Formal Analysis of Information Flow in HOL

Ghassen Helali[1(✉)], Sofiène Tahar[1], Osman Hasan[1], and Tsvetan Dunchev[2]

[1] Electrical and Computer Engineering, Concordia University, Montreal, Canada
{helali,o_hasan}@encs.concordia.ca,
tahar@ece.concordia.ca

[2] Computer Science and Engineering, University of Bologna, Bologna, Italy
tsvetan.dunchev@unibo.it

Abstract. Protecting information has become very important due to the safety-critical nature of many computer-based applications. Information flow analysis plays a very important role in quantifying information-related properties under external attacks. Traditionally, information flow analysis is performed using paper-and-pencil based proofs or computer simulations but due to their inherent nature, these methods are prone to errors and thus cannot guarantee accurate analysis. As an accurate alternative, we propose to conduct the information flow analysis within the sound core of a higher-order-logic theorem prover. For this purpose, some of the most commonly used information flow measures, including Shanon entropy, mutual information, min-entropy, belief min-entropy, have been formalized. In this paper, we use the Shannon entropy and mutual information formalizations to formally verify the Data Processing and Jensen's inequalities. Moreover, we extend the security model for the case of the partial guess scenario to formalize the gain min-entropy. These formalizations allow us to reason about the information flow of a wide range of systems within a theorem prover. For illustration purposes, we perform a formal comparison between the min-entropy leakage and the gain leakage.

Keywords: Information flow · Entropy · Gain function · g-Leakage · Theorem proving · Higher-order logic

1 Introduction

Information flow analysis mainly consists of using information measures to evaluate the amount of information an attacker could get by observing the low output of a system or a protocol. Examples of this analysis include the evaluation of anonymity protocols [27] and security networks [31]. Protecting the confidentiality of sensitive information and guaranteeing a perfect level of anonymity are increasingly being required in numerous fields such as electronic payments [17], auctioning [32] and voting [7].

Various techniques for analyzing the information flow have been used. The possibilistic approach [3] consists of using non-deterministic behaviors to model

K.G. Larsen et al. (Eds.): SETTA 2017, LNCS 10606, pp. 283–299, 2017.
https://doi.org/10.1007/978-3-319-69483-2_17

the given system. Information flow analysis based on epistemic logic [11] and process calculi [28] fall into the category of possibilistic analysis. This approach is limited in terms of distinguishing between systems of varying degrees of protection [10]. As a solution for this limitation, probabilistic approaches, based on information and statistics, are considered as a more reliable alternative for computing information flow. In a threat model where the secret should be guessed in one try, the main objective of the attacker is to maximize the probability of guessing the right value of the high input (secret), in one try, by betting on the most probable element. To cater for this particular threat model, Renyi's entropy metrics [26], i.e., min-entropy and belief min-entropy are employed [30]. These measures are commonly used to effectively reason about deterministic and probabilistic systems.

Due to the difficulty of preventing the information leakage completely, "small" leaks are usually tolerated [19,29] by the above-mentioned information flow measures. With respect to the partial guess, g-leakage [2] is introduced as a generalization of the min-entropy model. The main idea of this notion is to extend the vulnerability in order to take into consideration the so called *gain function g*. The gain function models the profit that an attacker gets by using a certain guess z over the secret x. The gain value ranges from 0, when the guess has no corresponding secret value, to 1, in the case of an ideal guess. Hence the vulnerability (g-vulnerability) is redefined as the maximum expected gain over all possible guesses [2].

Traditionally, the quantitative analysis of information flow has been conducted using paper-and-pencil and computer simulation. The paper-and-pencil technique cannot cope with complex systems due to the high chances of human error while dealing with large models. On the other hand, the computer simulation approach cannot be considered accurate due to the use of numeric approximations. In order to overcome those shortcomings, formal methods [12] have been proposed as a sound technique to enhance accuracy of safety-critical systems. For instance, in [19], the probabilistic mode checker PRISM has been used to reason about several information systems, e.g., the Dining Cryptographers protocol. However, the state-space explosion problem of model checking limits the scope of its usage in information flow analysis. In contrast, higher-order-logic theorem proving can be used for the analysis of information flow to overcome these limitations.

In [8], Coble has formalized the conditional mutual information in the higher-order-logic theorem prover HOL4 [1] based on the Lebesgue integration. These fundamentals have been later used to formally analyze the privacy and the anonymity guarantees and proposed the Dining Cryptographers. However, Coble's formalization of Lebesgue integrals can only consider finite-valued measures, functions and integrals. Considering this fact, Mhamdi et al. [22] generalized the formalizations of the probability and information theories by introducing the notions of extended real numbers and formalizing Borel sigma algebra that covers larger classes of functions in terms of integrability and convergence. The authors further used these fundamentals to formalize the measures

of entropy, relative entropy and mutual information [9]. In the same context, information and conditional information leakage degree have been formalized [23] in HOL4 to assess security and anonymity protocols. Similarly, Hölzl [15,16] formalized a generic version of the measure, probability and information theories in Isabelle/HOL. This definition is very similar to Coble's work. Hölzl used the measure and the probability theories to define the Kullback-Leibler divergence, entropy, conditional entropy, mutual information and conditional mutual information and verify the properties related to the quantification of the information represented by a random variable [24].

Most of our work is based on the probability and information theories, formalized in Mhamdi's work [21], due to their completeness and availability in HOL4. We previously used these fundamentals to develop formal reasoning support for information flow using min-entropy and belief min-entropy [14], which we are extending to the gain min-entropy (g-leakage), which considers the model where the secret is totally guessed based on a partial gain about the secret using a certain guess. We will also use the formalized information measures in [23] to conduct the formal verification of the Data Processing [4] and Jensen's [18] inequalities which are major properties in information flow analysis. In the information flow context, the Data Processing Inequality (DPI) states that any post-processing of data does not increase the information leakage while Jensen's Inequality shows a relation between data averaging and data processing.

To the best of our knowledge, these measures have not been formalized before. We apply them to conduct an information leakage analysis of a threat scenario to compare min-entropy leakage and g-leakage (based on gain min-entropy) and since a small/partial leak can be tolerated, we show that the min-entropy leakage can be arbitrarily greater than the g-leakage.

2 Preliminaries

This section describes the HOL4 environment as well as the formalization of probability and information theories, which we would be building upon to formalize the DPI and Jensen's Inequality as well as the gain Min-Entropy notions.

2.1 HOL Theorem Prover

The HOL system is an environment for interactive theorem proving in higher order logic. Higher-order logic is a system of deduction with a precise semantics and is expressive enough to be used for the specification of almost all classical mathematics theories. In order to ensure secure theorem proving, the logic in the HOL system is represented in the strongly-typed functional programming language ML. An ML abstract data type is used to represent higher-order-logic theorems and the only way to interact with the theorem prover is by executing ML procedures that operate on values of these data types.

Soundness is assured as every new theorem must be verified by applying the basic axioms and primitive inference rules or any other previously verified theorems/inference rules. The HOL system has been used to formalize pure mathematics and verify industrial software and hardware systems.

2.2 Probability and Information Theory

Probability and information theories provide mathematical models to evaluate the uncertainty of random phenomena. These concepts are commonly used in different fields of engineering and computer sciences, such as signal processing, data compression and data communication, to quantify the information. Recently, the probability and information theories have been widely used for cryptographic and information flow analysis [29]. Some foundational notions of these formalizations are described below.

Let X and Y denote discrete random variables, with x and y and $\mathcal{X}$ and $\mathcal{Y}$ denoting their specific values and set of all possible values, respectively. Similarly, the probabilities of X and Y being equal to x and y is denoted by $p(x)$ and $p(y)$, respectively.

- Probability Space: *a measure space such that the measure of the state space is* 1.
- Independent Events: *Two events X and Y are independent iff* $p(X \cap Y) = p(X)p(Y)$.
- Random Variable: $X : \Omega \rightarrow \mathcal{R}$ *is a random variable iff X is $(F, \mathcal{B}(\mathcal{R}))$ measurable, where Ω is the state space, F denotes the set of events and $\mathcal{B}$ is the Borel sigma algebra of real valued functions.*
- Joint Probability: *A probabilistic measure where the likelihood of two events occurring together and at the same point in time is calculated. Joint probability is the probability of event Y occurring at the same time event X occurs. It is mathematically expressed as* $p(X \cap Y)$ *or* $p(X, Y)$.
- Conditional Probability: *A probabilistic measure where an event X will occur, given that one or more other events Y have occurred. Mathematically* $p(X|Y)$ *or* $\frac{p(X \cap Y)}{p(Y)}$.
- Expected Value: $E[X]$ of a random variable X is its Lebesgue integral with respect to the probability measure. The following properties of the expected value have been verified in HOL4 [22]:
  1. $E[X + Y] = E[X] + E[Y]$
  2. $E[aX] = aE[X]$
  3. $E[a] = a$
  4. $X \leq Y$ then $E[X] \leq E[Y]$
  5. X and Y are independent then $E[XY] = E[X]E[Y]$
- Variance and Covariance: Variance and covariance have been formalized in HOL4 using the formalization of expectation. The following properties have been verified [22]:
  1. $Var(X) = E[X^2] - E[X]^2$
  2. $Cov(X, Y) = E[XY] - E[X]E[Y]$

3. $Var(X) \geq 0$
4. $\forall a \in R, Var(aX) = a^2 Var(X)$
5. $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$

The above-mentioned definitions and properties have been utilized to formalize the foundations of information theory in HOL4 [22]. The widely used information theoretic measures can be defined as:

– The Shannon Entropy: *It measures the uncertainty of a random variable*

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

– The Conditional Entropy: *It measures the amount of uncertainty of X when Y is known*

$$H(X|Y) = -\sum_{y \in \mathcal{Y}} p(y) \sum_{x \in \mathcal{X}} p(x|y) log\, p(x|y)$$

– The Mutual Information: *It represents the amount of information that has been leaked*

$$I(X;Y) = I(Y;X) = H(X) - H(X|Y)$$

– The Relative Entropy or Kullback Leiber Distance: *It measures the inaccuracy or information divergence of assuming that the distribution is q when the true distribution is p*

$$D(p\|q) = \sum_{x \in \mathcal{X}} p(x) log \frac{p(x)}{q(x)}$$

– The Guessing Entropy: *It measures the expected number of tries required to guess the value of X optimally*

$$G(X) = \sum_{1 \leq i \leq n} i p(x_i)$$

– The Rényi Entropy: *It is related to the difficulty of guessing the value of X*

$$H_\alpha(X) = \frac{1}{1-\alpha} log\, (\sum_{x \in \mathcal{X}} P[X = x]^\alpha)$$

Among the measures listed above, Mhamdi [21] and Coble [8] formalized the Entropy, Conditional Entropy, Relative Entropy and Mutual Information in HOL4 and Hölzl [15] formalized similar concepts in Isabelle/HOL.

3 Shannon Based Information Flow

In this section, we will use the most common measures to quantify information flow, such as Shannon entropy, related entropy and mutual information formalized in [23] to formally verify the Data Processing Inequality as well as Jensen's Inequality properties.

3.1 Data Processing Inequality

According to the Data Processing Inequality (DPI), post-processing cannot increase information. Quantitatively, considering three random variables X, Y and Z satisfying the Markov property [5], the DPI states that Z cannot have more information about X than Y has about X; which is

$$I(X, Z) \leq I(X, Y)$$

Our formalization is based on the Discrete Time Markov Chain formalization (DTMC) [20], formalized information measures and probability theory [21].

The motivation behind this definition relies on the fact that the three random variables X, Y and Z satisfy the Markov property and thus

$$p(x, y, z) = p(x).p(y, z|x) = p(x).p(y|x).p(z|x, y)$$

Similarly, we can also deduce that

$$p(z|x, y) = p(z|y)$$

In order to formally verify the DPI, we first formalized the conditional mutual information

Definition 1. *(Conditional Mutual Information)*
For discrete random variables X, Y, and Z, conditional mutual information is defined as

$$\begin{aligned} I(X;Y|Z) &= \sum_{z\in Z}\sum_{y\in Y}\sum_{x\in X} P_{X,Y,Z}(x,y,z)\, log\frac{P_z(Z).P_{X,Y,Z}(x,y,z)}{P_{X,Z}(x,z).P_{Y,Z}(y,z)} \\ &= H(X,Z) + H(Y,Z) - H(X,Y,Z) - H(Z) \\ &= H(X|Z) - H(X|Y,Z) \end{aligned}$$

Then, using the commutativity of the distribution function which says that $P_{Y,Z}((y,z)) = P_{Z,Y}((z,y))$, we get the following equality: $I(X;Y,Z) = I(X;Z,Y)$ Therefore, the following result can be deduced:

$$I(X;Y|Z) + I(X;Z) = I(X;Y,Z) = I(X;Z,Y) = I(X;Z|Y) + I(X;Y)$$

Theorem 1. *(Symmetry of Mutual Information Property)*

```
⊢ ∀ b p X Y Z.
(POW (p_space p) = events p) ∧ prob_space p ∧
random_variable X p s1 ∧ random_variable Y p s2 ∧
random_variable Z p s3 ∧ random_variable (λ x.(Z x, Y x)) p s32 ∧
FINITE (p_space p) ∧
(mutual_information b p s1 s2 X Y ≠ −∞ ∧
 mutual_information b p s1 s2 X Y ≠ +∞ ∧
 mutual_information b p s1 s3 X Z ≠ −∞ ∧
```

```
  mutual_information b p s1 s3 X Z ≠ +∞) ⇒
    conditional_mutual_information b p s1 s3 s2 X Z Y +
    mutual_information b p s1 s2 X Y =
    conditional_mutual_information b p s1 s2 s3 X Y Z +
    mutual_information b p s1 s3 X Z
```

where `POW` and `FINITE` refer to the *power set* operator and *finiteness* tester in HOL4 respectively.

The proof of the property above relies on the associativity of the joint distribution, namely $P(X,(Y,Z)) = P(X,(Z,Y))$ as well as the symmetry of the additivity. Now we formally verify our main goal, DPI, as follows

Theorem 2. *(Data Processing Inequality: DPI)*
For all random variables X, Y and Z satisfying the Markov property, the DPI states that $I(X;Z) \leq I(X;Y)$

which is formalized in HOL as follows:

```
⊢ ∀b p X Y Z.(POW (p_space p) = events p) ∧ prob_space p ∧
  random_variable X p s1 ∧ random_variable Y p s2 ∧
  random_variable Z p s3 ∧
  random_variable (λ x.(Y x, Z x)) p s23 ∧
  random_variable (λ x.(Z x, Y x)) p s32 ∧
  FINITE (p_space p) ∧   mc p X Y Z∧
 (mutual_information b p s1 s2 X Y ≠ −∞ ∧
  mutual_information b p s1 s2 X Y ≠ +∞ ∧
  mutual_information b p s1 s3 X Z ≠ −∞ ∧
  mutual_information b p s1 s3 X Z ≠ +∞) ⇒
      mutual_information b p s1 s2 X Y ≥
      mutual_information b p s1 s3 X Z
```

where `mc p X Y Z` denotes the Markov property and the assertions related to the mutual information are constraints to avoid the infinite bounds of the information leakage.

For proving this theorem, we first need to prove the following two properties:

- $\forall X, Y$ random variables X and Y, the mutual information between X and Y is non-negative, $I(X;Y) \geq 0$
- if X, Y and Z form a Markov chain, then $I(X;Z|Y) = 0$

Applying the above properties to the equality:

$$I(X;Y|Z) \ + \ I(X;Z) \ = \ I(X;Z|Y) \ + \ I(X;Y)$$

as well as the previously verified property which states

$$I(X;Y,Z) = I(X;Z,Y)$$

our result can be proved.

The above result states that any transformation of the output channel Y will not give more information about the input X than itself. This concept also states that the information content of a signal cannot be increased via a local physical operation: post-processing cannot increase information. The main challenges of proving this result in HOL is to use the formalized notions of probability and information theories and reason about one of the major applications of the information theory. By proving the DPI, we show the usefulness of the theoretic information framework formalized in HOL.

3.2 Jensen's Inequality

Jensen's inequality has applications in many fields of applied mathematics and specifically information theory. For example, it plays a key role in the proof of the information inequality, $0 \leq D(p||q)$. In the following, we prove Jensen's inequality in its measure theoretic form as an application for information theory formalized in HOL. We first formalize in HOL4 the notion of convex functions:

Definition 2. *(Convex function)*

```
⊢ conv_func = ∀ x y z. (x<y ∧ y<z) ⇒
    ((f(y)-f(x))/(y-x) ≤ (f(z)-f(y))/(z-y))
```

Now, let Ω be a probability space, μ is a measure funcion on Ω, and g and f be arbitrary convex functions on the real numbers, respectively. Then according to Jensen's inequality: $\int_{\Omega} f(g(x))\,\mathrm{d}\mu \geq f(\int_{\Omega} g(x)\,\mathrm{d}\mu)$.

The most challenging part of the proof of Jensen's inequality is to prove the existence of subderivatives a and b of f, such that for all x, $a.x + b \leq f(x)$, where for $x_0 = \int_{\Omega} g(x)\,\mathrm{d}\mu$ we reach the equality $a.x_0 + b = f(x_0)$. This follows from the following two facts:

- According to the Mean value theorem, there exists ν such that if $x < \nu < \xi$, then: $\frac{f(x)-f(\xi)}{x-\xi} = f'(\nu)$
- Since f is convex, then its derivative increases, i.e. $f'(\nu) \leq f'(\xi)$

Having a and b, the proof of Jensen's inequality is straightforward:

$$\begin{aligned}
\int_{\Omega} f(g(x))\,\mathrm{d}\mu &\geq \int_{\Omega} (a.g(x) + b)\,\mathrm{d}\mu \\
&\geq a.\int_{\Omega} g(x)\,\mathrm{d}\mu + b.\int_{\Omega} 1\,\mathrm{d}\mu \\
&\geq a.x_0 + b \\
&\geq f(x_0) \\
&= f(\int_{\Omega} g(x)\,\mathrm{d}\mu)
\end{aligned}$$

Since μ is a measure, it holds that $\mu(\Omega) = 1$. Therefore $\int_{\Omega} 1\,\mathrm{d}\mu = 1$.

Using the monotonicity of sub-derivatives and the existence of a convex function properties, we formalize Jensen's inequality for the continuous case:

Theorem 3. *(Jensen's Inequality)*

```
⊢ ∀f g m. measure_space m ∧ integrable m g ∧ (b = integral m λy.b)∧
  (a *(integral m λx.g(x)) + b = f((integral m λx. g(x)))) ∧
  (∀x.a * x + b ≤ f x) ⇒
    f (integral m λx.g(x)) ≤ integral m λx.(f(g(x)))
```

The result verified above is a relation between the integral of a convex function and the value of a convex function of an integral. In the information theoretic context, Jensen's inequality relates the averaging of data to the transformation of data. This result is then formally verified in HOL4.

4 Partial Guess, Gain Function and g-Leakage

In this section, we analyze the threat scenarios where the secret is totally guessed in one try by using min-entropy measures [14]. This model is extended with the presence of the attacker's belief leading to the concept of the belief min-entropy [14]. Since the guess of the sensitive information can be partial, we formalize the gain function and the gain min-entropy. We first start by the formalization of the gain function and the related leakage properties. Compared to min-entropy and belief min-entropy measures, where the secret is assumed to be guessed in one try, the new model assumes that the secret can be partially guessed. We then introduce the notion of gain functions, which range from 0 to 1 and operate over a guess z and a secret x. Then $g(z, x)$ models the gain that an attacker gets about the secret x using the guess z.

Definition 3. *(Gain Function)*
Given a set $\mathcal{X}$ of possible secrets and a finite and non-empty set of guesses $\mathcal{Z}$, a gain function is defined as: $g : \mathcal{Z} \times \mathcal{X} \rightarrow [0,\ 1]$

For the rest of the paper, H_g, V_g and IL_g will respectively denote *gain min entropy* (also called as g-min-entropy), *(prior/posterior)gain vulnerabilty* (also called g-vulnerability) and *gain information leakage* (also called g-leakage).

Based on the gain function, we define the prior vulnerability:

Definition 4. *(Prior g-Vulnerability)*
Given a gain function g, a random variable X modeling the a-priori behavior, the prior g-vulnerability is

$$V_g(X) = \max_{z \in \mathcal{Z}} \sum_{x \in \mathcal{X}} p(X = x).g(z, x)$$

which is formalized in HOL4 as follows

```
⊢ ∀ p X g Z. prior_g_vulnerability = extreal_max_set
        (IMAGE (λz. Σ_{x∈X} distribution p X {x}.g(z,x))
           (IMAGE X (p_space p)) Z)
```

where `IMAGE f s` in HOL denotes the image of the set s by the function f which in our case is $X(\Omega)$ and `extreal_max_set (IMAGE f s)` refers to the *max* of the set `IMAGE f s` which in our case is the maximum probability over the distributions set.

Compared to the previous definition of vulnerability, the above definition shows that the gain is weighted by the probability of the secret itself, which means that the adversary $\mathcal{A}$ tries to make a guess maximizing the gain about every x from $\mathcal{X}$.

Definition 5. *(Posterior g-Vulnerability)*
Given a gain function g, a high input behavior modeled by the random variable X and a low output modeled by Y, the posterior g-vulnerability is

$$\begin{aligned} V_g(X|Y) &= \sum_{y\in\mathcal{Y}} \max_{z\in\mathcal{Z}} \sum_{x\in\mathcal{X}} p(X=x).p(Y=y|X=x).g(z,x) \\ &= \sum_{y\in\mathcal{Y}} \max_{z\in\mathcal{Z}} \sum_{x\in\mathcal{X}} p(X=x,Y=y).g(z,x) \end{aligned}$$

This definition can be formalized in HOL4 as

```
⊢ ∀p X Y g Z. posterior_g_vulnerability =
  ∑ y∈𝒴 extreal_max_set(IMAGE (λz. ∑ x∈𝒳 distribution p Y {y}.
    conditional_distribution p X Y ({x},{y}).g(z,x))
        (IMAGE X (p_space p)))
```

Now we define the uncertainty measures; g-min-entropy (initial uncertainty), and conditional g-min-entropy, (remaining uncertainty) which will be used to define the g-leakage.

Definition 6. *(g-Min-Entropy, g-Conditional-Min-Entropy and g-Leakage)*

```
⊢ g_min_entropy p X g Z = -log(prior_g_vulnerability p X g Z)
⊢ g_conditional_min_entropy p X Y g Z =
   -log(posterior_g_vulnerability p X Y g Z)
⊢ g_information_leakage p X Y g Z =
   g_min_entropy p X g Z - g_conditional_min_entropy p X Y g Z
```

We next consider a model where the attacker can get partial knowledge about the secret using a certain guess. The gain function models the benefit that the attacker gets about the secret. We then verify that the prior g-vulnerability cannot exceed the posterior g-vulnerability. Thus the g-leakage is positive:

Theorem 4. *(Positive g-Leakage)*

```
⊢ ∀ p X Y g Z. prob_space p ∧ FINITE (p_spac p) ∧ FINITE Z ∧
  Z ≠ ∅ ∧ p_space p ≠ ∅ ∧   ∀x. x ∈ p_space p ⇒
  {x} ∈ events p ∧ events p = POW(p_space p) ∧
  ∀x z. 0 ≤ g(z, x) ∧ g(z, x) ≤ 1⇒
  0≤ g_information_leakage p X Y g Z
```

Proof. First, note that the gain information leakage (g-leakage) is $IL_g = H_g(X) - H_g(X|Y) = log(V_g(X|Y)) - log(V_g(X))$. After simplification, our goal will be reduced to $V_g(X) \leq V_g(X|Y)$. Then

$$\begin{aligned} V_g(X) &= \max_z \sum_x \sum_y P(X = x, Y = y).g(z, x) \\ &\leq \sum_y \max_z \sum_x P(X = x, Y = y).g(z, x) \\ &\leq V_g(X|Y) \end{aligned}$$

Next, we will study the case when the g-leakage is equal to zero. We will evaluate the condition under which this result occurs. Before stating this property formally we need first to define the notion of the expected gain of a guess z. With respect to the same configuration, the prior and posterior expected gains are defined as:

Definition 7. *(Prior Expected Gain)*

$$E_g(z) \ = \ \sum_x P(X \ = \ x).g(z, x)$$

which is formalized in HOL4 as follows

$\vdash \forall$`p X g z. prior_expected_gain p X g Z =`
$\sum_{x \in X(\Omega)}$ `distribution p X {x}.g(z,x)`

Definition 8. *(Posterior Expected Gain)*
Given an output y the expected gain of a guess z is

$$E_g(z, y) \ = \ \sum_x P(X \ = \ x)P(Y \ = \ y \mid X \ = \ x).g(z, x)$$

The HOL4 formalization of this definition is

$\vdash \forall$`p X Y y g z. posterior_expected_gain p X Y y g z =`
$\sum_{x \in X(\Omega)}$ `distribution p X {x}.`
`conditional_distribution p Y X ({y},{x}).g(z,x)`

In the context of vulnerabilities and information flow, these definitions satisfy the following properties:

Theorem 5. *(Expected Gain and Vulnerabilities)*

- $V_g(X) = \max_z E_g(z)$
- $V_g(X \mid Y) = \sum_y \max_z E_g(z,y)$
- $E_g(z) = \sum_y E_g(z, y)$

We prove these results in the HOL4 theorem prover as follows

```
⊢ ∀p X g Z. prior_g_vulnerability p X g Z =
   extreal_max_set (IMAGE (λz. prior_expected_gain p X g z) Z)
```

```
⊢ ∀p X Y g Z. FINITE Ω ∧ prob_space p ∧
   (∀x. x ∈ Ω ⇒ {x} ∈ events p) ⇒
      posterior_g_vulnerability p X Y g Z = ∑ extreal_max_set
                                            y∈Y(Ω)
      (IMAGE (λz. posterior_expected_gain p X Y y g z) Z)
```

```
⊢ ∀p X Y y g z. prob_space p ∧ FINITE Ω ∧ Ω ≠ ∅ ∧
   (∀x. x∈ Ω ⇒{x} ∈ events p) ∧
   (∀x. x∈(X(Ω)) ⇒ 0≤g(z,x) ∧ g(z,x)≤1) ⇒
      prior_expected_gain p X g z =
      ∑ posterior_expected_gain p X Y y g z
      y∈Y(Ω)
```

We will later use these properties in order to verify the zero valued g-leakage result. We prove the fact that the g-leakage of 0 is related to the expected gain of all outputs, i.e., this statement occurs if there exists a guess z' maximizing the expected gain for all outputs y.

Theorem 6. *(Zero Gain Information Leakage)*
Given a random variable X modeling the initial uncertainty, a random variable Y modeling the remaining uncertainty and a gain function g, the g-leakage is 0 *if there exists a guess $z' \in Z$ such that: $\forall z\ y.\ E_g(z', y) \geq E_g(z, y)$*

In the HOL4 environment, this property is formalized as follows:

```
⊢ ∀p X Y y g Z. (prob_space p ∧ FINITE (p_space p) ∧
   ((p_space p) ≠ ∅) ∧
   (∀x. x ∈ p_space p ⇒ {x} ∈ events p) ∧ (FINITE Z) ∧
   (events p = POW (p_space p)) ∧ (Z ≠ ∅) ∧
   (∀x z. (0 ≤ (g (z,x))) ∧ ((g (z, x)) ≤ 1)) ∧
   (0 < prior_g_vulnerability p X g Z)) ⇒
   ((∃z'. (z'∈Z) ∧ (∀z y. (posterior_expected_gain p X Y y g z) ≤
   (posterior_expected_gain p X Y y g z'))) ⇒
     (g_information_leakage p X Y g Z = 0))
```

Proof. If such a guess exists, then we first prove that it corresponds to the maximum prior expected gain ∀z. $E_g(z') \geq E_g(z)$. Then, using the previous results, it follows that the posterior g-vulnerability is equal to the prior expected gain of the best guess

$$V_g(X,Y) = \sum_y E_g(z,y) = \sum_y \max_z E_g(z,y) = \sum_y E_g(z',y) = E_g(z')$$

However, since $E_g(z')$ is the prior g-vulnerability $V_g(X)$, from Theorem 5, so it follows from the definition of the g-leakage that this measure is 0.

Based on the soundness of theorem proving, the above-mentioned formally verified theorems are guaranteed to be accurate and contain all the required assumptions. Moreover, these results can be built upon to reason about information flow analysis of various applications within the sound core of a theorem prover.

5 Min-Entropy Leakage and g-Leakage

In this section, we illustrate the practical usefulness of the theoretical foundations developed in this paper so far. We will present a threat scenario in which we conduct a comparison between the min-entropy leakage and the g-leakage and show that the g-leakage can be smaller than min-entropy leakage. Consider the channel (Matrix of transitional probabilities), described in Table 2, where x_i are the high inputs and y_i are the outputs modelled, respectively, with the random variables X and Y. We assume for this example that inputs and outputs are *uniformly* distributed.

Table 1. Transition channel

	y_1	y_2
x_1	$\frac{1}{2}$	$\frac{1}{2}$
x_2	1	0
x_3	0	1

Table 2. Gain function

g_d	x_1	x_2	x_3
z_1	1	0	0
z_2	0	1	0.98
z_3	0	0.98	1

For our particular example, we consider the gain function called the distance gain function between the secrets, assuming that $\mathcal{X} = \mathcal{Z}$, $g_d(z, x) = 1 - \mathrm{d}(z, x)$ where $\mathrm{d}(z, x)$ is the normalized distance between z and x. Using this configuration, we prove that the min-entropy leakage is equal to $log\ 2 = 1$ and g-leakage is equal to $log\ \frac{2}{1.98}$. The formalization of this theorem in HOL4 is

Theorem 7. *(Comparing Min-Entropy Leakage and g-Leakage)*

```
⊢ ∀p X Y Z g. (prob_space p) ∧ (FINITE (p_space p)) ∧
   (∀x. (x ∈ p_space p) ⇒ {x} ∈ events p) ∧
   ((IMAGE X (p_space p)) = {0;1;2}) ∧
   ((IMAGE Y (p_space p)) = {0;1}) ∧ (Z = {0;1;2}) ∧
   (∀x. x ∈ (IMAGE X (p_space p)) ⇒ distribution p X {x} =
     (1/|IMAGE X (p_space p)|) ∧
   (∀y. y ∈ (IMAGE Y (p_space p)) ⇒
   distribution p Y {y} = (1/|IMAGE Y (p_space p)|) ∧
   (conditional_distribution p Y X ({0},{0}) = (1/2)) ∧
   (conditional_distribution p Y X ({0},{1}) = 1) ∧
   (conditional_distribution p Y X ({0},{2}) = 0) ∧
   (conditional_distribution p Y X ({1},{0}) = (1/2)) ∧
```

```
(conditional_distribution p Y X ({1},{1}) = 0) ∧
(conditional_distribution p Y X ({1},{2}) = 1) ∧
(g(0,0) = 1) ∧ (g(0,1) = 0) ∧ (g(0,2) = 0) ∧ (g(1,0) = 0) ∧
(g(1,1) = 1) ∧ (g(1,2) = Normal (0.98)) ∧ (g(2,0) = 0) ∧
(g(2,1) = Normal (0.98)) ∧ (g(2,2) = 1)) ⇒
  ((information_leakage p X Y = 1) ∧
  (g_information_leakage p X Y g Z = log (2/(Normal(1.98)))))
```

where `g` refers to the gain function, `conditional_distribution` denotes the transition distribution with respect to Tables 1 and 2 and the term `Normal` is used for the extended real numbers theory.

The proof of this result is conducted using the vulnerability properties, probability reasoning and real analysis. The first part of the theorem is proved using Theorem 4.2. The second part of the goal is verified by computing the values of the initial and remaining vulnerabilities. We find that

$$V_{gd}(\mathrm{X}) = \frac{1}{3}\max\{1+0+0, 0+1+0.98, 0+0.98+1\} = 0.66$$

Now for the posterior vulnerability, we calculate the posterior distribution $P(X = x|Y = y_1) = (\frac{1}{3}, \frac{2}{3}, 0)$.

$$V_{gd}(X = x|Y = y_1) = max \left\{ \begin{array}{l} \frac{1}{3}.1 + \frac{2}{3}.0 + 0.0, \\ \frac{1}{3}.0 + \frac{2}{3}.1 + 0.0.98, \\ \frac{1}{3}.0 + \frac{2}{3}.0.98 + 0.1 \end{array} \right\} = \frac{2}{3}$$

Similarly, we prove that $V_{gd}(X = x|Y = y_2) = \frac{2}{3}$ and then by rewriting these values on their corresponding quantities, we get $IL_{gd}(X, Y) = log\frac{\frac{2}{3}}{0.66} \approx log\frac{2}{1.98}$. Here the min-entropy leakage is greater than the g-leakage. Perceptively, this example differentiates between x_2 and x_3. Due to the relations (z_2, x_3) and (z_3, x_2), under the distance gain function, it follows that x_2 and x_3 are so close (a gain of 0.98). Thus the g-vulnerability hardly increases. The proof of this result required 550 lines of HOL4 code [13] and around 60 man-hours in terms of reasoning effort. These results are considered to be accurate and the analysis covers any type of systems (in terms of state space size).

6 Conclusion

This paper presents a formalization of some of the most commonly used properties of information flow in higher-order logic. These properties, depending on the threat model, are based on Shannon entropy and gain min-entropy. This formalization provides a more reliable and richer information flow analysis framework compared to the traditional definitions of quantitative information flow analysis as formalized measures cover a wide variety of threat scenarios. We used the formalized notions of Shannon entropy to verify the Data Processing Inequality, which states that leakage cannot be increased by post processing of the information, and Jensen's Inequality, which is a relation between the averaging and the

processing of information flow. The g-leakage is introduced as a generalization of the min-entropy leakage and belief min-entropy leakage to assess the case where the secret is guessed partially using a gain function, which models the benefit that an attacker gets about the secret. Gain functions engender the possibility to cover a variety of operational scenarios. The proposed formalization can be built upon to conduct the information flow analysis within the sound core of a theorem prover and thus the analysis is guaranteed to be free of approximation and precision errors. For illustration purposes, we performed a comparison analysis between the min-entropy leakage and the g-leakage using the HOL4 theorem prover and the analysis results were found to be generic and accurate.

This work is conducted as a formal framework that can be used to formally verify many information flow aspects depending on the threat model. It provides a reasonable foundation for information flow in HOL. Many applications can be analyzed using our formalization, such as the Crowds protocol [25] and Freenets [6]. We are aiming to extend this work for the formal analysis of channel capacity (min-capacity and gain-capacity) and compare them with Shanon capacity. Starting from a specific leakage bound, our work can be used to evaluate the input set based on the output set. This formalization can in turn be used to formally ensure a specific level of security of critical information.

References

1. HOL4, hol.sourceforge.net (2017)
2. Alvim, M.S., Chatzikokolakis, K., Palamidessi, C., Smith, G.: Measuring information leakage using generalized gain functions. In: IEEE Symposium on Computer Security Foundations, pp. 265–279 (2012)
3. Andrea, S.: Possibilistic information theory: a coding theoretic approach. Fuzzy Sets Syst. **132**(1), 11–32 (2002)
4. Beaudry, N.J., Renner, R.: An intuitive proof of the data processing inequality. Quantum Inform. Comput. **12**(5–6), 432–441 (2012)
5. Chung, K.L.: Markov Chains with Stationary Transition Probabilities (1967)
6. Clarke, I., Sandberg, O., Wiley, B., Hong, T.W.: Freenet: a distributed anonymous information storage and retrieval system. In: Federrath, H. (ed.) Designing Privacy Enhancing Technologies. LNCS, vol. 2009, pp. 46–66. Springer, Heidelberg (2001). doi:10.1007/3-540-44702-4_4
7. Clarkson, M.R., Chong, S., Myers, A.C.: Civitas: toward a secure voting system. In: IEEE Symposium on Security and Privacy, pp. 354–368. IEEE Computer Society (2008)
8. Coble, A.R.: Anonymity, Information, and Machine-Assisted Proof. Ph.D. thesis, King's College, University of Cambridge, UK (2010)
9. Cover, T.M., Thomas, J.: Entropy, relative entropy and mutual information. In: Elements of Information Theory. Wiley-Interscience (1991)
10. Dubois, D., Nguyen, H.T., Prade, H.: Possibility theory, probability and fuzzy sets: misunderstandings, bridges and gaps. In: Dubois, D., Prade, H. (eds.) Fundamentals of Fuzzy Sets. The Handbooks of Fuzzy Sets Series, pp. 343–438. Kluwer, Boston (2000)
11. Halpern, J., O'Neill, K.: Anonymity and information hiding in multiagent systems. J. Comput. Secur. **13**(3), 483–514 (2005)

12. Hasan, O., Tahar, S.: Formal verification methods. In: Encyclopedia of Information Science and Technology, pp. 7162–7170. IGI Global Pub. (2015)
13. Helali, G., Dunchev, C., Hasan, O., Tahar, S.: Towards The Quantitative Analysis of Information Flow in HOL, HOL4 code (2017). http://hvg.ece.concordia.ca/projects/prob-it/gainMinEntropy.php
14. Helali, G., Hasan, O., Tahar, S.: Formal analysis of information flow using min-entropy and belief min-entropy. In: Iyoda, J., de Moura, J. (eds.) SBMF 2013. LNCS, vol. 8195, pp. 131–146. Springer, Heidelberg (2013). doi:10.1007/978-3-642-41071-0_10
15. Hölzl, J.: Construction and Stochastic Applications of Measure Spaces in Higher-Order Logic. Ph.D. thesis, Institut für Informatik, Technische Universität München, Germany (2012)
16. Hölzl, J., Heller, A.: Three chapters of measure theory in Isabelle/HOL. In: van Eekelen, M., Geuvers, H., Schmaltz, J., Wiedijk, F. (eds.) ITP 2011. LNCS, vol. 6898, pp. 135–151. Springer, Heidelberg (2011). doi:10.1007/978-3-642-22863-6_12
17. Hua, J., Jing, Y.: On-line payment and security of e-commerce. In: International Conference on Computer Engineering and Applications, pp. 545–550. CEA, WSEAS (2007)
18. Jebara, T., Pentland, A.: On Reversing Jensen's Inequality. In: Advances in Neural Information Processing Systems 13. MIT Press (2000)
19. Chatzikokolakis, K., Palamidessi, C., Panangaden, P.: Anonymity protocols as noisy channels. Inf. Comput. **206**(2–4), 378–401 (2008)
20. Liu, L.: Formalization of Discrete-time Markov Chains in HOL. Ph.D. thesis, Dept. of Electrical and Computer Engineering, Concordia University, Canada (2013)
21. Mhamdi, T.: Information-Theoretic Analysis using Theorem Proving. Ph.D. thesis, Dept. of Electrical and Computer Engineering, Concordia University, Canada (2012)
22. Mhamdi, T., Hasan, O., Tahar, S.: Formalization of entropy measures in HOL. In: Eekelen, M., Geuvers, H., Schmaltz, J., Wiedijk, F. (eds.) ITP 2011. LNCS, vol. 6898, pp. 233–248. Springer, Heidelberg (2011). doi:10.1007/978-3-642-22863-6_18
23. Mhamdi, T., Hasan, O., Tahar, S.: Quantitative analysis of information flow using theorem proving. In: Aoki, T., Taguchi, K. (eds.) ICFEM 2012. LNCS, vol. 7635, pp. 119–134. Springer, Heidelberg (2012). doi:10.1007/978-3-642-34281-3_11
24. Mhamdi, T., Hasan, O., Tahar, S.: Formalization of measure theory and lebesgue integration for probabilistic analysis in HOL. ACM Trans. Embedded Comput. Syst. **12**(1) (2013)
25. Reiter, M.K., Rubin, A.D.: Crowds: anonymity for web transactions. ACM Trans. Inform. Syst. Secur. **1**(1), 66–92 (1998)
26. Rényi, A.: On measures of entropy and information. In: Berkeley Symposium on Mathematics, Statistics and Probability, pp. 547–561 (1961)
27. Sassone, V., ElSalamouny, E., Hamadou, S.: Trust in crowds: probabilistic behaviour in anonymity protocols. In: Wirsing, M., Hofmann, M., Rauschmayer, A. (eds.) TGC 2010. LNCS, vol. 6084, pp. 88–102. Springer, Heidelberg (2010). doi:10.1007/978-3-642-15640-3_7
28. Schneider, S., Sidiropoulos, A.: CSP and anonymity. In: Bertino, E., Kurth, H., Martella, G., Montolivo, E. (eds.) ESORICS 1996. LNCS, vol. 1146, pp. 198–218. Springer, Heidelberg (1996). doi:10.1007/3-540-61770-1_38
29. Smith, G.: On the foundations of quantitative information flow. In: Alfaro, L. (ed.) FoSSaCS 2009. LNCS, vol. 5504, pp. 288–302. Springer, Heidelberg (2009). doi:10.1007/978-3-642-00596-1_21

30. Smith, G.: Quantifying information flow using min-entropy. In: IEEE International Conference on Quantitative Evaluation of Systems, pp. 159–167 (2011)
31. Syverson, P., Goldschlag, D., Reed, M.: Anonymous connections and onion routing. In: IEEE Symposium on Security and Privacy, Oackland, California, pp. 44–54 (1997)
32. Trevathan, J.: Privacy and Security in Online Auctions. Ph.D. thesis, School of Mathematics, Physics and Information Technology, James Cook University, Australia (2007)