Refining a Quantitative Information Flow Metric

Sari Haj Hussein\(^1\)

\(^1\)Department of Computer Science
Aalborg University

2012-01-13
1 Introduction

2 Problem

3 Size-consistent QIF Quantifier

4 Accuracy-based Information Flow Analysis

5 Refining The Divergence

6 Refining The Metric
1 Introduction

2 Problem

3 Size-consistent QIF Quantifier

4 Accuracy-based Information Flow Analysis

5 Refining The Divergence

6 Refining The Metric
Information Flow Analysis

- **Information flow analysis** aims at keeping track of a program’s secret input during the execution of that program.
Information Flow Analysis Techniques

- **Qualitative techniques.** prohibit flow from a program’s secret input to its public output
  - Expensive or rarely satisfied by real programs
  - No distinction between acceptable and unacceptable flows
  - Conceptual and boring

- **Quantitative techniques.** establish limits on the number of bits that might be revealed from a program’s secret input
  - Mainly based on information theory
  - More tangible

**Literature Observation**

Much work on qualitative, less on quantitative
Information Flow Analysis Techniques

- **Qualitative techniques.** prohibit flow from a program’s secret input to its public output
  - Expensive or rarely satisfied by real programs
  - No distinction between acceptable and unacceptable flows
  - Conceptual and boring

- **Quantitative techniques.** establish limits on the number of *bits* that might be revealed from a program’s secret input
  - Mainly based on information theory
  - More tangible

**Literature Observation**

Much work on qualitative, less on quantitative
Information Flow Analysis Techniques

- **Qualitative techniques.** prohibit flow from a program’s secret input to its public output
  - Expensive or rarely satisfied by real programs
  - No distinction between acceptable and unacceptable flows
  - Conceptual and boring

- **Quantitative techniques.** establish limits on the number of *bits* that might be revealed from a program’s secret input
  - Mainly based on information theory
  - More tangible

**Literature Observation**

Much work on qualitative, less on quantitative
Problem Description

- The quantitative metric by Clarkson et al.
- It is the first to address attacker’s belief in quantifying information flow
- This metric reports counter-intuitive flow quantities that are inconsistent with the size of a program’s secret input.
Problem Impact

- We cannot determine the space of the exhaustive search that should be carried out in order to reveal the residual part of a program’s secret input.
Informal Reasoning

- There is a **flaw** in the design of the metric
- We need to spot the source of that flaw
- Then we need to fix it!
1. Introduction

2. Problem

3. Size-consistent QIF Quantifier

4. Accuracy-based Information Flow Analysis

5. Refining The Divergence

6. Refining The Metric
Uncertainty-based Information Flow Analysis

- $U$ attacker’s pre-uncertainty
- $U'$ attacker’s post-uncertainty
- Flow = reduction in uncertainty
- $R = U - U'$
- $R \leq 0 \Rightarrow$ increase in uncertainty $\Rightarrow$ absence of flow
- $R > 0 \Rightarrow$ decrease in uncertainty $\Rightarrow$ we have flow
- Notice that $R$ ignores reality by measuring $U$ and $U'$ against each other, instead of against reality
If attacker’s belief is captured using a probability distribution, uncertainty is computed using Shannon uncertainty functional.

**Shannon Uncertainty Functional**

- $X$ a discrete random variable with alphabet $\mathcal{X}$
- $p$ a probability distribution function on $X$
- $S(p) = - \sum_{x \in \mathcal{X}} p(x) \log p(x)$

The range of $S$ is $[0, \log |\mathcal{X}|]$ $\Rightarrow \varrho_R = [-\log |\mathcal{X}|, \log |\mathcal{X}|]$.

This is plausible since $\log |\mathcal{X}|$ is the size of a program’s secret input.
Plausible Range

- If attacker’s belief is captured using a probability distribution, uncertainty is computed using Shannon uncertainty functional.

### Shannon Uncertainty Functional

- $X$ a discrete random variable with alphabet $\mathcal{X}$
- $p$ a probability distribution function on $X$
- $S(p) = - \sum_{x \in \mathcal{X}} p(x) \log p(x)$

- The range of $S$ is $[0, \log |\mathcal{X}|] \Rightarrow \varrho_{R} = [- \log |\mathcal{X}|, \log |\mathcal{X}|]$
- This is plausible since $\log |\mathcal{X}|$ is the size of a program’s secret input.
If attacker’s belief is captured using a probability distribution, uncertainty is computed using Shannon uncertainty functional.

**Shannon Uncertainty Functional**

- $X$ a discrete random variable with alphabet $\mathcal{X}$
- $p$ a probability distribution function on $X$
- $S(p) = - \sum_{x \in \mathcal{X}} p(x) \log p(x)$

The range of $S$ is $[0, \log |\mathcal{X}|] \Rightarrow \varrho_R = [- \log |\mathcal{X}|, \log |\mathcal{X}|]$.

This is plausible since $\log |\mathcal{X}|$ is the size of a program’s secret input.
Size-consistent QIF Quantifier

- $QUAN$ a QIF quantifier
- $\eta$ the size of a program’s secret input
- $QUAN$ is size-consistent if
  $QUAN_{max} \leq \eta$ and $QUAN_{min} \geq -\eta$
1. Introduction
2. Problem
3. Size-consistent QIF Quantifier
4. Accuracy-based Information Flow Analysis
5. Refining The Divergence
6. Refining The Metric
Clarkson Observation

- **PWC**: if $p = g$ then $a := 1$ else $a := 0$
- Password space is $W_p = \{A, B, C\} \Rightarrow$ password size is $\log |W_p| = \log 3 = 1.5849$ bits
- The correct password (the reality) is $C$
- Attacker’s prebelief $b_H = [(A : 0.98), (B : 0.01), (C : 0.01)]$
- Attacker (naturally) feeds PWC with $g = A$ and gets $a = 0$
- Attacker’s postbelief $b'_H = [(A : 0), (B : 0.5), (C : 0.5)]$
- $R = -0.8386$ bits $\Rightarrow$ absence of flow
- But $b'_H$ is nearer to reality than $b_H \Rightarrow$ attacker has learnt something $\Rightarrow$ we have flow
Uncertainty-based analysis is **inadequate** if input distributions represent attacker’s beliefs.
Accuracy-based Information Flow Analysis

- **Respect** reality by measuring $b_H$ and $b'_H$ against it, instead of against each other only.
- **Reality** is denoted as $\sigma_H$ (password is $C$).
- **Certainty** about reality is then $\dot{\sigma}_H$ (password is $C$ with a probability of 1).
- Accuracy of $b_H = D(b_H \rightarrow \dot{\sigma}_H)$
- Accuracy of $b'_H = D(b'_H \rightarrow \dot{\sigma}_H)$
- Flow = **improvement** in accuracy
- Clarkson metric $Q = D(b_H \rightarrow \dot{\sigma}_H) - D(b'_H \rightarrow \dot{\sigma}_H)$
Clarkson Choice of $D$

- Clarkson chose Kullback-Leibler divergence
- $D(b \rightarrow b') = \sum_{\sigma \in \mathcal{W}_p} b'(\sigma) \cdot \log \frac{b'(\sigma)}{b(\sigma)}$
- $Q = D(b_H \rightarrow \dot{\sigma}_H) - D(b'_H \rightarrow \dot{\sigma}_H)$
- $Q = \sum_{\sigma \in \mathcal{W}_p} \dot{\sigma}_H(\sigma) \cdot \log \frac{\dot{\sigma}_H(\sigma)}{b_H(\sigma)} - \sum_{\sigma \in \mathcal{W}_p} \dot{\sigma}_H(\sigma) \cdot \log \frac{\dot{\sigma}_H(\sigma)}{b'_H(\sigma)}$
- $Q = -\log b_H(\sigma_H) + \log b'_H(\sigma_H)$
Puzzling Result

- \( Q = -\log 0.01 + \log 0.5 = 6.6438 - 1 = 5.6438 \text{ bits} \)
- But the plausible range is
  \[ q_R = [-\log 3, -\log 3] = [-1.5849, 1.5849] \]
- \( Q \) is not a size-consistent QIF quantifier
Clarkson Argument

- $b_H$ is more erroneous than a uniform belief ascribing $1/3$ probability to each password $A$, $B$, and $C$
- Therefore a larger amount of information is required to correct $b_H$
- If $b_H$ is uniform, the attacker would learn a total of log 3 bits
Our Arguments

- We have shown that Clarkson argument is valid for deterministic programs, but incomplete for probabilistic ones.
- We have further shown that the range of $Q$ is $\varrho_Q = (-\infty, -\log b_H(\sigma_H)]$. 
Replacing the Construct

Original Construct

\[ I_{Dis}(\sigma) = \log \frac{b'(\sigma)}{b(\sigma)} \]

Proposed Construct

\[ I'_{Dis}(\sigma) = \log \frac{b'(\sigma)}{\frac{b'(\sigma)}{2} + b(\sigma)} \]

Replacement Effect

\[ I'_{Dis}(\sigma) \leq \frac{1}{2} I_{Dis}(\sigma) \]
Replacing the Construct

<table>
<thead>
<tr>
<th>Original Construct</th>
<th>Proposed Construct</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathcal{I}_{Dis}(\sigma) = \log \frac{b'(\sigma)}{b(\sigma)}$</td>
<td>$\mathcal{I}'_{Dis}(\sigma) = \log \frac{b'(\sigma)}{\frac{b'(\sigma)}{2} + b(\sigma)}$</td>
</tr>
</tbody>
</table>

Replacement Effect

$\mathcal{I}'_{Dis}(\sigma) \leq \frac{1}{2} \mathcal{I}_{Dis}(\sigma)$
## Replacing the Construct

<table>
<thead>
<tr>
<th><strong>Original Construct</strong></th>
<th><strong>Proposed Construct</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>[ \mathcal{I}_{Dis}(\sigma) = \log \frac{b'(\sigma)}{b(\sigma)} ]</td>
<td>[ \mathcal{I}'_{Dis}(\sigma) = \log \frac{b'(\sigma)}{\frac{b'(\sigma)}{2} + b(\sigma)} ]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Replacement Effect</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>[ \mathcal{I}'<em>{Dis}(\sigma) \leq \frac{1}{2} \mathcal{I}</em>{Dis}(\sigma) ]</td>
</tr>
</tbody>
</table>
Replacing the Divergence

**Original Divergence**

- $D(b \rightarrow b') = \sum_{\sigma \in \mathcal{W}_p} b'(\sigma) \cdot \log \frac{b'(\sigma)}{b(\sigma)}$
- Average number of bits that are wasted by encoding events from a distribution $b'$ with a code based on a not-quite-right distribution $b$
- Information gain

**Proposed Divergence**

- $D'(b \rightarrow b') = \sum_{\sigma \in \mathcal{W}_p} b'(\sigma) \cdot \log \frac{b'(\sigma)}{\frac{b'(\sigma) + b(\sigma)}{2}}$
- How much information is lost if we describe the two random variables that correspond to $b$ and $b'$ with their average distribution $(b' + b)/2$?
- Information radius
Replacing the Divergence

Original Divergence

\[ D(b \to b') = \sum_{\sigma \in \mathcal{W}_p} b' (\sigma) \cdot \log \frac{b'(\sigma)}{b(\sigma)} \]

- Average number of bits that are wasted by encoding events from a distribution \( b' \) with a code based on a not-quite-right distribution \( b \)
- Information gain

Proposed Divergence

\[ D'(b \to b') = \sum_{\sigma \in \mathcal{W}_p} b' (\sigma) \cdot \log \frac{b'(\sigma)}{b'(\sigma)+b(\sigma)} \]

- How much information is lost if we describe the two random variables that correspond to \( b \) and \( b' \) with their average distribution \((b' + b)/2\)?)
- Information radius
Plot

The Kullback-Leibler divergence plotted against $t \in [0,1]$, $b(\sigma) = t$, and $b'(\sigma) = 1 - b(\sigma)$

The proposed divergence plotted against $t \in [0,1]$, $b(\sigma) = t$, and $b'(\sigma) = 1 - b(\sigma)$
Refining to Normalization

**Normalized Metric**

- \( Q' = D'(b_H \rightarrow \dot{\sigma}_H) - D'(b'_H \rightarrow \dot{\sigma}_H) \)
- \( Q' = \sum_{\sigma \in \mathcal{W}_p} \dot{\sigma}_H(\sigma) \cdot \log \left( \frac{\dot{\sigma}_H(\sigma)}{\dot{\sigma}_H(\sigma) + b_H(\sigma)} \right) - \sum_{\sigma \in \mathcal{W}_p} \dot{\sigma}_H(\sigma) \cdot \log \left( \frac{\dot{\sigma}_H(\sigma)}{\dot{\sigma}_H(\sigma) + b'_H(\sigma)} \right) \)
- \( Q' = \log \left( \frac{2}{1 + b_H(\sigma_H)} \right) - \log \left( \frac{2}{1 + b'_H(\sigma_H)} \right) \)
- \( Q' = -\log(1 + b_H(\sigma_H)) + \log(1 + b'_H(\sigma_H)) \)

- We have shown that the range of \( Q' \) is \( \varrho_{Q'} = [-1, 1] \)
- This does **not** make \( Q' \) size-consistent
- Nonetheless, \( \varrho_{Q'} \) is a plausible normalization (flow percentage) that is invariant with respect to the choice of the measurement unit
Refining to Normalization

Normalized Metric

\[ Q' = D'(b_H \rightarrow \dot{\sigma}_H) - D'(b'_H \rightarrow \dot{\sigma}_H) \]
\[ Q' = \sum_{\sigma \in \mathcal{W}_p} \dot{\sigma}_H(\sigma) \cdot \log \frac{\dot{\sigma}_H(\sigma)}{\dot{\sigma}_H(\sigma) + b_H(\sigma)} - \sum_{\sigma \in \mathcal{W}_p} \dot{\sigma}_H(\sigma) \cdot \log \frac{\dot{\sigma}_H(\sigma)}{\dot{\sigma}_H(\sigma) + b'_H(\sigma)} \]
\[ Q' = \log \frac{2}{1 + b_H(\sigma_H)} - \log \frac{2}{1 + b'_H(\sigma_H)} \]
\[ Q' = -\log(1 + b_H(\sigma_H)) + \log(1 + b'_H(\sigma_H)) \]

- We have shown that the range of \( Q' \) is \( \varrho_{Q'} = [-1, 1] \)
- This does not make \( Q' \) size-consistent
- Nonetheless, \( \varrho_{Q'} \) is a plausible normalization (flow percentage) that is invariant with respect to the choice of the measurement unit
We want bit as the measurement unit
Let $\eta$ be the size of a program’s secret input in bits
$Q'' = \eta \cdot Q' = \eta \cdot [\log(1 + b_H(\sigma_H)) + \log(1 + b'_H(\sigma_H))]$

We have shown that the range of $Q''$ is
$q_{Q''} = [-\eta \cdot \log(1 + b_H(\sigma_H)), \eta \cdot [1 - \log(1 + b_H(\sigma_H))]]$

$log(1 + b_H(\sigma_H)) \leq 1 \Rightarrow Q''_{max} \leq \eta$ and $Q''_{min} \geq -\eta \Rightarrow Q''$ is size-consistent
Refining to Actuality

**Actual Metric**

- We want bit as the measurement unit
- Let $\eta$ be the size of a program’s secret input in bits
- $Q'' = \eta. Q' = \eta. [-\log(1 + b_H(\sigma_H)) + \log(1 + b'_H(\sigma_H))]$

We have shown that the range of $Q''$ is

$q_{Q''} = [-\eta. \log(1 + b_H(\sigma_H)), \eta. [1 - \log(1 + b_H(\sigma_H))]]$

$\log(1 + b_H(\sigma_H)) \leq 1 \Rightarrow Q''_{max} \leq \eta$ and $Q''_{min} \geq -\eta \Rightarrow Q''$ is size-consistent
What does it mean to leak $k$ bits according to $Q''$?

$Q'' = k$

$\eta \left[- \log(1 + b_H(\sigma_H)) + \log(1 + b'_H(\sigma_H))\right] = k$

$\frac{\log(1+b'_H(\sigma_H))}{\log(1+b_H(\sigma_H))} = \frac{k}{\eta}$

$\frac{1+b'_H(\sigma_H)}{1+b_H(\sigma_H)} = 2^{k/\eta}$

$b'_H(\sigma_H) = 2^{k/\eta} \cdot b_H(\sigma_H) + 2^{k/\eta} - 1$

This corresponds to the increase in the likelihood of the attacker's correct guess.
Meaningfulness of the Bounds

- An **informing** flow equal to the upper bound of $Q''$ is sufficient to make a fully uncertain attacker fully certain about the correct high state.

$$b_H(\sigma_H) = 0 \rightarrow Q''_{\text{max}} = \eta \cdot [1 - \log(1 + b_H(\sigma_H))] \rightarrow b'_H(\sigma_H) = 1$$

- A **misinforming** flow equal to the lower bound of $Q''$ is sufficient to make a fully certain attacker fully uncertain about the correct high state.

$$b_H(\sigma_H) = 1 \rightarrow Q''_{\text{min}} = -\eta \cdot \log(1 + b_H(\sigma_H)) \rightarrow b'_H(\sigma_H) = 0$$
Exhaustive Search Effort

- Assuming a program with a secret input of size $\eta$ bits.
- Assuming an informing flow of $k$ bits to an attacker
- $Q''_{\text{max}} = \eta \cdot [1 - \log(1 + b_H(\sigma_H))]$ tells us that $k \leq \eta$
- The space of the exhaustive search is $2^{\eta-k}$
- $Q_{\text{max}} = -\log b_H(\sigma_H)$ tells us that $k > \eta$ is possible
- The exhaustive search space cannot be established, albeit that the secret input might have been partially revealed to the attacker
We presented a refinement of a QIF metric that bounds its reported results by a plausible range.

The results reported by the refined metric are easily associated with the exhaustive search effort.

We believe that the same can be done with other QIF quantifiers.
Thank You!