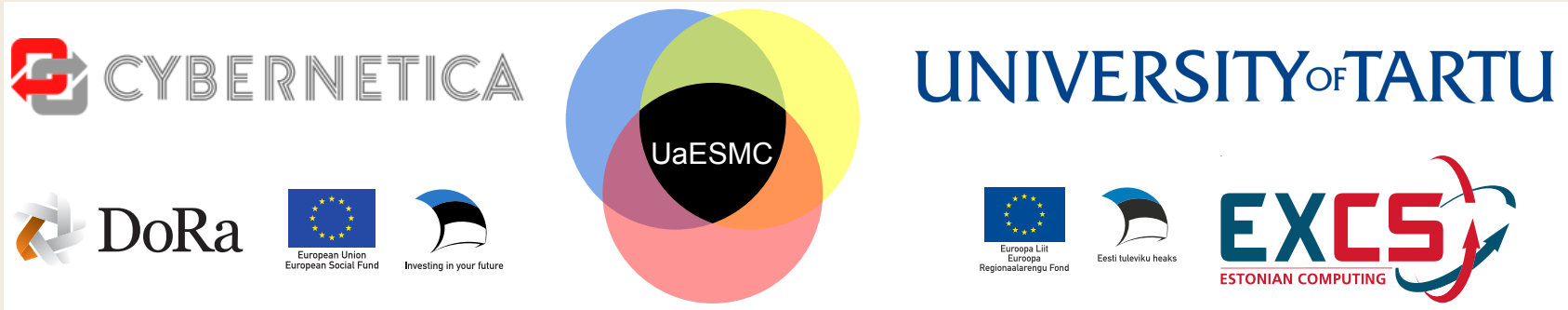
Practical linking of databases using secure multiparty computation

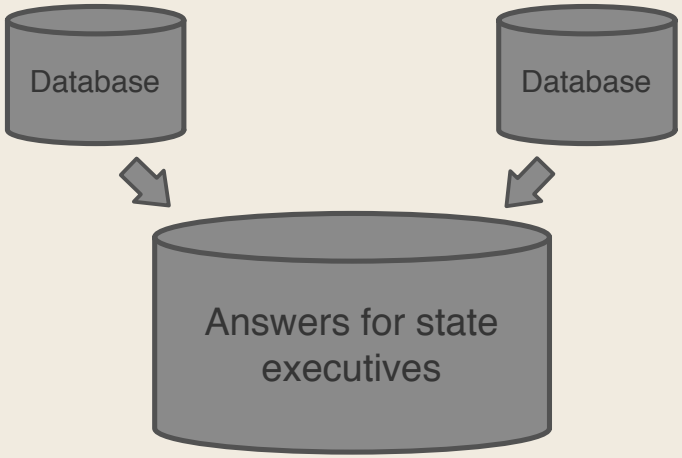
Riivo Talviste
riivo@cyber.ee
Cybernetica, Team Sharemind

February 21, 2014



Problem statement

- State is interested in data-driven decisions, thus needs to analyze combined data

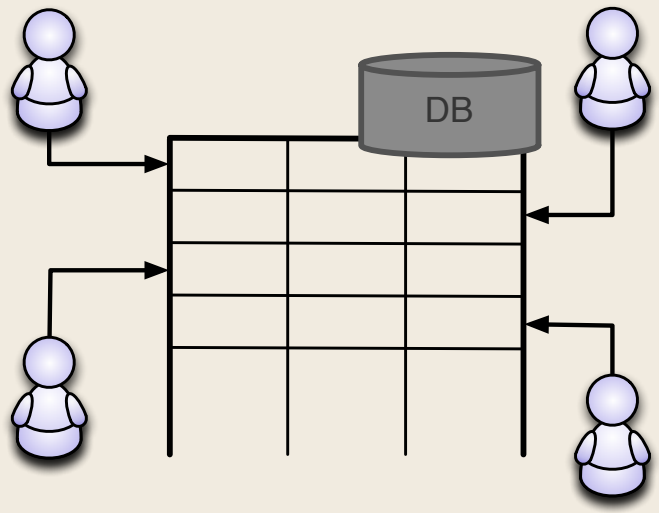


- Databases contain sensitive information
- Combined "super databases" are risky
- Replace combined databases with MPC

sharemind

Practical application

- 2013: Income analysis of public sector
  - Data sources: local governments and ministries
  - All sources have same data structure:
    `(job_title, count, salary)`

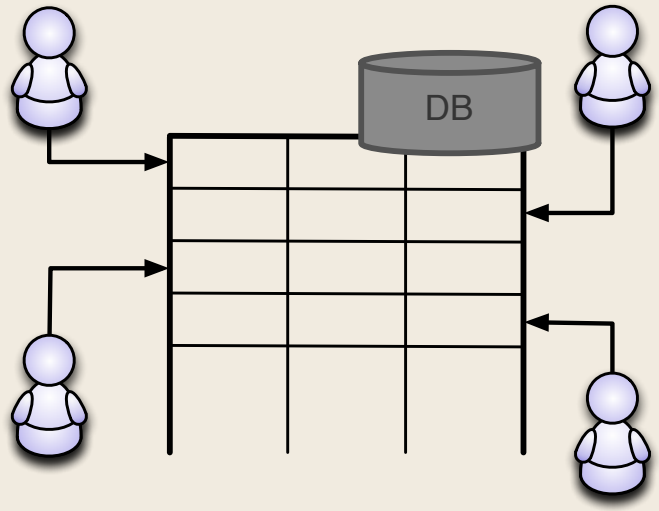


- Web-based MPC application running on public cloud

sharemind

Practical application

- 2013: Income analysis of public sector
  - Data sources: local governments and ministries
  - All sources have same data structure:
    `(job_title, count, salary)`



- Web-based MPC application running on public cloud
  - Public demo: `https://sharemind.cyber.ee/clouddemo/`

sharemind

Practical application: PRIST

- More complicated scenario: different data structure
- Universities vs. companies: *Does employment during studies have a negative effect?*
- PRIST: "Privacy-preserving statistical studies on linked databases"
  - Funded by European Union, 2013 – 2015
  - Answer that question for information technology
  - How? Link income data from Tax Office with education data from Education Information System

Practical application: PRIST

- Classical approach
  - Requires approval from Data Proteciton Agency
  - Tax Office gives out data with k-anonymity
  - Data loss with grouping education info, sex, age is 76 – 98%

sharemind

Practical application: PRIST

- Classical approach
  - Requires approval from Data Proteciton Agency
  - Tax Office gives out data with k-anonymity
  - Data loss with grouping education info, sex, age is 76 – 98%
- Using secure multipaty computation
  - Input data sets are secret shared
  - Data sets are linked using privacy-preserving database linking using personal ID codes
  - Data Protection Agency: "*No approval required as no sensitive information is gathered.*"
  - MPC platform: Sharemind Application Server

sharemind

Sharemind Application Server

- Modular desing with *protection domains*
  - e.g. additively shared 2-party with active security, n-party with Shamir sharing
- Additive 3-party protection domain:
  - Data types: [u]int{8,16,32,64}, boolean, float{32,64}, strings (known- and bounded-length)
  - Oblivious sorting (sorting networks, radix sort [BLT13]; quicksort [H$^+$12])
  - Oblivious shuffle [LWZ11]
  - Privacy-preserving database linking (SQL equi-join) [LTW13]

[BLT13] D. Bogdanov, S. Laur, R. Talviste. "Oblivious Sorting of Secret-Shared Data". Cybernetica research report T-4-19. `http://research.cyber.ee/`. 2013

[H$^+$12] K. Hamada, R. Kikuchi, D. Ikarashi, K. Chida, K. Takahashi. "Practically Efficient Multi-party Sorting Protocols from Comparison Sort Algorithms". ICISC'12.

[LWZ11] S. Laur, J. Willemson, and B. Zhang. "Round-Efficient Oblivious Database Manipulation". ISC'11.

[LTW13] S. Laur, R. Talviste, J. Willemson, "From Oblivious AES to Efficient and Secure Database Join in the Multiparty Setting". ACNS'13'.

sharemind

Sharemind Application Server

- Data persistence layer
- Programmable

sharemind

SecreC programming language

Hybrid model

```
1  // Import a PDK module called 'additive3pp'
2  import additive3pp;
3  // Create a domain 'private' from the PDK
4  domain private additive3pp;
5
6  void main () {
7    // Perform secure computations using the PDK
8    private int a = 2, b = 3;
9    private int c = a * b;
10   // We need a special function to publish 'c'
11   print (declassify (c));
12 }
```

sharemind

```
1 // Protection domain kind polymorphism
2 template<domain D>
3 D uint sum (D uint[[1]] vector) {
4   D uint result = 0;
5   for (uint i = 0; i < size (vector); i++) {
6     result[i] = result[i] + vector[i];
7   }
8   return result;
9 }
```

```
1 // Specialization to a PDK
2 template<domain D: additive3pp>
3 D uint sum (D uint[[1]] vec) {
4   D uint result = 0;
5   __syscall ("additive3pp::sum_uint64_vec",
6              __domainid (D),
7              vec, result);
8   return result;
9 }
```

sharemind

SecreC standard library

- EU-funded project "Usable and Efficient Secure Multiparty Computation" (UaESMC)
- Statistics suite
  - table filtering, linking and sorting,
  - descriptive statistics (mean, variance, standard deviation),
  - percentiles (minimum, maximum, mean, custom percentiles),
  - five-number summary and box-plots, histograms,
  - t-tests, paired t-tests, χ^2-tests, Wilcoxon tests.
- All algorithms support oblivious filters
- Build an R-like statistics application
- Why statistics?
  - We performed 25 interviews internationally
  - We explained what MPC can do and asked where it could be applied?
  - Most popular answers: statistics and optimization

Thank you!

https://sharemind.cyber.ee

The work of Riivo Talviste is supported by European Social Fund Doctoral Studies and Internationalisation Programme DoRa.

"Privacy-preserving statistical studies on linked databases" (PRIST) project is funded by the European Regional Development Fund through the Implementing Agency Archimedes Foundation.
http://cyber.ee/en/research/research-projects/prist/

"Usable and Efficient Secure Multiparty Computation" (UaESMC) project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no FP7-284731. http://www.usable-security.eu/en

sharemind