Secret Key Agreement and Secure Omniscience of Tree-PIN Source with Linear Wiretapper

Praneeth Kumar Vippathalla, Chung Chan, Navin Kashyap and Qiaoqiao Zhou

***Abstract*—In this paper, we obtain a single-letter characterization of the *wiretap secret key capacity* for a large class of multiterminal source models (namely, tree-PIN models) with a *linear wiretapper* that can observe arbitrary linear combinations of the source. For this class of sources, we also show a duality between the problems of wiretap secret key agreement and secure omniscience, which suggests that such duality potentially holds for more general sources.**

2021 IEEE International Symposium on Information Theory (ISIT) | 978-1-5386-8209-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISIT45174.2021.9518075

I. Introduction

The problem of multiterminal secret key agreement was studied by Csiszár and Narayan in [1]. They derived a single-letter expression for the secret key capacity C_{S} when the wiretapper has no side information. Remarkably, they established a duality between the problem of secret key agreement and the problem of communication for omniscience, which means that attaining omniscience by users is enough to extract a secret key of maximum rate. However, the characterization of secret key capacity when the wiretapper has side information C_{W} was left open, and they only gave some upper bounds for it. Later, Gohari and Anantharam, in [2], provided strengthened upper bounds and lower bounds. Furthermore, they proved a duality between secret key agreement with wiretapper side information and the problem of communication for omniscience by a neutral observer, where the neutral observer attains omniscience instead of the users. But this equivalence does not give an exact single-letter characterization of C_{W}. Nevertheless in some special cases, it is known exactly. In particular, [3] studied a pairwise independent network (PIN) source model defined on trees with wiretapper side information obtained by passing the edge random variables through independent channels. For this model, C_{W} was characterized using the conditional minimum rate of communication for omniscience characterization given in [1] together with an achieving scheme. The final form of C_{W} is similar to that of C_{S} except for the conditioning with respect to wiretap side information.

C. Chan (email: chung.chan@cityu.edu.hk) is with the Department of Computer Science, City University of Hong Kong. His work is supported by a grant from the University Grants Committee of the Hong Kong Special Administrative Region, China (Project No. 21203318).

Q. Zhou (email: zq115@ie.cuhk.edu.hk) is with the Institute of Network Coding and the Department of Information Engineering, The Chinese University of Hong Kong.

N. Kashyap (nkashyap@iisc.ac.in) and Praneeth Kumar V. (praneethv@iisc.ac.in) are with the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore 560012. Their work was supported in part by a Swarnajayanti Fellowship awarded to N. Kashyap by the Department of Science & Technology (DST), Government of India.

Recently, Chan et al. [4] studied the problem of secure omniscience in the context of multiterminal secure information exchange, and explored its connection to the problem of wiretap secret key agreement. In the secure omniscience problem, every user tries to attain omniscience by communicating interactively using their private observations from a correlated source; however, the goal here is to minimize the information leakage to a wiretapper who has side information about the source. Interestingly, in the case of a finite linear source (FLS) involving two active users and a wiretapper, they provided an explicit characterization of the wiretap secret key capacity and the minimum leakage rate for omniscience R_{L}. In fact, the achievable communication scheme for wiretap secret key capacity involves secure omniscience. Motivated by this result, they conjectured that such a duality holds for the entire class of FLSs. In this paper, we address this question and completely resolve it in the subclass of tree-PIN model but with a linear wiretapper, which is the most general wiretapper in the class of FLSs.

The PIN sources have received a wide attention in the secret key agreement problem without wiretapper side information, see [5–7]. The main motivation for studying PIN sources is that they model the problem of generating a global key out of locally generated keys by user pairs. In the study of general PIN sources, the subclass of tree-PIN sources play an important role. For the tree-PIN model [5], secret key capacity is achieved by using a linear and non-interactive communication scheme that propagates a key across the tree. This protocol indeed serves as a building block in the tree-packing protocol for the general PIN model. It was proved in [6] that the tree-packing protocol is even optimal for the secrecy capacity under any given total discussion rate. The optimality was shown by deriving a matching converse bound. Recently, [7] identified a large class of PIN models where the tree-packing protocol achieves the entire rate region where each point is a tuple of achievable key rate and individual discussion rates.

A problem that is closely related to secure omniscience is the coded cooperative data exchange (CCDE) problem with a secrecy constraint; see, for e.g., [8, 9]. The problem of CCDE considers a hypergraphical source and studies one-shot omniscience. The hypergraphical model generalizes the PIN model within the class of FLSs. [9] studied the secret key agreement in the CCDE context and characterized the number of transmissions required versus the number of SKs generated. On the other hand, [8] considered the same model but with wiretapper side information, and explored the leakage aspect

of an omniscience protocol. However, the security notion considered therein does not allow the eavesdropper to recover even one hyperedge of the source from the communication except what is already available. But the communication scheme can still reveal information about the source. In this paper we are interested to minimize the leakage of the total information to the wiretapper. Though we consider the asymptotic notion, the designed optimal communication scheme uses only a finite number of realizations of the source. Hence this scheme can find application even in CCDE problems.

The contribution of this paper is to show that secure omniscience achieves wiretap secret key capacity for the entire class of tree-PIN models with linear wiretapper. As a result of this duality between secure omniscience and wiretap secret key capacity problems, we could also characterize the single-letter expressions for both R_{L} and C_{W}, which was an open problem. The main novel ingredient is the construction of an optimal linear (non-interactive) omniscience communication scheme that completely spans the wiretapper side-information, thereby leaking as little information as possible.

II. Problem formulation

In this section, we describe two different scenarios in the context of multiterminal setting where the terminals communicate publicly using their correlated observations to perform a task securely from the eavesdropper, who has access to the public communication along with side information. More precisely, let $V = [m] := \{1, \ldots, m\}$ be the set of users and w denotes the wiretapper. Let $\mathsf{Z}_1, \ldots \mathsf{Z}_m$ and Z_{w} be the random variables taking values in finite alphabets $\mathcal{Z}_1, \ldots \mathcal{Z}_m$ and $\mathcal{Z}_{\mathrm{w}}$ respectively, and their joint distribution is given by $P_{\mathsf{Z}_1 \ldots \mathsf{Z}_m \mathsf{Z}_{\mathrm{w}}}$. Let $\mathsf{Z}_V := (\mathsf{Z}_i : i \in V)$ and Z_i^n denote the n i.i.d. realizations of Z_i. Each user has access to the corresponding random variable. Upon observing n i.i.d. realizations, the terminals communicate interactively using their observations and possibly independent private randomness on the noiseless and authenticated channel. In other words, the communication made by an user in any round depends on all the previous rounds communication and user's observations. Let $\mathsf{F}^{(n)}$ denotes this interactive communication. We say $\mathsf{F}^{(n)}$ is *non-interactive*, if it is of the form $(\tilde{\mathsf{F}}_i^{(n)} : i \in V)$, where $\tilde{\mathsf{F}}_i^{(n)}$ depends only on Z_i^n and the private randomness of user i. Note that the eavesdropper has access to the pair $(\mathsf{F}^{(n)}, \mathsf{Z}_{\mathrm{w}}^n)$. At the end of the communication, users output a value in a finite set using their observations and $\mathsf{F}^{(n)}$. For example, user i outputs $\mathsf{E}_i^{(n)}$ using $(\mathsf{F}^{(n)}, \mathsf{Z}_i^n)$ and its private randomness.

A. Secure Omniscience

In the secure omniscience scenario, each user tries to recover the observations of the other users except wiretapper's. We say that $(\mathsf{F}^{(n)}, \mathsf{E}_1^{(n)}, \ldots, \mathsf{E}_m^{(n)})_{n \geq 1}$ is an omniscience scheme if it satisfies the recoverability condition for omniscience

$$\liminf_{n \to \infty} \Pr(\mathsf{E}_1^{(n)} = \ldots = \mathsf{E}_m^{(n)} = \mathsf{Z}_V^n) = 1. \tag{1}$$

The minimum leakage rate for omniscience is defined as

$$R_{\mathrm{L}} := \inf \left\{ \limsup_{n \to \infty} \frac{1}{n} I(\mathsf{F}^{(n)} \wedge \mathsf{Z}_V^n | \mathsf{Z}_{\mathrm{w}}^n) \right\} \tag{2}$$

where the infimum is over all omniscience schemes. We sometimes use $R_{\mathrm{L}}(\mathsf{Z}_V || \mathsf{Z}_{\mathrm{w}})$ instead of R_{L} to make the source explicit. When there is no wiretapper side information, then the above notion coincides with the minimum rate of communication for omniscience, R_{CO} [1]. And the conditional minimum rate of communication for omniscience, $R_{\mathrm{CO}}(\mathsf{Z}_V | \mathsf{J})$, is used in the case when all the users have the shared randomness J^n along with their private observations. This means that user i observes $(\mathsf{J}^n, \mathsf{Z}_i^n)$.

B. Secret Key Agreement

In the secure secret key agreement, each user tries to recover a common randomness that is kept secure from the wiretapper. Specifically, we say that $(\mathsf{F}^{(n)}, \mathsf{E}_1^{(n)}, \ldots, \mathsf{E}_m^{(n)})_{n \geq 1}$ is a secret key agreement (SKA) scheme if there exists a sequence $(\mathsf{K}^{(n)})_{n \geq 1}$ such that

$$\liminf_{n \to \infty} \Pr(\mathsf{E}_1^{(n)} = \ldots = \mathsf{E}_m^{(n)} = \mathsf{K}^{(n)}) = 1, \tag{3a}$$

$$\limsup_{n \to \infty} \left[\log |\mathcal{K}^{(n)}| - H(\mathsf{K}^{(n)} | \mathsf{F}^{(n)}, \mathsf{Z}_{\mathrm{w}}^n) \right] = 0, \tag{3b}$$

where (3a) is the key recoverability condition and (3b) is the secrecy condition of the key and $|\mathcal{K}^{(n)}|$ denotes the cardinality of the range of $\mathsf{K}^{(n)}$. The wiretap secret key capacity is defined as

$$C_{\mathrm{W}} := \sup \left\{ \liminf_{n \to \infty} \frac{1}{n} \log |\mathcal{K}^{(n)}| \right\} \tag{4}$$

where the supremum is over all SKA schemes. The quantity C_{W} is also sometimes written as $C_{\mathrm{W}}(\mathsf{Z}_V || \mathsf{Z}_{\mathrm{w}})$. In (4), we use C_{S} instead of C_{W}, when the wiretap side information is set to a constant. Similarly, we use $C_{\mathrm{P}}(\mathsf{Z}_V | \mathsf{J})$ in the case when wiretap side information is $\mathsf{Z}_{\mathrm{w}} = \mathsf{J}$ and all the users have the shared random variable J along with their private observations Z_i. The quantities C_{S} and $C_{\mathrm{P}}(\mathsf{Z}_V | \mathsf{J})$ are referred to as secret key capacity of Z_V and private key capacity of Z_V with compromised-helper side information J respectively.

C. Tree PIN source with linear wirtapper

A source Z_V is said to be *Tree-PIN* if there exists a tree $T = (V, E, \xi)$ and for each edge $e \in E$, there is a non-negative integer n_e and a random vector $\mathsf{Y}_e = (\mathsf{X}_{e,1}, \ldots, \mathsf{X}_{e,n_e})$. We assume that the collection of random variables $\mathsf{X} := (\mathsf{X}_{e,k} : e \in E, k \in [n_e])$ are i.i.d. and each component is uniformly distributed over a finite field, say $\mathbb{F}_q$. For $i \in V$,

$$\mathsf{Z}_i = (\mathsf{Y}_e : i \in \xi(e)) .$$

The linear wiretapper's side information Z_{w} is defined as

$$\mathsf{Z}_{\mathrm{w}} = \mathsf{X} \boldsymbol{W},$$

where X is a $1 \times (\sum_{e \in E} n_e)$ vector and $\boldsymbol{W}$ is a $(\sum_{e \in E} n_e) \times n_w$ full column-rank matrix over $\mathbb{F}_q$. We sometimes refer to X as the base vector. We refer to the pair $(\mathsf{Z}_V, \mathsf{Z}_{\mathrm{w}})$ defined

as above as the *Tree-PIN source with linear wiretapper*. This is a special case of finite linear sources [10] where both Z_V and Z_w can be written as $\mathsf{X}\boldsymbol{M}$ and $\mathsf{X}\boldsymbol{W}$ respectively for some matrices $\boldsymbol{M}$ and $\boldsymbol{W}$. In the context of FLS, we say a communication scheme $\mathsf{F}^{(n)}$ is *linear*, if each user's communication is a linear function of its observations and the previous communication on the channel. Without loss of generality, linear communication can also be assumed to be non-interactive. In the rest of the paper, we consider only matrices over $\mathbb{F}_q$ unless otherwise specified.

D. Motivating example

The following example of a tree-PIN source with linear wiretapper appeared in our earlier work [4], where we constructed an optimal secure omniscience scheme. Let $V = \{1,2,3,4\}$ and

$$\mathsf{Z}_\mathrm{w} = \mathsf{X}_a + \mathsf{X}_b + \mathsf{X}_c, \tag{5}$$

$$\mathsf{Z}_1 = \mathsf{X}_a,\ \mathsf{Z}_2 = (\mathsf{X}_a, \mathsf{X}_b),\ \mathsf{Z}_3 = (\mathsf{X}_b, \mathsf{X}_c),\ \mathsf{Z}_4 = \mathsf{X}_c, \tag{6}$$

where X_a, X_b and X_c are uniformly random and independent bits. The tree here is a path of length 3 (Fig. 1) and the wiretapper observes the linear combination of all the edge random variables. For secure omniscience, terminals 2 and 3, using $n = 2$ i.i.d. realizations of the source, communicate linear combinations of their observations. The communication is of the form, $\mathsf{F}^{(2)} = (\tilde{\mathsf{F}}_2^{(2)}, \tilde{\mathsf{F}}_3^{(2)})$, where $\tilde{\mathsf{F}}_2^{(2)} = \mathsf{X}_a^2 + \boldsymbol{M}\mathsf{X}_b^2$ and $\tilde{\mathsf{F}}_3^{(2)} = (\boldsymbol{M}+\boldsymbol{I})\mathsf{X}_b^2 + \mathsf{X}_c^2$ with $\boldsymbol{M} := \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$. Since the matrices $\boldsymbol{M}$ and $\boldsymbol{M}+\boldsymbol{I}$ are invertible, all the terminals can recover Z_V^2 using this communication. For example, user 1 can first recover X_b^2 from $(\mathsf{X}_a^2, \tilde{\mathsf{F}}_2^{(2)})$ as $\mathsf{X}_b^2 = (\boldsymbol{M}+\boldsymbol{I})(\mathsf{X}_a^2 + \tilde{\mathsf{F}}_2^{(2)})$, then X_b^2 can be used along with $\tilde{\mathsf{F}}_3^{(2)}$ to recover X_c^2 as $\mathsf{X}_c^2 = (\boldsymbol{M}+\boldsymbol{I})\mathsf{X}_b^2 + \tilde{\mathsf{F}}_3^{(2)}$. More interestingly, this communication is aligned with the eavesdropper's observations, since $\mathsf{Z}_\mathrm{w}^2 = \tilde{\mathsf{F}}_2^{(2)} + \tilde{\mathsf{F}}_3^{(2)}$. This scheme achieves R_L, which is 1 bit.

For minimizing leakage, this kind of alignment must happen. For example, if Z_w^2 were not contained in the span of $\tilde{\mathsf{F}}_2^{(2)}$ and $\tilde{\mathsf{F}}_3^{(2)}$, then the wiretapper could infer a lot more from the communication. Ideally if one wants zero leakage, then $\mathsf{F}^{(n)}$ must be within the span of Z_w^n, which is not feasible in many cases because with that condition, the communication might not achieve omniscience in the first place. Therefore keeping this in mind, it is reasonable to assume that there can be components of $\mathsf{F}^{(n)}$ outside the span of Z_w^n. And we look for communication schemes which span as much of Z_w as possible. Such an alignment condition is used to control the leakage. In this particular example, it turned out that an omniscience communication that achieves R_CO can be made to completely align with the wiretapper side information. With the motivation from this example, we in fact showed that such an alignment phenomenon holds true in the entire class of tree-PIN with linear wiretapper.

III. Main results

The following two propositions give upper and lower bounds on minimum leakage rate for a general source $(\mathsf{Z}_V, \mathsf{Z}_\mathrm{w})$. The lower bound on R_L in terms of wiretap secret key capacity is obtained by using the idea of privacy amplification on the recovered source, while the multi-letter upper bound is given in terms of any communication made using first n i.i.d. realizations.

Proposition 1 ([4], Theorem 1) *For the secure omniscience scenario with $|V| \geq 2$,*

$$R_\mathrm{L} \geq H(\mathsf{Z}_V|\mathsf{Z}_\mathrm{w}) - C_\mathrm{W}. \tag{7}$$

Proposition 2 ([4], Theorem 2) *For the secure omniscience scenario,*

$$R_\mathrm{L} \leq \frac{1}{n}[R_\mathrm{CO}(\mathsf{Z}_V^n|\mathsf{F}^{(n)}) + I(\mathsf{Z}_V^n \wedge \mathsf{F}^{(n)}|\mathsf{Z}_\mathrm{w}^n)] \leq R_\mathrm{CO}, \tag{8}$$

where the inequality holds for any integer n and valid public discussion $\mathsf{F}^{(n)}$ for block length n. □

Before we present our result, we will discuss some notions related to Gács-Körner common information, which play an important role in proving the result. The Gács-Körner common information of X and Y with joint distribution $P_{\mathsf{X},\mathsf{Y}}$ is defined as

$$J_\mathrm{GK}(\mathsf{X},\mathsf{Y}) := \max\{H(\mathsf{G}) : H(\mathsf{G}|\mathsf{X}) = H(\mathsf{G}|\mathsf{Y}) = 0\} \tag{9}$$

A G that satisfies the constraint in (9) is called a common function (c.f.) of X and Y. An optimal G in (9) is called a *maximal common function* (m.c.f.) of X and Y, and is denoted by $\mathrm{mcf}(\mathsf{X},\mathsf{Y})$. Similarly, for n random variables, $\mathsf{X}_1, \mathsf{X}_2, \ldots, \mathsf{X}_n$, we can extend these definitions by replacing the condition in (9) with $H(\mathsf{G}|\mathsf{X}_1) = H(\mathsf{G}|\mathsf{X}_2) = \ldots = H(\mathsf{G}|\mathsf{X}_n) = 0$. For a finite linear source pair $(\mathsf{Z}_1, \mathsf{Z}_2)$, i.e., $\mathsf{Z}_1 = \mathsf{X}\boldsymbol{M}_1$ and $\mathsf{Z}_2 = \mathsf{X}\boldsymbol{M}_2$ for some matrices $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ where X is a $1 \times n$ row vector uniformly distributed on $\mathbb{F}_q^n$, it was shown in [11] that the $\mathrm{mcf}(\mathsf{Z}_1, \mathsf{Z}_2)$ is a linear function of X. This means that there exists a matrix $\boldsymbol{M}_g$ such that $\mathrm{mcf}(\mathsf{Z}_1, \mathsf{Z}_2) = \mathsf{X}\boldsymbol{M}_g$.

The main result of this paper is the following theorem.

Theorem 1 *For a Tree-PIN source Z_V with linear wiretapper observing Z_w,*

$$C_\mathrm{W} = \min_{e \in E} H(\mathsf{Y}_e|\,\mathrm{mcf}(\mathsf{Y}_e, \mathsf{Z}_\mathrm{w})),$$

$$R_\mathrm{L} = \left(\sum_{e \in E} n_e - n_w\right)\log_2 q - C_\mathrm{W} \text{ bits}.$$

In fact, a linear non-interactive scheme is sufficient to achieve both C_W and R_L simultaneously. □

The above theorem shows that the intrinsic upper bound on C_W holds with equality. In the multiterminal setting, the intrinsic bound that follows from [1, Theorem 4] is given by

$$C_\mathrm{W}(\mathsf{Z}_V||\mathsf{Z}_\mathrm{w}) \leq \min_{\mathsf{J}-\mathsf{Z}_\mathrm{w}-\mathsf{Z}_V} C_\mathrm{P}(\mathsf{Z}_V|\mathsf{J}).$$

This is analogous to the intrinsic bound for the two terminal case [12]. For the class of tree-PIN sources with linear wiretapper, when $\mathsf{J}^* = (\mathrm{mcf}(\mathsf{Y}_e, \mathsf{Z}_\mathrm{w}))_{e \in E}$, it can be shown that $C_\mathrm{P}(\mathsf{Z}_V|\mathsf{J}^*) = \min_{e \in E} H(\mathsf{Y}_e|\,\mathrm{mcf}(\mathsf{Y}_e, \mathsf{Z}_\mathrm{w}))$. This can be

derived using the characterization in [1] of the conditional minimum rate of communication for omniscience, $R_{\mathrm{CO}}(\mathsf{Z}_V|\mathsf{J}^*)$. In fact, the same derivation can also be found in [3] for a J that is obtained by passing the edge random variables through independent channels. In particular, J^* is a function of edge random variables $(\mathsf{Y}_e)_{e\in E}$ because $\mathrm{mcf}(\mathsf{Y}_e, \mathsf{Z}_{\mathrm{w}})$ is a function of Y_e. Therefore, we can see that $C_{\mathrm{P}}(\mathsf{Z}_V|\mathsf{J}^*)$, which is an upper bound on $\min_{\mathsf{J}-\mathsf{Z}_{\mathrm{w}}-\mathsf{Z}_V} C_{\mathrm{P}}(\mathsf{Z}_V|\mathsf{J})$, matches with the C_{W} obtained from Theorem 1.

Furthermore, the theorem guarantees that in the tree-PIN case with linear wiretapper, we can achieve the wiretap secret key capacity through a linear secure omniscience scheme, which establishes the duality between the two problems. This shows that omniscience can be useful even beyond the case when there is no wiretapper side information.

Our proof of Theorem 1 is through a reduction to the particular subclass of *irreducible* sources, which we defined next.

Definition 1 A Tree-PIN source with linear wiretapper is said to be *irreducible* iff $\mathrm{mcf}(\mathsf{Y}_e, \mathsf{Z}_{\mathrm{w}})$ is a constant function for every edge $e \in E$. □

Whenever there is an edge e such that $\mathsf{G}_e := \mathrm{mcf}(\mathsf{Y}_e, \mathsf{Z}_{\mathrm{w}})$ is a non-constant function, the user corresponding to a vertex incident on e can reveal G_e to the other users. This communication does not leak any additional information to the wiretapper, because G_e is a function of Z_{w}. Intuitively, for the further communication, G_e is not useful and hence can be removed from the source. After the reduction the m.c.f. corresponding to e becomes a constant function. In fact, we can carry out the reduction until the source becomes irreducible. This idea of reduction is illustrated through the following example.

Example 1 Let us consider a source Z_V defined on a path of length 3, which is shown in Fig. 1. Let $\mathsf{Y}_a = (\mathsf{X}_{a1}, \mathsf{X}_{a2})$, $\mathsf{Y}_b = \mathsf{X}_{b1}$ and $\mathsf{Y}_c = \mathsf{X}_{c1}$, where X_{a1}, X_{a2}, X_{b1} and X_{c1} are uniformly random and independent bits. If $\mathsf{Z}_{\mathrm{w}} = \mathsf{X}_{b1} + \mathsf{X}_{c1}$,

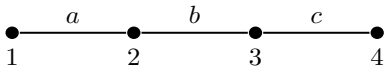

Fig. 1. A path of length 3

then the source is irreducible because $\mathrm{mcf}(\mathsf{Y}_e, \mathsf{Z}_{\mathrm{w}})$ is a constant function for all $e \in \{a, b, c\}$.

However if $\mathsf{Z}_{\mathrm{w}} = (\mathsf{X}_{a1} + \mathsf{X}_{a2}, \mathsf{X}_{b1} + \mathsf{X}_{c1})$, then the source is not irreducible, as $\mathrm{mcf}(\mathsf{Y}_a, \mathsf{Z}_{\mathrm{w}}) = \mathsf{X}_{a1} + \mathsf{X}_{a2}$, which is a non-constant function. An equivalent representation of the source is $\mathsf{Y}_a = (\mathsf{X}_{a1}, \mathsf{G}_a)$, $\mathsf{Y}_b = \mathsf{X}_{b1}$, $\mathsf{Y}_c = \mathsf{X}_{c1}$ and $\mathsf{Z}_{\mathrm{w}} = (\mathsf{G}_a, \mathsf{X}_{b1} + \mathsf{X}_{c1})$, where $\mathsf{G}_a = \mathsf{X}_{a1} + \mathsf{X}_{a2}$, which is also a uniform bit independent of $(\mathsf{X}_{a1}, \mathsf{X}_{b1}, \mathsf{X}_{c1})$. So, for omniscience, user 2 initially can reveal G_a without affecting the information leakage as it is completely aligned to Z_{w}. Since everyone has G_a, users can just communicate according to the omniscience scheme corresponding to the source without G_a. Note that this new source is irreducible. □

The next lemma shows that the kind of reduction to an irreducible source used in the above example is indeed optimal in terms of R_L and C_{W} for all tree-PIN sources with linear wiretapper.

Lemma 1 *If the Tree-PIN source with linear wiretapper $(\mathsf{Z}_V, \mathsf{Z}_{\mathrm{w}})$ is not irreducible then there exists an irreducible source $(\tilde{\mathsf{Z}}_V, \tilde{\mathsf{Z}}_{\mathrm{w}})$ such that*

$$C_{\mathrm{W}}(\mathsf{Z}_V||\mathsf{Z}_{\mathrm{w}}) = C_{\mathrm{W}}(\tilde{\mathsf{Z}}_V||\tilde{\mathsf{Z}}_{\mathrm{w}}),$$
$$R_{\mathrm{L}}(\mathsf{Z}_V||\mathsf{Z}_{\mathrm{w}}) = R_{\mathrm{L}}(\tilde{\mathsf{Z}}_V||\tilde{\mathsf{Z}}_{\mathrm{w}}),$$
$$H(\mathsf{Y}_e|\,\mathrm{mcf}(\mathsf{Y}_e, \mathsf{Z}_{\mathrm{w}})) = H(\tilde{\mathsf{Y}}_e),$$

for all $e \in E$. □

Note that, in the above lemma, the scheme that achieves $R_{\mathrm{L}}(\mathsf{Z}_V||\mathsf{Z}_{\mathrm{w}})$ involves revealing the reduced m.c.f. components first and then communicating according to the scheme that achieves $R_{\mathrm{L}}(\tilde{\mathsf{Z}}_V||\tilde{\mathsf{Z}}_{\mathrm{w}})$. As a consequence of Lemma 1, to prove Theorem 1, it suffices to consider only irreducible sources. For ease of reference, we re-state the theorem for irreducible sources below.

Theorem 2 *If Tree-PIN source with linear wiretapper is irreducible then*

$$C_{\mathrm{W}} = \min_{e\in E} H(\mathsf{Y}_e) = C_{\mathrm{S}},$$
$$R_{\mathrm{L}} = \left(\sum_{e\in E} n_e - n_w\right)\log_2 q - C_{\mathrm{S}} \text{ bits},$$

where C_{S} is the secret key capacity of Tree-PIN source without the wiretapper side information [1]. □

Theorem 2 shows that, for irreducible sources, even when the wiretapper has side information, the users can still extract a key at rate C_{S}. In terms of secret key generation, the users are not really at a disadvantage if the wiretapper has linear observations.

IV. Proofs

In this section we provide the essential proof ideas while the full proofs are available in the longer version [13, Sec. IV].

A. Proof sketch of Lemma 1

First we identify an edge e such that $\mathsf{G}_e := \mathrm{mcf}(\mathsf{Y}_e, \mathsf{Z}_{\mathrm{w}})$ is a non-constant function. Then, by appropriately transforming the random vector Y_e, we can separate out G_e from the random variables corresponding to the edge and the wiretapper. Later we argue that the source $(\mathsf{Z}_V, \mathsf{Z}_{\mathrm{w}})$ can be reduced into $(\tilde{\mathsf{Z}}_V, \tilde{\mathsf{Z}}_{\mathrm{w}})$ by removing G_e entirely without affecting C_{W} and R_{L}. And we repeat this process until the source becomes irreducible. At each stage, to show that $\mathrm{mcf}(\tilde{\mathsf{Y}}_b, \tilde{\mathsf{Z}}_{\mathrm{w}}) = \mathrm{mcf}(\mathsf{Y}_b, \mathsf{Z}_{\mathrm{w}})$, for $b \neq e$, and $\mathrm{mcf}(\tilde{\mathsf{Y}}_e, \tilde{\mathsf{Z}}_{\mathrm{w}})$ is a constant function, we use the following lemma which is proved in [13, Appendix].

Lemma 2 *If (X, Y) is independent of Z, then $\mathrm{mcf}(\mathsf{X}, (\mathsf{Y}, \mathsf{Z})) = \mathrm{mcf}(\mathsf{X}, \mathsf{Y})$ and $\mathrm{mcf}((\mathsf{X}, \mathsf{Z}), (\mathsf{Y}, \mathsf{Z})) = (\mathrm{mcf}(\mathsf{X}, \mathsf{Y}), \mathsf{Z})$.* □

B. Proof sketch of Theorem 2

Converse part. An upper bound on C_{W} is C_{S}. It was shown in [1, Example 5] that if the random variables of a source form a Markov chain on a tree, then $C_{\mathrm{S}} = \min_{(i,j):\{i,j\}=\xi(e)} I(\mathsf{Z}_i \wedge \mathsf{Z}_j)$. In the tree-PIN case, which satisfies the Markov property, this turns out to be $C_{\mathrm{S}} = \min_{e\in E} H(\mathsf{Y}_e)$. As consequence, we have $C_{\mathrm{W}} \leq \min_{e\in E} H(\mathsf{Y}_e)$, which implies that

$$R_{\mathrm{L}} \geq \Big(\sum_{e\in E} n_e - n_w\Big)\log_2 q - \min_{e\in E} H(\mathsf{Y}_e) \tag{10}$$

where the inequality follows from Proposition 1 and the full column-rank assumption on $\boldsymbol{W}$.

Achievability part. The achievable communication scheme for irreducible sources involves the following key components:

1) *Perfect omniscience [14]:* For a fixed $n \in \mathbb{N}$, $\mathsf{F}^{(n)}$ is said to achieve perfect omniscience if terminals can recover the source Z_V^n perfectly, i.e., $H(\mathsf{Z}_V^n|\mathsf{F}^{(n)}, \mathsf{Z}_i^n) = 0$ for all $i \in V$. If we do not allow any private randomness, then $H(\mathsf{F}^{(n)}|\mathsf{Z}_V^n) = 0$, which implies $\frac{1}{n} I(\mathsf{Z}_V^n \wedge \mathsf{F}^{(n)}|\mathsf{Z}_{\mathrm{w}}^n) = \frac{1}{n} H(\mathsf{F}^{(n)}|\mathsf{Z}_{\mathrm{w}}^n)$.
2) *Perfect alignment:* For an $n \in \mathbb{N}$, we say that $\mathsf{F}^{(n)}$ perfectly aligns with $\mathsf{Z}_{\mathrm{w}}^n$ if $H(\mathsf{Z}_{\mathrm{w}}^n|\mathsf{F}^{(n)}) = 0$. Note that $\mathsf{Z}_{\mathrm{w}}^n$ is only recoverable from $\mathsf{F}^{(n)}$ but not the other way around. In this case, $H(\mathsf{F}^{(n)}|\mathsf{Z}_{\mathrm{w}}^n) = H(\mathsf{F}^{(n)}) - H(\mathsf{Z}_{\mathrm{w}}^n)$. In an FLS, the wiretapper side information is $\mathsf{Z}_{\mathrm{w}}^n = \mathsf{X}^n \boldsymbol{W}^{(n)}$ where X is the base vector. Suppose the communication is of the form $\mathsf{F}^{(n)} = \mathsf{X}^n \boldsymbol{F}^{(n)}$, for some matrix $\boldsymbol{F}^{(n)}$. Then, the condition of perfect alignment is equivalent to the condition that the column space of $\boldsymbol{F}^{(n)}$ contains the column space of $\boldsymbol{W}^{(n)}$. This is in turn equivalent to the condition that the left nullspace of $\boldsymbol{W}^{(n)}$ contains the left nullspace of $\boldsymbol{F}^{(n)}$, i.e., if $\mathsf{y}\boldsymbol{F}^{(n)} = 0$ for some vector y then $\mathsf{y}\boldsymbol{W}^{(n)} = 0$.

As a consequence, the leakage rate of a perfect omniscience and alignment scheme (deterministic) is $\frac{1}{n} I(\mathsf{Z}_V^n \wedge \mathsf{F}^{(n)}|\mathsf{Z}_{\mathrm{w}}^n) = \frac{1}{n} H(\mathsf{F}^{(n)}|\mathsf{Z}_{\mathrm{w}}^n) = \frac{1}{n}[H(\mathsf{F}^{(n)}) - H(\mathsf{Z}_{\mathrm{w}}^n)] = \frac{1}{n} H(\mathsf{F}^{(n)}) - n_w \log_2 q$. To show the desired rate (10), it is enough to have $\frac{1}{n} H(\mathsf{F}^{(n)}) = \big(\sum_{e\in E} n_e\big)\log_2 q - \min_{e\in E} H(\mathsf{Y}_e)$. We show the existence of a scheme with that rate in the sub-case of PIN model defined on a path graph with $n_e = s$ for all $e \in E$. This can be extended to the tree-PIN case by using the fact that there exists a unique path from any vertex to a distinguished root node of the tree. The full proof of this case along with the most general model can be found in [13, Sec. IV].

Consider a PIN model defined on a path with length L and $n_e = s$ for all $e \in E$. Let $V = \{0, 1, \ldots, L\}$ be the set of vertices and $E = \{1, \ldots, L\}$ be the edge set such that edge i is incident on vertices $i-1$ and i. Since $n_e = s$ for all $e \in E$, $\min_{e\in E} H(\mathsf{Y}_e) = s \log_2 q$. Fix the number of i.i.d. realization of the source $n > \log_q(sL)$. The communication made by the terminals is as follows. Leaf nodes 0 and L do not communicate. The internal node i communicates $\tilde{\mathsf{F}}_i^{(n)} = \mathsf{Y}_i^n + \mathsf{Y}_{i+1}^n \boldsymbol{A}_i$, where $\mathsf{Y}_i^n = [\mathsf{X}_{i,1}^n \ldots \mathsf{X}_{i,s}^n] \in (\mathbb{F}_{q^n})^s$ and $\boldsymbol{A}_i$ is an $s \times s$ matrix with elements from $\mathbb{F}_{q^n}$. This communication is of the form $\mathsf{F}^{(n)} = \big[\tilde{\mathsf{F}}_1^{(n)} \cdots \tilde{\mathsf{F}}_{L-1}^{(n)}\big] = \big[\mathsf{Y}_1^n \cdots \mathsf{Y}_L^n\big]\boldsymbol{F}^{(n)}$ where $\boldsymbol{F}^{(n)}$ is

$$\begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{A}_1 & \boldsymbol{I} & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_2 & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{A}_{L-2} & \boldsymbol{I} \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{A}_{L-1} \end{bmatrix}.$$

Here $\boldsymbol{F}^{(n)}$ is an $sL \times s(L-1)$ matrix over $\mathbb{F}_{q^n}$ with $\mathrm{rank}_{\mathbb{F}_{q^n}}(\boldsymbol{F}^{(n)}) = s(L-1)$, which implies that $H(\mathsf{F}^{(n)}) = (sL - s)\log_2 q^n$ and the dimension of the left nullspace of $\boldsymbol{F}^{(n)}$ is s. Now the communication coefficients, $(\boldsymbol{A}_i : 1 \leq i \leq L-1)$, have to be chosen such that $\mathsf{F}^{(n)}$ achieves both perfect omniscience and perfect alignment.

It is not difficult to show that the perfect omniscience is equivalent to the condition that the $\boldsymbol{A}_i$'s are invertible. For perfect alignment, we require that the left nullspace of $\boldsymbol{F}^{(n)}$ is contained in the left nullspace of $\boldsymbol{W}^{(n)}$. Observe that

$$\underbrace{\big[\boldsymbol{S}_1 \quad -\boldsymbol{S}_1\boldsymbol{A}_1^{-1} \quad \cdots \quad (-1)^{L-1}\boldsymbol{S}_1\boldsymbol{A}_1^{-1}\ldots\boldsymbol{A}_{L-1}^{-1}\big]}_{:=\boldsymbol{S}} \boldsymbol{F}^{(n)} = \boldsymbol{0}.$$

where $\boldsymbol{S}_1$ is some invertible matrix and $\boldsymbol{S}_{i+1} := (-1)^i \boldsymbol{S}_1 \boldsymbol{A}_1^{-1} \ldots \boldsymbol{A}_i^{-1}$ for $1 \leq i \leq L-1$. Notice that the $\boldsymbol{S}_i$'s are invertible and $\boldsymbol{A}_i = -\boldsymbol{S}_{i+1}^{-1}\boldsymbol{S}_i$ for $1 \leq i \leq L-1$. The dimension of the left nullspace of $\boldsymbol{F}^{(n)}$ is s and all the s rows of $\boldsymbol{S}$ are independent, so these rows span the left nullspace of $\boldsymbol{F}^{(n)}$. Therefore for the inclusion, we must have $\boldsymbol{S}\boldsymbol{W}^{(n)} = \boldsymbol{0}$.

Thus, proving the existence of communication coefficients $\boldsymbol{A}_i$'s that achieve perfect omniscience and perfect alignment is equivalent to proving the existence of $\boldsymbol{S}_i$'s that are invertible and satisfy $[\boldsymbol{S}_1 \ldots \boldsymbol{S}_L]\boldsymbol{W}^{(n)} = \boldsymbol{0}$. The condition $n > \log_q(sL)$ guarantees the existence of $\boldsymbol{S}_i$'s, which follows from Schwartz-Zippel lemma. It should be noted that the assumption of irreducibility of the source is crucially used in the alignment condition and in the existence argument.

V. Conclusion and Future Direction

For a tree-PIN model with linear wiretapper, we have characterized R_{L} and C_{W}. It is worth noting that since the evaluation of the lower bound on C_{W} given in [2, Th. 7] is not explicit, the optimality of the bound is not very clear. But we showed an even stronger result for our particular model that a linear secure omniscience scheme achieves C_{W}, which establishes the duality between the two problems. So, our result gives evidence for the conjecture that, for FLSs, secure omniscience achieves C_{W}. However, proving this even for a general PIN model turned out to be quite challenging. Another interesting part of the proof is the communication scheme that achieves perfect omniscience and perfect alignment with the wiretapper. Though we used a random coding approach in the proof, there is an explicit optimal protocol construction in the case $n_e = 1$ for all $e \in E$ [13, Sec. V]. Such deterministic protocols, which have potential application in the CCDE context, are not known in the general tree-PIN case.

REFERENCES

[1] I. Csiszar and P. Narayan, "Secrecy capacities for multiple terminals," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3047–3061, 2004.

[2] A. A. Gohari and V. Anantharam, "Information-theoretic key agreement of multiple terminals—Part I," *IEEE Transactions on Information Theory*, vol. 56, no. 8, pp. 3973–3996, 2010.

[3] A. Poostindouz and R. Safavi-Naini, "Wiretap secret key capacity of tree-PIN," in *2019 IEEE International Symposium on Information Theory (ISIT)*, 2019, pp. 315–319.

[4] C. Chan, N. Kashyap, P. K. Vippathalla, and Q. Zhou, "Secure information exchange for omniscience," in *2020 IEEE International Symposium on Information Theory (ISIT)*, 2020, pp. 966–971.

[5] S. Nitinawarat, C. Ye, A. Barg, P. Narayan, and A. Reznik, "Secret key generation for a pairwise independent network model," *IEEE Transactions on Information Theory*, vol. 56, no. 12, pp. 6482–6489, 2010.

[6] C. Chan, M. Mukherjee, N. Kashyap, and Q. Zhou, "Upper bounds via lamination on the constrained secrecy capacity of hypergraphical sources," *IEEE Transactions on Information Theory*, vol. 65, no. 8, pp. 5080–5093, 2019.

[7] Q. Zhou, C. Chan, and R. W. Yeung, "On the discussion rate region for the PIN model," in *2020 IEEE International Symposium on Information Theory (ISIT)*, 2020, pp. 955–959.

[8] M. Yan and A. Sprintson, "Algorithms for weakly secure data exchange," in *2013 International Symposium on Network Coding (NetCod)*, 2013, pp. 1–6.

[9] T. A. Courtade and T. R. Halford, "Coded cooperative data exchange for a secret key," *IEEE Transactions on Information Theory*, vol. 62, no. 7, pp. 3785–3795, 2016.

[10] Chung Chan and Lizhong Zheng, "Mutual dependence for secret key agreement," in *2010 44th Annual Conference on Information Sciences and Systems (CISS)*, 2010, pp. 1–6.

[11] C. Chan, M. Mukherjee, N. Kashyap, and Q. Zhou, "Multiterminal secret key agreement at asymptotically zero discussion rate," in *2018 IEEE International Symposium on Information Theory (ISIT)*, 2018, pp. 2654–2658.

[12] U. M. Maurer and S. Wolf, "Unconditionally secure key agreement and the intrinsic conditional information," *IEEE Transactions on Information Theory*, vol. 45, no. 2, pp. 499–514, 1999.

[13] P. K. Vippathalla, C. Chan, N. Kashyap, and Q. Zhou, "Secret key agreement and secure omniscience of tree-PIN source with linear wiretapper," 2021. [Online]. Available: https://arxiv.org/abs/2102.01771

[14] S. Nitinawarat and P. Narayan, "Perfect omniscience, perfect secrecy, and Steiner tree packing," *IEEE Transactions on Information Theory*, vol. 56, no. 12, pp. 6490–6500, 2010.