Deploying Secure Computing for Real-world Applications

Dan Bogdanov, PhD
Head of Privacy Technology Development
Cybernetica
dan@cyber.ee
The Sharemind Privacy-preserving Computing Platform
Components for Privacy

- **Encrypted computing**
  - MPC
  - FHE
  - Trusted hardware

- **Privacy policies**
  - Multi-party consensus
  - Disclosure control

- **Audit support**
  - Online verification
  - Offline audit
Secure Computing Model

Input parties

\[ IP_1 \quad \ldots \quad IP_k \]

Computing parties

\[ CP_1 \quad \ldots \quad CP_l \]

Result parties

\[ RP_1 \quad \ldots \quad RP_m \]

\begin{align*}
& x_1 \rightarrow x_{11} \rightarrow x_{1i} \rightarrow x_{1l} \\
& x_k \rightarrow x_{k1} \rightarrow x_{ki} \rightarrow x_{kl} \\
& \vdots \\
& \vdots \\
& \vdots \\
& \vdots \\
& CP_1 \rightarrow y_1 \rightarrow y_i \rightarrow y_l \\
& \vdots \\
& \vdots \\
& \vdots \\
& \vdots \\
& \vdots \\
& \vdots \\
& RP_1 \rightarrow y \rightarrow RP_m
\end{align*}

Step 1: upload and storage of inputs
Step 2: Sharemind servers
Step 3: publishing of results
Programmable Architecture

- **Sharemind**
  - Interfaces
  - Application servers
  - Database backends

- **Java/JavaScript/C/C++/Haskell**
  - Mobile apps
  - Web apps
  - Desktop apps
  - SQL queries
  - Rmind statistics package

- **Host 1**
- **Host 2**
- **Host n**
# Sharemind’s Protocols

<table>
<thead>
<tr>
<th>Name</th>
<th>num of input parties</th>
<th>num of computing parties</th>
<th>num of result parties</th>
<th>Technology</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>shared3p</td>
<td>any</td>
<td>3</td>
<td>any</td>
<td>LSS/MPC</td>
<td>In commercial use</td>
</tr>
<tr>
<td>shared2p</td>
<td>any</td>
<td>2</td>
<td>any</td>
<td>LSS/MPC</td>
<td>Under development</td>
</tr>
<tr>
<td>sharednp</td>
<td>any</td>
<td>3 or more</td>
<td>any</td>
<td>LSS/MPC</td>
<td>Under development</td>
</tr>
</tbody>
</table>

More are being planned
1. Pick random number $a_1 = 57$
2. Pick random number $a_2 = 13$
3. Find $a_3 = 25 - 57 - 13 \equiv 55 \mod 100$
4. Send $a_k$ to Server $k$, ($k \in \{1, 2, 3\}$)

1. Pick random number $b_1 = 44$
2. Pick random number $b_2 = 57$
3. Find $b_3 = 33 - 44 - 57 \equiv 32 \mod 100$
4. Send $b_k$ to Server $k$, ($k \in \{1, 2, 3\}$)

Student C learns that the sum of A’s and B’s score is 58 without learning the scores of either student.
Getting More Operations

- (continued example)
- Addition derives from the homomorphic property of additive secret sharing.
- Further operations require network communication.
- The challenge is finding non-trivial ways to simplify the more complex protocols to make them efficient and keep them composable.

Coding for Sharemind
Analytics with Sharemind
Demo Contents

• Programming SMC using SecreC
• Parallel operations
• Security protocol polymorphism
• Usability of SMC
• The Rmind statistics tool


Secure Computing for Governmental Statistics
It’s a Good Time to be in IT

Software developer shortage transcends international boundaries

Shortage brings demand for overseas engineers

The Myth of America's Tech-Talent Shortage

Computer science graduates struggle to find work despite IT skills shortage

The fact that up to 900 000 jobs in the ICT sector remain unfilled because of a skills gap gives the clearest indication possible of what needs to be done,” says Manuel Kohnstamm, Liberty Global’s senior vice president and chief policy officer.

IT Training has a Failure Rate

By 2012, a total of 43% of students enrolled in the four largest IT higher learning institutions in Estonia during 2006-2012 had quit their studies. Source: Estonian Ministry of Education and Research, CentAR.
Government has the Data

**Tax records**
- Has the student worked?
- In which period?
- In an IT company?

**Barriers**
- Data Protection
- Tax Secrecy

**Education records**
- How is working related to not graduating on time?
- When did the student enrol?
- When did he or she graduate?
- In an IT curriculum?
Sharemind Deployment

600,000 records

Education records
Ministry of Education and Research

Employment tax records
Estonian Tax and Customs Board

10,000,000 records

Estonian Information System’s Authority
Ministry of Finance IT Center
Cybernetica

... collected data in an encrypted form,
... prevented any server from opening the data,
... ran queries without removing encryption
and enforced restrictions on result publishing.

Statistician from Centar
Universities Companies Policymakers

Secure Computing for Tax Fraud Prevention
VAT Evasion is a Problem

![Bar chart showing VAT evasion by different taxes]

- VAT: €300.6M
- Social tax: €67.5M
- Income tax: €9.3M
- Alcohol excise: €9.4M
- Tobacco excise: €11.2M
- Fuel excise: €10.7M
- Packaging excise: €14.4M
The Story of the 1000 € Law

• In 2013, the Estonian parliament ratified the Value-Added Tax Act and the Accounting Act Amendment Act that would force enterprises to report all invoices above 1000 € to the Tax and Customs Board (MTA).

• MTA then matches outgoing invoices to the incoming invoices reported by others and find companies trying to get refunds for fraudulently declared input VAT.

• President Ilves refused to proclaim the law, as “…creating a database containing almost all of Estonia’s business secrets cannot be justified with a hypothetical, unproven conjecture that the tax hole would diminish.”

http://news.err.ee/v/politics/5b358dbd-8836-43ca-992c-973d206a3ac6
Prototype with SMC

Benefits
- Analyze, combine and build reports without decrypting data.
- Confidentiality is guaranteed against all servers and against malicious hackers.
- Values are only decrypted when all hosts agree to do so.

Benefits
- Encryption is applied on the data directly at the source.
- The data is cryptographically protected during processing.
- No need to unconditionally trust a single organization.

Tax Office
- Tax Office server
- Risk queries
- Risk scores
- Tax Office

Taxpayer's association's server
- Transactions

Watchdog NGO server
-

Sharemind

Secure multi-party computation system with database

Large-scale Benchmarks

12 computing nodes running a total of 80 Sharemind processes

Client instance

eu-central

eu-west
Even Larger Data Size

### Table 3. Descriptions of the three data sets used in the experiments.

<table>
<thead>
<tr>
<th>No. of companies</th>
<th>No. of transaction partner pairs</th>
<th>Total no. of transactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 000</td>
<td>200 000</td>
<td>25 000 000</td>
</tr>
<tr>
<td>40 000</td>
<td>400 000</td>
<td>50 000 000</td>
</tr>
<tr>
<td>80 000</td>
<td>800 000</td>
<td>100 000 000</td>
</tr>
</tbody>
</table>

The source data for 100 000 000 transactions had a total size of 35 GB in XML format (about 1 GB in the secret-shared database).
# Computing Environment

<table>
<thead>
<tr>
<th>Setup</th>
<th>Client</th>
<th>Computing parties</th>
<th>Latency (round-trip)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>us-east – c3.8xlarge</td>
<td>us-east – 12x c3.8xlarge</td>
<td>&lt; 0.1ms between all nodes</td>
</tr>
<tr>
<td>2</td>
<td>eu-west – c3.8xlarge</td>
<td>eu-west – 8x c3.8xlarge, eu-central – 4x c3.8xlarge</td>
<td>&lt; 0.1ms inside eu-west 19ms (eu-west/eu-central)</td>
</tr>
<tr>
<td>3</td>
<td>us-east – c3.8xlarge</td>
<td>us-east – 4x c3.8xlarge, us-west – 4x c3.8xlarge, eu-west – 4x c3.8xlarge</td>
<td>77ms (us-east/us-west), 133ms (us-west/eu-west), 76ms (us-east/eu-west)</td>
</tr>
</tbody>
</table>
Cross-ocean SMC Runtime

<table>
<thead>
<tr>
<th>Computation phase</th>
<th>20k</th>
<th>40k</th>
<th>80k</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk analysis</td>
<td>38:44</td>
<td>01:23:10</td>
<td>02:47:53</td>
</tr>
<tr>
<td>Aggregation</td>
<td>01:14:36</td>
<td>02:25:12</td>
<td>05:05:16</td>
</tr>
<tr>
<td>Upload</td>
<td>04:26:15</td>
<td>20:53:00</td>
<td></td>
</tr>
</tbody>
</table>
Rather Acceptable Costs

Dan Bogdanov, Marko Jõemets, Sander Siim, Meril Vaht. **Privacy-preserving tax fraud detection in the cloud with realistic data volumes.** Real World Crypto 2016 Lightning Talk. https://drive.google.com/file/d/0Bzm_4XrWnlSvVnRTRF9wT0EtvUW8/view?pref=2&pli=1
Brute force risk analysis

![Bar chart showing computation time for different numbers of companies and regions.](chart.png)

- **Computation phase**
  - Risk analysis
  - Aggregation
  - Upload

- **Number of companies**
  - 20k
  - 40k
  - 80k

- **Computation time**
  - 02:55:40
  - 09:29:57
  - 33:34:07
  - 22:38:25
  - 48:41:02
  - 111:16:25

---

**Legend**
- **us**: United States
- **2-eu**: Europe (2)
Cost of using brute force

Number of companies

Deployment region

us

2-eu

$36.96

$89.04

$221.76

$197.9

$415.41

$1028.67

Deployment regions

- $36.96
- $89.04
- $221.76
- $197.9
- $415.41
- $1028.67
Take-home Messages

• Sharemind is designed to be a privacy platform that use secure computing as component.

• It used to focus on three-party secure computing, but this less the case as time goes on.

• Sharemind also includes other privacy techniques like side-channel-safe statistics and audit features.

• Cybernetica is continuously developing privacy technologies for use in real-world applications.
We Build Applications

Learn about Sharemind
http://sharemind.cyber.ee/

Open source prototyping tools (under development)
http://sharemind-sdk.github.io/

Contact us for more information and collaborations
E-mail: sharemind@cyber.ee
Twitter: @sharemind