |G|$.

Therefore, by re-ordering groups and channels, the channel assignment matrix can be put in the following form

$$S' = \left[\begin{array}{c} I \\ \hline \widetilde{S'} \end{array}\right] \tag{10}$$

This form is special due to the identity submatrix with dimension $|G| \times |G|$, corresponding to distinguishable channels of each group. Note that in this case node sets are not necessarily distinguishable but M' is a full Boolean rank matrix with $\text{rank}_B(M') = |G|$.

The product of S' and M' yields

$$S'M' = A' = \left[\begin{array}{c} M' \\ \hline \widetilde{S'M'} \end{array}\right]. \tag{11}$$

Since M' is full Boolean rank, the rows of M' are linearly independent. Thus, the upper submatrix of A' with $|G|$ rows are also linearly independent. The rows of lower submatrix of A' are the linear combinations of rows of M', since $A'_{k,\cdot} = S'_{k,\cdot}M'$ and $S'_{k,\cdot}$, $\forall k > |G|$ is either one of unit row vectors e^g or contained more than 1's in its entries.

Having said that, an activity matrix A is not necessarily in the from given in (11). To recover the node sets, one needs to identify $|G|$ linearly independent rows of A, where the rest of the rows are linear combinations of them. These linearly independent rows comprise M.

The proof for the case that the node sets satisfy the distinguishability condition and S is full Boolean rank with $\text{rank}_B(S) = |G|$ is similar. □

Under the distinguishability condition, we propose an efficient algorithm for finding unique factorization of the activity matrix A into the product of matrices S and M. Let $A_{\cdot,n}$ denote the column vector of A corresponding to channels used by node n, and likewise let $S_{\cdot,g}$ be the column vector of S corresponding to channels used by group g. In addition to this, let $x \succeq y$ denote the component-wise inequality between vectors x and y, i.e., $x_i \geq y_i$ for every index i. If node n is a member of group g, then $A_{\cdot,n} \succeq S_{\cdot,g}$. Note that in general node n can participate in more than one group; if n is a distinguishable node of group g, then $A_{\cdot,n} = S_{\cdot,g}$. Also note that the activity vectors of distinguishable nodes (columns of matrix A) are linearly independent and all other activity vectors can be formed as their linear combinations.

Algorithm 1 Activity Matrix Factorization

Input: Activity matrix A.
Initialize S, M matrices to 0; number of groups $\gamma = 0$
Preprocessing: Calculate $h_n = \|A_{\cdot,n}\|_1, \forall n = 1, \dots, |N|$
Relabel channels such that $n < n'$ if $h_n \leq h_{n'}$
for $n = 1, \cdots, |N|$ **do**
  for $g = 1, \cdots, \gamma$ **do**
    if $A_{\cdot,n} \succeq S_{\cdot,g}$ **then** $M_{g,n} = 1$
  if all prior comparisons fail **then**
    define a new group with $\gamma = \gamma + 1$
    $S_{\cdot,g} = A_{\cdot,n}$
    $M_{g,n} = 1$

Due to distinguishability condition on channels, each group has at least one channel that other groups do not use. Hence, the channel set of a distinguishable node cannot include the channel set of another distinguishable node, i.e., $A_{\cdot,n_1} \not\succeq A_{\cdot,n_2}$, where nodes n_1 and n_2 are distinguishable and belong to different groups. Therefore, we conclude that among the nodes of a group, the nodes that are active on a minimal number of channels are distinguishable. We use this observation in our iterative algorithm to decompose A corresponding to a distinguishable network.

Starting from the node using the smallest number of channels, Algorithm 1 processes columns of A sequentially. Since each group has at least one distinguishable node, the column of A with the smallest norm corresponds to one of the distinguishable nodes. The channel set of this node specifies the unique group that the node belongs to. The algorithm adds this column to matrix S and sets the corresponding entry of M to 1 in order to reflect the group membership of the node. In the next iterations, the channel sets of the remaining nodes are compared with the detected channel sets of distinguishable nodes, i.e., the columns of S. If the channel set vector of a node is componentwise greater than a column of S then it means that it is a member of a previously detected group. If not, then the node is a distinguishable node of a new group, and channel set of this node is added to S. Based on these

comparisons, 1's are added to the proper entries of the matrix M. When all the columns of A are processed, both S and M are obtained.

The complexity of the algorithm is driven by two steps: calculating the column norms (h_n), and the comparison of vectors in the *for* loop. Although h_n's are computed by a vector multiplication, they can actually be found by counting 1's in $A_{\cdot,n}$, which incurs complexity $\mathcal{O}(|C|)$). Each iteration of the outer *for* loop goes through $|C|$-dimensional vectors $|G|$ times. Thus, the complexity of one iteration is $\mathcal{O}(|C||G|)$, and the overall complexity of the loop is $\mathcal{O}(|C||G||N|)$. This dominates the time complexity of the algorithm.

In summary, although the binary decomposition problem is generally NP-complete, under the distinguishability condition it can be solved in $\mathcal{O}(|C||G||N|)$ time.

IV. Sampling The Activity Matrix

In this section, we briefly consider the time required to collect a complete data set under a given sensing and node activity model. This is tied to the well known coupon collector problem. Indeed, under random frequency hopping, one needs to observe the activity in all of each node's channels — collect all the coupons. We will study this under a uniform activity and observation model.

A. Uniform Activity and Observation Model

To capture activities of nodes and observations, we consider a discrete time system. In each time slot, either one of the nodes in group g is active at one of the channels in C^g or no node is active in group g. Meanwhile, the sensor system takes observations in a finer granularity, such that it observes one of the channels in C in each mini time slots. We denote the number of mini time slots in a time slot as γ, where $\gamma \in \mathbb{N}^+$. The duration of a mini time slot is set to be δ, and consequently, the duration of a time slot equals to $\gamma\delta$. The activity of node n at channel c in time slot t is detected, if the channel c is observed by the sensor system in any of the mini time slots of time slot t. The underlying assumption used here is that spectrum monitoring is an easier task than transmitting in a frequency hopping scheme. In other words, the rate of channel sweeping is higher than the rate of the frequency hopping. Note that a more capable sensor system would have a larger γ. For example, a perfect sensor system detecting every activity with probability 1, would have $\gamma = |C|$ and sweep all channels during a time slot to detect any activity.

The activity of nodes in a group g is denoted by an IID discrete time process $\Phi^g = (\Phi^g_t)_t$. Φ^g_t is either IDLE state or a two-tuple state (K^g_t, F^g_t), where K^g_t and F^g_t denote the active node and channel in time slot t, and $K^g_t \in N^g$ and $F^g_t \in C^g$.

Under the uniform activity model, nodes in a group are active with the same probability and the channel used for the activity is selected uniformly. Therefore, K^g_t and F^g_t are set to be independent i.i.d random variables. We assume that the group g is in IDLE state (no node is active) with probability π^g_{IDLE}. When the group is not IDLE, a node n becomes active at a channel c with equal probability, i.e., $\mathrm{P}(K^g_t = n \mid \Phi^g_t \neq \texttt{IDLE}) = \frac{1}{|N^g|}$ and $\mathrm{P}(F^g_t = c \mid \Phi^g_t \neq \texttt{IDLE}) = \frac{1}{|C^g|}$. Thus, for all $n \in N^g$ and $c \in C^g$, the probability density function of Φ^g_t is

$$\mathrm{P}(\Phi^g_t = \phi) = \begin{cases} \frac{1-\pi^g_{IDLE}}{|N^g||C^g|}, & \phi = (K^g_t = n, F^g_t = c), \\ \pi^g_{IDLE}, & \phi = \texttt{IDLE}. \end{cases} \quad (12)$$

We denote the observed channels by an IID discrete time process $\Psi = (\Psi_{t,\tau})_{t,\tau}$, and $\{\Psi_{t,\tau} = c\}$ is the event that channel c is observed in mini slot τ of time slot t, where $\Psi_{t,\tau} \in C$.

Under the uniform observation model, in each mini slot, the sensor system randomly chooses a channel to observe. Thus, we assume $\mathrm{P}(\Psi_{t,\tau} = c) = \frac{1}{|C|}$ for all $t > 0$ and $1 \leq \tau \leq \gamma$.

In Figure 2, we show sample paths for activity and observation processes, where $\gamma = 2$. As can be seen in the figure, the first activity is not detected since observed channels c_1 and c_5 are different than the active channel c_2. On the other hand, the activity in the third slot is detected since c_3 is observed in one of the mini time slots.

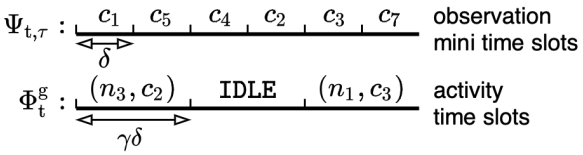

Fig. 2: Active channels and observed channels under uniform activity and observation models, γ = 2.

By the uniform activity and observation model, we have that $\mathrm{P}(\Psi_{t,\emptyset} = c) = 1/|C|, \forall t, \emptyset$ and $\mathrm{P}(\Phi^g_t = (K^g_t, F^g_t = c) \mid K^g_t = n) = 1/|C^g|, \forall t$. Consequently, the probability that an activity at time slot t is detected, denoted by q_g becomes

$$\begin{aligned} q_g &= \mathrm{P}(\{\Psi_{t,\tau} = c : \exists j \in [1,\gamma]\} \mid F^g_t = c) \\ &= (1 - (1 - \frac{1}{|C|})^\gamma) \approx \frac{\gamma}{|C|} \end{aligned}$$

Note that, in general, depending on the activity and observation model q_g can take values between 0 and 1.

For example, the aforementioned perfect sensing system with $\gamma = |C|$, that deterministically observes every channel in each time slot. In this case, since every activity is detected with probability one, and $q_g = 1$ for all $g \in G$.

Under the uniform activity and observation model to capture the complete dataset for group g, all node-channel pairs of the group should be detected. In group g, there are $|N^g||C^g|$ node-channel pairs and the probability that one of them gets detected is q_g. Given these and probability of being idle π^g_{IDLE}, we can bound the required time to capture a complete data of a group as follows:

Theorem 3. *To capture all possible node-channel activity pairs for group g under the uniform activity and observation model with a probability higher than $(1 - \epsilon)$, the total observation time required is*

$$\begin{aligned} \gamma\delta R(\epsilon) &\geq \frac{\gamma\delta|C^g||N^g|}{q_g(1-\pi^g_{IDLE})} \ln(\frac{|C^g||N^g|}{\epsilon}) \\ &\approx \delta\frac{|C^g||C||N^g|}{(1-\pi^g_{IDLE})} \ln(\frac{|C^g||N^g|}{\epsilon}), \end{aligned}$$

where $R(\epsilon)$ denotes the required number of time slots and $\gamma\delta$ is the duration of a time slot.

Proof of Theorem 3. Under uniform activity and observation model, the node-channel pairs are observed with the same probability $\frac{1-\pi^g_{IDLE}}{|C^g||N^g|}$. Thus, the probability that a certain node-channel pair is detected is $\frac{1-\pi^g_{IDLE}}{|C^g||N^g|}q_g$.

Let $Z^{g,r}_{n,c}$ denote the event that the node-channel pair $(n.c)$ in group g is not detected in the first r time slots. The probability of this event is given by

$$\begin{aligned}\mathrm{P}(Z^{g,r}_{n,c}) &= \left(1-\frac{1-\pi^g_{IDLE}}{|C^g||N^g|}q_g\right)^r \\ &\leq \exp\left(-\frac{rq_g(1-\pi^g_{IDLE})}{|C^g||N^g|}\right).\end{aligned}$$

Let T^g denote the number of time slots to detect all node-channel pairs in the group. We can write a tail estimate for T^g based on $Z^{g,r}_{n,c}$'s.

$$\begin{aligned}\mathrm{P}(T^g > r) &= \mathrm{P}\Big(\bigcup_{(n,c)} Z^{g,r}_{n,c}\Big) \\ &\leq |C^g||N^g|\exp\left(-\frac{rq_g(1-\pi^g_{IDLE})}{|C^g||N^g|}\right) \leq \epsilon\end{aligned}$$

By substituting r with $R(\epsilon)$ and after some mathematical manipulations on the last inequality, for large $|C^g||N^g|$, we obtain a bound on the required number of time slots to detect all node-channel pairs with probability higher than $(1-\epsilon)$, $R(\epsilon)$ as follows:

$$R(\epsilon) \geq \frac{\gamma\delta|C^g||N^g|}{q_g(1-\pi^g_{IDLE})}\ln(\frac{|C^g||N^g|}{\epsilon}) \tag{13}$$

By using the approximation for $q_g \approx \frac{\gamma}{|C||C^g|}$, and multiplying the number of time slots $R(\epsilon)$ with the duration of a slot $\gamma\delta$, we obtain the expression in the Theorem 3. □

By the Theorem 3, we see that required duration to collect complete data of a group is proportional to $|C^g||C| \geq |C^g|^2$. Other than that, the number of nodes in the group and the probability of being active affects the duration linearly. The accuracy parameter ϵ has an effect in order of $\ln(1/\epsilon)$, which is not much costly. Note that, under another activity and observation model, in general q_g can be as high as 1. However, in a model that the nodes active non-uniformly, the duration to detect all the node-channel pairs in a group would be longer, since detecting the less active nodes would take longer.

V. Conclusion

We have characterized and presented approaches for network inference of ad-hoc networks, with overlapping node groups. Our characterization includes necessary and sufficient conditions, called distinguishability conditions, for network recovery from channel usage data, thereby enabling the efficient inference of the network configuration. We also provide an analysis for the required observation time to collect channel usage data which is necessary to recover the network. As should be clear network inference based solely on channel usage is quite difficult, however one can also exploit the dynamics (activity correlation) among nodes sharing frequencies to construct a correlation matrix, which can further be incorporated in the inference process. We plan to explore this ıdea in our future work.

VI. Acknowledgement

This research is supported by the US Army Futures Command. The contents of this document represent the views of the authors and are not necessarily the official views of, or endorsed by, the US Government, Department of Defense, Department of the Army or US Army Futures Command.

The work of G. De Veciana and H. Vikalo was also partially supported NSF Award ECC-1809327.

References

[1] Y. Arjoune and N. Kaabouch, "A comprehensive survey on spectrum sensing in cognitive radio networks: Recent advances, new challenges, and future research directions," *Sensors*, 2019.

[2] M. Kulin, T. Kazaz, I. Moerman, and E. De Poorter, "End-to-end learning from spectrum data: A deep learning approach for wireless signal identification in spectrum monitoring applications," *IEEE Access*, 2018.

[3] J. Wenli, G. Teng, J. Meiyin, and L. Dan, "Researching topology inference based on end-to-end date in wireless sensor networks," in *ICICTA*, 2011.

[4] T. Kontos, G. S. Alyfantis, Y. Angelopoulos, and S. Hadjiefthymiades, "A topology inference algorithm for wireless sensor networks," in *IEEE Symposium on Computers and Communications (ISCC)*, 2012.

[5] Y. Gao, W. Dong, C. Chen, J. Bu, W. Wu, and X. Liu, "iPath: Path inference in wireless sensor networks," *IEEE/ACM Transactions on Networking*, 2016.

[6] C. Partridge, D. Cousins, A. W. Jackson, R. Krishnan, T. Saxena, and W. T. Strayer, "Using signal processing to analyze wireless data traffic," in *1st ACM Workshop on Wireless Security*, 2002, p. 67–76.

[7] M. Laghate and D. Cabric, "Learning wireless networks' topologies using asymmetric granger causality," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 233–247, 2018.

[8] P. Tilghman and D. Rosenbluth, "Inferring wireless communications links and network topology from externals using granger causality," in *IEEE Military Communications Conference*, 2013, pp. 1284–1289.

[9] E. Testi and A. Giorgetti, "Blind wireless network topology inference," *IEEE Transactions on Communications*, vol. 69, no. 2, 2021.

[10] D. Napoletani and T. D. Sauer, "Reconstructing the topology of sparsely connected dynamical networks," *Phys. Rev. E*, vol. 77, Feb 2008.

[11] M. Stecklein, H. B. Beytur, G. de Veciana, and H. Vikalo, "Optimizing resource constrained distributed collaborative sensing," in *IEEE International Conference on Communications Workshops*, 2021, pp. 1–6.

[12] Y. Guddeti, R. Subbaraman, M. Khazraee, A. Schulman, and D. Bharadia, "SweepSense: Sensing 5 GHz in 5 milliseconds with low-cost radios," in *Networked Systems Design and Implementation*, Feb. 2019.

[13] L. Shi, P. Bahl, and D. Katabi, "Beyond sensing: Multi-GHz realtime spectrum analytics," in *Symposium on Networked Systems Design and Implementation*, Oakland, CA, May 2015.

[14] H. Lu, J. Vaidya, and V. Atluri, "Optimal boolean matrix decomposition: Application to role engineering," in *IEEE 24th International Conference on Data Engineering*, 2008.

[15] J. Vaidya, V. Atluri, and Q. Guo, "The role mining problem: Finding a minimal descriptive set of roles," in *Proceedings of the 12th ACM Symposium on Access Control Models and Technologies*, 2007.

[16] F. Geerts, B. Goethals, and T. Mielikäinen, "Tiling databases," in *Discovery Science*, Berlin, Heidelberg, 2004.

[17] J. Orlin, "Contentment in graph theory: Covering graphs with cliques," *Indagationes Mathematicae (Proceedings)*, 1977.

[18] V. L. Watts, "Boolean rank of kronecker products," *Linear Algebra and its Applications*, 2001.