Quantitative Analysis of Information Flow using Theorem Proving

Tarek Mhamdi, Osman Hasan, and Sofiène Tahar

ECE Department, Concordia University, Montreal, QC, Canada
{mhamdi,o_hasan,tahar}@ece.concordia.ca

Abstract. Quantitative analysis of information flow is widely used to measure how much information was leaked from the secret inputs to the outputs or public inputs of a program. We propose to conduct the quantitative analysis of information flow within the trusted kernel of a higher-order-logic theorem prover in order to overcome the inaccuracy limitations of traditional analysis techniques used in this domain. For this purpose, we present the formalization of the Kullback-Leibler divergence that can be used as a unified measure of information leakage. Furthermore, we propose two new measures of information leakage, namely the information leakage degree and the conditional information leakage degree. We also formalize the notion of anonymity-based single MIX and use the channel capacity as a measure of information leakage in the MIX. Finally, for illustration purposes, we show how our framework allowed us to find a counter-example for a theorem that was reported in the literature to describe the leakage properties of the anonymity-based single MIX.

1 Introduction

Quantitative information flow [19, 17] allows to measure how much information about the high security inputs of a system can be leaked, accidentally or maliciously, by observing the systems outputs and possibly the low security inputs. Unlike non-interference analysis, which only determines whether a system is completely secure or not completely secure, quantitative information flow provides an information theoretic measure on how secure or insecure a system is. Quantitative information flow is extensively used for analyzing anonymity protocols and secure communications using various measures of information flow. Serjantov [18] and Diaz et al. [6] independently proposed to use the entropy to define the quality of anonymity and to compare different anonymity systems. Malacaria [12] defined the leakage of confidential information in a program as the conditional mutual information between its outputs and secret inputs, given the knowledge of its low security inputs. Deng [5] proposed relative entropy as a measure of the amount of information revealed to the attacker after observing the outcomes of the protocol, together with the a priori information. Chatzikokolakis [1] modeled anonymity protocols as noisy channels and used the channel capacity as a measure of the loss of anonymity.

Traditionally, paper-and-pencil based analysis or computer simulations have been used for quantitative analysis of information flow. Paper-and pencil analysis does not scale well to complex systems and is prone to human error. Computer simulation, on the other hand, lacks in accuracy due to numerical approximations. These analysis inaccuracies may result in compromising national security and finances given the safety and security-critical nature of systems where information flow analysis is usually used.

As an alternative approach, we propose a machine-assisted analysis of information flow by conducting the analysis within the trusted kernel of a higher-order-logic theorem prover [9]. Higher-order logic is a system of deduction with precise semantics and, due to its high expressiveness, can be used to describe any mathematical relationship. Interactive theorem proving is the field of computer science and mathematical logic concerned with computer-based formal proof tools that require human assistance. We argue that the high expressiveness of higher-order logic can be leveraged to formalize the commonly used information leakage measures by building upon the existing formalization of measure, integration and probability [13], and information theories [14]. This foundational formalization can hence be used to formally reason about quantitative properties of information flow analysis within the sound core of a theorem prover and thus guarantee accuracy of the analysis.

In particular, this paper presents an extension of existing theories of measure, Lebesgue integration and probability [13] to cater for measures involving multiple random variables. Building upon this formalization, we present a higher-order-logic formalization of the Kullback-Leibler (KL) divergence [4] from which we can derive the formalization of most of the information leakage measures presented in the literature so far. Furthermore, we propose two novel measures of information leakage termed as degrees of information leakage. We will show that they are somehow related to the existing measures but have the advantage that they not only quantify the information leakage but also describe the quality of leakage by normalizing the measure by the maximum leakage that the system allows under extreme situations. The formalization reported in this paper has been done using the HOL4 [10] theorem prover. The prime motivation behind this choice is the availability of the measure, probability and integration theories [13, 14].

We illustrate the usefulness of the framework for formal analysis of quantitative information flow by tackling an anonymity-based single MIX application [20]. We provide a higher-order-logic formalization of the single MIX as well as the channel capacity which we use as a measure of information leakage within the MIX. We then formally verify that a sender using the MIX as a covert channel, can transmit information through the MIX at maximum capacity without having to communicate with all the receivers. This result allowed us to identify a flaw in the paper-and-pencil based analysis of a similar problem [20] which clearly indicates the usefulness of the proposed technique.

The rest of the paper is organized as follows: In Section 2 we briefly present the proposed extensions to the formalization of measure, integration and probability theories [13]. In Section 3, we describe our formalization of KL divergence and how we use it to formalize various measures of information leakage as well as prove their properties in HOL. We introduce novel information leakage degrees in Section 4. In Section 5, we present an analysis of an anonymity-based single MIX. We discuss related work in Section 6 and conclude the paper in Section 7.

2 Measure, Integration and Probabilities

In this section, we extend the formalization of measure theory [13] and Lebesgue integration [13] as well as probability theory [14] to support products of measure spaces and joint distributions of two or more random variables. Both are necessary to define measures over multiple random variables.

Products of Measure Spaces. Let $m_1 = (X_1, \mathcal{S}_1, \mu_1)$ and $m_2 = (X_2, \mathcal{S}_2, \mu_2)$ be two measure spaces. The product of m_1 and m_2 is defined to be the measure space $(X_1 \times X_2, \mathcal{S}, \mu)$, where $\mathcal{S}$ is the sigma algebra on $X_1 \times X_2$ generated by subsets of the form $A_1 \times A_2$ where $A_1 \in \mathcal{S}_1$, and $A_2 \in \mathcal{S}_2$. The measure μ is defined for σ-finite measure spaces as

$$\mu(A) = \int_{X_1} \mu_2(\{y \in X_2 | (x, y) \in A\})\, d\mu_1$$

and $\mathcal{S}$ is defined using the `sigma` operator which returns the smallest sigma algebra containing a set of subsets, i.e., the product subsets in this case.

Let $g(s_1)$ be the function $s_2 \to (s_1, s_2)$ and `PREIMAGE` denote the HOL function for inverse image, then the product measure is formalized as

```
⊢ prod_measure m1 m2 =
    (λa. integral m1 (λs1. measure m2 (PREIMAGE g(s1) a)))
```

We verified in HOL that the product measure can be reduced to $\mu(a_1 \times a_2) = \mu_1(a_1) \times \mu_2(a_2)$ for finite measure spaces.

```
⊢ prod_measure m1 m2 (a1 × a2) = measure m1 a1 × measure m2 a2
```

We use the above definitions to define products of more than two measure spaces as follows. $X_1 \times X_2 \times X_3 = X_1 \times (X_2 \times X_3)$ and $\mu_1 \times \mu_2 \times \mu_3$ is defined as $\mu_1 \times (\mu_2 \times \mu_3)$. We also define the notion of absolutely continuous measures where μ_1 is said to be absolutely continuous w.r.t μ_2 iff for every measurable set A, $\mu_2(A) = 0$ implies $\mu_1(A) = 0$. Further details about this formalization can be found in [15].

Joint Distribution. The joint distribution of two random variables defined on the same probability space is defined as,

$$p_{XY}(a) = p(\{(X, Y) \in a\})$$

```
⊢ joint_distribution p X Y =
        (λa. prob p (PREIMAGE (λx. (X x,Y x)) a ∩ Ω))
```

Here the intersection with the sample space Ω is required because HOL functions are total and should be defined on all variables of the specific type instead of only on Ω. The joint distribution of any number of variables can be defined in a similar way. We formally verified a number of joint distribution properties in HOL [15] and some of the useful ones are given below:

```
⊢ 0 ≤ joint_distribution p X Y a
⊢ joint_distribution p X Y = joint_distribution p Y X
⊢ joint_distribution p X Y (a × b)  ≤ distribution p X a
⊢ joint_distribution p X Y (a × b)  ≤ distribution p Y b
```

We also verified that the joint distribution is absolutely continuous w.r.t to the product of marginal distributions and the following useful properties in HOL.

$$p_X(a) = \sum_{y \in Y(\Omega)} p_{XY}(a \times \{y\})$$

$$p_Y(b) = \sum_{x \in X(\Omega)} p_{XY}(\{x\} \times b)$$

The formalization of joint distributions and products of measures spaces, presented in the next section, play a vital role in formalizing information-theoretic measures with multiple random variables.

3 Measures of Information Leakage

In this section, we first provide a formalization of the Radon-Nikodym derivative [8] which is then used to define the KL divergence. Based on the latter, we define most of the commonly used measures of information leakage. We start by providing general definitions which are valid for both discrete and continuous cases and then prove the corresponding reduced expressions where the measures considered are absolutely continuous over finite spaces.

3.1 Radon-Nikodym Derivative

The Radon-Nikodym derivative of a measure ν with respect to the measure μ is defined as a non-negative measurable function f, satisfying the following formula, for any measurable set A.

$$\int_A f \, d\mu = \nu(A)$$

We formalize the Radon-Nikodym derivative in HOL as

```
⊢ RN_deriv m v =
   @f. f IN measurable (X, S) Borel ∧
   ∀x ∈ X, 0 ≤ f x ∧
   ∀a ∈ S, integral m (λx. f x * indicator_fn a x) = v a
```

where @ denotes the Hilbert-choice operator in HOL. The existence of the Radon-Nikodym derivative is guaranteed for absolutely continuous measures by the Radon-Nikodym theorem.

Theorem 1. *If ν is absolutely continuous with respect to μ, then there exists a non-negative measurable function f such that for any measurable set A,*

$$\int_A f \, d\mu = \nu(A)$$

We proved the Radon-Nikodym theorem in HOL for finite measures which can be easily generalized to σ-finite measures.

```
⊢ ∀m v s st.
    measure_space (s,st,m) ∧ measure_space (s,st,v) ∧
    measure_absolutely_continuous (s,st,m) (s,st,v) ∧
    v s ≠ ∞ ∧ m s ≠ ∞ ⇒
       ∃f. f ∈ measurable (s,st) Borel ∧
       ∀x ∈ s, 0 ≤ f x < ∞ ∧
       ∀a ∈ st,
          integral m (λx. f x * indicator_fn a x) = v a
```

The formal reasoning about the above theorem is primarily based on the Lebesgue monotone convergence and the following lemma which, to the best of our knowledge, has not been referred to in paper-and-pencil based mathematical texts before.

Lemma 1. *If P is a non-empty set of extended-real valued functions closed under the max operator, g is monotone over P and $g(P)$ is upper bounded, then there exists a monotonically increasing sequence $f(n)$ of functions, elements of P, such that*

$$\sup_{n\in\mathbb{N}} g(f(n)) = \sup_{f\in P} g(f)$$

Finally, we formally verified various properties of the Radon-Nikodym derivative. For instance, we prove that for absolutely continuous measures defined over a finite space, the derivative reduces to

```
⊢ ∀x ∈ s, u{x} ≠ 0 ⇒ RN_deriv u v x = v{x} / u{x}
```

The following properties play a vital role in formally reasoning about the Radon-Nikodym derivative and have also been formally verified.

```
⊢ ∀x ∈ s, 0 ≤ RN_deriv m v x < ∞
⊢ RN_deriv ∈ measurable (s,st) Borel
⊢ ∀a ∈ st, integral m (λx. RN_deriv m v x * indicator_fn a x) = v a
```

3.2 Kullback-Leibler Divergence

The Kullback-Leibler (KL) divergence [4] $D_{KL}(\mu||\nu)$ is a measure of the distance between two distributions μ and ν. It can be used to define most information-theoretic measures such as the mutual information and entropy and can, hence, be used to provide a unified framework to formalize most information leakage measures. It is because of this reason that we propose to formalize the KL divergence in this paper as it will facilitate formal reasoning about a wide variety of information flow related properties. The KL divergence is defined as

$$D_{KL}(\mu||\nu) = -\int_X log\frac{d\nu}{d\mu}\, d\mu$$

where $\frac{d\nu}{d\mu}$ is the Radon-Nikodym derivative of ν with respect to μ. The KL divergence is formalized in HOL as

```
⊢ KL_divergence b m v = -integral m (λx. logr b((RN_deriv m v)x))
```

where b is the base of the logarithm. D_{KL} is measured in *bits* when $b = 2$. We formally verify various properties of the KL divergence. For instance, we prove that for absolutely continuous measures over a finite space, it reduces to

$$D_{KL}(\mu||\nu) = \sum_{x\in s} \mu\{x\} \log \frac{\mu\{x\}}{\nu\{x\}}$$

```
⊢ KL_divergence b u v = SIGMA (λx. u{x} logr b (u{x} / v{x})) s
```

We also prove the following properties

```
⊢ KL_divergence b u u = 0
⊢ 1 ≤ b  ⇒  0 ≤ KL_divergence b u v
```

The non-negativity of the KL divergence for absolutely continuous probability measures over finite spaces is extensively used to prove the properties of information theory measures like the mutual information and entropy. To prove this result, we use the Jensen's inequality and the concavity of the logarithm function.

We show in the subsequent sections how we use the KL divergence to formalize the mutual information, Shannon entropy, conditional entropy and the conditional mutual information, which are some of the most commonly used measures of information leakage.

3.3 Mutual Information and Entropy

The mutual information has been proposed as a measure of information leakage [20] from the secure inputs S of a program to its public outputs O as it represents the mutual dependence between the two random variables S and O. The mutual information is defined as the KL divergence between the joint distribution and the product of marginal distributions. The following is a formalization of the mutual information in HOL.

```
⊢ I(X;Y) = KL_divergence b (joint_distribution p X Y)
                           prod_measure (distribution p X)
                                        (distribution p Y)
```

We prove various properties of the mutual information in HOL, such as the non-negativity, symmetry and reduced expression for finite spaces, using the result that the joint distribution is absolutely continuous w.r.t the product of marginal distributions.

```
⊢ 0 ≤ I(X;Y)
⊢ I(X;Y) = I(Y;X)
⊢ I(X;Y) = 0 ⇔ X and Y independent
⊢ I(X;Y) = SIGMA (λ(x,y). p{(x,y)} logr b (p{(x,y)}/p{x}p{y})) s
```

The Shannon entropy `H(X)` was one of the first measures to be proposed to analyze anonymity protocols and secure communications [18, 6] as it intuitively measures the uncertainty of a random variable `X`. It can be defined as the expectation of p_X or simply as $I(X;X)$.

```
⊢ H(X) = I(X;X)
```

We prove that it can also be expressed in terms of the KL divergence between p_X and the uniform distribution p_X^u, where N is the size of the alphabet of X.

```
⊢ H(X) = log(N) - KL_divergence b (distribution p X)
                                  (uniform_dist p X)
```

The cross entropy $H(X, Y)$ is the entropy of the random variable (X, Y) and hence there is no need for a separate formalization of the cross entropy.
The conditional entropy is defined in terms of the KL divergence as follows:

```
⊢ H(X|Y) = log(N) - KL_divergence b (joint_distribution p X Y)
                             prod_measure (uniform_dist p X)
                                          (distribution p Y)
```

The entropy properties that we prove in HOL include:

```
⊢ 0 ≤ H(X) ≤ log(N)
⊢ max(H(X),H(Y)) ≤ H(X,Y) ≤ H(X) + H(Y)
⊢ H(X|Y) = H(X,Y) - H(Y)
⊢ 0 ≤ H(X|Y) ≤ H(X)
⊢ I(X;Y) = H(X) + H(Y) - H(X,Y)
⊢ I(X;Y)  ≤ min(H(X),H(Y))
⊢ H(X)   = -SIGMA (λx. p{x} logr b (p{x})) s
⊢ H(X|Y) = -SIGMA (λ(x,y). p{(x,y)} logr b (p{(x,y)}/p{y})) s
```

3.4 Conditional Mutual Information

The conditional mutual information $I(X;Y|Z)$ allows one to measure how much information about the secret inputs X is leaked to the attacker by observing the outputs Y of a program given knowledge of the low security inputs Z. This property was used by Malacaria [12] to introduce the conditional mutual information as a measure of information flow for a program with high security inputs and low security inputs and outputs. The conditional mutual information is defined as the KL divergence between the joint distribution p_{XYZ} and the product measure $p_{X|Z}p_{Y|Z}p_Z$. The HOL formalization is as follows.

```
⊢ conditional_mutual_information b p X Y Z =
   KL_divergence b (joint_distribution p X Y Z)
                   (prod_measure (conditional_distribution p X Z)
                                 (conditional_distribution p Y Z)
                                 (distribution p Y))
```

We formally verify the following reduced form of the conditional mutual information for finite spaces by first proving that p_{XYZ} is absolutely continuous w.r.t $p_{X|Z}p_{Y|Z}p_Z$ and then apply the reduced form of the KL divergence.

$$I(X;Y|Z) = \sum_{(x,y,z)\in\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}} p(x,y,z)\log\frac{p(x,y,z)}{p(x|z)p(y|z)p(z)}$$

When the two random variables X and Y are independent given Z, the conditional mutual information $I(X;Y|Z) = 0$. In fact, in this case,
$\forall x,y,z.\ p(x,y,z) = p(x,y|z)p(z) = p(x|z)p(y|z)p(z)$.

```
⊢ indep_rv_cond p X Y Z ⇒ I(X;Y|Z) = 0
```

We also prove a few other important results regarding the conditional mutual information which will be useful later in our work.

```
⊢ 0 ≤ I(X;Y|Z)
⊢ I(X;Y|Z) = H(X|Z) - H(X|Y,Z)
⊢ I(X;Y|Z) = I(X;(Y,Z)) - I(X;Z)
⊢ I(X;Y|Z) ≤ H(X|Z)
```

So far, we have provided a higher-order-logic formalization of the KL divergence which we used to define various measures of quantitative information flow. This framework, along with the formalization of measure and probability theories, allows us to conduct many analyses of quantitative information flow using a theorem prover and hence guaranteeing the soundness of the analysis.

4 Degrees of Information Leakage

We introduce two new measures of information leakage that can be used to describe the anonymity properties of security systems and protocols, namely the information leakage degree and the conditional information leakage degree.

4.1 Information Leakage Degree

We define the information leakage degree between random variables X and Y representing the secret inputs and public outputs of a program, respectively, as

$$D = \frac{H(X|Y)}{H(X)}$$

```
⊢ information_leakage_degree p X Y =
        conditional_entropy p X Y / entropy p X
```

To better understand the intuition behind this definition, let us consider the two extreme cases of a completely secure program and a completely insecure program. Complete security, intuitively, happens when the knowledge of the public output Y of a program does not affect the uncertainty about the secret input X. This is equivalent to the requirement that X is independent of Y. In this case $H(X|Y) = H(X)$ and the information leakage degree is equal to 1. On the other hand, when the output of the program completely identifies its secret input, the entropy $H(X|Y)$ is equal to 0 and hence the information leakage degree is equal to 0 in this case of perfect identification. For situations between the two extremes, the information leakage degree lies in the interval $(0, 1)$. We formally verified this result as the following theorem in HOL which provides the bounds of the degree of information leakage.

```
⊢ 0 ≤ information_leakage_degree p X Y ≤ 1
```

Using the properties of the mutual information we prove that the information leakage degree is also equal to

$$D = 1 - \frac{I(X;Y)}{H(X)}$$

This result illustrates the significance of the information leakage degree definition since the mutual information measures how much information an adversary can learn about the input X after observing the output Y. This also allows to compare our definition to the anonymity degree proposed in [20] as

$$D' = 1 - \frac{I(X;Y)}{logN}$$

where N is the size of the alphabet of X. Our definition is more general. In fact, when X is uniformly distributed, the two measures coincide $D = D'$. However in the general case we believe that our definition is more accurate since in the perfect identification scenario, for instance, D is always equal to 1 regardless of the input distribution. On the other hand, D' is equal to 1 only in the special case of a uniform distribution. In [20] the authors considered using $H(X)$ as a normalization factor instead of $logN$ but opted for the latter arguing that the input distribution is already accounted for in the mutual information. But as stated previously, with the definition of D', the proof for perfect identification is only valid for uniformly distributed inputs.

4.2 Conditional Information Leakage Degree

We propose another variation of information leakage degree that is more general and can cover a wider range of scenarios. First, consider a program which has a set of high security inputs S, a set of low security inputs L and a set of public outputs O. The adversary wants to learn about the high inputs S by observing the outputs O given the knowledge of the low inputs L. To capture this added information to the adversary (low inputs), we propose the following definition, which we call the conditional information leakage degree.

$$D_c = \frac{H(S|(O,L))}{H(S|L)}$$

This can be formalized in HOL as

```
⊢ conditional_information_leakage_degree p S L O =
     conditional_entropy p S (O,L) / conditional_entropy p S L
```

Just like the previous case, consider the two extremes of perfect security and perfect identification. When the outputs and the secret inputs are independent, given L, the conditional entropy $H(S|(O,L))$ is equal to $H(S|L)$ which results in a conditional leakage degree equal to 1 for perfect security. However, if the public inputs and outputs completely identify the secret inputs, then $H(S|(O,L))$ is equal to 0 and so is the conditional leakage degree in the case of perfect identification. As in the case of leakage degree, we are also able to prove that the conditional information leakage degree lies in the interval $[0,1]$.

```
⊢ 0 ≤ conditional_information_leakage_degree p X Y Z ≤ 1
```

We also prove that the conditional information leakage degree can be written in terms of the conditional mutual information and the conditional entropy.

$$D = 1 - \frac{I(S;O|L)}{H(S|L)}$$

This shows that this definition is clearly a generalization of the information leakage degree for the case of programs with additional low security inputs. We provide more intuition to interpret this definition by proving the data processing inequality (DPI) [4].

Definition 1. *Random variables X, Y, Z are said to form a Markov chain is that order (denoted by $X \to Y \to Z$) if the conditional distribution of Z depends only on Y and is conditionally independent of X. Specifically, X, Y and Z form a Markov chain $X \to Y \to Z$ if the joint probability mass function can be written as $p(x,y,z) = p(x)p(y|x)p(z|y)$.*

We formalize this in HOL as follows.

```
⊢ markov_chain p X Y Z =
   ∀ x y z. joint_distribution p X Y Z {(x,y,z)} =
                 distribution p X {x} *
                 conditional_distribution p Y X {(y,x)} *
                 conditional_distribution p Z Y {(z,y)}
```

We prove that $X \to Y \to Z$ is equivalent to the statement that X and Z are conditionally independent given Y. In fact, $p(x)p(y|x)p(z|y) = p(x,y)p(z|y) = p(x|y)p(z|y)p(y)$. This in turn is equivalent to $I(X;Z|Y) = 0$. This result will allow us to prove the DPI theorem.

Theorem 2. *(DPI) if $X \to Y \to Z$ then $I(X;Z) \leq I(X;Y)$*

```
⊢ markov_chain p X Y Z ⇒
    mutual_information b p X Z ≤ mutual_information b p X Y
```

We prove the DPI theorem using the properties of the mutual information. In fact, as shown previously, $I(X;(Y,Z)) = I(X;Z) + I(X;Y|Z)$. By symmetry of the mutual information, we also have $I(X;(Y,Z)) = I(X;Y) + I(X;Z|Y) = I(X;Y)$. The last equality results from the fact that $I(X;Z|Y) = 0$ for a Markov Chain. Using the non-negativity of the conditional mutual information, which we proved previously, it is straightforward to conclude that $I(X;Z) \leq I(X;Y)$.

The data processing inequality is an important result in information theory that is used, for instance, in statistics to define the notion of sufficient statistic. We make use of the DPI to interpret the conditional information leakage degree. For a system with high security inputs S, low security inputs L and outputs O, if the outputs depend only on the low inputs, i.e., $p(O|S,L) = p(O|L)$ then $S \to L \to O$ and S and O are conditionally independent given L. This is the perfect security scenario, for which $D_c = 1$. Using the DPI, we conclude that $I(S;O) \leq I(S;L)$. This means that when the conditional mutual information leakage is equal to 1, no clever manipulation of the low inputs, by the attacker, deterministic or random, can increase the information that L contains about S ($I(S;L)$).

We have presented so far our higher-order-logic formalization of measures of information flow building upon the extension of measure, Lebesgue integration and probability formalization in HOL. Overall the HOL definitions and proof scripts of the above formalization required around 15,000 lines of code [15]. These results can now be readily used to reason about information flow analysis of real-world protocols and programs.

5 Application

In this section, we use our formalization to reason about an anonymity-based single MIX, designed to hide the communication links between a set of senders and a set of receivers. We model a single MIX as a communication node connecting m senders $(s_1, \ldots, s_m)$ to n receivers $(r_1, \ldots, r_n)$. The single MIX is determined by its inputs (senders), outputs (receivers) and the transition probabilities. We can also add clauses in the specification to capture additional information about the MIX like structural symmetry. The following is the formalization of the single MIX given in Figure 1.

```
⊢ MIX_channel s m X Y =
   (IMAGE X s = {0;1}) ∧ (IMAGE Y s = {0;1;2;3})  ∧
   (conditional_distribution (s,POW s,m) Y X {0} {0} = 1/2) ∧
   (conditional_distribution (s,POW s,m) Y X {1} {0} = 1/2) ∧
   (conditional_distribution (s,POW s,m) Y X {2} {1} = 1)
```

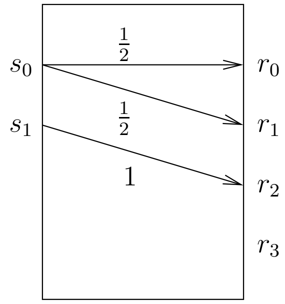

Fig. 1. Single MIX

Zhu and Bettati [20] used the single MIX to model an anonymity-based covert-channel where a sender is trying to covertly send messages through the MIX. They used the channel capacity as a measure of the maximum information that can be leaked through the MIX and can be used as a measure of the quality of anonymity of the network. A communication between a sender s_i and a receiver r_j is denoted by $[s_i, r_j]$. The term $p([s_u, r_v]_s | [s_i, r_j]_a)$ represents the probability that the communication $[s_u, r_v]$ is suspected given that $[s_i, r_j]$ is actually taking place. This model describes attacks on sender-receiver anonymity. The input symbols of the covert-channel are the actual sender-receiver pairs $[s, r]_a$ and the output symbols are the suspected pairs $[s, r]_s$. In this case, $p([s, r]_s | [s, r]_a)$ represents the result of the anonymity attack. We consider the case where an attacker can establish a covert-channel by having 1 sender s_1 communicate with any combination of j receivers. The same reasoning can be applied to multiple senders. The authors claim the following result [20]

Lemma 2. *For a single sender s_1 on a single mix, the maximum covert-channel capacity is achieved when s_1 can communicate to all receivers.*

We initially tried to formally verify this result, using the foundational results presented in the previous two sections of this paper, but we found a counter-example for an assumption upon which the paper-and-pencil proof of Lemma 2 is based [20]. The erroneous assumption states that the maximum of the mutual information is achieved when all input symbols have non-zero probabilities regardless of the transition probabilities (the results of the anonymity attack). We are able to prove in HOL that it is not necessary for the sender s_1 to communicate with all receivers to achieve capacity.

First, we provide a higher-logic-formalization of the channel capacity which is defined as the maximum, over all input distributions, of the mutual information between the input and the output of the channel. We formalize it in HOL using the Hilbert-choice operator; i.e., if it exists, the capacity is some c such that $c = I_m(X;Y)$ for some probability distribution m and for any input distribution p, $I_p(X;Y) \leq c$.

```
⊢ capacity s X Y = @c.
      ∃m. c = mutual_information (s, POW s, m) X Y  ∧
      ∀m. mutual_information (s, POW s, m) X Y ≤ c
```

Next, consider the covert-channel depicted in Figure 2. To simplify the notation, let $x_i = [s_1, r_i]_a$ and $y_i = [s_1, r_i]_s$. This covert-channel is formalized in HOL as

```
⊢ MIX_channel_1 s m X Y =
   (IMAGE X s = {0;1;2}) ∧ (IMAGE Y s = {0;1;2})  ∧
   (distribution(s,POW s,m) X{0} = distribution(s,POW s,m) X{2}) ∧
   (conditional_distribution (s,POW s,m) Y X {0} {0} = 1) ∧
   (conditional_distribution (s,POW s,m) Y X {0} {1} = 1 / 2) ∧
   (conditional_distribution (s,POW s,m) Y X {0} {2} = 0) ∧
   (conditional_distribution (s,POW s,m) Y X {1} {0} = 0) ∧
   (conditional_distribution (s,POW s,m) Y X {1} {1} = 0) ∧
   (conditional_distribution (s,POW s,m) Y X {1} {2} = 0) ∧
   (conditional_distribution (s,POW s,m) Y X {2} {0} = 0) ∧
   (conditional_distribution (s,POW s,m) Y X {2} {1} = 1 / 2) ∧
   (conditional_distribution (s,POW s,m) Y X {2} {2} = 1)
```

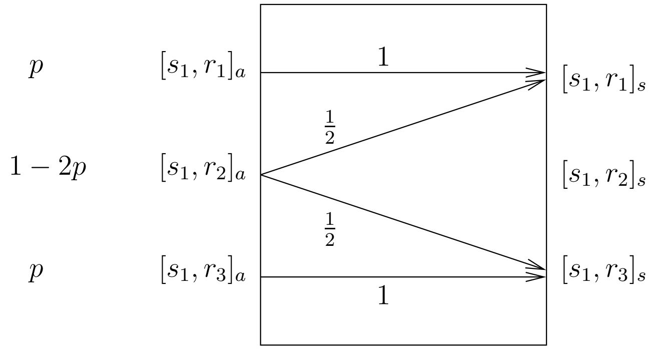

Fig. 2. Single MIX example

We prove that its mutual information is equal to $2p$.

```
⊢ ∀X Y s. MIX_channel_1 s m X Y ⇒
       mutual_information 2 (s, POW s, m) X Y =
          2 * distribution (s, POW s, m) X {0}
```

We also prove that the capacity is equal to 1 and corresponds to $p = \frac{1}{2}$. This means that the input distribution that achieves the channel capacity is $[p\{x_0\} = \frac{1}{2}, p\{x_1\} = 0, p\{x_2\} = \frac{1}{2}]$. Hence, we prove that the sender s_1 does not need to communicate with the receiver r_2 and still achieve maximum capacity, contradicting Lemma 2. Notice that with $p = \frac{1}{2}$, $I(X;Y) = H(X) = 1$ which implies that the degree of information leakage $D = 0$. So for this covert-channel, the maximum capacity corresponds to perfect identification.

Unlike the paper-and-pencil based analysis, a machine-assisted analysis of quantitative information flow using theorem proving guarantees the accuracy of the results. In fact, the soundness of theorem proving inherently ensures that only valid formulas are provable. The requirement that every single step of the proof needs to be derived from axioms or previous theorems using inference rules, allows us to find missing assumptions and even sometimes wrong statements as was the case in the single MIX application. We were able to detect the problem with the reasoning and confirm the result using our formalization in HOL.

6 Related Work

The underlying theories over which we built this work are mainly from [13] and [14]. In [13], we provided a formalization of the measure theory and Lebesgue integration in HOL and proved some classical probability results like the Weak Law of Large Numbers. In [14], we formalized extended reals and based on them provided a more extensive formalization of measure and Lebesgue integration. We also formalized the Shannon entropy and Relative entropy and proved the Asymptotic Equipartition Property. In the current paper, we enrich the underlying theories by adding, for instance, products of measure spaces and joint distributions. The main difference, however, is that in this paper we propose new measures of information leakage and formalize various other measures like mutual information and conditional mutual information based on a unified definition of the KL divergence. We also formalize the channel capacity and the notion of single MIX and use the framework for an illustrative example.

Coble [3] formalized some information theory in higher-order logic and used Malacaria's measure of information leakage, i.e., the conditional mutual information [12], to formally analyse the anonymity properties of the Dining Cryptographers protocol. Our formalization of information theory is an extended version of Coble's formalization, i.e., it supports Borel spaces and extended real numbers which allowed us to prove the Radon Nikodym theorem. Coble's formalization of information theory does not offer these capabilities and thus cannot be used to formally verify the Radon Nikodym theorem.

Zhu and Bettati [20] proposed the notion of degree of anonymity which is close to our definition of information leakage degree but we showed that our definition is more general and the two are equal in the case of uniform distribution. Besides, we proposed the conditional information leakage degree, suitable for programs with low security inputs and proved the data processing inequality to give more insight into the intuition behind this new definition. Moreover, our work is based on higher-order-logic theorem proving, which is arguably more sound than the paper-and-pencil based analysis of Zhu and Bettati. In fact, with our analysis we were able to detect the aforementioned problem with the analysis in [20] and provide a counter-example using theorem proving.

Chatzikokolakis [1] modeled anonymity protocols as noisy channels and used the channel capacity as a measure of the loss of anonymity. In the case where some leakage is intended by design, like in an election protocol, they introduced the notion of conditional capacity which is related to the conditional mutual information. They used the PRISM model checker [11] to assist in computing the transition probabilities and capacity of two protocols, namely the Dining cryptographers and the Crowds protocol. This probabilistic model checking based analysis technique inherits the state-space explosion limitation of model checking. Similarly, it cannot be used to verify universally quantified generic mathematical relationships like we have been able to verify in the reported work.

7 Conclusions

In this paper, we conducted the quantitative analysis of information flow within the sound core of higher-order-logic theorem prover. For this purpose, we provided a formalization of the Kullback-Liebler divergence in the HOL4 theorem prover and used it to formalize various measures of information leakage that have been proposed in the literature such as the entropy, mutual information and conditional mutual information. We proposed two novel measures of information leakage which we called information leakage degree and gave some insight into the intuition behind the definitions.

We also provided a higher-order-logic formalization of channel capacity and the single MIX and used our framework in a small example to show the usefulness of using a theorem prover in this context. In fact, we were able to come up with a counter-example to a result that appeared in [20] related to the single MIX and proved in HOL that the senders need not communicate with all receivers to achieve channel capacity. Our results have been confirmed by Prof. Gallager from MIT, a well-known name in Information Theory and the author of the book Information Theory and Reliable Communication [7]. Catching this significant problem in the paper-and-pencil proofs clearly indicates the usefulness of using higher-order-logic theorem proving for conducting information flow analysis.

Our future plans include using this framework and new measures of information leakage to study the security properties of various protocols in HOL like the Dining Cryptographers [2] and Crowds protocols [16].

References

1. K. Chatzikokolakis, C. Palamidessi, and P. Panangaden. Anonymity Protocols as

Noisy Channels. In *Trustworthy Global Computing*, volume 4661 of *LNCS*, pages 281–300. Springer, 2007.

2. D. Chaum. The Dining Cryptographers Problem: Unconditional Sender and Recipient Untraceability. *Journal of Cryptology*, 1(1):65–75, 1988.
3. A. R. Coble. Formalized Information-Theoretic Proofs of Privacy using the HOL4 Theorem-Prover. In *Privacy Enhancing Technologies*, volume 5134 of *LNCS*, pages 77–98. Springer, 2008.
4. T. M. Cover and J. A. Thomas. *Elements of Information Theory.* Wiley-Interscience, 1991.
5. Y. Deng, J. Pang, and P. Wu. Measuring Anonymity with Relative Entropy. In *Formal Aspects in Security and Trust*, volume 4691 of *LNCS*, pages 65–79. Springer, 2007.
6. C. Diaz, S. Seys, J. Claessens, and B. Preneel. Towards Measuring Anonymity. In *Privacy Enhancing Technologies*, volume 2482 of *LNCS*, pages 54–68. Springer, 2003.
7. Robert G. Gallager. *Information Theory and Reliable Communication.* John Wiley & Sons, Inc., 1968.
8. R. R. Goldberg. *Methods of Real Analysis.* Wiley, 1976.
9. M. J. C. Gordon. Mechanizing Programming Logics in Higher-Order Logic. In *Current Trends in Hardware Verification and Automated Theorem Proving*, pages 387–439. Springer, 1989.
10. M. J. C. Gordon and T. F. Melham. *Introduction to HOL: A Theorem Proving Environment for Higher-Order Logic.* Cambridge University Press, 1993.
11. M. Kwiatkowska, G. Norman, and D. Parker. Quantitative Analysis with the Probabilistic Model Checker PRISM. *Electronic Notes in Theoretical Computer Science*, 153(2):5–31, 2005.
12. P. Malacaria. Assessing Security Threats of Looping Constructs. *SIGPLAN Notes*, 42(1):225–235, 2007.
13. T. Mhamdi, O. Hasan, and S. Tahar. On the Formalization of the Lebesgue Integration Theory in HOL. In *Interactive Theorem Proving*, volume 6172 of *LNCS*, pages 387–402. Springer, 2010.
14. T. Mhamdi, O. Hasan, and S. Tahar. Formalization of Entropy Measures in HOL. In *Interactive Theorem Proving*, volume 6898 of *LNCS*, pages 233–248. Springer, 2011.
15. T. Mhamdi, O. Hasan, and S. Tahar. Quantitative Information Flow Analysis in HOL. http://hvg.ece.concordia.ca/code/hol/information-flow/, 2012.
16. M. K. Reiter and A. D. Rubin. Crowds: Anonymity for Web Transactions. *ACM Transactions on Information and System Security*, 1(1):66–92, 1998.
17. Andrei Sabelfeld and Andrew C. Myers. Language-Based Information-Flow Security. *IEEE Journal on Selected Areas in Communications*, 21(1):5–19, 2003.
18. A. Serjantov and G. Danezis. Towards an Information Theoretic Metric for Anonymity. In *Privacy Enhancing Technologies*, volume 2482 of *LNCS*, pages 259–263. Springer, 2003.
19. G. Smith. On the Foundations of Quantitative Information Flow. In *Foundations of Software Science and Computational Structures*, volume 5504 of *LNCS*, pages 288–302. Springer, 2009.
20. Ye Zhu and Riccardo Bettati. Information Leakage as a Model for Quality of Anonymity Networks. *IEEE Transactions on Parallel and Distributed Systems*, 20(4):540–552, 2009.