Oct.2015-Sep.2021

Secure Data Sharing and Distribution Platform for Integrated Big Data Utilization

- Handling all data with encryption -

ビッグデータ統合利用のためのセキュアなコンテンツ共有・流通基盤の構築

Joint project by 6 Universities

Waseda University Hayato YAMANA
Institute of Information Security Atsuhiro GOTO
Ochanomizu University Masato OGUCHI
Kogakuin University Saneyasu YAMAGUCHI
The University of Electro-Communications Takahiko SHINTANI
Meiji Pharmaceutical University Tamotsu NOGUCHI

Waseda University
INSTITUTE of INFORMATION SECURITY
Ochanomizu University
工学院大学
UEC TOKYO
MEIJI PHARMACEUTICAL UNIVERSITY
JST 国立研究開発法人 科学技術振興機構 Japan Science and Technology Agency

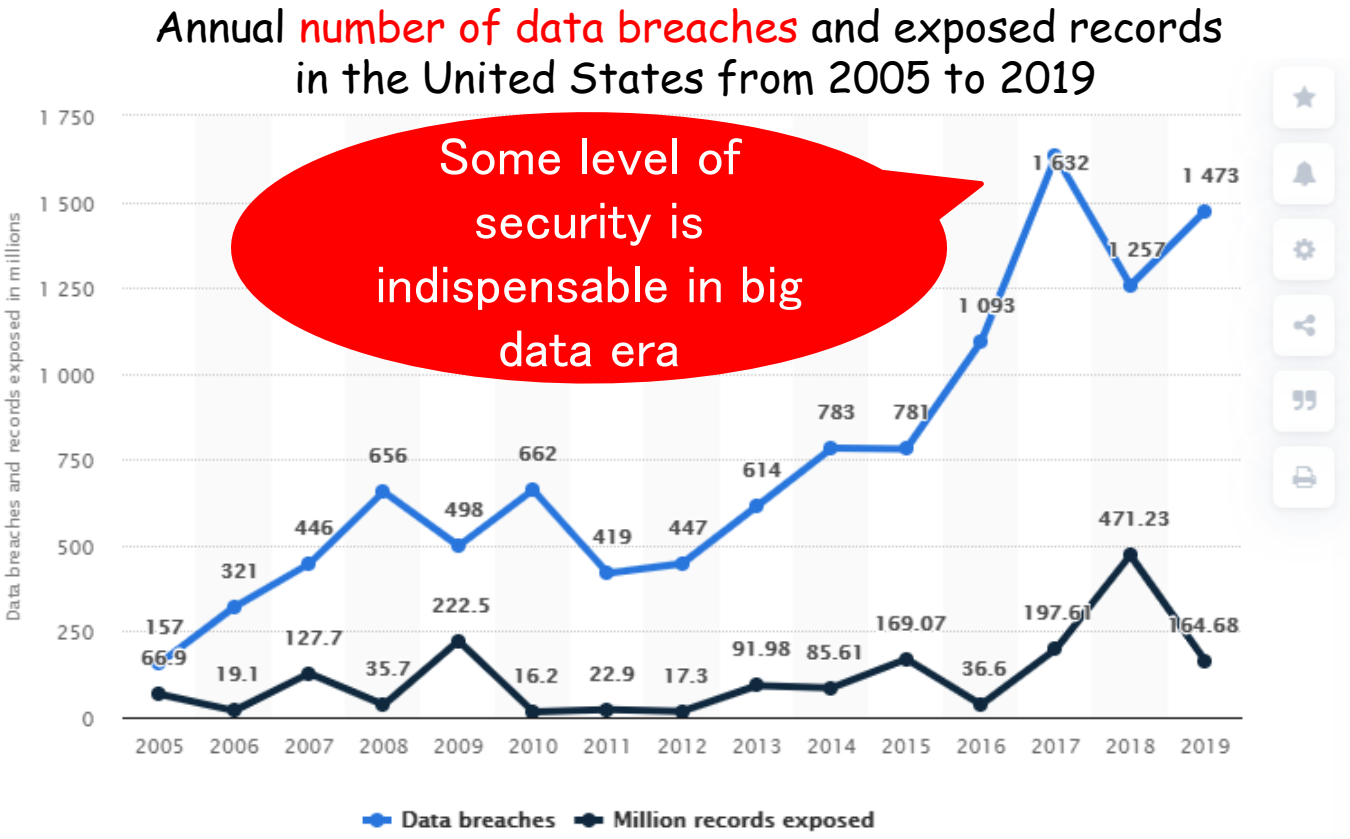
OVERVIEW OF THE PROJECT

AGENDA

Background

1. Goal and Achievements(**目標と成果概要**)
   - 1-1 Research Goal
   - 1-2 Achievements
2. Research Outcome(**研究成果**)
   - 2-1 Publications
   - 2-2 Contribution to Innovation
3. Summary(**今後の展開**)

Background



https://www.statista.com/statistics/273550/data-breaches-recorded-in-the-united-states-by-number-of-breaches-and-records-exposed/

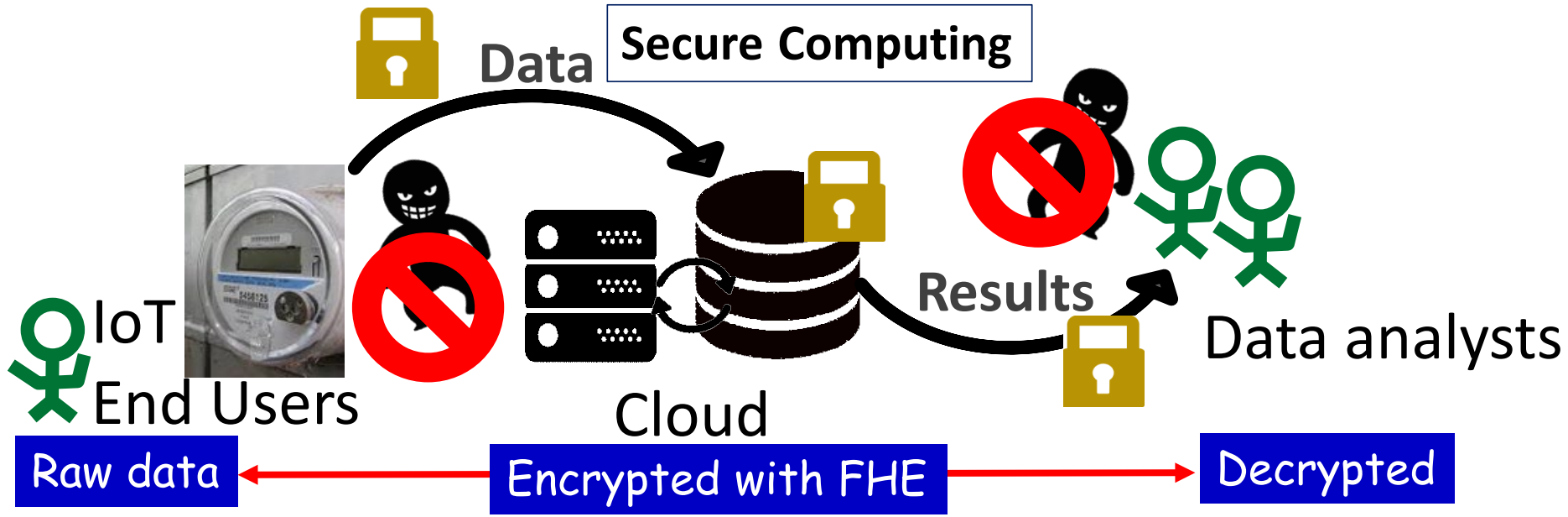
Based on IDC white paper (Mar.2017), by 2025, almost 90% of all data will require some level of security.

1. GOAL AND ACHIEVEMENTS
（目標と成果概要）

1-1 Research Goal
1-2 Achievements

1-1 Research Goal

HANDLING ALL DATA WITH ENCRYPTION
THROUGHOUT DATA LIFE CYCLE



Fully Homomorphic Encryption (FHE)
Outsourcing and calc. over encrypted data
too slow to adopt → Goal: 1,000 times speed-up

1-1 Research Goal

- Protecting the **PRIVACY** of Big Data even in its calculation by adopting Fully Homomorphic Encryption (FHE)
- Speeding-up encrypted calculation, especially data mining, **1,000 times*** over HElib1.3.
  - Adopting new theory, I/O optimization, parallelization and distributed computing
- Constructing a Platform including library for Secure Computation with **FHE**

***Calculation w/ FHE was originally 10^7 to 10^{10} times slower than calculation w/o FHE.**

1-2 Achievements

OVERVIEW

- Exceeded the original goal
  - **2,912 times faster** than original HElib. (new subring homomorphic encryption 26 times faster x Middleware 112 times faster)
- 13 open-source software
  - 1 new homomorphic encryption library
  - 7 middleware libraries to speedup FHE
  - 5 application libraries
- Because of COVID-19 situation, reduced the scale of the demonstration experiments

Preliminaries on FHE

- Given Integers X and Y,

$$X \times Y = Dec\big(Enc(X) \times Enc(Y)\big),$$
$$X + Y = Dec\big(Enc(X) + Enc(Y)\big)$$
$$,where\ Enc(X) \neq Enc(X)\ and\ Enc(Y) \neq Enc(Y)$$

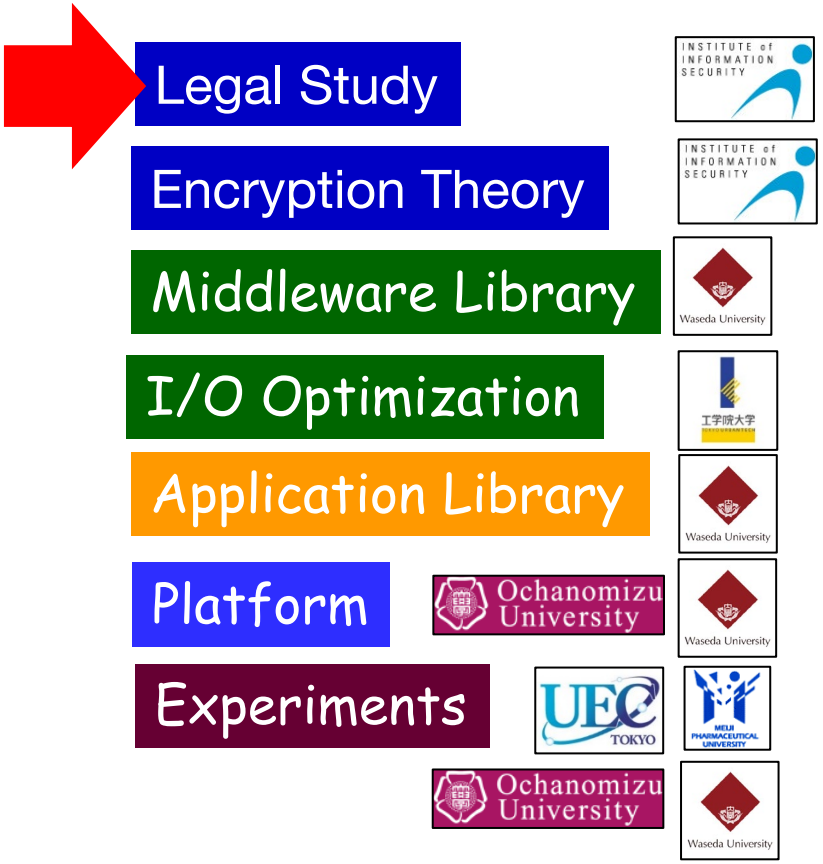
- We cannot use **branch (if)** operations.
- Other operations will be implemented as **bit-operations**.

→ **slow**

- Long execution time
  - After a given number of multiplications, noise-reset (bootstrapping) is required to keep the cypher text decryptable.
  - For FHE - B/FV Scheme, cypher text size increases with the number of multiplications, which results in long execution time.(relinearize problem)
  - Cypher text size becomes several tens to hundreds MB.

1-1 Achievements

- This project was organized by seven sub-projects.



Clarifying the legal treatment of encrypted data

1-1 Achievements

Legal Study

Research on the legal systems of various countries
→ Clarification of the handling method for **personal data/information**

Policy Proposal on handling personal data/information with encryption
in corporation with JILIS(Japan Inst. of Law and Information System)

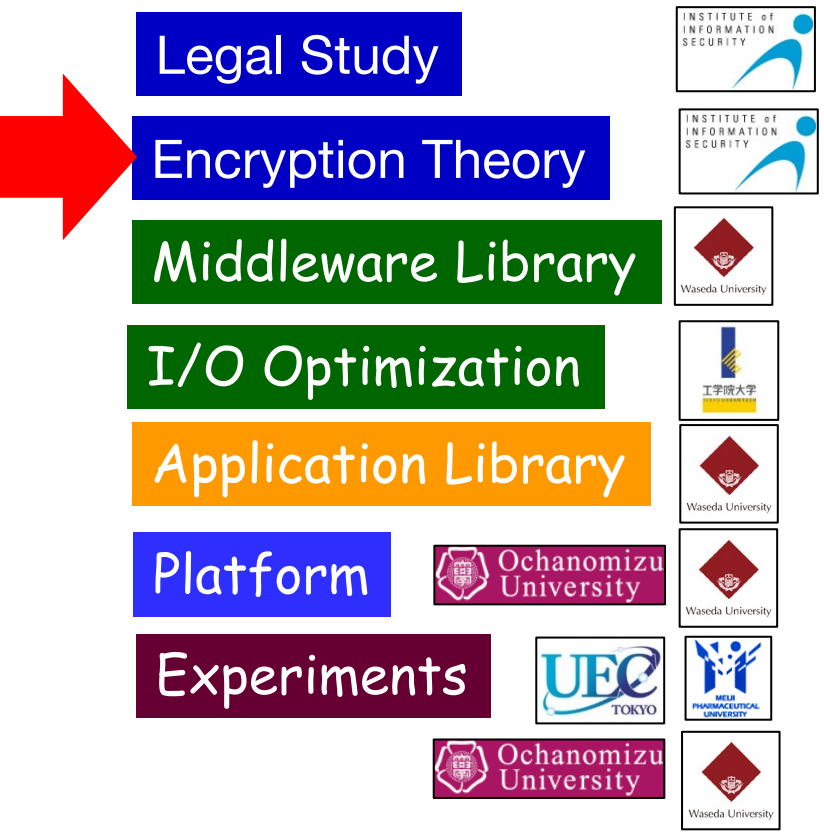
Act on the Protection of Personal Information Amendment Bill (2020)

- Establishment of the legal notion of "Pseudonymization".
  - c.f. GDPR endorses pseudonymization when processing personal data. (Encryption is regarded as pseudonymization)
- Creating the notion of pseudonymization will endorse encryption to process personal data and personal information in Japan.

→ Affected the revision of Japanese Personal Information Protection Law
→ Significant advance in utilization of encrypted data

1-1 Achievements

- This project was organized by seven sub-projects.



Goal:Speeding up 30 times by encryption scheme

1-1 Achievements

Encryption Theory

INSTITUTE of INFORMATION SECURITY

- FHE4FX/FHE4FL[AN16] the world's first
  - enabled floating point number computation
  - enabled greater-than comparison (several 10 sec)
- Subring-HE/FV and Subring-HE/BGV[AH17][AH20] OSS
  - enabled SIMD execution with power of 2 slots
  - accelerated bootstrapping

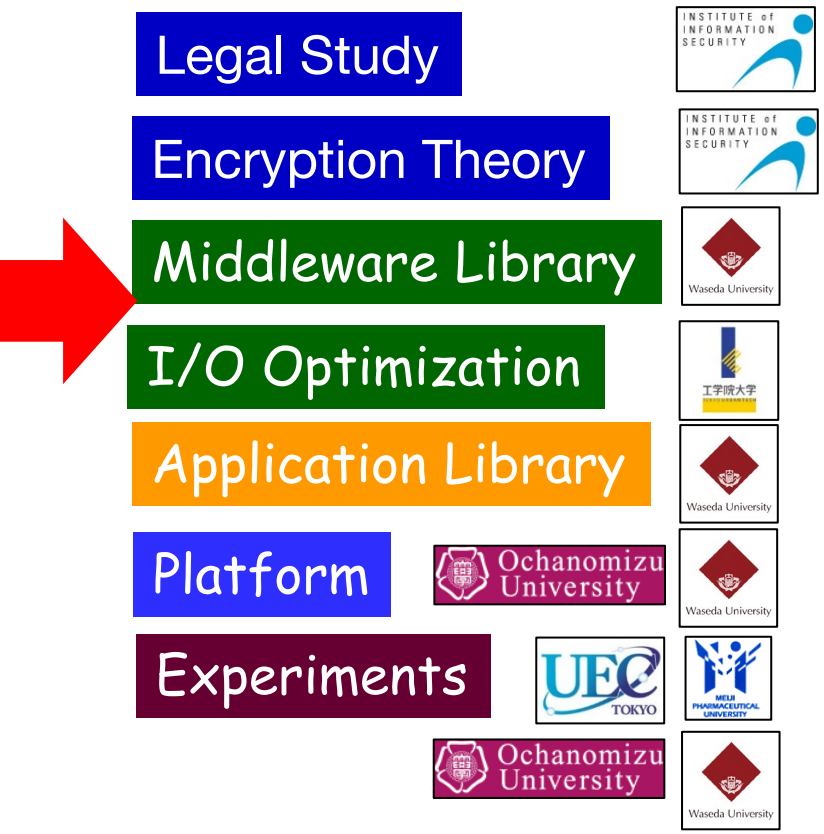
→ 26 times faster than HElib1.3

  - adopting FPGA acceleration for "basis conversion between ring and slots" (15% speed-up from w/o FPGA)

[AN 16] S.Arita and S.Nakasato, Fully Homomorphic Encryption For Point Numbers, In Proc. of INSCRYPT 2016, Beijing, China, pp.253-270, 2016.
[AH17] S.Arita and S. Handa, "Subring Homomorphic Encryption," In: Kim H., Kim DC. (eds) Information Security and Cryptology – ICISC 2017. ICISC 2017. Lecture Notes in Computer Science, vol 10779. Springer, Cham, pp.112-136 (2017.12)
[AH20] S.Arita and S. Handa, "Fully Homomorphic Encryption Scheme Based on Decomposition Ring", IEICE TRANS., Vol.E103-A,No.1,pp.195-211 (2020.1)

1-1 Achievements

- This project was organized by seven sub-projects.



Goal: Speeding up 30 times by middlewares

1-1 Achievements

Middleware Library 

I/O Optimization 

- De facto standardizing the techniques to speed-up FHE

Before (HElib1.3(IBM)/2015)

- Provided primitives of calculation only

After

- Provided seven middlewares to speed-up FHE **OSS**

→ 112 times faster than w/ only HElib1.3

- Provided five application models w/ FHE **OSS**
- Proposed other three techniques/applications w/FHE
- Proposed other four application models w/ FHE

1-1 Achievements

Middleware Library

I/O Optimization

OSS		years	speed-up ratio	description
OSS	**(M-1)** FCMalloc	2015-2017	4.72*	Fast Memory Allocator optimized for FHE
OSS	MCMalloc			Optimized version of FCMaclloc for Many-Core Shared-Memory Machine
	(M-2) Decresing #bootstrapping	2017-2018	2.67*	Optimization of decreasing the # of bootstrapping by loop unfolding.
OSS	**(M-3)** Opti. of bootstrapping / relinerization	2018-2020	1.3	Near-optimal solution for optimization of decreasing the # of bootstrapping and relinearization
	(M-4) Revocation of ABE	2016-2017	3.7	Speeding-up of revoking on attributed based encryption
OSS	**(M-5)** Prv_cmp	2017-2018	2.7*	Non-Interactive and fully output expressive private comparison (enabling comparison w/o decryption)
OSS	**(M-6)** DAMCREM	2018-2020	2.75*	Dynaminc FHE task scheduling: Dynamic allocation of computation resource to macro-tasks for FHE
OSS	**(M-7)** FHE-Table-Search	2018-2020	infinite	FHE-Table-Search function evaluation for any number of inputs using table lookup (3.4μs/entry)
OSS	**(M-8)** Blkctrl	2015-2020	1.2*	Block control for Ext2/3/4 to utilize outer tracks

* by adopting (M-1)(M-2)(M-5)(M-6)(M-8) simultaneously, we will have 112 times (4.72x2.67x2.7x2.75x1.2) speed-up from the model only w/ HElib

1-1 Achievements

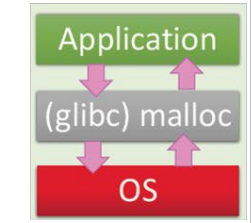
(M-1) FCMalloc/MCMalloc

Waseda University

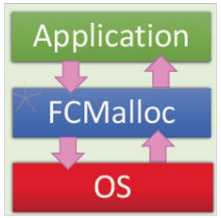
OSS

FHE equipped applications tend to request similar sizes of memory frequently

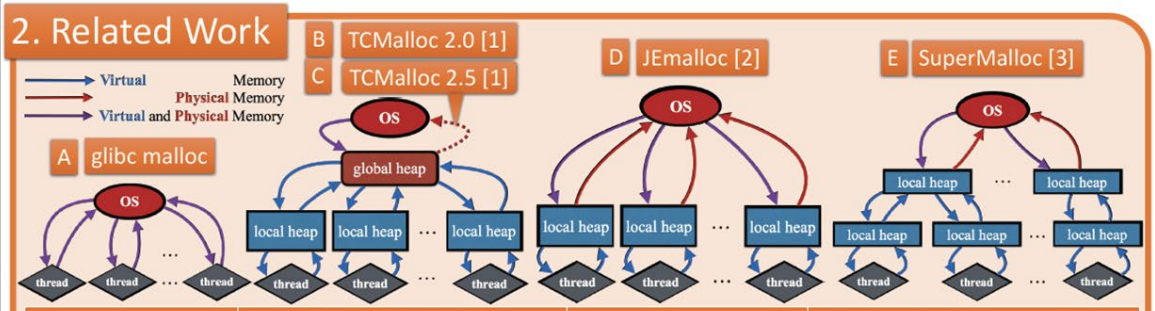
Speed-up by 1) Managing Local Heap individualy insted of using Central Heap, 2) Implemeting Pseudo Free mechanism.



Replace



2. Related Work



Name	Freeing Mechanism		Heap		Communication among Heaps	
	avoids freeing Virtual Memory	avoids freeing Physical Memory	adopts Global Heap	adopts Local Heap	through Global Heap or OS	directly among Local Heaps
A: glibc malloc						
B: TCMalloc [1] (before 2.2rc)	✓	✓	✓	✓	✓	
C: TCMalloc [1] (2.2rc or after)	✓		✓	✓	✓	
D: JEmalloc [2]	✓			✓	✓	
E: SuperMalloc [3]	✓			✓	✓	✓
Proposed	✓	✓		✓		✓

4.72 times speed-up (MCMalloc) in comparison with JEmalloc [UY17]

[UY17] A.Umayabara, H. YAMANA, "A Scalable Memory Allocator for Multithreaded Applications on a Many-Core Shared-Memory Machine," Proc. of IEEE Big Data 2017 (2017)

1-1 Achievements

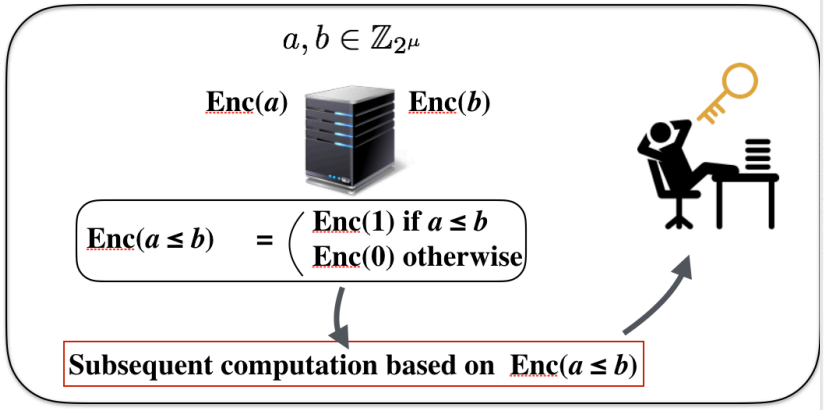
(M-5) Private Comparison with FHE

Waseda University

OSS

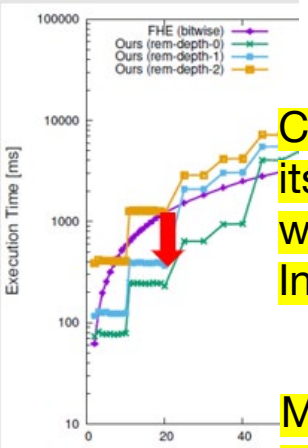
- Bitwise comparison (w/TFHE) requires long execution time
- No calculations after comparison w/ FHE

Speed-up and further calculation enabled comparison w/ FHE



$$(f() \boxdot \mathrm{Enc}(a \le b)) \boxplus (g() \boxdot (1 - \mathrm{Enc}(a \le b)))$$

before: no handling of the above calculation w/o decryption



Comparison itself w/ 40bit Integer

2.7 times speed-up in comparison with bitwise comparison [IY18]

Multiplication after a Comparison

10,000 times speed-up for a mul. after the comparison [IY18]

after: Enabled

[IY18] Y.Ishimaki, H.Yamana:"Non-Interactive and Fully Output Expressive Private Comparison," Proc. of INDOCRYPT 2018, pp.355-374 (2018)

1-1 Achievements

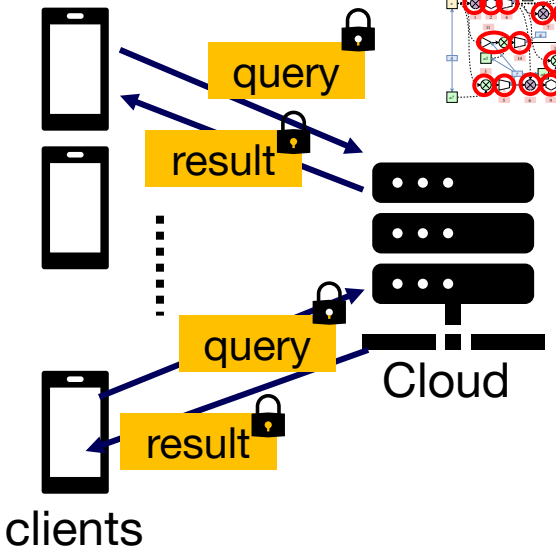
(M-6) DAMCREM

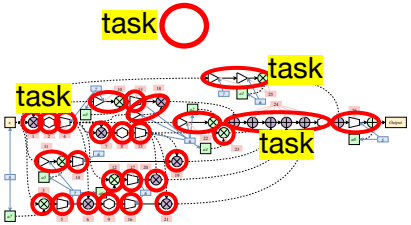
Waseda University

OSS

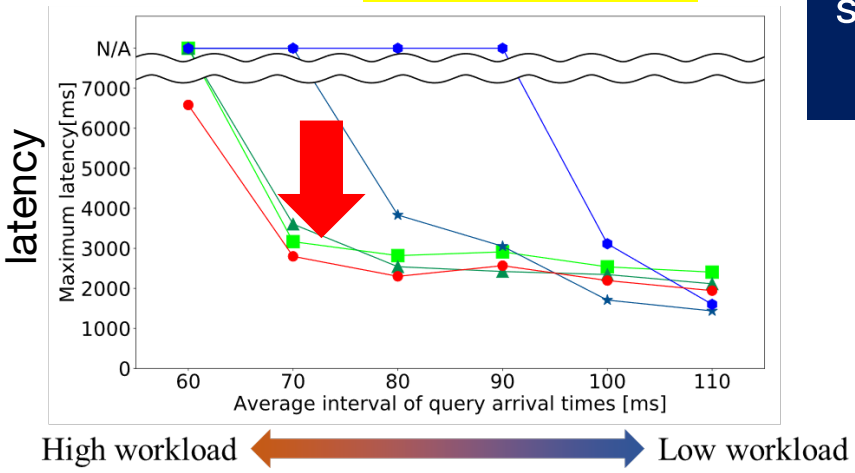
PROS of FHE tasks : predictable execution time / #threads

Speed-up by deciding #threads dynamically based on #free-cores





Pre-calculate the execution time of each task and dynamically choose the best #threads



2.75 times speed-up in comparison with static scheduling [SI20]

[SI20] T.Suzuki, Y.Ishimaki, H. Yamana, "DAMCREM: Dynamic Allocation Method of Computation REsource to Macro-Tasks for Fully Homomorphic Encryption Applications," Proc. of IEEE BITS2020 (2020)

1-1 Achievements

(M-7) FHE Table Search

OSS

- FHE cannot calculate complex functions, e.g. logarithm, trigonometric f.
- Bitwise calculations require long execution time

Speed-up by adopting look-up table for complex functions w/o calc.

$f(x, y, \ldots\ldots) =$ any combination of calculations

index	Input values v_{in}	Output values v_{out}
0	x_0, y_0	$f(x_0, y_0)$
1	x_0, y_1	$f(x_0, y_1)$
...	...	...
$(n+1)^2$-1	x_n, y_n	$f(x_n, y_n)$

$Enc(x_0, y_1)$ → 1 → $f(x_0, y_1)$ → $Enc(x_0, y_1)$

3 sec for 8K inputs,
186 sec for 64M (8K x 8K) inputs

Infinite
speed-up

(one input ver.[LY19],
multi-input ver.[LY20])

Enabled non-exact match searches

Intel(R) Core(TM) i7-8700@3.20GHz12 with
15.4 GB main memory (8 threads were used)

[LY19] R.Li, Y.shimaki, H. Yamana, "Fully Homomorphic Encryption with Table Lookup for Privacy-Preserving Smart Grid," Proc. of IEEE BITS 2019 (2019)
[LY20] R.Li, Y.Ishimaki, H.Yamana, "Privacy Preserving Calculation in Cloud using Fully Homomorphic Encryption with Table Lookup," Proc. of IEEE ICBDA2020 (2020)
[LY21] R.Li , H.Yamana, "Fast and Accurate Function Evaluation with LUT over Integer-based Fully Homomorphic Encryption," Proc. of the 35th International Conference on Advanced Information Networking and Applications (AINA-2021 (2021)

1-1 Achievements

(M-8) Block-control for ext2/3/4

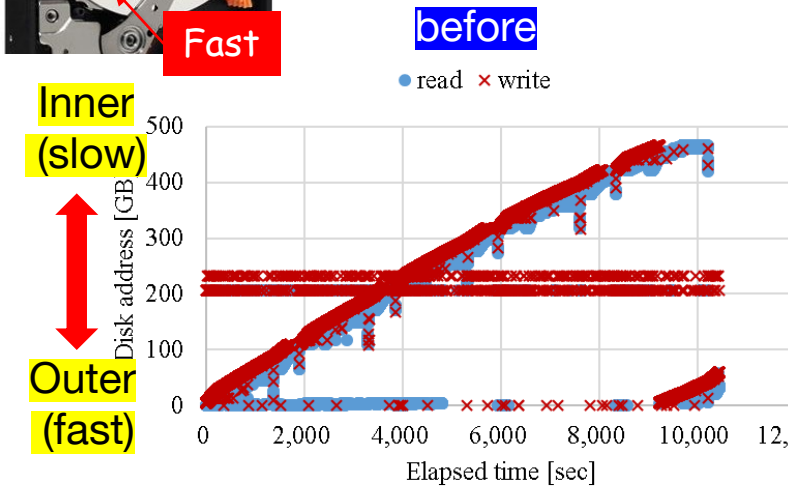
工学院大学

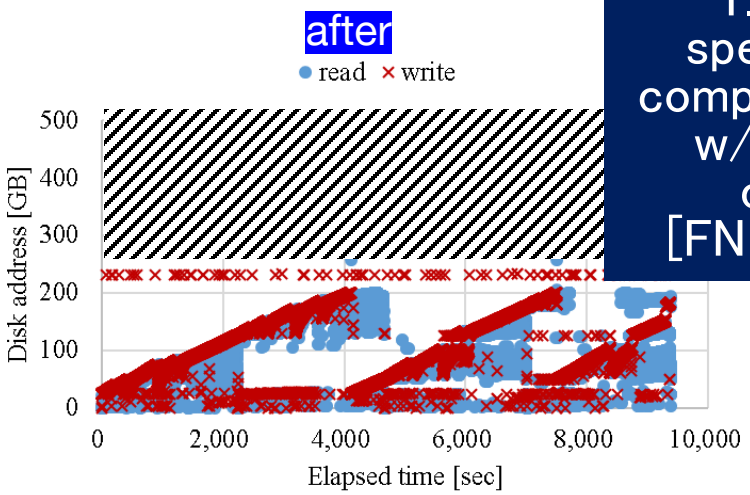
OSS

– Huge temporary data reading/writing from/to hard disks → overheads

Speed-up by placing temporal data into outer zone of platters.

1.2 times speed-up in comparison with w/o block-control [FN17][NF19]



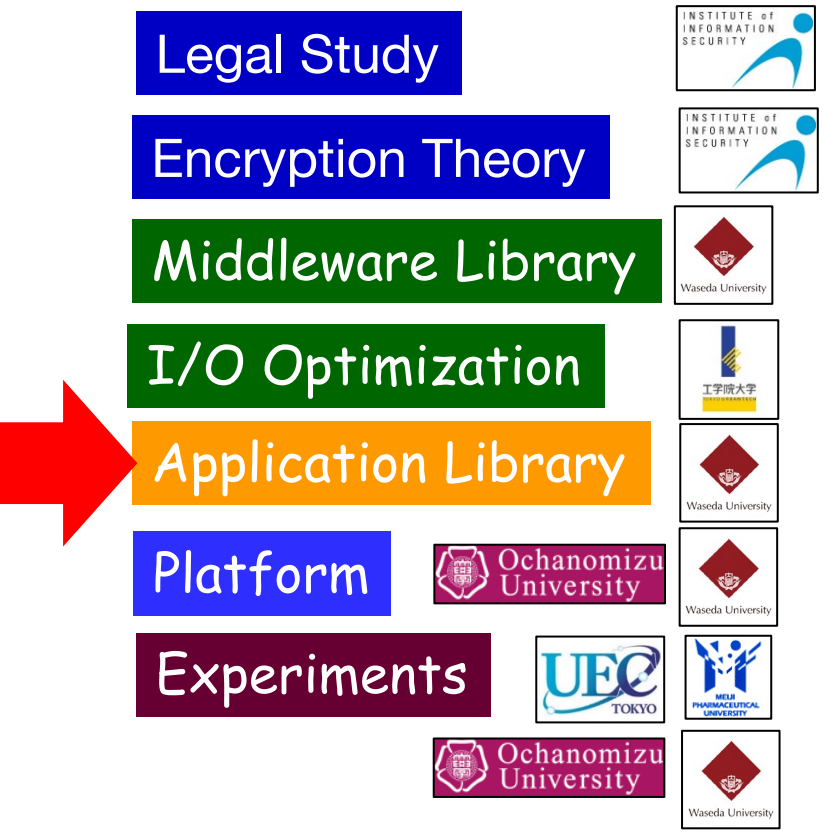


Files are place in the area with lower addresses

[FN17] E.FUJISHIMA, K. NAKASHIMA, S. YAMAGUCHI, "Hadoop I/O Performance Improvement by File Layout Optimization", IEICE TRANSACTIONS on Information and Systems, Vol.E101-D, No.2, pp.415-427 (2017)
[NF20] M. Nakagami, J. Fortes, S. Yamaguchi, "Job-aware File-storage Optimization for Improved Hadoop I/O Performance," IEICE TRANSACTIONS on Information and Systems, Vol. E103.D Issue 10, pp. 2083-2093 (2020)

1-1 Achievements

- This project was organized by seven sub-projects.



Real-world application-specific acceleration

1-1 Achievements

Application Library

- De facto standardizing the implementation protocols

Before

- A few implemented applications w/ FHE

After

- Provided five application models w/ FHE
- Proposed other four application models w/ FHE

OSS

1-1 Achievements

Application Library

Waseda University

OSS		years	speed-up ratio	description
	(A-1) Apriori	2015-2016	**430**	Speed-up of Frequent pattern mining(Apriori) by 1) packing, 2) caching specialized for Aprori, 3) cache pruning, 4) stream processing.
	(A-2) IR	2015-2017	**100**	Speed-up of IR by 1) packing, 2) decresing #bootstrapping with loop unfolding.
* OSS	**(A-3)** OPSI-CA	2016-2018	(5min /100 data)	Outsourced Private Set Intersection Cardinality with FHE(non-interaction required)
	(A-4) Outsourced union	2018-2019	(103s (50data x 2party))	Outsourced union with multi-attribute data by adopting Cartesian-join in Bloom filter
* OSS	**(A-5)** IbsRecommend	2016-2018	(28s /100 POI)	Privacy-preserving recommendation for location-based services by adopting collaborate filteing w/ FHE
* OSS	**(A-6)** NB-Classify	2018-2019	(0.252s /4 classes)	Secure naive bayes classification protocol w/ FHE
* OSS	**(A-7)** PP-CNN	2019-2020	(80.47% precision (CIFAR-10), 0.2s/infer)	Privacy-preserving CNN. (World 2nd ranked precision with fast execution, c.f. World 1st rank:81.5%, 160s/infer)
	(A-8) Power grid	2020	(every 10 sec update)	Privacy-preserving anomaly-based attack detection against data falsification in Smart Grid w/o recall degradation.
OSS	(A-9) Drug Adverse Analysis	2021		will be released

1-1 Achievements

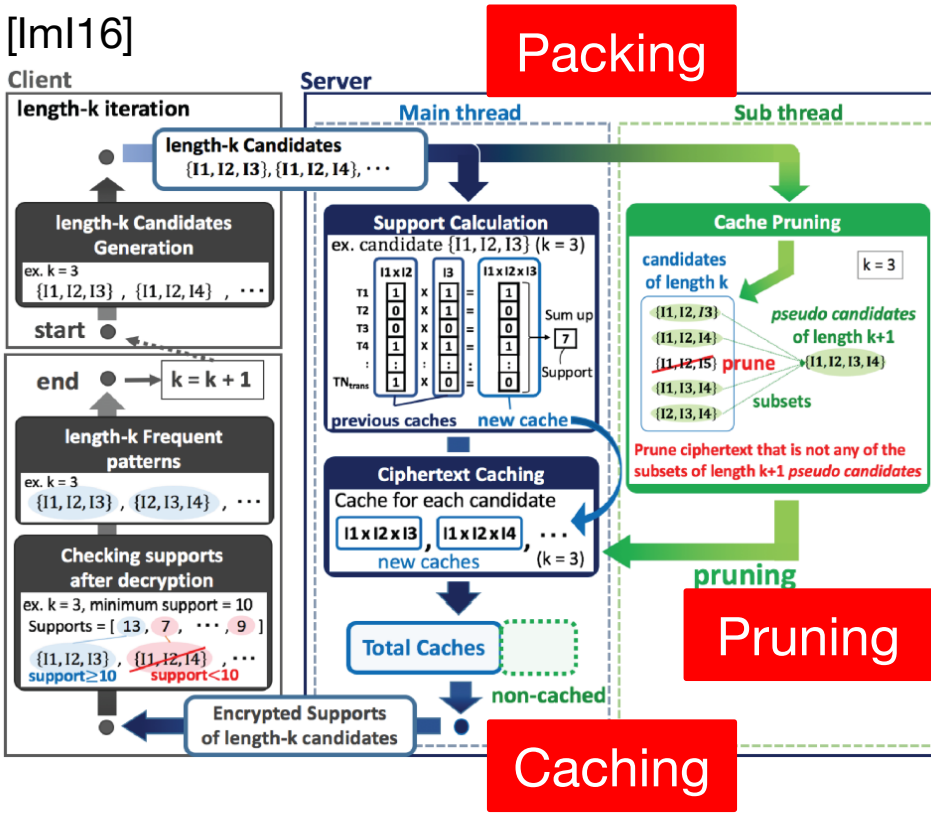
(A-1) Apriori (frequent pattern mining)

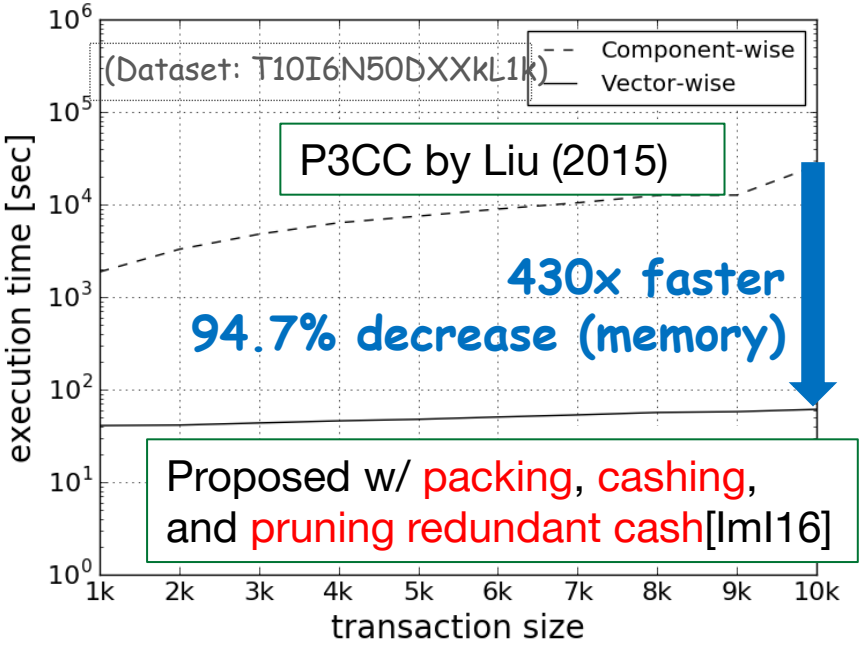
Waseda University

- Frequent Item-set Mining (Apriori) with FHE

speed-up by packing(vector operation), original cashing, and pruning

[ImI16]





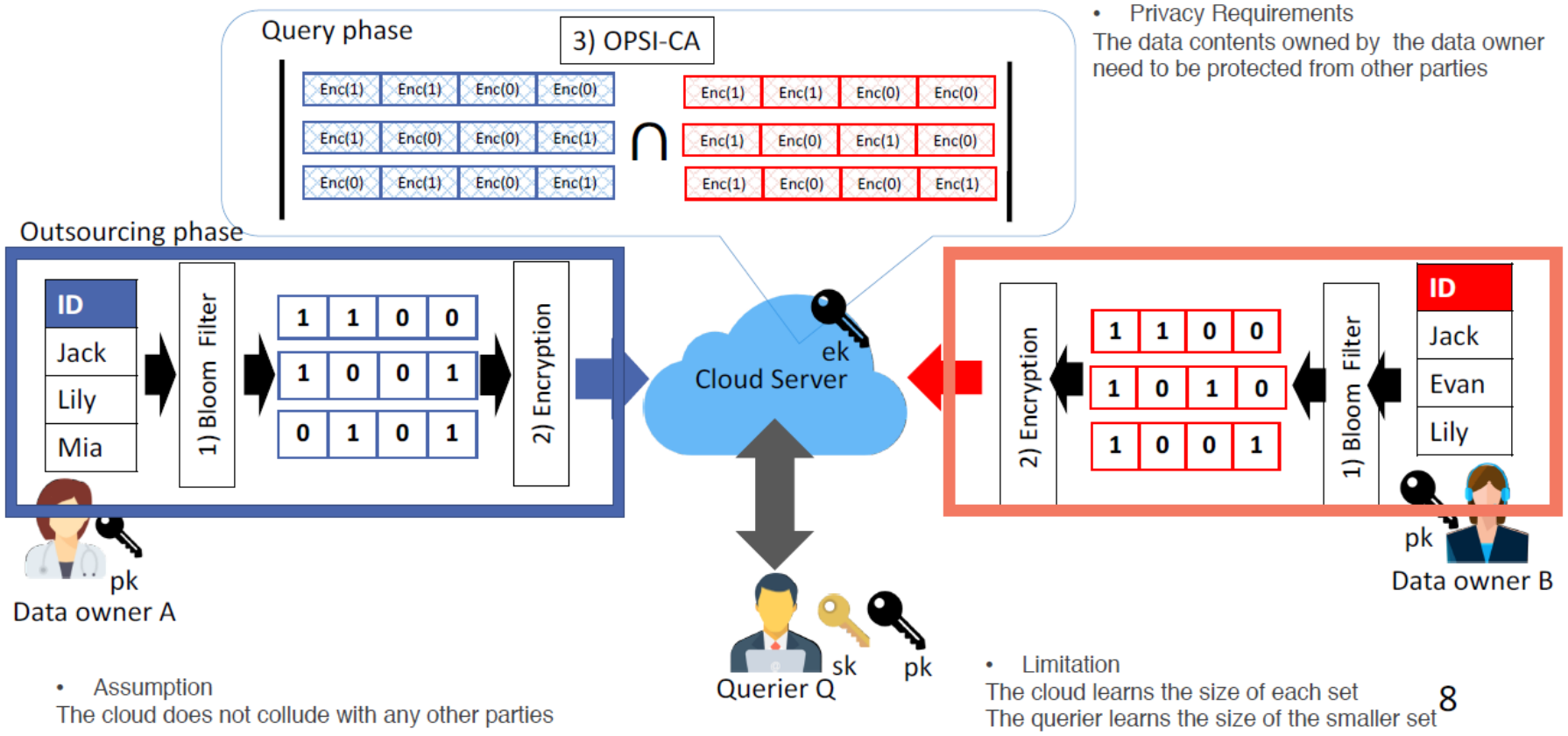
[ImI16] H.Imabayashi, Y.Ishimaki, A.Umayabara and H.Yamana, “Fast and Space-Efficient Secure Frequent Pattern Mining by FHE,” Proc. of IEEE BigData 2016 (2016)

1-1 Achievements

(A-3) Outsourced private set intersection cardinality

Waseda University

OSS



- Privacy Requirements
The data contents owned by the data owner need to be protected from other parties

- Assumption
The cloud does not collude with any other parties

- Limitation
The cloud learns the size of each set
The querier learns the size of the smaller set

8

Non-interaction required. 5min/100data

[TSY18] A.Tajima, H.Sato, H.Yamana:"Outsourced Private Set Intersection Cardinality with Fully Homomorphic Encryption," Proc. of ICMCS2018 (2018)

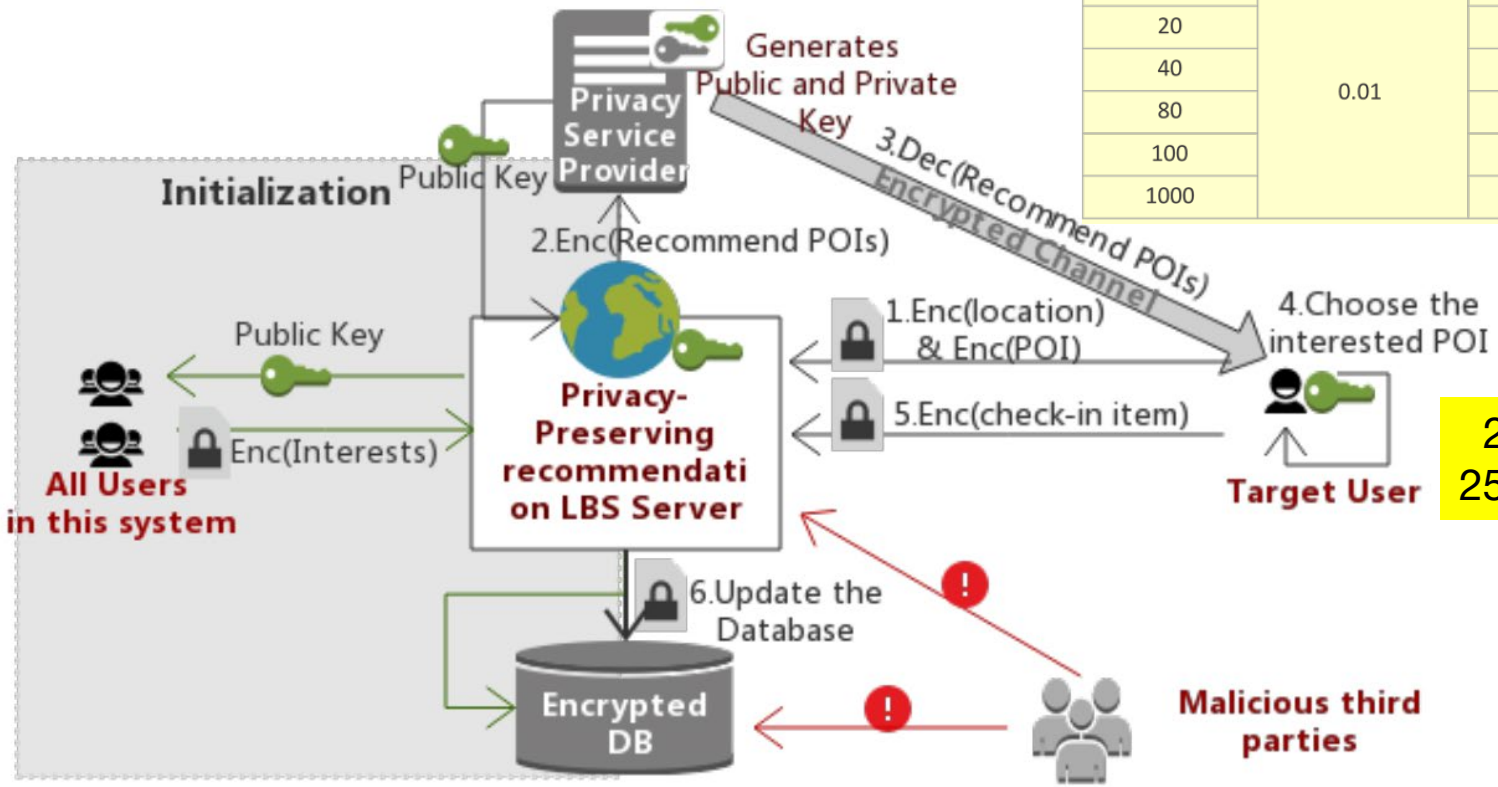
1-1 Achievements

(A-5) Privacy-preserving recommendation on location-based service

Waseda University

OSS

#of elements	T_E : Encryption[s]	T_R : Recommendation[s]	T_D : Decryption[s]
10	0.01	2.73	0.05
20		5.35	0.12
40		10.92	0.20
80		21.91	0.40
100		27.45	0.57
1000		263.96	5.50



28 sec / 100 POIs
259 sec / 1000 POIs

server side: 64-bit CentOS 6, Intel Xeon E7-8880 v3 @ 2.30 GHz x 72, and 1 TB memory.

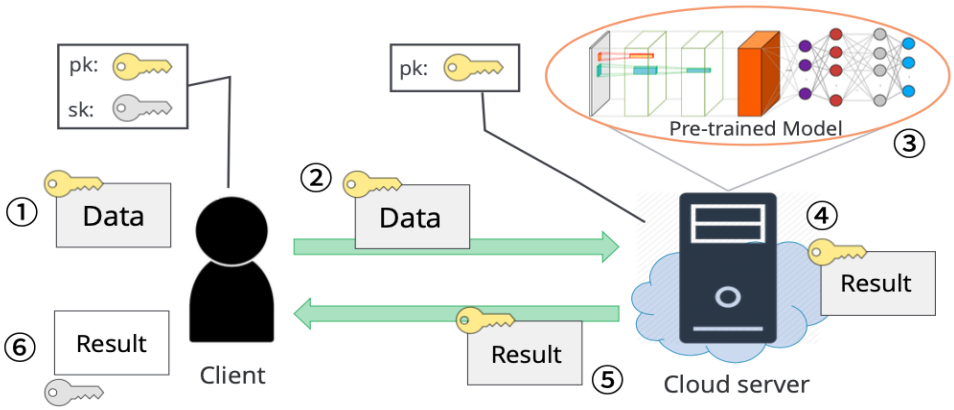
[LI19] Q.Lyu, Y.Ishimaki, H.Yamana: "Privacy-Preserving Recommendation for Location-Based Services," Proc. of IEEE ICBDA2019 (2019)

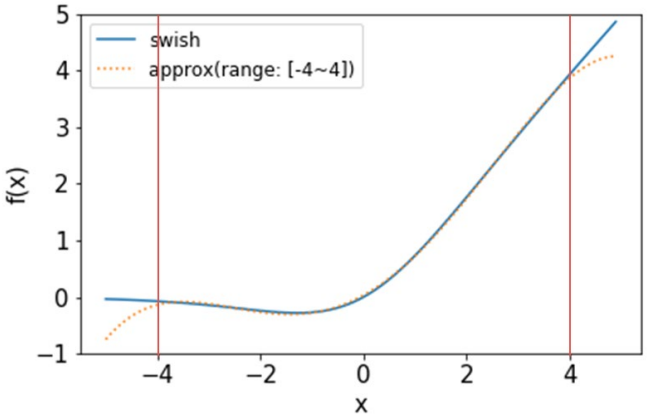
1-1 Achievements

(A-7) Privacy-preserving CNN Inferencing using Approximate Activation Functions

Waseda University

OSS





Background:

Sigmoid, ReLU
Max Pooling

Cannot be adopted

Previous:

$f(x) = x^2$, Average Pooling

Proposed:

adopting Swish $f(x) = \frac{x}{1+e^{-x}}$

and Batch normalization

+ Average Pooling

For $x: [-6, 6])$

$$f(x) = 0.1198 + 0.5x + 0.1473x^2 - 0.002012x^4$$

(CIFAR-10 dataset) 80.47%, 0.2 sec / infer

World 2nd rank in 2020

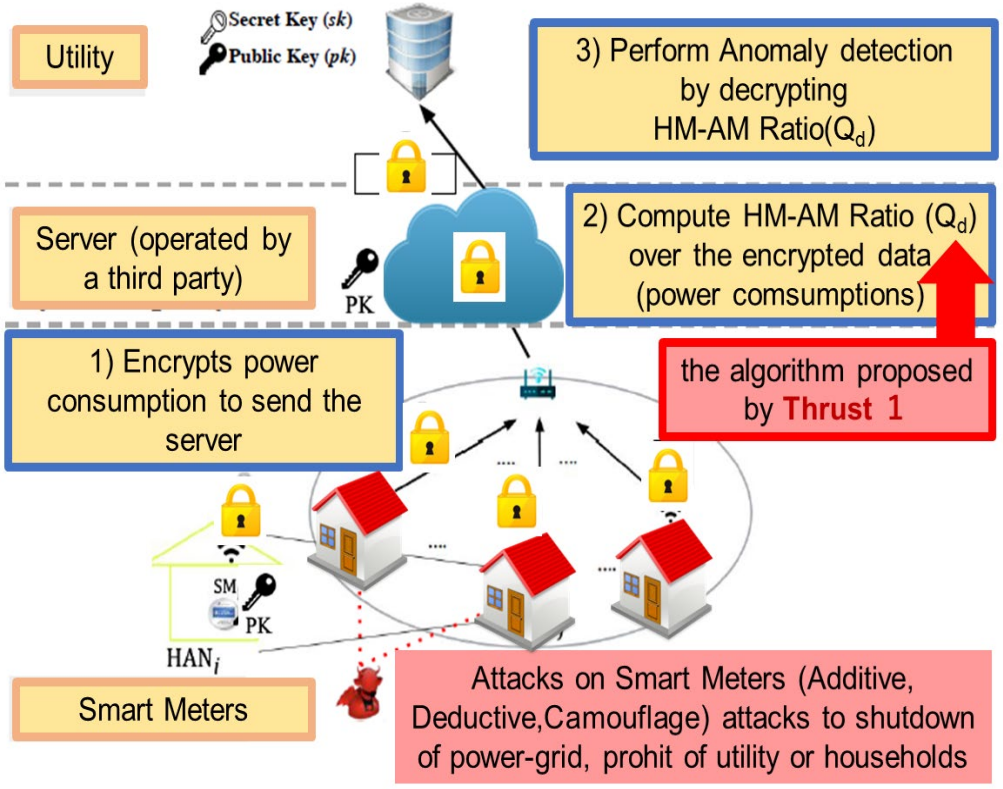
[IS20] T.Ishiyama, T.Suzuki, H.Yamana: "Highly Accurate CNN Inference Using Approximate Activation Functions over Homomorphic Encryption," Proc. of IEEE PSPD2020 (in conjunction with IEEE BigData2020) (2020)

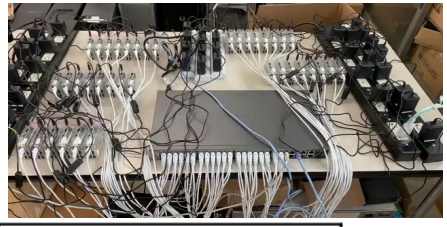
1-1 Achievements

(A-8) Privacy-preserving Anomaly-based Attack Detection

collaborated with

Waseda University, Missouri S&T, W





Execution time

	time [sec]
1) Aggregator	9.935
2) Utility	0.022
Total time	9.957

Aggregator : Intel Xeon E5-1620v4 3.5GHz(single thread)
Encrypted data size: 11,784KB(household→aggregator)
288KB(aggregator→utility)

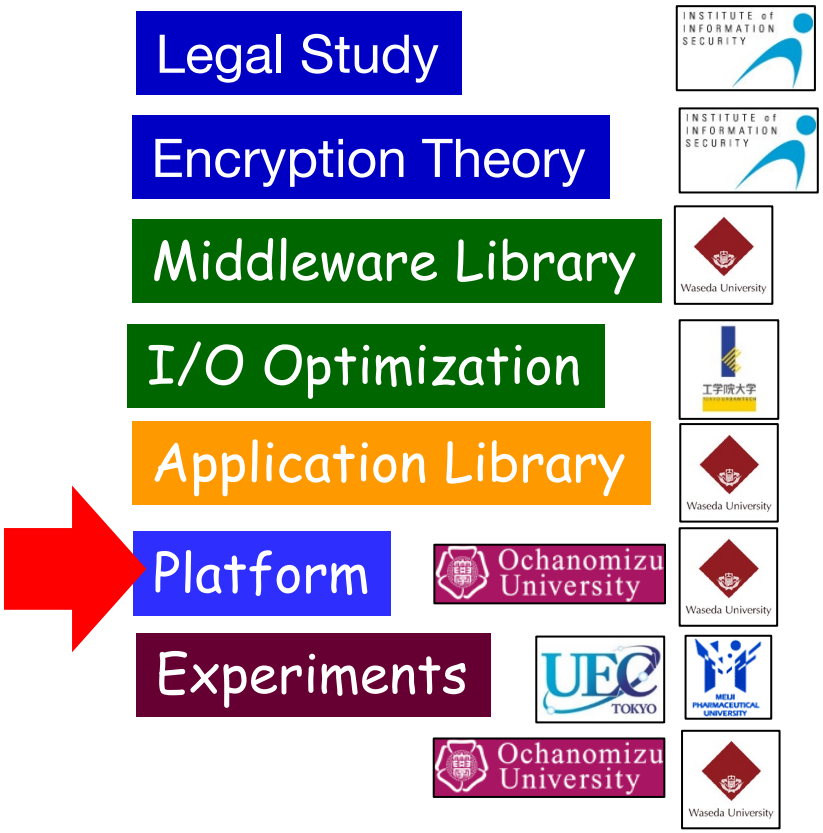
※execution time at power meter is 0.112 sec/data

<Experimental setup>
#households 200
timeslot/day 24
training : US Texs 2014-2015 data
testing : US Texs 2016 data

[IB20] Y.Ishimaki, S.Bhattacharjee, H.Yamana, S.Das : "Towards Privacy-preserving Anomaly-based Attack Detection against Data Falsification in Smart Grid," Proc. of IEEE SmartGridComm 2020 (2020)

1-1 Achievements

- This project was organized by seven sub-projects.



Implementing the platform on cloud

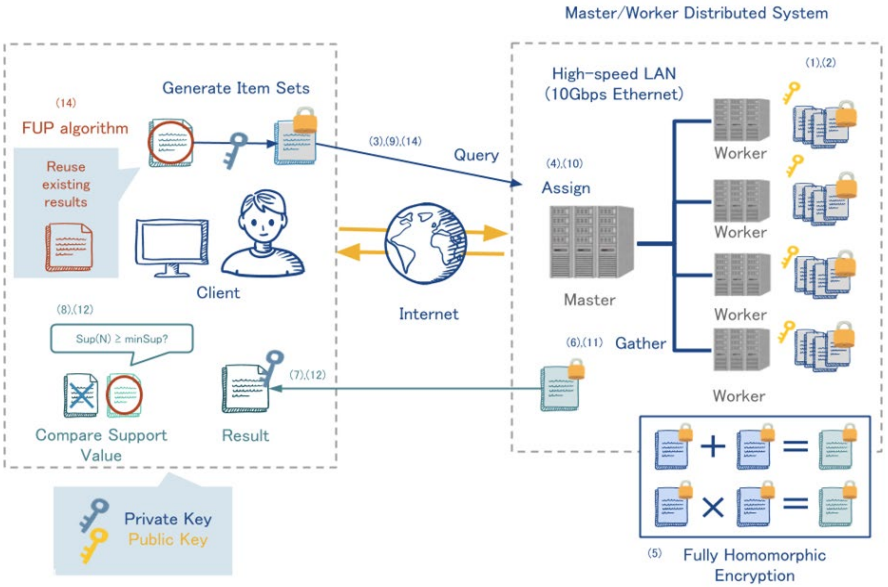
1-1 Achievements

Platform

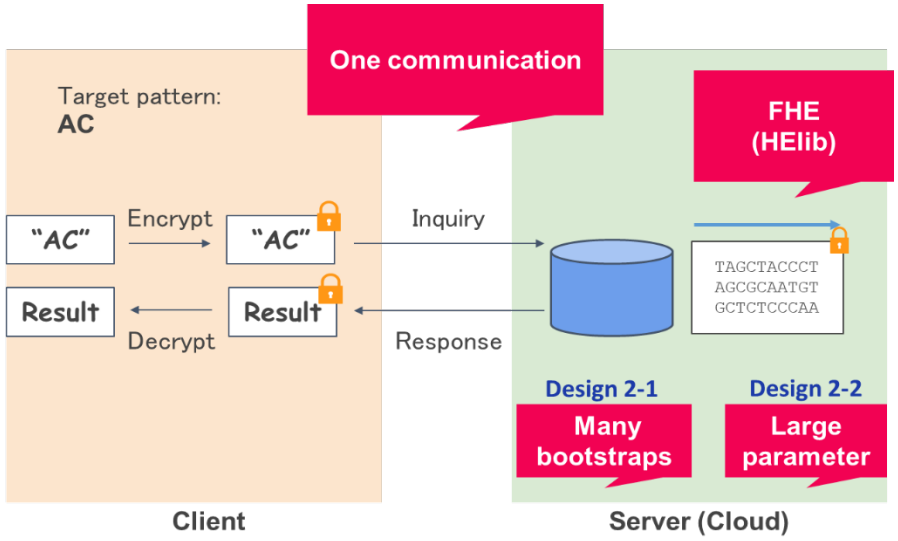
Ochanomizu University

Waseda University

Cloud Platform: Privacy-Preserving Data Mining



Cloud Platform: Privacy-Preserving Search



Apriori

- Database is encrypted and stored at Server side
- Candidate itemset is created at Master/Worker distributed environment
- Frequent itemset is determined at Client side by comparison with minimum support

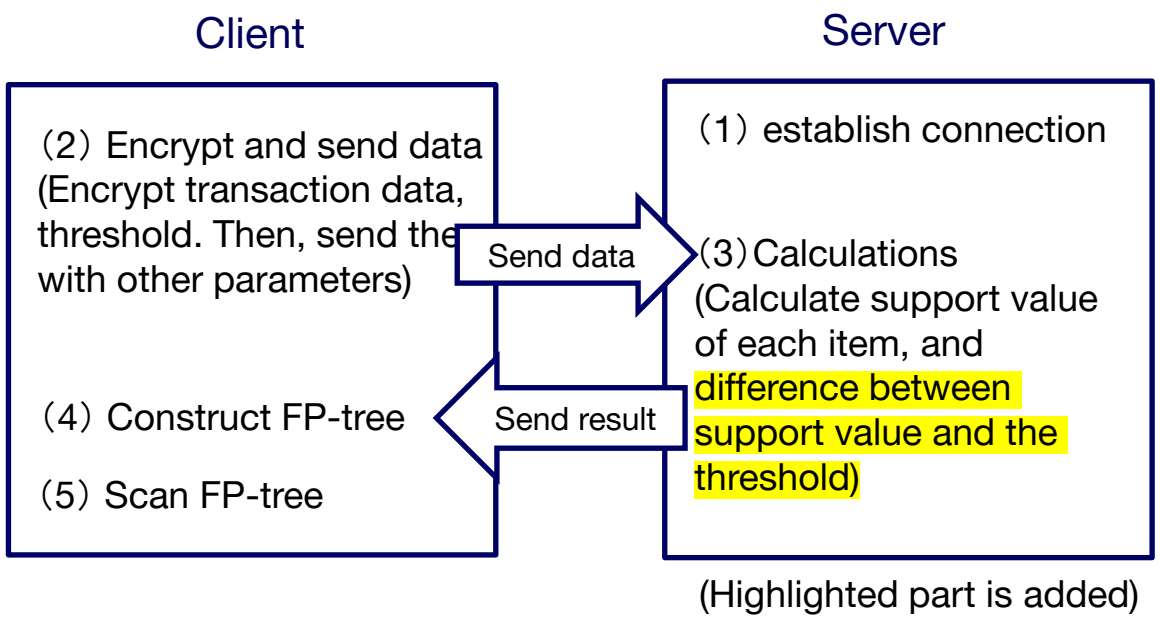
- One-time communication between client and server
- Two implementation versions (w/ bootstrapping, w/o bootstrapping (large leveled HE))

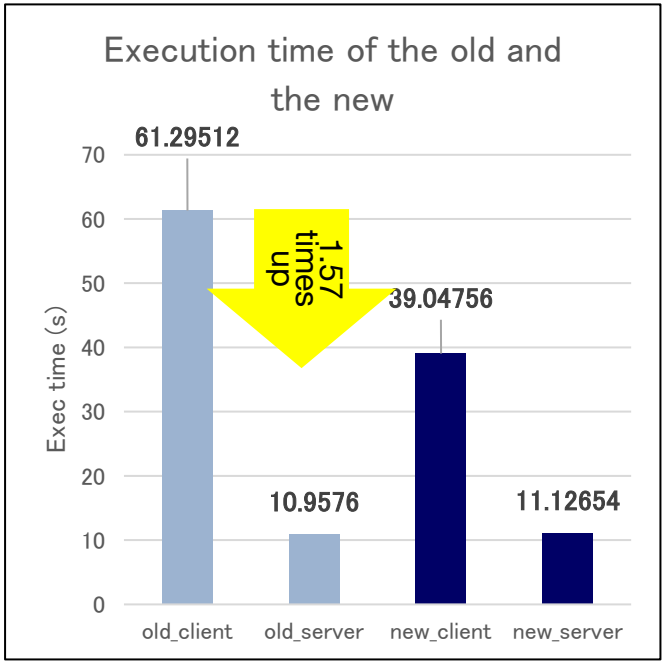
1-1 Achievements

Platform

Ochanomizu University

- Speeding-up of FP-growth





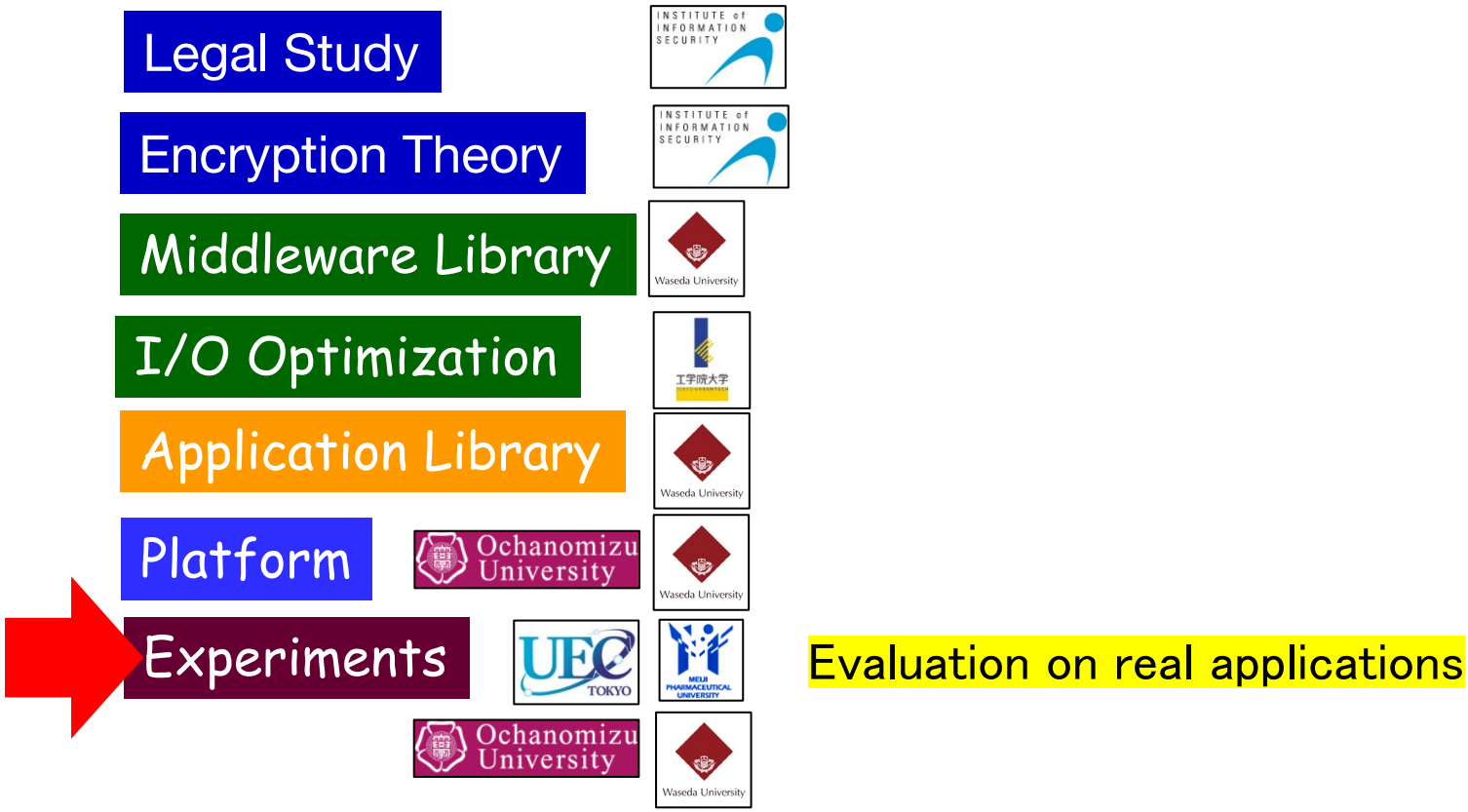
Preprocessing part of FP-growth is executed on the server

- **because calculation cost for mining is high for encrypted data using FHE**

[TO19] M.Tanemura and M.Oguchi, "A Study about FP-growth on a Distributed System Using Homomorphic Encryption," Proc. of SECURWARE2019, pp.96-100 (2019)

1-1 Achievements

- This project was organized by seven sub-projects.



Evaluation on real applications

1-1 Achievements

Experiments | **Drug Adverse Analysis**

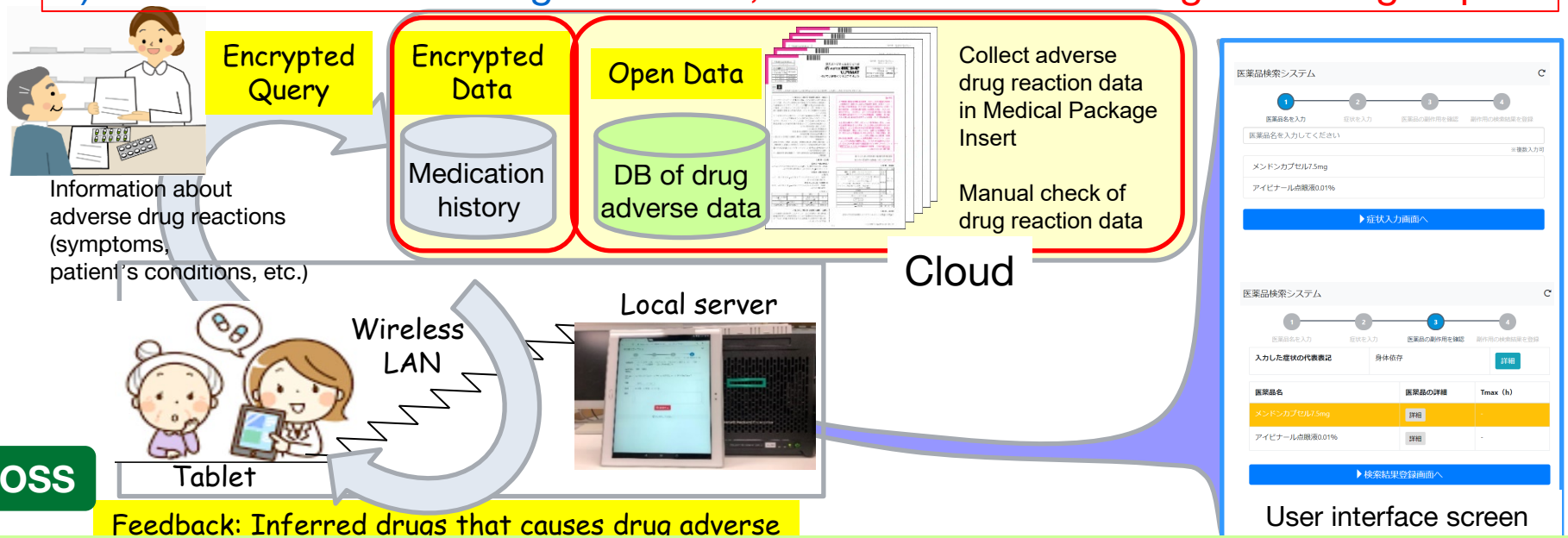
Objective: confirming the adoption of searching w/ FHE to a real application – inferring drug adverse

Dataset Creation for Drug Adverse Analysis

1) Collected adverse drug reaction data in Medical Package 15,498 files

2) Manually checked the drug reaction data 15,498 files

3) Re-classification of drug reaction 9,650 terms into 556 drug reaction group



Y.Jiang, T.Noguchi, N.Kanno, Y.Yasumura, T.Suzuki, Y.Ishimaki, H.Yamana: "A Privacy-Preserving Query System using Fully Homomorphic Encryption with Real-World Implementation for Medicine-Side Effect Search," Proc. of the 21st iiWAS2019 (2019.12)

1-1 Achievements

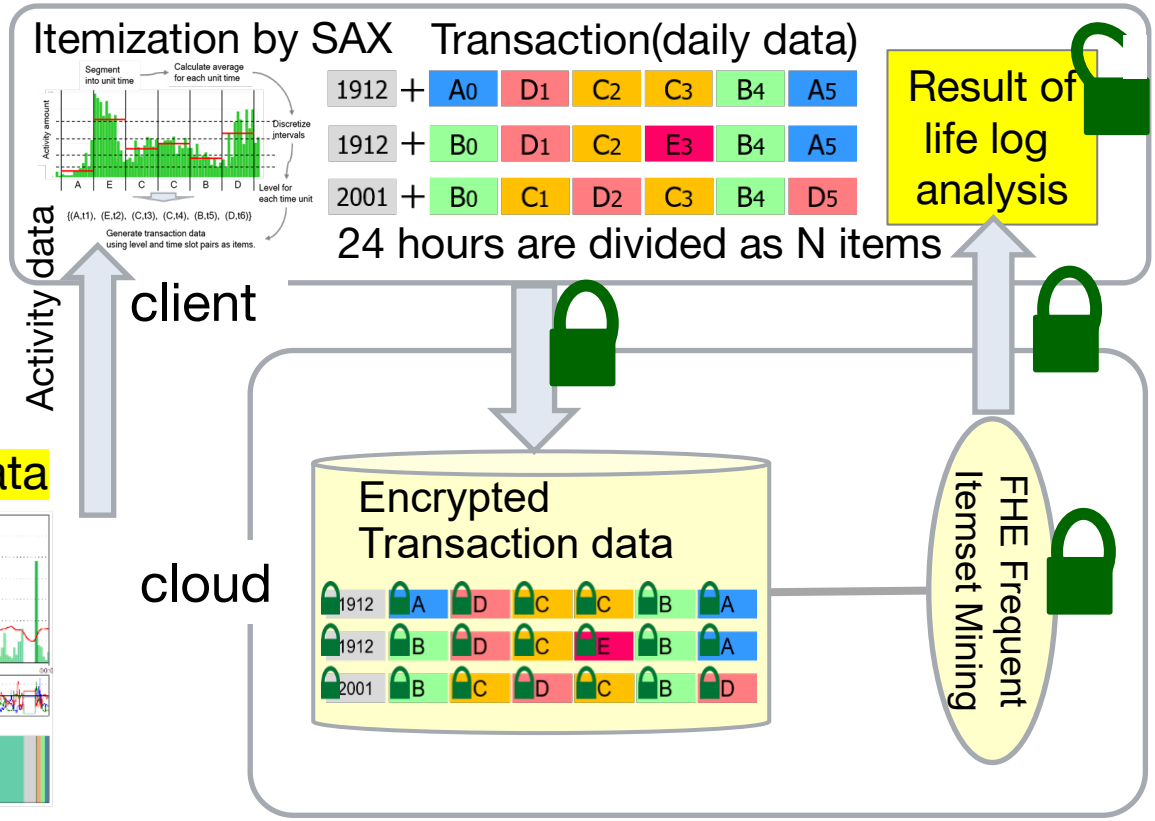
Experiments | Life log Analysis

Objective: confirming the adoption of frequent-itemset mining w/ FHE to a real application – life log analysis

- Collected life log
  1) three-axis acceleration,
  2) activity amount,
  3) segmented motion status

2016/2～now

collected 5,200 person・day data



[SO19] T.Shintani, T.Ohmori, H.Fujita, "Comparison Method of Long-term Daily Life Considering the Manner of Spending a Day," Proc. of IC3K, pp.347-354 (2019)

2. RESEARCH OUTCOME
（研究成果）

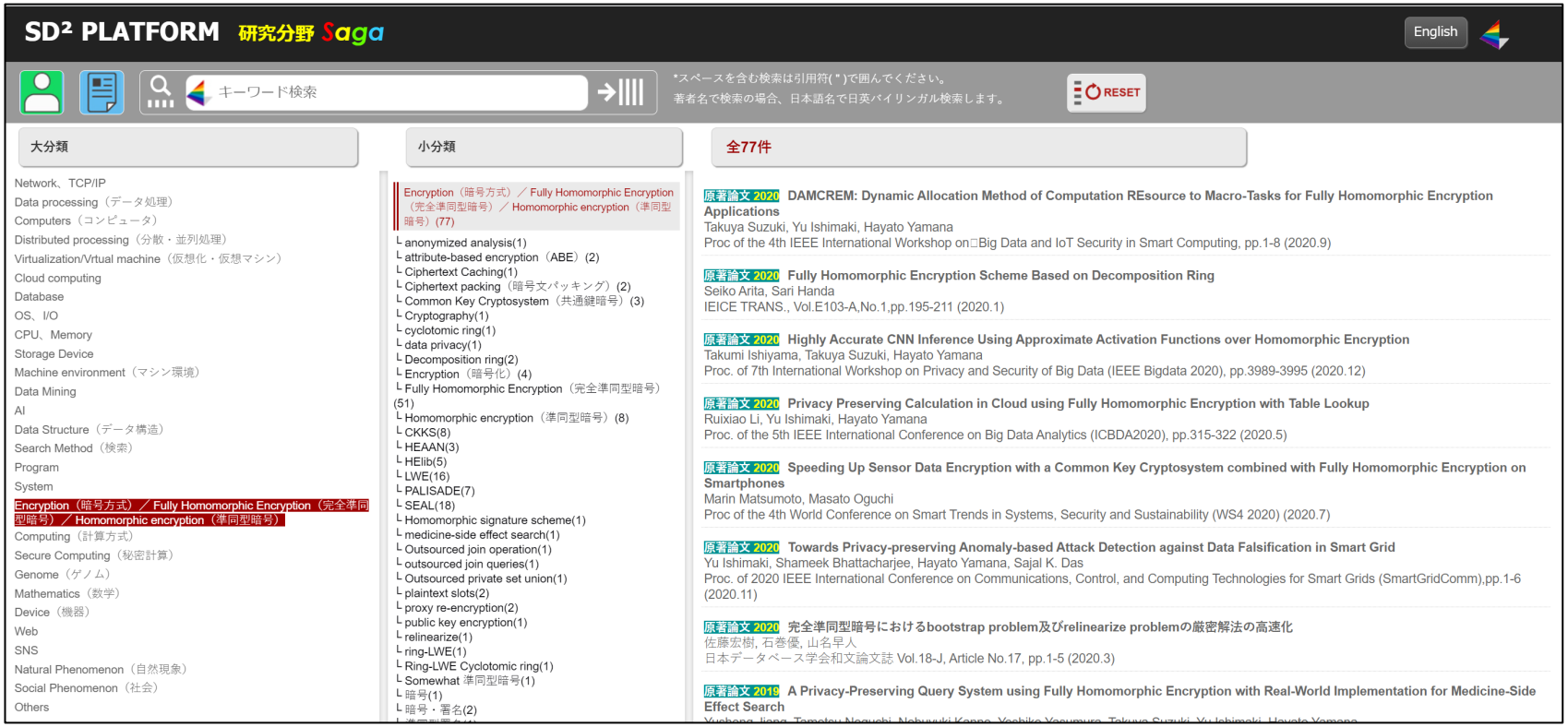
2-1 Publications
2-2 Intellectual property

2-1 Publications

SD² Platform for Integrated Big Data Utilization

Publications

	Invited Talk	Paper	Article /Book	Oral Presentation	Poster	Award	Patent	Public Comment
Total	11	95	7	136	13	26	0	2



2-2 Contribution to Innovation

A doctor course students starts a startup company, EAGLYS.

EAGLYS

Enjoy these and more features with

DataArmor

suite of encryption products:

DataArmor GATE DB

Always-On encryption of databases with sensitive data in non-trusted zones such as the cloud.

DataArmor GATE AI

Allows your AI and ML models to be encrypted and to take advantage of secure computing.

DataArmor ROOM

A safe and secure data access environment for projects with multiple organizations and stakeholders.

2-3 Establishing WS

B) Held four International workshops with other teams

Co-organized by Sajal K. Das (Missouri S&T)
and H.YAMANA(Waseda Univ)

IEEE BITS 2017 (May 2017) SMARTCOMP 2017
Corroborated with Sakuma-team and Miyaji-team

IEEE BITS 2018 (June 2018) SMARTCOMP 2018 Smart Living through Computi
Corroborated with Sakuma-team and Miyaji-team

IEEE BITS 2019 (June 2019) SMARTCOMP 201 Smart Living through Comput
Corroborated with Miyaji-team

IEEE BITS 2020 (Sep. 2020) SMARTCOMP 2020 Smart Living through Computing
Corroborated with Miyaji-team

IEEE SMART COMP 2021

IEEE BITS 2021（Aug.2021）

3. SUMMARY
（総合評価と今後の展開）

Summary

- Exceeded the original goal
  - **2,912 times faster** than original HElib. (new subring homomorphic encryption 26 times faster x Middleware 112 times faster)
- 13 open-source software
  - 1 new homomorphic encryption library
  - 7 middleware libraries to speedup FHE
  - 5 application libraries

THANK YOU

https://www.yama.info.waseda.ac.jp/crest/#open_sources

OPEN SOURCE

- FCMalloc: A Fast Memory Allocator for Fully Homomorphic Encryption
- MCMalloc: A Scalable Memory Allocator for Multithreaded Applications on a Many-Core Shared-Memory Machine
- lbsRecommend: Privacy-Preserving Recommendation for Location-Based Service
- OPSI-CA: Outsourced Private Set Intersection Cardinality with Fully Homomorphic Encryption
- NB-Classify: Naive Bayes with Fully Homomorphic Encryption
- blkctrl: Block Control for Ext2/3/4
- prv_cmp: Private Comparison library
- FHE-Table-Search: FHE-Table-Search function evaluation of fully homomorphic encryption using table lookup
- Subring-HE/FV: The decomposition ring based homomorphic encryption scheme
- PP-CNN: Privacy Preserving CNN Inference over Homomorphic Encryption
- DAMCREM: Dynamic Allocation Method of Computation REsource to Macro-Tasks for Fully Homomorphic Encryption Applications
- BRPS for FHE: Bootstrapping and Relinearization Problem Solver: Solvers for Fully Homomorphic Encryption