CS6113 NOTES ON QUANTITATIVE INFORMATION FLOW

LECTURER: ANDREW MYERS – SCRIBES: SAM HOPKINS, JED LIU

We read [3] and [2].

1. Introduction

First proposed by Denning and Gray in the early 1990s, the basic goal of a theory of quantitative information flow is to measure the information leaked by a program to an attacker by magnitude of the mutual information between some true distribution on secret data and the attacker's best guess at the distribution of that data. Early models of quantitative information flow directly used the Shannon mutual information $I(X;Y)$. Roughly, $I(X;Y)$ measures the (expected) reduction in uncertainty about the value of random variable X given by learning the value of Y (formal definitions follow).

Doing information flow quantitatively immediately raises the following objections, among others:

- Naïvely, $I(X;Y)$ seems to value all secret data equally. However, different pieces of high-security data might be worth more than others—not all leaked bits are equal. This problem is beginning to be addressed in work of Scedrov, but is still at least a partially-open question.
- It's not at all clear how to analyze code to measure its information leakage quantitatively. There is a POPL '07 paper on this problem, as well as a body of work on sampling executions of randomized programs and programs with randomized input data, as well as a POPL '07 paper on this problem.
- Dealing with nondeterminism in the program may pose a problem to a theory of quantitative information flow. Orthogonally, using Shannon mutual information directly does not appear to result in a theory that allows compositional reasoning about program executions: if k bits are leaked in the first execution of a program and k' in the second, we would like our theory to say that together the executions leak $k + k'$ bits, but that need not be true if we work directly with mutual information. The work of Clarkson et al. in CSF '05 addresses both of these problems simultaneously [2].
- Geoffry Smith claims that Shannon entropy is the wrong measure, because it fails to accurately capture what he calls the *vulnerability* of a program, which is the probability that attacker's first guess at the high data is correct. This quantity is captured by *min-entropy*. The work is discussed further below.

2. Information Theory – Definitions

Definition 1. *Let X be a discrete random variable over a probability space $\mathcal{X}$. The Shannon entropy of X, written $H(X)$, is*

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x) = E_{x \sim X}[-\log p(x)].$$

Date: November 12, 2013.

$H(X)$ measures "surprisal" the average amount that one should be surprised by seeing the value of X. The definition originates in coding theory from the 1940s.

Definition 2. *The conditional entropy of X given Y, written $H(X|Y)$, is $E_{y\sim Y}H(X|Y=y)$, where $H(X|Y=y)$ is the entropy of the conditional distribution $p(x|Y=y)$.*

Definition 3. *The mutual information $I(X;Y)$ between variables X, Y is the entropy reduction in X given by learning Y. Equivalent definitions are*

$$I(X;Y) = H(X) + H(Y) - H(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X).$$

Mutual information is symmetric and nonnegative. Conditional mutual information is defined analogously to conditional entropy.

3. Geoffrey Smith: Vulnerability and Min-Entropy

4. Clarkson et al: Beliefs

5. Summary of Discussion

The discussion seemed to split into three questions.

- There seem to be lots of proposed models of quantitative information flow, all of which have nontrivial differences. Is there a taxonomy of such models? Is there a list of desiderata for them? I.e., is there a list of theorems that should be provable about a model of quantitative information flow in order for it to qualify as such? The answer seems to be: no, but that would be pretty nice. Maybe some theorem that relates the leakage measured by the model to some notion in formulated in the language of differential privacy?
- Is there a connection between these leakage measures and differential privacy? Given that much of the privacy community seems to have settled on differential privacy as the right notion, this seems like an important question to address. One paper which seems to address this issue is [1].
- There are often natural coding interpretations of information-theoretic quantities which are enlightening to think about when using an entropic measure in the wild. How can we interpret the use of relative entropy in the Clarkson et al. paper? Since relative entropy measures efficiency losses when coding a message according to the wrong distribution, one proposal is to interpret it as follows: suppose we have an inside attacker who learns the high-security output but believes it to be distributed differently than it really is, and suppose that this attacker must exfiltrate the data on some bounded communication channel. Then the relative entropy between the actual distribution of the high data and the attacker's guess at the distribution should somehow measure how successfully the attacker can exfiltrate the data on his bounded channel.

References

[1] Gilles Barthe and Boris Kopf. Information-theoretic bounds for differentially private mechanisms. In *Computer Security Foundations Symposium (CSF), 2011 IEEE 24th*, pages 191–204. IEEE, 2011.

[2] Michael R Clarkson, Andrew C Myers, and Fred B Schneider. Quantifying information flow with beliefs. *Journal of Computer Security*, 17(5):655–701, 2009.

[3] Geoffrey Smith. On the foundations of quantitative information flow.