T-79.514 Special Course on Cryptology

Seminar 1: Introduction

Helger Lipmaa
Helsinki University of Technology
http://www.tcs.hut.fi/~helger

Overview of Seminar

- Introduction to the area
- Practicalities

Introduction to the Area: Buzzwords

Thanks to www.googlism.com!

- *Datamining* is an automated process for discovering information in large data sets to be used in decision, datamining is alive and well on the internet, datamining is all about counting, datamining is perfectly legal, *datamining is using a database to gain more information about your business*

- *Cryptography* is related with the communication or computation involving two or more parties who may not trust one another, cryptography is the most powerful single tool that users can use to secure the internet, cryptography is outlawed, *cryptography is the art of hiding the meaning of information*

Introduction to the Area: Huh?

(by a cryptographer)

- *Datamining*: build models of natural data from available (huge) datasets, without having to store them or having access to all them. Includes ability to predict, classify, cluster, ...
- In many applications, data is owned by a participant (a person or a company, Alice) who might be not willing to reveal it to the evil data-miner, Bob
- If Alice suspects that Bob is misusing her information, she will hide information or lie
- Result: Bob cannot obtain truthful models

Solutions

(by a cryptographer, mimicking a data-miner)

- Assume Bob is honest! — Really?!
- Assume there is a trusted third party! — Really?!
- Give up, and go to a pub — Why not, but could we. . .

  introduce cryptography?

Cryptography: In Its Modern Meaning

- Main result for us: all efficiently computable functions can also computed securely
- Assume there are n participants, and the ith participant has input x_i. Assume f is a function $f(x_1, \ldots, x_n) = (y_1, \ldots, y_n)$.
- There is a way (*multi-party computation*) to compute f so that at the end of the protocol, the ith participant will get the know value of y_i and nothing else, except what she could compute from (x_i, y_i) herself.

Cryptography: Drawbacks

- MPC requires that a majority (often $2/3$ of the participants) are honest
- If there are only two participants, one can guarantee privacy but not fairness: that is, one of the participants may disconnect after obtaining his private output
- The general solutions are not very efficient

Solutions so Far

- Avoid cryptography at any cost — lightweight solutions that depend on statistical truths
- Use toy cryptography — “our scheme is secure against an underpaid cryptographer”
- Devise specific secure cryptographic schemes for concrete problems. For some problems, the protocols are efficient.

Practicalities: Goals

- Goal (minimum): to study existing solutions
- Goal (medium): to break weak solutions, to improve upon them
- Goal (maximum): conquer the world (start a research program on the PPDM)

Practicalities: Form

- Seminars, every week one or two presentations. (+ written survey)
- Every presentation is reviewed by another student. All goes to the Internet (for the maximum impact)
- Ideal: every survey done by two students, one data-miner and one cryptographer
- If not enough students, we will have one student doing every survey, but the reviewer will represent another community

Practicalities: Form

- One survey + presentation + review is ≈ 1.5 credits
- Possible to do more than one!
- Surveys must be returned a few days before the presentation, so that the reviewer can read and comment on it before the presentation

Practicalities: Next Lecture

- Ella Bingham: data-mining for dummies
- Helger Lipmaa: cryptography for smarties

Questions?

?