UFMG/ICEx/DCC

Course summary: *Quantitative Information Flow* - 2021/01

The course

- **Title:** Quantitative Information Flow (DCC 030/049/831)
- **Level:** Graduate and Advanced undergraduate
- **Credits:** 04 (60 hours)
- **Lecturer:** Mário S. Alvim
- **Date and time:** Tuesdays and Thursdays, 14:55-16:35

General information

- **Objectives:** *Quantitative information flow (QIF)* is the area of research that aims to understand fundamentally how sensitive input information (e.g., personal data) "flows" or "leaks" as it is processed by an authorized entity (e.g., a computer program), and to ensure that those flows are acceptable to us in terms of the quantified damage they might cause. Historically, the development of QIF has been motivated primarily by concerns about security. However, information flow is a general phenomenon with broader relevance, and QIF can undoubtedly be applied fruitfully in diverse contexts such as machine learning, recommendation systems, and robotics. (Interestingly, in those contexts information flow would typically be seen as a good thing.) For this reason, students outside the field of security may also profit from taking this course.

  In this course we will cover the fundamentals of quantitative information flow (e.g., mathematical models of knowledge, uncertainty, and information; as well as models of information channels, information flow, and robustness), and examples of applications of the theory to security problems.

  We will tackle questions such as:

  a) Distance is measured in meters, time is measured in seconds, mass is measured in kilograms. What is "secrecy", and in what units should we measure it?
  b) Is there a distinction between "knowledge" and "information", and, if so, why does it matter?
  c) How information-theoretic channels can model computational systems as knowledge-transforming devices?
  d) How to measure how much information a computational system may leak?
  e) There are many ways to measure "information"; is there a unique framework that encompasses all "reasonable" measures?
  f) How can this whole theory be applied to security and privacy?

- **Pre-requisites:** Good knowledge about discrete mathematics (e.g., sets, functions, basic proof techniques) and discrete probability (e.g., random variables, joint, marginal and conditional probabilities, expected value) is assumed.
- **Language:** This course will be taught in English.
- **Bibliography:** *The Science of Quantitative Information Flow.* Mário. S. Alvim, Konstantinos Chatzikokolakis, Annabelle McIver, Carroll Morgan, Catuscia Palamidessi, Geoffrey Smith. Springer International Publishing (2020).
- **Grading:**
  - Exams (50%);
  - Seminar (30%);
  - Problem sets, homework, in-class participation (20%).

Syllabus

1. **Introduction:** *An overview of quantitative information flow.*
2. **Modeling secrets:** What is a "secret" and how can we measure "secrecy"?
   (a) Secrets and probability distributions.
   (b) Shannon entropy.
   (c) Bayes vulnerability.
   (d) A more general view.
3. **On g-vulnerability:** *A general framework to quantify "information".*
   (a) Basic definitions.
   (b) A catalog of gain functions.
   (c) Classes of gain functions.
   (d) Mathematical properties.
   (e) On "absolute" versus "relative" security.
4. **Channels:** *Models for how a system updates "knowledge".*
   (a) Channel matrices.
   (b) The effect of a channel on the adversary's knowledge.
   (c) From joint distributions to hyper-distributions.
   (d) Abstract channels.
   (e) A first look on channel compositions.
5. **Posterior vulnerability and leakage:** *Measuring how much information is leaked by a channel.*
   (a) Posterior g-vulnerability and its basic properties.
   (b) Multiplicative and additive g-leakage.
   (c) A closer look at posterior Bayes vulnerability and Bayes leakage.
   (d) Measuring leakage using Shannon entropy.
   (e) More properties of posterior g-vulnerability and g-leakage.
   (f) Example channels and their leakage.
   (g) Max-case posterior g-vulnerability.
6. **Robustness:** *Making sure quantitative analyses of leakage hold up in real life.*
   (a) The need for robustness.
   (b) Approaches to robustness (capacity, composition of channels, refinement, the Dalenius perspective).
7. **Axiomatics:** *How do we know we have the "right" measures of information?*
   (a) An axiomatic view of vulnerability.
   (b) Axiomatization of prior vulnerabilities.
   (c) Axiomatization of posterior vulnerabilities.
   (d) Applications of axiomatization to understanding leakage measures.
8. **The geometry of hypers, gains and losses:** *Visualizing the mathematical properties of QIF.*
   (a) Barycentric representation of gain/loss functions.
   (b) Barycentric representation of hypers and their refinement.
   (c) Primitive hyper-distributions.
9. **Applications:** *Using QIF to analyze real systems.*
   (a) Examples of applications to security (such as the Crowds protocol; and differential privacy).