Secure statistical computation system on encrypted data:

An empirical study of secure regression analysis for official statistics

September 22, 2017

Satoshi Tanaka*, Yutaka Abe**, Satoshi Takahashi*,
Atsushi Doi*, Ryo Kikuchi*, Koji Chida*, Kiyomi Shirakawa***

* NTT Secure Platform Labs.
** National Statistics Center, Japan
*** Hitotsubashi Univ.

Copyright©2017 NTT corp. All Rights Reserved.

Overview

- NTT Lab. and Hitotsubashi Univ. have developed “Secure Computation” system for a secondary use of official statistics.

- We introduce our secure computation system.

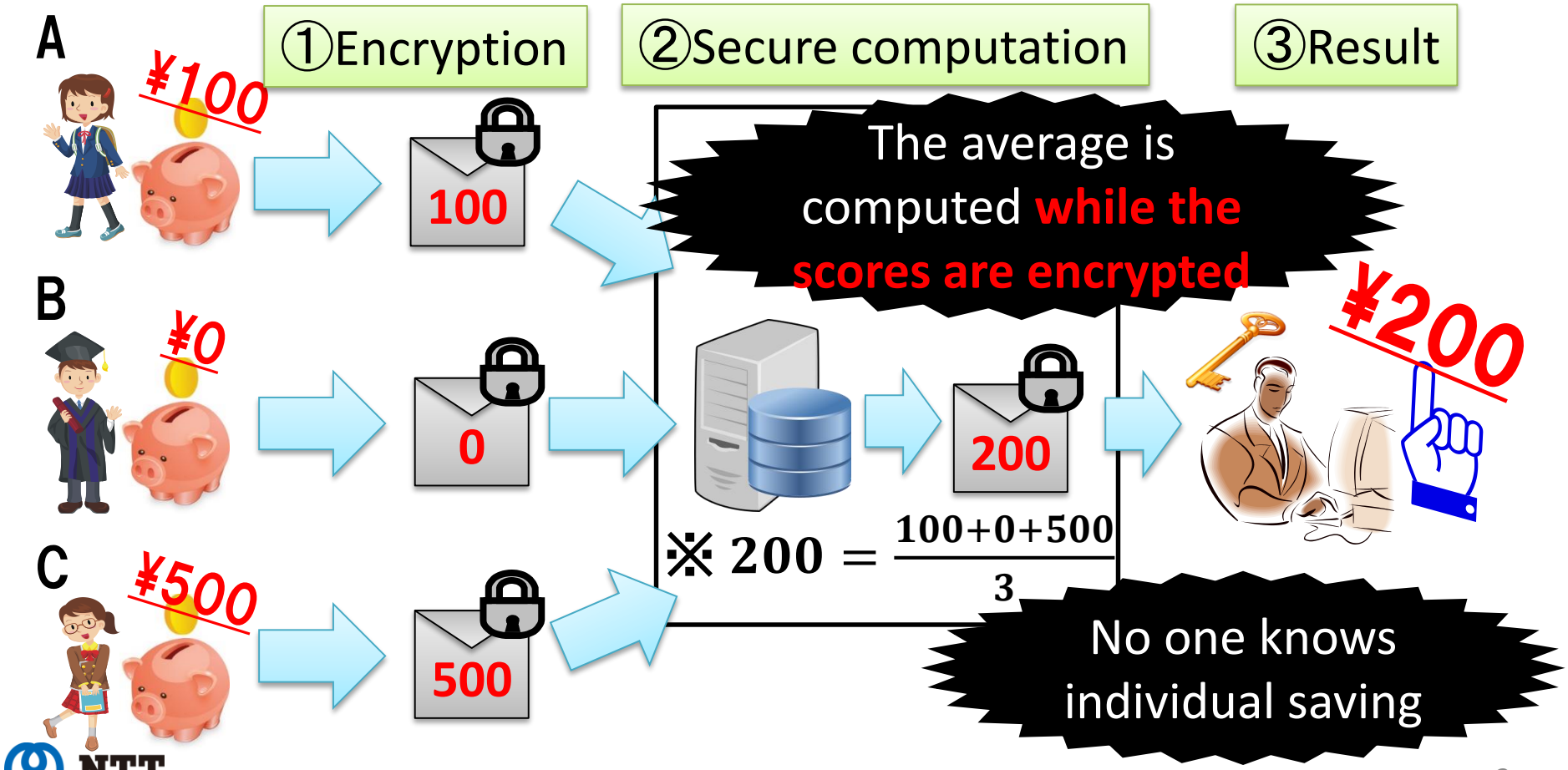
- We consider how to make a function for regressions on secure computation and show experimental results with a demonstration.

Copyright©2017 NTT corp. All Rights Reserved.

What is secure computation

A method to analyze data **while the data is encrypted**

– Example: computing average savings of three students

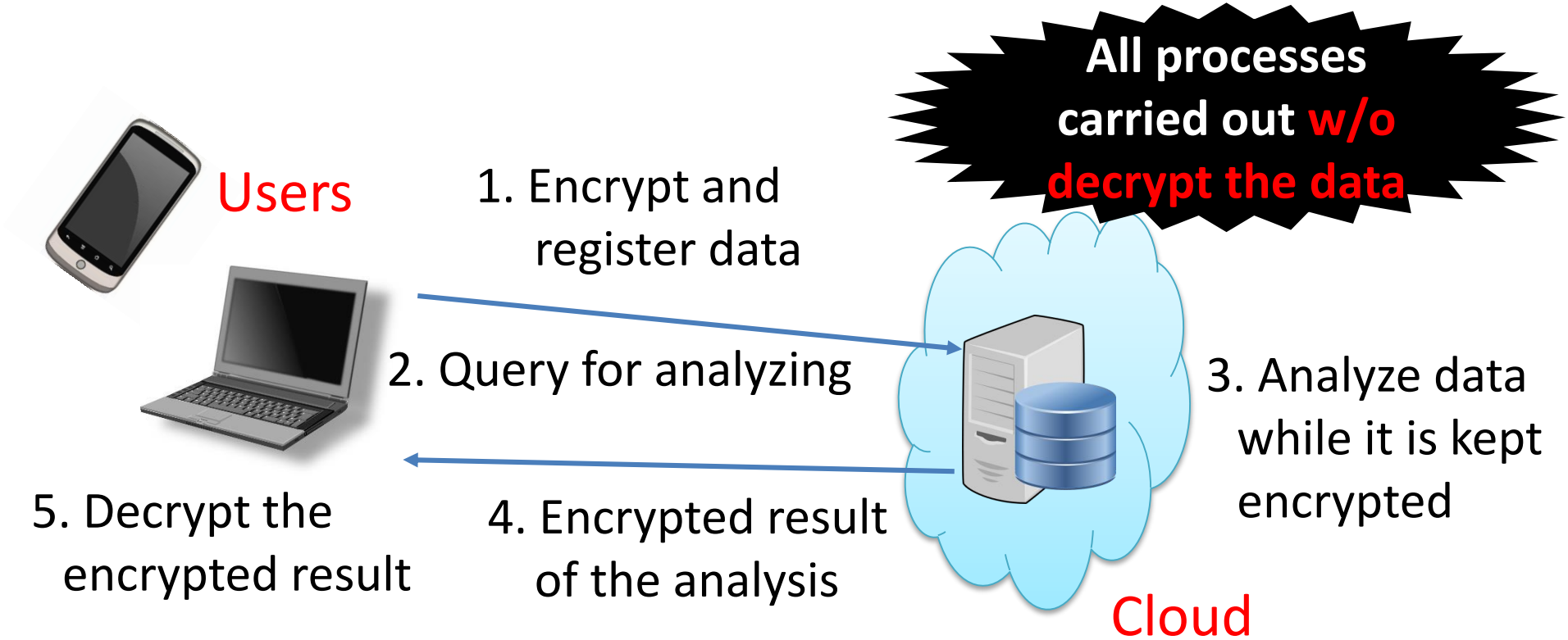


Copyright©2017 NTT corp. All Rights Reserved.

Prospect of secure computation (1)

Can analyze data **without disclosing them** to cloud servers

– Keeping Confidentiality, Privacy-preservation, and Countermeasure against information leakage



Copyright©2017 NTT corp. All Rights Reserved.

Prospect of secure computation (2)

Secure secondary use (for Stat.) of sensitive data

Data providers

Sensitive data have never known to both the server and analyst

1. Encrypt data and register them

Server

3. Compute the statistic from the encrypted sensitive data

2. Request a statistic

Data analysts

4. Send the encrypted result

5. Decrypt the result

Risk of information leakage is much reduced

Copyright©2017 NTT corp. All Rights Reserved.

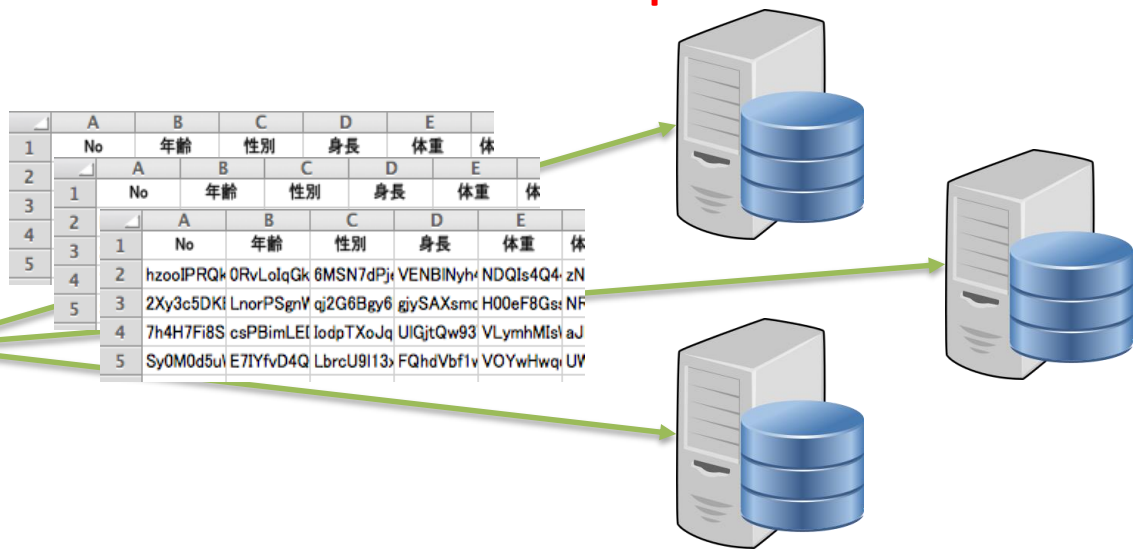
NTT secure computation system

- NTT Lab. developed a secure computation system for statistic analysis using sensitive micro data
- Data is encrypted cell-wise through the technique called “Secret Sharing”
  - The data is divided into 3 data and each sent to different server

	A	B	C	D	E
1	No	年齢	性別	身長	体重
2	1	57	男	154.8	71.7
3	2	45	男	179.9	70.0
4	3	53	男	156.3	81.1
5	4	77	男	159.2	63.0
6	5	41	女	170.6	67.5

Secure computation servers

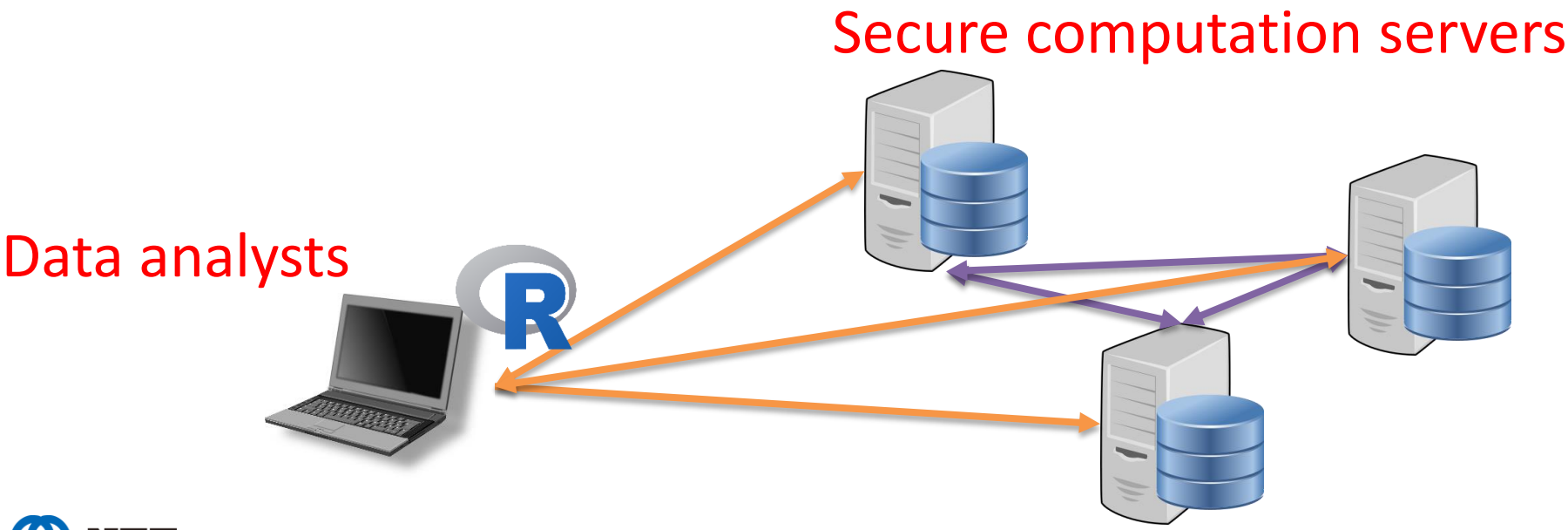
Data providers



Copyright©2017 NTT corp. All Rights Reserved.

NTT secure computation (Cont.)

- Statistical analysis on secure computation can be performed through the software 'R'
  - Data analysts request a desired analysis to the servers
  - Servers process the analysis with incorporating each other
  - The servers send an encrypted result to the data analyst



Copyright©2017 NTT corp. All Rights Reserved.

Functions of our system (Example)

- Basic statistics
  - average, variance, covariance
- Aggregation
  - cross table of up to three categorical attributes
- Primitives
  - addition, subtraction, multiplication, logical operations, equation, comparison, sorting, mapping

Copyright©2017 NTT corp. All Rights Reserved.

Collaboration of NTT and Hitotsubashi Univ.

Innovative R&D by NTT

Secure computation could be a solution for secure secondary use of official statistics

- Want to compute regression on secure computation
  - The regression was not implemented on NTT secure computation system
- Develop some regression functions by combining basic statistics and primitives implemented on NTT secure computation system **(this work)**

NTT

Copyright©2017 NTT corp. All Rights Reserved.

Main topic

- Motivation
  - Make regression functions on secure computation
- Problem
  - Some operations are expensive on secure computation (e.g. division)
- Approach
  - Secure computation computes only aggregated values

Copyright©2017 NTT corp. All Rights Reserved.

Example: secure linear regression

- Compute coefficients $w_1, \dots, w_k$ and intercepts w_0 by solving equation system:

$$\begin{pmatrix} N & S_{\boldsymbol{x}_1} & \cdots & S_{\boldsymbol{x}_k} \\ S_{\boldsymbol{x}_1} & S_{\boldsymbol{x}_1^2} & \cdots & S_{\boldsymbol{x}_1\boldsymbol{x}_k} \\ \vdots & \vdots & \ddots & \vdots \\ S_{\boldsymbol{x}_k} & S_{\boldsymbol{x}_1\boldsymbol{x}_k} & \cdots & S_{\boldsymbol{x}_k^2} \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_k \end{pmatrix} = \begin{pmatrix} S_{\boldsymbol{y}} \\ S_{\boldsymbol{x}_1\boldsymbol{y}} \\ \vdots \\ S_{\boldsymbol{x}_k\boldsymbol{y}} \end{pmatrix}$$

N: number of records,
$\boldsymbol{x}_1, \dots, \boldsymbol{x}_k$: regressors
$\boldsymbol{y}$: regressand
$S_{\boldsymbol{v}}$: sum of elements in a vector $\boldsymbol{v}$

Copyright©2017 NTT corp. All Rights Reserved.

Example: secure linear regression(cont.)

- Each value in the equation system is not personal data, but an aggregated value
  - S_{x_i} requires only **additions**,
  - $S_{x_i x_j}$, $S_{x_i y}$ require **multiplications** additionally
- Secure linear regression:
  1. Compute every S_{x_i}, $S_{x_i x_j}$, $S_{x_i y}$ on secure computation and decrypt them
  2. Solve the equation system with S_{x_i}, $S_{x_i x_j}$, $S_{x_i y}$ on non-encrypted data

Copyright©2017 NTT corp. All Rights Reserved.

Experiments

- Data: Synthetic Microdata for Educational Use
  - Generated with pseudorandom numbers from 2004, National Survey on Family Income and Expenditure in Japan
  - Contains 32,027 records with 197 attributes
- Details:
  1. Aggregate frequency and magnitude by the number of members and workers
  2. Linear regression between outcome and income.
  3. Chow test in the different number of workers

Copyright©2017 NTT corp. All Rights Reserved.

Demo

Please see a demonstration

Data: Synthetic Microdata for Educational Use
(32,027 records, 197 attributes)

Copyright©2017 NTT corp. All Rights Reserved.

Discussion

- Performance
  - Have sufficient accuracy compared with non-encrypted data
  - All experiments were finished within few seconds
- Security
  - Depends on the assumption:
    aggregated values do not leak privacy information
    - Some cases (e.g. only 1 record) break the assumption
    - Expect statistical disclosure control (SDC) protects privacy in such bad cases

Copyright©2017 NTT corp. All Rights Reserved.

Conclusions and Future works

- NTT Lab. and Hitotsubashi Univ. have developed "Secure Computation" system for a secondary use of official statistics
- Introduced our secure computation system
- Presented an examples to make a function for regressions on secure computation
- Showed experimental results on a demo

- Future works
  - Implement other regressions and statistical analysis
  - Consider SDC for secure computation

Copyright©2017 NTT corp. All Rights Reserved.