

# Secrecy Capacities for Multiterminal Channel Models

Imre Csiszár, *Fellow, IEEE*, and Prakash Narayan, *Fellow, IEEE*

**Abstract**—Shannon-theoretic secret key generation by several parties is considered for models in which a secure noisy channel with one input terminal and multiple output terminals and a public noiseless channel of unlimited capacity are available for accomplishing this goal. The secret key is generated for a set  $A$  of terminals of the noisy channel, with the remaining terminals (if any) cooperating in this task through their public communication. Single-letter characterizations of secrecy capacities are obtained for models in which secrecy is required from an eavesdropper that observes only the public communication and perhaps also a set of terminals disjoint from  $A$ . These capacities are shown to be achievable with noninteractive public communication, the channel input terminal sending no public message and each output terminal sending at most one public message, not using randomization. Moreover, when the input terminal belongs to the set  $A$ , it can generate the secret key at the outset and transmit it over the noisy channel, suitably encoded, whereupon the output terminals in  $A$  securely recover this key using public communication as above. For models in which the eavesdropper also possesses side information that is not available to any of the terminals cooperating in secrecy generation, an upper bound for the secrecy capacity and a sufficient condition for its tightness are given.

**Index Terms**—Multiple source, multiterminal channel, private key, secrecy capacity, secret key, wiretap side information.

## I. INTRODUCTION

IT is now well recognized that separate terminals, with the means to transact over a secure noisy channel as well as a public noiseless channel, might devise a secret key more effectively than by using the secure channel alone. A secret key, in the Shannon-theoretic sense, is common randomness of near-uniform distribution regarding which an eavesdropper, which observes the public communication and perhaps also possesses additional observations available or unavailable to the terminals engaged in secrecy generation, can glean only a negligible amount of information.

The first Shannon-theoretic model for generating a secret key over a noisy channel was Wyner's wiretap channel [18], generalized by Csiszár and Körner [4]. This model did not allow

Manuscript received November 22, 2006; revised November 29, 2007. The work of I. Csiszár was supported by the Hungarian National Foundation for Scientific Research under Grant T046376. The work of P. Narayan was supported by the National Science Foundation under Grants CCF0515124 and CCF0635271. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Nice, France, June 2007.

I. Csiszár is with the A. Rényi Institute of Mathematics, Hungarian Academy of Sciences, POB 127, H-1364 Budapest, Hungary (e-mail: csiszar@renyi.hu).

P. Narayan is with the Department of Electrical and Computer Engineering and the Institute for Systems Research, University of Maryland, College Park, MD 20742 USA (e-mail: prakash@eng.umd.edu).

Communicated by U. Maurer, Guest Editor for Special Issue on Information Theoretic Security.

Digital Object Identifier 10.1109/TIT.2008.921705

for public communication, and secret key generation was tantamount to secure transmission over the noisy channel, when the eavesdropper had access to wiretap side information. The fact that secrecy generation could be enhanced by public communication was first illustrated by Maurer [12]. Models for secrecy generation which entailed two terminals communicating over a public noiseless channel, were examined in detail by Maurer [13] and Ahlswede and Csiszár [1]. These models involve either a discrete memoryless multiple source (DMMS) with two components accessible to one terminal each, or a discrete memoryless channel (DMC) with one input terminal and one output terminal. In both types of models, an additional "wiretapped" terminal may or may not be present. The sumptuous literature on such models includes Maurer [14], Bennett, Brassard, Crépeau, and Maurer [2], Csiszár [3], Maurer and Wolf [15], [16], Csiszár and Narayan [7], and Renner and Wolf [17]. A single-letter characterization of the secrecy capacity—the largest rate at which a secret key can be generated—is known in special cases, e.g., when a wiretapped terminal is absent or when the wiretapped terminal reveals itself to the parties generating secrecy.

In an earlier paper [7], we have studied secrecy generation for a multiterminal source model where each participating terminal had access to one component of a DMMS, which was followed by public noiseless communication. To our knowledge, multiterminal channel models for generating secrecy have not been studied heretofore. In this paper, we examine channel models for secret key generation which involve an underlying DMC with a single input and  $m \geq 1$  outputs. Terminal 0 governs the input and terminals  $1, \dots, m$  observe the corresponding outputs. Following each symbol transmission over the DMC, communication over a public noiseless channel of unlimited capacity is allowed between all the terminals, which may be interactive and which is observed by all the terminals.<sup>1</sup> The goal is to generate secret common randomness shared by a given set  $A \subset \{0, 1, \dots, m\}$  of terminals at the largest rate possible. Thus, the resulting key must be accessible to every terminal in  $A$ . It need not be accessible to the terminals not in  $A$ , but nor is it required to be concealed from them, with the possible exception of a set  $D$  of terminals which are "wiretapped" by the eavesdropper (where  $A \cap D = \emptyset$ ). The DMC input terminal 0 may or may not belong to the set  $A$  or  $D$ .

Our emphasis is on models where all the terminals cooperate, including those that are wiretapped (if  $D \neq \emptyset$ ), in generating a secret key for the terminals in  $A$ , with secrecy being required from the eavesdropper which has access to only the public communication and the information available to the wiretapped terminals in  $D$ .

<sup>1</sup>For ease of distinction between the use of the DMC and the use of the public channel, hereafter the former will be termed "transmission" while the latter will be referred to as "communication."

A situation that is much more difficult to analyze but is practically important, arises when the eavesdropper has access to additional side information. A model for this situation assumes that the underlying DMC also has an additional output terminal that is wiretapped by the eavesdropper, whereas it is not accessible to any of the terminals cooperating in secrecy generation. This model will be referred to as the model with wiretap side information (WSI). For this model, unlike for that described earlier, we have only partial results. For the analogous problem of source models with WSI, an upper bound for secrecy capacity was given in [7], which easily extends to channel models. A sufficient condition for the tightness of this bound had also been given in [7], regrettably with a flawed proof; it remains an open question whether the condition indeed suffices. Here we shall present a more restrictive sufficiency condition for both source and channel models.

It should be mentioned that all the models considered in this paper assume the eavesdropper to be passive, i.e., unable to tamper with the communication of the legitimate terminals.

One possible operational strategy in a channel model as above is for terminal 0 to transmit an independent and identically distributed (i.i.d.) sequence of random variables (rvs) over the DMC, thereby emulating a source model. The main technical result of this paper is that for channel models with no WSI, the best possible secrecy rate can always be achieved in this manner provided that randomization is allowed at terminal 0 (Theorem 4.1). This theorem includes a single-letter characterization of the secrecy capacity in question, obtained from a dual representation of the corresponding capacity of the source model that was determined in [7]. It is expressed in two equivalent forms as a max-min and as a min-max, with the latter being apt for a proof of optimality. This proof uses familiar techniques from Shannon theory, but it is difficult. In particular, it takes recourse to two entropy inequalities (Lemmas A.2 and B.1 in Appendices A and B) that may be of independent interest.

In order to generate secrecy at an optimal rate by emulating a source model as above, in general it is not adequate for terminal 0 to merely transmit over the DMC, and it has to communicate over the public channel as well. We show, however, that a secret key can be generated at an optimal rate also in such a manner that terminal 0 sends no public message, each of the other terminals sends at most one public message (not using randomization) and that after transmission over the DMC has been completed. If  $0 \in A$ , then no rate optimality is lost if the secret key is generated at the outset by the input terminal and then transmitted over the DMC, whereupon the output terminals in  $A$  recover it securely with the help of public communication as above (Theorem 4.2).

The case  $0 \in D$  models situations when the input terminal transmits over the DMC with the objective of enabling the receiving terminals to generate common randomness (shared by some or all of them) which remains concealed even from itself. We show that in this case, there is no need for randomization at terminal 0 either. Moreover, the secrecy capacity for any channel model with no WSI, subject to the constraint that no randomization is allowed at terminal 0, coincides with the secrecy capacity for the modified model obtained upon replacing

$D$  by  $D \cup \{0\}$  whenever  $0 \notin A$  (Theorem 4.3); the secrecy capacity under this constraint is trivially zero if  $0 \in A$ .

The paper is organized as follows. Section II contains the preliminaries. In Section III, we briefly recount relevant prior results for source models from [7], and present a new result. In Section IV, we present our main results on secrecy capacities for channel models with no WSI, including necessary and sufficient conditions for positivity. Section V contains partial results on source as well as channel models with WSI. We close with a discussion in Section VI.

## II. PRELIMINARIES

In this work, we consider channel models of the following kind. Terminal 0, with finite alphabet  $\mathcal{X}_0$ , is connected to terminals  $1, \dots, m$ , with finite alphabets  $\mathcal{X}_1, \dots, \mathcal{X}_m$ , respectively, by a DMC  $W : \mathcal{X}_0 \rightarrow \mathcal{X}_1 \times \dots \times \mathcal{X}_m$ . Terminal 0 governs the input of the DMC over which it transmits a sequence of length  $n$ , while terminals  $1, \dots, m$  observe the corresponding output sequences of length  $n$ . In between consecutive symbol transmissions over the DMC, the terminals in  $\mathcal{M} = \{0, 1, \dots, m\}$  are allowed to communicate over a public noiseless channel, possibly in several rounds; in any transmission or communication by a terminal, randomization may or may not be permitted. When considering models with WSI, the underlying DMC will be taken to be  $W : \mathcal{X}_0 \rightarrow \mathcal{X}_1 \times \dots \times \mathcal{X}_m \times \mathcal{Z}$ , the additional terminal  $\mathcal{Z}$  not participating in the public communication but feeding side information to the eavesdropper.

A formal description is provided next. All rvs are assumed to take values in finite sets, even if not stated explicitly. The range of an rv denoted by an uppercase letter, will be given by the corresponding script capital unless stated otherwise. The cardinality of a finite set  $\mathcal{X}$  is denoted by  $|\mathcal{X}|$ . Logarithms are with respect to the base 2.

When randomization is permitted at terminal  $i \in \mathcal{M}$ , we shall assume that it generates at the outset an rv  $U_i$ ; the rvs  $U_0, U_1, \dots, U_m$  are mutually independent. When such randomization is forbidden, we take  $U_i = \text{a constant}$  with probability 1. Terminal 0 transmits  $n$  symbols  $X_{01}, X_{02}, \dots, X_{0n}$ , over the DMC  $W$  at time instants  $\tau_1 < \tau_2 < \dots < \tau_n$ , and every terminal  $i \in \mathcal{M} \setminus \{0\}$  observes (instantaneously) the corresponding output symbols  $X_{i1}, X_{i2}, \dots, X_{in}$ ; in a model with WSI, also the terminal  $\mathcal{Z}$  observes the output symbols  $Z_1, \dots, Z_n$ . In addition, communication among the terminals in  $\mathcal{M}$  over the public channel occurs—possibly interactively—during the time intervals  $(\tau_t, \tau_{t+1})$ , for  $t = 1, \dots, n-1$ , and immediately following  $\tau_n$ , which hereafter will be referred to simply as the intervals  $t = 1, \dots, n$ . The public communication of all the terminals in interval  $t$  is depicted collectively as  $F_t$ , and we denote  $\mathbf{F} = (F_1, \dots, F_n)$ .

In general, terminal 0 determines the  $t$ th input  $X_{0t}$  of the DMC  $W$  as a function of  $U_0$  for  $t = 1$ , and of  $(U_0, F_1, \dots, F_{t-1})$  for  $t = 2, \dots, n$ . Also, the communication of terminal  $i \in \mathcal{M}$  in the interval  $t$  is allowed to depend on  $U_i$  and the symbols  $(X_{i1}, \dots, X_{it})$ , previously generated at the outset, or observed, by terminal  $i$ , and on all previous communication. Formally, assuming without any loss of generality that the communication of the terminals in  $\mathcal{M}$  in interval  $t$  takes place sequentially in  $r$  rounds ( $r$  being arbitrary, possibly

depending on  $n$ ), this communication  $F_t$  comprises a sequence of messages

$$f_{10}, f_{11}, \dots, f_{1m}, f_{20}, f_{21}, \dots, f_{2m}, \dots, f_{r0}, f_{r1}, \dots, f_{rm}$$

as in Definition 1 below. To be exact,  $F_t$  is defined as  $F$  in Definition 1, with the role of the initial knowledge  $H_i$  of terminal  $i$  being played by  $(U_i, X_i^t, F^{t-1})$ , the knowledge of terminal  $i$  at the beginning of interval  $t$ . Here,  $X_i^t$  is shorthand for  $(X_{i1}, \dots, X_{it})$ . Also, we use the notation  $X_B = (X_i, i \in B)$ ,  $X_{Bt} = (X_{it}; i \in B)$  and  $X_B^t = (X_{B1}, \dots, X_{Bt})$  for sets  $B \subset \mathcal{M}$  and  $1 \leq t \leq n$ .

*Definition 1:* Given rvs  $H_i, i \in \mathcal{M}$ , interactive communication of the terminals in  $\mathcal{M}$ , with terminal  $i \in \mathcal{M}$  possessing initial knowledge  $H_i$ , is a sequence

$$F = (f_{10}, \dots, f_{1m}, f_{20}, \dots, f_{2m}, \dots, f_{r0}, \dots, f_{rm})$$

where  $f_{ji}$ , the message sent in round  $j$  by terminal  $i$ , is a function of  $H_i$  and of the prior communication

$$\phi_{ji} = \{f_{kl} : k < j, l \in \mathcal{M}, \text{ or } k = j, l < i\}.$$

While the description above admits complex transmission and communication protocols, it will transpire that noninteractive communication suffices for our purposes with each terminal  $i \in \mathcal{M} \setminus \{0\}$  sending one public message  $f_i = f_i(X_i^n)$  upon completion of the  $n$  transmissions over the DMC, and terminal 0 sending no public message at all; in this case  $\mathbf{F} = (f_i(X_i^n), i \in \mathcal{M} \setminus \{0\})$ .

The following concepts introduced in [7] will also be used. Given  $\epsilon > 0$ , a rv  $U$  is  $\epsilon$ -recoverable from  $V$  if  $\Pr\{U \neq f(V)\} \leq \epsilon$  for some function  $f(V)$  of  $V$ . For rvs  $K$  and  $Y$ , to be interpreted as representing a secret key and the eavesdropper's knowledge, respectively, the information-theoretic *security index* is

$$s(K; Y) = \log |\mathcal{K}| - H(K|Y).$$

Smallness of this security index is tantamount jointly to a nearly uniform distribution for  $K$  (i.e.,  $\log |\mathcal{K}| - H(K)$  is small) and to the near independence of  $K$  and  $Y$  (i.e., the mutual information  $I(K \wedge Y)$  is close to 0).

*Definition 2:* Given any set  $A \subset \mathcal{M}$  of size  $|A| \geq 2$ , a rv  $K$  constitutes an  $(\epsilon, \delta)$ -secret key  $((\epsilon, \delta)$ -SK) for the set of terminals  $A$ , achievable with  $n$  uses of the DMC  $W$ , randomization  $U_{\mathcal{M}}$ , and public communication  $\mathbf{F}$ , if  $K$  is  $\epsilon$ -recoverable from  $(U_i, X_i^n, \mathbf{F})$  for each  $i \in A$  and, in addition, it satisfies the secrecy condition

$$s(K; \mathbf{F}) \leq \delta. \tag{1}$$

An  $(\epsilon, \delta)$ -SK as above is called an  $(\epsilon, \delta)$ -private key  $((\epsilon, \delta)$ -PK) for the terminals  $A$ , private from the set of terminals  $D \subset \mathcal{M}$  with  $A \cap D = \emptyset$ , if it satisfies the stronger secrecy condition

$$s(K; U_D, X_D^n, \mathbf{F}) \leq \delta. \tag{2}$$

By definition, an  $(\epsilon, \delta)$ -SK is recoverable at the terminals in  $A$ , and is nearly uniformly distributed and effectively concealed from an eavesdropper with access to the public communication  $\mathbf{F}$ ; it need not be concealed from the terminals in  $A^c = \mathcal{M} \setminus A$ . On the other hand, an  $(\epsilon, \delta)$ -PK is effectively concealed from an eavesdropper with access—in addition to the public communication  $\mathbf{F}$ —also to a set  $D$  of “wiretapped” terminals; the latter, however, do cooperate in the secrecy generation through their public communication. This  $(\epsilon, \delta)$ -PK need not be concealed from the terminals in  $A^c \setminus D$ .

*Remark:* The security index in (2) is clearly equal to  $s(K; U_D, X_D^n, \tilde{\mathbf{F}})$  where  $\tilde{\mathbf{F}}$  denotes the public communication of the terminals in  $D^c$ . In turn, the latter security index equals  $s(K; \mathbf{F}')$  for the communication  $\mathbf{F}'$  defined by letting each terminal  $i \in D$  send the message  $(U_i, X_{i1})$  in the interval  $t = 1$ , and  $X_{it}$  in the intervals  $t = 2, \dots, n$ , with the communication  $\tilde{\mathbf{F}}$  of the terminals in  $D^c$  unchanged. Hence, when considering a PK, it can be assumed without loss of generality (w.l.o.g.) that the terminals in  $D$  reveal, as above, all the information in their possession (which, anyway, is accessible to the eavesdropper). This assumption will be made usually without explicit mention. Then the secrecy condition (2) reduces to (1), rendering the latter a proper secrecy condition also for a PK.

*Definition 3:* A number  $R$  is an achievable SK rate for a set of terminals  $A \subset \mathcal{M}$  if there exist  $(\epsilon_n, \delta_n)$ -SKs  $K^{(n)}$  achievable for  $A$  with  $n$  uses of the DMC  $W$ , suitable randomization  $U_{\mathcal{M}}$ , and public communication  $\mathbf{F}^{(n)}$ , such that

$$\begin{aligned} \epsilon_n &\rightarrow 0, & \delta_n &\rightarrow 0 \\ \text{and } \frac{1}{n} \log |\mathcal{K}^{(n)}| &\rightarrow R, & \text{ as } n &\rightarrow \infty. \end{aligned} \tag{3}$$

The largest achievable SK rate for  $A$  is the SK-capacity  $C_S(A)$ . Achievable PK rates and PK capacity  $C_P(A|D)$  are defined similarly.

*Remark:* The definitions above are analogous to those for the source models in [7], with the slight difference that there the roles of both  $\epsilon$  and  $\delta$  had been played by the same  $\epsilon$ . The formal modification here is convenient when comparing the “strong” secrecy concept [14], [2] adopted above with the earlier “weak” one [13], [1] which only requires that  $\delta_n = o(n)$ , while  $\epsilon_n$  still decays to 0. Note that our converse proofs stand under the weaker requirement, while the achievability results hold with both  $\epsilon_n$  and  $\delta_n$  decaying to 0 exponentially rapidly.

Definitions 2 and 3 are for models without WSI, the brunt of this paper. For models with WSI, the only modification required in these definitions is to strengthen the secrecy conditions (1) and (2) to  $s(K; \mathbf{F}, Z^n) \leq \delta$  and  $s(K; U_D, X_D^n, \mathbf{F}, Z^n) \leq \delta$ , respectively, where  $Z^n$  denotes the DMC output sequence at the additional output terminal  $\mathcal{Z}$  that is observed by the eavesdropper.

An obvious lower bound for  $C_S(A)$ , of course, is the capacity of the compound DMC (cf. e.g., [5]) consisting of the family of DMCs  $\{W_i : \mathcal{X}_0 \rightarrow \mathcal{X}_i, i \in A \setminus \{0\}\}$  with

$$W_i(x_i|x_0) = \sum_{x_j \in \mathcal{X}_j, j \in \{1, \dots, m\} \setminus \{i\}} W(x_1, \dots, x_m|x_0). \tag{4}$$

The capacity of this compound DMC, which is equal to  $\max_Q \min_{i \in A \setminus \{0\}} I(Q, W_i)$ , is always an achievable SK rate for the set of terminals  $A \subset \mathcal{M}$ . Terminal 0 simply forms the SK as a randomly chosen codeword from a capacity-achieving codebook for the compound DMC, which it then transmits over the DMC  $W$  for reliable decoding by the terminals in  $A \setminus \{0\}$ ; no public communication is used. In general, public communication between the terminals can serve to improve achievable SK rate, as seen in the following example.

*Example 1:* Consider the DMC  $W : \mathcal{X}_0 \rightarrow \mathcal{X}_1 \times \mathcal{X}_2$  where  $\mathcal{X}_0 = \mathcal{X}_1 = \mathcal{X}_2 = \{0, 1\}$  and

$$W(x_1, x_2 | x_0) = \frac{1}{2} 1(x_0 = x_1 + x_2 \bmod 2)$$

where  $1(\cdot)$  denotes the indicator function. Here, terminal 1 (resp., terminal 2) observes a DMC output symbol which is uniform on  $\{0, 1\}$  regardless of the input, while terminal 2 (resp., terminal 1) observes the binary sum of the input and the other output.

Consider first SK generation for  $A = \mathcal{M} = \{0, 1, 2\}$ . The lower bound for  $C_S(\{0, 1, 2\})$  provided by the capacity of the compound DMC is clearly equal to 0 since  $W_1, W_2$  given by (4) are binary symmetric channels with crossover probability 0.5. However, as seen below, 1 bit of perfect SK (i.e.,  $(\epsilon, \delta)$ -SK with  $\epsilon = \delta = 0$ ) for the terminals in  $\{0, 1, 2\}$  can be achieved with  $n = 2$  transmissions over the DMC  $W$  followed by public messages sent by terminals 1 and 2. Terminal 0 transmits  $(X_{01}, X_{02})$  as  $(0, 0)$  or  $(1, 1)$  with probabilities 0.5. Terminals 1 and 2, respectively, send public messages  $f_1 = f_1(X_{11}, X_{12}) = X_{11}$  and  $f_2 = f_2(X_{21}, X_{22}) = X_{22} = X_{02} + X_{12} \bmod 2$ . Clearly, all the terminals can perfectly recover  $K = X_{01}$ , say, from any  $X_i^2$  and the communication  $\mathbf{F} = (f_1, f_2)$  while satisfying the secrecy condition (1)

$$s(K; \mathbf{F}) = 1 - H(X_{01} | X_{11}, X_{22}) = 1 - H(X_{01}) = 0.$$

Thus,  $K = X_{01}$  is a perfect SK for the terminals 0, 1, 2, of rate 0.5. The results of the paper will show that this achievable SK rate cannot be bettered, i.e.,  $C_S(\{0, 1, 2\}) = 0.5$ .

Next, consider SK generation for  $A = \{1, 2\}$  and PK generation for  $A = \{1, 2\}$  with privacy from  $D = \{0\}$ . It is easy to see that 1 bit of perfect SK, which is also perfect PK, can be achieved with  $n = 1$  transmission over the DMC  $W$  and no public communication. Terminal 0 transmits  $X_{01} = 0$  with probability (w.p.) 1, and terminals 1 and 2 observe  $X_{11} = X_{21}$  equal to 0 or 1 with probabilities 0.5. Clearly,  $K' = X_{11}$ , of rate equal to  $H(X_{11}) = 1$ , is a perfect SK for  $A = \{1, 2\}$  as well as a perfect PK for  $A = \{1, 2\}$  with privacy from  $D = \{0\}$ . Since at most 1 bit per channel transmission can be recovered at any terminal,  $K'$  has the best rate possible, i.e.,  $C_S(\{1, 2\}) = C_P(\{1, 2\} | \{0\}) = 1$ .  $\square$

Several features of the schemes used for achieving the SK and PK capacities above will turn out to hold in general (Theorem 4.2). Specifically, these capacities are achieved with terminal 0 not communicating publicly, and with the remaining terminals communicating publicly at most once, without randomization, and only upon the completion of terminal 0's transmissions over the DMC  $W$ . Also, when  $0 \in A$ , the SK can

be formed first by terminal 0 and transmitted over the DMC with suitable encoding, upon which the output terminals in  $A$  securely recover it with the assistance of public communication. When privacy from terminal 0 is sought, the PK capacity is achieved with terminal 0 transmitting without randomization over the DMC.  $\square$

### III. SOURCE MODELS REVISITED

Consider a DMMS with generic rvs  $\mathcal{X}_{\mathcal{M}} = (X_0, X_1, \dots, X_m)$ . Given  $n$  i.i.d. repetitions  $X_{\mathcal{M}}^n$  of  $\mathcal{X}_{\mathcal{M}}$ , we assume that each terminal  $i \in \mathcal{M}$  observes the  $i$ th component  $X_i^n$  of  $X_{\mathcal{M}}^n$ , whereupon all the terminals are allowed to communicate over a public noiseless channel, possibly interactively in several rounds, with all such communication observed by all the terminals. The goal is to generate secret common randomness for a given set of terminals  $A \subset \mathcal{M}$ , perhaps with privacy from a set of terminals  $D \subset \mathcal{M}$  (disjoint from  $A$ ). The definitions of  $(\epsilon, \delta)$ -SK,  $(\epsilon, \delta)$ -PK, and SK and PK capacities  $C_S(A)$  and  $C_P(A|D)$  are similar to those for the channel models described in Section II. However, the situation here is simplified by the fact that all the communication occurs only following the observation by the terminals of their respective components  $X_i^n$  of  $X_{\mathcal{M}}^n$ , which is now an i.i.d. repetition of  $\mathcal{X}_{\mathcal{M}}$ .

The capacities  $C_S(A)$  and  $C_P(A|D)$  for source models have been determined in [7]. Slight formal differences appear therein, in that the set  $\mathcal{M}$  was  $\{1, \dots, m\}$  rather than  $\{0, 1, \dots, m\}$ , and instead of  $(\epsilon, \delta)$ -SK and  $(\epsilon, \delta)$ -PK,  $\epsilon$ -SK and  $\epsilon$ -PK were defined corresponding to  $\epsilon = \delta$ . The following theorem reproduces the main result of [7] in a modified form which is better suited for the purposes of this paper.

Throughout the paper, the following notation will be used. For  $A \subset \mathcal{M}$ , let

$$\mathcal{B}(A) = \{B \subset \mathcal{M} : B \neq \emptyset, A \not\subset B\}$$

and  $\mathcal{B}_i(A)$ ,  $i \in \mathcal{M}$ , its subset consisting of those  $B \in \mathcal{B}(A)$  that contain  $i$ . Let  $\Lambda(A)$  be the set of all collections  $\lambda = \{\lambda_B : B \in \mathcal{B}(A)\}$  of weights  $0 \leq \lambda_B \leq 1$ , satisfying

$$\sum_{B \in \mathcal{B}_i(A)} \lambda_B = 1, \quad \text{for all } i \in \mathcal{M}.$$

Similarly, for  $A \subset \mathcal{M}$  and  $D \subset A^c$ , let

$$\mathcal{B}(A|D) = \{B \subset D^c : B \neq \emptyset, A \not\subset B\}.$$

The set  $\Lambda(A|D)$  of collections of weights  $\lambda = \{\lambda_B : B \in \mathcal{B}(A|D)\}$  is defined analogously as above with the roles of  $\mathcal{B}(A)$  and  $\mathcal{M}$  now played by  $\mathcal{B}(A|D)$  and  $D^c$ , respectively.

*Remark:* For any partition  $\mathcal{P}$  of  $\mathcal{M}$ , i.e., a family of disjoint sets whose union equals  $\mathcal{M}$ , in case  $\mathcal{P} \subset \mathcal{B}(A)$ , it is clear that the collection of weights defined by  $\lambda_B = 1$  or 0 according to whether or not  $B \in \mathcal{P}$ , belongs to  $\Lambda(A)$ . In particular,  $\Lambda(A)$  is always nonempty if  $|A| \geq 2$ . Any  $\lambda \in \Lambda(A)$  can be regarded as a "fractional partition" of  $\mathcal{M}$ , and similarly,  $\lambda \in \Lambda(A|D)$  as a fractional partition of  $D^c$ . This notion of fractional partition is a special case of fractional packing that has recently been used in information theory in [11].

**Theorem 3.1:** For a source model with generic rvs  $X_{\mathcal{M}} = (X_0, X_1, \dots, X_m)$ , the SK capacity for a set of terminals  $A \subset \mathcal{M}$ , with  $|A| \geq 2$ , is

$$C_S(A) = H(X_{\mathcal{M}}) - R_{\text{CO}}(A),$$

where

$$R_{\text{CO}}(A) = \max_{\lambda \in \Lambda(A)} \sum_{B \in \mathcal{B}(A)} \lambda_B H(X_B | X_{B^c}).$$

Similarly, the PK capacity for  $A$  with privacy from a set of terminals  $D \subset A^c$  is

$$C_P(A|D) = H(X_{\mathcal{M}} | X_D) - R_{\text{CO}}(A|D),$$

where

$$R_{\text{CO}}(A|D) = \max_{\lambda \in \Lambda(A|D)} \sum_{B \in \mathcal{B}(A|D)} \lambda_B H(X_B | X_{B^c}).$$

For later reference, we note that  $R_{\text{CO}}(A)$  is the smallest rate of “communication for omniscience” (CO rate) for  $A$ , namely, the smallest number  $R$  such that, for suitable communication  $\mathbf{F}^{(n)} = (f_i(X_i^n), i \in \mathcal{M})$  of the terminals in  $\mathcal{M}$ , and  $\epsilon_n$  with

$$\epsilon_n \rightarrow 0, \quad \frac{1}{n} \log |\mathbf{F}^{(n)}| \rightarrow R,$$

$X_{\mathcal{M}}^n$  is  $\epsilon_n$ -recoverable from  $(X_j^n, \mathbf{F}^{(n)})$  for each  $j \in A$ . Similarly,  $R_{\text{CO}}(A|D)$  is the smallest achievable CO rate for  $A$  when each terminal  $i \in D$  reveals all of  $X_i^n$ . Such achievable CO rates are defined with the obvious modification of the above that the communication in  $\mathbf{F}^{(n)}$  has to satisfy  $f_i(X_i^n) = X_i^n$  for  $i \in D$ , and the rate condition applies to not all of  $\mathbf{F}^{(n)}$  but to  $\tilde{\mathbf{F}}^{(n)} = (f_i(X_i^n), i \in D^c)$ .

*Proof:* By [7, Theorems 1, 3], the first equality holds with  $R_{\text{CO}}(A)$  denoting the smallest CO rate for  $A$ . By [7, Proposition 1],  $R_{\text{CO}}(A)$  equals the minimum of  $\sum_{i=0}^m R_i$  subject to the Slepian–Wolf constraints

$$\sum_{i \in B} R_i \geq H(X_B | X_{B^c}), \quad B \in \mathcal{B}(A).$$

By the duality theorem of linear programming as stated in [7, Sec. 5], that minimum is equal to the maximum of  $\sum_{B \in \mathcal{B}(A)} \lambda_B H(X_B | X_{B^c})$ , for  $\lambda \in \Lambda(A)$ . This proves the first assertion.

The second assertion holds similarly: by [7, Theorems 2, 3], the formula for  $C_P(A|D)$  holds with  $R_{\text{CO}}(A|D)$  denoting the smallest achievable CO rate for  $A$  when the terminals  $i \in D$  reveal all of  $X_i^n$ . By [7, Proposition 1], this  $R_{\text{CO}}(A|D)$  equals the minimum of  $\sum_{i \in D^c} R_i$  subject to

$$\sum_{i \in B} R_i \geq H(X_B | X_{B^c}), \quad B \in \mathcal{B}(A|D).$$

Again, by the duality theorem of linear programming, that minimum is equal to the maximum of  $\sum_{B \in \mathcal{B}(A|D)} \lambda_B H(X_B | X_{B^c})$  for  $\lambda \in \Lambda(A|D)$ .  $\square$

*Remark:* The maximizing  $\lambda$  in Theorem 3.1 need not be unique in either case. As discussed in [7, Sec. V], it can be

chosen in such a manner that  $\lambda_B > 0$  holds only for a collection of sets  $B$  whose incidence vectors are linearly independent. This fact facilitates the actual computation of SK and PK capacities but will not be used in this paper.

The next theorem includes a new result that the SK or PK capacity can be achieved with the key being generated at any chosen terminal  $i \in A$ . In addition to being of independent interest, it will play a role in the achievability of SK and PK capacities for certain source and channel models with WSI, in Section V. The proof of this result uses a technical lemma which is provided first.

**Lemma 3.1:** If n rv  $U$  is  $\epsilon$ -recoverable from a rv  $V$ , then for i.i.d. repetitions of  $(U, V)$  it holds that  $U^n$  is  $\epsilon_n$ -recoverable from  $(V^n, g(U^n))$  for a suitable function  $g : \mathcal{U}^n \rightarrow \{1, \dots, 2^{n\eta}\}$ , where  $\eta = \epsilon \log |\mathcal{U}| + h(\epsilon)$ , and  $\epsilon_n$  vanishes to 0 exponentially as  $n \rightarrow \infty$ .

*Proof:* The  $\epsilon$ -recoverability assumption implies by Fano’s inequality that  $H(U|f(V)) < \eta$  for some function  $f(V)$  of  $V$ . Hence, by the Slepian–Wolf theorem,  $U^n$  can be recovered from  $f(V_1), \dots, f(V_n)$  and a code for  $U^n$  of rate at most  $\eta$ , with probability of error approaching 0 exponentially rapidly.  $\square$

**Theorem 3.2:** For a source model, the SK and PK capacities can be achieved with noninteractive communication without randomization, with each terminal  $i \in \mathcal{M}$  sending a single message  $f_i(X_i^n)$  which equals  $X_i^n$  if  $i \in D$ . Further, the SK or PK capacity can be achieved with the key generated at any particular terminal  $i \in A$ , obviously of the public communication. Precisely, for any  $\zeta > 0$  and observation length  $n$ , for each  $i \in A$  there exists  $K^{(n)} = K^{(n)}(X_i^n)$  with  $\frac{1}{n} \log |\mathcal{K}^{(n)}|$  exceeding  $C_S(A) - \zeta$  or  $C_P(A|D) - \zeta$ , respectively, and  $\mathbf{F}^{(n)} = \{f_j(X_j^n), j \in \mathcal{M}\}$  (with  $f_j(X_j^n) = X_j^n$  if  $j \in D$ ), such that  $X_i^n$  is  $\epsilon_n$ -recoverable from  $X_j^n$  and  $\mathbf{F}^{(n)}$  for each  $j \in A$  different from  $i$ , where both  $\epsilon_n$  and the security index  $s(K^{(n)}; \mathbf{F}^{(n)})$  vanish exponentially as  $n \rightarrow \infty$ .

*Proof:* The first assertion is proved in [7, Theorems 1–3]. The second assertion, which is new, is proved next. We note that the following proof gives the first assertion as well.

Consider first the case of SK capacity. Fix  $i \in A$ , and some  $\epsilon > 0$  which will be specified later. According to the passage after Theorem 3.1, for any  $\zeta > 0$  and  $k$  sufficiently large, there exists communication  $\mathbf{F}^{(k)} = \{f_j(X_j^k), j \in \mathcal{M}\}$  with  $\log \|\mathbf{F}^{(k)}\| < k \left( R_{\text{CO}}(A) + \frac{\zeta}{2} \right)$  such that  $X_{\mathcal{M}}^k$  is  $\epsilon$ -recoverable from  $X_j^k$  and  $\mathbf{F}^{(k)}$  for each  $j \in A$ . The proof will rely on this  $\mathbf{F}^{(k)}$ . Note that

$$\begin{aligned} & H(X_i^k | \mathbf{F}^{(k)}) \\ &= H(X_{\mathcal{M}}^k | \mathbf{F}^{(k)}) - H(X_{\mathcal{M}}^k | X_i^k, \mathbf{F}^{(k)}) \\ &= kH(X_{\mathcal{M}}) - H(\mathbf{F}^{(k)}) - H(X_{\mathcal{M}}^k | X_i^k, \mathbf{F}^{(k)}) \\ &\geq kH(X_{\mathcal{M}}) - H(\mathbf{F}^{(k)}) - \epsilon k \sum_{j \in \mathcal{M}} \log |\mathcal{X}_j| - h(\epsilon) \\ &\geq k \left( C_S(A) - \frac{\zeta}{2} - \epsilon \sum_{j \in \mathcal{M}} \log |\mathcal{X}_j| \right) - h(\epsilon). \end{aligned} \quad (5)$$

Here, the first inequality is a consequence of the  $\epsilon$ -recoverability of  $X_{\mathcal{M}}^k$  from  $(X_i^k, \mathbf{F}^{(k)})$ , by Fano's inequality, and the second inequality follows from

$$H(\mathbf{F}^{(k)}) \leq \log |\mathbf{F}^{(k)}| < k \left( R_{\text{CO}}(A) + \frac{\zeta}{2} \right)$$

since  $H(X_{\mathcal{M}}) - R_{\text{CO}}(A) = C_S(A)$  by Theorem 3.1.

It suffices to prove the assertion for block lengths equal to integer multiples of  $k$ . To this end, consider  $n$  i.i.d. repetitions of  $(X_{\mathcal{M}}^k, \mathbf{F}^{(k)})$ , i.e.,  $(X_{\mathcal{M}}^{kn}, \mathbf{F}^{(k)n})$ . Then, for each  $j \in A \setminus \{i\}$ , Lemma 3.1 above applied to  $U = X_i^k, V = (\mathbf{F}^{(k)}, X_j^k)$  gives that  $X_i^{kn}$  is  $\epsilon_n$ -recoverable from  $(X_j^{kn}, \mathbf{F}^{(k)n}, g_j(X_i^{kn}))$ , for a suitable function  $g_j : \mathcal{X}_i^{kn} \rightarrow \{1, \dots, 2^{n\eta}\}$  where  $\eta = \epsilon k \log |\mathcal{X}_i| + h(\epsilon)$ , and  $\epsilon_n$  decays to 0 exponentially rapidly as  $n \rightarrow \infty$ . It follows, setting  $g(X_i^{kn}) = \{g_j(X_i^{kn}), j \in A \setminus \{i\}\}$  that  $\Phi^{(kn)} = \{\mathbf{F}^{(k)n}, g\}$  satisfies the recoverability assertion of the theorem on  $\mathbf{F}^{(n)}$ , with  $n$  replaced by  $kn$ .

It remains to show that there exists a function  $K^{(kn)}$  of  $X_i^{kn}$  that satisfies the assertions on  $K^{(n)}$ , again with  $n$  replaced by  $kn$  if  $\Phi^{(kn)}$  plays the role of  $\mathbf{F}^{(n)}$ . To this end, we apply [7, Lemma B3] with  $U = X_i^k, V = \mathbf{F}^{(k)}$ , and  $R = (|A| - 1)\eta$ , where  $\eta$  and the function  $g$  of  $U^n = X_i^{kn}$  are as above. With this choice, by (5) and the definition of  $\eta$

$$\begin{aligned} H(U|V) - R &= H(X_i^k | \mathbf{F}^{(k)}) - (|A| - 1)\eta \\ &\geq k \left[ C_S(A) - \frac{\zeta}{2} \right. \\ &\quad \left. - \epsilon \left( |A| \log |\mathcal{X}_i| + \sum_{j \in \mathcal{M}} \log |\mathcal{X}_j| \right) \right] \\ &\quad - |A|h(\epsilon) \\ &> k(C_S(A) - \zeta) \end{aligned}$$

if  $\epsilon > 0$  (unspecified until now) has been chosen sufficiently small. Hence [7, Lemma B3] gives that there exists  $K^{(kn)} = K^{(kn)}(X_i^{kn})$  with  $\log |\mathcal{K}^{(kn)}| \geq kn(C_S(A) - \zeta)$  and  $s(K^{(kn)}; \Phi^{(kn)})$  vanishing to 0 exponentially rapidly as  $n \rightarrow \infty$ . This completes the proof for SK capacity.

For PK capacity, the following minor modifications are needed: the communication  $\mathbf{F}^{(k)}$  is now chosen as  $\tilde{\mathbf{F}}^{(k)} = (X_D^n, \tilde{\mathbf{F}}^{(k)})$ , with  $\tilde{\mathbf{F}}^{(k)} = (f_j(X_j^k), j \in D^c)$  satisfying

$$\log \|\tilde{\mathbf{F}}^{(k)}\| < k \left( R_{\text{CO}}(A|D) + \frac{\zeta}{2} \right).$$

Then the bound in (5) with  $C_P(A|D)$  replacing  $C_S(A)$  is obtained noting that  $H(\mathbf{F}^{(k)})$  in the third line of (5) is bounded as

$$\begin{aligned} H(\mathbf{F}^{(k)}) &\leq H(X_D^k) + H(\tilde{\mathbf{F}}^{(k)}) \\ &< kH(X_D) + k \left( R_{\text{CO}}(A|D) + \frac{\zeta}{2} \right) \end{aligned}$$

and using that

$$H(X_{\mathcal{M}}) - H(X_D) - R_{\text{CO}}(A|D) = C_P(A|D)$$

by Theorem 3.1. In the final step of the proof, the security index  $s(K^{(kn)}; \Phi^{(kn)})$  is the apt one also in the PK case, since the communication  $\Phi^{(kn)}$  contains all of  $X_j^{kn}, j \in D$ , by construction.  $\square$

#### IV. MAIN RESULTS

Our main results are single-letter characterizations of SK and PK capacities for channel models with no WSI.

Given a channel model described in Section II, for a probability mass function (pmf)  $Q$  on  $\mathcal{X}_0$  and  $\lambda \in \Lambda(A)$  (resp.,  $\lambda \in \Lambda(A|D)$ ), denote

$$G_A(Q, \lambda) = H(X_{\mathcal{M}}) - \sum_{B \in \mathcal{B}(A)} \lambda_B H(X_B | X_{B^c}) \quad (6)$$

and

$$G_{A|D}(Q, \lambda) = H(X_{\mathcal{M}} | X_D) - \sum_{B \in \mathcal{B}(A|D)} \lambda_B H(X_B | X_{B^c}) \quad (7)$$

where  $X_{\mathcal{M}} = (X_0, X_1, \dots, X_m)$  is a collection of rvs with  $P_{X_0} = Q$  and  $P_{X_1, \dots, X_m | X_0} = W$ .

*Theorem 4.1:* The SK capacity  $C_S(A)$  of the channel model for a set of terminals  $A \subset \mathcal{M}$ , and the PK capacity  $C_P(A|D)$  for  $A$  with privacy from a set of terminals  $D \subset A^c$ , are, respectively, equal to the maxima over  $Q$  of the SK and PK capacities of the corresponding emulated source models. Specifically

$$\begin{aligned} C_S(A) &= \max_Q \min_{\lambda \in \Lambda(A)} G_A(Q, \lambda) \\ &= \min_{\lambda \in \Lambda(A)} \max_Q G_A(Q, \lambda) \end{aligned} \quad (8)$$

and

$$\begin{aligned} C_P(A|D) &= \max_Q \min_{\lambda \in \Lambda(A|D)} G_{A|D}(Q, \lambda) \\ &= \min_{\lambda \in \Lambda(A|D)} \max_Q G_{A|D}(Q, \lambda). \end{aligned} \quad (9)$$

*Proof:* The proof is given in two steps: (i) the equality of the max-min and min-max expressions in (8), (9), and the achievability part; and (ii) the converse, which is the hard part.

(i) Both  $G_A(Q, \lambda)$  and  $G_{A|D}(Q, \lambda)$  are continuous functions of  $Q$  and  $\lambda$ , defined over convex compact sets, concave in  $Q$  by Lemma A.1 in Appendix A, and affine in  $\lambda$ . Hence, the claimed equalities in (8), (9), follow from the minimax theorem (cf., e.g., [10]).

The achievability part of Theorem 4.1 is obtained by observing that terminal 0 can emulate a source model by transmitting i.i.d. rvs with pmf  $P_{X_0} = Q$  over the DMC  $W$ . By Theorem 3.1, the SK capacity of the emulated source model is equal to  $\min_{\lambda \in \Lambda(A)} G_A(Q, \lambda)$ , and, therefore, the maximum over  $Q$  of that minimum is an achievable SK-rate for the channel model. Similarly, from Theorem 3.1, the maximum over  $Q$  of  $\min_{\lambda \in \Lambda(A|D)} G_{A|D}(Q, \lambda)$  is an achievable PK-rate for the channel model.

(ii) The converse proof for PK capacity is presented. The analogous result for SK capacity then follows as the special case  $D = \emptyset$ .

Suppose that the rv  $K^{(n)}$  represents an  $(\epsilon_n, \delta_n)$ -PK for  $A$  with privacy from  $D \subset A^c$ , achievable with randomization  $U_{\mathcal{M}}$  and public communication  $\mathbf{F}^{(n)}$ , where  $\delta_n = o(n)$  and  $\epsilon_n \rightarrow 0$ ; see Definition 3 and the remark following it. Supposing w.l.o.g. that  $\mathbf{F}^{(n)} = (U_D, X_D^n, \tilde{\mathbf{F}}^{(n)})$ , where  $\tilde{\mathbf{F}}^{(n)}$  comprises the communication of all the terminals  $i \in D^c$  (see the remark preceding Definition 3), the secrecy condition (2) is

$$\log |\mathcal{K}^{(n)}| - H(K^{(n)} | U_D, X_D^n, \tilde{\mathbf{F}}^{(n)}) < \delta_n.$$

Thus, using the Corollary of Lemma A.2 in Appendix A with  $(U_i, X_i^n)$  in the role of  $X_i, i \in \mathcal{M}$ , and  $\tilde{\mathbf{F}}^{(n)}$  in the role of  $Y$ , we get that for every  $\lambda = \{\lambda_B : B \in \mathcal{B}(A|D)\} \in \Lambda(A|D)$

$$\begin{aligned} \log |\mathcal{K}^{(n)}| &\leq H(U_{\mathcal{M}}, X_{\mathcal{M}}^n | U_D, X_D^n, \tilde{\mathbf{F}}^{(n)}) \\ &\quad - \sum_{B \in \mathcal{B}(A|D)} \lambda_B H(U_B, X_B^n | U_{B^c}, X_{B^c}^n, \tilde{\mathbf{F}}^{(n)}) \\ &\quad + (m+2)\nu_n + \delta_n \end{aligned} \quad (10)$$

where  $\nu_n = \epsilon_n \log |\mathcal{K}^{(n)}| + h(\epsilon_n)$ . Rearranging terms, we get

$$\begin{aligned} &\frac{1}{n} \log |\mathcal{K}^{(n)}| \\ &\leq \frac{\alpha_n}{n} \left[ \left\{ H(U_{\mathcal{M}}, X_{\mathcal{M}}^n | U_D, X_D^n, \tilde{\mathbf{F}}^{(n)}) \right. \right. \\ &\quad \left. \left. - \sum_{B \in \mathcal{B}(A|D)} \lambda_B H(U_B, X_B^n | U_{B^c}, X_{B^c}^n, \tilde{\mathbf{F}}^{(n)}) \right\} \right] \\ &\quad + \beta_n \end{aligned} \quad (11)$$

where

$$\begin{aligned} \alpha_n &= \frac{1}{1 - (m+2)\epsilon_n} \rightarrow 1 \quad \text{and} \\ \beta_n &= \frac{(m+2)h(\epsilon_n) + \delta_n}{n(1 - (m+2)\epsilon_n)} \rightarrow 0, \quad \text{as } n \rightarrow \infty. \end{aligned}$$

A key ingredient of the proof is to show that the expression within  $[\cdot \cdot \cdot]$  in (11) is bounded above by

$$\begin{aligned} &\frac{1}{n} \sum_{t=1}^n \left( H(X_{\mathcal{M}t} | X_{D^c t}) - \sum_{B \in \mathcal{B}(A|D)} \lambda_B H(X_{Bt} | X_{B^c t}) \right) \\ &= \frac{1}{n} \sum_{t=1}^n G_{A|D}(P_{X_{0t}}, \lambda) \end{aligned} \quad (12)$$

(see (7)). Our proof of this bound involves a rather tedious manipulation of information quantities, and therefore is relegated to Appendix B. Using the mentioned bound, (11) implies that

$$\limsup_n \frac{1}{n} \log |\mathcal{K}^{(n)}| \leq \max_Q G_{A|D}(Q, \lambda) \quad (13)$$

for every

$$\lambda = \{\lambda_B : B \in \mathcal{B}(A|D)\} \in \Lambda(A|D).$$

Hence,  $\min_{\lambda \in \Lambda(A|D)} \max_Q G_{A|D}(Q, \lambda)$  is an upper bound for achievable PK rates.  $\square$

*Theorem 4.2:* The secrecy capacities  $C_S(A)$  and  $C_P(A|D)$  can be achieved with all public communication occurring after completion of transmission over the DMC  $W$  and terminal 0 not communicating over the public channel, and each terminal in  $\mathcal{M} \setminus \{0\}$  communicating publicly only once, without interaction or randomization. If  $0 \in A$ , then, in addition, the SK or PK can be generated at terminal 0 at the outset as a uniformly distributed rv, and then transmitted over the DMC by means of a suitable randomized encoding. If  $0 \in D$ , the PK capacity can be achieved with terminal 0 transmitting a deterministic sequence over the DMC.

*Remark:* The last assertion of Theorem 4.2 cannot be strengthened to terminal 0 transmitting a constant sequence; see Example 2 below.

*Proof:* As shown in Theorem 3.2, for a source model the SK and PK capacities can be attained in such a way that each terminal  $i \in \mathcal{M}$  sends a single public message  $F_i = f_i(X_i^n)$ , where in the PK case,  $f_i(X_i^n) = X_i^n$  if  $i \in D$ . Moreover, if  $0 \in A$ , the key can be generated at terminal 0 oblivious of the public communication, as a function of  $X_0^n$ . It follows by step (i) of the proof of Theorem 4.1 that the SK and PK capacities for the channel model can be achieved with only terminal 0 randomizing (transmitting over the DMC  $W$  an i.i.d. sequence  $X_0^n$  that can be identified with the randomization  $U_0$ ) and with each terminal sending a single public message as above.

To show that terminal 0 need not actually send any public message, we concentrate on the PK case, as in part (ii) of the proof of Theorem 4.1. Consider an  $(\epsilon_n, \delta_n)$ -PK achieved with randomization  $U_0$  at terminal 0 and public communication  $\mathbf{F} = (f_0, f_1, \dots, f_m)$  as above, with  $\epsilon_n \rightarrow 0, \delta_n \rightarrow 0$  as  $n \rightarrow \infty$ . Thus,  $K$  is  $\epsilon_n$ -recoverable from  $(X_i^n, \mathbf{F})$  if  $i \in A$ , i.e.,

$$\Pr\{K \neq K_i(X_i^n, \mathbf{F})\} < \epsilon_n, \quad i \in A$$

and  $K$  has security index

$$s(K; \mathbf{F}) = \log |\mathcal{K}| - H(K | \mathbf{F}) < \delta_n.$$

In particular

$$\begin{aligned} &\frac{1}{\epsilon_n} \sum_{i \in A} \Pr\{K \neq K_i(X_i^n, \mathbf{F})\} + \frac{1}{\delta_n} (\log |\mathcal{K}| - H(K | \mathbf{F})) \\ &< |A| + 1. \end{aligned}$$

By the expressions at the top of the next page, it follows that for some possible value  $j$  of  $f_0$

$$\begin{aligned} &\frac{1}{\epsilon_n} \sum_{i \in A} \Pr\{K \neq K_i(X_i^n, \mathbf{F}) | f_0 = j\} \\ &\quad + \frac{1}{\delta_n} (\log |\mathcal{K}| - H(K | f_0 = j, f_1, \dots, f_m)) < |A| + 1. \end{aligned}$$

This shows that if terminal 0 changes the pmf of  $X_0^n$  to a (no longer i.i.d.) pmf which is the conditional pmf of the i.i.d. sequence  $X_0^n$  conditioned on  $f_0 = j$ , the same public communication as above renders  $K$  an  $((|A| + 1)\epsilon_n, (|A| + 1)\delta_n)$ -achievable PK. For this modified non-i.i.d.  $X_0^n$ , the public message

$$\Pr\{K \neq K_i(X_i^n, \mathbf{F})\} = \sum_j \Pr\{f_0 = j\} \Pr\{K \neq K_i(X_i^n, \mathbf{F}) | f_0 = j\}$$

and

$$H(K|\mathbf{F}) = \sum_j \Pr\{f_0 = j\} H(K|f_0 = j, f_1, \dots, f_m)$$

$f_0$  becomes the constant  $j$  which means, in effect, no public communication from terminal 0. If  $0 \in A$ , then on account of the cited consequence of Theorem 3.2, the conclusion regarding achievability with no public communication from terminal 0 holds for  $K$  equal to a function of  $X_0^n$  where  $X_0^n$  is non-i.i.d., in general. In order to generate such  $K$  and  $X_0^n$  (with given joint pmf), we can generate  $K$  at the outset and then  $X_0^n$  by randomized encoding as a function of  $K$  and an auxiliary randomizing rv. By Pinsker's inequality, the variation distance of the pmf of  $K$  from the uniform pmf on  $\mathcal{K}$  is exponentially close to 0 as  $n \rightarrow \infty$  since so is the corresponding I-divergence close to  $\log |\mathcal{K}| - H(K)$ . Hence,  $K$  can be obtained by randomized encoding from an rv  $\tilde{K}$  which is uniformly distributed on  $\mathcal{K}$  such that the probability that  $K \neq \tilde{K}$  is exponentially close to 0. Then this  $\tilde{K}$  is an SK or PK with the asserted properties.

If  $0 \in D$ , then the previous argument applies with  $f_0 = X_0^n$ , and changing the i.i.d. input pmf of the DMC  $W$  to the conditional pmf with the condition  $f_0 = j$ , means that the i.i.d. input sequence is replaced by a deterministic sequence.

This completes the proof.  $\square$

**Theorem 4.3:** If no randomization is allowed at terminal 0, then for a set of terminals  $A \subset \mathcal{M} \setminus \{0\}$ , the SK capacity equals  $C_P(A|\{0\})$ , and the PK capacity with privacy from  $D \subset A^c$  equals  $C_P(A|D \cup \{0\})$ . Both the capacities can be achieved with terminal 0 sending a deterministic sequence over the DMC  $W$  and not communicating over the public channel, and the remaining terminals using the public channel as in Theorem 4.2.

**Remark:** In Theorem 4.3, the case  $0 \in A$  has been excluded as then, with randomization not allowed at terminal 0, the capacities in question are zero. Indeed, since the secret key must be  $\epsilon$ -recoverable from  $(U_0, \mathbf{F})$ , in the case  $U_0 = \text{constant}$ , it cannot be nearly independent of  $\mathbf{F}$  while having a positive rate.

**Proof:** Suppose that no randomization is allowed at terminal 0, and let  $K$  be an  $(\epsilon, \delta)$ -SK for a set of terminals  $A \subset \mathcal{M} \setminus \{0\}$  achieved with public communication  $\mathbf{F}$ . The previous results do not directly justify the assumption that this communication is noninteractive or that it does not use randomization. Nevertheless, since  $U_0 = \text{constant}$  by hypothesis,  $X_0^n$  must be uniquely determined by the communication  $\mathbf{F}$ , and therefore the security indices  $s(K; \mathbf{F})$  and  $s(K; \mathbf{F}, U_D, X_D^n)$  with  $D = \{0\}$  are equal. Hence,  $K$  is also an  $\epsilon$ -PK for  $A$  with privacy from terminal 0, and the SK capacity in question does not exceed  $C_P(A|\{0\})$ . On the other hand, the PK capacity  $C_P(A|\{0\})$  can be achieved with terminal 0 transmitting a deterministic sequence over the DMC  $W$  by Theorem 4.2, and hence the previous SK capacity is no smaller than  $C_P(A|\{0\})$ .

A similar argument yields that the PK capacity for  $A \subset \mathcal{M} \setminus \{0\}$  with privacy from  $D \subset A^c$  equals  $C_P(A|D \cup \{0\})$ .

Both the capacities in question can be achieved with terminal 0 sending a deterministic sequence over the DMC  $W$ , since  $C_P(A|\{0\})$  and  $C_P(A|D \cup \{0\})$  can be attained likewise by Theorem 4.2.

The remaining assertions also follow by Theorem 4.2.  $\square$

The next theorem provides necessary and sufficient conditions for the positivity of the SK capacity  $C_S(A)$  and the PK capacity  $C_P(A|\{0\})$ . It is not difficult to give a similar condition for the positivity of  $C_P(A|D)$  when  $A$  and  $D$  are arbitrary (disjoint) subsets of  $\mathcal{M}$ ; however, we omit this as the case  $D = \{0\}$  appears to be of main interest, and the positivity condition for a general  $D$  is rather ungainly.

For subsets  $A$  and  $B$  of  $\mathcal{M}$ ,  $B$  is said to *split*  $A$  if both  $A \cap B$  and  $A \cap B^c$  are nonempty. If  $B$  is a nonempty subset of  $\mathcal{M} \setminus \{0\}$ , the  $B$ -component of the DMC  $W : \mathcal{X}_0 \rightarrow \mathcal{X}_1 \times \dots \times \mathcal{X}_m$  is the DMC  $W_B : \mathcal{X}_0 \rightarrow \prod_{i \in B} \mathcal{X}_i$  defined by

$$W_B(\{x_i, i \in B\} | x_0) = \sum_{x_j \in \mathcal{X}_j, j \in B^c \setminus \{0\}} W(x_1, \dots, x_m | x_0).$$

This  $W_B$  is said to be an *independent component* of  $W$  if either  $B = \mathcal{M} \setminus \{0\}$  (thus,  $W_B = W$ ) or else

$$\begin{aligned} & W(x_1, \dots, x_m | x_0) \\ &= W_B(\{x_i, i \in B\} | x_0) W_{B^c \setminus \{0\}}(\{x_j, j \in B^c \setminus \{0\}\} | x_0) \end{aligned}$$

for every  $(x_0, x_1, \dots, x_m)$  in  $\prod_{i=0}^m \mathcal{X}_i$ .

**Theorem 4.4:**

(i) For  $A \subset \mathcal{M}$ , the SK capacity  $C_S(A)$  is positive iff for each  $B \subset \mathcal{M} \setminus \{0\}$  that splits  $A$ , the DMC  $W_B$  either has positive Shannon capacity or is not an independent component of  $W$ .

(ii) For  $A \subset \mathcal{M} \setminus \{0\}$ , the PK capacity  $C_P(A|\{0\})$  is positive iff for no  $B \subset \mathcal{M} \setminus \{0\}$  that splits  $A$  is  $W_B$  an independent component of  $W$ .

**Proof:** By Theorem 4.1, the SK and PK capacities of the channel model are obtained by maximizing the corresponding SK and PK capacities of the emulated source models with generic rvs  $X_{\mathcal{M}} = (X_0, X_1, \dots, X_m)$  satisfying  $P_{X_1, \dots, X_m | X_0} = W$ , with respect to  $P_{X_0} = Q$ . By [7, Theorem 5], the SK capacity of the source model is positive iff  $I(X_B \wedge X_{B^c}) > 0$  for every  $B$  that splits  $A$ . For the channel model, owing to symmetry, consideration can be restricted to  $B \subset \mathcal{M} \setminus \{0\}$ . Then,  $I(X_B \wedge X_{B^c})$  equals  $I(X_B \wedge X_0) + I(X_B \wedge X_{B^c \setminus \{0\}} | X_0)$  (or  $I(X_{\mathcal{M} \setminus \{0\}} \wedge X_0)$  if  $B = \mathcal{M} \setminus \{0\}$ , which splits  $A$  iff  $0 \in A$ ), and hence this mutual information is positive for some  $P_{X_0} = Q$  iff either of the two conditions in part (i) holds.

Also, by [7, Theorem 5], the source model with  $D = \{0\}$  has positive PK capacity iff  $I(X_B \wedge X_{B^c} | X_0) > 0$  for every  $B \subset \mathcal{M} \setminus \{0\}$  that splits  $A$ . Clearly, this positivity holds for

some  $P_{X_0} = Q$  iff  $W_B$  is not an independent component of  $W$ .  $\square$

*Example 2:* Consider the DMC  $W : \mathcal{X}_0 \rightarrow \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3$  where  $\mathcal{X}_0 = \mathcal{X}_1 = \mathcal{X}_2 = \mathcal{X}_3 = \{0, 1\}$  and

$$W(x_1, x_2, x_3|x_0) = \frac{1}{4}(1(x_0 = 0, x_3 = x_1)) + \frac{1}{4}(1(x_0 = 1, x_3 = x_2)).$$

Here, terminals 1 and 2 observe DMC outputs which are independent and uniform on  $\{0, 1\}$  regardless of the DMC input, while the observation of terminal 3 coincides with that of terminal 1 or terminal 2 according to whether  $x_0 = 0$  or  $x_0 = 1$ .

Consider PK generation for  $A = \{1, 2, 3\}$  with privacy from  $D = \{0\}$ . Letting  $Q = (1 - q, q)$ , to compute the PK capacity given by Theorem 4.1, note first that

$$H(X_{\{0,1,2,3\}}|X_{\{0\}}) = H(X_1, X_2|X_0) + H(X_3|X_0, X_1, X_2) = H(X_1) + H(X_2) = 2$$

and that  $\mathcal{B}(\{1, 2, 3\}|\{0\})$  consists of all nonempty proper subsets  $B$  of  $\{1, 2, 3\}$ ; further,  $H(X_B|X_{B^c})$  equals  $q$ ,  $1 - q$ , and  $0$  for  $B = \{1\}$ ,  $\{2\}$ , and  $\{3\}$ , respectively, and it equals  $1$  for  $B = \{1, 2\}$ ,  $\{1, 3\}$ , and  $\{2, 3\}$ . By Theorem 4.1

$$C_P(\{1, 2, 3\}|\{0\}) = 2 - \min_{0 \leq q \leq 1} \max_{\lambda} \left( \begin{matrix} \lambda_{\{1\}}q + \lambda_{\{2\}}(1 - q) \\ + \lambda_{\{1,2\}} + \lambda_{\{1,3\}} + \lambda_{\{2,3\}} \end{matrix} \right)$$

with the maximum taken under the constraints

$$\begin{aligned} \lambda_{\{1\}} + \lambda_{\{1,2\}} + \lambda_{\{1,3\}} &= 1 \\ \lambda_{\{2\}} + \lambda_{\{1,2\}} + \lambda_{\{2,3\}} &= 1 \\ \lambda_{\{3\}} + \lambda_{\{1,3\}} + \lambda_{\{2,3\}} &= 1. \end{aligned}$$

It can be seen that a saddle point exists above with  $q = 0.5$  and

$$\lambda_{\{1\}} = \lambda_{\{2\}} = \lambda_{\{1,3\}} = \lambda_{\{2,3\}} = 0.5, \lambda_{\{3\}} = \lambda_{\{1,2\}} = 0.$$

Thus,  $C_P(\{1, 2, 3\}|\{0\}) = 0.5$ .

The following scheme achieves 1 bit of perfect PK for  $A = \{1, 2, 3\}$  with privacy from  $D = \{0\}$  using  $n = 2$  transmissions over the DMC  $W$  followed by a public message from terminal 3. Terminal 0 transmits  $(X_{01}, X_{02}) = (0, 1)$ . Terminal 3 sends the public message  $f_3 = f_3(X_{31}, X_{32}) = X_{31} + X_{32} \bmod 2 = X_{11} + X_{22} \bmod 2$ . Clearly, terminals  $i = 1, 2, 3$  can perfectly recover  $K = X_{22}$  from  $X_i^2$  and  $f_3$ , while satisfying the secrecy condition (2)

$$\begin{aligned} s(K; X_{01}, X_{02}, f_3) &= 1 - H(X_{22}|X_{11} + X_{22} \bmod 2) \\ &= 1 - H(X_{22}) = 0. \end{aligned}$$

Note that this scheme is in concordance with the assertions of Theorem 4.2 above. In particular, with  $D = \{0\}$ , the PK capacity is achieved with terminal 0 transmitting a deterministic sequence over the DMC. This deterministic sequence, however, cannot be a constant sequence; in fact, the latter would allow no

secrecy generation at all for the set of terminals  $A = \{1, 2, 3\}$ .  $\square$

*Example 3:* Consider the DMC  $W : \mathcal{X}_0 \rightarrow \mathcal{X}_1 \times \dots \times \mathcal{X}_m$  with

$$W(x_1, \dots, x_m|x_0) = \prod_{i=1}^m W_i(x_i|x_0)$$

so that  $W$  consists of independent components  $W_i, i = 1, \dots, m$ . We shall compute the PK capacity  $C_P(A|D)$  with  $0 \notin D$  (otherwise, clearly  $C_P(A|D) = 0$ ).

By the conditional independence of the  $X_i$ 's given  $X_0$ , observe that if  $0 \notin B$

$$\begin{aligned} H(X_B|X_{B^c}) &= H(X_B|X_0, X_{B^c \setminus \{0\}}) \\ &= H(X_B|X_0) = \sum_{i \in B} H(X_i|X_0) \end{aligned}$$

and, if  $0 \in B$

$$\begin{aligned} H(X_B|X_{B^c}) &= H(X_0, X_{B \setminus \{0\}}|X_{B^c}) \\ &= H(X_0|X_{B^c}) + H(X_{B \setminus \{0\}}|X_0, X_{B^c}) \\ &= H(X_0|X_{B^c}) + H(X_{B \setminus \{0\}}|X_0). \end{aligned}$$

Then, for  $\lambda \in \Lambda(A|D)$

$$\begin{aligned} &\sum_{B \in \mathcal{B}(A|D)} \lambda_B H(X_B|X_{B^c}) \\ &= \sum_{B \notin \mathcal{B}_0(A|D)} \lambda_B \sum_{i \in B} H(X_i|X_0) \\ &\quad + \sum_{B \in \mathcal{B}_0(A|D)} \lambda_B \sum_{i \in B \setminus \{0\}} H(X_i|X_0) \\ &\quad + \sum_{B \in \mathcal{B}_0(A|D)} \lambda_B H(X_0|X_{B^c}) \\ &= \sum_{i \in D^c \setminus \{0\}} \left( \sum_{B \in \mathcal{B}_i(A|D)} \lambda_B \right) H(X_i|X_0) \\ &\quad + \sum_{B \in \mathcal{B}_0(A|D)} \lambda_B H(X_0|X_{B^c}). \end{aligned}$$

Here, the last term is bounded above by

$$\left( \sum_{B \in \mathcal{B}_0(A|D)} \lambda_B \right) \max_{i \in A} H(X_0|X_i, X_D)$$

noting that  $D \subset B^c$  for  $B \in \mathcal{B}_0(A|D)$ . Using the definition of  $\Lambda(A|D)$ , it follows that

$$\begin{aligned} &\sum_{B \in \mathcal{B}(A|D)} \lambda_B H(X_B|X_{B^c}) \\ &\leq H(X_{D^c \setminus \{0\}}|X_0) + \max_{i \in A} H(X_0|X_i, X_D) \end{aligned}$$

and equality is achieved if  $\lambda$  is given by  $\lambda_B = 1$  for  $B = \{i^*\}$  and  $B = D^c \setminus \{i^*\}$ , and  $\lambda_B = 0$  otherwise, where  $i^* \in A$  maximizes  $H(X_0|X_i, X_D)$ . This proves that

$$\begin{aligned} &\min_{\lambda \in \Lambda(A|D)} G_{A|D}(Q, \lambda) \\ &= H(X_{\mathcal{M}}|X_D) - H(X_{D^c \setminus \{0\}}|X_0) \\ &\quad - \max_{i \in A} H(X_0|X_i, X_D) \\ &= H(X_0|X_D) - \max_{i \in A} H(X_0|X_i, X_D) \\ &= \min_{i \in A} I(X_0 \wedge X_i|X_D). \end{aligned}$$

Thus, by Theorem 4.1

$$C_P(A|D) = \max_{P_{X_0}} \min_{i \in A} I(X_0 \wedge X_i | X_D).$$

With  $D = \emptyset$ , we obtain the SK capacity as  $C_S(A) = \max_{P_{X_0}} \min_{i \in A} I(X_0 \wedge X_i)$ , which is seen to coincide with the lower bound mentioned in the passage preceding Example 1 in Section II. Observe that the SK and PK capacities remain unchanged if  $A$  is replaced by  $A \cup \{0\}$ , which is to be expected in view of the manner of SK generation given in that passage.

Finally, we remark that for  $P_{X_0} = Q$ , the emulated source models corresponding to this example have, as their underlying DMMS, a Markov chain on a tree [7, Example 7] (see also the passage preceding Theorem 5.1 below) with vertex set  $\{0, 1, \dots, m\}$  and edge set  $\{(0, i), i = 1, \dots, m\}$ . Hence,  $C_S(A)$  and  $C_P(A|D)$  above are also obtained as the maximum over  $Q$  of the corresponding capacities for the emulated source models determined in [7, eqs. (36), (37)].  $\square$

## V. MODELS WITH WSI

In this section, source and channel models with WSI are addressed. Our results are partial.

Formally, a source model with WSI involves an underlying DMMS with generic rvs  $(X_0, X_1, \dots, X_m, Z)$ , where the terminals in  $\mathcal{M} = \{0, 1, \dots, m\}$  cooperating in secrecy generation observe  $X_0^n, X_1^n, \dots, X_m^n$ , respectively, whereas they remain ignorant of  $Z^n$  which is observed as WSI by the eavesdropper. Similarly, a channel model with WSI involves an underlying DMC  $W : \mathcal{X}_0 \rightarrow \mathcal{X}_1 \times \dots \times \mathcal{X}_m \times \mathcal{Z}$ , where terminal 0 governs the input, terminals  $1, \dots, m$  observe the corresponding outputs, while all remain ignorant of the  $\mathcal{Z}$ -outputs which are observed as WSI by the eavesdropper. In both these models, the permissible behavior of the terminals in  $\mathcal{M}$  is the same as in a model without WSI. The goal of the terminals  $0, 1, \dots, m$ , once again, is to generate secret common randomness for a given set  $A \subset \mathcal{M}$  of terminals, which is concealed from the eavesdropper that, in addition to observing the public communication of the terminals in  $\mathcal{M}$  and wiretapping the terminals in a given set  $D \subset \mathcal{M}$  (disjoint from  $A$  and possibly empty), now also possesses the WSI  $Z^n$ . Accordingly, in the definitions of  $(\epsilon, \delta)$ -SK and  $(\epsilon, \delta)$ -PK, the secrecy conditions (1) and (2) must be strengthened to

$$s(K; \mathbf{F}, Z^n) \leq \delta \quad \text{and} \quad s(K; U_D, X_D^n, \mathbf{F}, Z^n) \leq \delta \quad (14)$$

respectively. Except for this modification, the definitions of SK and PK capacities remain unchanged; these capacities for models with WSI will be denoted by  $C_{WS}(A)$  and  $C_{WP}(A|D)$ .

Unlike for SK and PK capacities for models without WSI, no single-letter characterization of  $C_{WS}(A)$  or  $C_{WP}(A|D)$  is available. Such a characterization of  $C_{WS}(A)$  is lacking even in the simplest case of source models with  $\mathcal{M} = A = \{0, 1\}$ , which has been extensively studied in the literature. One of the partial results available for this case is an upper bound for

$C_{WS}(A)$  given in [1]; see also [15], where the expression appearing in the bound has been termed ‘‘intrinsic conditional mutual information.’’ This bound has been improved upon in [17]. A further improvement has been reported recently in [9] as a particularization of a new upper bound for SK capacity derived therein for  $\mathcal{M} = A = \{0, 1, \dots, m\}$  (in our terminology). A generalization is also provided in [9] of the notion of CO rate introduced in [7], in terms of which a nonsingle-letter characterization of secrecy capacities for source models with WSI is given.

The simplest upper bound for secrecy capacity known for  $\mathcal{M} = A = \{0, 1\}$  easily extends to the general case, as noted for source models in [7], where a sufficient condition for its tightness has been stated, regrettably with a flawed proof. Below, we describe this bound for both source and channel models with WSI and, as a partial remedy for the flaw, prove its tightness under a sufficient condition which is more restrictive than that in [7], whose sufficiency remains unresolved.

Consider first source models with WSI. Given a model whose underlying DMMS has generic rvs  $X_0, X_1, \dots, X_m, Z$ , we associate with it a model *without* WSI determined by the same DMMS but with one more terminal  $m+1$  observing  $Z^n$  which is wiretapped by the eavesdropper. The set  $A \subset \mathcal{M}$  of terminals for which the secret key is generated remains as such, while the set  $D \subset \mathcal{M}$  of wiretapped terminals cooperating in secrecy generation grows to  $D \cup \{m+1\}$ .

*Lemma 5.1:* The PK capacity  $C_{WP}(A|D)$ , or the SK capacity  $C_{WS}(A)$  if  $D = \emptyset$ , for the model with WSI is bounded above by the PK capacity of the associated model without WSI.

*Proof:* As using  $Z^n$  can only help in secrecy generation, any  $(\epsilon, \delta)$ -achievable PK rate (or SK rate, if  $D = \emptyset$ ) for the model with WSI is an  $(\epsilon, \delta)$ -achievable PK rate for the associated model without WSI.  $\square$

Note that if in the original source model with WSI, the rv  $Z$  were altered to another rv  $Z'$  which is conditionally independent of  $X_{\mathcal{M}}$  given  $Z$ , this change can only increase the capacities  $C_{WS}(A)$  and  $C_{WP}(A|D)$  since a key satisfying a secrecy condition in (14) for  $Z$ , will also do so for  $Z'$  replacing  $Z$ . The latter capacities, with  $Z'$  playing the role of  $Z$ , can also be bounded above using Lemma 5.1. This leads to a family of upper bounds for the original  $C_{WS}(A)$  or  $C_{WP}(A|D)$ , where all the bounds have single-letter characterizations. These bounds can then be optimized with respect to the channel  $V : \mathcal{Z} \rightarrow \mathcal{Z}'$  that gives the conditional pmf of  $Z'$  given  $Z$ . In the special case  $\mathcal{M} = A = \{0, 1\}$ , the intrinsic conditional mutual information bound mentioned above is recovered.

We consider next channel models with WSI. Given a model whose underlying DMC is  $W : \mathcal{X}_0 \rightarrow \mathcal{X}_1 \times \dots \times \mathcal{X}_m \times \mathcal{Z}$ , associate with it the model without WSI determined by the same DMC but with the  $\mathcal{Z}$ -outputs observed not only by the eavesdropper but also by a terminal  $m+1$  which cooperates in secrecy generation; the set  $A$  is left unchanged while  $D$  is replaced by  $D \cup \{m+1\}$  as above. Lemma 5.1 applies to channel models as well.

Similarly as for source models, Lemma 5.1 leads to a family of upper bounds for channel models as well. Indeed, if in the

original model with WSI, the underlying DMC  $W : \mathcal{X}_0 \rightarrow \mathcal{X}_1 \times \dots \times \mathcal{X}_m \times \mathcal{Z}$  is changed to the DMC  $W' : \mathcal{X}_0 \rightarrow \mathcal{X}_1 \times \dots \times \mathcal{X}_m \times \mathcal{Z}'$  defined by

$$W'(x_1, \dots, x_m, z'|x_0) = \sum_{z \in \mathcal{Z}} W(x_1, \dots, x_m, z|x_0) V(z'|z)$$

for an arbitrary channel  $V : \mathcal{Z} \rightarrow \mathcal{Z}'$ , this change can only result in increasing the capacities  $C_{WS}(A)$  and  $C_{WP}(A|D)$ . For each such  $V$ , these capacities can be bounded above by means of Lemma 5.1; the resulting bounds can be optimized with respect to the channel  $V : \mathcal{Z} \rightarrow \mathcal{Z}'$ .

The next theorem provides a sufficient condition for the tightness of the bound in Lemma 5.1. This condition involves the concept of a Markov chain on a tree; see [8], [7].

Let  $G$  be a connected undirected graph with no cycles, i.e., a tree, with vertex set  $\{0, 1, \dots, l\}$ . For an edge  $(j, k)$  of  $G$ , denote by  $B(j \rightarrow k)$  the set of all vertices of  $G$  that are connected with  $j$  by a path containing the edge  $(j, k)$ . A family of rvs  $X_j, 0 \leq j \leq l$  form a Markov chain on the tree  $G$  if for all edges  $(j, k)$  of  $G$ , the conditional pmf of  $X_j$  given the collection of rvs  $X_{B(j \rightarrow k)}$  coincides with the conditional pmf of  $X_j$  given  $X_k$ . Equivalently, the rvs  $X_j, 0 \leq j \leq l$ , constitute a Markov chain on  $G$  if their joint pmf has the form

$$P(x_0, x_1, \dots, x_l) = P(x_0) \prod_{j=1}^l P_j(x_j|x_{i_j})$$

where  $i_j$  denotes that vertex of  $G$  for which  $(i_j, j)$  is an edge belonging to the path from vertex 0 to vertex  $j$ . Note that a Markov chain in the usual sense is a Markov chain on a tree  $G$  comprising a single path.

*Theorem 5.1:* Let  $G$  be a tree with vertex set  $\{0, 1, \dots, m+1\}$ , and  $A$  a subset of  $\mathcal{M} = \{0, 1, \dots, m\}$  of size  $|A| \geq 2$ , such that the subtree of  $G$  spanned by  $A \cup \{m+1\}$  is a path. Then a sufficient condition for the tightness of the bound in Lemma 5.1, with this  $\mathcal{M}$  and  $A$  and any  $D \subset \mathcal{M} \setminus A$ , is:

(i) for a source model, that the generic rvs  $X_0, X_1, \dots, X_m, Z$  of the underlying DMMS form a Markov chain on the tree  $G$  (with  $X_{m+1} = Z$ );

(ii) for a channel model, that the underlying DMC  $W : \mathcal{X}_0 \rightarrow \mathcal{X}_1 \times \dots \times \mathcal{X}_m \times \mathcal{Z}$  is of the form

$$W(x_1, \dots, x_{m+1}|x_0) = \prod_{j=1}^{m+1} P_j(x_j|x_{i_j}) \quad (\text{where } x_{m+1} = z) \quad (15)$$

with  $i_j$  defined by  $G$  as above.

*Remark:* For source models without WSI such that the generic rvs of the underlying DMMS form a Markov chain on a tree, a simple explicit formula for PK capacity is available [7, eq. (37)]. Hence, for models with WSI that meet the sufficient condition in Theorem 5.1(i), the SK or PK capacity  $C_{WS}(A)$  or  $C_{WP}(A|D)$  is also given by a simple expression. In the special case when  $X_0 \text{---} X_1 \text{---} \dots \text{---} X_m \text{---} Z$  is a Markov chain and  $i_{\min} < i_{\max}$  denote the smallest and largest elements of  $A$ , we obtain using [7, eq. (33)] that

$$C_{WS}(A) = \min_{i_{\min} \leq i < i_{\max}} I(X_i \wedge X_{i+1}|Z). \quad (16)$$

Unfortunately, Theorem 5.1 does not answer the case when  $Z_1 \text{---} X_0 \text{---} X_1 \text{---} \dots \text{---} X_m \text{---} Z_2$  and  $Z = (Z_1, Z_2)$ . In [7], based on the mentioned flawed result, it was claimed that (16) held in this case too. It remains unresolved whether this is true, even if  $m = 1, A = \{0, 1\}$ . ( $C_W(A)$  in [7] denotes the present  $C_{WS}(A)$ .)

*Proof:*

(i) *Source model:* It can be assumed that  $m+1$  is an endpoint of the path that is the subtree of  $G$  spanned by  $A \cup \{m+1\}$ , for else both the capacities claimed to be equal are 0. Let the other endpoint of this path be  $i$ ; clearly,  $i \in A$ .

For the model without WSI which is associated with the given model with WSI according to the passage preceding Lemma 5.1, the PK capacity can be achieved with the key generated by terminal  $i$  as  $K^{(n)} = K^{(n)}(X_i^n)$ , by Theorem 3.2. By the same theorem, this  $K^{(n)}$  is  $\epsilon_n$ -recoverable at each  $j \in A \setminus \{i\}$  because the entire  $X_i^n$  is, using communication  $\mathbf{F}^{(n)} = (f_l(X_l^n), l \in \{0, 1, \dots, m+1\})$ , where  $X_{m+1}^n = Z^n$  and also  $f_{m+1}(X_{m+1}^n) = Z^n$ . It suffices to show that in the present case, a knowledge of  $Z^n$  is not needed to recover  $X_i^n$ .

Now, given  $j \in A \setminus \{i\}$ , let  $(j, k)$  be the starting edge of the path from  $j$  to  $i$ . Then, since  $j$  belongs to the path connecting  $i$  and  $m+1$ , it holds that  $i \in B(j \rightarrow k), m+1 \in B(k \rightarrow j)$ . The Markov assumption implies that the collections of rvs  $X_{B(j \rightarrow k)}$  and  $X_{B(k \rightarrow j)}$  are conditionally independent given  $X_j$ , and hence the same holds also for their i.i.d. repetitions. It follows, in particular, that  $(X_i^n, f_l(X_l^n), l \in B(j \rightarrow k))$  is conditionally independent of  $(f_l(X_l^n), l \in B(k \rightarrow j))$  given  $X_j^n$ . Hence, in order to recover  $X_i^n$  from  $X_j^n$  and the communication  $\mathbf{F}^{(n)}$ , the part  $(f_l(X_l^n), l \in B(k \rightarrow j))$  of this communication containing  $f_{m+1}(X_{m+1}^n) = Z^n$  is redundant. This proves that a knowledge of  $Z^n$  is unnecessary to recover  $X_i^n$ .

(ii) *Channel model:* Consider a channel model with WSI, with underlying DMC  $W : \mathcal{X}_0 \rightarrow \mathcal{X}_1 \times \dots \times \mathcal{X}_m \times \mathcal{Z}$ , that satisfies the condition in Theorem 5.1. The PK capacity of the associated channel model without WSI is equal, by Theorem 4.1, to the maximum with respect to  $Q$  of the PK capacities of the source models (without WSI) whose generic rvs  $X_0, X_1, \dots, X_m, Z$  have joint pmf of the form  $Q(x_0)W(x_1, \dots, x_m, z|x_0)$ . On account of (15), the latter means that  $X_0, X_1, \dots, X_m, Z$  form a Markov chain on the tree  $G$ . Hence, taking the maximizing  $Q$ , the PK capacity of the corresponding source model can be achieved without any knowledge of  $Z^n$ , by the proof of part (i). Thus, the strategy that achieves this PK capacity also achieves the (SK or) PK capacity of the original channel model with WSI.  $\square$

We illustrate Theorem 5.1 and its limitations by two examples involving the graphs shown in Figs. 1 and 2. For both graphs  $G$ , we consider source models with WSI whose underlying DMMS has generic rvs  $X_0, X_1, X_2, Z$ , which form a Markov chain on  $G$ , and channel models with WSI whose underlying DMC is of the form (15). In both cases,  $\mathcal{M} = \{0, 1, 2\}$  and vertex 3 represents the additional terminal needed in the associated model without WSI.

*Example 4:* For the graph of Fig. 1, the hypothesis of Theorem 5.1 is valid for  $A = \{0, 1\}$  and  $A = \{0, 2\}$ , but not for

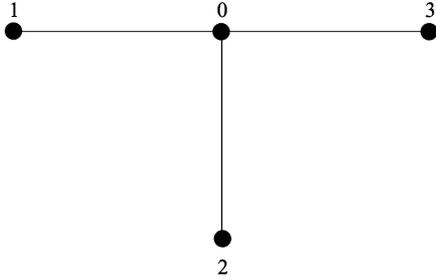


Fig. 1. Example 4.



Fig. 2. Example 5.

$A = \{1, 2\}$  or  $A = \{0, 1, 2\}$ . When  $A = \{0, 1\}$ , Theorem 5.1 gives that the SK capacity  $C_{WS}(\{0, 1\})$  for the source model with WSI is equal to the PK capacity of the associated source model without WSI that has  $\mathcal{M} = \{0, 1, 2, 3\}$ ,  $A = \{0, 1\}$ ,  $D = \{3\}$ . According to [7, Example 7], this PK capacity equals  $I(X_0 \wedge X_1|Z)$ . Similarly, Theorem 5.1 gives that the PK capacity  $C_{WP}(\{0, 1\}|\{2\})$  is equal to the PK capacity of the associated model without WSI, again with  $\mathcal{M} = \{0, 1, 2, 3\}$ ,  $A = \{0, 1\}$ , but now with  $D = \{2, 3\}$ . Hence,  $C_{WP}(\{0, 1\}|\{2\}) = I(X_0 \wedge X_1|X_2, Z)$ .

For a channel model with WSI, the corresponding capacities  $C_{WS}(\{0, 1\})$  and  $C_{WP}(\{0, 1\}|\{2\})$  are obtained by maximizing, with respect to  $Q$ , the source model capacities  $I(X_0 \wedge X_1|Z)$  and  $I(X_0 \wedge X_1|X_2, Z)$  for rvs  $X_0, X_1, X_2, Z$  with joint pmf  $Q(x_0)W(x_1, x_2, z|x_0)$ . The source or channel model with WSI, corresponding to  $A = \{1, 2\}$  is not covered by Theorem 5.1. A whiff of the potential difficulty of determining the SK capacity  $C_{WS}(\{1, 2\})$  (or the PK capacity  $C_{WP}(\{1, 2\}|\{0\})$ ) is given by the close connection of this problem to that studied in [13], [15], whose solution remains elusive. The latter asks, in the “three independent channel” situation of Fig. 1 with  $\mathcal{X}_0 = \{0, 1\}$ , for the SK capacity (in our terminology)  $C_{WS}(\{1, 2\})$  for the source model with WSI for which  $\mathcal{M} = \{1, 2\}$ . This differs from the problem above of determining  $C_{WS}(\{1, 2\})$  for the source model with  $\mathcal{M} = \{0, 1, 2\}$  only in that the model with  $\mathcal{M} = \{1, 2\}$  excludes terminal 0 from cooperating in secrecy generation.  $\square$

*Example 5:* The graph of Fig. 2 is a path, and  $X_0, X_1, X_2, Z$  form a Markov chain on  $G$  iff they form a Markov chain in the usual sense in the order  $Z \text{---} 0 \text{---} X_0 \text{---} 0 \text{---} X_1 \text{---} 0 \text{---} X_2$ . The hypothesis of Theorem 5.1 is valid for any choice of  $A$ , and we concentrate on the case  $A = \{0, 1\}$ ,  $D = \{2\}$ . By Theorem 5.1, the PK capacity  $C_{WP}(\{0, 1\}|\{2\})$  for the source model with WSI is equal to the PK capacity for the associated model without WSI, with  $\mathcal{M} = \{0, 1, 2, 3\}$ ,  $A = \{0, 1\}$ ,  $D = \{2, 3\}$ , which is  $I(X_0 \wedge X_1|X_2, Z)$ . As before, the PK capacity  $C_{WP}(\{0, 1\}|\{2\})$  for the analogous channel model with WSI is obtained by maximization with respect to the pmf of  $X_0$ . Recall that the reason for the equality of the PK capacities for the source model with WSI and the associated model without WSI is that the latter PK capacity can be achieved without any use of the knowledge of  $Z^n$ ; see the

proof of Theorem 5.1(i). If the latter PK capacity also could be achieved upon dispensing with the knowledge of both  $X_2^n$  and  $Z^n$ , it could be concluded similarly for the source model with WSI represented by  $(X_2, Z)$ , i.e., with  $\mathcal{M} = A = \{0, 1\}$ , and underlying DMMS with generic rvs  $X_0, X_1, (X_2, Z)$ , that the SK capacity also would be equal to  $I(X_0 \wedge X_1|X_2, Z)$ . It remains open whether the equality is true; see the remark after Theorem 5.1.  $\square$

## VI. DISCUSSION

Channel models for generating a Shannon-theoretic secret key for several parties have been studied, in which the permitted transactions comprise transmissions over a secure DMC with one input and multiple output terminals, as well as unrestricted communication among all the terminals over a public noiseless channel. The objective is to generate secret common randomness at the largest possible rate for a given set  $A$  of secret key-seeking terminals, which is concealed from an eavesdropper with access to the public communication and perhaps also to a set  $D$  of “wiretapped” terminals disjoint from  $A$ . All the terminals, including those in the set  $D$ , cooperate in achieving this objective. The corresponding optimal rates (secrecy capacities) have been called the SK and PK capacity, respectively. In more complex models, the eavesdropper may have additional “wiretap side information,” not available to any of the cooperating terminals.

Our primary focus has been on models without WSI, for which single-letter characterizations of SK and PK capacities have been derived using our previous results for source models [7]. By our main result, when randomization is allowed at the input terminal, the secrecy capacities for the channel model can be achieved by emulating a source model upon transmitting over the DMC an i.i.d. sequence of rvs, with the best pmf. This need not be the most efficient way to achieve a secrecy capacity, for in general, it requires terminal 0 to use the public channel too. However, it is shown that a key of optimal rate can be generated with the input terminal not having to send any public message, and with each of the remaining terminals sending at most one public message, without interaction or randomization. When the input terminal belongs to the set  $A$ , it can generate the key at the outset and transmit it over the DMC without loss of rate optimality. The last assertion is a consequence of a new result for source models: the SK or PK capacity can be achieved with the key being generated by any chosen terminal in the set  $A$ , obviously of the public communication. When no randomization is permitted at terminal 0, the SK or PK capacity equals the PK capacity of the counterpart previous model modified to include 0 in the set  $D$  of wiretapped terminals, and is achieved with terminal 0 transmitting a deterministic sequence over the DMC; a constant sequence, however, may be inadequate.

We note that our results extend to models with constraints imposed on the DMC inputs. Suppose that the  $n$  symbols transmitted by terminal 0 over the DMC  $W$  are subject to the input constraint  $\frac{1}{n} \sum_{t=1}^n c(X_{0t}) \leq \Gamma$  for a nonnegative function  $c(x)$  defined on  $\mathcal{X}_0$ , with  $\Gamma > \min_{x \in \mathcal{X}_0} c(x)$ . For a pmf  $P_{X_0} = Q$ , denote, with a slight abuse of notation,  $c(Q) = \sum_{x \in \mathcal{X}_0} Q(x)c(x)$ . Theorem 4.1 holds under this input constraint, with the maxima over  $Q$  in (8) and (9) now being

subject to  $c(Q) \leq \Gamma$ . In the achievability part of Theorem 4.1, the rv  $X_0$  is chosen with pmf  $P_{X_0} = Q$  that satisfies  $c(Q) < \Gamma$ . Then the i.i.d. rvs  $X_{01}, \dots, X_{0n}$ , transmitted by terminal 0 in emulating a source model, satisfy the input constraint above with probability approaching 1 as  $n \rightarrow \infty$ . In the converse part, observe that by the concavity in  $Q$  of the function  $G_{A|D}(Q, \lambda)$  (see Lemma A.1 in Appendix A), the right-hand side of (12) can be further bounded above by  $G_{A|D}(\frac{1}{n} \sum_{t=1}^n P_{X_{0t}}, \lambda)$ , where  $X_{01}, \dots, X_{0n}$  satisfy the input constraint. Then, since

$$c\left(\frac{1}{n} \sum_{t=1}^n P_{X_{0t}}\right) = \frac{1}{n} \sum_{t=1}^n c(P_{X_{0t}}) \leq \Gamma$$

the maximum in (13) can be taken over  $Q$  satisfying  $c(Q) \leq \Gamma$ .

For models with WSI, single-letter characterizations of secrecy capacities are elusive, in general, even in the extensively examined case of SK capacity for two terminals. We have provided partial results for (source and) channel models with WSI. Specifically, upper bounds for SK and PK capacities have been given in terms of their counterpart capacities for associated models without WSI; the latter models are obtained by assuming the wiretapped terminal with side information to cooperate in generating the key. Sufficient conditions for the tightness of these bounds have also been provided. The conditions are more restrictive than the sufficient condition for tightness of the bound for source models stated in [7] with a flawed proof.

We emphasize that for general source and channel models with WSI for which single-letter characterizations of SK or PK capacities remain elusive, we make no claims on the manner of transmission and communication strategies by which the capacities could be achieved. In the special circumstance of Theorem 5.1 when the SK and PK capacities of a model with WSI have been determined, these capacities can be achieved by means of the same simple manner of strategies as for models without WSI, except that the key cannot necessarily be generated by terminal 0. In the presence of channel input constraints, the results of Theorem 5.1 can be extended as for channel models without WSI.

#### APPENDIX A

In this appendix,  $A$  and  $D$  denote disjoint subsets of  $\mathcal{M} = \{0, 1, \dots, m\}$  with  $|A| \geq 2$ . The notation introduced in Section III is followed; in particular,  $\Lambda(A)$  and  $\Lambda(A|D)$  denote, respectively, the sets of collections of weights  $\lambda = \{\lambda_B : B \in \mathcal{B}(A)\}$  and  $\lambda = \{\lambda_B : B \in \mathcal{B}(A|D)\}$  defined there.

*Lemma A.1:* For any  $\lambda \in \Lambda(A)$ ,  $G_A(Q, \lambda)$  defined by (6) for a given DMC  $W : \mathcal{X}_0 \rightarrow \mathcal{X}_1 \times \dots \times \mathcal{X}_m$  is a concave function of  $Q$ ; and for any  $\lambda \in \Lambda(A|D)$ , so is  $G_{A|D}(Q, \lambda)$  defined by (7).

*Proof:* Since  $\lambda \in \Lambda(A)$  implies  $\sum_{B \in \mathcal{B}_0(A)} \lambda_B = 1$  by definition, we have by (6) that

$$\begin{aligned} G_A(Q, \lambda) &= H(X_{\mathcal{M}}) - \sum_{B \in \mathcal{B}(A)} \lambda_B H(X_B|X_{B^c}) \\ &= \sum_{B \in \mathcal{B}_0(A)} \lambda_B [H(X_{\mathcal{M}}) - H(X_B|X_{B^c})] \\ &\quad - \sum_{B \in \mathcal{B}(A) \setminus \mathcal{B}_0(A)} \lambda_B H(X_B|X_{B^c}) \end{aligned}$$

$$\begin{aligned} &= \sum_{B \in \mathcal{B}_0(A)} \lambda_B H(X_{B^c}) \\ &\quad - \sum_{B \in \mathcal{B}(A) \setminus \mathcal{B}_0(A)} \lambda_B [H(X_{\mathcal{M}}|X_0) - H(X_{B^c}|X_0)]. \end{aligned}$$

As a function of  $P_{X_0} = Q$ , the pmf of  $X_{B^c}$  is affine and therefore  $H(X_{B^c})$  is concave, and the conditional entropies  $H(X_{\mathcal{M}}|X_0)$  and  $H(X_{B^c}|X_0)$  are affine; this proves the first assertion.

The second assertion follows similarly by an analogous identity from (7) that

$$\begin{aligned} G_{A|D}(Q, \lambda) &= H(X_{\mathcal{M}}|X_D) - \sum_{B \in \mathcal{B}(A|D)} \lambda_B H(X_B|X_{B^c}) \\ &= \sum_{B \in \mathcal{B}_0(A|D)} \lambda_B H(X_{B^c}|X_D) \\ &\quad - \sum_{B \in \mathcal{B}(A|D) \setminus \mathcal{B}_0(A|D)} \lambda_B [H(X_{\mathcal{M}}|X_0) - H(X_{B^c}|X_0)] \end{aligned}$$

using the fact that  $H(X_{B^c}|X_D)$  is a concave function of the joint pmf of  $(X_{B^c}, X_D)$ .  $\square$

*Lemma A.2:* Let  $X_{\mathcal{M}} = (X_0, X_1, \dots, X_m)$  and  $K, Y$  be rvs such that  $K$  is  $\epsilon$ -recoverable from  $(X_i, Y)$  for each  $i \in A$ . Then for an arbitrary  $\lambda \in \Lambda(A)$

$$\begin{aligned} H(K|Y) &\leq H(X_{\mathcal{M}}|Y) - \sum_{B \in \mathcal{B}(A)} \lambda_B H(X_B|X_{B^c}, Y) + (m+2)\nu \end{aligned}$$

where  $\nu = \epsilon \log |\mathcal{K}| + h(\epsilon)$ .

*Proof:* If  $B \in \mathcal{B}(A)$ , then by definition,  $A \cap B^c \neq \emptyset$ . As  $K$  is  $\epsilon$ -recoverable from  $(X_i, Y)$  if  $i \in A \cap B^c$ , by Fano's inequality

$$H(K|X_{B^c}, Y) \leq H(K|X_i, Y) \leq \nu$$

for  $\nu$  as in the lemma. Hence, for  $B \in \mathcal{B}(A)$

$$\begin{aligned} H(X_B|X_{B^c}, Y) &= H(X_B|X_{B^c}, K, Y) + I(K \wedge X_B|X_{B^c}, Y) \\ &\leq H(X_B|X_{B^c}, K, Y) + \nu. \end{aligned}$$

This, and

$$\begin{aligned} H(X_B|X_{B^c}, K, Y) &= \sum_{i \in B} H(X_i|X_{[0, i-1] \cap B}, X_{B^c}, K, Y) \\ &\leq \sum_{i \in B} H(X_i|X_{[0, i-1]}, K, Y) \end{aligned}$$

with  $X_{[0, i-1]} = (X_0, \dots, X_{i-1})$  imply, since  $\sum_{B \in \mathcal{B}_i(A)} \lambda_B = 1, i \in \mathcal{M}$ , that

$$\begin{aligned} &\sum_{B \in \mathcal{B}(A)} \lambda_B H(X_B|X_{B^c}, Y) \\ &\leq \sum_{i \in \mathcal{M}} \sum_{B \in \mathcal{B}_i(A)} \lambda_B (H(X_i|X_{[0, i-1]}, K, Y) + \nu) \\ &\leq \sum_{i \in \mathcal{M}} H(X_i|X_{[0, i-1]}, K, Y) + (m+1)\nu \\ &= H(X_{\mathcal{M}}|K, Y) + (m+1)\nu. \end{aligned}$$

Hence

$$\begin{aligned} H(X_{\mathcal{M}}|Y) - \sum_{B \in \mathcal{B}(A)} \lambda_B H(X_B|X_{B^c}, Y) \\ \geq I(K \wedge X_{\mathcal{M}}|Y) - (m+1)\nu \\ \geq H(K|Y) - (m+2)\nu \end{aligned}$$

the last step again by Fano's inequality. Rearranging terms, the proof is completed.  $\square$

*Corollary:* Suppose that for  $X_{\mathcal{M}}$ ,  $Y$ , and  $K$ , the rv  $K$  is  $\epsilon$ -recoverable from  $(X_i, X_D, Y)$  for each  $i \in A$ . Then, for an arbitrary  $\lambda \in \Lambda(A|D)$

$$\begin{aligned} H(K|X_D, Y) \leq H(X_{\mathcal{M}}|X_D, Y) \\ - \sum_{B \in \mathcal{B}(A|D)} \lambda_B H(X_B|X_{B^c}, Y) + (m+2)\nu. \quad (\text{A1}) \end{aligned}$$

*Proof:* By Lemma A.2 applied to  $(X_D, Y)$  in the role of  $Y$

$$\begin{aligned} H(K|X_D, Y) \leq H(X_{\mathcal{M}}|X_D, Y) \\ - \sum_{B \in \mathcal{B}(A)} \lambda'_B H(X_B|X_{B^c}, X_D, Y) + (m+2)\nu \quad (\text{A2}) \end{aligned}$$

for arbitrary  $\lambda' = \{\lambda'_B : B \in \mathcal{B}(A)\} \in \Lambda(A)$ . If  $\lambda = \{\lambda_B : B \in \mathcal{B}(A|D)\} \in \Lambda(A|D)$ , then  $\lambda'$  defined by

$$\lambda'_B = \begin{cases} \lambda_B, & \text{if } B \in \mathcal{B}(A|D) \\ 0, & \text{if } B \cap D \neq \emptyset, B \neq D \\ 1, & \text{if } B = D \end{cases}$$

belongs to  $\Lambda(A)$ , and (A2) applied to this  $\lambda'$  reduces to (A1).  $\square$

## APPENDIX B

In order to complete the proof of Theorem 4.1, we prove that in (11)

$$\begin{aligned} H(U_{\mathcal{M}}, X_{\mathcal{M}}^n | U_D, X_D^n, \tilde{\mathbf{F}}^{(n)}) \\ - \sum_{B \in \mathcal{B}(A|D)} \lambda_B H(U_B, X_B^n | U_{B^c}, X_{B^c}^n, \tilde{\mathbf{F}}^{(n)}) \\ \leq \sum_{t=1}^n \left( H(X_{\mathcal{M}t} | X_{Dt}) - \sum_{B \in \mathcal{B}(A|D)} \lambda_B H(X_{Bt} | X_{B^ct}) \right). \quad (\text{B1}) \end{aligned}$$

First, observe that the left-hand side of (B1) is

$$\begin{aligned} H(U_{\mathcal{M}}, X_{\mathcal{M}}^n) - H(U_D, X_D^n, \tilde{\mathbf{F}}^{(n)}) \\ - \sum_{B \in \mathcal{B}(A|D)} \lambda_B \left( H(U_{\mathcal{M}}, X_{\mathcal{M}}^n) - H(U_{B^c}, X_{B^c}^n, \tilde{\mathbf{F}}^{(n)}) \right) \\ = (1 - \lambda_{\text{sum}}) H(U_{\mathcal{M}}, X_{\mathcal{M}}^n) - H(U_D, X_D^n, \tilde{\mathbf{F}}^{(n)}) \\ + \sum_{B \in \mathcal{B}(A|D)} \lambda_B H(U_{B^c}, X_{B^c}^n, \tilde{\mathbf{F}}^{(n)}) \quad (\text{B2}) \end{aligned}$$

where

$$\lambda_{\text{sum}} = \sum_{B \in \mathcal{B}(A|D)} \lambda_B. \quad (\text{B3})$$

Now, in (B2)

$$\begin{aligned} H(U_{\mathcal{M}}, X_{\mathcal{M}}^n) &= H(U_{\mathcal{M}}) + \sum_{t=1}^n H(X_{\mathcal{M}t} | U_{\mathcal{M}}, X_{\mathcal{M}}^{t-1}) \\ &= H(U_{\mathcal{M}}) + \sum_{t=1}^n H(X_{\mathcal{M}t} | X_{0t}) \quad (\text{B4}) \end{aligned}$$

where the last equality holds because  $X_{0t}$  is a function of  $(U_0, F^{t-1}) = (U_0, U_D, X_D^{t-1}, \tilde{F}^{t-1})$  and so is determined by  $(U_{\mathcal{M}}, X_{\mathcal{M}}^{t-1})$ , and  $X_{\mathcal{M}t}$  is conditionally independent of  $(U_{\mathcal{M}}, X_{\mathcal{M}}^{t-1})$  given  $X_{0t}$ .

Next, in (B2)

$$\begin{aligned} H(U_D, X_D^n, \tilde{\mathbf{F}}^{(n)}) \\ = H(U_D) + \sum_{t=1}^n H(X_{Dt}, \tilde{F}_t | U_D, X_D^{t-1}, \tilde{F}^{t-1}) \\ = H(U_D) + \sum_{t=1}^n H(X_{Dt} | U_D, X_D^{t-1}, \tilde{F}^{t-1}) \\ + \sum_{t=1}^n H(\tilde{F}_t | U_D, X_D^t, \tilde{F}^{t-1}) \quad (\text{B5}) \end{aligned}$$

where, for the same reason as above,

$$H(X_{Dt} | U_D, X_D^{t-1}, \tilde{F}^{t-1}) = H(X_{Dt} | X_{0t}), \quad \text{if } 0 \in D. \quad (\text{B6})$$

Similarly, in (B2)

$$\begin{aligned} H(U_{B^c}, X_{B^c}^n, \tilde{\mathbf{F}}^{(n)}) \\ = H(U_{B^c}) + \sum_{t=1}^n H(X_{B^ct} | U_{B^c}, X_{B^c}^{t-1}, \tilde{F}^{t-1}) \\ + \sum_{t=1}^n H(\tilde{F}_t | U_{B^c}, X_{B^c}^t, \tilde{F}^{t-1}) \quad (\text{B7}) \end{aligned}$$

where

$$H(X_{B^ct} | U_{B^c}, X_{B^c}^{t-1}, \tilde{F}^{t-1}) = H(X_{B^ct} | X_{0t}) \quad (\text{B8})$$

if  $0 \in B^c$ . By (B2)–(B5) and (B7), the left-hand side of (B1) decomposes as  $E_1 + E_2 + E_3$  with

$$\begin{aligned} E_1 &= (1 - \lambda_{\text{sum}}) H(U_{\mathcal{M}}) - H(U_D) \\ &\quad + \sum_{B \in \mathcal{B}(A|D)} \lambda_B H(U_{B^c}); \\ E_2 &= \sum_{t=1}^n \left[ (1 - \lambda_{\text{sum}}) H(X_{\mathcal{M}t} | X_{0t}) \right. \\ &\quad \left. - H(X_{Dt} | U_D, X_D^{t-1}, \tilde{F}^{t-1}) \right] \end{aligned}$$

$$E_3 = \sum_{t=1}^n \left[ -H(\tilde{F}_t|U_D, X_D^t, \tilde{F}^{t-1}) + \sum_{B \in \mathcal{B}(A|D)} \lambda_B H(X_{B^c t}|U_{B^c}, X_{B^c}^{t-1}, \tilde{F}^{t-1}) \right];$$

$$+ \sum_{B \in \mathcal{B}(A|D)} \lambda_B H(\tilde{F}_t|U_{B^c}, X_{B^c}^t, \tilde{F}^{t-1}).$$

Each of the terms  $E_1$ ,  $E_2$ , and  $E_3$  will be dealt with separately.

First

$$\sum_{B \in \mathcal{B}(A|D)} \lambda_B H(U_{B^c})$$

$$= \sum_{B \in \mathcal{B}(A|D)} \lambda_B \left( \sum_{i \in B^c} H(U_i) \right)$$

$$= \sum_{i=0}^m \left( \sum_{B \in \mathcal{B}(A|D): i \in B^c} \lambda_B \right) H(U_i)$$

$$= \sum_{i=0}^m \left( \sum_{B \in \mathcal{B}(A|D) \setminus \mathcal{B}_i(A|D)} \lambda_B \right) H(U_i)$$

$$= \sum_{i=0}^m \left( \lambda_{\text{sum}} - \sum_{B \in \mathcal{B}_i(A|D)} \lambda_B \right) H(U_i).$$

Here,  $\mathcal{B}_i(A|D) = \{B \in \mathcal{B}(A|D) : i \in B\}$  is nonempty iff  $i \in D^c$ , in which case  $\sum_{B \in \mathcal{B}_i(A|D)} \lambda_B = 1$  by definition. Thus, the last term above equals  $\lambda_{\text{sum}} H(U_{\mathcal{M}}) - H(U_{D^c})$ , and therefore  $E_1 = 0$ .

Second,  $E_3 \leq 0$  since each of the  $n$  terms in the sum is nonpositive by the Corollary of Lemma B.1 below, when applied to the communication  $\tilde{F}_t$  of the terminals  $i \in D^c$  in interval  $t$ , with each terminal  $i \in D^c$  possessing initial knowledge  $(U_i, X_i^t, U_D, X_D^t, \tilde{F}^{t-1})$ .

Third, we claim that for each  $1 \leq t \leq n$ , the  $t$ th term of  $E_2$  satisfies

$$(1 - \lambda_{\text{sum}}) H(X_{\mathcal{M}t}|X_{0t})$$

$$- H(X_{Dt}|U_D, X_D^{t-1}, \tilde{F}^{t-1})$$

$$+ \sum_{B \in \mathcal{B}(A|D)} \lambda_B H(X_{B^c t}|U_{B^c}, X_{B^c}^{t-1}, \tilde{F}^{t-1})$$

$$\leq H(X_{\mathcal{M}t}|X_{Dt}) - \sum_{B \in \mathcal{B}(A|D)} \lambda_B H(X_{Bt}|X_{B^c t}). \quad (\text{B9})$$

Proving this will establish (B1).

In the case  $0 \in D$ , (B9) holds with equality. For then, by (B6) and (B8), the left-hand side of (B9) is

$$(1 - \lambda_{\text{sum}}) H(X_{\mathcal{M}t}|X_{0t}) - H(X_{Dt}|X_{0t})$$

$$+ \sum_{B \in \mathcal{B}(A|D)} \lambda_B H(X_{B^c t}|X_{0t})$$

$$= H(X_{\mathcal{M}t}|X_{0t}) - H(X_{Dt}|X_{0t})$$

$$- \sum_{B \in \mathcal{B}(A|D)} \lambda_B (H(X_{\mathcal{M}t}|X_{0t}) - H(X_{B^c t}|X_{0t})).$$

Using  $0 \in D \subset B^c$ , the previous expression clearly equals the right-hand side of (B9).

Turning to the case  $0 \notin D$ , write the sum on the left-hand side of (B9) as two sums over  $B \in \mathcal{B}(A|D)$ , for  $0 \in B^c$  and  $0 \in B$ , respectively. Since  $\sum_{B \in \mathcal{B}(A|D): 0 \in B} \lambda_B = 1$  and so  $\sum_{B \in \mathcal{B}(A|D): 0 \in B^c} \lambda_B = \lambda_{\text{sum}} - 1$ , it follows by using (B8) that the left-hand side of (B9) equals the expression of (B10), shown at the bottom of the page. Here, for  $B$  with  $0 \in B^c$

$$- H(X_{\mathcal{M}t}|X_{0t}) + H(X_{B^c t}|X_{0t})$$

$$= -H(X_{\mathcal{M}t}) + H(X_{B^c t})$$

$$= -H(X_{Bt}|X_{B^c t}).$$

Further, since  $B \in \mathcal{B}(A|D)$  implies  $B^c \supset D$

$$H(X_{B^c t}|U_{B^c}, X_{B^c}^{t-1}, \tilde{F}^{t-1})$$

$$- H(X_{Dt}|U_D, X_D^{t-1}, \tilde{F}^{t-1})$$

$$\leq H(X_{B^c t}|U_{B^c}, X_{B^c}^{t-1}, \tilde{F}^{t-1})$$

$$- H(X_{Dt}|U_{B^c}, X_{B^c}^{t-1}, \tilde{F}^{t-1})$$

$$= H(X_{B^c t}|X_{Dt}, U_{B^c}, X_{B^c}^{t-1}, \tilde{F}^{t-1})$$

$$\leq H(X_{B^c t}|X_{Dt})$$

$$= H(X_{B^c t}) - H(X_{Dt}).$$

The last entropy difference is equal to  $H(X_{\mathcal{M}t}|X_{Dt}) - H(X_{Bt}|X_{B^c t})$ , and so (B10) is bounded above by

$$- \sum_{B \in \mathcal{B}(A|D): 0 \in B^c} \lambda_B H(X_{Bt}|X_{B^c t})$$

$$+ \sum_{B \in \mathcal{B}(A|D): 0 \in B} \lambda_B (H(X_{\mathcal{M}t}|X_{Dt}) - H(X_{Bt}|X_{B^c t}))$$

$$= H(X_{\mathcal{M}t}|X_{Dt}) - \sum_{B \in \mathcal{B}(A|D)} \lambda_B H(X_{Bt}|X_{B^c t})$$

thereby establishing (B9) also for the case  $0 \notin D$ . This completes the proof of (B1).

$$\sum_{B \in \mathcal{B}(A|D): 0 \in B^c} \lambda_B (-H(X_{\mathcal{M}t}|X_{0t}) + H(X_{B^c t}|X_{0t}))$$

$$+ \sum_{B \in \mathcal{B}(A|D): 0 \in B} \lambda_B \left[ H(X_{B^c t}|U_{B^c}, X_{B^c}^{t-1}, \tilde{F}^{t-1}) - H(X_{Dt}|U_D, X_D^{t-1}, \tilde{F}^{t-1}) \right]. \quad (\text{B10})$$

*Lemma B.1:* Given rvs  $X_0, X_1, \dots, X_m, Y$ , for interactive communication  $F$  of the terminals  $i \in \mathcal{M}$  in the sense of Definition 1, with terminal  $i \in \mathcal{M}$  possessing initial knowledge  $(X_i, Y)$ , the inequality

$$\sum_{B \in \mathcal{B}} \lambda_B H(F|X_{B^c}, Y) \leq H(F|Y) \quad (\text{B11})$$

holds for an arbitrary family  $\mathcal{B}$  of subsets of  $\mathcal{M}$  and numbers  $\lambda_B \geq 0, B \in \mathcal{B}$ , that satisfy  $\sum_{B \in \mathcal{B}_i} \lambda_B = 1$  for each  $i \in \mathcal{M}$ , where  $\mathcal{B}_i = \{B \in \mathcal{B} : i \in B\}$ .

*Proof:* By Definition 1,  $f_{ji}$  is a function of  $(X_i, Y)$  and  $\phi_{ji}$ , and hence in the decomposition

$$H(F|X_{B^c}, Y) = \sum_{j=1}^r \sum_{i=0}^m H(f_{ji}|X_{B^c}, Y, \phi_{ji})$$

the terms  $i \in B^c$  vanish. Thus

$$\begin{aligned} H(F|X_{B^c}, Y) &= \sum_{j=1}^r \sum_{i \in B} H(f_{ji}|X_{B^c}, Y, \phi_{ji}) \\ &\leq \sum_{j=1}^r \sum_{i \in B} H(f_{ji}|Y, \phi_{ji}). \end{aligned}$$

It follows that

$$\begin{aligned} \sum_{B \in \mathcal{B}} \lambda_B H(F|X_{B^c}, Y) &\leq \sum_{B \in \mathcal{B}} \sum_{j=1}^r \sum_{i \in B} \lambda_B H(f_{ji}|Y, \phi_{ji}) \\ &= \sum_{j=1}^r \sum_{i=0}^m \sum_{B \in \mathcal{B}_i} \lambda_B H(f_{ji}|Y, \phi_{ji}) \\ &= \sum_{j=1}^r \sum_{i=0}^m H(f_{ji}|Y, \phi_{ji}) \\ &= H(F|Y) \end{aligned}$$

as claimed.  $\square$

*Corollary:* Given rvs  $X_0, X_1, \dots, X_m, Y$ , and a set  $D \subset \mathcal{M}$ , for interactive communication  $\tilde{F}$  of the terminals  $i \in D^c$ , with terminal  $i \in D^c$  possessing initial knowledge  $(X_i, X_D, Y)$ , the inequality

$$\sum_{B \in \mathcal{B}} \lambda_B H(\tilde{F}|X_{B^c}, Y) \leq H(\tilde{F}|X_D, Y)$$

holds for an arbitrary family  $\mathcal{B}$  of subsets of  $D^c$  and numbers  $\lambda_B \geq 0, B \in \mathcal{B}$ , that satisfy  $\sum_{B \in \mathcal{B}_i} \lambda_B = 1$  for each  $i \in D^c$ .

*Proof:* Apply Lemma B.1 with  $D^c$  in the role of  $\mathcal{M}$ ,  $\tilde{F}$  in the role of  $F$ , and with the role of  $X_i, i \in \mathcal{M}$ , now played

by  $X_i, i \in D^c$ , and that of  $Y$  by  $(X_D, Y)$ . As  $\mathcal{M}$  has been replaced by  $D^c$ , we must replace—in (B11)— $B^c$  by  $D^c \setminus B$  to get the inequality

$$\sum_{B \in \mathcal{B}} \lambda_B H(\tilde{F}|X_{D^c \setminus B}, X_D, Y) \leq H(\tilde{F}|X_D, Y).$$

Since  $\mathcal{B}$  consists of subsets of  $D^c$ , here  $(X_{D^c \setminus B}, X_D) = X_{B^c}$ , so that the previous inequality is the same as the one claimed.  $\square$

## REFERENCES

- [1] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography, Part I: Secret sharing," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1121–1132, Jul. 1993.
- [2] C. H. Bennett, G. Brassard, C. Crépeau, and U. M. Maurer, "Generalized privacy amplification," *IEEE Trans. Inf. Theory*, vol. 41, no. 6, pp. 1915–1923, Nov. 1995.
- [3] I. Csiszár, "Almost independence and secrecy capacity," *Probl. Pered. Inform.*, vol. 32, no. 1, pp. 48–57, 1996.
- [4] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 3, pp. 339–348, May 1978.
- [5] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1982.
- [6] I. Csiszár and P. Narayan, "Common randomness and secret key generation with a helper," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 344–366, Mar. 2000.
- [7] I. Csiszár and P. Narayan, "The secret key capacity of multiple terminals," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3047–3061, Dec. 2004.
- [8] H. O. Georgii, *Gibbs Measures and Phase Transitions*. Berlin, Germany: de Gruyter, 1988.
- [9] A. Gohari and V. Anantharam, "Communication for omniscience by a neutral observer and information-theoretic key agreement of multiple terminals," in *Proc. 2007 IEEE Int. Symp. Information Theory*, Nice, France, Jun. 2007, pp. 2056–2060.
- [10] S. Karlin, *Mathematical Methods and Theory in Game, Programming and Economics*. Reading, MA: Addison-Wesley, 1959.
- [11] M. Madiman and A. Barron, "Generalized entropy power inequalities and monotonicity properties of information," *IEEE Trans. Inf. Theory*, vol. 53, no. 7, pp. 2317–2329, Jul. 2007.
- [12] U. M. Maurer, "Provably secure key distribution based on independent channels," presented at the IEEE Workshop Information Theory, Eindhoven, The Netherlands, Jun. 1990.
- [13] U. M. Maurer, "Secret key agreement by public discussion from common information," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 733–742, May 1993.
- [14] U. M. Maurer, "The strong secret key rate of discrete random triples," in *Communications and Cryptography: Two Sides of One Tapestry*, R. E. Blahut, Ed. *et al.* Norwell, MA: Kluwer, 1994, ch. 26, pp. 271–285.
- [15] U. M. Maurer and S. Wolf, "Unconditionally secure key agreement and the intrinsic conditional information," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 499–514, Mar. 1999.
- [16] U. M. Maurer and S. Wolf, "Information-theoretic key agreement: From weak to strong secrecy for free," in *Proc. EUROCRYPT 2000 (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 2000, pp. 352–368.
- [17] R. Renner and S. Wolf, "New bounds in secret-key agreement: The gap between formation and secrecy extraction," in *Proc. EUROCRYPT 2003 (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 2003, vol. 2656, pp. 526–577.
- [18] A. D. Wyner, "The wiretap channel," *Bell Syst. Tech. J.*, vol. 54, pp. 1355–1387, 1975.