Understanding Information Disclosure from Secure Computation Output: A Comprehensive Study of Average Salary Computation

ALESSANDRO BACCARINI, University at Buffalo, Buffalo, United States
MARINA BLANTON, University at Buffalo, Buffalo, United States
SHAOFENG ZOU, Arizona State University, Tempe, United States

Secure multi-party computation has seen substantial performance improvements in recent years and is being increasingly used in commercial products. While a significant amount of work was dedicated to improving its efficiency under standard security models, the threat models do not account for information leakage from the output of secure function evaluation. Quantifying information disclosure about private inputs from observing the function outcome is the subject of this work. Motivated by the City of Boston gender pay gap studies, in this work, we focus on the computation of the average of salaries and quantify information disclosure about private inputs of one or more participants (the target) to an adversary via information-theoretic techniques. We study a number of distributions including log-normal, which is typically used for modeling salaries. We consequently evaluate information disclosure after repeated evaluation of the average function on overlapping inputs, as was done in the Boston gender pay study that ran multiple times, and provide recommendations for using the sum and average functions in secure computation applications. Our goal is to develop mechanisms that lower information disclosure about participants' inputs to a desired level and provide guidelines for setting up real-world secure evaluation of this function.

CCS Concepts: • **Security and privacy → Information-theoretic techniques**; *Information flow control*; • **Mathematics of computing → Information theory**;

Additional Key Words and Phrases: Secure function evaluation, information disclosure, entropy, average salary computation

1 Introduction

Secure multi-party computation and other forms of computing on cryptographically protected data (such as homomorphic encryption) open up possibilities for great utilization and analysis of private data distributed across different domains, which otherwise might not be feasible due to the

sensitive nature of the data. For example, analysis of health-related records and medical images distributed across different medical facilities and extracting cues from them lead to medical advances without the need to see the records themselves. Today, data analysis practices by researchers are hindered by laws regulating access to health data in different countries, and analyzing medical data at scale presents challenges. Similarly, analyzing sensitive data such as salaries to understand disparities by gender, race, or other types of marginalization can supply decision-makers with important information and empower them to address the disparities. This was the case with the Boston area gender pay gap surveys [13, 14, 42–44] that were initiated in 2015 and ran through 2017 with more participants and data analysis by additional categories including race. More broadly, wider adoption of privacy-preserving technologies, and secure computation in particular, can lead to higher security standards and practices for a broad range of different aspects of our society.

Secure computation techniques have seen significant advances in recent decades in terms of their speed, as well as the availability of implementations and tools to facilitate their use for a variety of applications. Tech giants such as Google and Apple started using secure computation techniques in their products [12, 37, 41, 63] and the number of start-up companies offering related products is growing (see, e.g., References [36, 45, 53, 54]). However, a number of fundamental questions still need to be addressed by the research community to make secure computing practices commonplace.

One of the fundamental questions is how much information about a participant's private input(s) might be available as a result of evaluating a desired function on private inputs. Standard security definitions adopted in the cryptographic community require that no information about private inputs is disclosed during function evaluation. That is, given a function f that we evaluate on private inputs $x_1, x_2, \ldots$ coming from different sources, security is achieved if a participant does not learn more information than the function output and any information that can be deduced from the output and its private input. However, there are no constraints on types of functions that can be evaluated in this framework, and thus the information a participant can deduce from the output and its private input about another participant's private input is potentially large. This problem is typically handled by assuming that the function being evaluated is agreed upon by and acceptable to the data owners as not to reveal too much information about private inputs. However, our ability to evaluate functions in this aspect and determine what functions might be acceptable is currently limited. This question is the subject of this work.

Intuitively, what we want is to guarantee that the function being evaluated on private data is non-invertible, i.e., observing the output does not reveal its private input. Cryptography uses the notion of one-way functions—and assumes this property for hash functions—to model non-invertibility. However, what is needed in this case is to ensure that the possible space for a target private input is still large after the adversary observes the result of function evaluation. This notion of non-invertibility was first used in the context of secure multi-party computation in solutions for business applications such as supply chain management and component procurement [24–26] and was formulated as the inability to narrow down the (private) input of another party to a single value or a small set of possible values. Consequently, a series of publications by Ah-Fat and Huth [1–4] put forward formal definitions that use entropy to measure the amount of uncertainty about one or more participants' private inputs after using them in secure multi-party computation. The definitions are framed from (a) an attacker's perspective who aims to maximize information disclosure of a target's private input and (b) from a target's perspective who determines the maximum information disclosure about their inputs when deciding whether to contribute their inputs to secure evaluation of a particular function. The above formulations are general and applicable to any function, while application-specific formulations of what constitutes sufficient input protection and function non-invertibility also emerged. One example is building machine learning

models resilient to membership inference attacks [58, 61] that guarantee that it is infeasible to determine whether someone's data was used for training the model.

Our contributions. In this work, we use the entropy-based definitions from Reference [1] as our starting point and analyze a specific function of significant practical relevance. In particular, we focus on the case of average salary computation as used in the Boston gender pay gap study [43]. When the total number of inputs is known (which is typically the case), the average computation is equivalent to computing the sum. We intuitively understand that the larger the number of inputs used in the computation of the average is, the better protection each individual contributing its input obtains. In the extreme case of two participants,[1] no protection can be achieved. This was understood by the designers of the Boston gender pay gap study who recommended running the computation with at least five contributors [44]. However, the information disclosure was not quantified, which we remedy in this work.

We start by analyzing the function itself and formally show that the amount of information an attacker learns is independent of their own inputs. This is consistent with our intuition that, given a sum, one can always remove their contribution to the sum and analyze the resulting value. Thus, the protection depends on the number of spectators, i.e., input parties distinct from those controlled by the adversary and the party or parties being targeted.

We analyze the target's input entropy remaining after participating in the computation (and consequently the entropy loss as a result of participation) for a number of discrete and continuous distributions including uniform, Poisson, normal (Gaussian), and log-normal. Log-normal is typically used for modeling salary data but is the least trivial to analyze. An unexpected finding of our analysis is that for a given distribution, the absolute entropy loss is normally independent of the distribution parameters, and the absolute entropy loss remains very close for different distributions as we vary the number of participants/spectators. Quantifying the information loss allows us to devise a mechanism to lower information disclosure to any desired level (e.g., 1% of original entropy, 0.05 bit of entropy).

We extend our analysis of information loss to the case when the computation is run more than once (as was the case for the Boston gender pay gap study) and examine the case with two evaluations. This corresponds to (i) the target participating in two computations with the same input where the set of participants differs between the executions and (ii) the target participating in one computation, where the other is run without the target's input. We observe that information loss increases as a result of multiple computations, regardless of whether the target participates once or twice. Furthermore, the protection is maximized when one-half of the original contributors are replaced, i.e., 50% of the initial participants remain and the other 50% are replaced with new participants. Our multi-execution analysis is based on the normal distribution, but we expect the outcome to be similar for other distributions as well.

We proceed with generalizing our analysis to three and more, M, executions. An interesting finding is that the best configuration that minimizes information loss is determined by pairwise overlaps of participants between the executions and not by other parameters and sizes. This allows us to determine the optimal setup for a single and repeated execution of the average function.

We empirically validate our findings throughout this work and provide recommendations for securely evaluating the average function in real-world applications. In particular, in all of our experiments, the cost of participating in the average computation, i.e., the difference in the entropy before and after the computation, is a fraction of a bit (for both Shannon entropy used with discrete

---

[1]We use the term "participants" to denote parties contributing inputs to the computation. The computation itself can be performed by a different set of parties, but our result is independent of the mechanism used to realize secure function evaluation.

distributions and differential entropy used with continuous distributions). This translates to a small relative entropy loss in practice. When modeling salary data using the log-normal distribution with the parameters specifically chosen for salaries [17], the entropy loss is below 5% with at least 5 non-adversarial participants or spectators, and achieving 1% entropy loss requires 24 spectators. These numbers are also surprisingly similar across different distributions. Furthermore, when the computation is repeated (we use a normal distribution to adequately approximate the log-normal setup), engaging in the computation the second time with an overlapping set of 50% participants whose inputs do not change results in only 30% entropy loss of the first participation. These and other findings lead to a number of recommendations for running this computation in practice, which we provide at the end of this work.

This work is the full version of our conference publication [8].

On the choice of metric. Our analysis uses Shannon entropy. One might argue that this is not the best metric because it does not distinguish between, e.g., leaking the least significant vs. most significant bit of one's salary, while learning the latter is much more valuable to an adversary than learning the former. However, as we show throughout this work, information leakage for this application is always small, regardless of the setup. In particular, the most favorable for the adversary setup across all distributions discloses only about 0.7 bit of entropy, i.e., the adversary cannot learn even a single bit of the target's salary. Furthermore, we derive effective mechanisms for reducing information loss to a controlled low level such that the worst-case scenario will not realize. We conduct similar analysis using min-entropy (see Section 4.4) and show that Shannon entropy trends are consistent with those for min-entropy. A primary advantage of using Shannon entropy is that we are able to go much further in our analysis and ultimately derive close-form expressions, which cannot be accomplished for other metrics.

2 Related Work

In what follows, we review prior literature on information disclosure from function output in the context of computing on private data and related techniques that limit information disclosure.

2.1 Quantitative Information Flow

The field of *quantitative information flow* is closely related to our work. Denning [23] is credited as the first to quantify information flow as a measure of the interference between variables at two stages during a program's execution (typically denoted by "high-" and "low-security" variables, which equates to the target's inputs and output in our setting, respectively). Smith [60] formally established the foundations of quantifying the information leakage under the threat model that an attacker can recover a secret in one attempt (denoted by the notion of *vulnerability*). It has been shown by Massey [48] that the Shannon entropy cannot capture this information under the guessing assumption, and Smith recommends min-entropy in its place. Alvim et al. [6] generalized the min-entropy into the g-leakage to incorporate gain functions to model the *benefit* an adversary gains from making guesses about the secret. Subsequent works encompassed variations on the g-leakage [5]. Other works in differential privacy feature derivations of leakage bounds [19], leakage analysis in the case of an adaptive adversary [39], and knowledge-based approaches for measuring risk [47, 55].

The fundamental advantage of our Shannon-based approach is the ability to derive closed-form expressions for the information leakage of the average salary computation, while other metrics do not share this characteristic. For example, the chain rule of entropy (a simple, yet critical component of our analysis) is not satisfied if min-entropy is used [38, 59] in place of Shannon entropy. Our reductions would no longer hold, and we would be forced to resort to complete enumeration

or approximation methods to compute the entropy. However, in Section 4.4, we provide supplementary analysis that demonstrates similarities between Shannon entropy and min-entropy-based analyses. We also remain open to evaluating other metrics in the future.

An additional distinction between our work and existing literature on (quantitative) information flow is that we do not consider possible leakage from intermediate aspects of a computation's execution. Whereas other works may examine a program's loops [47], side-channel vectors [39], or inter-dependent structures [7], we strictly investigate the relationship between the output and target's input, since function itself is assumed to be evaluated using secure multi-party protocols.

2.2 Function Information Disclosure

Existing literature on information leakage from the output of a secure function evaluation is limited, relative to the rest of the field of secure computation. Secure multi-party protocols are designed to guarantee no information is disclosed throughout a computation but do not ensure input protection after the output is revealed. The work of Deshpande et al. [24–26] was pioneering in that respect and designed secure multi-party protocols for business applications that ensured that the function being evaluated is *non-invertible*, i.e., no participant can infer other participants' inputs from the output. A trivially invertible example is the average salary calculation between two individuals, since either party can recover the other's input exactly. Deshpande et al. [25, 26] first addressed non-invertibility in the context of secure supply chain processes. The proposed protocols offered protection from inference of future inputs to a repeated calculation after a result is disclosed. A later work by Deshpande et al. [24] achieved non-invertibility for a framework designed for secure price masking for outsourcing manufacturing. The authors argued information leakage was minimal by analyzing mutual information between correlated normal random variables but did not consider other distributions or entropy metrics.

Ah-Fat and Huth [1] provided the first in-depth analysis of information leakage from the outputs of secure multi-party computations. The authors formalized two metrics to measure expected information flow from the attacker's and target's perspectives, namely, the ***attacker's weighted average entropy*** (awae) and ***target's weighted average entropy*** (twae), respectively. Participants' inputs are modeled using probability distributions and were specified to be uniform, but this constraint can be relaxed. The inherent difficulty of this entropy-based approach is the requirement to enumerate every possible input combinations from all parties, which scales poorly as the input space and number of participants grow. We utilize their definitions for our analysis and demonstrate their utility to computation designers to determine potential disclosure about participants' inputs

This model was expanded in Reference [2] to encompass the Rényi, min-, and *g*-entropy. The extension is presented in combination with a technique for distorting secure computation outputs to limit information disclosure from the output and achieve balance between accuracy and privacy. This was further developed in Reference [3] with a fuzzing method based on randomized approximations. A closed-form expression for the min-entropy of a two- and three-party auction was derived in Reference [4] alongside a conjecture for the case with an arbitrary number of parties.

Conceptually, the notion of *output privacy* is related to our work. The terminology was introduced in the field of data mining [15, 40, 50, 51, 64] with the goal of designing techniques to protect inputs from inference attacks on the output model. Information about the inputs that can be obtained from the output includes, but is not limited to, properties that can be uniquely attributed to a small number of input participants. Conventional approaches for minimizing disclosure involve applying transformations on the result via monotonic functions [15] or even proactive learning [64]. These techniques have little to no impact on the result of the computation. This direction differs from our work, since the type of disclosure they aim to rectify is not quantified.

There is also literature that uses specific formulations to demonstrate that computation does not disclose sensitive information about participants. This includes resilience to *membership inference attacks* [58, 61] in the context of machine learning training and *differential privacy* [27, 29] for statistical databases. In particular, differential privacy ensures the output of a query is negligibly dependent on a single individual's record in the database, and resilience to membership inference attacks prevents one from determining whether a specific individual's data were used for model training. These concepts have no direct relationship to our work, aside from designing mechanisms for lower information disclosure as a result of computation on private data. In this work, we do so by varying the number of participants in the computation, while other methods augment the function to produce, e.g., a differentially private output. We discuss differential privacy in the context of the average salary in more detail in Section 4.5.

3   Preliminaries

Following Reference [1]'s notation, let P denote the set of all participants in a computation with $|P| = m$. All participants P are partitioned into three groups: parties controlled by an attacker $A \subset P$, a group of parties being targeted $T \subseteq P \setminus A$, and the remaining participants called spectators $S = P \setminus (A \cup T)$. Let the random variable X_{P_i} correspond to the input of a single participant P_i and x_{P_i} denotes a value that X_{P_i} takes. In addition, the notation $\vec{X}_P = (X_{P_1}, \ldots, X_{P_m})$ denotes a multidimensional random variable and $\vec{x}_P$ is a vector of the individual values of the same size. We also let $X_P = \sum_i X_{P_i}$ define a new random variable representing the sum of the participants' random variables. The same notation applies to the sets A, T, and S. Our present analysis is based upon the assumption that all participants' inputs are independent and identically distributed, which we consequently relax.

For discrete distributions, we use Shannon entropy $H(X)$ to measure the information of a discrete random variable X with mass function $\Pr(X = x)$ defined over domain D_X. Specifically,

$$H(X) = - \sum_{x \in D_X} \Pr(X = x) \cdot \log \Pr(X = x),$$

where all logarithms are to the base 2. If we are dealing with continuous distributions, then we shift to the differential entropy $h(X)$ with density function $f(x)$ over the support set $\mathcal{X}$, defined as

$$h(X) = - \int_{\mathcal{X}} f(x) \log f(x) dx.$$

We study information leakage of the computation of the average:

$$o = f(\vec{x}_A, \vec{x}_T, \vec{x}_S) = \frac{1}{m} \left(\sum_i x_{T_i} + \sum_j x_{A_j} + \sum_k x_{S_k} \right),$$

where o denotes the output of the function. We model the output o by the random variable O defined over the domain D_O, namely,

$$O = \frac{1}{m} \left(\sum_i X_{T_i} + \sum_j X_{A_j} + \sum_k X_{S_k} \right).$$

The $1/m$ factor can be ignored in the final expression, since the number of participants is typically known by all parties and can trivially be removed from the output. We omit it throughout the remainder of the article.

In this work, we consider distributions where the sum of independent individual random variables is well studied and their mass or density functions have closed-forms expressions or can be reasonably approximated. This includes the following distributions:

— *Discrete uniform* $\mathcal{U}(a, b)$, where a and b are integers corresponding to the minimum and maximum of the range of the support set $\{a, a+1, \ldots, b-1, b\}$.

— *Poisson* $\text{Pois}(\lambda)$, where $\lambda \in \mathbb{R}_{>0}$ is the shape parameter that indicates the expected (average) rate of an event occurring over a given interval.

— *Normal (Gaussian)* $\mathcal{N}(\mu, \sigma^2)$, where $\mu \in \mathbb{R}$ and $\sigma^2 \in \mathbb{R}_{>0}$ correspond to the mean and squared standard deviation, respectively.

— *Log-normal* $\log \mathcal{N}(\mu, \sigma^2)$ with parameters $\mu \in \mathbb{R}$ and $\sigma^2 \in \mathbb{R}_{>0}$, which correspond to the mean and squared standard deviation of the random variable's natural logarithm.

$X \sim \text{Dist}$ indicates that random variable X has distribution Dist.

As stated earlier, Ah-Fat and Huth [1] provided multiple information-theoretic measures to quantify information disclosure after a function evaluation, which we use here:

Definition 1 ([1]). The **joint weighted average entropy (jwae)** of a target T attacked by parties A is defined over all $\vec{x}_A \in D_A$ and $\vec{x}_T \in D_T$ as

$$\text{jwae}(\vec{x}_A, \vec{x}_T) = \sum_{o \in D_O} \Pr(O = o \mid \vec{X}_A = \vec{x}_A, \vec{X}_T = \vec{x}_T) \cdot H(\vec{X}_T \mid \vec{X}_A = \vec{x}_A, O = o).$$

This metric measures the information an attacker would learn (on average) about the target when the input vectors are $\vec{x}_A$ and $\vec{x}_T$. One can subsequently define the average of the jwae over all possible $\vec{x}_T$ or $\vec{x}_A$ vectors weighted by their respective prior probabilities.

Definition 2 ([1]). The **target's weighted average entropy (twae)** of a target T attacked by parties A is defined for all $\vec{x}_T \in D_T$ as

$$\text{twae}(\vec{x}_T) = \sum_{\vec{x}_A \in D_A} \Pr(\vec{X}_A = \vec{x}_A) \cdot \text{jwae}(\vec{x}_A, \vec{x}_T).$$

The twae informs a target how much information an attacker can learn about its input when the input is $\vec{x}_T$.

Definition 3 ([1]). The **attacker's weighted average entropy (awae)** of a target T attacked by parties A is defined for all $\vec{x}_A \in D_A$ as

$$\text{awae}(\vec{x}_A) = \sum_{\vec{x}_T \in D_T} \Pr(\vec{X}_T = \vec{x}_T) \cdot \text{jwae}(\vec{x}_A, \vec{x}_T).$$

The awae informs an attacker about how much information it can learn about the target's input when the attacker's input vector is $\vec{x}_A$. The attacker can consequently compute the awae on all values in D_A to determine which input maximizes the information learned about the target's input (and thus what should be entered into the computation). Using the definition of jwae, it follows that:

$$\text{awae}(\vec{x}_A) = \sum_{\vec{x}_T \in D_T} \Pr(\vec{X}_T = \vec{x}_T) \sum_{o \in D_O} \Pr(O = o | \vec{X}_A = \vec{x}_A, \vec{X}_T = \vec{x}_T) \cdot H(\vec{X}_T \mid \vec{X}_A = \vec{x}_A, O = o)$$

$$= \sum_{\vec{x}_T \in D_T} \sum_{o \in D_O} \Pr(O = o, \vec{X}_T = \vec{x}_T \mid \vec{X}_A = \vec{x}_A) \cdot H(\vec{X}_T \mid \vec{X}_A = \vec{x}_A, O = o).$$

Since $\vec{X}_T$ is independent of $\vec{X}_A$, we derive that awae equals to conditional entropy:

$$\text{awae}(\vec{x}_A) = \sum_{o \in D_O} \Pr(O = o \mid \vec{X}_A = \vec{x}_A) \cdot H(\vec{X}_T \mid \vec{X}_A = \vec{x}_A, O = o) = H(\vec{X}_T \mid \vec{X}_A = \vec{x}_A, O),$$

where the last equality is due to the definition of conditional entropy.

4   Single Execution

In this section, we analyze a single execution of the average function on private inputs and the associated information disclosure of the target's inputs. Recall that the computation is modeled by

$$O = f(\vec{X}_A, \vec{X}_T, \vec{X}_S) = \sum_i X_{T_i} + \sum_j X_{A_j} + \sum_k X_{S_k}, \tag{1}$$

and we let $n = |S|$ denote the number of spectators.

As a first step, we plot the values of awae and twae for our function of interest. Figure 1 illustrates these values with a single adversarial participant, a single target, and a varying number of spectators (1–3). All inputs follow the uniform distribution $\mathcal{U}(0, 15)$. Calculating the twae and awae values using Definitions 2 and 3 requires enumerating all input and output combinations. This quickly becomes computationally inefficient as the input space grows.

Each participant, acting as a target, can utilize the twae prior to the computation to determine how much information an attacker can learn (on average) from the output for a specific input that the participant enters into the computation. As the figure illustrates, the target's remaining average entropy is maximized when the input is in the middle of the range, indicating that those values have better protection than inputs near the extrema. As the number of spectators increases, the curves shift upwards, i.e., the uncertainty about the target's input increases and the gap in the uncertainty between different input values reduces.

The awae, however, gives an adversary the ability to determine which input to enter into the computation that leads to maximum information disclosure about a target's input (without knowing what input the target used). As displayed in the figure, the adversarial knowledge does not change by varying its inputs into the computation. This is consistent with our intuition that, given the output, the adversary can remove their contribution to the computation and possess information about the sum of the inputs of the remaining parties. We formalize this as the following result:

CLAIM 1.   awae($\vec{x}_A$) *is independent of attacker's input vector* $\vec{x}_A$.

PROOF.   According to the chain rule of entropy, which states that $H(X, Y) = H(X \mid Y) + H(Y)$ [22, Ch. 2.5], we have that:

$$H(\vec{X}_T \mid \vec{X}_A = \vec{x}_A, O) = H(\vec{X}_T, O \mid \vec{X}_A = \vec{x}_A) - H(O \mid \vec{X}_A = \vec{x}_A)$$

$$= H(\vec{X}_T \mid \vec{X}_A = \vec{x}_A) + H(O \mid \vec{X}_T, \vec{X}_A = \vec{x}_A) - H(O \mid \vec{X}_A = \vec{x}_A)$$

$$= H(\vec{X}_T) + H\left(\sum_i X_{S_i}\right) - H\left(\sum_i X_{T_i} + \sum_j X_{S_j}\right), \tag{2}$$

which is independent of $\vec{x}_A$.   □

Using our notation from Section 3, the above expression for awae($\vec{x}_A$) simplifies to

$$H(\vec{X}_T) + H(X_S) - H(X_T + X_S) = H(\vec{X}_T \mid X_T + X_S). \tag{3}$$

The next step is to determine which measure (awae or twae) we should use in our analysis of the average salary computation. Ah-Fat and Huth [1] argued that the awae served as a more precise metric for measuring information leakage of a secure function evaluation than twae for their choice of function and used awae in their subsequent work [2]. Our perspective also aligns with that conclusion. In particular, while the twae informs the target of the amount of information leakage for the input they possess, the target may not be technically savvy enough to be able to apply the metric and make an informed decision regarding computation participation (plus, the choice to participate or not participate can leak information about their input). Perhaps more importantly, a
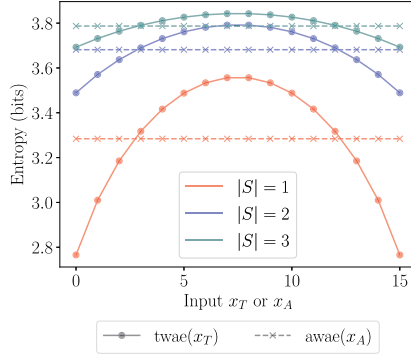
Fig. 1. The twae($\vec{x}_T$) and awae($\vec{x}_A$) using inputs over $\mathcal{U}(0, 15)$ with a different number of spectators $|S|$.

function needs to be analyzed by the computation designers in advance and without access to the inputs of future computation participants to determine a safe setup for the participants. Thus, the available mechanism for this purpose is the attacker's perspective or awae, and we focus on this metric in the rest of this work.

Based on the above, in what follows, we use $H(\vec{X}_T \mid X_T + X_S)$ to measure the leakage, and the simplified function is

$$f(\vec{X}_T, \vec{X}_S) = \sum_i X_{T_i} + \sum_j X_{S_j} = X_T + X_S.$$

This refines the parameters we can vary in our analysis to (1) the number of participants in the target and spectators groups and (2) the types of distributions and statistical parameters of the inputs. Furthermore, the computational difficulty associated with directly computing the awae is absent when using $H(\vec{X}_T \mid X_T + X_S)$. Instead, the computation simplifies to calculating the entropy of sums of random variables. We examine the behavior of the conditional entropy for several characteristic probability distributions next.

4.1 Discrete Distributions

We start with discrete input modeled using the uniform and Poisson distributions. The sum of n identical independent Poisson random variables $X_i \sim \text{Pois}(\lambda)$ is equivalent to a single Poisson random variables $X = \sum_i X_i \sim \text{Pois}(n\lambda)$ with the mass function $\Pr(X = x) = (n\lambda)^x \, e^{-n\lambda}/(x!)$. Note that the Poisson distribution is defined over all non-negative integers, hence the distribution has infinite support. We choose to halt the calculation of $H(X)$ when $\Pr(X = x) < 10^{-7}$, as the contribution of events beyond this point to the entropy is infinitesimal.

Conversely, the sum of n identical independent uniform random variables $X_i \sim \mathcal{U}(0, N-1)$ is not immediately obvious. Caiado and Rathie [16] derived several equivalent expressions for the mass function of the sum of n uniform random variables, one of which we use in our analysis and is defined as:

$$\Pr(X = x) = \frac{n}{N^n} \left(\sum_{p=0}^{\lfloor x/N \rfloor} \frac{\Gamma(n + x + pN)(-1)^p}{\Gamma(p+1)\Gamma(n-p+1)\Gamma(x - pN + 1)} \right),$$

where $\Gamma(n) = (n-1)!$ is the Gamma function. The domain of X is $\{0, \ldots, n(N-1)\}$.

Our analysis of awae for these two distribution is given in Figures 2 and 3, respectively. We compute and display
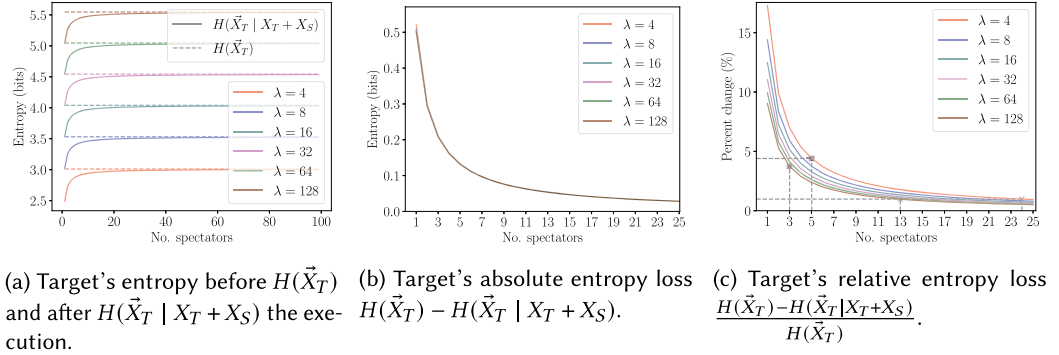
(a) Target's entropy before $H(\vec{X}_T)$ and after $H(\vec{X}_T \mid X_T + X_S)$ the execution.

(b) Target's absolute entropy loss $H(\vec{X}_T) - H(\vec{X}_T \mid X_T + X_S)$.

(c) Target's relative entropy loss $\frac{H(\vec{X}_T) - H(\vec{X}_T \mid X_T + X_S)}{H(\vec{X}_T)}$.

Fig. 2. Analysis of target's entropy loss using the Poisson distribution with $\text{Pois}(\lambda)$, varying λ with $|T| = 1$.

(a) Target's entropy before $H(\vec{X}_T)$ and after $H(\vec{X}_T \mid X_T + X_S)$ the execution.

(b) Target's absolute entropy loss $H(\vec{X}_T) - H(\vec{X}_T \mid X_T + X_S)$.

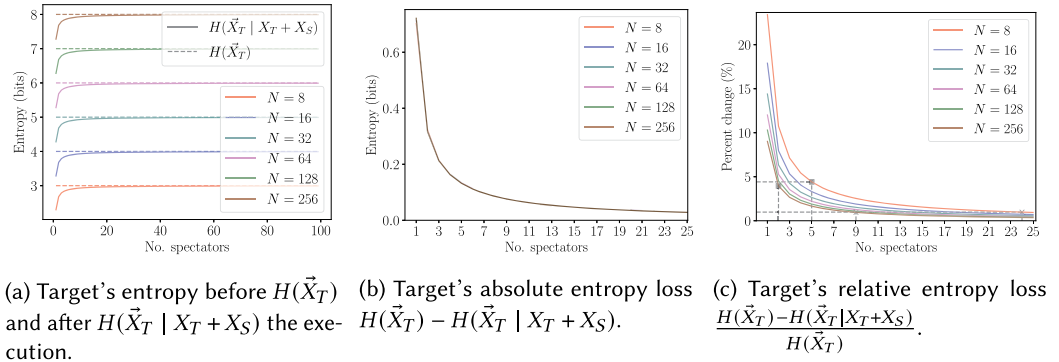(c) Target's relative entropy loss $\frac{H(\vec{X}_T) - H(\vec{X}_T \mid X_T + X_S)}{H(\vec{X}_T)}$.

Fig. 3. Analysis of target's entropy loss using the uniform distribution with $\mathcal{U}(0, N-1)$, varying N with $|T|=1$.

— the original entropy of target's inputs prior to the computation $H(\vec{X}_T)$ (subfigure a),
— the awae or target's remaining entropy after the computation $H(\vec{X}_T \mid X_T + X_S)$ (subfigure a),
— their difference of the two that represents the absolute entropy loss $H(\vec{X}_T) - H(\vec{X}_T \mid X_T + X_S)$ (subfigure b), and
— the entropy loss relative to the original entropy prior to the execution $(H(\vec{X}_T) - H(\vec{X}_T \mid X_T + X_S))/H(\vec{X}_T)$ (subfigure c)

with a single target ($|T| = 1$), a varying number of spectators, and varying distribution parameters. Relative entropy loss is included to demonstrate to potential input contributors, who are likely non-experts, that information disclosure is small. That is, disclosure of, e.g., 5% of input's information is easier to explain to non-experts than 0.1 bit of entropy. The absolute loss is equivalent to the mutual information between the target input and the output:

$$I(\vec{X}_T; O) = H(\vec{X}_T) - H(\vec{X}_T \mid X_T + X_S).$$

Figure 2 presents this information for the Poisson distribution with $\lambda \in \{4, 8, \ldots, 128\}$. In Figure 2(a), entropy after the execution converges toward the corresponding entropy prior to the execution for all values of λ as the number of spectators increases. Increasing λ by a factor of two repeatedly yields an upward shift of these two curves by a constant amount while preserving their respective shapes. The increase is expected as a result of the inputs having more entropy as λ increases, but the shape of the remaining entropy is notable, as λ does not appear to impact the

entropy loss. This is further confirmed when displaying the absolute entropy loss in Figure 2(b): The resultant curves overlap each other, regardless of λ.

The relative entropy loss in Figure 2(c), calculated as a percentage of the target's initial entropy, demonstrates how many spectators the computation needs to include to lower the entropy loss to the desired level. The larger the original entropy is (larger λ), the fewer spectators will be needed to stay within the desired percentage. For example, 5 spectators are needed with $\lambda = 4$ to limit relative loss to 5% (marked by ■), and 24 spectators are needed to cap the loss at 1% (marked by ×). When $\lambda = 128$, the number of spectators reduces to 3 and 13 to maintain loss tolerances of 5% and 1%, respectively.

The same trends hold for the uniform distribution in Figure 3, where we use $N \in \{8, 16, \ldots, 256\}$, but the values themselves slightly differ. For example, the absolute entropy loss in Figure 3(b) is slightly larger than the loss in Figure 2(b) when the number of spectators is small. When $N = 8$ with 3 bits of original entropy, 5 and 24 spectators are needed to achieve at most 5% and 1% relative loss, respectively. This is the same as what was observed for Poisson distribution with 3-bit inputs ($\lambda = 4$).

4.2 Continuous Distributions

For continuous distributions, we shift to differential entropy and analyze normal and log-normal distributions, the latter of which is typically used to model salaries. While there is no direct relationship between differential and Shannon entropy (see Reference [22, Chapter 8.3]), we demonstrate that they exhibit very similar behavior for the average computation.

The differential entropy of a normal random variable $X_i \sim \mathcal{N}(\mu, \sigma^2)$ is $h(X_i) = \frac{1}{2}\log\left(2\pi e\sigma^2\right)$ [22, Ch. 8.1]. The sum of n identical normal random variables is also normal, namely, $X \sim \mathcal{N}(n\mu, n\sigma^2)$. This enables us to directly apply the differential entropy definition to the sum.

The log-normal distribution is a well-established means of modeling salary data for 99% of the population [20], with the top 1% modeled by the Pareto distribution [62]. The differential entropy of a log-normal random variable $X_i \sim \log\mathcal{N}(\mu, \sigma^2)$ is $h(X_i) = \log(e^{\mu+\frac{1}{2}}\sqrt{2\pi\sigma^2})$. However, the sum of n log-normal random variables has no closed form and is an active area of research [9–11, 21, 31, 56, 57, 65]. We adopt the **Fenton-Wilkinson (FW)** approximation[2] [21, 31] that specifies a sum of n identical independent log-normal random variables $X_i \sim \log\mathcal{N}(\mu, \sigma^2)$ as another log-normal random variable $X \sim \log\mathcal{N}(\hat{\mu}, \hat{\sigma}^2)$ with parameters

$$\hat{\sigma}^2 = \ln\left(\frac{\exp(\sigma^2) - 1}{n} + 1\right), \ \hat{\mu} = \ln(n \cdot \exp(\mu)) + \frac{1}{2}\left(\sigma^2 - \hat{\sigma}^2\right).$$

This enables us to compute differential entropy using a closed-form expression. Unlike prior distributions, we use a single set of μ and σ^2 parameters calculated from real salary data in Reference [17]; namely, $\mu = 1.6702$ and $\sigma^2 = 0.145542$.

Figures 4 and 5 present experimental evaluation of entropy loss with a single target and a varying number of spectators for normal and log-normal distributions, respectively. As before, we report target's entropy before and after the execution, the difference of the two as the absolute entropy loss, and the entropy loss relative to the entropy before the execution.

In Figure 4 (normal), we set the mean $\mu = 0$ for all experiments (since differential entropy does not depend on μ) and vary σ^2 from 4 to 128. The results are consistent with the discrete

---

[2]Other approximations for the sum of log-normal random variables are difficult to translate into an expression for the differential entropy, and hence we choose the FW approximation. Its disadvantage is that the FW approximation deteriorates for $\sigma^2 > 4$ and small values of x in the density function [10, 65]. Fortunately, our σ^2 is sufficiently small, allowing us to use the FW approximation free of consequence.
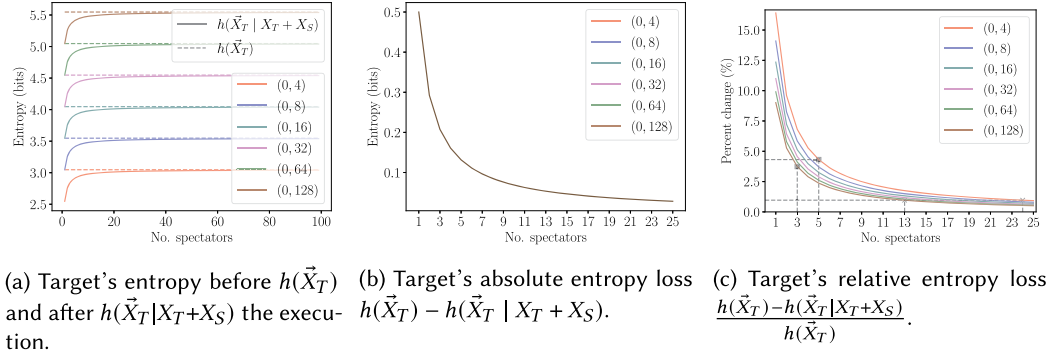
(a) Target's entropy before $h(\vec{X}_T)$ and after $h(\vec{X}_T|X_T+X_S)$ the execution.

(b) Target's absolute entropy loss $h(\vec{X}_T) - h(\vec{X}_T \mid X_T + X_S)$.

(c) Target's relative entropy loss $\frac{h(\vec{X}_T) - h(\vec{X}_T|X_T+X_S)}{h(\vec{X}_T)}$.

Fig. 4. Analysis of target's entropy loss using the normal distribution with $\mathcal{N}(0, \sigma^2)$, varying σ^2 with $|T| = 1$.

(a) Target's entropy before $h(\vec{X}_T)$ and after $h(\vec{X}_T \mid X_T + X_S)$ the execution.

(b) Target's absolute entropy loss $h(\vec{X}_T) - h(\vec{X}_T \mid X_T + X_S)$.

(c) Target's relative entropy loss $\frac{h(\vec{X}_T) - h(\vec{X}_T|X_T+X_S)}{h(\vec{X}_T)}$.
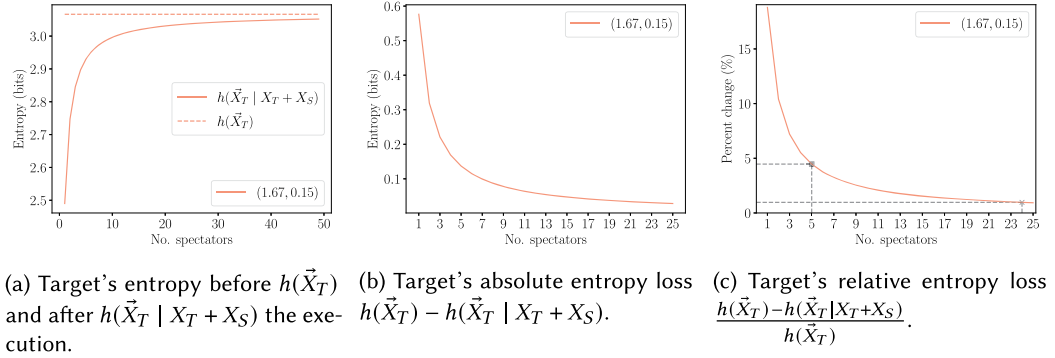
Fig. 5. Analysis of target's entropy loss using the log-normal distribution with $\log \mathcal{N}(1.6702, 0.145542)$ and $|T|=1$.

counterparts in terms of the trends, curve shapes, and specific values. The absolute loss in Figure 4(b) is once again constant for any σ^2, and the relative loss is dictated by the amount of input's entropy in Figure 4(c). When $\sigma^2 = 4$ and inputs have 3 bits of entropy, the number of spectators required to maintain at most 5% and 1% entropy loss (5 and 24 spectators, respectively) is the same as for Poisson and uniform distributions with 3-bit inputs ($\lambda = 4$ and $N = 8$, respectively). With 5.5-bit inputs ($\sigma^2 = 128$), 3 and 13 spectators are needed to achieve at most 5% and 1% loss, respectively, which is the same as for Poisson distribution with 5.5-bit inputs ($\lambda = 128$).

The results in Figure 5 (log-normal with real salary parameters) are consistent with both the discrete and continuous distributions. Surprisingly, we observe the same 5 and 24 spectators achieve at most 5% and 1% relative loss, as observed with all other distributions (with input original entropy being slightly over 3 bits).

Before concluding our discussion of continuous distributions, we are able to show one more result. We experimentally demonstrated that the amount of absolute entropy loss is parameter-independent for several distributions, but we can formally prove this for normally distributed inputs:

CLAIM 2. *If the inputs are modeled by independent identically distributed normal random variables, then the absolute entropy loss $h(\vec{X}_T) - h(\vec{X}_T \mid X_T + X_S)$ depends only on the number of target $|T| = t$ and spectator $|S| = n$ inputs and is $\frac{1}{2} \log \left(\frac{t}{n} + 1 \right)$.*
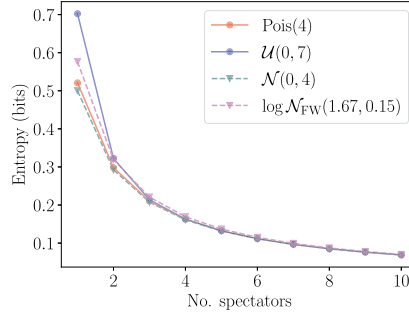
Fig. 6. Comparing target's absolute entropy loss for discrete $H(\vec{X}_T) - H(\vec{X}_T \mid X_T + X_S)$ and continuous $h(\vec{X}_T) - h(\vec{X}_T \mid X_T + X_S)$ distributions.

PROOF. Let $|T| = t$ and $|S| = n$, such that $X_T \sim \mathcal{N}(0, t\sigma^2)$ and $X_S \sim \mathcal{N}(0, n\sigma^2)$. The absolute entropy loss is therefore

$$h(\vec{X}_T) - h(\vec{X}_T \mid X_T + X_S) = h(\vec{X}_T) - (h(\vec{X}_T) + h(X_S) - h(X_T + X_S)) = h(X_T + X_S) - h(X_S)$$

$$= \frac{1}{2} \log 2\pi e(t+n)\sigma^2 - \frac{1}{2} \log 2\pi e n\sigma^2 = \frac{1}{2} \log \left(\frac{t}{n} + 1 \right) = \Theta \left(\log \left(\frac{t}{n} + 1 \right) \right),$$

which depends only on n and t. □

4.3 Discrete vs. Continuous Distributions

We next compare the information loss across all four (discrete and continuous) distributions. We choose parameters to maintain the initial entropy of an input, $H(X_i)$ or $h(X_i)$, to be approximately 3 bits, as to reasonably correspond to the log-normal distribution. This leads to Pois(4), $\mathcal{U}(0, 7)$, and $\mathcal{N}(0, 4)$. We plot this information for a single target and a varying number of spectators in Figure 6.

In the figure, all distributions converge with ≥ 4 spectators and are very close even with 3 spectators. This convergence on large values is expected as a consequence of the central limit theorem. From the four distributions, the closest are the Poisson results with $\lambda = 4$ (discrete) and the normal distribution $\mathcal{N}(0, 4)$ (continuous). Unlike normal, log-normal, and the single-variate uniform distributions, an exact expression of the entropy of a Poisson distribution has not been derived. Instead, when computing the necessary values in Section 4.1, we directly applied the definition of Shannon entropy. To draw a parallel between discrete and continuous distributions, and specifically show a similarity between Poisson and normal distributions, we turn to an approximation of Poisson distribution's entropy computation.

It was conjectured that for sufficiently large λ (e.g., $\lambda > 10$), the Poisson distribution's Shannon entropy can be approximated by $H(X_i) = \frac{1}{2} \log(2\pi e \lambda)$, which resembles $h(X_i) = \frac{1}{2} \log(2\pi e \sigma^2)$ used for normal distributions. Evans and Boersma [30] proposed a tighter bound (further formalized by Cheraghchi in Reference [18]), to be

$$H(X_i) = \frac{1}{2} \log(2\pi e \lambda) - \frac{1}{12\lambda} - \frac{1}{24\lambda^2} - \frac{19}{360\lambda^3} + O(\lambda^4)$$

and remains close to that of normal distribution with $\sigma^2 = \lambda$.

One implication of this result for us is that Claim 2, which we demonstrated for normal distributions, would apply to the approximation of Poisson distributions as well. As a result, we obtain independence of the (absolute) entropy loss of distribution parameters for both discrete and continuous distributions and almost identical behavior across the distributions as a function of the number of spectators.

4.4 Min-entropy Analysis

We treat min-entropy as an alternative to Shannon entropy, which was studied in the context of information flow by Smith [60]. While we are unable to go as far in our analysis as in the case of (Shannon) entropy, we certainly observe similar trends. We begin by defining the concept of *vulnerability*:

Definition 4 (Vulnerability, [60]). Given a discrete random variable X with support $\mathcal{X}$, the *vulnerability* of X, denoted by $V_\infty(X)$, over the unit interval $[0, 1]$ is given by

$$V_\infty(X) = \max_{x \in \mathcal{X}} \Pr(X = x).$$

The vulnerability $V_\infty(X)$ is interpreted as the worst-case probability that an adversary could guess the value of X in one attempt. If m guesses are allowed, then the adversary's success probability is at most $mV_\infty(X)$. The implication is that if the vulnerability with a practical number of m guesses is significant, then $V_\infty(X)$ must also be significant. Since the vulnerability is a probability, we can convert it to an entropy measure (in bits) by taking the logarithm of $V_\infty(X)$. Conveniently, this is exactly the definition of *min-entropy* $H_\infty(X) = \log \frac{1}{V_\infty(X)}$. Smith's [60] motivation for departing from Shannon entropy stems from its ineffectiveness of properly assessing the threat the output Y has on its input X.

Since our analysis studies the relationship between input and output random variables (i.e., X_T and O), a necessary extension is the *conditional vulnerability*, which specifies the expected probability of guessing X in one try, given that Y is observed:

Definition 5. Given two random variables X and Y with supports $\mathcal{X}$ and $\mathcal{Y}$, respectively, the *conditional vulnerability* $V_\infty(X \mid Y)$ is

$$V_\infty(X \mid Y) = \sum_{y \in \mathcal{Y}} \Pr(Y = y) \cdot V_\infty(X \mid Y = y), \text{ where } V_\infty(X \mid Y = y) = \max_{x \in \mathcal{X}} (\Pr(X = x \mid Y = y)).$$

Having established the necessary foundations of min-entropy, we are equipped to extend our single-execution analysis of Section 4 from the perspective of min-entropy:

Definition 6. The attacker's weighted average min-entropy (awae$_\infty$) of a target $\vec{X}_T$ attacked by parties A is defined for all $\vec{x}_A \in D_A$ as

$$\text{awae}_\infty(\vec{x}_A) = H_\infty(X_T \mid O, \vec{X}_A = \vec{x}_A) = -\log \sum_{o \in D_O} \Pr(O = o \mid \vec{X}_A = \vec{x}_A) \cdot V_\infty(\vec{X}_T \mid \vec{X}_A = \vec{x}_A, O = o),$$

where $V_\infty(\vec{X}_T \mid \vec{X}_A = \vec{x}_A, O = o)$ is the conditional vulnerability defined above.

The above definition is a concrete min-entropy specification of Ah-Fat and Huth's [2] generalized awae, which is parameterized by α and a gain function g. We can manipulate Definition 6 into terms consistent with Section 4 by plugging in the expression for conditional vulnerability:

$$
\begin{aligned}
\text{awae}_\infty(\vec{x}_A) &= -\log \sum_{o \in D_O} \Pr(O = o \mid \vec{X}_A = \vec{x}_A) \cdot \left(\max_{\vec{x}_T \in D_T} \Pr(\vec{X}_T = \vec{x}_T \mid \vec{X}_A = \vec{x}_A, O = o) \right) \\
&= -\log \sum_{o \in D_O} \Pr(O = o \mid \vec{X}_A = \vec{x}_A) \cdot \left(\max_{\vec{x}_T \in D_T} \frac{\Pr(O = o \mid \vec{X}_T = \vec{x}_T, \vec{X}_A = \vec{x}_A) \cdot \Pr(\vec{X}_T = \vec{x}_T)}{\Pr(O = o \mid \vec{X}_A = \vec{x}_A)} \right) \\
&= -\log \sum_{o \in D_O} \left(\max_{\vec{x}_T \in D_T} \Pr(O = o \mid \vec{X}_T = \vec{x}_T, \vec{X}_A = \vec{x}_A) \cdot \Pr(\vec{X}_T = \vec{x}_T) \right).
\end{aligned}
$$
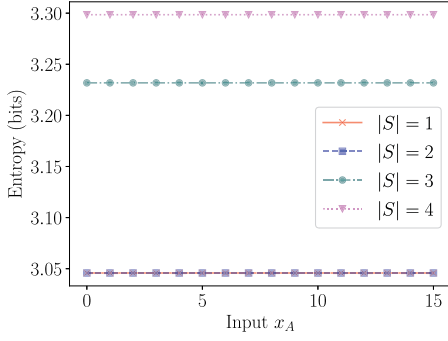
(a) The awae$_\infty(\vec{x_A})$ using uniformly distributed inputs over $\mathcal{U}\,(0, 15)$ with a different number of spectators $|S|$.
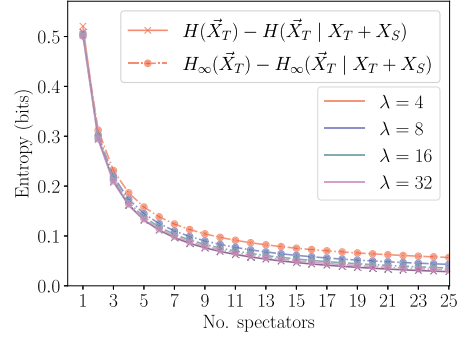
(b) Comparing the Shannon and min-entropy absolute losses using Poisson distribution with Pois(λ), varying λ with $|T| = 1$.

Fig. 7. Min-entropy analysis.

In the second line, we invoked Bayes' theorem, and in the third line, we observed that the denominator is a constant factor in the max expression and could thus be factored out and subsequently canceled with the leading $\Pr(O = o \mid \vec{X}_A = \vec{x}_A)$.

In Claim 1, we proved awae($\vec{x}_A$) was independent of the attacker's input $\vec{x}_A$. Conversely, awae$_\infty(\vec{x}_A)$ cannot be simplified further to prove the claim holds. Hence, we conjecture the following:

CONJECTURE 1. awae$_\infty(\vec{x_A})$ is independent of attacker's input vector $\vec{x_A}$.

We can, however, repeat the calculation of Figure 1 using min-entropy. In Figure 7(a), we once again observe the same behavior that the adversarial knowledge does not change by varying its inputs into the computation.[3] This suggests the conjecture holds for the average salary computation. Hence, we assume such in our subsequent analysis.

The next logical step is to examine how the effect transitioning from Shannon entropy to min-entropy has on the absolute loss. We compute and display both absolute losses in Figure 7(b), where participants' inputs are modeled by the Poisson distribution (as in Section 4.1). As observed previously in Figure 2(b), the Shannon absolute loss curves all overlap each other. Interestingly, we observe the min-entropy absolute loss curves converge towards their Shannon counterparts as λ grows. This suggests that for a sufficiently large statistical parameter, the choice of metric used to represent information disclosure is less impactful.

4.5 Comparison to Differential Privacy

The purpose of this work is to measure information disclosure from function output, which is the first necessary step to determine whether the function is suitable for evaluation on private data. Once it is determined that it is not, the second question to answer is how the function or the setup is to be modified to reduce information disclosure to a controlled sufficiently small level. This can be achieved by different means, e.g., by enrolling more participants as suggested in this work or by modifying the function to be evaluated (e.g., by injecting noise in the output). In the context of the average computation we consider in this work, the information disclosure can be reduced to any desired level by controlling the number of participants. This eliminates the

---

[3]Interestingly, awae$_\infty(\vec{x_A})$ is the same for $|S| = 1$ and $|S| = 2$ and the curves overlap on the plot.

need for function modification. Nevertheless, we provide a detailed analysis of the suitability of **differential privacy (DP)** for this application. In short, when the number of participants is large, information disclosure from the output is very low and DP is not needed, but when the number of participants is small, applying a DP mechanism results in distortion that impairs utility.

We start with a brief overview of the fundamentals of DP. The core tenet of DP involves restricting the information learned about a single individual within a sensitive dataset; this is normally accomplished by introducing noise to the result. Formally, DP is defined as follows:

Definition 7 ((ε, δ)-differential privacy, [28]). Let $\mathcal{M}$ be a mechanism that takes an input database D of size n supported by $\mathcal{D}$ and produces a randomized output in the set S. Then, $\mathcal{M}$ is considered to be (ε, δ)-*differentially private* for $\varepsilon, \delta \geq 0$ if for all *adjacent* databases D and D' (differ by a single entry) and all sets $S \subseteq \text{Range}(\mathcal{M})$ if $\Pr[\mathcal{M}(D) \in S] \leq \exp(\varepsilon) \cdot \Pr[\mathcal{M}(D') \in S] + \delta$, where $\text{Range}(\mathcal{M})$ is the set of all possible outputs of the mechanism $\mathcal{M}$.

The above definition encapsulates both "pure" DP ($\delta = 0$) and "approximate" DP, which allows an additive privacy loss of $\delta > 0$, where δ is negligible in the size of the database. We formulate our arguments under the stronger assumption of pure DP.

Though many differentially private mechanisms exist (see, e.g., References [46, 49]), we restrict our view to the *Laplace mechanism* [28] due to its well-established nature and reliable performance for low-dimensional queries [52]. We construct the one-dimensional Laplace mechanism, since the output of the average function is a single value. We require the notion of *function sensitivity*, i.e., the maximum difference in the output of f when applied to two adjacent datasets:

Definition 8 ([27]). Let $f : \mathcal{D}^n \to \mathbb{R}$. The *sensitivity* of f is $\Delta f = \max_{D, D'} |f(D) - f(D')|$, where D and D' are adjacent.

The sensitivity dictates the scale parameter of a Laplace random variable, which we can now formally define:

Definition 9 (Laplace mechanism, [27]). Let $f : \mathcal{D}^n \to \mathbb{R}$. Given $\varepsilon \geq 0$ and sensitivity Δf, the *Laplace mechanism* is defined as $\mathcal{M}(D) = f(D) + L$, where $L \sim \text{Lap}(\Delta f / \varepsilon)$ is a Laplace random variable.

We compute how much noise is required to maintain differential privacy for a given ε. The absolute error introduced by a mechanism $\mathcal{M}$ (or, analogously, the "utility" of the mechanism in the literature) is measured via a *loss function* ℓ. The function $\ell(f(D), \mathcal{M}(D))$ refers to the loss of an individual user when the output of the function is $f(D)$ and the mechanism's (perturbed) output is $\mathcal{M}(D)$. The choice of ℓ is arbitrary, but most DP literature [34] uses the mean error $\ell(f(D), \mathcal{M}(D)) = |f(D) - \mathcal{M}(D)|$, and thus we use it for our analysis. Given a statistical significance α, it can be shown (via application of Chernoff bounds) that the error introduced by the Laplace mechanism is bounded by

$$\ell(f(D), \mathcal{M}(D)) = |f(D) - \mathcal{M}(D)| \leq (\Delta f / \varepsilon) \ln(1/\alpha),$$

with probability $1 - \alpha$ (i.e., the confidence level). We interpret this bound as the "worst-case" upper bound on the distance between the true and the perturbed outputs. To transform this bound into a measurable quantity, we compare it against the expected value of the function with no distortion applied as a percentage, i.e., the relative error.

If we model inputs by the uniform distribution $X_i \sim \mathcal{U}(0, N - 1)$, then we can derive an exact value for the sensitivity of our f. For the average with uniformly distributed inputs, the function sensitivity is $\Delta f = N - 1$, since the output of the function o can differ by at most $N - 1$ if one individual does not participate. Given the expected value for the sum of n uniform random variables

of $\mathbb{E}[\sum_i X_i] = n\frac{N-1}{2}$, the maximum possible error ∇ introduced by the mechanism is

$$\nabla = \frac{(\Delta f/\varepsilon)\ln(1/\alpha)}{n(N-1)/2} = \frac{2\ln(1/\alpha)}{\varepsilon n}.$$

This quantity is monotonically decreasing in n such that, as n increases, the bound on ∇ shrinks to zero. This echoes the sentiment in DP literature that databases are typically assumed to be large.

Consider the example of 6 participants (5 spectators, 1 target) with $X_i \sim \mathcal{U}(0,7)$, $\varepsilon = 1$, and a standard [35, 52] 95% confidence level ($\alpha = 0.05$). From our analysis in Section 4.1, this ensures the relative entropy loss for the target is at most 5%. However, the relative error introduced by the Laplace mechanism is at most $\nabla = 99.8\%$, implying that the output of the computation can vary so drastically as to render the output unusable. Placing this in the context of the Boston gender pay gap study, where the goal was to determine the difference between male and female average salaries, the error is too large for reliable decision-making. Imposing a stricter bound on the relative entropy loss of 1% by increasing the number of participants to 25 reduces the upper bound on the error to $\nabla = 24.0\%$. We find disclosure of 1% of the input's entropy (e.g., about 0.03 bit for this application) to be acceptable, at which point there is no longer a need to use DP, and we can output precise results.

To summarize, increasing the number of participants is a natural mechanism for lowering information disclosure for the average computation. DP is of limited utility in this context, as applying it to the setting where the number of participants cannot be increased results in utility loss.

4.6   Mixed Distribution Parameters

Up to this point, we have assumed that all participants' inputs are sampled from identically distributed random variables. However, we can relax this assumption and investigate if/how the information disclosure changes if parties' inputs are non-identically distributed. For example, employee salaries may differ slightly from company to company while still following the same distribution. We can model this by adjusting the statistical parameters of individual participants.

We begin by formalizing the notion of participant "groups." Define $\mathbb{G}$ as a finite set of statistical distributions, of which participants' inputs can be sourced from. For example, if we have two groups B and C of normally distributed inputs parameterized by $\mathcal{N}(0, \sigma_B^2)$ and $\mathcal{N}(0, \sigma_C^2)$ (where $\sigma_B^2 \neq \sigma_C^2$), respectively, then $\mathbb{G} = \{B, C\}$. This formulation poses two interesting directions for introducing participant group identities, i.e., correspondence of a participant to one of the distribution groups, into our analysis:

— **Group identities of individual participants are known.** The first setting we consider is that the identity, i.e., the group, of each individual participant is known. In practice, this is realized by multiple entities with inputs modeled by different statistical distributions contribute to a computation, where the number of inputs submitted by each is publicly available.

— **Group identities of individual participants are unknown.** Conversely, we have the scenario where we have knowledge of the possible distribution groups participants can belong to with anticipated likelihoods, but the group identity of an individual party is not known. This is objectively more general than the first category but requires knowledge in the form of the probabilities of an arbitrary participant belonging to each group.

It is therefore of interest to revisit our prior conclusions under the known *and* unknown group identity generalizations (denoted by Cases 1 and 2, respectively), since both formulations bear operational significance.

Entropy loss as a result of computation participation. The first conclusion we revisit is Claim 1, since it is integral to our analysis as a whole. The claim states that the information

disclosure from the average function output is independent of the attackers' inputs. Based on this result, our subsequent analysis enabled us to derive expressions for awae.

In the current generalized setting, Claim 1 remains true for both Cases 1 and 2, since the derivation in the proof of Claim 1 itself remains unchanged. However, we must adjust Equation (3) such that participant group membership is captured by our entropy measure. We recall our definitions of entropy remaining after participation and the absolute entropy loss:

$$H(\vec{X}_T \mid X_T + X_S) = H(\vec{X}_T) + H(X_S) - H(X_T + X_S) \tag{4}$$

$$H(\vec{X}_T) - H(\vec{X}_T \mid X_T + X_S) = H(X_T + X_S) - H(X_S). \tag{5}$$

Accounting for group identities, we introduce the *participant identity random variable* ID_{P_i} supported by $\mathbb{G}$. This corresponds to the group identity of participant P_i, and we denote $\mathsf{id}_{P_i} \in \mathbb{G}$ as the value ID_{P_i} takes. We similarly denote $\vec{\mathsf{ID}}_P = (\mathsf{ID}_{P_1}, \ldots, \mathsf{ID}_{P_m})$ as a multidimensional random variable, with $\vec{\mathsf{id}}_P$ as the vector of individual values of the same size.

At this point, our analysis splits into two directions based on the knowledge of individual group identities.

Case 1: If participant group identities are available (i.e., $\vec{\mathsf{ID}}_T = \vec{\mathsf{id}}_T$ and $\vec{\mathsf{ID}}_S = \vec{\mathsf{id}}_S$), then Equation (4) becomes:

$$H(\vec{X}_T \mid X_T + X_S, \vec{\mathsf{ID}}_T = \vec{\mathsf{id}}_T, \vec{\mathsf{ID}}_S = \vec{\mathsf{id}}_S) = H(\vec{X}_T \mid \vec{\mathsf{ID}}_T = \vec{\mathsf{id}}_T) + H(X_S \mid \vec{\mathsf{ID}}_S = \vec{\mathsf{id}}_S)$$
$$- H(X_T + X_S \mid \vec{\mathsf{ID}}_T = \vec{\mathsf{id}}_T, \vec{\mathsf{ID}}_S = \vec{\mathsf{id}}_S). \tag{6}$$

Since we have exact knowledge of each participant's identity, we can explicitly partition the input random variables accordingly. Therefore, all the above quantities are computable with minimal deviation from our original analysis. For instance, if we recall our earlier example with two participant groups B and C, then the term $H(X_S \mid \vec{\mathsf{ID}}_S = \vec{\mathsf{id}}_S)$ can be computed as:

$$H(X_S \mid \vec{\mathsf{ID}}_S = \vec{\mathsf{id}}_S) = H\left(\sum_{j \in S} X_{S_j} \mid \vec{\mathsf{ID}}_S = \vec{\mathsf{id}}_S\right) = H\left(\left(\sum_{j \in S:\, \mathsf{id}_{S_j} = \mathsf{B}} X_{S_j} + \sum_{j \in S:\, \mathsf{id}_{S_j} = \mathsf{C}} X_{S_j}\right) \mid \vec{\mathsf{ID}}_S = \vec{\mathsf{id}}_S\right).$$

The absolute entropy loss directly follows from Equation (6) and is computed as:

$$H(\vec{X}_T \mid \vec{\mathsf{ID}}_T = \vec{\mathsf{id}}_T) - H(\vec{X}_T \mid X_T + X_S, \vec{\mathsf{ID}}_T = \vec{\mathsf{id}}_T, \vec{\mathsf{ID}}_S = \vec{\mathsf{id}}_S)$$
$$= H(X_T + X_S \mid \vec{\mathsf{ID}}_T = \vec{\mathsf{id}}_T, \vec{\mathsf{ID}}_S = \vec{\mathsf{id}}_S) - H(X_S \mid \vec{\mathsf{ID}}_S = \vec{\mathsf{id}}_S). \tag{7}$$

Under this generalization, the group to which a target belongs impacts how much information is disclosed from the computation, i.e., the disclosure can fall within a range based on the values $\vec{\mathsf{id}}_T$ can take. We determine the worst-case information disclosure by iterating over all possible target identities and taking the maximum:

$$\max_{\vec{\mathsf{id}}_T} \left(H(X_T + X_S \mid \vec{\mathsf{ID}}_T = \vec{\mathsf{id}}_T, \vec{\mathsf{ID}}_S = \vec{\mathsf{id}}_S) - H(X_S \mid \vec{\mathsf{ID}}_S = \vec{\mathsf{id}}_S) \right).$$

We can further refine our earlier notation established in Section 3 to encompass participant group identities (applicable to both targets and spectators). Let $P_{\mathsf{G}} \subset P$ be the set of participants belonging to group $\mathsf{G} \in \mathbb{G}$. The sum of random variables modeling participant inputs is given as

$$X_S = \sum_{\mathsf{G} \in \mathbb{G}} \sum_{i \in P_{\mathsf{G}}} X_{P_i} = \sum_{\mathsf{G} \in \mathbb{G}} X_{P_{\mathsf{G}}},$$

where $X_{P_{\mathsf{G}}} = \sum_{i \in P_{\mathsf{G}}} X_{P_i}$.

Case 2: When the group identities of individual inputs are not known and only the probability of belonging to a given group is given, the procedure for evaluating the information disclosure changes. The probability mass and density functions, respectively, for the participant inputs random variables are now:

$$\Pr(\vec{X}_P = \vec{x}_P) = \sum_{\vec{\mathrm{id}}_P} \Pr(\vec{\mathrm{ID}}_P = \vec{\mathrm{id}}_P)\Pr(\vec{X}_P = \vec{x}_P \mid \vec{\mathrm{ID}}_P = \vec{\mathrm{id}}_P)$$

$$f(\vec{x}_P) = \sum_{\vec{\mathrm{id}}_P} \Pr(\vec{\mathrm{ID}}_P = \vec{\mathrm{id}}_P)f(\vec{x}_P \mid \vec{\mathrm{ID}}_P = \vec{\mathrm{id}}_P).$$

For a participant set P, there are $|\mathbb{G}|^{|P|}$ possible identity configurations, such that the number of terms in the summation is exponential in the number of participants and/or size of the group identity set. The Shannon and differential entropies are now computed as:

$$H(\vec{X}_P) = - \sum_{\vec{x}_P \in D_{X_P}} \left(\sum_{\vec{\mathrm{id}}_P} \Pr(\vec{\mathrm{ID}}_P = \vec{\mathrm{id}}_P)\Pr(\vec{X}_P = \vec{x}_P | \vec{\mathrm{ID}}_P = \vec{\mathrm{id}}_P) \right)$$

$$\cdot \log \left(\sum_{\vec{\mathrm{id}}_P} \Pr(\vec{\mathrm{ID}}_P = \vec{\mathrm{id}}_P)\Pr(\vec{X}_P = \vec{x}_P | \vec{\mathrm{ID}}_P = \vec{\mathrm{id}}_P) \right), \tag{8}$$

$$h(\vec{X}_P) = - \int_{X_P} \left(\sum_{\vec{\mathrm{id}}_P} \Pr(\vec{\mathrm{ID}}_P = \vec{\mathrm{id}}_P)f(\vec{x}_P \mid \vec{\mathrm{ID}}_P = \vec{\mathrm{id}}_P) \right)$$

$$\cdot \log \left(\sum_{\vec{\mathrm{id}}_P} \Pr(\vec{\mathrm{ID}}_P = \vec{\mathrm{id}}_P)f(\vec{x}_P \mid \vec{\mathrm{ID}}_P = \vec{\mathrm{id}}_P) \right) d\vec{x}_P. \tag{9}$$

The fundamental difference between the entropy calculation under this generalization and the analysis conducted in Sections 4.1 and 4.2 is that *the entropy of these random variables is no longer exactly modeled by the input distribution itself* (e.g., Poisson, uniform, Gaussian, log-normal). Furthermore, for continuous input distributions, our previous approach of leveraging closed-form expressions is no longer applicable when group identities are unknown —the information disclosure must be computed numerically, rather than exactly.

Parameter independence of the absolute loss for normally distributed inputs. The next conclusion we revisit is Claim 2, which previously stated that for normally distributed inputs, the absolute entropy loss depends only on the number of targets and spectators present in the computation. This conclusion changes in the generalized setting when the participants' inputs are no longer identically distributed, as we demonstrate below.

Case 1: Let σ_{G}^2 be the standard deviation of the participants inputs that belong to group G. Using the definitions from Section 4.2 for the entropy sums of identically distributed normal random variables, the entropy of $X_{P_{\mathrm{G}}}$ is $h(X_{P_{\mathrm{G}}}) = \frac{1}{2} \log \left(2\pi e \sigma_{\mathrm{G}}^2 |P_{\mathrm{G}}|\right)$. We similarly derive the following expressions needed to compute the absolute entropy loss (given in Equation (7)):

$$h(X_T + X_S \mid \vec{\mathrm{ID}}_T = \vec{\mathrm{id}}_T, \vec{\mathrm{ID}}_S = \vec{\mathrm{id}}_S) = \frac{1}{2} \log 2\pi e \left(\sum_{G \in \mathbb{G}} \left(\sigma_{\mathrm{G}}^2 \cdot |T_G| + \sigma_{\mathrm{G}}^2 \cdot |S_G| \right) \right)$$

$$h(X_S \mid \vec{\mathrm{ID}}_S = \vec{\mathrm{id}}_S) = \frac{1}{2} \log 2\pi e \left(\sum_{G \in \mathbb{G}} \left(\sigma_{\mathrm{G}}^2 \cdot |S_G| \right) \right).$$

(a) Absolute entropy loss $h(\vec{X}_T) - h(\vec{X}_T | X_T + X_S)$.

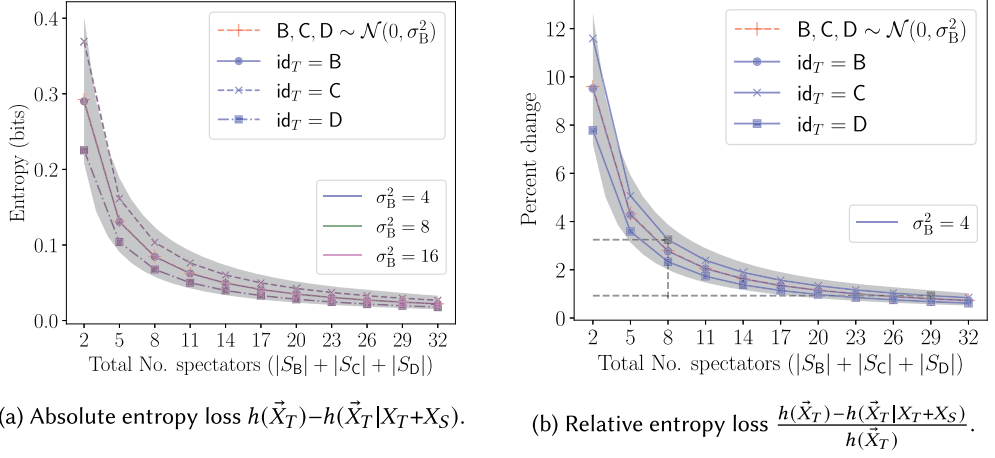(b) Relative entropy loss $\frac{h(\vec{X}_T) - h(\vec{X}_T | X_T + X_S)}{h(\vec{X}_T)}$.

Fig. 8. Mixed distribution analysis under Case 1. The red dashed curves correspond to our baseline where all groups are identically distributed (B, C, D $\sim \mathcal{N}(0, \sigma_B^2)$), while the remaining curves indicate the target belonging to distinct groups distributed by B $\sim \mathcal{N}(0, \sigma_B^2)$, C $\sim \mathcal{N}(0, 1.1^2\sigma_B^2)$, and D $\sim \mathcal{N}(0, 0.9^2\sigma_B^2)$. The shaded regions illustrate the full space for the absolute entropy loss, generated from every possible spectator and group configurations.

Plugging in these equations to our expression for the absolute entropy loss and simplifying yields:

$$h(X_T + X_S \mid \vec{\mathsf{ID}}_T = \vec{\mathsf{id}}_T, \vec{\mathsf{ID}}_S = \vec{\mathsf{id}}_S) - h(X_S \mid \vec{\mathsf{ID}}_S = \vec{\mathsf{id}}_S) = \frac{1}{2}\log\left(\frac{\sum_{G\in\mathbb{G}}\left(\sigma_G^2 \cdot |T_G|\right)}{\sum_{G\in\mathbb{G}}\left(\sigma_G^2 \cdot |S_G|\right)} + 1\right).$$

Unlike the analysis in the proof of Claim 2, the standard deviations do not cancel. However, we can reformulate our interpretation of the sums of standard deviations when accounting for group identities. Let us define $\sigma_B^2 \in \mathbb{R}_{>0}$ as the "base standard deviation" for all input random variables. Then, for all G $\in \mathbb{G}$, there exists some $\delta_G > 0$ such that $\sigma_G^2 = \delta_G \cdot \sigma_B^2$. Substituting into the above expression yields:

$$h(X_T + X_S \mid \vec{\mathsf{ID}}_T = \vec{\mathsf{id}}_T, \vec{\mathsf{ID}}_S = \vec{\mathsf{id}}_S) - h(X_S \mid \vec{\mathsf{ID}}_S = \vec{\mathsf{id}}_S) = \frac{1}{2}\log\left(\frac{\sum_{G\in\mathbb{G}}\left(\delta_G \cdot \sigma_B^2 \cdot |T_G|\right)}{\sum_{G\in\mathbb{G}}\left(\delta_G \cdot \sigma_B^2 \cdot |S_G|\right)} + 1\right)$$

$$= \frac{1}{2}\log\left(\frac{\sum_{G\in\mathbb{G}}\left(\delta_G \cdot |T_G|\right)}{\sum_{G\in\mathbb{G}}\left(\delta_G \cdot |S_G|\right)} + 1\right).$$

The key conclusion from the above equation is that the absolute entropy loss is not directly affected by the statistical parameter σ_G^2 but rather the *relationship* each σ_G^2 has (via the scaling factor δ_{ID}) to the base standard deviation σ_B^2.

To demonstrate this phenomenon, we compute the absolute entropy loss when the target belongs to one of three possible groups differing by ±10% in their average salary. Concretely, we have B, C, or D with deviations σ_B^2, $1.1^2\sigma_B^2$, and $0.9^2\sigma_B^2$, respectively. The target is interpreted to "move" from group to group as to maintain a consistent group size, e.g., $|S_B \cup T| = |S_C| = |S_D|$ when the target is in group B. This constitutes the curves displayed in Figure 8. To further illustrate the best- and worst-case information disclosure, we compute the absolute entropy loss for all possible spectator-group configurations and compose the shaded region from the maximums and minimums. Figure 8(a) reflects our observation that, regardless of the base standard deviation σ_B^2 (4, 8, or 16), the curves fall on top of each other. Moreover, the absolute loss when $\mathsf{id}_T = $ B is equivalent
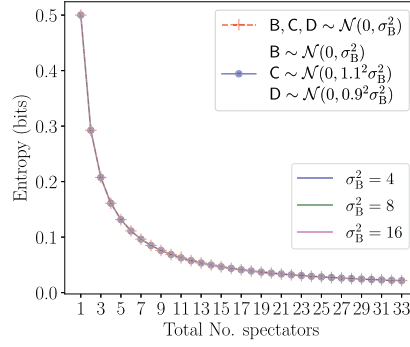
Fig. 9. Mixed distribution analysis under Case 2, where the probability of an arbitrary participant belonging to any specific group is equally likely, i.e., $\Pr(\text{ID}_P = B) = \Pr(\text{ID}_P = C) = \Pr(\text{ID}_P = D) = 1/3$.

to our original computation when all groups are identically distributed. We also reproduce our relative loss experiment using the same $\pm10\%$ salary configuration for $\sigma_B^2 = 4$ in Figure 8(b). Achieving maximum relative losses of 5% and 1% now requires at least 6 and 27 spectators, respectively.

Case 2: When group identities of individual inputs are unknown, we refer to the definition of absolute entropy loss (Equation (5)) alongside the expressions for the differential entropy we derived in Equation (9) for the required quantities and obtain:

$$
h(X_T+X_S) = -\int_{\mathcal{X}_T \cup \mathcal{X}_S} \left(\sum_{\vec{\text{id}}_T, \vec{\text{id}}_S} \Pr(\vec{\text{ID}}_T=\vec{\text{id}}_T, \vec{\text{ID}}_S=\vec{\text{id}}_S) f(x_T+x_S | \vec{\text{ID}}_T=\vec{\text{id}}_T, \vec{\text{ID}}_S=\vec{\text{id}}_S) \right)
$$

$$
\cdot \log \left(\sum_{\vec{\text{id}}_T, \vec{\text{id}}_S} \Pr(\vec{\text{ID}}_T=\vec{\text{id}}_T, \vec{\text{ID}}_S=\vec{\text{id}}_S) f(x_T+x_S | \vec{\text{ID}}_T=\vec{\text{id}}_T, \vec{\text{ID}}_S=\vec{\text{id}}_S) \right) d(x_T+x_S)
$$

$$
h(X_S) = -\int_{\mathcal{X}_S} \left(\sum_{\vec{\text{id}}_S} \Pr(\vec{\text{ID}}_S=\vec{\text{id}}_S) f(x_S | \vec{\text{ID}}_S=\vec{\text{id}}_S) \right) \cdot \log \left(\sum_{\vec{\text{id}}_S} \Pr(\vec{\text{ID}}_S=\vec{\text{id}}_S) f(x_S | \vec{\text{ID}}_S=\vec{\text{id}}_S) \right) dx_S.
$$

These values need to be computed numerically, since, as previously stated, the random variables that represent the target and spectators' inputs no longer exactly translate to the input distribution itself. Utilizing the same group configuration as specified above under Case 1 and assuming the probability for each identity is equally likely (for convenience), we compute the absolute loss in Figure 9 alongside our baseline where every participant belongs to a single group. The most interesting observation is that all the curves overlap each other, a trend originally observed in Sections 4.1 and 4.2. We note that this is likely a consequence of the experimental configuration itself (groups' salaries differ by $\pm10\%$, identity probabilities are equally likely).

5 Two Executions

A natural generalization of the results of the prior section is to consider executing the average salary computation more than once. For example, after running the Boston gender pay gap study once, the same computation was executed the following year with an extended set of participants. In this case, if the time interval between the executions is small enough such that the inputs do not change between the executions or change minimally, then one would expect that repeated participations would lead to additional information disclosure compared to a single execution. Thus, in this section, we analyze the case of two executions and demonstrate their impact on the

participants. We consider both the cases when a target contributes its input to both executions and when the target participates only in one of the executions and other takes place without the target but on related inputs. Both cases result in additional information disclosure compared to a single execution, which we quantify in this section.

We partition the set of spectators S into the following subsets:

— spectators present only in the first execution $S_1 \subset S$,
— spectators present only in the second execution $S_2 \subseteq S \setminus S_1$,
— and spectators present in both executions $S_{12} = S \setminus (S_1 \cup S_2)$.

A person participating more than once (target or spectator) enters the same input into both executions.

When the target participates in both executions, we have:

$$O_1 = \sum_i X_{T_i} + \sum_{i \in S_{12}} X_i + \sum_{i \in S_1} X_i = X_T + X_{S_{12}} + X_{S_1}$$

$$O_2 = \sum_i X_{T_i} + \sum_{i \in S_{12}} X_i + \sum_{i \in S_2} X_i = X_T + X_{S_{12}} + X_{S_2}.$$

The random variables O_1 and O_2 are *not* independent, as they both are composed of X_T and $X_{S_{12}}$. We therefore want to compute the conditional entropy (using differential entropy notation):

$$h(\vec{X}_T \mid O_1, O_2) = h(\vec{X}_T, O_1, O_2) - h(O_1, O_2). \tag{10}$$

CLAIM 3. *The above conditional entropy can be expressed as*

$$h(\vec{X}_T | O_1, O_2) = h(\vec{X}_T) + h(X_{S_{12}} + X_{S_1}, X_{S_{12}} + X_{S_2}) - h(O_1, O_2). \tag{11}$$

PROOF. Simplifying the first term of Equation (10) using the chain rule of entropy $h(X, Y) = h(X \mid Y) + h(Y)$ [22], we obtain:

$$h(\vec{X}_T, O_1, O_2) = h(\vec{X}_T, X_T + X_{S_{12}} + X_{S_1}, X_T + X_{S_{12}} + X_{S_2})$$
$$= h(\vec{X}_T) + h(X_T + X_{S_{12}} + X_{S_1} \mid \vec{X}_T) + h(X_T + X_{S_{12}} + X_{S_2} \mid X_T + X_{S_{12}} + X_{S_1}, \vec{X}_T).$$

Using the fact that all participants' inputs are independent, we have:

$$h(\vec{X}_T, O_1, O_2)$$
$$= h(\vec{X}_T) + h(X_{S_{12}} + X_{S_1}) + h(X_T + X_{S_{12}} + X_{S_2}, X_T + X_{S_{12}} + X_{S_1} \mid \vec{X}_T) - h(X_T + X_{S_{12}} + X_{S_1} \mid \vec{X}_T)$$
$$= h(\vec{X}_T) + h(X_{S_{12}} + X_{S_1}) + h(X_{S_{12}} + X_{S_2}, X_{S_{12}} + X_{S_1}) - h(X_{S_{12}} + X_{S_1})$$
$$= h(\vec{X}_T) + h(X_{S_{12}} + X_{S_1}, X_{S_{12}} + X_{S_2}).$$

The second term of Equation (10) can be rewritten as:

$$h(O_1, O_2) = h(X_T + X_{S_{12}} + X_{S_1}, X_T + X_{S_{12}} + X_{S_2})$$
$$= h(X_T + X_{S_{12}} + X_{S_1}) + h(X_T + X_{S_{12}} + X_{S_2} \mid X_T + X_{S_{12}} + X_{S_1}),$$

but cannot be simplified further. Therefore, the final expression of the conditional entropy is

$$h(\vec{X}_T | O_1, O_2) = h(\vec{X}_T) + h(X_{S_{12}} + X_{S_1}, X_{S_{12}} + X_{S_2}) - h(O_1, O_2). \qquad \square$$

In the special case when no spectators participate in both executions (i.e., $S_{12} = \emptyset$), the middle term simplifies to $h(X_{S_1}) + h(X_{S_2})$.

When the target participates only in one of the experiments, we define executions O'_1 and O'_2, which are the same as O_1 and O_2, respectively, except that the target's inputs are not included. For instance, $O'_1 = X_{S_{12}} + X_{S_1}$. The relevant entropies in that case are $h(\vec{X}_T | O'_1, O_2)$ and $h(\vec{X}_T | O_1, O'_2)$.

The above requires us to introduce the definition of joint entropy of correlated random variables. Now, the normal distribution stands out among those considered in Section 4 as a suitable candidate for our analysis. The generalized multivariate normal distribution is well-studied and has a closed-form differential entropy, which we discuss next.

5.1 Bivariate Normal Distributions

Evaluating Equation (11) requires defining the differential entropy of a multivariate normal random variable. We then derive the necessary core parameters for our distributions and use them to compute the conditional entropy.

Let $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ be a single normal random variable as defined in Section 3. We define $\vec{X} = (X_1, \ldots, X_k)^{\mathsf{T}}$ to be a general multivariate normal distribution of a k-dimensional random vector, with $\vec{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$. Here, $\boldsymbol{\mu} \in \mathbb{R}^k$ is the mean vector specified as $\boldsymbol{\mu} = \mathbb{E}[\vec{X}] = (\mathbb{E}[X_1], \mathbb{E}[X_2], \ldots, \mathbb{E}[X_k])^{\mathsf{T}} = (\mu_1, \mu_2, \ldots, \mu_k)^{\mathsf{T}}$, and $\Sigma \in \mathbb{R}^{k \times k}$ is the $k \times k$ covariance matrix with each element defined as $\Sigma_{i,j} = \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)] = \mathrm{Cov}[X_i, X_j]$. The differential entropy of the multivariate normal distribution $\vec{X}$ is given by $h(\vec{X}) = \frac{1}{2} \log((2\pi e)^k \det \Sigma)$ [22, Chapter 8.4], where $\det \Sigma$ is the determinant of the covariance matrix. The next step is to characterize our multivariate distributions and determine their covariance matrices. We also derive their mean vectors, which are used for intermediate results.

To compute the second and third terms of Equation (11), we formalize the bivariate distributions $\vec{S} = (X_{S_{12}} + X_{S_1}, X_{S_{12}} + X_{S_2})^{\mathsf{T}}$ and $\vec{O} = (O_1, O_2)^{\mathsf{T}}$. We denote $\mu_P = \sum_i \mu_{P_i}$ and $\sigma_P^2 = \sum_i \sigma_{P_i}^2$ as the sum of the means and standard deviations, respectively, of all participants within a group P. Note that the mean is absent from the formula for the differential entropy, and therefore we can safely assume all $\mu_i = 0$. Starting with $\vec{O}$, we invoke the linearity of the expectation for the mean vector:

$$\boldsymbol{\mu}_{\vec{O}} = \begin{pmatrix} \mathbb{E}[O_1] \\ \mathbb{E}[O_2] \end{pmatrix} = \begin{pmatrix} \mathbb{E}[X_T + X_{S_{12}} + X_{S_1}] \\ \mathbb{E}[X_T + X_{S_{12}} + X_{S_2}] \end{pmatrix} = \begin{pmatrix} \mu_T + \mu_{S_{12}} + \mu_{S_1} \\ \mu_T + \mu_{S_{12}} + \mu_{S_2} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}.$$

For the covariance matrix, using the properties $\mathrm{Cov}[X, X] = \mathrm{Var}[X] = \sigma_X^2$ and $\mathrm{Cov}[X, Y] = \mathrm{Cov}[Y, X]$ yields

$$\begin{aligned} \Sigma_{\vec{O}} &= \begin{pmatrix} \mathrm{Cov}[O_1, O_1] & \mathrm{Cov}[O_1, O_2] \\ \mathrm{Cov}[O_2, O_1] & \mathrm{Cov}[O_2, O_2] \end{pmatrix} = \begin{pmatrix} \mathrm{Var}[O_1] & \mathrm{Cov}[O_1, O_2] \\ \mathrm{Cov}[O_1, O_2] & \mathrm{Var}[O_2] \end{pmatrix} \\ &= \begin{pmatrix} \sigma_T^2 + \sigma_{S_{12}}^2 + \sigma_{S_1}^2 & \mathrm{Cov}[O_1, O_2] \\ \mathrm{Cov}[O_1, O_2] & \sigma_T^2 + \sigma_{S_{12}}^2 + \sigma_{S_2}^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \mathrm{Cov}[O_1, O_2] \\ \mathrm{Cov}[O_1, O_2] & \sigma_2^2 \end{pmatrix}. \end{aligned}$$

The expression for $\mathrm{Cov}[O_1, O_2]$ can be stated as follows:

CLAIM 4. $\mathrm{Cov}[O_1, O_2] = \sigma_T^2 + \sigma_{S_{12}}^2$ if S_{12} is non-empty, and $\mathrm{Cov}[O_1, O_2] = \sigma_T^2$ otherwise.

PROOF.

$\mathrm{Cov}[O_1, O_2]$

$= \mathbb{E}[(O_1 - \mu_1)(O_2 - \mu_2)] = \mathbb{E}[O_1 O_2 - \mu_2 O_1 - \mu_1 O_2 + \mu_1 \mu_2]$

$= \mathbb{E}[(X_T + X_{S_{12}} + X_{S_1})(X_T + X_{S_{12}} + X_{S_2})] - \mathbb{E}[\mu_2(X_T + X_{S_{12}} + X_{S_1})] - \mathbb{E}[\mu_1(X_T + X_{S_{12}} + X_{S_2})] + \mathbb{E}[\mu_1 \mu_2]$

$= \mathbb{E}[X_T^2] + \mathbb{E}[X_{S_{12}}^2] + 2\mathbb{E}[X_T X_{S_{12}}] + \mathbb{E}[X_T X_{S_1}] + \mathbb{E}[X_T X_{S_2}] + \mathbb{E}[X_{S_{12}} X_{S_1}] + \mathbb{E}[X_{S_{12}} X_{S_2}]$

$\qquad + \mathbb{E}[X_{S_1} X_{S_2}] - \mu_2 \underbrace{(\mathbb{E}[X_T] + \mathbb{E}[X_{S_{12}}] + \mathbb{E}[X_{S_1}])}_{\mu_1} - \mu_1 \underbrace{(\mathbb{E}[X_T] + \mathbb{E}[X_{S_{12}}] + \mathbb{E}[X_{S_2}])}_{\mu_2} + \mu_1 \mu_2.$

Exploiting the definition of variance $\mathbb{E}[X^2] = \sigma_X^2 + \mu_X^2$ and fundamental properties of expectation:

$$\text{Cov}[O_1, O_2] = \sigma_T^2 + \sigma_{S_{12}}^2 - \mu_1\mu_2 + \underbrace{\mu_T^2 + \mu_{S_{12}}^2 + 2\mu_T\mu_{S_{12}} + \mu_T\mu_{S_1} + \mu_T\mu_{S_2} + \mu_{S_{12}}\mu_{S_1} + \mu_{S_{12}}\mu_{S_2} + \mu_{S_1}\mu_{S_2}}_{=\mu_1\mu_2}$$

$$= \sigma_T^2 + \sigma_{S_{12}}^2.$$

Clearly, if $S_{12} = \emptyset$, then the above result simplifies to $\text{Cov}\,[O_1, O_2] = \sigma_T^2$. This result is intuitive, since the covariance measures the strength of correlation between two random variables, and O_1 and O_2 are both composed of X_T and $X_{S_{12}}$. $\qquad\square$

The final parameters of the bivariate distribution $\vec{O}$ are

$$\boldsymbol{\mu}_{\vec{O}} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Sigma_{\vec{O}} = \begin{pmatrix} \sigma_1^2 & \sigma_T^2 + \sigma_{S_{12}}^2 \\ \sigma_T^2 + \sigma_{S_{12}}^2 & \sigma_2^2 \end{pmatrix}.$$

Repeating this procedure for the spectator joint distribution $\vec{S}$ yields a similar set of parameters:

$$\boldsymbol{\mu}_{\vec{S}} = \begin{pmatrix} \mu_{S_{12}} + \mu_{S_1} \\ \mu_{S_{12}} + \mu_{S_2} \end{pmatrix}, \Sigma_{\vec{S}} = \begin{pmatrix} \sigma_{S_{12}}^2 + \sigma_{S_1}^2 & \sigma_{S_{12}}^2 \\ \sigma_{S_{12}}^2 & \sigma_{S_{12}}^2 + \sigma_{S_2}^2 \end{pmatrix}.$$

Equipped with expressions for $\Sigma_{\vec{O}}$ and $\Sigma_{\vec{S}}$, we are prepared to begin our experimental analysis of $h(X_T \mid O_1, O_2)$.

5.2 Experimental Evaluation

The above allows us to experimentally evaluate the target's entropy loss for when inputs are normally distributed. We use normal distribution $\mathcal{N}(0, 4)$ to reasonably approximate the log-normal distribution with real data. Once again, $|T| = 1$ for concreteness, and we let $|S_1| = |S_2|$ in all experiments, i.e., the number of spectators is the same in both executions.

It is informative to analyze information loss as the fraction of shared spectators changes, and we do so for three different computation sizes. To be as close to the setup that guarantees 1%–5% entropy loss for the log-normal distribution (5–24 spectators), we choose to execute our experiments with 6, 10, and 24 spectators (where having an even number is beneficial for illustration purposes). This corresponds to the number of non-adversarial participants when the target is absent and the number of non-adversarial participants is one higher when the target is participating.

We display the following information in Figure 10:

— the target's initial entropy $h(\vec{X}_T)$,
— the target's entropy after a single execution $h(\vec{X}_T \mid O_1)$,
— the target's entropy after participating twice $h(\vec{X}_T \mid O_1, O_2)$,
— the target's entropy after participating in one of the two executions, i.e., $h(\vec{X}_T \mid O_1, O_2')$ and $h(\vec{X}_T \mid O_1', O_2)$

and plot the values as a function of the fractional overlap between two executions for a given number of spectators.

Naturally, the value of $h(\vec{X}_T \mid O_1)$ remains constant when the number of participants is fixed. We observe that when participating twice, $h(\vec{X}_T \mid O_1, O_2)$ converges to $h(\vec{X}_T \mid O_1)$ as the fraction of shared spectators increases. This is expected, because at 100% overlap, we are functionally calculating $h(\vec{X}_T \mid O_1, O_1) = h(\vec{X}_T \mid O_1)$. We formalize this into the claim:

CLAIM 5. *If the target participates in both evaluations and 100% of the spectators are reused, then* $h(\vec{X}_T \mid O_1, O_2) = h(\vec{X}_T \mid O_1)$.

(a) $n = 6$ spectators per execution. (b) $n = 10$ spectators per execution. (c) $n = 24$ spectators per execution.

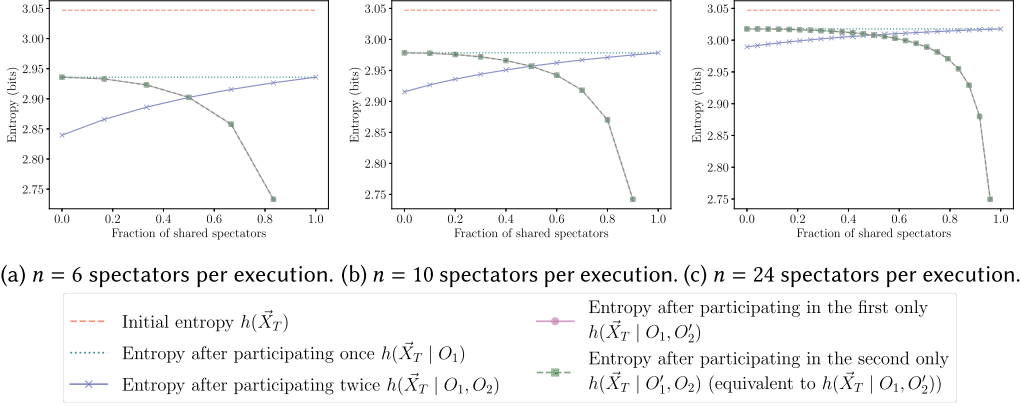|	
- - - - Initial entropy $h(\vec{X}_T)$	Entropy after participating in the first only $h(\vec{X}_T \mid O_1, O_2')$
⋯⋯ Entropy after participating once $h(\vec{X}_T \mid O_1)$	Entropy after participating in the second only $h(\vec{X}_T \mid O_1', O_2)$ (equivalent to $h(\vec{X}_T \mid O_1, O_2')$)
—×— Entropy after participating twice $h(\vec{X}_T \mid O_1, O_2)$	

Fig. 10. Target information loss after participating in one or two computations. Omitted: If the target participates in one experiment and all the shared spectators are reused, then $h(X_T \mid O_1, O_2') = 0$.

PROOF. We begin by analyzing the absolute loss between the first and second evaluations when the target participates twice, namely:

$$h(\vec{X}_T \mid O_1) - h(\vec{X}_T \mid O_1, O_2).$$

Assume all participants' inputs are normally distributed ($X_i \sim \mathcal{N}(0, \sigma^2)$). Denote $p = |P|$ as the size of an arbitrary group P (e.g., $s_{12} = |S_{12}|$), such that $X_P \sim \mathcal{N}(0, p\sigma^2)$. Simplifying the absolute loss between the first and second evaluations, we obtain:

$$h(\vec{X}_T \mid O_1) - h(\vec{X}_T \mid O_1, O_2) = h(\vec{X}_T) + h(X_{S_{12}} + X_{S_1}) - h(X_T + X_{S_{12}} + X_{S_1})$$
$$- (h(\vec{X}_T) + h(X_{S_{12}} + X_{S_1}, X_{S_{12}} + X_{S_2}) - h(O_1, O_2))$$
$$= h(X_{S_{12}} + X_{S_1}) - h(X_T + X_{S_{12}} + X_{S_1}) + h(O_1, O_2) - h(X_{S_{12}} + X_{S_1}, X_{S_{12}} + X_{S_2}).$$

Using the definitions from Section 5.1, we calculate the remaining terms as

$$h(X_{S_{12}} + X_{S_1}) + h(X_T + X_{S_{12}} + X_{S_1}) = \frac{1}{2} \log \left(\frac{s_{12} + s_1}{t + s_{12} + s_1} \right)$$

$$h(O_1, O_2) = \frac{1}{2} \log \left((2\pi e)^2 ((t + s_{12})(s_1 + s_2) + s_1 s_2) \sigma^2 \right)$$

$$h(X_{S_{12}} + X_{S_1}, X_{S_{12}} + X_{S_2}) = \frac{1}{2} \log \left((2\pi e)^2 (s_{12}(s_1 + s_2) + s_1 s_2) \sigma^2 \right)$$

$$h(O_1, O_2) - h(X_{S_{12}} + X_{S_1}, X_{S_{12}} + X_{S_2}) = \frac{1}{2} \log \left(\frac{(t + s_{12})(s_1 + s_2) + s_1 s_2}{s_{12}(s_1 + s_2) + s_1 s_2} \right).$$

Therefore, the absolute entropy loss between the first and second evaluations is

$$h(\vec{X}_T \mid O_1) - h(\vec{X}_T \mid O_1, O_2) = \left(\frac{1}{2} \log \left(\frac{s_{12} + s_1}{t + s_{12} + s_1} \right) \right) + \frac{1}{2} \log \left(\frac{(t + s_{12})(s_1 + s_2) + s_1 s_2}{s_{12}(s_1 + s_2) + s_1 s_2} \right)$$
$$= \frac{1}{2} \log \left(\left(\frac{s_{12} + s_1}{t + s_{12} + s_1} \right) \left(\frac{t(s_1 + s_2) + s_{12}(s_1 + s_2) + s_1 s_2}{s_{12}(s_1 + s_2) + s_1 s_2} \right) \right).$$

Since we assume $s_1 = s_2$, the above expression simplifies to

$$h(\vec{X}_T \mid O_1) - h(\vec{X}_T \mid O_1, O_2) = \frac{1}{2} \log \left(\left(\frac{s_{12} + s_1}{t + s_{12} + s_1} \right) \left(\frac{2t + 2s_0 + s_1}{2s_0 + s_1} \right) \right).$$

This function is monotonically decreasing when the total number of spectators is fixed to $s_{12} + s_1$, and we vary the ratio $\frac{s_{12}}{s_1} \in [0, 1]$, which is consistent with our observation that the absolute loss will converge to $h(\vec{X}_T \mid O_1)$.   □

Conversely, increasing the fraction of the overlap has the inverse effect for $h(\vec{X}_T \mid O_1, O_2')$, causing it to trend downward. At 100% overlap, $h(\vec{X}_T \mid O_1, O_2') = 0$ (point omitted from the plots). This is a consequence of effectively computing $h(\vec{X}_T \mid O_1, X_{S_{12}})$:

$$h(\vec{X}_T | O_1, X_{S_{12}}) = h(\vec{X}_T, O_1, X_{S_{12}}) - h(O_1, X_{S_{12}}) = h(\vec{X}_T) + h(X_{S_{12}}) - (h(X_T + X_{S_{12}} \mid X_{S_{12}}) + h(X_{S_{12}}))$$

$$= h(\vec{X}_T) + h(X_{S_{12}}) - (h(X_T) + h(X_{S_{12}})) = h(\vec{X}_T) - h(X_T).$$

When $|T| = 1$, then $h(\vec{X}_T) = h(X_T)$, thus reducing the above equation to zero. This informs us that the output of the second computation O_2' without any unique spectators reveals the target's information entirely. We state this observation as follows:

CLAIM 6. *If the target participates in one evaluation and 100% of the spectators are reused, then* $h(\vec{X}_T \mid O_1, O_2') = 0$.

PROOF. Next, we examine the absolute entropy loss between the first and second evaluations when the target participates in only the first evaluation:

$$h(\vec{X}_T \mid O_1) - h(\vec{X}_T \mid O_1, O_2').$$

The only difference from the prior calculation arises from replacing $h(O_1, O_2)$ with $h(O_1, O_2')$, which evaluates to

$$h(O_1, O_2) = \frac{1}{2} \log \left((2\pi e)^2 (t(s_{12} + s_2) + s_{12}(s_1 + s_2) + s_1 s_2) \sigma^2 \right),$$

such that our final expression is

$$h(\vec{X}_T \mid O_1) - h(\vec{X}_T \mid O_1, O_2') = \frac{1}{2} \log \left(\left(\frac{s_{12} + s_1}{t + s_{12} + s_1} \right) \left(\frac{t(s_{12} + s_2) + s_{12}(s_1 + s_2) + s_1 s_2}{s_{12}(s_1 + s_2) + s_1 s_2} \right) \right)$$

$$= \frac{1}{2} \log \left(\left(\frac{s_{12} + s_1}{t + s_{12} + s_1} \right) \left(\frac{t(s_{12} + s_1) + s_1(2s_0 + s_1)}{s_1(2s_0 + s_1)} \right) \right).$$

This function equals to infinity when $s_1 = s_2 = 0$, which confirms that the output of the second computation O_2' without the presence of any unique spectators reveals the target's information entirely.   □

A passive result of both proofs is that all forms of absolute loss are parameter-independent, which is consistent with Claim 2.

Our next observation pertains to the point of intersection where $h(\vec{X}_T \mid O_1, O_2) = h(\vec{X}_T \mid O_1, O_2')$, which occurs when 50% of the spectators are shared across the computation. This appears for the special case when the total number of spectators in a single evaluation is even. Concretely, we compare

$$h(\vec{X}_T | O_1, O_2) = h(\vec{X}_T, O_1, O_2) - h(O_1, O_2),$$
$$h(\vec{X}_T | O_1, O_2') = h(\vec{X}_T, O_1, O_2') - h(O_1, O_2'). \tag{12}$$

It can be shown using the procedure outlined in Section 5 that $h(\vec{X}_T, O_1, O_2) = h(\vec{X}_T, O_1, O_2')$. Therefore, we prove the following:

CLAIM 7. *With normally distributed inputs, the terms $h(O_1, O_2)$ and $h(O_1, O_2')$ are equal when* $|S_{12}| = |S_1|$.

Table 1. Percentage of Information Loss after Two
Executions Relative to a Single Execution for $n = 10$.

Number of evaluations the target participates in	Spectator overlap		
	40%	50%	60%
One	18.0%	31.3%	52.3%
Two	40.1%	31.3%	23.5%

PROOF OF CLAIM 7. Following the steps used to derive the covariance matrix of $\vec{O} = (O_1, O_2)$, the covariance matrix of $\vec{O}' = (O_1, O_2')$ is

$$\Sigma_{\vec{O}'} = \begin{pmatrix} \sigma_T^2 + \sigma_{S_{12}}^2 + \sigma_{S_1}^2 & \sigma_{S_{12}}^2 \\ \sigma_{S_{12}}^2 & \sigma_{S_{12}}^2 + \sigma_{S_2}^2 \end{pmatrix}.$$

Recall that the differential entropy of the multivariate normal is $h(\vec{X}) = \frac{1}{2} \log((2\pi e)^k \det \Sigma)$. The sole object of interest is the $\det \Sigma$ term, as the remainder contribute a constant factor. We calculate

$$\det \Sigma_{\vec{O}} = \left(\sigma_T^2 + \sigma_{S_{12}}^2 + \sigma_{S_1}^2 \right) \left(\sigma_T^2 + \sigma_{S_{12}}^2 + \sigma_{S_2}^2 \right) - \left(\sigma_T^2 + \sigma_{S_{12}}^2 \right)^2$$

$$= \sigma_T^2 \left(\sigma_{S_1}^2 + \sigma_{S_2}^2 \right) + \sigma_{S_{12}}^2 \left(\sigma_{S_1}^2 + \sigma_{S_2}^2 \right) + \sigma_{S_1}^2 \sigma_{S_2}^2.$$

Similarly,

$$\det \Sigma_{\vec{O}'} = \left(\sigma_T^2 + \sigma_{S_{12}}^2 + \sigma_{S_1}^2 \right) \left(\sigma_{S_{12}}^2 + \sigma_{S_2}^2 \right) - \sigma_{S_{12}}^4 = \sigma_T^2 \left(\sigma_{S_{12}}^2 + \sigma_{S_2}^2 \right) + \sigma_{S_{12}}^2 \left(\sigma_{S_1}^2 + \sigma_{S_2}^2 \right) + \sigma_{S_1}^2 \sigma_{S_2}^2.$$

Therefore, the equality $h(\vec{X}_T \mid O_1, O_2) = h(\vec{X}_T \mid O_1, O_2')$ is satisfiable if and only if $\sigma_{S_{12}}^2 = \sigma_{S_1}^2$, which occurs when $|S_{12}| = |S_1|$. □

As computation designers, we can minimize information disclosure for all participants by targeting 50% of participants' overlap between the first and second executions. For the configurations in Figure 10, at 50% overlap, the percentages of information loss from the second evaluation relative to the first evaluation are comparable for the selected number of spectators n (30.18% for $n = 6$, 31.3% for $n = 10$, and 32.45% for $n = 24$). This corresponds to the intersection points in Figure 10.

As we may be unable to guarantee that exactly 50% of participants overlap between two executions, we can increase our tolerance for entropy loss by inviting more participants and building a buffer to accommodate overlaps in a range, e.g., 40%–60%. Using data in Figure 10, this information is convenient to gather for $n = 10$. That is, if we increase the fraction of overlapping spectators, then single-participation targets are most at risk. The converse is true if the overlap decreases—the target suffers less exposure from participating one evaluation. Table 1 summarizes the results. This means that performing two executions in the worst case costs a participant entropy loss 1.5 times higher than if only a single computation is executed. As a result, with the target entropy loss of 5% and 1%, we need to increase the number of spectators from 5 and 24 to 7 and 33, respectively.

We note that our analysis of repeated executions applies only when the inputs of the participants in the overlapping set of participants do not change. And if the executions are distant enough in time that the participants' inputs significantly change, then they would no longer be treated as repeated dependent executions.

5.3 Additional Two Executions Experiments

We examine the impact of shared spectators' presence on target's information loss. In Figure 11(a), we plot:
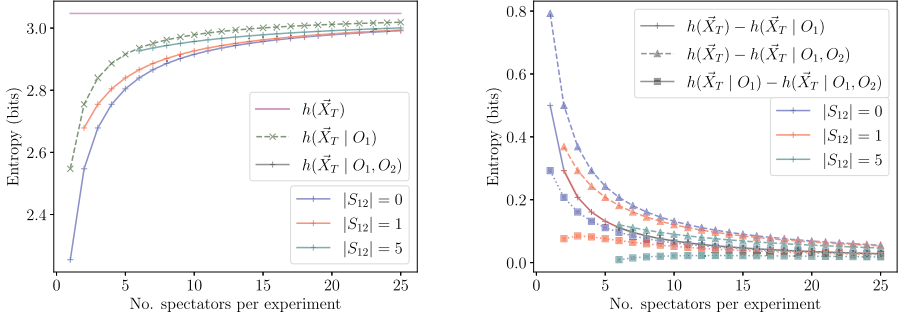
(a) Computing $h(X_T)$, $h(X_T \mid O_1)$, and $h(X_T \mid O_1, O_2)$ for several $|S_{12}|$ sizes.

(b) Absolute entropy loss.

Fig. 11. Comparing the relative and absolute entropy losses of participants with normally distributed inputs. The number of spectators per experiment on the x-axis is computed as $|S_{12} \cup S_1| = |S_{12} \cup S_2|$, starting with $|S_1| = |S_2| = 1$.

— the target's initial entropy $h(\vec{X}_T)$,
— the target's entropy after a single execution $h(\vec{X}_T \mid O_1)$, and
— the target's entropy after two executions $h(\vec{X}_T \mid O_1, O_2)$ with a different number of spectators participating in both executions.

We vary the total number of spectators per experiment $|S_{12} \cup S_1| = |S_{12} \cup S_2|$ on the x-axis, starting with one unique spectator per experiment $|S_1| = |S_2| = 1$. The $h(\vec{X}_T \mid O_1, O_2)$ curves correspond to awae after two executions and start when the number of participants reaches their respective number of shared spectators $|S_{12}|$ to make an accurate comparison. A single curve for $h(\vec{X}_T \mid O_1)$ suffices, since it does not use the notion of shared spectators.

We observe in Figure 11(a) that the larger the number of shared spectators for a given $|S_0|$ is, the less information is revealed about the target. These spectators function as "noise" that protects the target. The protection offered by a small number of shared spectators becomes less pronounced as the number of participants grows.

We also compute and present in Figure 11(b) the target's absolute entropy loss for the following experiments:

— after a single execution $h(\vec{X}_T) - h(\vec{X}_T \mid O_1)$,
— after two executions $h(\vec{X}_T) - h(\vec{X}_T \mid O_1, O_2)$, and
— after the second execution $h(\vec{X}_T \mid O_1) - h(\vec{X}_T \mid O_1, O_2)$

using a varying number of shared spectators $|S_{12}|$. We see that for each fixed number of shared spectators $|S_{12}|$, the absolute loss as a result of the first participation ($h(\vec{X}_T) - h(\vec{X}_T \mid O_1)$) is greater than the absolute loss of the second participation ($h(\vec{X}_T \mid O_1) - h(\vec{X}_T \mid O_1, O_2)$). With no shared spectators, the curves converge at about 15 participants per execution, while increasing the number of shared spectators causes the curves to converge at a slower rate.

5.4 Mixed Distribution Parameters for Two Executions

In Section 4.6, we examined how our conclusions changed when we generalize our analysis to non-identically distributed participant inputs. We are similarly interested in how this affects our two-execution analysis. Recalling the formulation of our two-evaluation setting where spectators are present in the first, second, or both executions, we now consider these spectator subsets can be

further partitioned into sub-subsets based on their group identities. Combining our two-evaluation notation and that of Section 4.6, we define $P_{k,G} \subseteq P$ as the set of participants present in execution(s) $k = \{1, 2, 12\}$ belonging to group $G \in \mathbb{G}$. For example, $k = 1$ means participation in the first execution, while $k = 12$ means participation in both. The sum of these participants' inputs is

$$X_{P_k} = \sum_{G \in \mathbb{G}} \sum_{i \in P_{k,G}} X_{P_{k,i}} = \sum_{G \in \mathbb{G}} X_{P_{k,G}},$$

where $X_{P_G} = \sum_{i \in P_{k,G}} X_{P_{k,i}}$.

Optimal setup for minimizing information disclosure. In Section 5.2, we determined that the point of intersection of the entropies $h(\vec{X}_T \mid O_1, O_2)$ and $h(\vec{X}_T \mid O_1, O_2')$ at 50% participant overlap provides the best level of protection for all types of targets, and moving in either direction (increasing or decreasing the overlap) causes the leakage to increase. We revisit our proof of Claim 7 in our generalized setting to determine whether the equality $h(\vec{X}_T \mid O_1, O_2) = h(\vec{X}_T \mid O_1, O_2')$ still holds when 50% of the spectators are shared across the computation, i.e., $|S_{12}| = |S_1|$.

Considering the expansions of $h(\vec{X}_T \mid O_1, O_2)$ and $h(\vec{X}_T \mid O_1, O_2')$ in Equation (12), it can be shown that the terms $h(\vec{X}_T, O_1, O_2)$ and $h(\vec{X}_T, O_1, O_2')$ are equal, leaving $h(O_1, O_2)$ and $h(O_1, O_2')$ for us to compare. We perform this analysis under Case 1 (participant group identities are known), since the proof of Claim 7 relies on the existence of the closed-form expression for the differential entropy.

Computing $h(O_1, O_2)$ and $h(O_1, O_2')$ in the generalized setting requires re-formalizing the covariance matrices of the joint random variables $\vec{O} = (O_1, O_2)$ and $\vec{O'} = (O_1, O_2')$. Performing the steps outlined in Section 5.1 yields the following:

$$\Sigma_{\vec{O}} = \begin{pmatrix} \sum_{G \in \mathbb{G}} \sigma_G^2 \left(|T_G| + |S_{12,G}| + |S_{1,G}| \right) & \sum_{G \in \mathbb{G}} \sigma_G^2 \left(|T_G| + |S_{12,G}| \right) \\ \sum_{G \in \mathbb{G}} \sigma_G^2 \left(|T_G| + |S_{12,G}| \right) & \sum_{G \in \mathbb{G}} \sigma_G^2 \left(|T_G| + |S_{12,G}| + |S_{2,G}| \right) \end{pmatrix}$$

$$\Sigma_{\vec{O'}} = \begin{pmatrix} \sum_{G \in \mathbb{G}} \sigma_G^2 \left(|T_G| + |S_{12,G}| + |S_{1,G}| \right) & \sum_{G \in \mathbb{G}} \sigma_G^2 |S_{12,G}| \\ \sum_{G \in \mathbb{G}} \sigma_G^2 |S_{12,G}| & \sum_{G \in \mathbb{G}} \sigma_G^2 \left(|S_{12,G}| + |S_{2,G}| \right) \end{pmatrix}.$$

Given the definition of the differential entropy of a multivariate normal as $h(\vec{X}) = \frac{1}{2} \log((2\pi e)^k \det \Sigma)$, we compute the determinants of the above matrices as:

$$\det \Sigma_{\vec{O}} = \left(\sum_{G \in \mathbb{G}} \sigma_G^2 \left(|T_G| + |S_{12,G}| + |S_{1,G}| \right) \right) \left(\sum_{G \in \mathbb{G}} \sigma_G^2 \left(|T_G| + |S_{12,G}| + |S_{2,G}| \right) \right) - \left(\sum_{G \in \mathbb{G}} \sigma_G^2 \left(|T_G| + |S_{12,G}| \right) \right)^2$$

$$= \sum_{G \in \mathbb{G}} \sigma_G^2 |T_G| \left(\sum_{G \in \mathbb{G}} \left(\sigma_G^2 \left(|S_{1,G}| + |S_{2,G}| \right) \right) \right) + \sum_{G \in \mathbb{G}} \sigma_G^2 |S_{(12,G)}| \left(\sum_{G \in \mathbb{G}} \left(\sigma_G^2 \left(|S_{1,G}| + |S_{2,G}| \right) \right) \right)$$

$$+ \sum_{G \in \mathbb{G}} \sigma_G^2 \left(|S_{1,G}| \right) \sum_{G \in \mathbb{G}} \sigma_G^2 \left(|S_{2,G}| \right)$$

$$\det \Sigma_{\vec{O'}} = \left(\sum_{G \in \mathbb{G}} \sigma_G^2 \left(|T_G| + |S_{12,G}| + |S_{1,G}| \right) \right) \left(\sum_{G \in \mathbb{G}} \sigma_G^2 \left(|S_{12,G}| + |S_{1,G}| \right) \right) - \left(\sum_{G \in \mathbb{G}} \sigma_G^2 \left(|S_{12,G}| \right) \right)^2$$

$$= \sum_{G \in \mathbb{G}} \sigma_G^2 |T_G| \left(\sum_{G \in \mathbb{G}} \left(\sigma_G^2 \left(|S_{12,G}| + |S_{2,G}| \right) \right) \right) + \sum_{G \in \mathbb{G}} \sigma_G^2 |S_{(12,G)}| \left(\sum_{G \in \mathbb{G}} \left(\sigma_G^2 \left(|S_{1,G}| + |S_{2,G}| \right) \right) \right)$$

$$+ \sum_{G \in \mathbb{G}} \sigma_G^2 \left(|S_{1,G}| \right) \sum_{G \in \mathbb{G}} \sigma_G^2 \left(|S_{2,G}| \right).$$

By inspection, we obtain that the equality $h(\vec{X}_T \mid O_1, O_2) = h(\vec{X}_T \mid O_1, O_2')$ is satisfiable if and only if $\sum_{G \in \mathbb{G}} \left(\sigma_G^2 \left|S_{12,G}\right|\right) = \sum_{G \in \mathbb{G}} \left(\sigma_G^2 \left|S_{1,G}\right|\right)$. However, this no longer implies that the optimal configuration is at 50% overlap. For $|\mathbb{G}| > 1$, there can be multiple solutions with respect to the individual group sizes, statistical parameters, and overlap percentages such that the equality can be satisfied.

6 Three Executions and Beyond

The next logical step is to further generalize our analysis to three and any number M executions.

6.1 Three Executions

For three evaluations, there are additional possibilities for spectators to overlap between experiments. Specifically, we have:

— spectators who participate in one experiment (S_1, S_2, S_3),
— spectators who participate in two experiments but not a third (S_{12}, S_{13}, S_{32}), and
— spectators who participate in all three experiments (S_{123}).

Let n be the (fixed) total number of spectators per experiment. For each evaluation, let superscript (τ_i) denote a target's participation flag defined as:

$$
\tau_i = \begin{cases} 0 & T \text{ does } not \text{ participate in evaluation } i \\ 1 & T \text{ participates in evaluation } i \end{cases} .
$$

We require $\sum_{i=1}^{3} \tau_i > 0$ to signify that the target participates at least once. Therefore, there are $2^3 - 1 = 7$ possible target configurations. For example, $(\tau_1, \tau_2, \tau_3) = (1, 0, 1)$ means the target participated in the first and third executions. We use this notation to generate expressions for all configurations of the targets' participation in evaluations. The random variables for each evaluation are:

$$
O_1^{(\tau_1)} = \tau_1 \cdot X_T + X_{S_1} + X_{S_{12}} + X_{S_{13}} + X_{S_{123}} = \tau_1 \cdot X_T + X_{\hat{S}_1}
$$
$$
O_2^{(\tau_2)} = \tau_2 \cdot X_T + X_{S_2} + X_{S_{12}} + X_{S_{23}} + X_{S_{123}} = \tau_2 \cdot X_T + X_{\hat{S}_2}
$$
$$
O_3^{(\tau_3)} = \tau_3 \cdot X_T + X_{S_2} + X_{S_{23}} + X_{S_{13}} + X_{S_{123}} = \tau_3 \cdot X_T + X_{\hat{S}_3},
$$

where $X_{\hat{S}_i}$ is the sum of all spectator configurations in evaluation i. If we denote $p = |P|$ as the size of an arbitrary group P such that $X_P \sim \mathcal{N}(0, p\sigma^2)$, then the covariance matrix for the random vector $\vec{O}_{1,2,3} = (O_1^{(\tau_1)}, O_2^{(\tau_2)}, O_3^{(\tau_3)})^{\mathrm{T}}$ is

$$
\Sigma_{\vec{O}_{1,2,3}} = \begin{pmatrix} \mathrm{Cov}[O_1^{(\tau_1)}, O_1^{(\tau_1)}] & \mathrm{Cov}[O_1^{(\tau_1)}, O_2^{(\tau_2)}] & \mathrm{Cov}[O_1^{(\tau_1)}, O_3^{(\tau_3)}] \\ \mathrm{Cov}[O_2^{(\tau_2)}, O_1^{(\tau_1)}] & \mathrm{Cov}[O_2^{(\tau_2)}, O_2^{(\tau_2)}] & \mathrm{Cov}[O_2^{(\tau_2)}, O_3^{(\tau_3)}] \\ \mathrm{Cov}[O_3^{(\tau_3)}, O_1^{(\tau_1)}] & \mathrm{Cov}[O_3^{(\tau_3)}, O_2^{(\tau_2)}] & \mathrm{Cov}[O_3^{(\tau_3)}, O_3^{(\tau_3)}] \end{pmatrix}
$$

$$
= \begin{pmatrix} \begin{pmatrix} \tau_1 \cdot t + s_1 + s_{12} \\ + s_{13} + s_{123} \end{pmatrix} & \tau_1 \tau_2 \cdot t + s_{12} + s_{123} & \tau_1 \tau_3 \cdot t + s_{13} + s_{123} \\ \tau_1 \tau_2 \cdot t + s_{12} + s_{123} & \begin{pmatrix} \tau_2 \cdot t + s_2 + s_{12} \\ + s_{23} + s_{123} \end{pmatrix} & \tau_2 \tau_3 \cdot t + s_{23} + s_{123} \\ \tau_1 \tau_3 \cdot t + s_{13} + s_{123} & \tau_2 \tau_3 \cdot t + s_{23} + s_{123} & \begin{pmatrix} \tau_3 \cdot t + s_3 + s_{23} \\ + s_{13} + s_{123} \end{pmatrix} \end{pmatrix} \sigma^2
$$

$$
= \begin{pmatrix} \tau_1 \cdot t + n & \tau_1 \tau_2 \cdot t + s_{12} + s_{123} & \tau_1 \tau_3 \cdot t + s_{13} + s_{123} \\ \tau_1 \tau_2 \cdot t + s_{12} + s_{123} & \tau_2 \cdot t + n & \tau_2 \tau_3 \cdot t + s_{23} + s_{123} \\ \tau_1 \tau_3 \cdot t + s_{13} + s_{123} & \tau_2 \tau_3 \cdot t + s_{23} + s_{123} & \tau_3 \cdot t + n \end{pmatrix} \sigma^2.
$$

The second covariance matrix required is for the random vector $\vec{S}_{1,2,3} = (X_{\hat{S}_1}, X_{\hat{S}_2}, X_{\hat{S}_3})^{\mathrm{T}}$ and is

$$\Sigma_{\vec{S}_{1,2,3}} = \begin{pmatrix} n & s_{12}+s_{123} & s_{13}+s_{123} \\ s_{12}+s_{123} & n & s_{23}+s_{123} \\ s_{13}+s_{123} & s_{23}+s_{123} & n \end{pmatrix} \sigma^2.$$

With these matrices, we are capable of computing the conditional entropy $h(\vec{X}_T \mid O_1^{(\tau_1)}, O_2^{(\tau_2)}, O_3^{(\tau_3)})$. It will be important later that the above covariance matrices only depend on pairwise spectator overlaps between the executions $(s_{13} + s_{123})$, $(s_{12} + s_{123})$, and $(s_{23} + s_{123})$, rather than individual sets s_{12}, s_{23}, s_{123}, and so on.

6.2 M Executions

We can generalize the prior section's analysis to obtain the target's conditional entropy for an arbitrary number of evaluations. Let M be to the total number of evaluations where $M \in \mathbb{Z}_{>0}$. Denote A as the set of integers from 1 to M, such that $A = \{1, \ldots, M\}$. We can generate the set of all subsets of spectators that overlap and do not overlap between evaluation using the power set of A (denoted by $\mathcal{P}(A)$). Specifically $\mathcal{S} = \mathcal{P}(A) \setminus \{\emptyset\}$, the empty set is excluded, as it corresponds to the target not participating in any computation. The number of spectator subsets and target participation configurations is $|\mathcal{S}| = 2^M - 1$. The output random variable of experiment $i \in \{1, \ldots, M\}$ is therefore

$$O_i^{(\tau_i)} = \tau_i \cdot X_T + \sum_{R \subseteq \mathcal{S}: i \in R} X_{S_R} = \tau_i \cdot X_T + X_{\hat{S}_i}.$$

We can generate elements of the $M \times M$ covariance matrix of the random vector $\vec{O}_{1,\ldots,M} = (O_1^{(\tau_1)}, \ldots, O_M^{(\tau_M)})^{\mathrm{T}}$ using the following expression for $i, j \in \{1, \ldots, M\}$:

$$\mathrm{Cov}\left[O_i^{(\tau_i)}, O_j^{(\tau_j)}\right] = \begin{cases} \tau_i \tau_j \cdot \sigma_T^2 + \sum_{R \subseteq \mathcal{S}:(i,j) \in R} \sigma_R^2 & \text{if } i \neq j \\ \tau_i \cdot \sigma_T^2 + \sum_{R \subseteq \mathcal{S}: i \in R} \sigma_R^2 & \text{if } i = j \end{cases}.$$

Similarly, elements of the covariance matrix of the random vector $\vec{S}_{1,\ldots,M} = (X_{\hat{S}_1}, \ldots, X_{\hat{S}_M})^{\mathrm{T}}$ can be generated as follows:

$$\mathrm{Cov}[X_{\hat{S}_i}, X_{\hat{S}_j}] = \begin{cases} \sum_{R \subseteq \mathcal{S}:(i,j) \in R} \sigma_R^2 & \text{if } i \neq j \\ \sum_{R \subseteq \mathcal{S}: i \in R} \sigma_R^2 & \text{if } i = j \end{cases}.$$

If the total number of spectators per evaluation is fixed to n, then $\sum_{R \subseteq \mathcal{S}: i \in R} \sigma_R^2 = \sigma^2 n$.

6.3 Experimental Evaluation

Unlike two executions, we can no longer graphically represent conditional entropy as a function of overlap sizes, as there are several dimensions to consider. Instead, we enumerate all possible spectator configurations for $n = 24$, and for each spectator configuration, we compute the minimum of the seven conditional entropies corresponding to valid target configurations τ_1, τ_2, τ_3:

$$\min_{\tau_1, \tau_2, \tau_3} h\left(\vec{X}_T \mid O_1^{(\tau_1)}, O_2^{(\tau_2)}, O_3^{(\tau_3)}\right).$$

We then determine the maximums across all spectator configurations that correspond to the optimal choices that minimize target's information disclosure.

We plot the top 500 spectator configurations, which yield 20 unique entropy values (displayed in the color map), in Figure 12 from different viewing angles. The axes correspond to pairwise overlap sizes, and a point with a fixed overlap, e.g., $s_{12} + s_{123}$, corresponds to different individual sizes of s_{12} and s_{123} that add to the same values. Recall that only the sum contributes to the entropy computation.

(a)

(b)

(c)

(d)

$$\min_{\tau_1,\tau_2,\tau_3} h\left(\vec{X}_T \mid O_1^{(\tau_1)}, O_2^{(\tau_2)}, O_3^{(\tau_3)}\right)$$

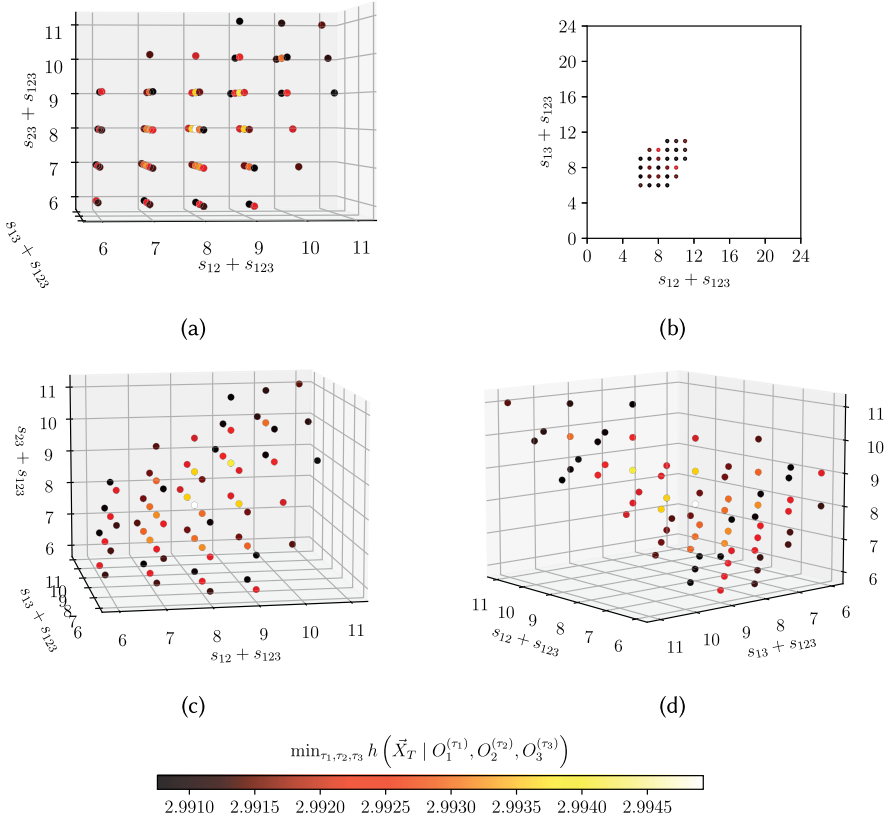2.9910   2.9915   2.9920   2.9925   2.9930   2.9935   2.9940   2.9945

Fig. 12. Configurations and values of minimal information disclosure as functions of the pairwise spectator overlaps for three evaluations.

The maximum conditional entropy (singular white point) occurs when the pairwise overlaps are 1/3 of n, i.e., when $s_{13} + s_{123} = s_{12} + s_{123} = s_{23} + s_{123} = 8$. Other top configurations are located nearby but do not deviate from the center evenly. In the projection of two of the three pairwise overlap dimensions ($s_{12} + s_{123}$ vs. $s_{13} + s_{123}$, Figure 12(b)), the top-500 configurations are concentrated in the 1/3 overlap region. The shape is preserved (and thus the figures are identical) in the other two projections. It is important to point out that the difference in entropy between the largest and smallest value plotted is less than 1/100th of a bit.

Having examined optimal configurations for two and three executions, we want to generalize the findings to any number of experiments and spectators n. In Figure 13, we plot the optimal pairwise overlap percentages as a function of n for 2, 3, and 4 executions. Information leakage is always the smallest when all pairwise overlaps are equal (i.e., for $M = 3$, $s_{13} + s_{123} = s_{12} + s_{123} = s_{23} + s_{123}$). The optimal overlap percentage for $M = 2$ is upper bounded by 50% and tends towards 50% as n grows. Interestingly, the optimal overlap for both $M = 3$ and $M = 4$ trend toward 1/3 overlap, while ideal overlaps are generally smaller for $M = 4$. Analysis of large M, while potentially interesting, is of limited practical value.

7 Conclusions and Recommendations

In this work, we study information disclosure associated with revealing the output of average salary computation on private inputs. Using the framework of Reference [1], we analyze the
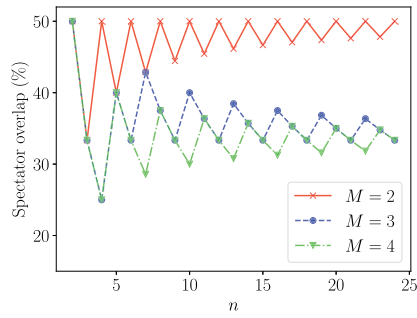
Fig. 13. The optimal shared spectators overlap relative to the total number of participants n for M evaluations.

function and derive several information-theoretic properties associated with the computation. Inputs are modeled using several discrete and continuous distributions, leading to multiple interesting conclusions about their entropy loss. We expand the scope to multiple executions on related inputs and determine the best configurations that minimize information disclosure. This leads to the following recommendations for computation designers:

— The amount of information disclosure about a target is independent of adversarial inputs. It was also experimentally shown to be independent of distribution parameters for three different distributions and analytically shown for normal distribution. All examined distributions produce nearly identical entropy loss curves.

— One can reduce the amount of entropy loss to a desired level by increasing the number of participants. For example, the computation designer can advertise at most 5% or 1% maximum entropy loss for the average salary application, which will require recruiting 6 or 25, respectively, non-adversarial participants when running only a single evaluation.

— In the presence of repeated computations, information disclosure about inputs continues for both participants who stay and participants who leave. With two executions, protection is the largest with 50% overlap in the participants, while both a small overlap and an overwhelming overlap result in undesirable information disclosure about different types of participants (i.e., those who stay vs. those who leave).

— With more executions, pairwise overlap sizes determine information disclosure. For three and four executions, optimal configurations have overlap sizes near 1/3 of the number of participants.

— Information disclosure about participants' inputs can still be kept at a desirable level by enrolling enough participants and restricting the percentage of reused inputs to be in a desired range. For example, with two executions and following the guidelines of keeping the overlap near 50%, the number of non-adversarial input contributors needs to be at least 8 to meet the target of 5% information loss.

This work provides the foundation for a number of interesting future directions, the first of which is expanding our analysis to more advanced descriptive statistics. This encompasses computations such as the max/min, median, and standard deviation. For these functions, it may not be possible to derive closed-form expressions for the information disclosure, and we would consequently pivot to leveraging simulation-based techniques to compute the entropy. Entropy estimators (such as those designed in References [32, 33]) are a highly accurate means of calculating the entropy of an unknown distribution, given a set of samples. These techniques provide a wide range of opportunities to measure the disclosure beyond statistical functions.

Acknowledgments

References

[1] P. Ah-Fat and M. Huth. 2017. Secure multi-party computation: Information flow of outputs and game theory. In *Principles of Security and Trust*. Springer, 71–92.

[2] P. Ah-Fat and M. Huth. 2019. Optimal accuracy-privacy trade-off for secure computations. *IEEE Trans. Inf. Theor.* 65, 5 (2019), 3165–3182.

[3] P. Ah-Fat and M. Huth. 2020. Protecting private inputs: Bounded distortion guarantees with randomised approximations. *Proc. Privac. Enhanc. Technol.* 2020, 3 (2020), 284–303.

[4] P. Ah-Fat and M. Huth. 2020. Two and three-party digital goods auctions: Scalable privacy analysis. arXiv preprint arXiv:2009.09524.

[5] M. Alvim, K. Chatzikokolakis, A. McIver, C. Morgan, C. Palamidessi, and G. Smith. 2014. Additive and multiplicative notions of leakage, and their capacities. In *IEEE Computer Security Foundations Symposium*. 308–322.

[6] M. Alvim, K. Chatzikokolakis, C. Palamidessi, and G. Smith. 2012. Measuring information leakage using generalized gain functions. In *IEEE Computer Security Foundations Symposium*. 265–279.

[7] M. Alvim, A. Scedrov, and F. Schneider. 2014. When not all bits are equal: Worth-based information flow. In *Principles of Security and Trust*. Springer, 120–139.

[8] A. Baccarini, M. Blanton, and S. Zou. 2024. Understanding information disclosure from secure computation output: A study of average salary computation. In *ACM Conference on Data and Application Security and Privacy (CODASPY'24)*. 187–198.

[9] R. Barakat. 1976. Sums of independent lognormally distributed random variables. *J. Optic. Societ. Amer.* 66, 3 (1976), 211–216.

[10] N. Beaulieu, A. Abu-Dayya, and P. McLane. 1995. Estimating the distribution of a sum of independent lognormal random variables. *IEEE Trans. Commun.* 43, 12 (1995), 2869–2873.

[11] N. Beaulieu and Q. Xie. 2004. An optimal lognormal approximation to lognormal sum distributions. *IEEE Trans. Vehic. Technol.* 53, 2 (2004), 479–489.

[12] A. Bhowmick, D. Boneh, S. Myers, and K. Tarbe. 2021. The Apple PSI system. Retrieved from https://www.apple.com/child-safety/pdf/Apple_PSI_System_Security_Protocol_and_Analysis.pdf

[13] Boston Women's Workforce Council (BWWC). 2017. 2016 Report. Retrieved from https://htv-prod-media.s3.amazonaws.com/files/bwwc-report-final-january-4-2017-1483635889.pdf

[14] Boston Women's Workforce Council (BWWC). 2018. 2017 Report. Retrieved from https://www.boston.gov/sites/default/files/document-file-01-2018/bwwc_2017_report.pdf

[15] S. Bu, L. Lakshmanan, R. Ng, and G. Ramesh. 2006. Preservation of patterns and input-output privacy. In *IEEE International Conference on Data Engineering*. 696–705.

[16] C. Caiado and P. Rathie. 2007. Polynomial coefficients and distribution of the sum of discrete uniform variables. In *International Conference on Special Functions Applications (SSFA'07)*.

[17] L. Cao, T. Tong, D. Trafimow, T. Wang, and X. Chen. 2022. The a priori procedure for estimating the mean in both log-normal and gamma populations and robustness for assumption violations. *Methodology* 18, 1 (2022), 24–43.

[18] M. Cheraghchi. 2019. Expressions for the entropy of basic discrete distributions. *IEEE Trans. Inf. Theor.* 65, 7 (2019), 3999–4009.

[19] D. Clark, S. Hunt, and P. Malacaria. 2002. Quantitative analysis of the leakage of confidential data. *Electron. Notes Theoret. Comput. Sci.* 59, 3 (2002), 238–251.

[20] F. Clementi and M. Gallegati. 2005. Pareto's law of income distribution: Evidence for Germany, the United Kingdom, and the United States. In *Econophysics of Wealth Distributions*. Springer, 3–14. DOI : https://doi.org/10.1007/88-470-0389-X_1

[21] B. Cobb, R. Rumí, and A. Salmerón. 2012. Approximating the distribution of a sum of log-normal random variables. *Stat. Comput.* 16, 3 (2012), 293–308.

[22] T. Cover and J. Thomas. 2006. *Elements of Information Theory*. Wiley-Interscience.

[23] D. Denning. 1982. *Cryptography and Data Security*. Addison-Wesley, Reading, MA.

[24] V. Deshpande, L. Schwarz, M. Atallah, M. Blanton, and K. Frikken. 2011. Outsourcing manufacturing: Secure price-masking mechanisms for purchasing component parts. *Product. Oper. Manag.* 20, 2 (2011), 165–180.

[25] V. Deshpande, L. Schwarz, M. Atallah, M. Blanton, K. Frikken, and J. Li. 2005. *Secure Collaborative Planning, Forecasting and Replenishment (SCPFR)*. Technical Report. CERIAS Tech Report 2006-65.

[26] V. Deshpande, L. Schwarz, M. Atallah, M. Blanton, K. Frikken, and J. Li. 2006. Secure collaborative planning, forecasting and replenishment (SCPFR). In *Multi-Echelon/Public Applications of Supply Chain Management Conference*. 165–180.

[27] C. Dwork. 2008. Differential privacy: A survey of results. In *International Conference on Theory and Applications of Models of Computation*. 1–19.

[28] C. Dwork, F. McSherry, K. Nissim, and A. Smith. 2006. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography Conference*. Springer, 265–284.

[29] C. Dwork and A. Roth. 2014. The algorithmic foundations of differential privacy. *Found. Trends Theoret. Comput. Sci.* 9, 3–4 (2014), 211–407.

[30] R. Evans and J. Boersma. 1988. The entropy of a Poisson distribution (C. Robert Appledorn). *SIAM Rev.* 30, 2 (1988), 314–317. DOI : https://doi.org/10.1137/1030059

[31] L. Fenton. 1960. The sum of log-normal probability distributions in scatter transmission systems. *IRE Trans. Commun. Syst.* 8, 1 (1960), 57–67.

[32] W. Gao, S. Kannan, S. Oh, and P. Viswanath. 2017. Estimating mutual information for discrete-continuous mixtures. In *Advances in Neural Information Processing Systems*, Vol. 30. Curran Associates Inc.

[33] W. Gao, S. Oh, and P. Viswanath. 2018. Demystifying fixed k-nearest neighbor information estimators. *IEEE Trans. Inf. Theor.* 64, 8 (2018), 5629–5661.

[34] A. Ghosh, T. Roughgarden, and M. Sundararajan. 2009. Universally utility-maximizing privacy mechanisms. In *ACM Symposium on Theory of Computing*. 351–360.

[35] J. Hsu, M. Gaboardi, A. Haeberlen, S. Khanna, A. Narayan, B. C. Pierce, and A. Roth. 2014. Differential privacy: An economic method for choosing epsilon. In *IEEE Computer Security Foundations Symposium*. IEEE, 398–410.

[36] Inpher. 2024. Inpher. Retrieved from https://inpher.io/

[37] M. Ion, B. Kreuter, A. Nergiz, S. Patel, S. Saxena, K. Seth, M. Raykova, D. Shanahan, and M. Yung. 2020. On deploying secure computing: Private intersection-sum-with-cardinality. In *IEEE European Symposium on Security and Privacy (EuroS&P'20)*. 370–389.

[38] M. Iwamoto and J. Shikata. 2013. Information theoretic security for encryption based on conditional Rényi entropies. In *International Conference on Information Theoretic Security*. 103–121.

[39] B. Köpf and D. Basin. 2011. Automatically deriving information-theoretic bounds for adaptive side-channel attacks. *J. Comput. Secur.* 19, 1 (2011), 1–31.

[40] R. Kotecha and S. Garg. 2017. Preserving output-privacy in data stream classification. *Prog. Artif. Intell.* 6 (2017), 87–104.

[41] B. Kreuter. 2017. Secure Multiparty Computation at Google. Real World Crypto. Retrieved from https://www.youtube.com/watch?v=ee7oRsDnNNc

[42] A. Lapets, F. Jansen, K. Albab, R. Issa, L. Qin, M. Varia, and A. Bestavros. 2018. Accessible privacy-preserving web-based data analysis for assessing and addressing economic inequalities. In *ACM COMPASS Conference*. 48:1–48:5.

[43] A. Lapets, N. Volgushev, A. Bestavros, F. Jansen, and M. Varia. 2016. Secure MPC for analytics as a web application. In *IEEE Secure Development Conference (SecDev'16)*. 73–74.

[44] A. Lapets, N. Volgushev, A. Bestavros, F. Jansen, and M. Varia. 2016. *Secure Multi-party Computation for Analytics Deployed as a Lightweight Web Application*. Technical Report BUCS-TR-2016-008. Boston University.

[45] Ligero Inc. 2022. Secure and Private Collaboration for Blockchains and Beyond. Retrieved from https://ligero-inc.com/

[46] F. Liu. 2018. Generalized Gaussian mechanism for differential privacy. *IEEE Trans. Knowl. Data Eng.* 31, 4 (2018), 747–756.

[47] P. Mardziel, M. Hicks, J. Katz, and M. Srivatsa. 2012. Knowledge-oriented secure multiparty computation. In *Workshop on Programming Languages and Analysis for Security*. 1–12.

[48] J. Massey. 1994. Guessing and entropy. In *IEEE International Symposium on Information Theory*. 204.

[49] F. McSherry and K. Talwar. 2007. Mechanism design via differential privacy. In *IEEE Symposium on Foundations of Computer Science*. 94–103.

[50] R. Mendes and J. Vilela. 2017. Privacy-preserving data mining: Methods, metrics, and applications. *IEEE Access* 5 (2017), 10562–10582.

[51] A. Monreale and W. Wang. 2016. Privacy-preserving outsourcing of data mining. In *IEEE Annual International Computer Software and Applications Conference (COMPSAC'16)*, Vol. 2. 583–588.

[52] J. P. Near, D. Darais, N. Lefkovitz, and G. Howarth. 2023. *Guidelines for Evaluating Differential Privacy Guarantees*. Technical Report BUCS-TR-2016-008. NIST.

[53] Nth party. 2024. Nth party. Retrieved from https://www.nthparty.com/

[54] Partisia. 2024. Partisia. Retrieved from https://partisia.com/

[55] A. Rastogi, P. Mardziel, M. Hicks, and M. Hammer. 2013. Knowledge inference for optimizing secure multi-party computation. In *SIGPLAN Workshop on Programming Languages and Analysis for Security*. 3–14.

[56] S. Schwartz and Y. Yeh. 1982. On the distribution function and moments of power sums with log-normal components. *Bell Syst. Technic. J.* 61, 7 (1982), 1441–1462.

[57] D. Senaratne and C. Tellambura. 2009. Numerical computation of the lognormal sum distribution. In *IEEE Global Communications Conference (GLOBECOM'09)*. 1–6.

[58] R. Shokri, M. Stronati, C. Song, and V. Shmatikov. 2017. Membership inference attacks against machine learning models. In *IEEE Symposium on Security and Privacy (S&P'17)*. 3–18.

[59] M. Skórski. 2019. Strong chain rules for min-entropy under few bits spoiled. In *IEEE International Symposium on Information Theory*. 1122–1126.

[60] G. Smith. 2009. On the foundations of quantitative information flow. In *International Conference on Foundations of Software Science and Computation Structures (FoSSaCS'09)*. 288–302.

[61] L. Song and P. Mittal. 2021. Systematic evaluation of privacy risks of machine learning models. In *USENIX Security Symposium*. 2615–2632.

[62] W. Souma. 2002. Physics of personal income. In *Empirical Science of Financial Fluctuations*. Springer, 343–352.

[63] A. Walker, S. Patel, and M. Yung. 2019. Helping organizations do more without collecting more data. *Google Security Blog* (June 2019). Retrieved from https://security.googleblog.com/2019/06/helping-organizations-do-more-without-collecting-more-data.html

[64] T. Wang and L. Liu. 2011. Output privacy in data mining. *ACM Trans. Datab. Syst.* 36, 1 (2011), 1–34.

[65] J. Wu, N. Mehta, and J. Zhang. 2005. Flexible lognormal sum approximation method. In *IEEE Global Communications Conference (GLOBECOM'05)*. 3413–3417.