

Precision variable anonymization method supporting transprecision computing

Keiya Harada, Hiroaki Nishi, Henri-Pierre Charles

▶ **To cite this version:**

Keiya Harada, Hiroaki Nishi, Henri-Pierre Charles. Precision variable anonymization method supporting transprecision computing. 22nd International Conference on Advanced Communications Technology (ICACT2020), Feb 2020, PyeongChang, South Korea. 10.23919/ICACT48636.2020.9061512 . cea-02556100

HAL Id: cea-02556100

https://cea.hal.science/cea-02556100

Submitted on 27 Apr 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Precision variable anonymization method supporting transprecision computing

Keiya Harada*, Henri-Pierre Charles**, Hiroaki Nishi*

*Graduate School of Science and Technology, Keio University 3-14-1 Hiyoshi Kouhoku Yokohama Kanagawa 223-8522 Japan

**Université Grenoble Alpes CEA, List, F-38000 Grenoble France

harada@west.sd.keio.ac.jp, Henri-Pierre.Charles@cea.fr, west@sd.keio.ac.jp

***Abstract*—Recently, the number of Internet of Things (IoT) sensors has been increasing rapidly; hence, various data are gathered. As a secondary use of the data, they are useful in providing new services, such as the demand response service in the Smart Grid. However, data services cause several problems in preserving privacy and during computation. This study focuses on these two significant problems. First, the invasion of privacy while using the data to provide such new services is problematic. A lot of private information is available in the data. For example, power consumption data may reveal the lifestyle of the residents, and the technique of obtaining information is known as nonintrusive load monitoring. Second, the penetration of IoT devices and sensors increases the computational and communicating energy consumption for processing the data and for providing various services using the data. In this paper, a new method is proposed to solve these two problems. This method is based on the fact that the anonymization process reduces the amount of information itself, as well as the quantity of computational resources required. This leads to a trade-off between anonymization level and computational cost. For example, raw data have a maximum amount of information and maximum computational cost. In contrast, fully generalized data (all zero data) have minimum amount of information and minimum computational cost. Compared to the conventional method, the proposed method demonstrated lower precision and a higher error rate. Therefore, the proposed method aims to control the trade-off and enables the provision of anonymized data with less information, the required anonymity level, and low computational cost compared to the conventional method. The proposed method is practiced using power consumption data gathered from the Urban Design Center Misiono (UDCMi) and the demand response service is evaluated as an experiment using the data. In this evaluation, a simple model of energy consumption was used in the calculation, which uses the required bit width of the arithmetic logic unit (ALU) for providing the service. The computational efficiency of the proposed method was increased by 60% when $k = 2$ and by 65% when $k = 3, 4, 5, 6$ compared to the conventional method. The method can also maintain an acceptable range of service error. The transprecision platform can restrict energy consumption by reducing the bit width of the data. Therefore, the proposed anonymization method can also reduce energy consumption by achieving lower usage of the ALU based on the transprecision architecture.**

Keywords—anonymization, k-anonymity, transprecision computing, approximate computing, power consumption data

I. INTRODUCTION

In recent years, the rapid growth of the Internet of Things (IoT) devices and small sensors has changed our lives dynamically. IoT devices are sensors connected to the Internet, and are often used to send measured data to a dedicated server. A typical example in this application of IoT devices is a smart meter. The smart meter is a sensor for gathering information related to power consumption. The conventional power meter involves the cost of a measurer for checking the power usage. However, a smart meter can cut down the measuring cost, as it enables power consumption data to be sent to power suppliers directly over the Internet. Conventional electricity meters are being replaced by smart meters. For example, in Great Britain, the number of installed smart meters in 2017 was 10 times larger than that in 2012 [1]. In France, 35 million Linky power meters will be deployed until 2021 [2].

Owing to their rapid growth, IoT sensors are beginning to gather various types of data. The data are used for various services. For example, power usage data are stored and used for the demand response service. A demand response service prompts consumers to reduce power consumption and cut down or shift the peak of power consumption. Moreover, the data can be used for recommendations on power usage and dynamic power pricing. The possible ways in which the data can be used are vast, and many new services are being offered.

Although such data bring several benefits to our daily life, there are problems in using the data. In this study, we consider two of these problems:

(1) Privacy violation. The data include a lot of information that may violate someone's privacy. For example, lifestyle-related information such as the number of residents in a household, the time they wake up, and the time they have a shower can be revealed by analyzing the power usage data [3]. This type of intrusion is known as nonintrusive load monitoring (NILM). Therefore, these data cannot be used for secondly use unless the data provider formally permits the use of the data or the data are appropriately anonymized. One well-known example of the problem of publicly releasing private data is the Netflix case [4]. Although Netflix released the movie rating dataset after removing information that could identify users such as name and birth date, sufficient data were published to identify a specific person. In this case, it is not enough to publish data just by removing sensitive information. The data are required to be anonymized by generalizing it such that data safety can be proven statistically. Such anonymization

technologies are gaining attention because they protect private information by appropriately deleting or generalizing the data.

(2) Increase in the processing cost of data. The processing cost brings the problem of increase in energy consumption into focus. In this paper, the term "energy" represents the consumption of computational power and the term "power" represents household power consumption. Energy consumption by information and communication technologies (ICT) has become a significant issue to be discussed because of the fact that it is increasing exponentially [5] and the amount of data involved in IP traffic is increasing year by year [6]. Considering these two facts, it is clear that the energy required for handling data increases with growth in computational networking. Moreover, the immense pressure to develop various services using the data also makes this situation worse. To mitigate this situation, the focus is increasingly shifting towards approximate or transprecision computing to provide a completely efficient and low-energy computing environment.

In this study, the transprecision method and its data processing architecture are proposed to solve the problems of both ensuring privacy and reducing computational cost. The proposed method uses a k -anonymization method for anonymizing data. This proposed method enables a trade-off between information loss and computational cost and accords the required anonymity level to anonymized data at an appropriate computational cost. The constraints of both k and precision differs by the application. k could differ due to where the data will be shared and used for the application. Precision will differ due to the error tolerance of the application. Therefore, a flexible method needs to be developed. Essentially, such anonymization methods ensure data privacy statistically and decrease the total energy consumption during the computation when anonymized data are used because they are generalized by reducing the bits involved. Our concept for reducing total energy consumption is as follows.

- For use in data services, all private data should be anonymized. That is, the cost for implementing anonymization is inevitable in the future and is considered as an unavoidable expenditure.
- Private information need not be stored when its use is not permitted.
- Therefore, private data is anonymized immediately upon creation or storage.
- In the future, the use of anonymized data for providing services safely will become inevitable. The total cost of transprecision computing will be reduced when it is used for anonymized data, as compared to the computational cost of using raw data for providing services.

The EU's General Data Protection Regulation (GDPR) and Japan's amended privacy-preserving law permit the provision of data services subsequent to the appropriate anonymization of private data. To estimate the energy reduction required in data transfer and storage, a baseline evaluation is required by comparing the bit width of the data, particularly focusing on the power consumed by the arithmetic and logical unit (ALU) [7]. In this work, power consumption data (float32) gathered from real residents are used for the evaluation.

II. RELATED WORKS

A. *k-anonymity*

k-anonymity is the standard in privacy protection for data anonymization [8]. Before discussing k-anonymity, we define three related terms here.

1) Data Table

A data table is a table used in creating a database. The row of a data table is termed a tuple and the column is termed a field.

2) Attributes

The heading of each field is termed an attribute. An identifier is an attribute that enables to the detection of privacy information. By combining certain attributes, privacy information can be detected. These attributes are called quasi-identifiers. Identifiers are deleted in anonymization; quasi-identifiers are anonymized using various methods such as masking and averaging.

3) Sensitive attribute

A sensitive attribute is a significant attribute for the data analyst. It cannot be anonymized. In contrast, the attribute that can be anonymized is called a non-sensitive attribute and has the same meaning as a quasi-identifier. The group of tuples that have the same quasi-identifier is called a q^*-block.

The definition of k-anonymity is that at least k tuples are observed in every q^*-block in the data table. For example, if Table 1 is an example of the original data, Table 2 is the 2-anonymized data of Table 1. Zip code, birth date, and gender are the quasi-identifiers and disease is the sensitive attribute in Table 1. The anonymization achieved in Table 2 is achieved by masking the numbers in the zip code from the right side, averaging the gender, and deleting some portions of the birth date.

Table 1 Example of original data

ID	Zip code	Birth date	Gender	Disease
$t1$	0123	1993.12.9	Female	Cancer
$t2$	0124	1993.7.5	Male	Cold
$t3$	0134	1993.12.13	Male	Flu
$t4$	1120	2004.8.12	Male	Cold
$t5$	1121	2004.8.17	Male	Flu

Table 2 Example of anonymized data

ID	Zip code	Birth date	Gender	Disease
$t1$	0***	1993	Human	Cancer
$t2$	0***	1993	Human	Cold
$t3$	0***	1993	Human	Flu
$t4$	112*	2004.8	Male	Cold
$t5$	112*	2004.8	Male	Flu

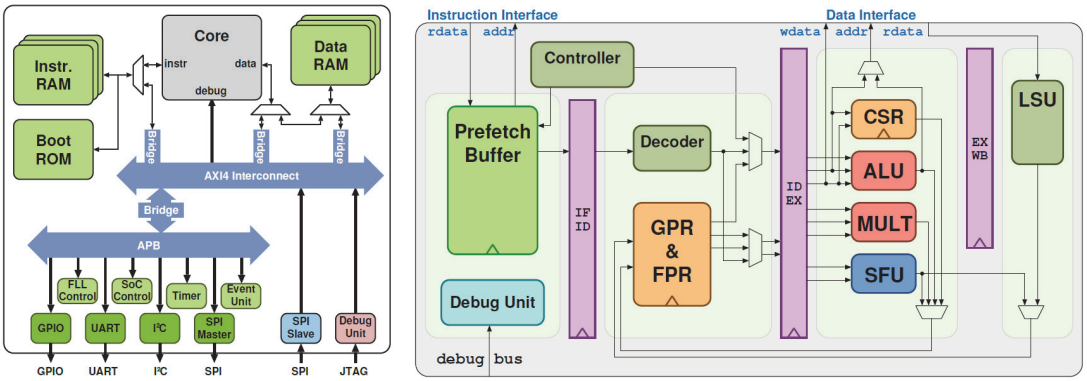


Figure 1 Simplified architectural overview of transprecision platform [7]

Differential privacy (DP) [9] is an alternative method of anonymization. DP has both merits and demerits compared to the *k*-anonymization method. The first merit of DP is that its implementation is simple and involves just adding appropriate noise. As a second merit, DP enables anonymization in on-demand processing. *k*-anonymity is based on batch processing, which requires q^*-blocks to be made from buffered data. Therefore, *k*-anonymity is not suitable for stream processing.

On the other hand, as the first demerit of DP, it has to consider a privacy budget and its distribution. When reusing data, the adjustments and control policy of two parameters of DP (ε and δ) are included in a privacy budget. As the second demerit, it is difficult to anonymize string data in DP. For example, *k*-anonymity can anonymize addresses simply, whereas DP requires appropriate definitions and methods to include noise when anonymizing string data.

Owing to the merits and demerits of differential privacy, both *k*-anonymization and differential privacy are effective methods. This study focuses on using transprecision architecture. As a privacy-preserving method, *k*-anonymization, which uses the cutting-down method in its generalization process, is more closely associated with transprecision computing compared to DP. Accordingly, we used *k*-anonymization as an anonymization method in our work.

B. *Transprecision computing*

Before we discuss transprecision computing, a brief introduction to approximate computing is required. Approximate computing is a computing method that enables efficiency gains in computation by processing "more inexactly" than usual [10]. This efficiency is mainly about reducing energy consumption, as the main purpose of approximate computing is to reduce computational energy consumption and storage size. The challenge is to balance computational errors and the level of approximation. Table 3 lists the various strategies used in approximate computing. Among these, transprecision computing is a type of precision-scaling approximation.

Transprecision computing enables the computation of various precision levels such as 32-bit, 24-bit, and 16-bit. By reducing the precision level of computing, the usage of the ALU is reduced, which leads to a reduction in energy consumption [11]. The balance between precision and computational error is a significant factor in transprecision computing.

A transprecision platform using a RISC machine is proposed in [7]; Figure 1 depicts its simplified architectural overview. This platform can compute in four different data types: float32, float16-alt, float16, and float8. The platform changes the data type from float32 to other data types and computes the result in the changed data type. A unit called SmallFloatUnit is used to carry out this operation. The platform was evaluated in six different applications. The platform was able to reduce energy consumption by 14%–18% compared to that required for achieving full precision. To advance the work of Mach et al. [7], the architecture may be developed further to support more data types. In this study, the focus is mainly on the usage of the ALU because 30% of the energy consumption by the core can be attributed to floating-point operations [12]. This platform propose only discrete data width (8, 16, 32 & 64 bits floating point arithmetic). Our study will not rely on this platform but will propose a more precise study on precision variation.

C. *Problems in related works*

The anonymization method originally focused on preserving privacy, and it was not used as a method for reducing energy consumption because of the calculation cost involved in it. In

Table 3 Strategies and definitions of approximate computing

Strategies	Definition
Precision scaling	Changes the precision of input intermediate operands
Loop perforation	Skipping some iterations of a loop
Load value approximation	Estimates load value using approximate nature to hide cache miss latencyfi
Skipping tasks and memory access	Skip memory references, tasks, or input portions
Using multiple inexact program versions	Utilize multiple versions of application code with different trade-offs
Inexact or faulty hardware	Using inexact/faulty circuits at architecture level

contrast, transprecision computing solves the energy consumption problem and does not consider the privacy problem. A merger of these two techniques can take advantage of both by accepting the computational error involved in processing anonymized data. In combining these two techniques, there is a trade-off between information loss and computational cost. For example, raw data have the maximum amount of information and maximum computational cost. In contrast, fully generalized data (all zero data) have a minimum amount of information and minimum computational cost. The aim of the proposed method is to balance this trade-off.

III. PROPOSED METHOD

The proposed method anonymizes the data using *k*-anonymization while reducing the mantissa bits. This method considers *k* (strength of the anonymization) and the level of precision as its parameters. It involves the three following steps:

1. Use *k*-member clustering to cluster the data.
2. Change the exponent bits in each cluster to satisfy *k*-anonymity.
3. Change the mantissa bits in each cluster to satisfy k-anonymity

Each step of the proposed method is explained in the following sections (A to C).

A. Clustering data

The *k*-member clustering method is used for clustering the data to maintain at least *k* data values in each cluster. This clustering method aims to minimize the defined total distance of the data in each cluster [13]. There are five steps to this method, as follows:

1. Choose the point furthest from a randomly chosen point.
2. Gather *k* data values nearest from the point chosen in 1.
3. Choose the furthest point from the center of the cluster and repeat Step 2.
4. Execute 3 repeatedly until there are less than $k - 1$ non-clustered points.
5. Add each left data value to the nearest cluster.

The distance in this method is the Euclidean distance. k is chosen according to the strength of anonymity required.

Table 4 Raw data table for one cluster

Value	Sign	Exponent	Mantissa	Cluster
18.12	0	1000011	001⋯11000010	1
17.56	0	1000011	000⋯11100001	1
15.17	0	1000010	111⋯01010001	1

Table 5 Anonymized data table for one cluster

Value	Sign	Exponent	Mantissa	Cluster
18.12	0	1000011	001⋯11000010	1
17.56	0	1000011	000⋯11100001	1
16.00	0	**1000011**	**000…00000000**	1

Table 6 Score calculation

Value	Exponent	Mantissa	10010	10011
18.12	10011	0010001000	273	0
17.56	10011	0001100100	200	0
15.17	10010	1110010110	0	105
14.32	10010	1100101001	0	150

Algorithm 1: Calculation of score

E_{xi} = original exponent in data *i*
M_{xi} = original mantissa in data *i*
E_y = updated exponent bit
cluster = the data inside the cluster
score = 0
for *i* in cluster:
    if $E_{xi} \neq \mathrm{E_y}$ and $E_{xi} < \mathrm{E_y}$:
        for j,m in enumerate(M_{xi}):
            if m == 1:
                score += $\mathrm{j} \times 2^{Exi-E_y}$
    elif $E_{xi} \neq \mathrm{E_y}$ and $E_{xi} > \mathrm{E_y}$:
        for j,m in enumerate(M_{xi}):
            if m == 0:
                score += $\mathrm{j} \times 2^{Exi-E_y}$

B. Change exponent bits

After clustering, it is confirmed that all the exponent bits have the same number in the cluster. If these are different, the exponent must be changed. The exponent is changed according to the number of exponent bits existing in the cluster. The exponent bit is changed into the most frequently appearing exponent bit inside the cluster. When the exponent bit is changed, all the mantissa bits are changed into either 0's or 1's. If the changed exponent bit is larger than before, all the mantissa bits are changed to 0's, and if the changed exponent bit is smaller than before, all the mantissa bits are changed to 1's. Table 4 and Table 5 provide an example of this approximation process. The data provided in a table are grouped in the same cluster and are of type float32. The bold and underlined numbers are the changed bits.

However, the exponent bit to be changed cannot be selected if the number of most appeared exponent bits in a cluster is two or more. In this case, the scores are calculated for all exponent bits that appear most frequently. The smallest score is selected for the exponent bits that appeared most frequently. The score provides the error when the original exponent bit is changed into the most frequently appearing exponent bit. The score is calculated using the following algorithm.

When $E_{xi} < E_y$, the error becomes larger based on the increase in the number of zeros in the mantissa because the value becomes larger by changing the exponent bit to a larger one and vice versa. $2^{E_{xi}-E_y}$ is multiplied to j to fix the displacement between the original mantissa and the changed one. Table 6 provides an example. The data in Table 6 are grouped into the same cluster and is of type float16. In Table 6,

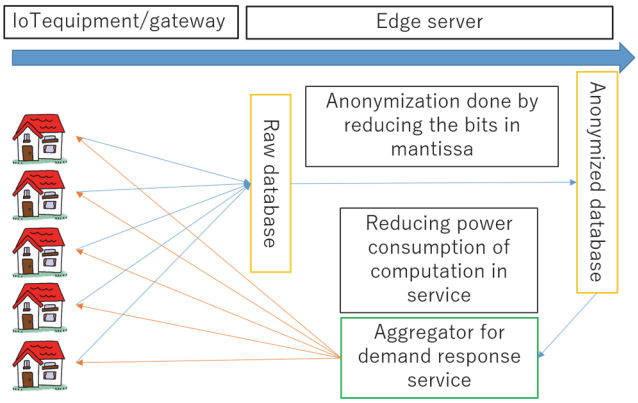


Figure 2 System flow

the 10010 column provides the score when the exponent is changed to 10010 and the 10011 column provides the score when the exponent is changed to 10011. In this case, the score for 10011 is smaller than that for 10010. Therefore, 10011 is chosen as the exponent bit used for changing.

C. *Change mantissa bits*

After changing the exponent bits, the mantissa bits are changed to satisfy k-anonymity. The change of the mantissa bits takes two steps. First, the mantissa bits are deleted according to the precision configured by a user. To delete the mantissa, the concerned mantissa bits are changed to 0, i.e., the number is truncated. Second, the mantissa bits are changed to satisfy k-anonymity by taking the following steps:

1. Set the attention bit to the least significant bit of the numbers in a cluster.
2. If all the attention bits of the mantissa in the cluster have the same number, the requirement of anonymization is satisfied.
3. If these are not the same, count the number of attention bits in a cluster separately. If there is only one maximum number among the counted bits, change the mantissa bit to that number for all the attention bits in the cluster.
4. If not, use the one-dimensional error diffusion method [14] for selecting the number to change all the attention bits into.
5. Repeat Steps 2 and 3 by changing the attention bit to the next significant bit.
6. Repeat Step 4 until the attention bit reaches the most significant bit of the mantissa.

Table 7 Example of mantissa anonymization

Value	Sign	Exponent	Mantissa	Cluster
17.23	0	1000011	000…11000**010**	1
17.23	0	1000011	000…11**0**000**001**	1
17.23	0	1000011	000…**11**000**001**	1

Table 8 Example of mantissa anonymization using error diffusion method

Value	Exponent	Mantissa
16.29	10011	**1**0101**1**00**1**00
16.29	10011	**1**0**1**0100100
16.29	**10011**	1**0**1010**1**0**1**0
16.29	**10011**	1**01**010**01**0

Table 7 provides an example of these steps using the data provided in Table 4. In this case, the given precision is 29 bits. The data in the table belongs to the same cluster and is of type float32. The bold and underlined numbers denote the changed numbers.

The error diffusion method is usually employed in image processing. Several error diffusion methods have been developed, such as that proposed by Floyd and Steinberg [15]. However, this method used the one-dimensional error diffusion method because the mantissa bit provides one-dimensional data. The following equation provides the actual binary number:

$$b(i) = \text{step}\left[f(i) - \sum_{n=1}^{i-1} e(i-n)\right] \quad (1)$$

$$e(i) = m_x \times 2^i + m_y \times -2^i$$

where i is the order of the mantissa bit and $f(i)$ is the difference between the number of 0's and 1's in the ith mantissa bit. $f(i)$ is always 0 in our method because the error diffusion method is used only when the counted numbers of 0's and 1's are the same. $e(i-n)$ is the error in the $(i-n)$th bit. m_x is the number of bits changed from 0 to 1 and m_y is the number of bits changed from 1 to 0. The step function in Equation (1) yields zero when the argument is negative, and vice versa. Table 8 provides the result when the mantissa bit of Table 6 is anonymized using the error diffusion method. The data in the table are assigned to the same cluster and is of float16 type. The bold and underlined numbers represent the changed numbers in this process. In the table, the 8th bit is changed to 1 because the errors in the 9th and 10th bits are –4 and –1, respectively, and the sum of the errors is –5.

IV. Evaluation

A. *System flow, dataset, conventional method, demand response service, and error rate*

The system flow for the evaluation is depicted in Figure 2. The edge server gathers the power consumption data from the smart meters. and anonymizes them using the proposed method. These anonymized data are finally used for the demand response service.

In this evaluation, the demand response service sends demand response messages to households that are supplied by the utility to reduce the power consumption. We assume that all

Table 9 Difference of ER with or without using scoring and error diffusion methods

k	2	3	4	5	6	7	8
D (%)	22.8	1.13	6.84	0.18	3.05	0.00	0.41

households accept the message and reduce their power consumption. The message is issued every 30 min in Japan [16]. In this evaluation, the demand response message is created to cut down the power consumption of the top 15% power users. The curtailment is done by restricting the power consumption of the top 15% power users to 85% of the maximum power consumption [17].

The data used in the experiment were gathered from seven households taking part in the smart community project conducted in Saitama, Japan referred to as Urban Design Center Misono (UDCMi). The data used in this experiment are two weeks of power usage data from October 12, 2018 to November 22, 2018. Each set of data values for a week is assumed to be data from a different household to enlarge the number of household data values for anonymization. Consequently, the data are composed of power usage values of 6 days from 49 households. The conventional anonymization method has three steps. First, the data are sorted by their values. Second, a set of top k data values are selected as a group. Third, all values in the group are changed into their mean values. These three steps are used continuously for all values.

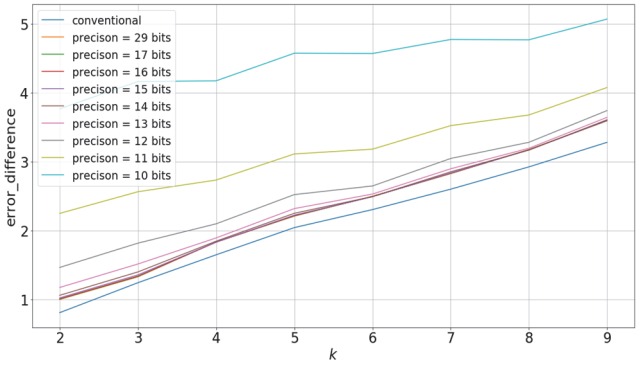


Figure 3 Relationship between ***ER*** and ***k***

Table 10 Coefficient for each precision value

Precision	29	19	14	13	12	11	10
C $(\times 10^{-3})$	3.70	3.69	3.63	3.52	3.24	2.6-	1.85

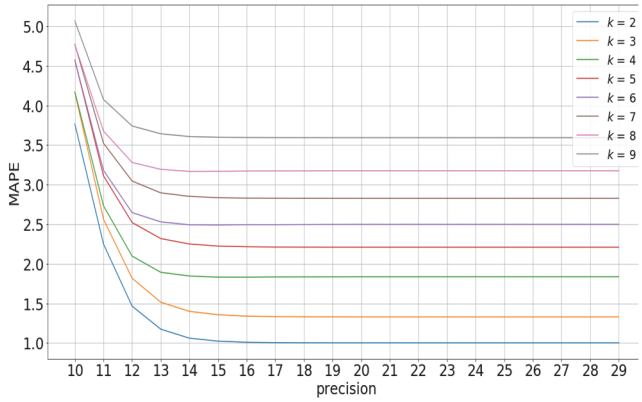


Figure 4 Relationship between ***ER*** and precision

Generally, for evaluating the quality of anonymized data, information loss (IL) is frequently used as the index of quality. It represents the amount of information lost in the process of anonymization. Normally, IL takes a value between 0 and 1. However, in this study, we extended the IL to the error rate (ER). For computing the ER, the mean absolute percentage error ($MAPE$) was used. The $MAPE$ is expressed as follows:

$$\boldsymbol{MAPE} = \frac{\sum\left|\frac{(y_t - \hat{y}_t)}{y_t}\right|}{\boldsymbol{n}} \times \mathbf{100} \quad (\boldsymbol{y_t \neq 0}) \tag{2}$$

where y_t denotes the original data, $\hat{y}_t$ denotes the anonymized data, and n denotes the number of the data points.

B. Model based on bit width

For the model based on bit width, Equation (3) is used. U represents the extent of usage of ALU, N is the number of floating operations, and $size_m$ is the size of the mantissa.

$$U = N * size_m \tag{3}$$

In this paper, the number of calculation cycles is modeled as the number of floating operations. This model is based on the calculation cycle of a typical RISC machine. Therefore, one cycle of instructions stands for one clock cycle. The size of the mantissa can be interpreted as the length of the critical path. This critical path stands for the computational complexity in one clock cycle. Therefore, the multiplication of the number of clocks and the computational complexity in each clock can represent the usage of ALU in a computation.

C. Evaluation of scoring and error diffusion methods

Table 9 displays the difference in ER with or without using the scoring and error diffusion methods. In Table 8, D denotes the amount by which ER has decreased using the scoring and error diffusion methods. It can observed in Table 9 that the error diffusion method leads to significant improvement when k is even and small. This is because the number of most appeared mantissa and exponent bits is often more than 1 when k is even and small. When $k = 7$, the scoring method was not used and the error diffusion method was used only a few times. Therefore, the error rate was not impacted.

D. Evaluation of ER vs k, and vs precision

Figure 3 depicts the relationship between the precision and error rate. It is clear from Figure 3 that, compared to the conventional method, the proposed method provides less precision and higher error rate. This means that the proposed method created anonymized data with lesser information at a lower computational cost. Therefore, the proposed method can balance the trade-off between the information loss and the computational cost. In Table 10, C represents the coefficient when we linearly approximate the results shown in Figure 3. C is referred to as the rate of ER increasing with k. From Table 10, the rate of ER increasing with k becomes less when the precision becomes lower. This is because some clusters are already anonymized by deleting only the mantissa bit. For example, if the data in one cluster had the same exponent and

the same upper 10 mantissa bits, the data in the cluster will be all the same by deleting 13 mantissa bits.

Figure 4 depicts the relationship between k and the ER. The figure identifies two specific features. First, 16 bits of the mantissa can be deleted in the anonymization with only a small error difference compared to the conventional method. The ER is 0.14% larger than that in the conventional method at an average when the precision is 16 bits. Therefore, the data can be reduced to float16 type with a small difference in the ER compared to that obtained in the conventional method. Second, the precision becomes smaller as k increases when the ER rises by more than 10% compared to the conventional method. Let pre_r be the precision when the ER rises by more than 10% compared to the conventional method. pre_r has 13 bits when $k = 2, 3$; 12 bits when $k = 4, 5$; and 11 bits when $k = 6, 7, 8, 9$. This is because the error of anonymization exceeds the error of lowering the precision.

E. Evaluation of error in service and percentage of ALU usage

Figure 5 displays the relationship between the error in the demand response service and the percentage of ALU usage compared to those in the conventional method. The error is calculated using the following equation:

$$\text{Error of service} = \frac{\sum_{i=1}^{n} \left| x_{real,i} - x_{anony,i} \right|}{\sum_{i=1}^{n} x_{real,i}} \times 100$$

where n denotes the number of households, $x_{real,i}$ denotes the raw power consumption data after the demand response service in the ith home, and $x_{anony,i}$ denotes the anonymized power consumption data after the demand response service in the ith home.

According to the Electricity Business Act of Japan, the error of balancing between the demand and supply of power consumption has to be lower than 3%. When we consider that the perfect working of the demand response cuts down the energy consumption, the error only occurs when the anonymization and transprecision mechanisms are used. In this case, the error of the service using these mechanisms has to be lower than 3%. It can be seen in Figure 5 that the threshold was not accomplished at any precision when $k = 6, 7, 8, 9$. For example, the demand control service can be achieved at a precision of 12 bits and the ALU usage can be reduced to 15% compared to the conventional method operation when $k = 2$. In the case of a 13-bit precision operation, the ALU usage can be reduced to 18%. In the case of certain values of k, the errors of service are eliminated by reducing the precision. This is because the anonymized power consumption happens to be greater than the real power consumption. Therefore, by ignoring the precision, the anonymized value is reduced and approaches the actual amount of power consumption.

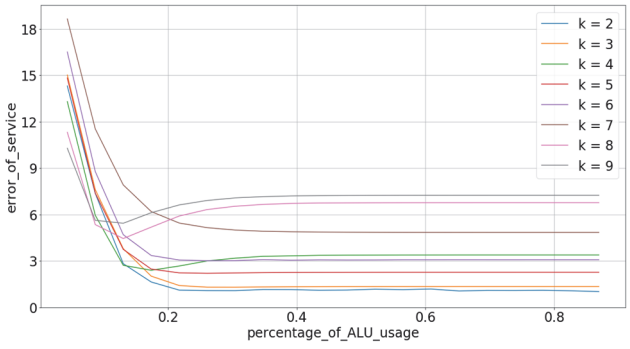


Figure 5 Relationship between error in service and percentage of ALU usage compared to conventional method

V. CONCLUSION

In this study, a precision-scalable anonymization method was proposed for transprecision computing. To the best of our knowledge, this is the first work that links two scientific domains: transprecision and anonymization. This method balances the trade-off between information loss and computational cost by deleting the mantissa bits for anonymizing the data. The effectiveness of the error diffusion method and the relationship between each parameter and ER were evaluated. The ER decreased when k is small and even when the scoring and error diffusion methods were used. When providing the demand response service, the ALU usage was calculated using the model with various precision values. The usage of the ALU was reduced to 15% when $k = 2$ and 18% when $k = 3, 4, 5$ within the acceptable service error. In the transprecision platform, the energy consumption of the computation is reduced owing to the reduction of the bit width [7]. Therefore, the energy consumption of the computation decreases by reducing the bit width and the ALU usage.

For future work, three tasks are essential. First, models should be made for modules other than the ALU. In this study, the usage of the ALU was measured. However, modules such as memories should also be simulated. Second, this method should be applied to multidimensional and string data. Currently, it can only be applied to one-dimensional numerical data. Third, a new data expression should be formulated to reduce storage requirements. Using our proposed method, there is a chance that the data can be compressed while anonymizing. However, for the computer to understand this compressed data, a new data expression is necessary.

ACKNOWLEDGMENT

This work was supported by MEXT/JSPS KAKENHI Grant (B) Number JP16H04455 and JP17H01739. Also, this work was supported by JST CREST Grant Number JPMJCR19K1, Japan.

REFERENCES

[1] BEIS, “Smart Meter Report Q1 2018,” March 2018.

[2] Smarter Together, “Report on Deployment of Linky Smart Power Meters in the Area,” no. 691876, pp. 1–32, 2018.

[3] NIST, "Guidelines for Smart Grid Cyber Security," vol. 1, September. 2010.

[4] A. Narayanan and V. Shmatikov, “Robust De-Anonymization of Large Sparse Datasets,” Proc. - IEEE Symp. Secur. Priv., pp. 111–125, 2008.

[5] ICT, “ICT Energy Strategic Research Agenda,” no. 611004, pp. 1–60, 2016.

[6] Cisco, “The Zettabyte Era: Trends and Analysis,” Cisco Vis. Netw. Index, May 2015, pp. 1–29, 2015.

[7] S. Mach, D. Rossi, G. Tagliavini, A. Marongiu, and L. Benini, “A Transprecision Floating-Point Architecture for Energy-Efficient Embedded Computing,” Proc. - IEEE Int. Symp. Circuits Syst., May, 2018.

[8] L. Sweeney, “k-Anonymity: A Model for Preserving Privacy,” Int. J. Uncertainty, Fuzziness Knowl-Based Syst., vol. 10, no. 05, pp. 557–570, 2002.

[9] C. Dwork and A. Roth, “The Algorithmic Foundations of Differential Privacy,” Found. Trends Theor. Comput. Sci., vol. 9, no. 3–4, pp. 211–487, 2013.

[10] S. Mittal, “A Survey of Techniques for Approximate Computing,” ACM Comput. Surv., vol. 48, no. 4, pp. 1–33, 2016.

[11] A. C. I. Malossi et al., “The Transprecision Computing Paradigm: Concept, Design, and Applications,” Proc. 2018 Des. Autom. Test Eur. Conf. Exhib. DATE 2018, Jaunuary 2018, pp. 1105–1110, 2018.

[12] G. Tagliavini, S. Mach, D. Rossi, A. Marongiu, and L. Benin, “A Transprecision Floating-Point Platform for Ultra-Low Power Computing,” Proc. 2018 Des. Autom. Test Eur. Conf. Exhib. DATE, January 2018, pp. 151–1056, 2018.

[13] J.-W. Byun, A. Kamra, E. Bertino, and N. Li, “Efficient k-Anonymization Using Clustering Techniques,” pp. 188–200, 2007.

[14] M. Kowalczyk, M. Martínez-Corral, T. Cichocki, and P. Andrés, “One-Dimensional Error-Diffusion Technique Adapted for Binarization of Rotationally Symmetric Pupil Filters,” Opt. Commun., vol. 114, no. 3–4, pp. 211–218, 1995.

[15] V. Ostromoukhov, “A Simple and Efficient Error-Diffusion Algorithm.”

[16] Martyn Williams, “Electricity System and Market in Japan,” IT World, March, 2011.

[17] O. Kengo and N. Hiroaki, “Big Data Anonymization Method for Demand Response,” Int. Conf. Internet Comput. Big Data, pp. 76–82, 2014.



Keiya Harada received his BS degree from Keio University, Japan, in 2018. He is now a master student in Keio University and a member of Nishi Laboratory, Keio University. He is currently interested in data privacy.



Henri-Pierre Charles is research director at CEA-LIST since September 2010. He was previously assistant professor in Versailles University during 17 years, authorized to supervise PhD students (HDR) since 2008. During this period he participated in many European and national projects, mainly in the high performance computing domain. He has supervised many PhD theses on code optimizations and High performance computing. Since his arrival in CEA he applied his knowledge on multi-processor system on chip (MPSoC) and embedded systems.



Hiroaki Nishi received his PhD from Keio University, Japan, in 1999. He is a professor at Keio University since 2014. He is the president of the Omotenashi ICT Consortium and the chairperson of Misono Town Management. He also chairs the IEEE P21451-1-6 Standard Working Group. The main theme of his research is building total network systems including the development of smart community data infrastructure.