Coded Cooperative Data Exchange for a Secret Key

Thomas A. Courtade, *Member, IEEE*, and Thomas R. Halford, *Member, IEEE*

***Abstract*— We consider a coded cooperative data exchange problem with the goal of generating a secret key. In particular, we investigate the number of public transmissions required for a set of clients to agree on a secret key with probability one, subject to the constraint that it remains private from an eavesdropper. Although the problems are closely related, we prove that secret key generation with the fewest number of linear transmissions is NP-hard, while it is known that the analogous problem in the traditional cooperative data exchange setting can be solved in polynomial time. In doing this, we completely characterize the best possible performance of linear coding schemes, and also prove that linear codes can be strictly suboptimal. Finally, we extend the single-key results to characterize the minimum number of public transmissions required to generate a desired integer number of statistically independent secret keys.**

***Index Terms*— Coded cooperative data exchange, secrecy, universal recovery, network coding.**

I. Introduction

In this paper, we consider a cooperative data exchange problem with the goal of generating a secret key. In particular, we study the number of public transmissions required for a set of clients to agree on a secret key, subject to the constraint that it remains private from an eavesdropper.

In an asymptotic setting, the reciprocal relationship between secret key (SK) capacity and communication for omniscience was revealed in the pioneering work [2] by Csiszár and Narayan. They showed that the maximum rate at which secrecy can be generated by a collection of terminals is in one-to-one correspondence with the minimum rate required for those same terminals to communicate for omniscience. Though they characterized the minimum communication rate required to attain omniscience, Csiszár and Narayan left characterizing the minimum communication rate required to generate a maximum-rate SK as an open problem [2, Sec. VI].

In [3], El Rouayheb *et al.* introduced a non-asymptotic, combinatorial version of Csiszár and Narayan's communication for omniscience problem, which they called *(coded) cooperative data exchange* (CCDE). In the CCDE problem, a set of clients initially hold a specified, finite set of messages coming from an ambient finite field $\mathbb{F}$, and the quantity of primary interest is the precise number of $\mathbb{F}$-valued public transmissions required for the clients to obtain omniscience (i.e., reproduce the messages collectively held by the clients). This contrasts with Csiszár and Narayan's formulation of communication for omniscience, wherein clients observe correlated memoryless sources and the quantity of interest is the *communication rate* (defined in the usual asymptotic sense) required for the clients to obtain omniscience.

Since its introduction, the CCDE problem has received significant attention from many researchers (see [4]–[14] and the references therein). Algorithms and heuristics for solving the CCDE problem were presented in [5]–[7] for broadcast networks and in [8] and [9] for multihop networks. Moreover, a number of authors have considered generalizations of the CCDE problem to model various practical system considerations [10]–[14].

Independent and contemporaneous with early work on CCDE, Chan and Zheng introduced an essentially equivalent source and network model, which they called the hypergraphical source and broadcast network, respectively [15]. While the CCDE works referenced above focus on communication for omniscience, [15] specifically addressed the problem of SK agreement (see also [16]). In follow up work, Chan [17] considered a closely related finite linear source model, and gave suboptimal bounds on the public transmission block length required for perfect SK agreement under linear transmission schemes. Despite their similarities, a distinctive feature of CCDE—which was not explicitly considered in [15]—is that public transmissions by clients are required to be $\mathbb{F}$-valued, and the figure of merit is the integer number of transmissions made by clients, rather than the public communication rate (defined under the assumption of a discrete memoryless hypergraphical source). This integrality constraint notwithstanding, [4], [8] showed that the omniscience-secrecy relations in [2] essentially carry over to the combinatorial CCDE setting. In particular, [4], [8] characterized the maximum number of independent, $\mathbb{F}$-valued SKs that clients can generate under the CCDE model. Notably, this quantity can be strictly smaller than the corresponding SK capacity if clients were to observe memoryless realizations according to the same hypergraphical source model. Despite the work in [4], [8], and [17], the communication requirements for agreeing on a desired number of $\mathbb{F}$-valued SKs under the CCDE model (equivalently, the hypergraphical source model of [15]) has remained open. This problem is the primary focus of the present paper, and we give a complete resolution for

Manuscript received August 1, 2014; revised June 1, 2015; accepted January 26, 2016. Date of publication March 8, 2016; date of current version June 14, 2016. This work was supported in part by the Division of Computing and Communication Foundations within the National Science Foundation Center for Science of Information under Grant CCF-0939370 and in part by the U.S. Defense Advanced Research Project Agency under Contract W15P7T-12-C-5013 and Contract W911QX-13-C-0010. This paper was presented in part at the 2014 IEEE International Symposium on Information Theory [1].

T. A. Courtade is with the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, CA 94720-1770 USA (e-mail: courtade@berkeley.edu).

T. R. Halford was with TrellisWare Technologies, Inc., San Diego, CA 92127-128 USA. He is now with WPL, Inc., Manhattan Beach, CA 90266-4447 USA (e-mail: tom.halford@wpli.net).

Communicated by M. Langberg, Associate Editor for Coding Theory.

Digital Object Identifier 10.1109/TIT.2016.2539347

the case of *linear* transmission schemes, essentially closing the gap between Chan's bounds in [17].

Related to the present work is the minimum communication rate required to generate a maximum-rate SK in the asymptotic setting (i.e., Csiszár and Narayan's open problem mentioned above). In [18], Tyagi gave a multi-letter expression characterization for this rate in the two-terminal case in terms of the *r-rounds interactive common information*. Recently, Mukherjee and Kashyap considered extensions to the multi-terminal case [19].

Despite the similarity in spirit, the asymptotic setting of [18] and [19] and the combinatorial setting of the present paper are considerably different in nature, and the proof techniques used are orthogonal. That said, all of these results shed light on the fundamentally different natures of SK generation at minimum communication rate and SK generation via communication for omniscience.

The weakly secure CCDE problem introduced in [20] is also related to our work. The goal of the weakly secure CCDE problem is to communicate for omniscience while revealing as little information as possible to an eavesdropper. This is closely related to the CCDE under a privacy constraint problem studied in [4]. Yan and Sprintson designed coding schemes that solve the weakly secure CCDE problem while revealing as little as information as possible to an eavesdropper [20]. Improvements to these schemes can be derived using the codes described in [21]. The primary distinction between the present setting and that of the weakly secure CCDE problem is that we only aim to generate a SK; we do not require that the nodes communicate for omniscience nor do we require that the SK corresponds to any given message.

Finally, we remark that in a recent paper [22], Halford *et al.* developed practical protocols for SK generation in ad hoc networks based on the CCDE problem. Briefly, a scenario was studied wherein the protocol designer controls the initial distribution of master keys so that secret keys can later be efficiently generated among arbitrary groups of clients. The results given in the present paper establish limits and suggest design rules for such protocols.

Summary of Contributions

We consider a cooperative data exchange problem with the goal of agreeing on a secret key of specified length. Specifically, for a fixed integer $\tau \geq 1$ and a hypergraphical source model where source symbols take values in a finite field $\mathbb{F}$, we investigate the number of $\mathbb{F}$-valued public transmissions required for a set of clients to agree on τ independent $\mathbb{F}$-valued random variables (i.e., secret keys), subject to the constraint that they remain private from an eavesdropper. For this model, our main contributions are the following:

- For linear encoding schemes, we characterize the number of public transmissions required to generate τ independent $\mathbb{F}$-valued SKs in terms of hypergraph connectivity. Moreover, we show that computing this quantity is NP-hard in general. This is surprising since the analogous CCDE problem (i.e., determining the minimum number of transmissions required for omniscience) can be solved in polynomial time. This dichotomy suggests that SK agreement with minimum-rate communication in general network models is a fundamentally more difficult problem than minimum-rate communication for omniscience.
- Further, we show that nonlinear encoding schemes can strictly outperform linear schemes (i.e., generate the same number of SKs with fewer public transmissions), and that this performance gap can be arbitrarily large. This is also surprising since for the same network model, linear schemes encoding schemes are known to be optimal in communication for omniscience (i.e., achieve fewest transmissions).

This paper is organized as follows. Section II formally defines our system model and reviews relevant results on CCDE. In Section III, we state and prove our main results for the generation of a single SK. Section IV characterizes the minimum number of public transmissions required to generate multiple SKs, and Section V delivers concluding remarks.

II. System Model and Preliminaries

We first establish basic notation. Throughout, we use calligraphic notation to denote sets. For two sets $\mathcal{A} \subset \mathcal{B}$, we write $\mathcal{B} \backslash \mathcal{A}$ to denote those elements in $\mathcal{B}$, but not in $\mathcal{A}$. If $\mathcal{A}$ is a singleton set (i.e., $\mathcal{A} = \{a\}$), then we often use the notation $\mathcal{B} - a \triangleq \mathcal{B} \backslash \{a\}$ for convenience. We define $\mathbb{Z}$ to be the set of integers. For positive $m \in \mathbb{Z}$, we use the shorthand notation $[m] \triangleq \{1, 2, \ldots, m\}$. Finally, for random variables X, Y, we write $I(X; Y)$ for the mutual information between X and Y.

A. System Model

Throughout, we consider networks defined by a set of n clients (i.e., terminals) $\mathcal{C} = \{c_1, c_2, \ldots, c_n\}$, a positive integer m, and a family of finite sets $\{\mathcal{I}_1, \mathcal{I}_2, \ldots, \mathcal{I}_n\}$ (each $\mathcal{I}_j \subseteq [m]$ and $\cup_{j=1}^{n} \mathcal{I}_j = [m]$) in the following way. Define the random (column) vector $\underline{X} \triangleq [X_1, X_2, \ldots, X_m]^T$, where each X_i is a discrete random variable with equiprobable distribution on a finite field $\mathbb{F}$, and $(X_1, X_2, \ldots, X_m)$ are mutually independent.[1] The random variables $\{X_i\}_{i=1}^{m}$ are called *messages*, and $\{X_i : i \in \mathcal{I}_j\}$ is the set of messages initially held by client $c_j \in \mathcal{C}$. In other words, $\mathcal{I}_j$ defines the indices of messages initially held by client c_j, for $j = 1, \ldots, n$. Throughout, n will always denote the number of clients; since the sets $\mathcal{I}_j$ are always indexed by $j \in [n]$, we will use the shorthand notation $\{\mathcal{I}_j\}$ to denote the family $\{\mathcal{I}_1, \mathcal{I}_2, \ldots, \mathcal{I}_n\}$.

We adopt the communication model which is standard in index coding and CCDE problems. That is, we consider transmission schemes consisting of a finite number of communication rounds. In each round, a single client broadcasts an element of $\mathbb{F}$ (which can be a function of the messages initially held by that client and all previous transmissions) to all other

[1] We assume throughout that $|\mathbb{F}| > n$. Indeed, the CCDE achievability results rely on the fact that linear network coding can achieve the cut-set bound in a multicast setting provided $|\mathbb{F}| > n$ (e.g., [23]). In practical settings, this will almost certainly be the case since the number of potential messages (e.g., packets) will typically far exceed the number of clients.

clients over an error-free channel. It is further assumed that all clients have knowledge of the index sets $\mathcal{I}_1, \ldots, \mathcal{I}_n$, and thus follow a protocol which is mutually agreed upon. We will elaborate on the definition of a transmission protocol in the next subsection.

B. Transmission Protocols

For a network defined by $\{\mathcal{I}_j\}$, a transmission protocol $\mathbf{P}$ (or simply, a protocol $\mathbf{P}$) consisting of t communication rounds is defined by n encoding functions $\{\mathsf{f}_1, \mathsf{f}_2, \ldots, \mathsf{f}_n\}$, and a t-tuple $(i_1, i_2, \ldots, i_t)$, where $i_k \in [n]$ indicates which client transmits during communication round k. More specifically, during communication round k, client c_{i_k} transmits

$$\mathsf{f}_{i_k}\left(\{X_j : j \in \mathcal{I}_{i_k}\}, k, \{\mathsf{f}_{i_\ell}\}_{\ell=1}^{k-1}\right) \in \mathbb{F}, \tag{1}$$

where we have abbreviated the transmitted symbols in rounds $\ell \in [k-1]$ by $\{\mathsf{f}_{i_\ell}\}_{\ell=1}^{k-1}$. For a given transmission protocol $\mathbf{P}$ requiring t communication rounds, we let $\mathbf{T}(\underline{X}, \mathbf{P}) \in \mathbb{F}^t$ be the column vector with k^{th} entry equal to $\mathsf{f}_{i_k}\left(\{X_j : j \in \mathcal{I}_{i_k}\}, k, \{\mathsf{f}_{i_\ell}\}_{\ell=1}^{k-1}\right)$. Letting $\|\cdot\|$ be the length function, we have $\|\mathbf{T}(\underline{X}, \mathbf{P})\| = t$. Note that $\mathbf{T}(\underline{X}, \mathbf{P})$ is a random variable since it is a function of the random vector $\underline{X}$. Generally, the transmission protocol under consideration will be clear from context. Hence, we abbreviate $\mathbf{T}(\underline{X}) \triangleq \mathbf{T}(\underline{X}, \mathbf{P})$ for convenience when there is no ambiguity.

A transmission protocol is said to be *linear* (over $\mathbb{F}$) if the encoding functions $\{\mathsf{f}_1, \mathsf{f}_2, \ldots, \mathsf{f}_n\}$ are of the form

$$\mathsf{f}_{i_k}\left(\{X_j : j \in \mathcal{I}_{i_k}\}, k, \{\mathsf{f}_{i_\ell}\}_{\ell=1}^{k-1}\right) = \sum_j \alpha_j^{(k)} X_j, \tag{2}$$

where $\alpha_j^{(k)} \in \mathbb{F}$ can be interpreted as the encoding coefficient for message j during communication round k. In this case, we can express $\mathbf{T}(\underline{X}) = A\underline{X}$, where $A \in \mathbb{F}^{t \times m}$ assuming the definitions $t \triangleq \|\mathbf{T}(\underline{X})\|$ and $m \triangleq |\cup_j \mathcal{I}_j|$. Hence, the *encoding matrix* A provides a succinct description of a linear transmission protocol. Note that the order of transmissions corresponding to a linear protocol is inconsequential.

C. Transmission Protocols for Omniscience

A transmission protocol $\mathbf{P}$ is said to achieve *omniscience* if there exist decoding functions $\{\mathsf{g}_1, \mathsf{g}_2, \ldots, \mathsf{g}_n\}$ which satisfy

$$\mathsf{g}_j\left(\{X_i : i \in \mathcal{I}_j\}, \mathbf{T}(\underline{X}, \mathbf{P})\right) = \underline{X} \text{ for each } j \in [n] \tag{3}$$

with probability 1.

Before proceeding, let $\mathsf{M}^\star\left(\{\mathcal{I}_j\}\right)$ denote the optimal value of the following integer linear program (ILP):

$$\begin{aligned} \textsf{minimize}: \ & \sum_{j \in [n]} a_j \\ \textsf{subject to}: \ & \sum_{j \in \mathcal{S}} a_j \geq \left| \bigcap_{j \in \bar{\mathcal{S}}} \bar{\mathcal{I}}_j \right| \text{ for all nonempty } \mathcal{S} \subset [n] \\ & a_j \in \mathbb{Z} \quad \text{for all } j \in [n], \end{aligned} \tag{4}$$

where $\bar{\mathcal{I}}_i \triangleq \left(\cup_j \mathcal{I}_j\right) \backslash \mathcal{I}_i$ and $\bar{\mathcal{S}} \triangleq [n] \backslash \mathcal{S}$. The quantity $\mathsf{M}^\star\left(\{\mathcal{I}_j\}\right)$ will play an important role in our treatment due to its inherent connection to the communication for omniscience, which is made explicit by the following theorem.[2]

Theorem 1 ([4, Th. 2]): If a protocol $\mathbf{P}$ *achieves omniscience, then* $\|\mathbf{T}(\underline{X}, \mathbf{P})\| \geq \mathsf{M}^\star\left(\{\mathcal{I}_j\}\right)$. *Conversely, there always exists a linear protocol* $\mathbf{P}_{\mathrm{L}}$ *that achieves omniscience and has* $\|\mathbf{T}(\underline{X}, \mathbf{P}_{\mathrm{L}})\| = \mathsf{M}^\star\left(\{\mathcal{I}_j\}\right)$.

Theorem 1 addresses the central issue in the CCDE problem, which primarily investigates the number of transmissions required to achieve omniscience. We remark that this is not equivalent to characterizing the minimum communication *rate* required for omniscience (as would be the case in Csiszár and Narayan's setting [2]) due in part to the integrality constraint on the number of transmissions.

D. Transmission Protocols for Secret Keys

A transmission protocol (with corresponding transmission sequence $\mathbf{T}(\underline{X})$) generates a secret key (SK) if there exist decoding functions $\{\mathsf{k}_1, \mathsf{k}_2, \ldots, \mathsf{k}_n\}$ which satisfy the following three properties:

(i) For all $j \in [n]$, and with probability 1,
$$\mathsf{k}_j\left(\{X_i : i \in \mathcal{I}_j\}, \mathbf{T}(\underline{X})\right) = \mathsf{k}_1\left(\{X_i : i \in \mathcal{I}_1\}, \mathbf{T}(\underline{X})\right).$$
(ii) $\mathsf{k}_1\left(\{X_i : i \in \mathcal{I}_1\}, \mathbf{T}(\underline{X})\right)$ is equiprobable on $\mathbb{F}$.
(iii) $I\left(\mathsf{k}_1\left(\{X_i : i \in \mathcal{I}_1\}, \mathbf{T}(\underline{X})\right); \mathbf{T}(\underline{X})\right) = 0$.

In words, requirement (iii) guarantees that the public transmissions $\mathbf{T}(\underline{X})$ reveal no information about $\mathsf{k}_1\left(\{X_i : i \in \mathcal{I}_1\}, \mathbf{T}(\underline{X})\right)$. Requirement (i) asserts that all clients $c_j \in \mathcal{C}$ can compute $\mathsf{k}_1\left(\{X_i : i \in \mathcal{I}_1\}, \mathbf{T}(\underline{X})\right)$. For these reasons, $\mathsf{k}_1\left(\{X_i : i \in \mathcal{I}_1\}, \mathbf{T}(\underline{X})\right)$ is called a *secret key*. Naturally, a secret key should be equiprobable on its domain to make guessing difficult, thus motivating requirement (ii).

It is not immediately clear whether any protocol $\mathbf{P}$ generates a SK. However, it turns out that such protocols exist in great abundance. In particular, the existence of protocols that generate a SK depends solely on the family $\{\mathcal{I}_j\}$.

Theorem 2 ([4, Th. 6]): For a network defined by $\{\mathcal{I}_j\}$, *there exists a protocol* $\mathbf{P}$ *which generates a SK if and only if*

$$\left|\cup_j \mathcal{I}_j\right| \geq \mathsf{M}^\star\left(\{\mathcal{I}_j\}\right) + 1. \tag{5}$$

It is important to point out that (5) closely parallels Csiszar and Narayan's work [2], which showed that a positive *rate* secret key can be generated if and only if the joint entropy of the encoder observations is strictly greater than the communication rate required for omniscience. In the combinatorial CCDE context of Theorem 2, $\left|\cup_j \mathcal{I}_j\right|$ plays the role of the joint source entropy and $\mathsf{M}^\star\left(\{\mathcal{I}_j\}\right)$ serves as a proxy for the communication rate required for omniscience. It is important to note that there is no concept of *rate* in the present setting, since the clients generate precisely one $\mathbb{F}$-valued SK.

Despite the fact that $\mathsf{M}^\star\left(\{\mathcal{I}_j\}\right)$ corresponds to the optimal value of an ILP, it can be computed in time polynomial in the number of messages $m = \left|\cup_j \mathcal{I}_j\right|$ (see [4], [11]). Therefore, for any family $\{\mathcal{I}_j\}$, we can efficiently test whether (5) holds. Hence, the essential remaining

[2]Theorem 1 essentially appeared in the given form in [8]. However, it was independently discovered by Milosavljevic et al. [11] and Chan [17] at roughly the same time.

question is: "*How many transmissions are needed to generate a SK?*"

To this end, let $\mathcal{P}(\{\mathcal{I}_j\})$ denote the set of protocols for $\{\mathcal{I}_j\}$ that generate a SK, and define

$$\mathsf{S}\left(\{\mathcal{I}_j\}\right) \triangleq \min\left\{\|\mathbf{T}(\underline{\mathrm{X}}, \mathbf{P})\| : \mathbf{P} \in \mathcal{P}(\{\mathcal{I}_j\})\right\}. \quad (6)$$

That is, $\mathsf{S}\left(\{\mathcal{I}_j\}\right)$ is the minimum number of transmissions needed to generate a SK. Similarly, let $\mathcal{P}_{\mathrm{L}}(\{\mathcal{I}_j\})$ denote the set of linear protocols for $\{\mathcal{I}_j\}$ that generate a SK, and define

$$\mathsf{S}_{\mathrm{L}}\left(\{\mathcal{I}_j\}\right) \triangleq \min\left\{\|\mathbf{T}(\underline{\mathrm{X}}, \mathbf{P})\| : \mathbf{P} \in \mathcal{P}_{\mathrm{L}}(\{\mathcal{I}_j\})\right\}. \quad (7)$$

In words, $\mathsf{S}_{\mathrm{L}}\left(\{\mathcal{I}_j\}\right)$ is the minimum number of transmissions required to generate a SK when we restrict our attention to linear protocols. If $\{\mathcal{I}_j\}$ does not satisfy (5), then we set $\mathsf{S}\left(\{\mathcal{I}_j\}\right) = \mathsf{S}_{\mathrm{L}}\left(\{\mathcal{I}_j\}\right) = \infty$.

Remark 1: We will often write "$\{\mathcal{I}_j\}$ generates a SK" instead of the more accurate, but cumbersome, "For the network defined by $\{\mathcal{I}_j\}$, there exists a protocol $\mathbf{P}$ *which generates a SK" whenever* (5) *holds.*

III. Generating a Single Secret Key

In this section, we investigate the number of transmissions required to generate a SK. In particular, we completely characterize $\mathsf{S}_{\mathrm{L}}\left(\{\mathcal{I}_j\}\right)$, and make progress toward characterizing $\mathsf{S}\left(\{\mathcal{I}_j\}\right)$. We will treat the more general case of generating multiple secret keys with minimum public communication in Section IV. Since the single-SK setting is arguably the most important in practice and the notation is less cumbersome than the general case, we find it beneficial to highlight the single-SK setting in the present section.

As demonstrated in the previous section, the CCDE and SK-generation problems are closely connected through the quantity $\mathsf{M}^\star\left(\{\mathcal{I}_j\}\right)$. Since Theorem 1 and the tractability of ILP (4) essentially resolve the CCDE problem, it is natural to conjecture that a similar result should hold for $\mathsf{S}\left(\{\mathcal{I}_j\}\right)$ and $\mathsf{S}_{\mathrm{L}}\left(\{\mathcal{I}_j\}\right)$. Unfortunately, there is a fundamental difference between the problems, which is revealed by the following two negative results:

Theorem 3: Computing $\mathsf{S}_{\mathrm{L}}\left(\{\mathcal{I}_j\}\right)$ is NP-hard.

Theorem 4: For any integer k, there exist families $\{\mathcal{I}_j\}$ for which $\mathsf{S}_{\mathrm{L}}\left(\{\mathcal{I}_j\}\right) > \mathsf{S}\left(\{\mathcal{I}_j\}\right) + k$.

For the CCDE problem, Theorem 1 asserts that linear protocols achieve optimal performance. Furthermore, the number of transmissions required by linear protocols is easily computed. For the problem of SK generation, the opposite is true. That is, linear protocols can be suboptimal, and the number of transmissions required by linear protocols is generally difficult to compute. This situation is parallel to that of multicast network coding and index coding. The two problems are closely related (cf. [24]), but exhibit the same dichotomy. See [25]–[27] and our remark at the end of this section for more details.

A. Proof of Theorem 3

Despite the negative results offered by Theorems 3 and 4, we can characterize several properties of $\mathsf{S}_{\mathrm{L}}\left(\{\mathcal{I}_j\}\right)$, $\mathsf{S}\left(\{\mathcal{I}_j\}\right)$, and $\mathsf{M}^\star\left(\{\mathcal{I}_j\}\right)$. Some of these properties are demonstrated in the following results, which are needed as we progress toward proving Theorem 3. A complete characterization of $\mathsf{S}_{\mathrm{L}}\left(\{\mathcal{I}_j\}\right)$ will be given in Theorem 5.

Lemma 1: If $\{\mathcal{I}_j\}$ generates a SK, then

$$\mathsf{S}\left(\{\mathcal{I}_j\}\right) \le \mathsf{S}_{\mathrm{L}}\left(\{\mathcal{I}_j\}\right) \le \mathsf{M}^\star\left(\{\mathcal{I}_j\}\right). \quad (8)$$

Proof: By definition, $\mathsf{S}\left(\{\mathcal{I}_j\}\right) \le \mathsf{S}_{\mathrm{L}}\left(\{\mathcal{I}_j\}\right)$ since $\mathcal{P}_{\mathrm{L}}(\{\mathcal{I}_j\}) \subseteq \mathcal{P}(\{\mathcal{I}_j\})$. The second inequality follows from the proof of [4, Th. 6], in which a linear transmission protocol $\mathbf{P}$ is constructed that generates a SK with $\|\mathbf{T}(\underline{\mathrm{X}}), \mathbf{P})\| = \mathsf{M}^\star\left(\{\mathcal{I}_j\}\right)$ communication rounds. □

We say that $\{\mathcal{J}_j\}$ is a *subfamily* of $\{\mathcal{I}_j\}$ if there is a set $\mathcal{S} \subset \cup_j \mathcal{I}_j$ such that $\mathcal{J}_j = \mathcal{I}_j \backslash \mathcal{S}$ for all $j \in [n]$.

Lemma 2: If $\{\mathcal{J}_j\}$ is a subfamily of $\{\mathcal{I}_j\}$, then

$$\mathsf{M}^\star\left(\{\mathcal{J}_j\}\right) \le \mathsf{M}^\star\left(\{\mathcal{I}_j\}\right), \quad (9)$$

$$\mathsf{S}\left(\{\mathcal{J}_j\}\right) \ge \mathsf{S}\left(\{\mathcal{I}_j\}\right), \textit{ and} \quad (10)$$

$$\mathsf{S}_{\mathrm{L}}\left(\{\mathcal{J}_j\}\right) \ge \mathsf{S}_{\mathrm{L}}\left(\{\mathcal{I}_j\}\right). \quad (11)$$

Proof: By De Morgan's law, it is easy to verify that

$$\left|\bigcap_{j\in\bar{\mathcal{S}}} \bar{\mathcal{I}}_j\right| \ge \left|\bigcap_{j\in\bar{\mathcal{S}}} \bar{\mathcal{J}}_j\right| \text{ for all nonempty } \mathcal{S} \subset [n], \quad (12)$$

where $\bar{\mathcal{I}}_i \triangleq \left(\cup_j \mathcal{I}_j\right) \backslash \mathcal{I}_i$ and $\bar{\mathcal{J}}_i \triangleq \left(\cup_j \mathcal{J}_j\right) \backslash \mathcal{J}_i$. Therefore, the constraints in ILP (4) are relaxed, and $\mathsf{M}^\star\left(\{\mathcal{J}_j\}\right) \le \mathsf{M}^\star\left(\{\mathcal{I}_j\}\right)$ by definition.

To show (10), observe that any transmission protocol which generates a SK for the subfamily $\{\mathcal{J}_j\}$ also generates a SK for the family $\{\mathcal{I}_j\}$ by ignoring the set of messages $\{X_i : i \notin \cup_j \mathcal{J}_j\}$. Hence, it follows that $\mathsf{S}\left(\{\mathcal{J}_j\}\right) \ge \mathsf{S}\left(\{\mathcal{I}_j\}\right)$. If $\{\mathcal{J}_j\}$ can not generate a SK, the inequality trivially holds. This argument also proves (11). □

Lemma 2 demonstrates monotonicity, but offers no insight into whether inequalities (9)-(11) are tight. The following lemma identifies settings under which (11) holds with equality, and will prove useful later on.

Lemma 3: If $\mathsf{S}_{\mathrm{L}}\left(\{\mathcal{I}_j\}\right) < \mathsf{M}^\star\left(\{\mathcal{I}_j\}\right)$, then there exists some $\ell \in \cup_j \mathcal{I}_j$ for which $\mathsf{S}_{\mathrm{L}}\left(\{\mathcal{I}_j - \ell\}\right) = \mathsf{S}_{\mathrm{L}}\left(\{\mathcal{I}_j\}\right)$.

Proof: By definition, there is a linear transmission protocol $\mathbf{P}_{\mathrm{L}}$ which generates a SK in $\mathsf{S}_{\mathrm{L}}\left(\{\mathcal{I}_j\}\right)$ communication rounds. Let $\mathbf{T}(\underline{\mathrm{X}}, \mathbf{P}_{\mathrm{L}}) = A\underline{\mathrm{X}}$ be the sequence of transmissions made by $\mathbf{P}_{\mathrm{L}}$, and let $\{\mathsf{k}_1, \dots, \mathsf{k}_n\}$ be valid decoding functions.

Since $\|\mathbf{T}(\underline{\mathrm{X}}, \mathbf{P}_{\mathrm{L}})\| = \mathsf{S}_{\mathrm{L}}\left(\{\mathcal{I}_j\}\right) < \mathsf{M}^\star\left(\{\mathcal{I}_j\}\right)$, Theorem 1 asserts that the protocol $\mathbf{P}_{\mathrm{L}}$ can not achieve omniscience. Therefore, by a possible permutation of clients, we can assume without loss of generality that there is no function g_1 for which

$$\mathsf{g}_1\left(\{X_i : i \in \mathcal{I}_1\}, A\underline{\mathrm{X}}\right) = \underline{\mathrm{X}} \text{ with probability 1.} \quad (13)$$

As a consequence, there must exist a nonzero vector $\underline{\mathrm{v}}$ such that $A\underline{\mathrm{v}} = 0$, and $v_i = 0$ for all $i \in \mathcal{I}_1$. Indeed, if there is no such $\underline{\mathrm{v}}$, then $A\underline{\mathrm{x}}_1 \ne A\underline{\mathrm{x}}_2$ for all $\underline{\mathrm{x}}_1 \ne \underline{\mathrm{x}}_2$ that agree on all coordinates in $\mathcal{I}_1$. Thus, client c_1, which knows all coordinates of $\underline{\mathrm{X}}$ in $\mathcal{I}_1$, can uniquely determine $\underline{\mathrm{X}}$ from $A\underline{\mathrm{X}}$, yielding a contradiction.

Since $\underline{\mathrm{v}}$ is not identically zero, there is some $\ell \notin \mathcal{I}_1$ for which $v_\ell \neq 0$. Considering any such ℓ, we define $\hat{X}_\ell \triangleq 0$, and $\hat{X}_i \triangleq X_i$ for $i \in \cup_j(\mathcal{I}_j - \ell)$. Also, define vectors $\underline{\hat{\mathrm{X}}} \triangleq [\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_m]^T$ and $\underline{\mathrm{X}}' \triangleq \underline{\hat{\mathrm{X}}} + X_\ell \cdot \underline{\mathrm{v}} = [X_1 + v_1 X_\ell, X_2 + v_2 X_\ell, \ldots, v_\ell X_\ell, \ldots, X_m + v_m X_\ell]^T$. Observe that, since the X_i's are independent and uniform on $\mathbb{F}$, the random vector $\underline{\mathrm{X}}'$ is also uniform over $\mathbb{F}^m$ and therefore equal in distribution to $\underline{\mathrm{X}}$. Next, note that

$$\mathsf{k}_j\left(\{\hat{X}_i : i \in \mathcal{I}_j\}, A\underline{\hat{\mathrm{X}}}\right) = \mathsf{k}_1\left(\{\hat{X}_i : i \in \mathcal{I}_1\}, A\underline{\hat{\mathrm{X}}}\right) \quad (14)$$

for all $j \in [n]$ since

$$\mathsf{k}_j\left(\{X_i : i \in \mathcal{I}_j\}, A\underline{\mathrm{X}}\right) = \mathsf{k}_1\left(\{X_i : i \in \mathcal{I}_1\}, A\underline{\mathrm{X}}\right) \quad (15)$$

with probability 1 by definition. Stated another way, (15) holds for all realizations $\underline{\mathrm{X}} = \underline{\mathrm{x}}$, and must therefore also hold for $\underline{\hat{\mathrm{X}}}$.

Next, observe that:

$$\begin{aligned} I&\left(\mathsf{k}_1\left(\{\hat{X}_i : i \in \mathcal{I}_1\}, A\underline{\hat{\mathrm{X}}}\right); A\underline{\hat{\mathrm{X}}}\right) \\ &= I\left(\mathsf{k}_1\left(\{\hat{X}_i + X_\ell \cdot v_i : i \in \mathcal{I}_1\}, A\underline{\mathrm{X}}'\right); A\underline{\mathrm{X}}'\right) && (16) \\ &= I\left(\mathsf{k}_1\left(\{X_i : i \in \mathcal{I}_1\}, A\underline{\mathrm{X}}\right); A\underline{\mathrm{X}}\right) = 0 && (17) \end{aligned}$$

In the above,

- (16) follows since $A\underline{\mathrm{X}}' = A(\underline{\hat{\mathrm{X}}} + X_\ell \cdot \underline{\mathrm{v}}) = A\underline{\hat{\mathrm{X}}}$, and $v_i = 0$ for all $i \in \mathcal{I}_1$.
- (17) follows from the above observation that $\underline{\mathrm{X}}'$ and $\underline{\mathrm{X}}$ are equal in distribution, and by definition of A and k_1.

Finally, by similar reasoning, we note that the random variable $\mathsf{k}_1\left(\{\hat{X}_i : i \in \mathcal{I}_1\}, A\underline{\hat{\mathrm{X}}}\right)$ is equiprobable on $\mathbb{F}$ since

$$\begin{aligned} \mathsf{k}_1&\left(\{\hat{X}_i : i \in \mathcal{I}_1\}, A\underline{\hat{\mathrm{X}}}\right) \\ &= \mathsf{k}_1\left(\{\hat{X}_i + X_\ell \cdot v_i : i \in \mathcal{I}_1\}, A\underline{\mathrm{X}}'\right) && (18) \\ &\stackrel{d}{=} \mathsf{k}_1\left(\{X_i : i \in \mathcal{I}_1\}, A\underline{\mathrm{X}}\right), && (19) \end{aligned}$$

and $\mathsf{k}_1\left(\{X_i : i \in \mathcal{I}_1\}, A\underline{\mathrm{X}}\right)$ is equiprobable on $\mathbb{F}$ by definition. In (19), the notation $\stackrel{d}{=}$ indicates equality in distribution.

Therefore, we can conclude that a SK can be generated by the subfamily $\{\mathcal{I}_j - \ell\}$ by applying the protocol $\mathbf{P}_{\mathrm{L}}$ and fixing $X_\ell \equiv 0$. This proves that $\mathsf{S}_{\mathrm{L}}\left(\{\mathcal{I}_j - \ell\}\right) \leq \mathsf{S}_{\mathrm{L}}\left(\{\mathcal{I}_j\}\right)$. By Lemma 2, the reverse inequality also holds. □

In order to proceed, we will need to introduce *critical* families. To this end, let $\tau \geq 1$ be an integer. A family $\{\mathcal{I}_j\}$ is *τ-critical* if the following hold:

(i) $\left|\cup_j \mathcal{I}_j\right| - \mathsf{M}^\star\left(\{\mathcal{I}_j\}\right) = \tau$, and
(ii) $\mathsf{M}^\star\left(\{\mathcal{I}_j - i\}\right) = \mathsf{M}^\star\left(\{\mathcal{I}_j\}\right)$ for all $i \in \cup_j \mathcal{I}_j$.

It is interesting to note that τ-criticality of $\{\mathcal{I}_j\}$ can be efficiently tested since $\mathsf{M}^\star\left(\{\mathcal{I}_j\}\right)$ is computable in polynomial time. Observe that 1-critical families enjoy a threshold property: families $\{\mathcal{I}_j\}$ that are 1-critical generate a secret key, and no proper subfamilies of $\{\mathcal{I}_j\}$ generate SKs. This is a consequence of Theorem 2 and the definition of 1-criticality.

Further, observe that if $\left|\cup_j \mathcal{I}_j\right| - \mathsf{M}^\star\left(\{\mathcal{I}_j\}\right) = \tau$, then $\{\mathcal{I}_j\}$ contains τ'-critical subfamilies for all $\tau' \leq \tau$. Indeed, $\mathsf{M}^\star\left(\{\mathcal{I}_j\}\right) \leq \mathsf{M}^\star\left(\{\mathcal{I}_j - i\}\right) + 1$ by definition, so if (ii) does not hold, then there exists $i \in \cup_j \mathcal{I}_j$ for which $\mathsf{M}^\star\left(\{\mathcal{I}_j\}\right) = \mathsf{M}^\star\left(\{\mathcal{I}_j - i\}\right) + 1$, and therefore

$$\begin{aligned} \left|\cup_j(\mathcal{I}_j - i)\right| - \mathsf{M}^\star\left(\{\mathcal{I}_j - i\}\right) &= \left|\cup_j \mathcal{I}_j\right| - \mathsf{M}^\star\left(\{\mathcal{I}_j\}\right) \\ &= \tau. && (20) \end{aligned}$$

This may be repeated so long as (ii) does not hold, and terminates upon finding a τ-critical subfamily of $\{\mathcal{I}_j\}$. Now, if $\{\mathcal{I}_j\}$ is τ-critical, then for any $i \in \cup_j \mathcal{I}_j$, we have

$$\begin{aligned} \left|\cup_j(\mathcal{I}_j - i)\right| - \mathsf{M}^\star\left(\{\mathcal{I}_j - i\}\right) &= \left|\cup_j \mathcal{I}_j\right| - 1 - \mathsf{M}^\star\left(\{\mathcal{I}_j\}\right) \\ &= \tau - 1. && (21) \end{aligned}$$

Hence, repeating the previous argument, we can conclude that $\{\mathcal{I}_j\}$ contains a $(\tau - 1)$-critical subfamily, so the claim follows by induction.

A *minimum τ-critical subfamily* $\{\mathcal{J}_j^\star\}$ of $\{\mathcal{I}_j\}$ satisfies

$$\left|\cup_j \mathcal{J}_j^\star\right| \leq \left|\cup_j \mathcal{J}_j\right| \quad (22)$$

for all other τ-critical subfamilies $\{\mathcal{J}_j\}$ of $\{\mathcal{I}_j\}$. By the above observation, if $\left|\cup_j \mathcal{I}_j\right| - \mathsf{M}^\star\left(\{\mathcal{I}_j\}\right) = \tau$, then $\{\mathcal{I}_j\}$ contains minimum τ'-critical subfamilies for all $\tau' \leq \tau$. Note that if $\{\mathcal{I}_j\}$ is τ-critical, then $\{\mathcal{I}_j\}$ is its own unique minimum τ-critical subfamily.

The following theorem demonstrates that minimum 1-critical subfamilies completely characterize $\mathsf{S}_{\mathrm{L}}\left(\{\mathcal{I}_j\}\right)$.

Theorem 5: If $\{\mathcal{I}_j\}$ generates a SK, then

$$\mathsf{S}_{\mathrm{L}}\left(\{\mathcal{I}_j\}\right) = \mathsf{M}^\star\left(\{\mathcal{J}_j^\star\}\right) = \left|\cup_j \mathcal{J}_j^\star\right| - 1, \quad (23)$$

where $\{\mathcal{J}_j^\star\}$ is a minimum 1-critical subfamily of $\{\mathcal{I}_j\}$.

Proof: By inductively applying Lemma 3, we can find a subfamily $\{\mathcal{T}_j\}$ of $\{\mathcal{I}_j\}$ for which

$$\mathsf{S}_{\mathrm{L}}\left(\{\mathcal{I}_j\}\right) = \mathsf{M}^\star\left(\{\mathcal{T}_j\}\right). \quad (24)$$

Let $\{\mathcal{J}_j\}$ be any 1-critical subfamily of $\{\mathcal{T}_j\}$. We have the following chain of inequalities

$$\begin{aligned} \mathsf{S}_{\mathrm{L}}\left(\{\mathcal{I}_j\}\right) \leq \mathsf{S}_{\mathrm{L}}\left(\{\mathcal{J}_j^\star\}\right) &\leq \mathsf{M}^\star\left(\{\mathcal{J}_j^\star\}\right) && (25) \\ &\leq \mathsf{M}^\star\left(\{\mathcal{J}_j\}\right) && (26) \\ &\leq \mathsf{M}^\star\left(\{\mathcal{T}_j\}\right) && (27) \\ &= \mathsf{S}_{\mathrm{L}}\left(\{\mathcal{I}_j\}\right). && (28) \end{aligned}$$

The above steps can be justified as follows:

- (25) follows from Lemmas 1 and 2.
- By definition of τ-criticality, (22) is equivalent to $\mathsf{M}^\star\left(\{\mathcal{J}_j^\star\}\right) \leq \mathsf{M}^\star\left(\{\mathcal{J}_j\}\right)$. Thus, (26) follows since $\{\mathcal{J}_j^\star\}$ is a minimum 1-critical subfamily of $\{\mathcal{I}_j\}$, and $\{\mathcal{J}_j\}$ is a 1-critical subfamily of $\{\mathcal{I}_j\}$.
- (27) follows from Lemma 2.
- (28) is the assertion of (24).

This proves that $\mathsf{S}_{\mathrm{L}}\left(\{\mathcal{I}_j\}\right) = \mathsf{M}^\star\left(\{\mathcal{J}_j^\star\}\right)$. Recalling the definition of 1-criticality completes the proof. □

The network defined by $\{\mathcal{I}_j\}$ has a natural representation as a hypergraph.[3] In particular, we make the following definition:

[3]We adopt the definition of a hypergraph that allows for repeated edges (i.e., multiple edges, with the same set of vertices, are permitted.

Definition 1: Consider a hypergraph $H = (\mathcal{V}, \mathcal{E})$ with vertex set $\mathcal{V} = \mathcal{C}$, and edge set $\mathcal{E} = \cup_j \mathcal{I}_j$. H is the hypergraph representation of $\{\mathcal{I}_j\}$ *if it has the following property: a vertex $c_j \in \mathcal{V}$ is contained in the edge $e \in \mathcal{E}$ if and only if $e \in \mathcal{I}_j$.*

Theorem 5 implies that $\mathsf{S}_{\mathrm{L}}\left(\{\mathcal{I}_j\}\right)$ is easily computed if we can identify a minimum 1-critical subfamily of $\{\mathcal{I}_j\}$. By Theorem 3, we know this must be NP-hard. In order to prove this to be the case, we require the following lemma which lends a hypergraph interpretation to 1-criticality. For a hypergraph $H = (\mathcal{V}, \mathcal{E})$, an edge set $\mathcal{E}' \subseteq \mathcal{E}$ is a *minimal connected dominating edge set* if the subhypergraph $H' = (\mathcal{V}, \mathcal{E}')$ is connected, and the removal of any edge from $\mathcal{E}'$ disconnects H'.

Lemma 4: Let $H = (\mathcal{V}, \mathcal{E})$ be the hypergraph representation of $\{\mathcal{I}_j\}$. H is connected if and only if

$$\mathsf{M}^\star\left(\{\mathcal{I}_j\}\right) < \left|\cup_j \mathcal{I}_j\right|. \tag{29}$$

In particular, $\{\mathcal{I}_j\}$ is 1-critical if and only if $\mathcal{E}$ is a minimal connected dominating edge set.

Proof: First, suppose H is not connected. By definition, there must exist a nontrivial partition $\mathcal{V} = (\mathcal{S}, \bar{\mathcal{S}})$ such that there is no edge $e \in \mathcal{E}$ which contains vertices from both $\mathcal{S}$ and $\bar{\mathcal{S}}$. Stated another way, $(\cup_{j\in\mathcal{S}}\mathcal{I}_j) \cap (\cup_{j\in\bar{\mathcal{S}}}\mathcal{I}_j) = \emptyset$. Hence, ILP (4) includes the two constraints

$$\sum_{j\in\mathcal{S}} a_j \geq \left|\bigcap_{j\in\bar{\mathcal{S}}} \bar{\mathcal{I}}_j\right| = \left|\bigcup_{j\in\mathcal{S}} \mathcal{I}_j\right| \tag{30}$$

$$\sum_{j\in\bar{\mathcal{S}}} a_j \geq \left|\bigcap_{j\in\mathcal{S}} \bar{\mathcal{I}}_j\right| = \left|\bigcup_{j\in\bar{\mathcal{S}}} \mathcal{I}_j\right|, \tag{31}$$

the sum of which imply $\mathsf{M}^\star\left(\{\mathcal{I}_j\}\right) \geq |\cup_j \mathcal{I}_j|$. By taking the contrapositive, we have proven

$$\mathsf{M}^\star\left(\{\mathcal{I}_j\}\right) < \left|\cup_j \mathcal{I}_j\right| \implies H \text{ is connected.} \tag{32}$$

Next, suppose H is connected, and assume without loss of generality that $\mathcal{E} = \cup_j \mathcal{I}_j \triangleq \{1, 2, \ldots, m\}$. Since H is connected, there is a transmission protocol for which the entries of $\mathbf{T}(\underline{X})$ are precisely $\{X_1 + X_j\}_{j=2}^m$. Indeed, by connectivity of H, there must be some client c initially holding X_1 and some X_e (say, X_2 without loss of generality), and can therefore transmit $X_1 + X_2$ during the first communication round. By induction, assume that $\{X_1+X_j\}_{j=2}^{m-1}$ are transmitted during the first $m-2$ communication rounds (permuting indices of the X_i's if necessary). Again, by connectivity of H, there must be a client c' which initially holds X_m and X_k, where $k < m$. Hence, in communication round $m-1$, client c' can transmit $(X_1 + X_k) - (X_k - X_m) = X_1 + X_m$. Noting that

$$(X_1, X_1 + X_2, \ldots, X_1 + X_m) \stackrel{d}{=} (X_1, X_2, \ldots, X_m),$$

we have $I(X_1; \mathbf{T}(\underline{X})) = 0$. If client $c \in e \in \mathcal{E}$, then it can recover X_1 from the transmission $X_1 + X_e$ by simply subtracting X_e. Since H is connected, each $c \in \mathcal{V}$ belongs to some edge in $\mathcal{E}$, and therefore all clients can recover X_1 losslessly. Since X_1 is equiprobable on $\mathbb{F}$ by definition, we can conclude that $\{\mathcal{I}_j\}$ generates a SK. Theorem 2 asserts that we must have $\mathsf{M}^\star\left(\{\mathcal{I}_j\}\right) < \left|\cup_j \mathcal{I}_j\right|$, and we have proven

$$\mathsf{M}^\star\left(\{\mathcal{I}_j\}\right) < \left|\cup_j \mathcal{I}_j\right| \iff H \text{ is connected.} \tag{33}$$

We now prove the second claim. To this end, suppose $\{\mathcal{I}_j\}$ is 1-critical. Then $\mathsf{M}^\star\left(\{\mathcal{I}_j\}\right) = \left|\cup_j \mathcal{I}_j\right| - 1$, which implies H is connected (and thus $\mathcal{E}$ is dominating) by (33). Consider the subhypergraph $H' = (\mathcal{V}, \mathcal{E}\backslash\{e\})$, which corresponds to the subfamily $\{\mathcal{I}_j - e\}$ of $\{\mathcal{I}_j\}$. Since $\{\mathcal{I}_j\}$ is 1-critical, we must have $\mathsf{M}^\star\left(\{\mathcal{I}_j - e\}\right) = \mathsf{M}^\star\left(\{\mathcal{I}_j\}\right) = \left|\cup_j \mathcal{I}_j\right| - 1 = \left|\cup_j (\mathcal{I}_j - e)\right|$. By (33), H' must be disconnected, and therefore $\mathcal{E}$ is a minimal connected dominating edge set.

On the other hand, suppose $\mathcal{E}$ is a minimal connected dominating edge set. Since H is connected, (33) implies

$$\mathsf{M}^\star\left(\{\mathcal{I}_j\}\right) \leq \left|\cup_j \mathcal{I}_j\right| - 1. \tag{34}$$

Since $\mathcal{E}$ is minimal, for any $e \in \mathcal{E}$, $H' = (\mathcal{V}, \mathcal{E}\backslash\{e\})$ is disconnected, and (33) implies

$$\mathsf{M}^\star\left(\{\mathcal{I}_j - e\}\right) \geq \left|\cup_j (\mathcal{I}_j - e)\right| = \left|\cup_j \mathcal{I}_j\right| - 1. \tag{35}$$

Applying Lemma 2, we must have $\mathsf{M}^\star\left(\{\mathcal{I}_j\}\right) = \mathsf{M}^\star\left(\{\mathcal{I}_j - e\}\right)$, and $\left|\cup_j \mathcal{I}_j\right| - \mathsf{M}^\star\left(\{\mathcal{I}_j\}\right) = 1$, which implies $\{\mathcal{I}_j\}$ is 1-critical. □

Remark 2: We note that the simple secret-sharing scheme in the proof of Lemma 4 has appeared several times in the literature. See, for example, [2], [15], [28].

We are finally in a position to prove Theorem 3.

Proof of Theorem 3: Let $H = (\mathcal{V}, \mathcal{E})$ be the hypergraph representation of $\{\mathcal{I}_j\}$. We can assume $\{\mathcal{I}_j\}$ generates a SK. By Theorem 5 and Lemma 4, computing $\mathsf{S}_{\mathrm{L}}\left(\{\mathcal{I}_j\}\right)$ is equivalent to computing the the number of edges in a minimum connected dominating edge set (i.e., a minimal connected dominating edge set with fewest possible edges). It is easy to see that the NP-complete SET COVER DECISION PROBLEM is a special case.

Indeed, consider any subsets $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_k$ whose union covers a finite set $\mathcal{U}$. For $u' \notin \mathcal{U}$, define $\mathcal{U}' = \mathcal{U} \cup \{u'\}$, and $\mathcal{A}'_j = \mathcal{A}_j \cup \{u'\}$ for $j \in [k]$. Clearly, $\{\mathcal{A}_{j_i}\}_{i=1}^m$ is a minimum cover of $\mathcal{U}$ if and only if $\{\mathcal{A}'_{j_i}\}_{i=1}^m$ is a minimum connected cover of $\mathcal{U}'$. □

Remark 3: Together, Theorem 5 and Lemma 4 give a succinct characterization of $\mathsf{S}_{\mathrm{L}}\left(\{\mathcal{I}_j\}\right)$ in terms of hypergraph connectivity. We extend this result to the generation of multiple secret keys at the end of Section IV using a stronger form of hypergraph connectivity.

B. Proof of Theorem 4

Before proving Theorem 4, consider the following constructive example: Let $n = 7$, and consider the family $\{\mathcal{I}_j\}$ defined by $\mathcal{I}_1 = \{1, 2, 3, 4\}$, and $\mathcal{I}_2, \ldots, \mathcal{I}_7$ are all $\binom{4}{2}$ distinct 2-element subsets of $\{1, 2, 3, 4\}$. By direct computation, we find that $\{\mathcal{I}_j - \{1\}\}$ is a minimum 1-critical subfamily, and hence $\mathsf{S}_{\mathrm{L}}\left(\{\mathcal{I}_j\}\right) = 2$ by Theorem 5. Suppose $\mathbb{F} = \{0, 1, \alpha, \beta\}^2 = \mathsf{GF}(4) \times \mathsf{GF}(4)$. Thus, we can express $X_j = (X_j^{(1)}, X_j^{(2)})$ for each $j = 1, \ldots, 4$, where $X_j^{(1)}, X_j^{(2)}$ are mutually independent, each equiprobable on $\mathsf{GF}(4)$. It is readily verified that the single transmission

$$\left(X_1^{(1)} + \alpha X_2^{(1)} + X_3^{(1)}, X_1^{(1)} + \beta X_2^{(1)} + X_4^{(1)}\right) \in \mathbb{F} \quad (36)$$

by client c_1 permits reconstruction of the SK

$$\mathsf{k}_1\left(\{X_i : i \in \mathcal{I}_1\}, \mathbf{T}(\underline{X})\right) = (X_3^{(1)}, X_4^{(1)}) \in \mathbb{F} \quad (37)$$

at all clients. Hence, we can conclude $1 = \mathsf{S}\left(\{\mathcal{I}_j\}\right) < \mathsf{S}_\mathrm{L}\left(\{\mathcal{I}_j\}\right) = \mathsf{M}^\star\left(\{\mathcal{I}_j\}\right) = 2$.

The above construction is a *vector-linear* transmission protocol, and cannot be realized by a protocol which is linear over $\mathbb{F}$. A natural question is whether it is possible to bound the gap between $\mathsf{S}\left(\{\mathcal{I}_j\}\right)$ and $\mathsf{S}_\mathrm{L}\left(\{\mathcal{I}_j\}\right)$. As asserted by Theorem 4, the answer to this is negative. Indeed, it is straightforward to generalize the previous construction and make the gap arbitrarily large.

To this end, consider a network of $n = \binom{m}{2} + 1$ clients such that $\mathcal{I}_1 = [m]$ and the other $\binom{m}{2}$ clients possess distinct pairs of messages. Observe that the 1-critical subfamilies of $\{\mathcal{I}_j\}$ are obtained by removing a single message – i.e., if any two messages $m_1, m_2 \in [m]$ are removed then the resulting hypergraph representation of $\{\mathcal{I}_j - \{m_1, m_2\}\}$ is no longer connected. This implies that $\mathsf{S}_\mathrm{L}\left(\{\mathcal{I}_j\}\right) = m - 2$. To show that there exists a nonlinear scheme that can do better, we show that $\mathsf{M}^\star(\{\mathcal{I}_j\}) = m - 2$:

- To show that $\mathsf{M}^\star(\{\mathcal{I}_j\}) \leq m - 2$, let client c_1 transmit $m - 2$ independent linear combinations of the messages. Provided the encoding matrix A is full rank (e.g., a Vandermonde matrix), every other node can use its own pair of messages to recover the other $m - 2$.
- We note that $\mathsf{M}^\star(\{\mathcal{I}_j\}) \geq \mathsf{S}_\mathrm{L}\left(\{\mathcal{I}_j\}\right) = m - 2$ by Lemma 1, and therefore $\mathsf{M}^\star(\{\mathcal{I}_j\}) = m - 2$ as claimed.

Now, we simply split the packets and apply the optimal transmission protocol over the first halves of the packets as we did previously. This vector-linear scheme generates a SK with $m/2 - 1$ transmissions, which is an improvement of $m/2 - 1$ transmissions over the best linear scheme. Since m was arbitrary, we have shown that the gap between $\mathsf{S}\left(\{\mathcal{I}_j\}\right)$ and $\mathsf{S}_\mathrm{L}\left(\{\mathcal{I}_j\}\right)$ cannot be bounded in general, proving Theorem 4.

Remark 4: Our proof that $\mathsf{S}\left(\{\mathcal{I}_j\}\right) < \mathsf{S}_\mathrm{L}\left(\{\mathcal{I}_j\}\right)$ is similar to the index coding problem, where the suboptimality of linear schemes was also shown by demonstrating a gap between the performance of linear and vector-linear coding schemes [25], [26]. For several years, it was unknown whether vector-linear coding schemes were optimal in the index coding problem. However, Blasiak et al. have since proved that even vector-linear coding is strictly suboptimal for the index coding problem [27]. We conjecture the same is true for the present setting.

IV. Generating Multiple Secret Keys

Until now, we have focused exclusively on protocols that generate a single SK. However, it is also natural to consider protocols that generate τ independent secret keys. Indeed, the secrecy capacity as defined in [2] translates to the maximum number of secret keys that can possibly be generated in the combinatorial setting we consider. Thus, it is interesting to study the tradeoff between the number of secret keys that can be generated and the number of public transmissions required to do so.

To this end, we say a transmission protocol $\mathbf{P}$ (with corresponding transmission sequence $\mathbf{T}(\underline{X})$) generates τ secret keys if there exist decoding functions $\{\mathsf{k}_1, \mathsf{k}_2, \ldots, \mathsf{k}_n\}$ which satisfy the following three properties:

(i) For all $j \in [n]$, and with probability 1,

$$\mathsf{k}_j\left(\{X_i : i \in \mathcal{I}_j\}, \mathbf{T}(\underline{X})\right) = \mathsf{k}_1\left(\{X_i : i \in \mathcal{I}_1\}, \mathbf{T}(\underline{X})\right).$$

(ii) $\mathsf{k}_1\left(\{X_i : i \in \mathcal{I}_1\}, \mathbf{T}(\underline{X})\right)$ is equiprobable on $\mathbb{F}^\tau$.

(iii) $I\left(\mathsf{k}_1\left(\{X_i : i \in \mathcal{I}_1\}, \mathbf{T}(\underline{X})\right); \mathbf{T}(\underline{X})\right) = 0$.

Note that (i)–(iii) are the same requirements for generating a single SK with one exception: we require that $\mathsf{k}_1\left(\{X_i : i \in \mathcal{I}_1\}, \mathbf{T}(\underline{X})\right)$ is uniformly distributed over $\mathbb{F}^\tau$. In other words, we require that each client recovers τ independent SKs, each known to all clients and private from any eavesdropper. As stated in [4, Th. 6], Theorem 2 can be generalized as follows:

Theorem 6: For a network defined by $\{\mathcal{I}_j\}$, there exists a protocol $\mathbf{P}$ which generates τ SKs if and only if

$$\left|\cup_j \mathcal{I}_j\right| \geq \mathsf{M}^\star\left(\{\mathcal{I}_j\}\right) + \tau. \quad (38)$$

Analogous to the definition of $\mathsf{S}_\mathrm{L}\left(\{\mathcal{I}_j\}\right)$ in (7), let $\mathsf{S}_\mathrm{L}^{(\tau)}\left(\{\mathcal{I}_j\}\right)$ denote the minimum number of transmissions required by a linear protocol to generate τ independent secret keys. A minor modification of our arguments for the single-SK setting yields:

Theorem 7: Let $\tau \geq 1$ be an integer. If there is a protocol $\mathbf{P}$ for $\{\mathcal{I}_j\}$ which generates τ independent secret keys, then

$$\mathsf{S}_\mathrm{L}^{(\tau)}\left(\{\mathcal{I}_j\}\right) = \mathsf{M}^\star\left(\{\mathcal{J}_j^\star\}\right) = \left|\cup_j \mathcal{J}_j^\star\right| - \tau, \quad (39)$$

where $\{\mathcal{J}_j^\star\}$ is a minimum τ-critical subfamily of $\{\mathcal{I}_j\}$.

In the single-SK setting, Lemma 4 gave a succinct interpretation of minimum 1-critical subfamilies of $\{\mathcal{I}_j\}$ as connectedness of H, the hypergraph representation of $\{\mathcal{I}_j\}$. When combined with Theorem 5, we find that $\mathsf{S}_\mathrm{L}\left(\{\mathcal{I}_j\}\right)$ is in one-to-one correspondence with the size of a minimum connected dominating edge-set of H. The chief difficulty in giving a similarly succinct characterization of $\mathsf{S}^{(\tau)}\left(\{\mathcal{I}_j\}\right)$ lies in generalizing Lemma 4 appropriately for $\tau \geq 2$. In order to do so, we will need to introduce a more general notion of hypergraph connectivity.

Toward this end, recall that a *multigraph* is a graph that is permitted to have multiple edges connecting a pair of nodes (note that this is distinct from a hypergraph, in which an edge connects multiple vertices). A classical result of Nash-Williams [29] and Tutte [30] is the following:

Theorem 8 [29], [30]: An undirected multigraph $G = (\mathcal{V}, \mathcal{E})$ contains τ edge-disjoint spanning trees iff for every partition $\mathscr{P}$ of $\mathcal{V}$ into disjoint sets $\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_{|\mathscr{P}|}$,

$$\sum_{e \in \mathcal{E}} (r(e\,;\mathscr{P}) - 1) \geq \tau(|\mathscr{P}| - 1), \quad (40)$$

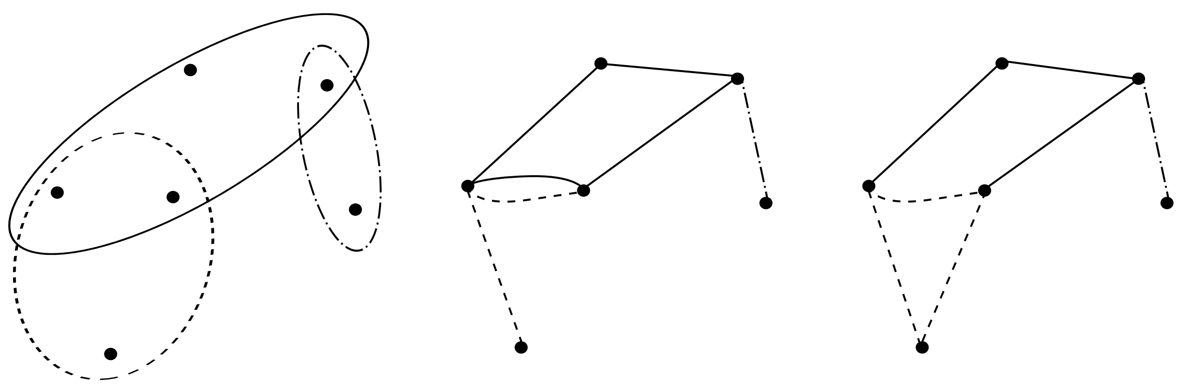

Fig. 1. An example of a hypergraph H (left) and two induced multigraphs (center, right). Line textures are used to emphasize the relationship between the hypergraph edges and the decomposition of the multigraphs into corresponding simple connected graphs.

where $r(e\,;\mathscr{P})$ denotes the number of parts in $\mathscr{P}$ that the edge e intersects.

Definition 2: A multigraph $G = (\mathcal{V}, \mathcal{E}_M)$ is induced by a hypergraph *$H = (\mathcal{V}, \mathcal{E})$ if it can be decomposed into a disjoint collection of simple graphs $\{G_e\}_{e\in\mathcal{E}}$, where $G_e = (e, E_e)$ is a connected graph on the vertex set $e \in \mathcal{E}$.*

Two examples of multigraphs induced by a hypergraph are given in Figure 1. We are now in a position to define our notion of hypergraph connectivity:

Definition 3: A hypergraph $H = (\mathcal{V}, \mathcal{E})$ is inherently τ-connected *if every induced multigraph contains at least τ edge-disjoint spanning trees.*

A pleasant generalization of Theorem 8 holds for inherently τ-connected hypergraphs.

Theorem 9: A hypergraph $H = (\mathcal{V}, \mathcal{E})$ is inherently τ-connected iff for any partition $\mathscr{P}$ of $\mathcal{V}$ into disjoint sets $\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_{|\mathscr{P}|}$,

$$\sum_{e\in\mathcal{E}}(r(e\,;\mathscr{P}) - 1) \geq \tau(|\mathscr{P}| - 1), \tag{41}$$

where $r(e\,;\mathscr{P})$ is the number of parts in $\mathscr{P}$ that the hyperedge e intersects.

Theorem 9 follows as an easy corollary of Theorem 8 and the definition of an inherently τ-connected hypergraph. However, a stronger version of Theorem 9 can be distilled from our proof of Lemma 5, which is stated shortly. Specifically, we will see that a hypergraph H is inherently τ-connected iff a relatively small subset of multigraphs induced by H contain τ edge-disjoint spanning trees. For a precise statement, see our remark following the proof of Lemma 5.

For a hypergraph $H = (\mathcal{V}, \mathcal{E})$, an edge set $\mathcal{E}' \subseteq \mathcal{E}$ is a minimal inherently τ-connected edge-set if the subhypergraph $H' = (\mathcal{V}, \mathcal{E}')$ is inherently τ-connected, and the removal of any edge from $\mathcal{E}'$ results in a subhypergraph that is not inherently τ-connected. Further, define

$$\varrho_\tau(H) = \min\Big\{ \left|\mathcal{E}'\right| : \mathcal{E}' \subseteq \mathcal{E} \text{ is an inherently } \tau\text{-connected edge-set}\Big\}. \tag{42}$$

In other words, $\varrho_\tau(H)$ is the minimum number of edges in an inherently τ-connected subhypergraph $H' = (\mathcal{V}, \mathcal{E}')$ of $H = (\mathcal{V}, \mathcal{E})$. Note that $\varrho_\tau(H)$ is the minimum number of edges in a connected dominating edge set when $\tau = 1$, and thus its computation is NP-hard in general.

Lemma 5: Let $H = (\mathcal{V}, \mathcal{E})$ be the hypergraph representation of $\{\mathcal{I}_j\}$. H is inherently τ-connected if and only if

$$\mathsf{M}^\star\left(\{\mathcal{I}_j\}\right) \leq \left|\cup_j \mathcal{I}_j\right| - \tau. \tag{43}$$

In particular, $\{\mathcal{I}_j\}$ is τ-critical if and only if $\mathcal{E}$ is a minimal inherently τ-connected edge-set.

Before we begin the proof of Lemma 5, we take a moment to describe a special class of multigraphs that are induced by H. For a hypergraph $H = (\mathcal{V}, \mathcal{E})$, let $\prec$ be a strict total order on $\mathcal{V}$. That is, if $\mathcal{V} = \{v_1, v_2, \ldots, v_n\}$, there is a permutation π on $\{1, \ldots, n\}$ for which $v_{\pi(1)} \prec v_{\pi(2)} \prec \cdots \prec v_{\pi(n)}$. Define the multigraph $G_{H,\prec}$ induced by H, with decomposition $\{G_e\}_{e\in\mathcal{E}}$, as follows: For each $e \in \mathcal{E}$, let G_e be a path that connects the vertices contained in e in ascending order (with respect to $\prec$). In other words, if $e = \{v_{i_1}, v_{i_2}, \ldots, v_{i_k}\}$, where $v_{i_j} \prec v_{i_\ell}$ for $i_j < i_\ell$, then the edge-set of G_e is precisely $\{v_{i_1}, v_{i_2}\}, \{v_{i_2}, v_{i_3}\}, \ldots, \{v_{i_{k-1}}, v_{i_k}\}$. An example is shown in Figure 2.

Proof of Lemma 5: Let $(a_1^\star, \ldots, a_n^\star)$ be an optimal solution to ILP (4). First, suppose $\mathsf{M}^\star\left(\{\mathcal{I}_j\}\right) \leq \left|\cup_j \mathcal{I}_j\right| - \tau$. Then, for any partition $\mathscr{P} = \{\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_k\}$ of $\mathcal{V}$, we have:

$$\begin{aligned}
(\left|\cup_j \mathcal{I}_j\right| - \tau)(k-1) &\geq \mathsf{M}^\star\left(\{\mathcal{I}_j\}\right)(k-1) && (44)\\
&= \sum_{i=1}^{k}\left(\mathsf{M}^\star\left(\{\mathcal{I}_j\}\right) - \sum_{j\in\mathcal{V}_i} a_j^\star\right) && (45)\\
&= \sum_{i=1}^{k}\sum_{j\in\bar{\mathcal{V}}_i} a_j^\star && (46)\\
&\geq \sum_{i=1}^{k}\left|\bigcap_{j\in\mathcal{V}_i}\bar{\mathcal{I}}_j\right| && (47)\\
&= k\left|\cup_j \mathcal{I}_j\right| - \sum_{i=1}^{k}\left|\bigcup_{j\in\mathcal{V}_i}\mathcal{I}_j\right|, && (48)
\end{aligned}$$

where (47) follows by feasibility of $(a_1^\star, \ldots, a_n^\star)$ for ILP (4).

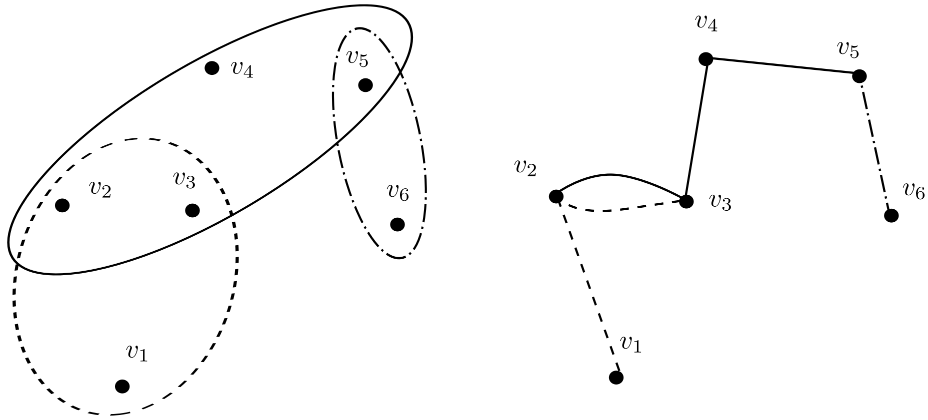

Fig. 2. A hypergraph H (left) and the induced multigraph $G_{H,\prec}$ (right) for the vertex-ordering $v_1 \prec v_2 \prec \cdots \prec v_6$. Line textures are used to emphasize the relationship between the hypergraph edges and the decomposition of the multigraphs into corresponding simple connected graphs.

Rearranging, we find

$$\sum_{i=1}^{k}\left|\bigcup_{j\in\mathcal{V}_i}\mathcal{I}_j\right| \geq \left|\cup_j\mathcal{I}_j\right| + \tau(k-1). \tag{49}$$

Now, let G be an arbitrary multigraph induced by H with decomposition given by $\{G_e\}_{e\in\mathcal{E}}$. Note that if $e \in \mathcal{E}$ intersects $r(e\,;\mathscr{P})$ parts of the partition $\mathscr{P}$, then at least $r(e\,;\mathscr{P}) - 1$ edges of G_e cross the partition $\mathscr{P}$. Therefore,

$$\Omega(G,\mathscr{P}) \geq \sum_{e\in\mathcal{E}}(r(e\,;\mathscr{P})-1) = \left(\sum_{i=1}^{k}\left|\bigcup_{j\in\mathcal{V}_i}\mathcal{I}_j\right|\right) - \left|\cup_j\mathcal{I}_j\right|, \tag{50}$$

where $\Omega(G,\mathscr{P})$ denotes the number of edges in G that cross the partition $\mathscr{P}$. Since the partition $\mathscr{P}$ and induced multigraph G were arbitrary, it follows from Theorem 8 and (49) that H is inherently τ-connected. Thus, we have shown:

$$\mathsf{M}^\star\left(\{\mathcal{I}_j\}\right) \leq \left|\cup_j\mathcal{I}_j\right| - \tau \implies H \text{ is inherently } \tau\text{-connected.} \tag{51}$$

Next suppose H is inherently τ-connected. By optimality of $(a_1^\star,\ldots,a_n^\star)$, there exists a partition $\mathscr{P}^\star = \{\mathcal{V}_1,\mathcal{V}_2,\ldots,\mathcal{V}_k\}$ of $\mathcal{V}$ (see [4, Appendix A], [31]) such that

$$\sum_{i=1}^{k}\sum_{j\in\bar{\mathcal{V}}_i} a_j^\star = \sum_{i=1}^{k}\left|\bigcap_{j\in\mathcal{V}_i}\bar{\mathcal{I}}_j\right|. \tag{52}$$

Now, consider an arbitrary order $\prec$ on $\mathcal{V}$ which satisfies $u \prec v$ if $u \in \mathcal{V}_i$, $v \in \mathcal{V}_j$ and $i < j$. In this case, if $e \in \mathcal{E}$ intersects $r(e\,;\mathscr{P}^\star)$ parts of the partition $\mathscr{P}^\star$, then the path in $G_{H,\prec}$ generated by the hyperedge e (i.e., G_e) will have precisely $r(e\,;\mathscr{P}^\star) - 1$ edges that cross $\mathscr{P}^\star$. Since H is inherently τ-connected, we have

$$\begin{aligned}\left(\sum_{i=1}^{k}\left|\bigcup_{j\in\mathcal{V}_i}\mathcal{I}_j\right|\right) - \left|\cup_j\mathcal{I}_j\right| &= \sum_{e\in\mathcal{E}}(r(e\,;\mathscr{P}^\star)-1)\\ &= \Omega\left(G_{H,\prec},\mathscr{P}^\star\right) \geq \tau(k-1)\end{aligned} \tag{53}$$

by Theorem 8. Proceeding in a fashion similar to before, we have for $\mathscr{P}^\star$ that

$$\begin{aligned}\mathsf{M}^\star\left(\{\mathcal{I}_j\}\right)(k-1) &= \sum_{i=1}^{k}\left(\mathsf{M}^\star\left(\{\mathcal{I}_j\}\right) - \sum_{j\in\mathcal{V}_i} a_j^\star\right) && (54)\\ &= \sum_{i=1}^{k}\sum_{j\in\bar{\mathcal{V}}_i} a_j^\star && (55)\\ &= \sum_{i=1}^{k}\left|\bigcap_{j\in\mathcal{V}_i}\bar{\mathcal{I}}_j\right| && (56)\\ &= k\left|\cup_j\mathcal{I}_j\right| - \sum_{i=1}^{k}\left|\bigcup_{j\in\mathcal{V}_i}\mathcal{I}_j\right| && (57)\\ &\leq (k-1)(\left|\cup_j\mathcal{I}_j\right| - \tau), && (58)\end{aligned}$$

where the final inequality follows from (53). Hence,

$$\begin{aligned}&\mathsf{M}^\star\left(\{\mathcal{I}_j\}\right) \leq \left|\cup_j\mathcal{I}_j\right| - \tau\\ &\quad\iff H \text{ is inherently } \tau\text{-connected.}\end{aligned} \tag{59}$$

We now prove the second claim. To this end, suppose $\{\mathcal{I}_j\}$ is τ-critical. Then $\mathsf{M}^\star\left(\{\mathcal{I}_j\}\right) = \left|\cup_j\mathcal{I}_j\right| - \tau$, which implies H is inherently τ-connected by (59). Consider the subhypergraph $H' = (\mathcal{V},\mathcal{E}\backslash\{e\})$, which corresponds to the subfamily $\{\mathcal{I}_j - e\}$ of $\{\mathcal{I}_j\}$. Since $\{\mathcal{I}_j\}$ is τ-critical, we must have $\mathsf{M}^\star\left(\{\mathcal{I}_j - e\}\right) = \mathsf{M}^\star\left(\{\mathcal{I}_j\}\right) = \left|\cup_j\mathcal{I}_j\right| - \tau = \left|\cup_j(\mathcal{I}_j - e)\right| - \tau + 1$. By (59), H' cannot be inherently τ-connected, and therefore $\mathcal{E}$ is a minimal inherently τ-connected edge-set.

On the other hand, suppose $\mathcal{E}$ is a minimal inherently τ-connected edge-set. Then, (59) implies

$$\mathsf{M}^\star\left(\{\mathcal{I}_j\}\right) \leq \left|\cup_j\mathcal{I}_j\right| - \tau. \tag{60}$$

Since $\mathcal{E}$ is a inherently τ-connected edge-set, for any $e \in \mathcal{E}$, $H' = (\mathcal{V},\mathcal{E}\backslash\{e\})$ is not inherently τ-connected, and (59) implies

$$\mathsf{M}^\star\left(\{\mathcal{I}_j - e\}\right) \geq \left|\cup_j(\mathcal{I}_j - e)\right| - \tau + 1 = \left|\cup_j\mathcal{I}_j\right| - \tau. \tag{61}$$

TABLE I

$\mathsf{S}_{\mathrm{L}}^{(\tau)}\left(\{\mathcal{I}_j\}\right)$ VS. τ FOR THE NETWORK GIVEN IN EXAMPLE 1. (Note that $\mathsf{S}_{\mathrm{L}}^{(\tau)}\left(\{\mathcal{I}_j\}\right) = \infty$ indicates that it is not possible to generate τ secret keys with any number of transmissions)

τ	1	2	3	4	5	6	≥ 7
$\mathsf{S}_{\mathrm{L}}^{(\tau)}(\{\mathcal{I}_j\})$	2	4	4	6	8	8	∞

Applying Lemma 2, we must have $\mathsf{M}^\star\left(\{\mathcal{I}_j\}\right) = \mathsf{M}^\star\left(\{\mathcal{I}_j - e\}\right)$, and $\left|\cup_j \mathcal{I}_j\right| - \mathsf{M}^\star\left(\{\mathcal{I}_j\}\right) = \tau$, which implies $\{\mathcal{I}_j\}$ is τ-critical. $\square$

Remark 5: From the proof of Lemma 5, we observe that a hypergraph H is inherently τ-connected if and only if $G_{H,\prec}$ contains τ edge-disjoint spanning trees for every strict order $\prec$. Hence, this apparently weaker condition is, in fact, necessary and sufficient for any multigraph induced by H to contain τ edge-disjoint spanning trees.

In summary, we have found the following characterization of $\mathsf{S}_{\mathrm{L}}^{(\tau)}\left(\{\mathcal{I}_j\}\right)$:

Theorem 10: If H is the hypergraph representation of the network defined by $\{\mathcal{I}_j\}$, then

$$\mathsf{S}_{\mathrm{L}}^{(\tau)}\left(\{\mathcal{I}_j\}\right) = \varrho_\tau(H) - \tau. \tag{62}$$

When we restrict ourselves to linear protocols, Theorem 10 elucidates a direct correspondence between the number of public transmissions required to generate τ SKs in a network and the inherent τ-connectivity of the representative hypergraph.

As an illustrative example of Theorem 10, consider the following network with 15 clients:

Example 1: Let $\mathcal{I}_1 = \{5, 7, 10, 11, 13, 14, 15\}$, and let $\{\mathcal{I}_j\}_{j=1}^{15}$ be the 14 different cyclic shifts of $\mathcal{I}_1$ (e.g., $\mathcal{I}_2 = \{1, 6, 8, 11, 12, 14, 15\}$, $\mathcal{I}_3 = \{1, 2, 7, 9, 12, 13, 15\}$, $\ldots$). Since the number of messages $m = 15$ is modestly small, we are able to compute $\varrho_\tau(H)$ explicitly for the hypergraph representation of the network defined by $\{\mathcal{I}_j\}$, and therefore also $\mathsf{S}_{\mathrm{L}}^{(\tau)}\left(\{\mathcal{I}_j\}\right)$ by invoking Theorem 10. Above, Table I gives $\mathsf{S}_{\mathrm{L}}^{(\tau)}\left(\{\mathcal{I}_j\}\right)$ for $\tau \geq 1$.

In another example, which generalizes a recent result due to Mukherjee and Kashyap [19], we give a complete characterization for $\mathsf{S}_{\mathrm{L}}^{(\tau)}\left(\{\mathcal{I}_j\}\right)$ when each pair of clients shares a unique message (i.e., $m = \binom{n}{2}$), and the hypergraph representation of the network defined by $\{\mathcal{I}_j\}$ is a complete (simple) graph on n vertices). This network model was called the PIN model by Nitinawarat and Narayan [28] who showed under this model that linear schemes can generate perfect secrecy at maximum-rate, which is characterized by a maximum packing of disjoint spanning trees in a multigraph.

Example 2: In the PIN model, $\mathsf{S}_{\mathrm{L}}^{(\tau)}\left(\{\mathcal{I}_j\}\right) = \tau(n-2)$, where $1 \leq \tau \leq \lfloor n/2 \rfloor$. Indeed, a simple graph is inherently τ-connected iff it contains τ edge-disjoint spanning trees by Theorem 8 (note that a minimal dominating edge set corresponds to a spanning tree in a simple graph). Thus, a simple counting argument gives $\varrho_\tau(H) = \tau(n-1)$. An application of Theorem 10 proves the claim.

Although the work [28] is primarily concerned with SK capacity in the PIN model (rather than communication requirements as a function of SK size), spanning trees play a prominent role. Therefore, since the PIN model is a special case of the hypergraph source model we consider, it is reassuring to observe that the quantity $\mathsf{S}_{\mathrm{L}}^{(\tau)}\left(\{\mathcal{I}_j\}\right)$ is closely related to packings of spanning trees. Finally, we remark that there is some similarity between Theorem 10 and the work of Tyagi *et al.* [32] which showed that PIN network which is τ-connected is capable of generating a SK even if $\tau - 1$ nodes refuse to participate.

V. Concluding Remarks

In this paper, we have completely characterized the number of public transmissions required to generate a specified number of SKs when linear transmission protocols are employed and a hypergraphical source model is considered. The minimum number of transmissions required by a linear protocol to generate τ secret keys is succinctly given in terms of the inherent τ-connectivity of the hypergraph associated with the source model. We have also shown that computing said minimum number of transmissions is NP-hard.

Moreover, we have established that there can be a gap between the number of transmissions required by a nonlinear transmission scheme and the number of transmissions required by the best linear transmission scheme, and that this gap can be arbitrarily large. The problem of characterizing the number of public transmissions required by a nonlinear scheme remains an open problem, and appears to be very challenging.

Finally, we note that it is an interesting combinatorial design problem to specify ideal message distributions amongst clients (subject to constraints) that allow SK agreement with fewest transmissions. For example, how many transmissions are required to generate a SK subject to the constraint that each message is initially held by at most t clients? This general problem is beyond the scope of the present paper, and is left for future work.

References

[1] T. A. Courtade and T. R. Halford, "Coded cooperative data exchange for a secret key," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun./Jul. 2014, pp. 776–780.

[2] I. Csiszár and P. Narayan, "Secrecy capacities for multiple terminals," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3047–3061, Dec. 2004.

[3] S. El Rouayheb, A. Sprintson, and P. Sadeghi, "On coding for cooperative data exchange," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Jan. 2010, pp. 1–5.

[4] T. A. Courtade and R. D. Wesel, "Coded cooperative data exchange in multihop networks," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 1136–1158, Feb. 2014.

[5] A. Sprintson, P. Sadeghi, G. Booker, and S. El Rouayheb, "A randomized algorithm and performance bounds for coded cooperative data exchange," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2010, pp. 1888–1892.

[6] A. Sprintson, P. Sadeghi, G. Booker, and S. El Rouayheb, "Deterministic algorithm for coded cooperative data exchange," in *Proc. ICST QShine*, Nov. 2010, pp. 282–289.

[7] T. A. Courtade and R. D. Wesel, "Efficient universal recovery in broadcast networks," in *Proc. 48th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Oct. 2010, pp. 1542–1549.

[8] T. A. Courtade and R. D. Wesel, "Weighted universal recovery, practical secrecy, and an efficient algorithm for solving both," in *Proc. 49th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Sep. 2011, pp. 1349–1357.

[9] M. Gonen and M. Langberg, "Coded cooperative data exchange problem for general topologies," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2012, pp. 2606–2610.
[10] N. Milosavljevic, S. Pawar, S. El Rouayheb, M. Gastpar, and K. Ramchandran. (2012). "Data exchange problem with helpers." [Online]. Available: http://arxiv.org/abs/1202.1612
[11] N. Milosavljevic, S. Pawar, S. El Rouayheb, M. Gastpar, and K. Ramchandran. (Apr. 2011). "Optimal deterministic polynomial-time data exchange for omniscience." [Online]. Available: http://arxiv.org/abs/1108.6046
[12] S. E. Tajbakhsh, P. Sadeghi, and R. Shams, "A generalized model for cost and fairness analysis in coded cooperative data exchange," in *Proc. Int. Symp. Netw. Coding (NetCod)*, Beijing, China, Jul. 2011, pp. 1–6.
[13] D. Ozgul and A. Sprintson, "An algorithm for cooperative data exchange with cost criterion," in *Proc. Inf. Theory Appl. Workshop (ITA)*, Feb. 2011, pp. 1–4.
[14] I.-H. Hou, Y.-P. Hsu, and A. Sprintson, "Truthful and non-monetary mechanism for direct data exchange," in *Proc. 51st Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Monticello, IL, USA, Oct. 2013, pp. 406–412.
[15] C. Chan and L. Zheng, "Mutual dependence for secret key agreement," in *Proc. 44th Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2010, pp. 1–6.
[16] C. Chan, "Generating secret in a network," Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Massachusetts Inst. Technol., Cambridge, MA, USA, 2010.
[17] C. Chan, "Linear perfect secret key agreement," in *Proc. IEEE Inf. Theory Workshop Process. (ITW)*, Oct. 2011, pp. 723–726.
[18] H. Tyagi, "Common information and secret key capacity," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5627–5640, Sep. 2013.
[19] M. Mukherjee and N. Kashyap. (2014). "On the communication complexity of secret key generation in the multiterminal source model." [Online]. Available: http://arxiv.org/abs/1401.1117
[20] M. Yan and A. Sprintson, "Algorithms for weakly secure data exchange," in *Proc. Int. Symp. Netw. Coding (NetCod)*, 2013, pp. 1–6.
[21] S. H. Dau, W. Song, and C. Yuen. (2014). "On the existence of MDS codes over small fields with constrained generator matrices." [Online]. Available: http://arxiv.org/abs/1401.3807
[22] T. R. Halford, T. A. Courtade, K. M. Chugg, X. Li, and G. Thatte, "Energy-efficient group key agreement for wireless networks," *IEEE Trans. Wireless Comms.*, vol. 14, no. 10, pp. 5552–5564, Oct. 2015.
[23] S. Jaggi *et al.*, "Polynomial time algorithms for multicast network code construction," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 1973–1982, Jun. 2005.
[24] M. Effros, S. El Rouayheb, and M. Langberg. "An equivalence between network coding and index coding," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2478–2487, 2015.
[25] N. Alon, E. Lubetzky, U. Stav, A. Weinstein, and A. Hassidim, "Broadcasting with side information," in *Proc. IEEE 49th Annu. IEEE Symp. Found. Comput. Sci. (FOCS)*, Oct. 2008, pp. 823–832.
[26] E. Lubetzky and U. Stav, "Nonlinear index coding outperforming the linear optimum," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3544–3551, Aug. 2009.
[27] A. Blasiak, R. Kleinberg, and E. Lubetzky, "Lexicographic products and the power of non-linear network coding," in *Proc. IEEE 52nd Annu. Symp. Found. Comput. Sci. (FOCS)*, Oct. 2011, pp. 609–618.
[28] S. Nitinawarat and P. Narayan, "Perfect omniscience, perfect secrecy, and Steiner tree packing," *IEEE Trans. Inf. Theory*, vol. 56, no. 12, pp. 6490–6500, Dec. 2010.
[29] C. S. J. A. Nash-Williams, "Edge-disjoint spanning trees of finite graphs," *J. London Math. Soc.*, vol. 1, no. 1, pp. 445–450, 1961.
[30] W. T. Tutte, "On the problem of decomposing a graph into n connected factors," *J. London Math. Soc.*, vol. 1, no. 1, pp. 221–230, 1961.
[31] A. Schrijver, *Combinatorial Optimization: Polyhedra and Efficiency*. Berlin, Germany: Springer-Verlag, 2003.
[32] H. Tyagi, N. Kashyap, Y. Sankarasubramaniam, and K. Viswanathan, "Fault-tolerant secret key generation," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2012, pp. 1787–1791.

Thomas A. Courtade (S'06–M'13) is an Assistant Professor in the Department of Electrical Engineering and Computer Sciences at the University of California, Berkeley. Prior to joining UC Berkeley in 2014, he was a postdoctoral fellow supported by the NSF Center for Science of Information. He received his Ph.D. and M.S. degrees from UCLA in 2012 and 2008, respectively, and he graduated summa cum laude with a B.Sc. in Electrical Engineering from Michigan Technological University in 2007. His honors include a Distinguished Ph.D. Dissertation award and an Excellence in Teaching award from the UCLA Department of Electrical Engineering, and a Jack Keil Wolf Student Paper Award for the 2012 International Symposium on Information Theory.

Thomas R. Halford (S'02–M'15) received the B.A.Sc. degree in Engineering Physics from Simon Fraser University, Burnaby, Canada in 2001, and the Ph.D. degree in Electrical Engineering from the University of Southern California, Los Angeles, CA, in 2007. From 2007 to 2015, he was with TrellisWare Technologies, Inc. He currently leads technology and business development at WPL, Inc. in Manhattan Beach, CA. His research interests span all aspects of military communications. He served as a track co-chair at the 2015 and 2016 Military Communications Conferences.