Full statistical analyses with secure multi-party computation

Dan Bogdanov, Liina Kamm, Ville Sokk

dan@cyber.ee
http://sharemind.cyber.ee/
The Sharemind model

**Input parties**

- $IP_1$
- $IP_k$

**Computing parties**

- $CP_1$
- $CP_2$
- $CP_3$

**Result parties**

- $RP_1$

---

**Step 1:** secret sharing and storage of inputs

**Step 2:** secure multi-party computation

**Step 3:** reconstruction of results
Secret sharing (simplified)

75 - 53 - 38 = 84 \mod 100

Reconstruction: 53 + 38 + 84 = 75 \mod 100
MPC from secret sharing

Inputs:
P1: X1
P2: X2
P3: X3

Computation:
(y1, y2, y3) = f(x1, x2, x3)

Outputs:
y1
y2
y3

All operations are composable.
Strengths / weaknesses

- **Easy to write code for.** Developers apply privacy patterns on classical algorithms.
- **Hybrid execution model** for balancing public and private computations.
- **Very high performance** for arithmetic circuits.
- **Small storage overhead** (3 times for 3 servers).
- **Requires three servers for best possible efficiency** (works with 2 to n servers as well).
- **Performance profile** not immediately intuitive.
- **Custom protocols** may perform better in some cases.
Genome data and MPC

Secure genome-wide association study workflow

Scenario 1: secure 23andMe
- Data acquisition
- Genotype & phenotype
- Secure storage & processing
- Case & control group index
- Secure statistical testing
- Results of the study

Scenario 2: international consortium study
- Gene bank 1
- Gene bank n
- Genotype/phenotype
- Secure storage & processing

Scenario 2: Phenotype-based filtering
- Research institution
- Secure storage & processing

Scenario 1: Extended clinical study
- Data acquisition and secure storage
- Available phenotype information
- Securely computed case/control index vector
- Research institution

A

B

C
Application development

Description of the data analysis task

- Business logic
- Data model
- UX requirements

SecretC language

Application Server package

Controller library

End user applications

end users (data owners, analysts etc)

secure application servers
Our competition entry

• **Task 2.1**
  • Importer (C++/SecreC), ~200 lines of code
  • Analyzer (C++/SecreC), ~200 lines of code
  • Secure operations used: secure integer arithmetic, floating point arithmetic, including division.

• **Task 2.2**
  • Importer (C++/SecreC), ~200 lines of code
  • Analyzer (C++/SecreC), ~300 lines of code
  • Secure operations used: secure integer arithmetic, shuffling, AES.
The Rmind tool

Type 'citation()' on how to cite R or R package.

Type 'demo()' for some demos, 'help()' for help

Type 'help.start()' for an HTML browser interface to

Type 'q()' to quit

Rmind
Copyright (C) Cybernetica AS
Type 'q()' to quit
Connected
Connecting to Sharemind...

> subject <- read.csv("subject1000.csv",
> salary <- read.csv("avg-salaries.csv",
> edu <- merge(subject, salary)
> age <- edu$age
> sal <- edu$avgSalary100
> plot(age, sal)
>
The RmInd tool
Features of Rmind

- **Data import**: CSV, anything with custom importers
- **Descriptive statistics**: stdev, var, cov, quantiles, histogram, frequency plots, heatmap
- **Quality assurance**: filtering, outlier removal with median absolute deviation
- **Transformations**: Sorting, merging, aggregation
- **Testing**: t-test, chi-square, Cochrane-Armitage, transmission disequilibrium, Wilcoxon, Mann-Whitney
- **Multiple testing**: Bonferroni correction, Benjamini-Hochberg procedure
- **Regressions**: linear, logistic
- We are continuously implementing new functions.
Legal situation

- In January 2014, the Estonian Data Protection Agency cleared the use of Sharemind/Rmind for education records of Estonian students.
- In January 2015, the Estonian Tax and Customs Board cleared the use of Sharemind/Rmind for analyzing tax records of working students.
- We also have experience in forming contracts with all associated parties under European law.


Acknowledgments

Our entry to the iDASH Privacy & Security Workshop Secure Genome Analysis Competition was prepared with support from

PRACTICE

http://practice-project.eu/

"The PRACTICE project has received funding from the European Union's Seventh Framework Programme ([FP7/2007-2013]) under grant agreement number ICT-609611."

The information in this document is provided “as is”, and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.