Secure Computing

Himanshu Tyagi*, Prakash Narayan* and Piyush Gupta†

*Dept. of Electrical and Computer Engineering
and Institute for Systems Research
University of Maryland
College Park, MD 20742, USA
Email: {tyagi, prakash}@umd.edu

†Bell Labs, Alcatel-Lucent
Murray Hill, NJ 07974, USA
Email: pgupta@research.bell-labs.com

***Abstract*—We study a problem of secure computation by multiple parties of a given function of their cumulative observations, using public communication but without revealing the value of the function to an eavesdropper with access to this communication. A Shannon theoretic formulation is introduced to characterize necessary and sufficient conditions for secure computability. Drawing on innate connections of this formulation to the problem of secret key generation by the same parties using public communication, we show that a function is securely computable if and only if its entropy is smaller than the secret key capacity. Conditions for secure computability at a lone terminal are also derived by association with an appropriate secret key generation problem.**

I. Introduction

Suppose that terminals $1, \ldots, m$ observe correlated signals and are required to compute "securely" a given function g of all the signals. To this end, following their observations, all the terminals are allowed to communicate interactively over a public noiseless channel of unlimited capacity, with all such communication being observed by all the terminals. The terminals seek to compute g in such a manner as to keep its value information theoretically secret from an eavesdropper that observes the public interterminal communication. See Fig. 1. A typical application arises in a wireless network of colocated sensors which must compute a given function of their correlated measurements using public communication that does not give away the value of the function.

Our goal is to characterize the necessary and sufficient conditions under which such secure computation is feasible. We formulate a new Shannon theoretic multiterminal source model that addresses the question: *When can a function g be computed so that its value is independent of the public communication used in its computation*? The answer to this question is innately connected to the problem of secret key (SK) generation in which the same terminals seek to generate "secret common randomness" at the largest rate possible, by means of public communication from which an eavesdropper can glean only a negligible amount of information [12], [1], [6], [7]. The largest rate of such a SK, which can be used for encrypted communication, is the SK capacity C_S. Since a securely computable function g will yield a SK of rate equal to its entropy H, it is clear that g necessarily must satisfy $H \leq C_S$. Surprisingly, $H < C_S$ is a sufficient condition for the existence of a protocol for the secure computation of g, constituting our main contribution.

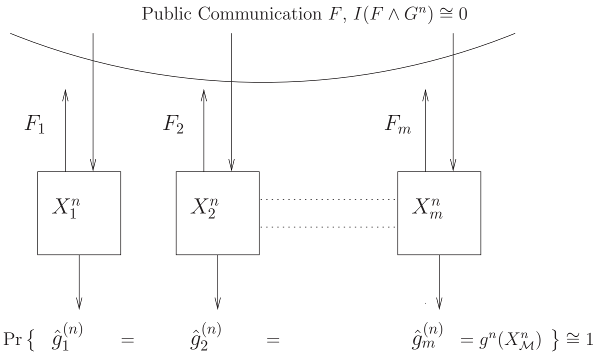
Public Communication F, $I(F \wedge G^n) \cong 0$
F_1
F_2
F_m
X_1^n
X_2^n
X_m^n
$\Pr\{\ \hat{g}_1^{(n)} = \hat{g}_2^{(n)} = \hat{g}_m^{(n)} = g^n(X_{\mathcal{M}}^n)\ \} \cong 1$

Fig. 1. Secure computation of g

Unlike in an SK generation model where the key must be shared by at least two terminals, the problem of the secure computation of g by a single terminal with the cooperation of the others, is also of interest. Our second contribution relates secure computability in this circumstance to an appropriately identified SK generation model.

We do not tackle the difficult problem of determining the minimum rate of public communication needed for the secure computation of g, which remains open even in the absence of a secrecy constraint [10]. Nor do we construct efficient protocols for this purpose. Instead, our objective in this work is merely to find conditions for the *existence* of such protocols.

The study of problems of function computation, with and without secrecy requirements, has a long and varied history to which we can make only a skimpy allusion here. Examples include: algorithms for exact function computation by multiple parties (cf. e.g., [18], [8], [9]); algorithms for asymptotically accurate (in observation length) function computation (cf. e.g., [16], [11]); exact function computation with secrecy (cf. e.g., [15]); and problems of oblivious transfer [14], [2].

A generalization of the problems considered here entails secure computation by a given subset $\mathcal{A}$ of the set of all terminals $\{1, ..., m\}$, with the remaining terminals serving as helpers. This general problem has been solved recently [17]. In contrast to the case $\mathcal{A} = \{1, ..., m\}$, an analogous necessary condition $H \leq C_S(\mathcal{A})$ where $C_S(\mathcal{A})$ is the secret key capacity for $\mathcal{A}$ [6], is no longer sufficient. Instead, a new secret key generation model is involved, with additional side information provided to the helper terminals.

Preliminaries and the problem formulation are contained in section II. Our main results, with outlines of proofs, are in section III. A brief discussion follows in section IV.

II. Preliminaries

Let $X_1, \ldots, X_m$, $m \geq 2$, be rvs with finite alphabets $\mathcal{X}_1, \ldots, \mathcal{X}_m$, respectively. For any nonempty set $A \subseteq \mathcal{M} = \{1, \ldots, m\}$, we denote $X_A = (X_i, \ i \in A)$. Similarly, for real numbers $R_1, \ldots, R_m$, and $A \subseteq \mathcal{M}$, we denote $R_A = (R_i, \ i \in A)$. Further, we denote n i.i.d. repetitions of $X_{\mathcal{M}} = (X_1, \ldots, X_m)$ by $X_{\mathcal{M}}^n = (X_1^n, \ldots, X_m^n)$. With $\mathcal{X}_{\mathcal{M}} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_m$, let $g : \mathcal{X}_{\mathcal{M}} \to \mathcal{Y}$ be a given mapping where $\mathcal{Y}$ is a finite alphabet. For $n \geq 1$, with $\mathcal{X}_{\mathcal{M}}^n = \mathcal{X}_1^n \times \ldots \times \mathcal{X}_m^n$, the mapping $g^n : \mathcal{X}_{\mathcal{M}}^n \to \mathcal{Y}^n$ is defined by

$$g^n(x_{\mathcal{M}}^n) = (g(x_{11}, \ldots, x_{m1}), \ldots, g(x_{1n}, \ldots, x_{mn})),$$
$$x_{\mathcal{M}}^n = (x_1^n, \ldots, x_m^n) \in \mathcal{X}_{\mathcal{M}}^n.$$

Following [6], given $\epsilon > 0$, for rvs U, V, we say that U is ϵ-*recoverable* from V if $\Pr\{U \neq f(V)\} \leq \epsilon$ for some function $f(V)$ of V. All logarithms and exponentials are with respect to the base 2.

We shall consider a multiterminal source model for secure computation with public communication; such a model was introduced in [6] in the context of SK generation with public transaction. Terminals $1, \ldots, m$ observe, respectively, the sequences $X_1^n, \ldots, X_m^n$, of observation length n. Randomization at the terminals is permitted; we assume that terminal i generates a rv $U_i, \ i \in \mathcal{M}$, such that $U_1, \ldots, U_m$ and $\mathcal{X}_{\mathcal{M}}^n$ mutually independent. The terminals are allowed to communicate over a noiseless public channel, possibly interactively in several rounds. Formally, assuming without any loss of generality that the communication of the terminals in $\mathcal{M}$ occurs in consecutive time slots in r rounds, such communication is described in terms of the mappings

$$f_{11}, \ldots, f_{1m}, f_{21}, \ldots, f_{2m}, \ldots, f_{r1}, \ldots, f_{rm},$$

with f_{ji} corresponding to a message in time slot j by terminal i, $1 \leq j \leq r$, $1 \leq i \leq m$; in general, f_{ji} is allowed to yield any function of (U_i, X_i^n) and of previous communication described in terms of $\{f_{kl} : k < j, \ l \in \mathcal{M} \text{ or } k = j, \ l < i\}$. The corresponding rvs representing the communication will be depicted collectively as

$$\mathbf{F} = \{F_{11}, \ldots, F_{1m}, F_{21}, \ldots, F_{2m}, \ldots, F_{r1}, \ldots, F_{rm}\}.$$

Definition 1. We say that g is ϵ_n-*securely computable* (ϵ_n-SC) by the terminals in $\mathcal{M}$ from observations of length n, randomization $U_{\mathcal{M}}$ and public communication $\mathbf{F} = \mathbf{F}^{(n)}$, if
(i) g^n is ϵ_n- recoverable from $(U_i, X_i^n, \mathbf{F})$ for every $i \in \mathcal{M}$, i.e., there exists $\hat{g}_i^{(n)}$ satisfying

$$\Pr\left\{\hat{g}_i^{(n)}(U_i, X_i^n, \mathbf{F}) \neq g^n(X_{\mathcal{M}}^n)\right\} \leq \epsilon_n, \quad i \in \mathcal{M}, \tag{1}$$

and
(ii) g^n satisfies the"strong" secrecy condition[1]

$$I(G^n \wedge \mathbf{F}) \leq \epsilon_n, \tag{2}$$

where $G^n = g^n(X_{\mathcal{M}}^n)$.

By definition, an ϵ_n-SC function g is recoverable (as g^n) at the terminals in $\mathcal{M}$ and is effectively concealed from an eavesdropper with access to the public communication $\mathbf{F}$.

Definition 2. We say that g is *securely computable* if g is ϵ_n- SC by $\mathcal{M}$ from observations of length n, suitable randomization $U_{\mathcal{M}}$ and public communication $\mathbf{F}^{(n)}$, such that $\lim\limits_n \epsilon_n = 0$.

We seek to answer the following question: When is a given g securely computable? The answer will be seen to be linked inherently to the concept of SK capacity for a multiterminal source model [6], [7].

Definition 3. [6], [7] A function K of $X_{\mathcal{M}}^n$ is an ϵ_n-*secret key* (ϵ_n-SK) for the terminals $\mathcal{M}$, achievable from observations of length n, randomization $U_{\mathcal{M}}$ and public communication $\mathbf{F}$ if
(i) K is ϵ_n-recoverable from $(U_i, X_i^n, \mathbf{F})$ for every $i \in \mathcal{M}$;
(ii) K satisfies the "strong" secrecy condition

$$\log|\mathcal{K}| - H(K \mid \mathbf{F}) = \log|\mathcal{K}| - H(K) + I(K \wedge \mathbf{F}) \leq \epsilon_n, \tag{3}$$

where $\mathcal{K} = \mathcal{K}^{(n)}$ denotes the set of possible values of K. The SK capacity $C_S(X_{\mathcal{M}})$ for $\mathcal{M}$ is the largest rate $\lim\limits_n (1/n)\log|\mathcal{K}^{(n)}|$ of ϵ_n-SKs for $\mathcal{M}$ as above, such that $\lim\limits_n \epsilon_n = 0$.

Remark. The secrecy condition (3) 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 $\mathbf{F}$ (i.e., $I(K \wedge \mathbf{F})$ is small).

A single-letter characterization of the SK capacity $C_S(X_{\mathcal{M}})$ is provided in [6], [7].

Theorem 1. [6], [7] *The SK capacity $C_S(X_{\mathcal{M}})$ equals*

$$C_S(X_{\mathcal{M}}) = H(X_{\mathcal{M}}) - R_{CO}(X_{\mathcal{M}}), \tag{4}$$

[1] The notion of strong secrecy for SK generation was introduced in [13], and developed further in [4], [5].

where

$$R_{CO}(X_{\mathcal{M}}) = \min_{R_{\mathcal{M}} \in \mathcal{R}(X_{\mathcal{M}})} \sum_{i=1}^{m} R_i \tag{5}$$

with

$$\mathcal{R}(X_{\mathcal{M}}) = \left\{ R_{\mathcal{M}} : \sum_{i \in B} R_i \geq H(X_B \mid X_{B^c}), B \subsetneq \mathcal{M} \right\}. \tag{6}$$

Furthermore, the SK capacity can be achieved with noninteractive communication and without recourse to randomization at the terminals in $\mathcal{M}$.

We note from [6] that $R_{CO}(X_{\mathcal{M}})$ has the operational significance of being the smallest rate of "communication for omniscience" for $\mathcal{M}$, namely the smallest rate $\lim_n (1/n) \log \|\mathbf{F}^{(n)}\|$ of suitable communication for the terminals in $\mathcal{M}$ whereby $X_{\mathcal{M}}^n$ is ϵ_n-recoverable from $(U_i, X_i^n, \mathbf{F}^n)$ at each terminal $i \in \mathcal{M}$, with $\lim_n \epsilon_n = 0$; here $\|\mathbf{F}^{(n)}\|$ denotes the cardinality of the set of values of $\mathbf{F}^{(n)}$. Thus, $R_{CO}(X_{\mathcal{M}})$ is the smallest rate of interterminal communication that enables every terminal in $\mathcal{M}$ to reconstruct all the sequences observed by all the other terminals. This notion of omniscience, which played a central role in SK generation for the multiterminal source model [6], will play a useful role in the secure computation of g, as well.

A comparison of the secrecy conditions in (2) and (3) that a securely computable g and an SK K must meet, respectively, shows that the latter is required additionally to be of near uniform distribution. However, if g is securely computable, then g^n can be rendered also to be nearly uniformly distributed for all n sufficiently large, for instance, by near-lossless data compression at each terminal $i \in \mathcal{M}$ of $\hat{g}_i^{(n)}$ (cf. Definition 1). The corresponding rvs, also of rate approximately equal to $H(G)$, clearly yield a SK for $\mathcal{M}$. Thus, for a given g to be securely computable, it is necessary that

$$H(G) \leq C_S(X_{\mathcal{M}}). \tag{7}$$

Remarkably, it transpires that $H(G) < C_S(X_{\mathcal{M}})$ is a sufficient condition for g to be securely computable, and constitutes our main result below.

A special case of secure computation arises when g must be securely computable at a sole terminal in $\mathcal{M}$, say terminal 1. This means that g is as in Definitions 1, 2, but with a requirement of recoverability *only* at terminal 1 replacing that at *every* $i \in \mathcal{M}$. This formulation does not relate immediately to that of SK generation (which is meaningful only when an SK is shared by at least two terminals). Nonetheless, by considering a model with $m = 2$ terminals, we show that conditions for g to be securely computable at only terminal 1 correspond to an appropriately identified SK generation problem.

III. Results

We begin with our main result which characterizes when a function g is securely computable by the terminals in $\mathcal{M}$.

Theorem 2. *A function g is securely computable by $\mathcal{M}$ if*

$$H(G) < C_S(X_{\mathcal{M}}). \tag{8}$$

Furthermore, under (8), g is securely computable with noninteractive communication and without recourse to randomization at the terminals.

Conversely, if g is securely computable by $\mathcal{M}$, then $H(G) \leq C_S(X_{\mathcal{M}})$.

Outline of proof: The necessity of $H(G) \leq C_S(X_{\mathcal{M}})$ for g to be securely computable has been outlined already in (7).

The achievability part consists of showing the existence of *noninteractive* source codes which enable ominiscience at all the terminals in $\mathcal{M}$, and thereby the computation of g. Furthermore, the corresponding codewords are selected so as to be simultaneously independent of G^n, thus assuring security. Formally, consider random mappings

$$F_i : \mathcal{X}_i^n \to \{1, \ldots \lceil \exp(nR_i) \rceil\}, \quad i \in \mathcal{M}, \tag{9}$$

with $R_{\mathcal{M}} = (R_1, \ldots, R_m) \in \mathcal{R}(X_{\mathcal{M}})$ in (6) chosen such that $\sum_{i=1}^m R_i \cong R_{CO}(X_{\mathcal{M}})$ in (5). By [6, Proposition 1], these mappings yield, with high probability for all n sufficiently large, the existence of codes for omniscience for the terminals in $\mathcal{M}$, whereupon g can be computed by them. Next, using condition (8) in conjunction with a variation of the "Balanced Coloring" lemma [1, Lemma 3.1], [6, Lemma B.2], stated in the appendix as Lemma BC, we show that $F_{\mathcal{M}} = (F_1, \ldots, F_m)$ is nearly independent of G^n with high probability. Specifically, we show that with high probability

$$I(F_i \wedge G^n, F_{\mathcal{M} \setminus \{i\}}) < \epsilon, \quad i \in \mathcal{M}, \tag{10}$$

which implies the smallness of

$$I(F_{\mathcal{M}} \wedge G^n) \leq \sum_{i}^{m} I(F_i \wedge G^n, F_{\mathcal{M} \setminus \{i\}}),$$

guaranteeing security. We mention that the main step in (10) is to show that

$$\Pr\left\{X_i^n \mid G^n, F_{\mathcal{M} \setminus \{i\}}\right\} = \Pr\left\{X_{\mathcal{M}}^n \mid G^n, F_{\mathcal{M} \setminus \{i\}}\right\}$$

with high probability, which is used then to specify the set $\mathcal{U}_0$ in Lemma BC. □

The sufficiency condition for secure computability of g in (8) raises the following natural question: Does the difference $C_S(X_{\mathcal{M}}) - H(G)$ possess an operational significance? The answer is in the affirmative. Indeed, the terminals in $\mathcal{M}$, over and above securely computing g with public communication $\mathbf{F}'$ (say), can generate simultaneously a SK $K_g = K_g^{(n)}$, possibly with additional public communication $\mathbf{F}''$ (say), satisfying Definition 3 with $(\mathbf{F}', G^n, \mathbf{F}'')$ in the role of $\mathbf{F}$. Denote the largest rate $\lim_n (1/n) \log |\mathcal{K}_g^{(n)}|$ of such a SK by $C_S^g(X_{\mathcal{M}})$.

Theorem 3. *It holds that*

$$C_S^g(X_{\mathcal{M}}) = C_S(X_{\mathcal{M}}) - H(G).$$

Remark. (i) Given a securely computable g with $H(G) < C_S(X_{\mathcal{M}})$, Theorem 3 says, in effect, that an optimum rate SK for $\mathcal{M}$ can be generated with two nearly independent parts: G^n and $K_g^{(n)}$.
(ii) Theorems 1 and 3 lead to the observation

$$H(X_{\mathcal{M}}) = R_{CO}(X_{\mathcal{M}}) + H(G) + C_S^g(X_{\mathcal{M}}),$$

which admits the following heuristic interpretation. The "total randomness" $X_{\mathcal{M}}^n$ that corresponds to omniscience decomposes into three "nearly mutually independent" components: a minimum-sized communication for omniscience and the independent parts of the optimum-rate SK just mentioned.

Outline of proof: The proof is based on the observation that the omniscience obtained in Theorem 2 with public communication $F_{\mathcal{M}}$ that is independent of G^n, constitutes "secure common randomness" for $\mathcal{M}$ which can be used further to generate the SK $K_g^{(n)}$ by applying Lemma BC. □

Next, we turn to the problem of secure computability of g at terminal 1 alone. Our partial results are for the case $m = 2$. This model, in the context of computability with interactive communication but without secrecy, has been studied in [16].

Theorem 4. *Consider the model with* $m = 2$. *Then,* g *is securely computable at terminal* 1 *if*

$$H(X_2 \mid X_1) < H(X_2 \mid G). \tag{11}$$

Conversely, if g *is securely computable at terminal* 1, *then*

$$H(X_2 \mid X_1) \leq H(X_2 \mid G). \tag{12}$$

Remark. Observe that (12) is equivalent to

$$H(G) \leq I(X_1 \wedge X_2, G) \tag{13}$$

which suggests a connection to a problem of SK generation since the right side above corresponds to the SK capacity of a modified source model with two terminals that have access, respectively, to i.i.d. repetitions of the rvs X_1 and (X_2, G).

Proof: In order to avoid the trivial, we assume that $H(X_2 \mid X_1) > 0$ and $H(G \mid X_1) > 0$. The converse part (12) is easy. Let g be securely computable at terminal 1. Consider an augmented model with $X_1' = X_1$ and $X_2' = (X_2, G)$. Clearly, g must be securely computable also for the augmented model and, hence, by Theorem 2, must satisfy

$$H(G) \leq I(X_1 \wedge X_2, G) \tag{14}$$

which is (12).

The achievability part entails showing that (11) is sufficient for the the existence of a Slepian-Wolf code for reconstructing X_2^n near-losslessly at terminal 1 using public communication (the Slepian-Wolf codeword) that is nearly independent of G^n. By (11), pick $0 < H(X_2 \mid X_1) < R < H(X_2 \mid G)$. Consider the random mapping $F : \mathcal{X}_2^n \to \{1, \ldots, \lceil \exp(nR) \rceil\}$ such that the rvs $F(x_2^n),\ x_2^n \in \mathcal{X}_2^n$, are mutually independent and uniform on $\{1, \ldots, \lceil \exp(nR) \rceil\}$. Given $\epsilon > 0$, by the Slepian-Wolf theorem (cf. e.g, [3, Lemma 3.1.13, pp. 252-253]), F yields an (n, ϵ)-source code with large probability, i.e.,

$$\Pr\left\{\phi(F(X_2^n), X_1^n) = X_2^n\right\} \geq 1 - \epsilon, \tag{15}$$

for some decoder $\phi = \phi_F^{(n)}$. Furthermore, using Lemma BC in the appendix with $U = X_2^n$, $V = G^n$, $h =$ constant and $d = \exp[n(H(X_2 \mid G) - \epsilon)]$, we get that

$$I(F(X_2^n) \wedge G^n) < \epsilon, \tag{16}$$

with probability approaching one as n tends to infinity. Therefore, for all n sufficiently large, there exists an encoder $f : \mathcal{X}_2^n \to \{1, \ldots, \lceil \exp(nR) \rceil\}$ and a decoder $\phi : \{1, \ldots, \lceil \exp(nR) \rceil\} \times \mathcal{X}_1^n \to \mathcal{X}_2^n$ satisfying (15) and (16) with f replacing F. Thus, terminal 1, having recovered X_2^n as above, can compute g and do so securely. □

Returning to secure computation by all the terminals in $\mathcal{M}$, observe in the proof of Theorem 2 that g was securely computed from *omniscience at all the terminals in* $\mathcal{M}$ that was attained using *noninteractive* public communication. However, omniscience is not necessary for the secure computation of g, and it is possible to make do with communication of rate less than $R_{CO}(X_{\mathcal{M}})$ using an *interactive* protocol. This is illustrated by the following example.

Example. Consider the secure computation of g by $m = 2$ terminals. In order to avoid the trivial, assume that $H(G|X_1) > 0$. The condition $H(G) < I(X_1 \wedge X_2)$ in (8) yields $H(G) < I(X_1 \wedge X_2) + H(G \mid X_2)$ which is equivalent to (11). It follows as in the proof of Theorem 4 that terminal 1 can reconstruct X_2^n, and therefore compute g^n, from public communication F_2 of rate $\cong H(X_2 \mid X_1)$ from terminal 2 (corresponding to F in that proof) that is nearly independent of G^n. Further, using Lemma BC in the appendix, terminals 1 and 2 can generate a SK $K = K(X_2^n)$ of rate $\cong H(X_2 \mid G) - H(X_2 \mid X_1)$ and with arbitrary small $I(K \wedge G^n, F_2)$ for all n sufficiently large.

Next, terminal 1 constructs a Slepian-Wolf codeword F_g for g with X_2^n as side-information, of rate $\cong H(G \mid X_2)$ and communicates to terminal 2 its encrypted version

$$F_1 = F_g + K \mod 2$$

(all represented in bits), with encryption feasible if

$$H(G \mid X_2) < H(X_2 \mid G) - H(X_2 \mid X_1),$$

a condition equivalent to (8). Terminal 2 first decrypts F_1 using K to recover F_g, and thereby also g^n using X_2^n.

The computation of g^n is secure since

$$I(G^n \wedge F_1, F_2) = I(G^n \wedge F_2) + I(G^n \wedge F_1 \mid F_2)$$

is small; specifically, the first term is negligible as noted above, while the second term is bounded by

$$\begin{aligned} I(G^n \wedge F_1 \mid F_2) &= H(F_g + K \mid F_2) - H(F_g + K \mid G^n, F_2) \\ &\leq H(K) - H(F_g + K \mid G^n, F_2) \\ &= I(K \wedge G^n, F_2), \end{aligned}$$

whose smallness too has been noted.

Observe that the communication F_1 from terminal 1 is interactive as it depends on the communication F_2 from terminal 2. Also, omniscience at both the terminals is not attained. While terminal 1 reconstructs X_2^n, terminal 2 cannot reconstruct X_1^n because the communication F_1 has rate $\cong H(G \mid X_2) < H(X_1 \mid X_2)$; the overall rate of public communication falls short of $R_{CO}(X_1, X_2) = H(X_1 \mid X_2) + H(X_2 \mid X_1)$. □

IV. Discussion

We have considered an elemental multiterminal source model of secure computation with public communication in which we seek conditions for computability with secrecy from an eavesdropper with access to only this communication. We obtain simple necessary and sufficient conditions on function entropy and SK capacity of an associated SK generation model that, in particular, do not involve auxiliary rvs.

A problem of considerably more difficulty is that of characterizing the minimum rate of public communication that is needed for secure computation. Similarly difficult is the practically important situation in which a function must be computed with secrecy from an eavesdropper that has access to "wiretapped" side information in addition to the public interterminal communication. These problems remain open, in general, as do single-letter answers to their counterparts in SK generation.

Appendix

Our proofs call for a balanced coloring of a set corresponding to a rv that differs from another rv for which probability bounds are used. However, both rvs agree with high probability when conditioned on a set of interest.

Consider rvs U, U', V with values in finite sets $\mathcal{U}, \mathcal{U}', \mathcal{V}$, respectively, and a mapping $h : \mathcal{U} \to \{1, \ldots, r'\}$. For $\lambda > 0$, let $\mathcal{U}_0$ be a subset of $\mathcal{U}$ such that
(i) $\Pr\{U \in \mathcal{U}_0\} > 1 - \lambda^2$;
(ii) given $U \in \mathcal{U}_0, h(U) = j, V = v, U' = u'$, there exists $u = u(u') \in \mathcal{U}$ satisfying

$$\begin{aligned}&\Pr\{U = u \mid h(U) = j, V = v, U \in \mathcal{U}_0\}\\ &\quad = \Pr\{U' = u' \mid h(U) = j, V = v, U \in \mathcal{U}_0\},\end{aligned}$$

for all $1 \leq j \leq r$ and $v \in \mathcal{V}$. Then the following holds.

Lemma BC. *(Balanced Coloring) Let the rvs U, U', V and the set $\mathcal{U}_0$ be as above. Further, assume that*

$$\mathrm{P}_{UV}\left(\left\{(u, v) : \Pr\{U = u \mid V = v\} > \frac{1}{d}\right\}\right) \leq \lambda^2.$$

Then, a randomly selected mapping $\phi : \mathcal{U}' \to \{1, \ldots, r\}$ fails to satisfy

$$\sum_{j=1}^{r'} \sum_{v \in \mathcal{V}} \sum_{i=1}^{r} \Pr\{h(U) = j, V = v\} \times \\ \left| \sum_{u' \in \mathcal{U}' : \phi(u') = i} \Pr\{U' = u' \mid h(U) = j, V = v\} - \frac{1}{r} \right| < 12\lambda,$$

with probability less than $2rr'|\mathcal{V}| \exp\left(-\frac{c\lambda^3 d}{rr'}\right)$ for a constant c.

Acknowledgements

The authors thank Sirin Nitinawarat for helpful discussions. The work of H. Tyagi and P. Narayan was supported by the U.S. National Science Foundation under Grants CCF0635271 and CCF0830697. P. Gupta acknowledges support from NSF Grant CNS-519535.

A part of this work was conducted during H. Tyagi's summer internship in 2009 at Bell Labs, Alcatel-Lucent, Murray Hill, NJ 07974, USA.

References

[1] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography–part i: Secret sharing," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1121–1132, 1993.

[2] R. Ahlswede and I. Csiszár, "On the oblivious transfer capacity," *ISIT, Proceedings of the IEEE International Symposium on Information Theory*, pp. 2061–2064, June 2007.

[3] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Channels*. Academic Press, 1981.

[4] I. Csiszár, "Almost independence and secrecy capacity," *Prob. Pered. Inform.*, vol. 32, no. 1, pp. 48–57, 1996.

[5] I. Csiszár and P. Narayan, "Common randomness and secret key generation with a helper," *IEEE Trans. Inform. Theory*, vol. 46, pp. 344–366, March 2000.

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

[7] ——, "Secrecy capacities for multiterminal channel models," *IEEE Trans. Inform. Theory*, vol. 54, no. 6, pp. 2437–2452, 2008.

[8] R. G. Gallager, "Finding parity in a simple broadcast nework," *IEEE Trans. Inform. Theory*, vol. 34, no. 2, pp. 176–180, 1988.

[9] A. Giridhar and P. Kumar, "Computing and communicating functions over sensor networks," *IEEE Journ. on Select. Areas in Commun.*, vol. 23, no. 4, pp. 755– 764, 2005.

[10] J. Körner and K. Marton, "How to encode the modulo-two sum of binary sources," *IEEE Trans. Inform. Theory*, vol. 25, no. 2, pp. 219– 221, 1979.

[11] N. Ma, P. Ishwar and P. Gupta, "Information-theoretic bounds for multiround function computation in collocated networks," *IEEE International Symposium on Information Theory*, Coex, Seoul, Korea, June - July 2009.

[12] U. M. Maurer, "Secret key agreement by public discussion from common information," *IEEE Trans. Inform. Theory*, vol. 39, pp. 733–742, May 1993.

[13] U. M. Maurer, *Communications and Cryptography: Two sides of One Tapestry*, R.E. Blahut et al., Eds. Norwell, MA: Kluwer, 1994, ch. 26, pp. 271–285.

[14] A. Nascimento and A. Winter, "On the oblivious transfer capacity of noisy correlations," *IEEE International Symposium on Information Theory*, pp. 1871–1875, July 2009.

[15] A. Orlitsky and A. E. Gamal, "Communication with secrecy constraints," *STOC*, pp. 217–224, 1984.

[16] A. Orlitsky and J. R. Roche, "Coding for computing," *IEEE Trans. Inform. Theory*, vol. 47, no. 3, pp. 903 – 917, 2001.

[17] H. Tyagi, P. Narayan and P. Gupta, "Secure Computing," *IEEE Trans. Inform. Theory*, to be submitted.

[18] A. C. Yao, "Some complexity questions related to distributive computing," *Proc. 11th Ann. Symp. on Theory of Computing*, pp. 209–213, May 1979.