

Composable Bounds on Information Flow from Distribution Differences

Megumi Ando^(✉) and Joshua D. Guttman

The MITRE Corporation, Bedford, USA
{mando,guttman}@mitre.org

Abstract. We define information leakage in terms of a “difference” between the *a priori* distribution over some remote behavior and the *a posteriori* distribution of the remote behavior conditioned on a local observation from a protocol run. Either a maximum or an average may be used. We identify a set of notions of “difference;” we show that they reduce our general leakage notion to various definitions in the literature. We also prove general composability theorems analogous to the data-processing inequality for mutual information, or cascading channels for channel capacities.

Keywords: Information flow · Non-disclosure · Limited disclosure · Information theory · Mutual information · Min-entropy leakage · Composability · Channel capacity

1 Introduction

One of us recently [11] introduced the *Frame Model* for studying information disclosure in distributed systems. Frames represent systems by directed graphs; their arcs represent the channels of communication permitted by the system. Disclosure occurs when the local behavior of one portion of the system (the “source”) affects what local behaviors may be observed at another portion of the system. That paper shows that limitations on disclosure respect a *cut principle*: Each limit on the disclosure from a source to a cut set of channels in the graph is also enforced on disclosure from the source to any more distant set of channels. This result furnishes a kind of compositionality for limited disclosure. However, the notion of limited disclosure in [11] was “possibilistic,” i.e. non-quantitative.

The purpose of this paper is to take key steps toward adapting its results to a quantitative treatment of disclosure. Given a probability distribution over the local behaviors of the system, we can generalize the cut principle to include a probabilistic analysis for quantifying leakage. As in the motivating example below, a quantitative analysis may capture insecurities that a possibilistic approach may overlook.

To focus our work, however, we have decided to omit one aspect of this problem. Namely, the frame model allows non-determinism. Generally, to obtain

Copyright © 2015 The MITRE Corporation. All rights reserved.

probability distributions on runs of a non-deterministic system, one must introduce a “scheduler” that chooses what events happen when different alternatives exist. Specifying these schedulers is subtle, essentially because a scheduler sensitive to the system’s secrets can signal them to the observer through its choices. In this paper, we ignore the resolution of non-determinism. Our analysis here applies in any case in which a probability distribution on executions is well-defined. In future work we will define methods for resolving non-determinism without giving the scheduler unfair ways to signal secrets.

A Motivating Example. David Chaum first introduced the Dining Cryptographers’ Protocol (DCP) as a means to study secure multi-party boolean-OR computation [5]. Chaum describes a scenario where a group of three cryptographers are at dinner, and either the Spymaster (their boss) or one of the cryptographers at the table pays for the meal. The protocol guarantees that each party can determine whether the Spymaster or one of the cryptographers at the table paid; and in the latter case, the identity of the payer remains hidden from the non-payers.

Let A , B , and C denote the three cryptographers. Without loss of generality, let us assume that A is a non-paying cryptographer, and consider her viewpoint. A flips a coin with B to get r_{AB} and flips another coin with C obtaining r_{AC} . She computes $m_A = r_{AB} \oplus r_{AC}$, and announces m_A to the table. (As the payer, she would have announced $m_A \oplus 1$.) From B and C ’s announcements, she surmises the overall parity $m = m_A \oplus m_B \oplus m_C$, from which she can determine whether the boss paid.

If m is odd, A learns that one of the other cryptographers paid but cannot know for sure which. Possibilistically, we say that the identity of the payer remains undisclosed to A . Further, *no* information is disclosed since the set of possibilities remains the same before and after a protocol run.

Despite provable non-disclosure, information (quantified in a certain way) can still leak if the coins are biased: Suppose that the payer is chosen from a fixed distribution. Conditioned on either B or C paying, the payer identity is a Bernoulli random variable $X_A \sim \text{Bern}(p)$ with probability p that player B is the payer, and probability $(1 - p)$ that player C is the payer. Suppose further that the coin flips are independent and identically distributed Bernoulli random variables: $R_{AB}, R_{BC}, R_{CA} \sim \text{Bern}(q)$.

The set of A ’s sent and received messages is another (multi-dimensional) random variable: $O_A = R_{AB}, R_{CA}, M_A, M_B, M_C$, where M_A , M_B , and M_C denote the cryptographers’ respective m -messages. Rather than merely confirming that the set of possible payers remains the same, we can compare the distributions of X_A pre- and post- protocol run. One way to do this is by computing the difference between the entropies of the a priori and a posteriori distributions:

$$\mathcal{I}(X_A; O_A) = \mathcal{H}(X_A) - \mathcal{H}(X_A | O_A), \quad (1)$$

where $\mathcal{I}(\cdot; \cdot)$, $\mathcal{H}(\cdot)$, and $\mathcal{H}(\cdot | \cdot)$ denote mutual information, entropy, and conditional entropy, resp. (All are formally defined in Sect. 2.1).

In the special case when there is an equal chance of B or C paying ($p = 0.5$), and the coin flips are fair ($q = 0.5$); the unconditional and conditional distributions O_A and $O_A|X_A$ are uniform, and there is no leakage. However, this is not generally true as shown by Chatzikokolakis *et al.* [4]. This example illustrates that information may leak even in scenarios where there is provably no disclosure. In this paper, we show how to generalize the results of [11] to include quantitative analyses such as the type presented above.

Other Related Work. In contrast to the possibilistic approach in [11], there are a number of well-cited papers that use information theoretic definitions for quantifying anonymity, information flow, or non-interference in distributed systems: Díaz *et al.* [9] and Serjantov and Danezis [9] use Shannon entropy; Clarkson *et al.* [7] and Deng *et al.* [8], relative entropy; Köpf and Basin [12], guessing entropy; Chatzikokolakis, Malacaria, Zhu, and others [4, 6, 13, 15], (conditional) mutual information or channel capacity; and Palamidessi, Smith, and others [1–3, 10, 14], min-entropy leakage. This list is not exhaustive.

Of these information theoretic concepts, we show that our leakage notion is reducible to mutual information and channel capacity [4, 6, 13, 15]. As seen in our motivating example, mutual information is a measure of reductions in uncertainty, where uncertainty is defined as the entropy of a distribution. For a specified a priori distribution, there is no leakage provided that the mutual information between X and O is zero. This idea is generalized by allowing for some intentionally revealed information (represented as a reveal random variable), such that security is achieved with zero *conditional* mutual information or capacity. This is the approach taken in by Chatzikokolakis *et al.* [4] and Clark *et al.* [6] and summarized in Sect. 2.1.

Reduction in uncertainty can also be measured by min-entropy leakage, which is defined as the difference between the min-entropy of the a priori distribution and the conditional min-entropy of the a posteriori distribution. Currently, there is no consensus on how conditional min-entropy should be defined. Indeed, Cachin [3] defines the conditional min-entropy $\mathcal{H}_\infty(X|Y)$ of $X|Y$ as

$$\mathcal{H}_\infty(X|Y) = - \sum_{y \in \mathcal{Y}} \mathbf{P}(Y = y) \cdot \log \max_{x \in \mathcal{X}} \mathbf{P}(X = x|Y = y). \quad (2)$$

whereas Palamidessi, Smith, and others [1, 2, 10, 14] define it with the logarithm and summation reversed. In Sect. 2.1, we provide a summary of min-entropy leakage as defined in [3], and we show how min-entropy leakage derived from this former conditional min-entropy also relates to our notion of leakage in Sect. 4.

Our Contributions. As in [11], we describe whether and (how much) information can leak from one portion of a distributed system to another. We also identify scenarios where the leakage provides an upperbound on information flow to more remote portions of the network. In these cases, compositions of local leakages bounds are meaningful globally.

In addition to providing a generalization of the cut-blur principle in [11], the contributions of this paper are the following:

- We define information flow in a distributed system very generally: Leakage is defined as the max (for worst-case) or average (for average-case) “difference” between the a priori distribution over some remote behavior and the a posteriori distribution of the remote behavior conditioned on a possible local observation from a protocol run.
- We identify a set of distribution differences that relate this unified notion of leakage to accepted definitions in the literature: namely mutual information, min-entropy leakage, and limited or non-disclosure.
- We also prove equivalence and implication relations between different leakage definitions. For zero leakage, we prove that zero mutual information provides the strongest security and implies zero leakage under all distribution differences satisfying the coincidence axiom.
- We identify a sufficient property (convexity) of distribution differences for the composability of leakage bounds analogous to one of the bounds in the data-processing inequality for mutual information, or cascading channels for channel capacities: Given a Markov chain $X \rightarrow Y \rightarrow Z$, the leakage from X to Z is bounded by the leakage from X to Y . If the leakage under the distribution distance is additionally symmetric, then we get the other bound: The leakage from X to Z is also bounded by the leakage from Y to Z . The composability property can also be seen as a generalization of the cut-blur principle for limited disclosure.

Road Map of Paper. Leakage definitions using mutual information, min-entropy leakage, and limited disclosure are described in Sect. 2. In Sect. 2.2, we provide informal descriptions of limited disclosure and the cut-blur principle (the main composability result of [11]). In Sect. 2.3, we formally define distribution differences, which are used in our leakage definitions in Sect. 4. Sections 3–5 contain our problem statement, leakage definitions, and results. We conclude with extensions to our results in Sect. 6.

2 Preliminaries

2.1 Mutual Information, Capacity, and Min-Leakage

Chatzikokolakis *et al.* [4] use conditional channel capacity for quantifying information leakage in anonymity protocols given intentionally revealed information. Below is their leakage definition, preceded by some information theoretic definitions.

Definition 1. Let $X : \mathcal{X} \rightarrow [0, 1]$, $Y : \mathcal{Y} \rightarrow [0, 1]$, and $Z : \mathcal{Z} \rightarrow [0, 1]$ be discrete random variables.

1. The entropy of X , denoted $\mathcal{H}(X)$, is given by

$$\mathcal{H}(X) = - \sum_{x \in \mathcal{X}} \mathbf{P}(X = x) \cdot \log \mathbf{P}(X = x),$$

where $\mathbf{P}(\cdot)$ denotes probability.

2. The conditional entropy of $X|Y$, denoted $\mathcal{H}(X|Y)$, is given by

$$\mathcal{H}(X|Y) = \sum_{y \in \mathcal{Y}} \mathbb{P}(Y = y) \cdot \mathcal{H}(X|Y = y)$$

3. The mutual information between X and Y , denoted $\mathcal{I}(X; Y)$, is given by

$$\mathcal{I}(X; Y) = \mathcal{H}(X) - \mathcal{H}(X|Y)$$

4. The mutual information $\mathcal{I}(X; Y|Z)$ between X and Y , conditioned on Z , is given by

$$\mathcal{I}(X; Y|Z) = \mathcal{H}(X|Z) - \mathcal{H}(X|Y, Z) = \sum_{z \in \mathcal{Z}} \mathbb{P}(Z = z) \cdot \mathcal{I}(X; Y|Z = z)$$

In [4], an anonymity protocol is modeled by a conditional distribution $p_{O|X}(\cdot|\cdot)$ over the space $\mathcal{O} \times \mathcal{X}$, where \mathcal{X} and \mathcal{O} are the domains of a secret random variable $X : \mathcal{X} \rightarrow [0, 1]$ and an observable random variable $O : \mathcal{O} \rightarrow [0, 1]$, resp. Every run of the protocol produces an independent observable sampled from this conditional distribution. Anonymity is achieved with zero capacity.

If the protocol intentionally reveals some information R , represented as a random variable, then the protocol is secure if it achieves *relative anonymity*, defined below.

Definition 2 (Informal). Given an anonymity protocol $p_{O|X}(\cdot|\cdot)$, we say it achieves relative anonymity if

$$\max_{p_X(\cdot)} \mathcal{I}(X; O|R) = 0, \quad (3)$$

where the maximization is over all input distributions $p_X(\cdot)$ on \mathcal{X} . In other words, $\mathcal{I}(X; O|R) = 0$ for all possible $p_X(\cdot)$.

See [4] for a formal treatment. Min-entropy leakage in [3] is defined analogously using min-entropy and conditional min-entropy:

Definition 3. Let $X : \mathcal{X} \rightarrow [0, 1]$, $Y : \mathcal{Y} \rightarrow [0, 1]$, and $Z : \mathcal{Z} \rightarrow [0, 1]$ be discrete random variables.

1. The min-entropy of X , denoted $\mathcal{H}_\infty(X)$, is given by

$$\mathcal{H}_\infty(X) = -\log \max_{x \in \mathcal{X}} \mathbb{P}(X = x) \quad (4)$$

2. The conditional min-entropy of $X|Y$, denoted $\mathcal{H}_\infty(X|Y)$, is given by

$$\mathcal{H}_\infty(X|Y) = \sum_{y \in \mathcal{Y}} \mathbb{P}(Y = y) \cdot \mathcal{H}_\infty(X|Y = y) \quad (5)$$

3. The min-entropy leakage from X to Y , denoted $\mathcal{M}(X; Y)$, is given by

$$\mathcal{M}(X; Y) = \mathcal{H}_\infty(X) - \mathcal{H}_\infty(X|Y), \quad (6)$$

where $\mathcal{H}_\infty(X|Y)$ is as defined in Definition 3.¹

4. The min-entropy leakage $\mathcal{M}(X; Y|Z)$ between X and Y , conditioned on Z , is given by

$$\mathcal{M}(X; Y|Z) = \mathcal{H}_\infty(X|Z) - \mathcal{H}_\infty(X|Y, Z) \quad (7)$$

2.2 Limited Disclosure in the Frame Model

The Frame Model was introduced in [11] as a means for studying composable information disclosure. Communication is modeled as point-to-point, and messages are delivered synchronously. Partial ordering on the message deliveries models true concurrency within a protocol run.

A frame $\mathcal{F} = (\mathcal{LO}, \mathcal{CH}, \mathcal{D}, \text{ends}, \text{traces})$ consists of a set of locations \mathcal{LO} , a set of channels \mathcal{CH} , a set of data \mathcal{D} , and methods $\text{ends}(\cdot)$ and $\text{traces}(\cdot)$ defined on $\ell \in \mathcal{LO}$. Graphically, a frame \mathcal{F} can be represented as a directed graph, where the nodes are the locations, and the (directed) edges are the channels. Each edge is labeled. The label represents the data that can be transmitted along that edge from the exit node to the entry node.

A channel endpoint is either an entry or an exit point of a channel; so $\text{ends}(\ell)$ returns the set of all endpoints that either enter into or exit from ℓ . A trace is an ordered sequence of local *events* that represents a location's interactions with the other locations; where calling the $\text{chan}(\cdot)$ method on an event object returns a channel, and calling the $\text{data}(\cdot)$ method returns a data. So $\text{traces}(\ell)$ returns all possible local sequences representing all the ways in which ℓ might participate in a (potentially incomplete) run of the protocol.

The authors of [11] provide a mathematical notion of an *execution* (a run of a protocol) within the Frame Model and define the portion of an execution relevant to a set C of channels as a C -run.

Definition 4 (Informal). A function $\text{blur} : \mathcal{P}(\mathcal{S}) \rightarrow \mathcal{P}(\mathcal{S})$ is a blur operator if it satisfies the properties:

1. Inclusion: $\mathcal{T} \subseteq \text{blur}(\mathcal{T})$
2. Idempotence: $\text{blur}(\text{blur}(\mathcal{T})) = \text{blur}(\mathcal{T})$
3. Union property: $\forall \Sigma \subseteq \mathcal{P}(\mathcal{S}). \text{blur}(\bigcup_{\mathcal{T} \in \Sigma} \mathcal{T}) = \bigcup_{\mathcal{T} \in \Sigma} \text{blur}(\mathcal{T})$, where $\mathcal{P}(\cdot)$ denotes the powerset.

Given a frame $\mathcal{F} = (\mathcal{LO}, \mathcal{CH}, \mathcal{D}, \text{ends}, \text{traces})$, a set $\text{src} \subseteq \mathcal{CH}$ of source channels, and a set $\text{obs} \subseteq \mathcal{CH}$ of observable channels; let \mathcal{S} be the set of source-runs (i.e., $\mathcal{S} = \text{src-runs}$), and let \mathcal{O} be the set of observable-runs (i.e., $\mathcal{O} = \text{obs-runs}$).

Information disclosure is restricted by a blur operator $\text{blur}(\cdot)$ if, for every observable $o \in \mathcal{O}$, the set $\mathcal{T} \subseteq \mathcal{S}$ of completed source-runs compatible with the observable o is blur-blurred , i.e., $\mathcal{T} = \text{blur}(\mathcal{T})$, where $\text{blur}(\cdot)$ is a blur operator.

¹ There is an alternative definition for conditional min-entropy [1, 2, 10, 14]. We will not be dealing with this alternative definition here.

The main result of the paper is the so-called cut-blur principle below. See [11] for a formal treatment.

Theorem 1 (Cut-blur Principle, Informal). *Given a frame*

$$\mathcal{F} = (\mathcal{LO}, \mathcal{CH}, \mathcal{D}, \text{ends}, \text{traces}), \quad (8)$$

a set $\text{src} \subseteq \mathcal{CH}$ of source channels, a set $\text{cut} \subseteq \mathcal{CH}$ of cut channels, and a set $\text{obs} \subseteq \mathcal{CH}$ of observable channels, such that cut is a cut-set partitioning src from obs , the source information disclosed at cut limits the source information disclosed at obs .

2.3 Distribution Differences

In our motivating example, we compared the a priori and a posteriori distributions in order to quantify how much information leaked. Intuitively, if the distributions are the same under some specified way of measuring, there is no measurable leakage. Stated as such, leakage is expressed in terms of “distribution differences,” which we define formally below.

Definitions 6–8 are distribution differences, which we use later on to relate our unified leakage notion in Sect. 4 to accepted leakage definitions in the literature, namely: conditional mutual information, min-entropy leakage, and limited information disclosure from [3, 4, 11]. These leakage definitions are also summarized in Sects. 2.1–2.2.

Definition 5. *Let $\mathbb{X}_{\mathcal{X}}$ denote a family of random variables defined over the same alphabet \mathcal{X} . A distribution difference $\Delta : \mathbb{X}_{\mathcal{X}} \times \mathbb{X}_{\mathcal{X}} \rightarrow \mathbb{R}$ is a function that takes two random variables defined over the same alphabet and returns a real number. In discussing this definition, we are often interested in the following properties:*

1. Coincidence axiom: $\forall \mathcal{X}, \forall X \in \mathbb{X}_{\mathcal{X}}. \Delta(X, X) = 0$
2. Nonnegativity: $\forall \mathcal{X}, \forall X_1, X_2 \in \mathbb{X}_{\mathcal{X}}. \Delta(X_1, X_2) \geq 0$
3. Convexity: $\forall \mathcal{X}, \forall X, X_1, X_2 \in \mathbb{X}_{\mathcal{X}}, \forall \alpha \in [0, 1].$

$$\Delta(X, (\alpha X_1 + (1 - \alpha)X_2)) \leq \alpha \cdot \Delta(X, X_1) + (1 - \alpha) \cdot \Delta(X, X_2) \quad (9)$$

Definition 6. *For any random variables X_1 and X_2 over the same alphabet \mathcal{X} , we say that the Shannon-difference between X_1 and X_2 , denoted $\Delta_S(X_1, X_2)$, is given by*

$$\Delta_S(X_1, X_2) = \mathcal{H}(X_1) - \mathcal{H}(X_2). \quad (10)$$

Definition 7. *For any random variables X_1 and X_2 over the same alphabet \mathcal{X} , we say that the minH-difference between X_1 and X_2 , denoted $\Delta_{\min}(X_1, X_2)$, is given by*

$$\Delta_{\min}(X_1, X_2) = \mathcal{H}_{\infty}(X_1) - \mathcal{H}_{\infty}(X_2). \quad (11)$$

Definition 8. For any random variables X_1 and X_2 over the same alphabet \mathcal{X} , we say that the *maxH-difference* between X_1 and X_2 , denoted $\Delta_{\max}(X_1, X_2)$, is given by

$$\Delta_{\max}(X_1, X_2) = \begin{cases} 0 & \text{supp}(X_1) = \text{supp}(X_2) \\ \infty & \text{otherwise,} \end{cases} \quad (12)$$

where $\text{supp}(X)$ denotes the support of a random variable X .

Note that Shannon-, minH-, and maxH-differences all satisfy the coincidence axiom and convexity. MaxH-difference additionally satisfies nonnegativity.

3 Problem Statement

While [11] presents purely set theoretic ideas, we generalize the cut-blur results to include information theoretic analyses. To do this, we shift from a possibilistic view of local behaviors to a probabilistic perspective. To begin with, we consider local behaviors from only completed executions, where a completed execution is the partially ordered entire set of messages from a completed protocol run. Our results in Sects. 4 and 5 cover leakages from complete observations. This allows us to present our work using cleaner notation. Extension to leakages from partial observations is covered in Sect. 6.

Below, we borrow the formalism from the Frame Model [11] to make our problem statement explicit. Our problem statement is defined with respect to a frame $\mathcal{F} = (\mathcal{LO}, \mathcal{CH}, \mathcal{D}, \text{ends}, \text{traces})$, a fixed set $\text{src} \subseteq \mathcal{CH}$ of source channels, and a fixed set $\text{obs} \subseteq \mathcal{CH}$ of observable channels.

Definition 9. Given a frame $\mathcal{F} = (\mathcal{LO}, \mathcal{CH}, \mathcal{D}, \text{ends}, \text{traces})$ and a location $\ell \in \mathcal{LO}$, let $\text{chans}(\ell)$ be the set of channels adjacent to ℓ :

$$\text{chans}(\ell) = \{c \in \mathcal{CH} : \text{entry}(c) \in \text{ends}(\ell) \vee \text{exit}(c) \in \text{ends}(\ell)\}, \quad (13)$$

where $\text{entry}(\cdot)$ and $\text{exit}(\cdot)$ return the entry and exit points of a channel, resp.

Definition 10. Given a frame $\mathcal{F} = (\mathcal{LO}, \mathcal{CH}, \mathcal{D}, \text{ends}, \text{traces})$ and a location $\ell \in \mathcal{LO}$, let $T(\ell) \subseteq \text{traces}(\ell)$ be defined by

$$T(\ell) = \{tr \in \text{traces}(\ell) : \exists tr' \in \text{traces}(\ell), \text{ } tr \text{ is a proper prefix of } tr'\} \quad (14)$$

($T(\ell)$ is the set of traces of ℓ that are proper prefixes of other traces of ℓ .) Let $\text{traces}^*(\ell) = \text{traces}(\ell) \setminus T(\ell)$, and call it the completed traces of ℓ .

Definition 11. An event set $\mathcal{E} = (E, \preceq)$ is a well-founded, partially ordered set E of events and is generally denoted by the name of the set and in curly-font.

Definition 12. Given an event set $\mathcal{E} = (E, \preceq)$ and a set C of channels, the restriction $\mathcal{E} \upharpoonright C$ of \mathcal{E} to C is the event set (E_c, \preceq_c) , where:

1. $E_c = \{e \in E : \text{chan}(e) \in C\}$, and
2. $\preceq_c = \preceq \cap E_c \times E_c$.

Definition 13. An event set $\mathcal{E} = (E, \preceq)$ is a completed execution in a frame $\mathcal{F} = (\mathcal{LO}, \mathcal{CH}, \mathcal{D}, \text{ends}, \text{traces})$, if for all $\ell \in \mathcal{LO}$:

1. $(\mathcal{E} \upharpoonright \text{chans}(\ell))$ is totally (linearly) ordered, and
2. $(\mathcal{E} \upharpoonright \text{chans}(\ell)) \in \text{traces}^*(\ell)$.

We call the set of all completed executions in a frame \mathcal{F} the completed execution set, denoted $\text{Exe}^*(\mathcal{F})$.

Definition 14. Given a frame $\mathcal{F} = (\mathcal{LO}, \mathcal{CH}, \mathcal{D}, \text{ends}, \text{traces})$ and a set of channels $C \subseteq \mathcal{CH}$, let $C\text{-runs}^*$ be the set of restrictions of completed executions to C :

$$C\text{-runs}^* = \{\mathcal{E} \upharpoonright C : \mathcal{E} \in \text{Exe}^*(\mathcal{F})\} \quad (15)$$

Let \mathcal{S} be the finite set of completed source-runs (i.e., $\mathcal{S} = \text{src-runs}^*$), and let \mathcal{O} be the finite set of completed observable-runs (i.e., $\mathcal{O} = \text{obs-runs}^*$).

We are also provided a partitioning function $f(\cdot, \cdot)$ which is a deterministic function over the joint space $\mathcal{S} \times \mathcal{O}$ and a joint probability mass function (pmf) $p_{\mathcal{S}, \mathcal{O}}(\cdot, \cdot)$ over $\mathcal{S} \times \mathcal{O}$, such that the supports for the corresponding marginal probabilities are \mathcal{S} and \mathcal{O} , resp. In other words, $p_{\mathcal{S}, \mathcal{O}}(\cdot, \cdot)$ written as a matrix has no all zero rows or columns. The partitioning function additionally has the property that for any $s_1, s_2 \in \mathcal{S}$ and $o \in \mathcal{O}$ such that $p_{\mathcal{S}, \mathcal{O}}(s_1, o), p_{\mathcal{S}, \mathcal{O}}(s_2, o) > 0$; $f(s_1, o) = f(s_2, o)$.

Definition 15. $S : \mathcal{S} \longrightarrow [0, 1]$ is the random variable on the completed source-runs, which maps to the marginal probabilities of the source-runs.

$$S(s) = \sum_{o \in \mathcal{O}} p_{\mathcal{S}, \mathcal{O}}(s, o). \quad (16)$$

S represents the a priori remote behavior.

Definition 16. $(S|O = o) : \mathcal{S} \longrightarrow [0, 1]$ is the conditional random variable on the completed source-runs, which maps to the probabilities of the source-runs conditioned on the observable-run $O = o$.

$$(S|O = o)(s) = \frac{p_{\mathcal{S}, \mathcal{O}}(s, o)}{\sum_{\omega \in \mathcal{S}} p_{\mathcal{S}, \mathcal{O}}(\omega, o)}. \quad (17)$$

$S|O$ represents the a posteriori remote behavior.

Definition 17. Let $f : \mathcal{S} \times \mathcal{O} \longrightarrow \mathcal{R}$ be any deterministic function on the joint space. $R_f : \mathcal{R} \longrightarrow [0, 1]$ is the random variable over the range of $f(\cdot, \cdot)$ whose probabilities are given by

$$R_f(r) = \sum_{(s, o) \in \mathcal{R}_r} p_{\mathcal{S}, \mathcal{O}}(s, o), \quad (18)$$

where $\mathcal{R}_r = \{(s, o) \in \mathcal{S} \times \mathcal{O} : f(s, o) = r\}$. Any intentionally revealed information may be represented by some R_f .

Given this set-up, we are interested in defining information flow in \mathcal{F} from source channels to observable channels. In this sense, this work is meant to generalize the main results of [11].

4 Leakage from Distribution Differences

Given a partitioning function $f(\cdot, \cdot)$ and a pmf $p_{S,O}(\cdot, \cdot)$ over the joint space $\mathcal{S} \times \mathcal{O}$; let S , O , and R_f representing remote behavior, local behavior, and intentionally revealed information be as defined in Sect. 3 above. Further, let \mathcal{V} denote the support of R_f , and for any $r \in \mathcal{V}$, let

$$\mathcal{O}_r = \{o \in \text{supp}(O) : \exists s \in \mathcal{S}, f(s, o) = r\}. \quad (19)$$

The definitions, theorems, and corollaries in Sect. 4 are with respect to this set-up. In Definitions 18 and 19, leakage is defined very generally as the max or average difference between the a priori and a posteriori distributions. These are definitions of leakage conditioned on some reveal random variable R_f . Note that unconditional leakage is captured by any all-to-one function $f(\cdot, \cdot)$.

Definition 18 (Worst-case Leakage). *The worst-case leakage $\mathcal{L}_{S;O|R_f}$ conditioned on R_f , is given by the maximum difference between the a priori distribution on $(S|R_f = r)$ and the a posteriori distribution $(S|R_f = r, O = o)$ over $r \in \mathcal{V}$ and $o \in \mathcal{O}_r$:*

$$\mathcal{L}_{S;O|R_f} = \max_{r \in \mathcal{V}} \max_{o \in \mathcal{O}_r} \Delta((S|R_f = r), (S|R_f = r, O = o)), \quad (20)$$

for some notion of distribution difference Δ . This is the worst-case leakage over all partitions. We say that there is zero conditional worst-case leakage when $\mathcal{L}_{S;O|R_f} = 0$.

Definition 19 (Average-case Leakage). *The average-case leakage $L_{S;O|R_f}$ conditioned on R_f is given by average difference between the a priori distribution on $(S|R_f = r)$ and the a posteriori distribution $(S|R_f = r, O = o)$ over $r \in \mathcal{V}$ and $o \in \mathcal{O}_r$:*

$$L_{S;O|R_f} = \sum_{r \in \mathcal{V}} \mathbb{P}(R_f = r) \cdot \sum_{o \in \mathcal{O}_r} \mathbb{P}(O = o | R_f = r) \cdot \Delta((S|R_f = r), (S|R_f = r, O = o)), \quad (21)$$

for some notion of distribution difference Δ . This is the average-case leakage over all partitions. We say that there is zero conditional average-case leakage when $L_{S;O|R_f} = 0$.

We chose to study nonstandard distribution differences instead of standard distribution distances, such as the Kullback-Leibler divergence (relative entropy) or statistical-closeness, because our general leakage definitions above reduce to accepted leakage notions in the literature under these distribution difference. Lemmas 1 and 2 and in Theorem 2 illustrate these equivalences.

Below, we show that average leakage under Shannon-difference is equivalent to mutual information between remote and locally observable behaviors. Likewise, average leakage under minH-difference is equivalent to min-entropy leakage.

Lemma 1. *Average-case conditional leakage under Shannon-difference is equivalent to conditional mutual information between S and O .*

Proof. Average leakage under Shannon-difference can be converted to conditional mutual information by pulling $\mathcal{H}(S|R_f = r)$ out from the summation and from the definitions of conditional entropy, mutual information, and conditional mutual information. \square

Lemma 2. *Average-case conditional leakage under minH-difference is equivalent to conditional min-entropy leakage from S to O .*

Proof. Same proof as above. \square

Zero leakage occurs when the a priori and a posteriori situations are equivalent under some specified distribution difference. Whereas unconditional zero leakage corresponds to no leakage in an absolute sense, conditional zero leakage corresponds to a somewhat weaker notion: Other than some intentionally revealed information, there is no leakage.

In Theorem 2 below, we prove that zero conditional worst-case leakage under maxH-difference is equivalent to limited disclosure, where the blur operator is related to the partitioning function. (Note that the equivalence is up to completed runs. See Sect. 6 for the extended results over partial observations.)

Theorem 2 (Non-disclosure over Blur-sets). *Zero conditional leakage under maxH-difference is equivalent to information disclosure restricted by a blur-operator $\text{blur}_{f,o} : \mathcal{P}(\mathcal{S}) \longrightarrow \mathcal{P}(\mathcal{S})$, given by*

$$\text{blur}_{f,o}(\mathcal{T}) = \bigcup_{t \in \mathcal{T}} \{s : f(s, o) = f(t, o) \wedge (S|R_f = f(t, o))(s) > 0\}. \quad (22)$$

Proof. Clearly, $\text{blur}_{f,o}(\cdot)$ is inclusive, idempotent, and satisfies the union property. So $\text{blur}_{f,o}$ is a blur-operator.

Let $\mathcal{T}_o = \text{supp}(S|O = o)$. For any fixed $o \in \mathcal{O}$, $\forall s \in \mathcal{S}$ where $p_{S,O}(s, o) > 0$, $f(s, o)$ maps to the same value, which we denote by r_o ; thus, $\mathcal{T}_o = \text{supp}(S|R_f = r_o, O = o)$.

(\implies) For any $\mathcal{T} \subseteq \mathcal{T}_o$, $\text{blur}_{f,o}(\mathcal{T}) = \text{supp}(S|R_f = r_o)$ by construction of the blur-operator; and $\text{supp}(S|R_f = r_o) = \text{supp}(S|R_f = r_o, O = o)$, by equality in the maxH-difference. So, $\text{blur}_{f,o}(\mathcal{T}) = \mathcal{T}_o$ as desired.

(\Leftarrow) Given any fixed $r \in \mathcal{V}$ where $R_f(r) > 0$ and any $o \in \mathcal{O}_r$, $\mathcal{T}_o = \text{supp}(S|R_f = r)$ by definition of blur-limited disclosure. So, $\text{supp}(S|R_f = r) = \text{supp}(S|R_f = r_o, O = o)$. In the case where $R_f(r) = 0$, this is vacuously true. \square

It can be shown that average zero leakage under Shannon-difference is equivalent to worst-case zero leakage under Shannon-difference.

Theorem 3. *Zero conditional worst-case leakage under Shannon-difference is equivalent to zero conditional mutual information between S and O :*

$$\max_{r \in \mathcal{V}, o \in \mathcal{O}_r} [\mathcal{H}(S|R_f = r) - \mathcal{H}(S|R_f = r, O = o)] = 0 \iff \mathcal{I}(S; O|R_f) = 0 \quad (23)$$

So from Lemma 1, both definitions are equivalent to zero mutual information between the remote and local behaviors.

Below, we prove that zero leakage under Shannon-difference is the strongest form of zero leakage, which trumps leakages under all other distribution differences that satisfy the coincidence axiom. Thus, while min-entropy captures a stronger notion of randomness compared with Shannon entropy and is often touted as the “correct” entropic notion for security analyses, zero mutual information capture a stronger notion of security than zero min-entropy leakage.

Corollary 1. *Zero conditional leakage under Shannon-difference implies: (i) zero conditional worst-case leakage $\mathcal{L}_{S;O|R_f}$ under Δ , and (ii) zero conditional average-case leakage $L_{S;O|R_f}$ under Δ , for any distribution difference Δ satisfying the coincidence axiom.*

Proof. Zero conditional leakage under Shannon-difference is equivalent to zero conditional mutual information between S and O (Lemma 1 and Theorem 3). For all $r \in \mathcal{V}$ and for all $o \in \mathcal{O}_r$, the distributions $(S|R_f = r)$ and $(S|R_f = r, O = o)$ are the same; and

$$\Delta((S|R_f = r), (S|R_f = r, O = o)) = 0, \quad (24)$$

by the coincidence axiom. \square

In Theorem 2, we proved that zero worst-case leakage under maxH-difference is equivalent to limited disclosure restricted by a blur operator. Theorem 4 below states that zero worst-case leakage is equivalent to zero average-case leakage for non-negative distribution differences. Since maxH-difference is non-negative, Theorem 4 establishes the equivalence of worst-case and average-case zero leakages under maxH-difference.

Theorem 4. *Zero conditional average-case leakage is equivalent to zero conditional worst-case leakage under any reasonable, non-negative distribution difference Δ .*

It can also be shown that worst-case leakage under minH-difference implies average-case leakage under minH-difference which, from Lemma 2, is equivalent to min-entropy leakage.

Theorem 5. *Zero conditional worst-case leakage under minH-difference implies zero conditional min-entropy leakage from S to O . (Note that the reverse implication does not hold, however.)*

Proof (of Theorem 5). From Lemma 2, it suffices to prove that zero worst-case conditional leakage implies zero conditional average-case leakage.

For a fixed $r \in \mathcal{V}$, let p' denote the largest probability mass in the a priori distribution, so that $\mathcal{H}_\infty(S|R_f = r) = -\log p'$. By the hypothesis, the difference between the a priori $\mathcal{H}_\infty(S|R_f = r)$ and a posteriori $\mathcal{H}_\infty(S|R_f = r, O = o)$, for any $o \in \mathcal{O}_r$, is bounded by zero

$$\mathcal{H}_\infty(S|R_f = r) - \mathcal{H}_\infty(S|R_f = r, O = o) \leq 0 \quad (25)$$

Thus, the largest probability mass in each of the a posteriori distributions is at most p' . Suppose that there exists an a posteriori distribution for which the largest probability mass is strictly less than p' . Then, in order for the largest probability mass in the marginals to be p' , there must exist another a posteriori distribution for which the largest probability mass is strictly greater than p' to compensate. This contradicts our earlier claim, and so the largest probability mass of every a posteriori distribution must be exactly p' . This is true over all r 's and o 's. \square

5 Composing Leakage Bounds

In order to generalize the cut-blur principle and more generally for composing leakage bounds, we desire a result similar to the data-processing inequality for mutual information, or cascading channels for channel capacities: Given a Markov chain of random variables $X \rightarrow Y \rightarrow Z$, we wish to bound the leakage from X to Z by the leakage from X to Y , as well as the leakage from Y to Z .

We prove that a sufficient property for achieving the first bound is the convexity of the distribution difference. If the leakage is also symmetric, we obtain the second bound.

Theorem 6. *Let $X : \mathcal{X} \rightarrow [0, 1]$, $Y : \mathcal{Y} \rightarrow [0, 1]$, and $Z : \mathcal{Z} \rightarrow [0, 1]$ be discrete, finite random variables; such that $X \rightarrow Y \rightarrow Z$ form a Markov chain in that order. Leakage from X to Z is upper bounded by leakage from X to Y under any convex distribution difference Δ .*

Proof (Worst-case leakage). By definition of worst-case leakage, there exists $y' \in \mathcal{Y}$, such that

$$\Delta(X, (X|Y = y')) = \mathcal{L}_{X;Y} \quad (26)$$

Let $\text{Dist}(\cdot)$ denote distribution. For any $z \in Z$,

$$\begin{aligned} \text{Dist}(X|Z=z) &= \sum_{y \in \mathcal{Y}} \mathbb{P}(Y=y|Z=z) \cdot \text{Dist}(X|Y=y, Z=z) \\ &= \sum_{y \in \mathcal{Y}} \mathbb{P}(Y=y|Z=z) \cdot \text{Dist}(X|Y=y) \end{aligned} \quad (27)$$

$$\Delta(X, (X|Z=z)) \leq \sum_{y \in \mathcal{Y}} \mathbb{P}(Y=y|Z=z) \cdot \Delta(X, (X|Y=y)) \quad (28)$$

$$\leq \Delta(X, (X|Y=y')) = \mathcal{L}_{X;Y}, \quad (29)$$

(27) holds from the conditional independence of X and Z by the definition of a Markov chain; so for all $(x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$

$$\mathbb{P}(X=x|Y=y) = \mathbb{P}(X=x|Y=y, Z=z). \quad (30)$$

(28) holds from the convexity of Δ . (29) follows from y' defined above. From (29), $\mathcal{L}_{X;Z} \leq \mathcal{L}_{X;Y}$. \square

Proof (Average-case leakage).

$$L_{X;Z} \leq \sum_{z \in \mathcal{Z}} \mathbb{P}(Z=z) \sum_{y \in \mathcal{Y}} \mathbb{P}(Y=y|Z=z) \cdot \Delta(X, (X|Y=y)) \quad (31)$$

$$= \sum_{y \in \mathcal{Y}} \mathbb{P}(Y=y) \cdot \Delta(X, (X|Y=y)) = L_{X;Y} \quad (32)$$

(31) holds from (28). \square

Theorem 7. *Let X , Y , and Z be discrete, finite random variables; such that $X \rightarrow Y \rightarrow Z$ form a Markov chain in that order. Leakage from X to Z is upper bounded by leakage from Y to Z for any symmetric leakage defined under any convex distribution difference Δ . (Symmetry is achieved when leakage computed in one direction is always equal to leakage in the opposite direction.)*

Proof. The proof follows the symmetry of the leakage and Theorem 6 above. \square

5.1 Generalized Cut-Blur Principle

We extend our original problem statement to the following scenario:

Given a frame $\mathcal{F} = (\mathcal{LO}, \mathcal{CH}, \mathcal{D}, \text{ends}, \text{traces})$, let $\text{src}, \text{cut}, \text{obs} \subseteq \mathcal{CH}$ be any three subsets of \mathcal{CH} , such that cut is a cut-set partitioning src from obs . Let \mathcal{S} , \mathcal{C} , and \mathcal{O} be the sets src-runs^* , cut-runs^* , and obs-runs^* , resp. We are given a partitioning function $f(\cdot, \cdot)$ and a pmf $p_{S,C}(\cdot, \cdot)$ over the joint space $\mathcal{S} \times \mathcal{C}$, and so the a posteriori distributions are now conditioned on $C = c$ for $c \in \mathcal{C}$. We assume that there are no all zero rows or columns in $p_{SC}(\cdot, \cdot)$; and we define the random variables S , $S|C$, and R_f , representing the a priori remote behavior, the a posteriori remote behavior at the cut set, and information intentionally revealed at the cut channels as before.

We can apply Theorem 6 and 7 to obtain a generalization of the cut-blur principle in cases where a cut-set imposes a Markov chain on the source, cut, and observable random variables. The leakage bounds below hold for any pmf $p_{S,C,O}(s, c, o)$ over the joint space $\mathcal{S} \times \mathcal{C} \times \mathcal{O}$. In other words, the leakage bounds hold for any conditional distribution $p_{O|C}(o|c)$.

Corollary 2. *Given a cut-set that imposes a Markov chain on the source, cut, and observable random variables; the leakage of the source behavior at the observable channels is bounded by the leakage of the source behavior at the cut when leakage is defined under a convex distribution difference Δ . If the leakage is additionally symmetric, it is also bounded by the leakage of the cut behavior at the observable channels.*

Corollary 3. *Suppose that the secret information X is not the remote behavior, but determined by the source behavior; so $X = g(S)$.*

In this case, the leakage of X at the cut is bounded by the leakage of X at the source channels when leakage is defined under a convex distribution difference Δ . If the leakage is additionally symmetric, it is also bounded by the leakage of the cut behavior at the observable channels.

Corollary 4. *Given a cut-set that imposes a Markov chain on the source, cut, and observable random variables and a leakage which is symmetric and defined under a convex distribution difference; zero leakage of the source behavior at the cut implies zero leakage of the source behavior at the observable channels, zero leakage of $X = g(S)$ at the cut, and zero leakage of X at the observable channels Δ .*

Since leakage under maxH-difference is symmetric and defined under a convex distribution difference, Theorems 6 and 7 and Corollaries 2–4 apply; in particular, the cut-blur principle is obtained from Cor. 4 under maxH-difference.

Corollary 3 bounds the leakage from X to C . Suppose we wish to compute the leakage from X to C instead, rather than merely obtaining a bound for it. Since we are given the function $g(\cdot)$ which relates S to X , we can compute the a priori and a posteriori distributions of X and $X|C = c$ from $g(\cdot)$ and the pmf $p_{SC}(\cdot, \cdot)$, and leakage is computable as the max or average difference between the a priori and a posteriori distributions.

6 Extensions to Our Results

Only completed runs were considered in Sects. 3–5: Zero leakage under maxH-difference and limited disclosure were proven equivalent only up to completed runs. Likewise, the generalization of the cut-blur principle applies only to completed runs.

Suppose that we are provided a function $h_t : \mathcal{O} \rightarrow \mathcal{O}_t$ mapping the completed observations to partial observations at some relativistic time t , so that every completed run maps to the unique partial run at time t from which it can

progress to completion. Then, we can define a random variable $O_t = h_t(O)$, and $S \rightarrow C \rightarrow O \rightarrow O_t$ form a Markov chain. Thus the leakage to the partial observations O_t is bounded above by the leakage to the completed observations, assuming that the leakage is symmetric and defined under a convex distribution difference. Moreover, we can compute a tighter bound on the leakage to O_t given $h_t(\cdot)$, in much the same way that we computed the leakage from X to C given $g(\cdot)$ in Sect. 5.1.

Suppose that we were provided the conditional probability $p_{C|S}(\cdot|\cdot)$. Then, leakage can be defined as the maximum (over all possible a priori distributions $p_S(\cdot)$) of the maximum or average (over the r 's and o 's) difference between the a priori and a posteriori distributions; and the results from the Sects. 4 and 5.1 trivially carry through. Under Shannon-difference, these correspond to channel capacity and cascading channel bounds.

Acknowledgments. We are grateful to Chris Eliopoulos Alicea, Joseph J. Ferraro, Vineet Mehta, Paul D. Rowe, John D. Ramsdell, Joe J. Rushanan, and the reviewers of this paper for helpful comments.

References

1. Alvim, M.S., Andrés, M.E., Chatzikokolakis, K., Degano, P., Palamidessi, C.: Differential privacy: on the trade-off between utility and information leakage. In: Barthe, G., Datta, A., Etalle, S. (eds.) FAST 2011. LNCS, vol. 7140, pp. 39–54. Springer, Heidelberg (2012)
2. Alvim, M.S., Chatzikokolakis, K., Palamidessi, C., Smith, G.: Measuring information leakage using generalized gain functions. In: Proceedings of the 25th Computer Security Foundations Symposium (CSF 2012) (2012)
3. Cachin, C.: Entropy Measures and Unconditional Security in Cryptography. Ph.D. thesis, Swiss Federal Institute of Technology Zürich (1997)
4. Chatzikokolakis, K., Palamidessi, C., Panangaden, P.: Anonymity protocols as noisy channels. *Inf. Comput.* **206**(2–4), 378–401 (2008)
5. Chaum, D.: The dining cryptographers problem: unconditional sender and recipient untraceability. *J. Cryptology* **1**, 65–75 (1988)
6. Clark, D., Hunt, S., Malacaria, P.: Quantitative information flow, relations and polymorphic types. *J. Logic Comput.* **15**(2), 181–199 (2005)
7. Clarkson, M.R., Myers, A.C., Schneider, F.B.: Belief in information flow. In: Proceedings of the 18th Computer Security Foundations, (CSFW-18 2005) (2005)
8. Deng, Y., Pang, J., Wu, P.: Measuring anonymity with relative entropy. In: Dimitrakos, T., Martinelli, F., Ryan, P.Y.A., Schneider, S. (eds.) FAST 2006. LNCS, vol. 4691, pp. 65–79. Springer, Heidelberg (2007)
9. Díaz, C., Seys, S., Claessens, J., Preneel, B.: Towards measuring anonymity. In: Dingledine, R., Syverson, P.F. (eds.) PET 2002. LNCS, vol. 2482, pp. 54–68. Springer, Heidelberg (2003)
10. Dodis, Y., Reyzin, L., Smith, A.: Fuzzy extractors: how to generate strong keys from biometrics and other noisy data. In: Cachin, C., Camenisch, J.L. (eds.) EUROCRYPT 2004. LNCS, vol. 3027, pp. 523–540. Springer, Heidelberg (2004)
11. Guttman, J.D., Rowe, P.D.: A cut principle for information flow. In: Proceedings of the 28th Computer Security Foundations Symposium (CSF 2015). IEEE, July 2015

12. Köpf, B., Basin, D.: An information-theoretic model for adaptive side-channel attacks. In: Proceedings of the 14th Computer and Communications Security (CCS 2007). ACM (2007)
13. Malacaria, P.: Assessing security threats of looping constructs. In: ACM SIGPLAN Notices, vol. 42. ACM (2007)
14. Smith, G.: Quantifying information flow using min-entropy. In: Proceedings of the 8th Quantitative Evaluation of Systems (QEST 2011), pp. 159–167, September 2011
15. Zhu, Y., Bettati, R.: Anonymity vs. information leakage in anonymity systems. In: Proceedings of the 25th Distributed Computing Systems (ICDCS 2005). IEEE (2005)

Data Privacy Management, and Security Assurance
10th International Workshop, DPM 2015, and 4th
International Workshop, QASA 2015, Vienna, Austria,
September 21-22, 2015. Revised Selected Papers
Garcia-Alfaro, J.; Navarro-Arribas, G.; Aldini, A.; Martinelli,
F.; Suri, N. (Eds.)
2016, XV, 291 p. 68 illus. in color., Softcover
ISBN: 978-3-319-29882-5