On Unconditionally Secure Multi-Party Sampling from Scratch

Ye Wang and Prakash Ishwar
Boston University
Boston, MA.
Email: {yw,pi}@bu.edu

***Abstract*—THIS PAPER IS ELIGIBLE FOR THE STUDENT PAPER AWARD. In the problem of secure multi-party sampling, n parties wish to securely sample an n-variate joint distribution, with each party receiving a sample of one the correlated variables. The objective is to correctly produces the samples using a distributed message passing protocol, while maintaining privacy against a coalition of passively cheating parties. In the two-party case, we fully characterize the joint distributions that can be securely sampled under perfect correctness and privacy requirements as well as under weakened correctness and privacy requirements. Furthermore, we show that the distributions that can be securely sampled can be produced with a protocol that only uses one round of unidirectional communication. For the n-party case, any distribution can be securely sampled with privacy against a strict minority coalition, due to well-known results in secure multi-party computation. However, when privacy against a majority coalition is required, not all distributions can be securely sampled. We give necessary conditions and sufficient conditions for distributions that can be securely sampled. However, the exact characterization of the distributions that can be securely sampled remains open.**

***Index Terms*—secure multi-party computation, secure sampling, unconditional security, common information**

I. Introduction

An important subclass of secure multi-party computation is the problem of secure multi-party sampling. In this problem, n parties wish to securely sample a set of n jointly distributed random variables, with each party uniquely obtaining one of these n variables. The objective is to correctly realize the desired joint distribution while ensuring that any coalition of up to c parties does not learn anything more about the other parties' samples than what can be inferred from the coalition's set of samples. This notion of privacy is formulated in an information theoretic sense, requiring unconditional security against parties with unbounded computational power. In order to achieve this objective, the parties execute a distributed protocol, in which they are allowed unlimited noiseless communication over multiple interactive rounds. The parties execute this protocol "from scratch", meaning that, while they can generate and utilize an unlimited amount of local independent randomness, they do not have access to any initial "setup" of correlated randomness. We assume that the parties are "semi-honest", that is, they will correctly execute the protocol and only passively attempt to extract information about other parties' samples.

In this work, we present results toward the characterization of the region of joint distributions that can be securely sampled. Analogously to the work of [1], which characterizes the boolean-valued functions that can be securely computed from scratch, we aim to characterize the joint distributions that can be securely sampled from scratch. Outer bounds for secure sampling also have implications for the outer bounds for specific secure multi-party computation problems (see [2]). The characterization of these regions address the question of feasibility. The questions of protocol efficiency and complexity are not the primary focus of this work, although to demonstrate the feasibility of some distributions, we construct protocols that use only one round of unidirectional communication.

For the two-party case, we fully characterize the region of distributions that can be securely sampled. We show that a pair of random variables can be securely sampled with perfect correctness and perfect privacy if, and only if, the common information of these variables is equal to their mutual information. Weakening the correctness or privacy requirements, increases this region corresponding to the degree of weakening. Furthermore, these distributions can be securely sampled by a protocol that uses only one round of unidirectional communication.

In the n-party case, the secure computability of any function (see [3] and [4]) with perfect privacy against coalitions of size up to $c < n/2$ implies that any n-variate joint distribution can be securely sampled with ϵ-correctness (for any $\epsilon > 0$) and perfect privacy against coalitions of size up to $c < n/2$. The parties can generate shares of a private common randomness and then compute the desired samples as a secure function of that common randomness. However, this technique is not applicable when privacy is required against coalitions of size $c \geq n/2$. For the regime of $c \geq n/2$, many distributions cannot be securely sampled and we present necessary conditions and sufficient conditions for distributions that can be securely sampled.

The problem of secure two-party sampling was first introduced in the specific form of "mental poker" in [5]. In the mental poker problem, Alice and Bob wish to sample random variables that simulate the act of randomly drawing cards from a shared deck, while ensuring that their drawn cards remain private from the other party. The impossibility of this task when unconditional security is required was proven and a cryptographic solution was proposed in [5].

The recent work of [2] and [6] has considered the general secure two-party sampling problem, with the modification that the parties have access to a "setup" of correlated random variables in order to assist them in generating the desired variables, and requiring perfect correctness and perfect privacy. Monotones for the secure two-party sampling problem, which are information measures that can only decrease from the measurement of the setup to the measurement of the output of the protocol, were presented in [2] and used as a technique for developing outer bounds. The notion of monotones was extended to monotone regions in [6], providing a tighter outer bound technique for the general problem. The outer bound of [2], when specialized to the scenario of secure sampling from scratch (i.e., using an independent setup), matches the two-variate distribution region that we characterize for perfect correctness and perfect privacy. Hence in this scenario, we have shown that their outer bound is tight.

II. Problem Formulation

The desired distribution is an n-variate joint distribution $P_{X_1,\ldots,X_n}$ over the finite alphabets $\mathcal{X}_1 \times \ldots \times \mathcal{X}_n$. The objective is to construct an n-party protocol that correctly and privately produce samples $(\hat{X}_1,\ldots,\hat{X}_n) \in \mathcal{X}_1 \times \ldots \times \mathcal{X}_n$, with party i generating $\hat{X}_i$.

A protocol may involve multiple rounds of error-free, interactive communication, with local random computation performed between rounds. We describe a very general family of permissible protocols to allow for stronger converse statements.

At the beginning of a protocol, each party i generates an arbitrary, independent random seed R_i, where $R_1,\ldots,R_n \sim P_{R_1,\ldots,R_n} = \prod_{i=1}^n P_{R_i}$.

Then, the parties interact for t rounds of communication, where in each round $k \in \{1,\ldots,t\}$, a message is sent between each pair of parties $i,j \in \{1,\ldots,n\}$, $i \neq j$, produced as a function of the sender's random seed and all messages received by the sender in previous rounds,

$$M_k(i,j) = f_{i,j,k}(R_i, M^{k-1}(*,i)),$$

where

$$M^{k-1}(*,i) := \{M_1(j,i),\ldots,M_{k-1}(j,i)\}_{i\neq j}$$

denotes every message received by party i prior to round k. Note that during a round, messages may even be sent simultaneously in both directions between a particular pair of parties. Also, any of these messages may be null, in order to capture simpler patterns of communication. In a particular round, if between each pair of parties, at least one of the messages sent in either direction is null, then we say that round uses unidirectional communication.

After t rounds of interaction, each party produces its sample as a function of its random seed and its received messages,

$$\hat{X}_i = g_i(R_i, M^t(*,i)).$$

Note that this computation can also make use of all of the messages sent by the party since the sent messages are only a function of the random seed and received messages.

From here on, we will use the following shorthand notation for messages over all t rounds. The set of all messages received by party i is denoted by

$$M_i := M^t(*,i),$$

and the set of all messages exchanged over all t rounds is denoted by

$$M = M_1,\ldots,M_n.$$

Correctness: A protocol is ϵ-correct (for $\epsilon \geq 0$) if the distribution of the produced samples is close to the desired distribution, that is,

$$d(P_{\hat{X}_1,\ldots,\hat{X}_n}, P_{X_1,\ldots,X_n}) \leq \epsilon,$$

where $d(P_U,P_V)$ is the variational distance between distributions P_U and P_V on the same alphabet $\mathcal{U}$:

$$\begin{aligned} d(P_U,P_V) &= \frac{1}{2}\sum_{u\in\mathcal{U}} |P_U(u) - P_V(u)| \\ &= \frac{1}{2}\|P_U - P_V\|_1. \end{aligned}$$

Privacy: A protocol is (δ,c)-private (for $\delta \geq 0$ and $c \in \{1,\ldots,n-1\}$), if

$$\sum_{T\subset\{1,\ldots,n\}:|T|\leq c} I\big(\{R_i,M_i\}_{i\in T};\{\hat{X}_i\}_{i\notin T}|\{\hat{X}_i\}_{i\in T}\big) \leq \delta.$$

In the two-party case ($n=2$), only coalitions of size $c=1$ are interesting, hence we simply say δ-private.

We say that an n-variate distribution $P_{X_1,\ldots,X_n}$ can be securely sampled with ϵ-correctness and (δ,c)-privacy if and only if there exists an n-party protocol that is both ϵ-correct and (δ,c)-private. The qualifiers of perfect correctness, perfect privacy, or perfect security respectively denote the cases when $\epsilon = 0$, $\delta = 0$, or both.

III. Main Results

The structure of the protocol implies the following property on the samples produced and the messages exchanged.

Lemma 1 *For any distributed protocol, the samples produced are independent conditioned on all of the messages exchanged, that is,*

$$P_{\hat{X}_1,\ldots,\hat{X}_n|M} = \prod_{i=1}^n P_{\hat{X}_i|M}$$

Proof: Consider the distribution of $(\hat{X}_1,\ldots,\hat{X}_n,M)$, which can be expressed as

$$\begin{aligned} P_{\hat{X}^n,M}(x^n,m) &= \sum_{r^n} P_{\hat{X}^n,M,R^n}(x^n,m,r^n) \\ &= \sum_{r^n} P_{R^n}(r^n)P_{\hat{X}^n,M|R^n}(x^n,m|r^n) \\ &= \sum_{r^n} P_{R^n}(r^n)\psi(x^n,m,r^n), \end{aligned}$$

where $\psi(x^n, m, r^n)$ is equal to one if for all $k \in \{1, \ldots, t\}$ and $i, j \in \{1, \ldots, n\}$ with $i \neq j$,

$$\begin{aligned} m_k(i,j) &= f_{i,j,k}(r_i, m^{k-1}(*, i)), \\ x_i &= g_i(r_i, m^t(*, i)), \end{aligned}$$

and equal to zero otherwise. This indicator function can be factorized as

$$\psi(x^n, m, r^n) = \prod_{i=1}^{n} \psi_i(x_i, m, r_i),$$

where $\psi_i(x_i, m, r_i)$ is equal to one if for all $k \in \{1, \ldots, t\}$ and $j \in \{1, \ldots, n\}$ with $i \neq j$,

$$\begin{aligned} m_k(i,j) &= f_{i,j,k}(r_i, m^{k-1}(*, i)), \\ x_i &= g_i(r_i, m^t(*, i)), \end{aligned}$$

and equal to zero otherwise. Thus, $P_{\hat{X}^n, M}(x^n, m)$ can be factorized as

$$\begin{aligned} P_{\hat{X}^n, M}(x^n, m) &= \sum_{r^n} P_{R^n}(r^n) \psi(x^n, m, r^n) \\ &= \prod_{i=1}^{n} \phi_i(x_i, m), \end{aligned}$$

where

$$\phi_i(x_i, m) = \sum_{r_i} P_{R_i}(r_i) \psi_i(x_i, m, r_i).$$

Hence, given this factorization, we have that $\hat{X}_1, \ldots, \hat{X}_n$ are independent given M. ■

A. Common Information

The notion of the "common information" between a pair of random variables has been widely explored in the literature (see for example [7], [8], [9], [10]). Several definitions, each with an operational significance to a coding problem, have been characterized, including the Gács-Körner common information [7] and the Wyner common information [9].

The Wyner common information [9] plays a significant role in the characterization of the region of two-variate distributions that can be securely sampled, and is defined by

$$W(X_1; X_2) := \min_{Y : I(X_1; X_2 | Y) = 0} I(X_1, X_2; Y),$$

where the minimum can be obtained by $Y \in \mathcal{Y}$, with $|\mathcal{Y}| \leq |\mathcal{X}_1| \cdot |\mathcal{X}_2|$ [9]. The Wyner common information is always larger than the mutual informaion [9],

$$I(X_1; X_2) \leq W(X_1; X_2),$$

and the following Lemma characterizes when Wyner common information is within δ of the mutual information.

Lemma 2 *For all $\delta \geq 0$, $W(X_1; X_2) \leq I(X_1; X_2) + \delta$ if, and only if, there exists Y such that $I(X_1; X_2 | Y) = 0$ and $I(X_1; Y | X_2) + I(X_2; Y | X_1) \leq \delta$. For the "only if" direction, one can find $Y \in \mathcal{Y}$, with $|\mathcal{Y}| \leq |\mathcal{X}_1| \cdot |\mathcal{X}_2|$.*

Proof: This follows from the definition of $W(X_1; X_2)$, the following identity

$$\begin{aligned} & I(X_1, X_2; Y) - I(X_1; X_2) \\ &\quad = I(X_1; Y | X_2) + I(X_2; Y | X_1) - I(X_1; X_2 | Y), \end{aligned}$$

and the cardinality bound on the optimization. ■

B. The Two-Party Case

In the two-party case, we fully characterize the region of distributions that can be securely sampled with ϵ-correctness and δ-privacy.

Theorem 1 *A two-variate distribution P_{X_1, X_2} can be securely sampled with ϵ-correctness and δ-privacy if, and only if, there exists $P_{\hat{X}_1, \hat{X}_2}$ such that $d(P_{\hat{X}_1, \hat{X}_2}, P_{X_1, X_2}) \leq \epsilon$ and*

$$W(\hat{X}_1; \hat{X}_2) \leq I(\hat{X}_1; \hat{X}_2) + \delta.$$

Furthermore, if P_{X_1, X_2} can be sampled with ϵ-correctness and δ-privacy, then it can be done with a protocol that uses only one round of unidirectional communication.

Proof of the "only if" part: An ϵ-correct and δ-private protocol samples $P_{\hat{X}_1, \hat{X}_2}$ such that $d(P_{\hat{X}_1, \hat{X}_2}, P_{X_1, X_2}) \leq \epsilon$ and

$$I(R_1, M_1; \hat{X}_2 | \hat{X}_1) + I(R_2, M_2; \hat{X}_1 | \hat{X}_2) \leq \delta.$$

We have

$$\begin{aligned} & I(M; \hat{X}_1 | \hat{X}_2) + I(M; \hat{X}_2 | \hat{X}_1) \\ &\quad \leq I(R_1, M; \hat{X}_2 | \hat{X}_1) + I(R_2, M; \hat{X}_1 | \hat{X}_2) \\ &\quad \overset{(a)}{=} I(R_1, M_1; \hat{X}_2 | \hat{X}_1) + I(R_2, M_2; \hat{X}_1 | \hat{X}_2) \\ &\quad \overset{(b)}{\leq} \delta, \end{aligned}$$

where (a) holds because M_2 is a function of (R_1, M_1), M_1 is a function of (R_2, M_2), and $M = (M_1, M_2)$, and (b) follows from the assumption that the protocol is δ-private. By Lemma 1, $I(\hat{X}_1; \hat{X}_2 | M) = 0$. From Lemma 2, with $Y = M$, it follows that $W(X_1; X_2) \leq I(X_1; X_2) + \delta$.

Proof of the "if" part: If there exists $P_{\hat{X}_1, \hat{X}_2}$ such that $W(\hat{X}_1; \hat{X}_2) \leq I(\hat{X}_1; \hat{X}_2) + \delta$, then by Lemma 2, there exists a $Y \in \mathcal{Y}$, with $|\mathcal{Y}| \leq |\mathcal{X}_1| \cdot |\mathcal{X}_2|$, such that $I(\hat{X}_1; \hat{X}_2 | Y) = 0$ and $I(\hat{X}_1; Y | \hat{X}_2) + I(\hat{X}_2; Y | \hat{X}_1) \leq \delta$. We can design a protocol that uses only one-round of unidirectional communication as follows:

1) Party 1 generates $R_1 = (\hat{X}_1, Y) \sim P_{\hat{X}_1, Y}$.
2) In the single round of communication, party 1 sends the message $M_1(1,2) = Y$ to party 2, and party 2 does not send a message, i.e., $M_1(2,1)$ is null.
3) Party 1 outputs $\hat{X}_1$.
4) Party 2 independently generates $R_2 = \{\hat{X}_2(y)\}_{y \in \mathcal{Y}}$, where $\hat{X}_2(y) \sim P_{\hat{X}_2 | Y}(\cdot | y)$.
5) Party 2 outputs $\hat{X}_2 = g_2(R_2, Y) = \hat{X}_2(Y)$.

Note that these last two steps result in $(\hat{X}_2, Y) \sim P_{\hat{X}_2, Y}$. Since also $(\hat{X}_1, Y) \sim P_{\hat{X}_1, Y}$ and $I(\hat{X}_1; \hat{X}_2 | Y) = 0$, we have that

$(\hat{X}_1, \hat{X}_2, Y) \sim P_{\hat{X}_1, \hat{X}_2, Y}$. Hence, the protocol is ϵ-correct.

We now show that the protocol is also δ-private. We have

$$\begin{aligned} I(R_1, M_1; \hat{X}_2 | \hat{X}_1) &\overset{(c)}{=} I(\hat{X}_1, Y; \hat{X}_2 | \hat{X}_1) \\ &= I(\hat{X}_2; Y | \hat{X}_1), \end{aligned}$$

where equality (c) is because $R_1 = (\hat{X}_1, Y)$ and M_1 is null. We also have

$$\begin{aligned} I(R_2, M_2; \hat{X}_1 | \hat{X}_2) &\overset{(d)}{=} I(R_2, Y; \hat{X}_1 | \hat{X}_2) \\ &= H(\hat{X}_1 | \hat{X}_2) - H(\hat{X}_1 | \hat{X}_2, R_2, Y) \\ &\overset{(e)}{=} H(\hat{X}_1 | \hat{X}_2) - H(\hat{X}_1 | R_2, Y) \\ &\overset{(f)}{=} H(\hat{X}_1 | \hat{X}_2) - H(\hat{X}_1 | Y) \\ &\overset{(g)}{=} H(\hat{X}_1 | \hat{X}_2) - H(\hat{X}_1 | \hat{X}_2, Y) \\ &= I(\hat{X}_1; Y | \hat{X}_2), \end{aligned}$$

where (d) is because $M_2 = Y$, (e) is because $\hat{X}_2$ is a deterministic function of R_2 and Y, (f) is because R_2 is independent of $(\hat{X}_1, Y)$, and (g) is because $I(\hat{X}_1; \hat{X}_2 | Y) = 0$, which is a consequence of the assumptions of the “if” part of the theorem. From this it follows that

$$\begin{aligned} &I(R_1, M_1; \hat{X}_2 | \hat{X}_1) + I(R_2, M_2; \hat{X}_1 | \hat{X}_2) \\ &\quad = \quad I(\hat{X}_2; Y | \hat{X}_1) + I(\hat{X}_1; Y | \hat{X}_2) \\ &\quad \leq \quad \delta. \end{aligned}$$

Hence, the protocol is δ-private. ■

An immediate corollary of the above theorem is the characterization for perfect security.

Corollary 1 *A two-variate distribution P_{X_1, X_2} can be securely sampled with perfect correctness and privacy ($\epsilon = \delta = 0$) if, and only if,*

$$W(X_1; X_2) = I(X_1; X_2).$$

The outer bound of [2] applies for secure two-party sampling with perfect security, but where the parties have access to a correlated setup. Specializing this outer bound to an independent setup corresponds to secure two-party sampling from scratch and yields the necessary condition that P_{X_1, X_2} can be securely sampled with perfect security only if the Gács-Körner common information (see [7]) between (X_1, X_2) is equal to the mutual information, which happens if, and only if, the Wyner common information is equal to the mutual information (see [9] or [10]). Thus, our corollary implies that the converse of [2] is tight for the scenario of secure two-party sampling with perfect security from scratch.

C. The n-Party Case

In the n-party case, when privacy is only required against a strict minority coalition ($c < n/2$), any distribution can be securely sampled as a consequence of the universality of secure multi-party computation with a passive coalition in the strict minority [3].

Theorem 2 *For any $c < n/2$ and $\epsilon > 0$, any n-variate distribution $P_{X_1, \ldots, X_n}$ can by securely sampled with ϵ-correctness and perfect $(0, c)$-privacy.*

Proof: For any distribution $P_{X_1, \ldots, X_n}$ and any $\epsilon > 0$, there exist random variables $(Z, \hat{X}_1, \ldots, \hat{X}_n)$ where $\hat{X}_i = h_i(Z)$ for deterministic functions h_i and Z uniform over a sufficiently large finite field, such that $d(P_{\hat{X}_1, \ldots, \hat{X}_n}, P_{X_1, \ldots, X_n}) \leq \epsilon$. Since $c < n/2$, the secure multi-party computation techniques of [3] can be used to enable the parties to securely sample $(\hat{X}_1, \ldots, \hat{X}_n)$. First, each party independently and uniformly generates Z_i and distributes shares of Z_i to all of the parties in the manner of [3]. By combining these shares, each party obtains shares $Z = Z_1 + \ldots + Z_n$, which is uniformly distributed. Then, using the secure computation techniques of [3], each party can securely compute $\hat{X}_i = h_i(Z)$ with perfect privacy. Thus, the parties will have securely sampled $P_{X_1, \ldots, X_n}$ with ϵ-correctness and perfect $(0, c)$-privacy. Note that the secure computation techniques of [3] are not applicable to the general case when privacy is required against coalitions of size $c \geq n/2$. Also, although ϵ-correctness can be obtained for any $\epsilon > 0$, perfect correctness might be impossible, since the procedure produces a joint distribution with only rational probability masses. ■

In the n-party case, with privacy required against a coalition of size $c \geq n/2$, the exact characterization of the distributions that can be securely sampled is not currently known. Necessary conditions for a distribution to be securely sampled can be derived from the fact that an n-party protocol, which is private against coalitions size $c \geq n/2$, can be transformed into a secure two-party protocol. Thus, the converse for the two-party case can be bootstrapped to produce an outer bound for the n-party case.

Theorem 3 *For $c \geq n/2$, an n-variate distribution $P_{X_1, \ldots, X_n}$ can by securely sampled with ϵ-correctness and (δ, c)-privacy only if there exists $P_{\hat{X}_1, \ldots, \hat{X}_n}$ such that $d(P_{\hat{X}_1, \ldots, \hat{X}_n}, P_{X_1, \ldots, X_n}) \leq \epsilon$ and for all $T \subset \{1, \ldots, n\}$, such that $|T| \leq c$ and $|\bar{T}| \leq c$,*

$$W(\hat{X}_T; \hat{X}_{\bar{T}}) \leq I(\hat{X}_T; \hat{X}_{\bar{T}}) + \delta,$$

where $\hat{X}_T := \{\hat{X}_i\}_{i \in T}$ and $\hat{X}_{\bar{T}} := \{\hat{X}_i\}_{i \in \bar{T}}$.

Proof: An ϵ-correct and (δ, c)-private protocol samples $P_{\hat{X}_1, \ldots, \hat{X}_n}$ such that $d(P_{\hat{X}_1, \ldots, \hat{X}_n}, P_{X_1, \ldots, X_n}) \leq \epsilon$ and

$$\sum_{T \subset \{1, \ldots, n\} : |T| \leq c} I\big(\{R_i, M_i\}_{i \in T}; \{\hat{X}_i\}_{i \notin T} | \{\hat{X}_i\}_{i \in T}\big) \leq \delta.$$

For any $T \subset \{1, \ldots, n\}$, such that $|T| \leq c$ and $|\bar{T}| \leq c$, this protocol can be converted to a secure two-party protocol for generating $P_{X_T, X_{\bar{T}}}$ with ϵ-correctness and δ-privacy. Party 1 simulates the parties in T and party 2 simulates the parties in $\bar{T}$, to produce $\hat{X}_T = \{\hat{X}_i\}_{i \in T}$ and $\hat{X}_{\bar{T}} = \{\hat{X}_i\}_{i \in \bar{T}}$. The parties generate the necessary seed randomness, $R'_1 = \{R_i\}_{i \in T}$ and $R'_2 = \{R_i\}_{i \in \bar{T}}$. The messages exchanged within each group become local computations and the messages exchanged between the groups is the actual communication between the two parties, $M_1' = \{M_1(j,i),\dots,M_t(j,i)\}_{i\in T, j\in \bar{T}}$ and $M_2' = \{M_1(j,i),\dots,M_t(j,i)\}_{i\in \bar{T}, j\in T}$. The two-party protocol is ϵ-correct since $d(P_{\hat{X}_T,\hat{X}_{\bar{T}}}, P_{X_T,X_{\bar{T}}}) = d(P_{\hat{X}_1,\dots,\hat{X}_n}, P_{X_1,\dots,X_n}) \le \epsilon$. The two-party protocol is δ-private since

$$\begin{aligned}
&I\big(R_1', M_1'; \hat{X}_{\bar{T}} | \hat{X}_T\big) + I\big(R_2', M_2'; \hat{X}_T | \hat{X}_{\bar{T}}\big) \\
&\quad \le I\big(\{R_i, M_i\}_{i\in T}; \{\hat{X}_i\}_{i\notin T} | \{\hat{X}_i\}_{i\in T}\big) \\
&\qquad + I\big(\{R_i, M_i\}_{i\in \bar{T}}; \{\hat{X}_i\}_{i\notin \bar{T}} | \{\hat{X}_i\}_{i\in \bar{T}}\big) \\
&\quad \le \sum_{S\subset\{1,\dots,n\}:|S|\le c} I\big(\{R_i, M_i\}_{i\in S}; \{\hat{X}_i\}_{i\notin S} | \{\hat{X}_i\}_{i\in S}\big) \\
&\quad \le \delta,
\end{aligned}$$

due to the privacy of the n-party protocol.

Since $P_{X_T,X_{\bar{T}}}$ can be sampled with ϵ-correctness and δ-privacy,

$$W(\hat{X}_T; \hat{X}_{\bar{T}}) \le I(\hat{X}_T; \hat{X}_{\bar{T}}) + \delta,$$

due to Theorem 1. ■

Sufficient conditions that enable a distribution to be securely sampled are given in the following theorem.

Theorem 4 *For any $c \in \{1,\dots,n-1\}$, an n-variate distribution $P_{X_1,\dots,X_n}$ can by securely sampled with ϵ-correctness and (δ,c)-privacy if there exists $P_{\hat{X}_1,\dots,\hat{X}_n,Y}$ such that $d(P_{\hat{X}_1,\dots,\hat{X}_n}, P_{X_1,\dots,X_n}) \le \epsilon$,*

$$P_{\hat{X}_1,\dots,\hat{X}_n|Y} = \prod P_{\hat{X}_i|Y},$$

and

$$\sum_{T\subset\{1,\dots,n\}:|T|\le c} I\big(Y; \{\hat{X}_i\}_{i\notin T} | \{\hat{X}_i\}_{i\in T}\big) \le \delta.$$

Furthermore, if the above conditions are satisfied then the protocol need only use one round of unidirectional communication.

Proof: If the conditions of the theorem are satisfied then we can construct the following protocol using only one round of unidirectional communication:

1) Party 1 generates $R_1 = (\hat{X}_1, Y) \sim P_{\hat{X}_1,Y}$.
2) In the single round of communication, party 1 sends message $M_1(1,i) = Y$ to every other party $i \in \{2,\dots,n\}$. All other messages are null.
3) Party 1 outputs $\hat{X}_1$.
4) For $i \in \{2,\dots,n\}$, party i independently generates $R_i = \{\hat{X}_i(y)\}_{y\in\mathcal{Y}}$, where $\hat{X}_i(y) \sim P_{\hat{X}_2|Y}(\cdot|y)$.
5) Party i outputs $\hat{X}_i = g_i(R_i, Y) = \hat{X}_i(Y)$.

This protocol produces samples such that $(\hat{X}_i, Y) \sim P_{\hat{X}_i,Y}$ for $i \in \{1,\dots,n\}$. Along with the given condition that $\hat{X}_i$ are independent given Y, this implies that $(\hat{X}_1,\dots,\hat{X}_n) \sim P_{\hat{X}_1,\dots,\hat{X}_n}$. Thus, by the given condition that $d(P_{\hat{X}_1,\dots,\hat{X}_n}, P_{X_1,\dots,X_n}) \le \epsilon$, the protocol is ϵ-correct.

Now we show the protocol is also (δ,c)-private, that is,

$$\sum_{T\subset\{1,\dots,n\}:|T|\le c} I\big(\{R_i, M_i\}_{i\in T}; \{\hat{X}_i\}_{i\notin T} | \{\hat{X}_i\}_{i\in T}\big) \le \delta.$$

The terms in the summation satisfy

$$\begin{aligned}
&I\big(\{R_i, M_i\}_{i\in T}; \{\hat{X}_i\}_{i\notin T} | \{\hat{X}_i\}_{i\in T}\big) \\
&\quad \overset{(a)}{=} I(R_T, Y; \hat{X}_{\bar{T}} | \hat{X}_T) \\
&\quad = H(\hat{X}_{\bar{T}} | \hat{X}_T) - H(\hat{X}_{\bar{T}} | \hat{X}_T, R_T, Y) \\
&\quad \overset{(b)}{=} H(\hat{X}_{\bar{T}} | \hat{X}_T) - H(\hat{X}_{\bar{T}} | R_T, Y) \\
&\quad \overset{(c)}{=} H(\hat{X}_{\bar{T}} | \hat{X}_T) - H(\hat{X}_{\bar{T}} | Y) \\
&\quad \overset{(d)}{=} H(\hat{X}_{\bar{T}} | \hat{X}_T) - H(\hat{X}_{\bar{T}} | Y, \hat{X}_T) \\
&\quad = I(Y; \hat{X}_{\bar{T}} | \hat{X}_T),
\end{aligned}$$

where (a) follows since Y is a function of R_1 and $M_i = Y$ for $i \neq 1$, (b) since $\hat{X}_T$ is a function of R_T, Y, (c) since R_T is independent of $\hat{X}_{\bar{T}}$ given Y, and (d) since $\hat{X}_i$ are independent given Y. Thus, the privacy condition summation is equal to the given condition

$$\sum_{T\subset\{1,\dots,n\}:|T|\le c} I\big(Y; \{\hat{X}_i\}_{i\notin T} | \{\hat{X}_i\}_{i\in T}\big) \le \delta,$$

and hence the protocol is (δ,c)-private. ■

Theorems 3 and 4 have given both necessary conditions and sufficient conditions for distributions that can be securely sampled for coalitions of size $c \ge n/2$. However, the exact region has not been characterized as these conditions are not tight. Perhaps, an appropriately defined set of conditions based on an n-variate notion of Wyner information, such as that proposed by [11], can fully characterize the n-party region, and possibly exhibit a zero-one law (analogous to [1]), if the region for $c = \lceil n/2 \rceil$ matches that for all $c \ge n/2$. However, this remains an open problem and is part of our ongoing work.

References

[1] B. Chor and E. Kushilevitz, "A zero-one law for boolean privacy," *SIAM Journal on Discrete Mathematics*, vol. 4, no. 1, pp. 36–47, February 1991.

[2] S. Wolf and J. Wullschleger, "New monotones and lower bounds in unconditional two-party computation," *IEEE Trans. Info. Theory*, vol. 54, no. 6, pp. 2792–2797, June 2008.

[3] M. Ben-Or, S. Goldwasser, A. Wigderson, "Completeness theorems for Non-Cryptographic Fault-Tolerate Distributed Computation," in *Proc. ACM STOC*, 1988, pp. 1–10.

[4] D. Chaum, C. Crépeau, I. Damgård, "Multi-Party Unconditionally Secure Protocols," in *Proc. ACM STOC*, 1988, pp. 11–19.

[5] A. Shamir, R. Rivest, and L. Adleman, "Mental poker," Massachusetts Institute of Technology, Technical Report LCS/TR-125, April 1979.

[6] V. Prabhakaran and M. Prabhakaran, "Assisted common information with applications to secure two-party computation," *http://arxiv.org/abs/1002.1916*, February 2010.

[7] P. Gács and J. Körner, "Common information is far less than mutual information," *Problems of Control and Information Theory*, vol. 2(2), pp. 149–162, 1973.

[8] R. Ahlswede and J. Körner, "On common information and related characteristics of correlated information sources," in *Proc. of the 7th Prague Conference on Information Theory*, 1974.

[9] A. D. Wyner, "The common information of two dependent random variables," *IEEE Trans. Info. Theory*, vol. IT-21, no. 2, pp. 163–179, March 1975.

[10] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, 1981, pp. 402–405.

[11] W. Liu, G. Xu, and B. Chen, "The common information of n dependent random variables," in *Proc. Allerton Conference on Communications, Control, and Computing*, September 2010.