



Precision variable anonymization method supporting transprecision computing

Henri-Pierre Charles, Keiya Harada, Hiroaki Nishi

► To cite this version:

Henri-Pierre Charles, Keiya Harada, Hiroaki Nishi. Precision variable anonymization method supporting transprecision computing. 2020. cea-02556188

HAL Id: cea-02556188

<https://cea.hal.science/cea-02556188>

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*

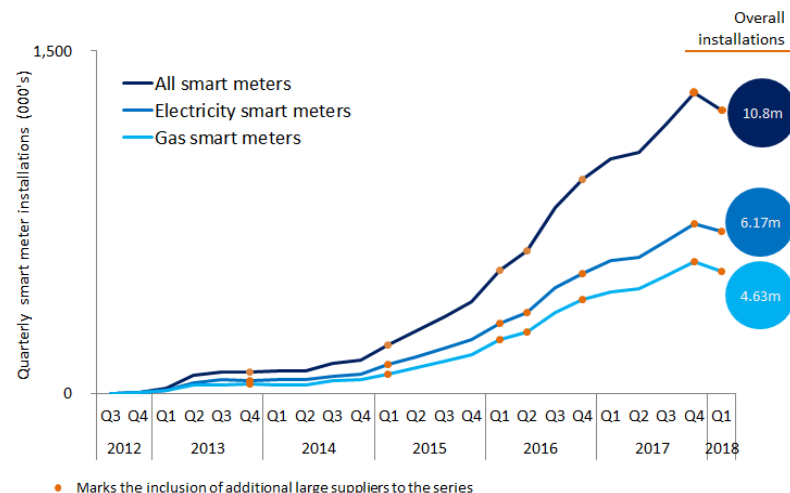
Keio University*, Japan

CEA Grenoble**, France



Background

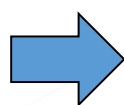
- The era of Big data
 - Many IoT equipment are built in our daily life and gathers various data
 - Smart meters
 - Automatically gathers power consumption data in every fixed time interval
 - The conventional power meters are changed into smart meters[1]
- The secondary use of the Big data is gaining attention
 - Power consumption data
 - Demand response service[2]
 - Gives out message to power users to save energy and cut the peak of the power consumption





Objective

- Issues in the secondary use of Big data
 - Privacy issue
 - Private information is invaded
 - Power consumption data
 - Violates one's lifestyle_[3]
 - Anonymization gaining attention for this issue
 - Energy issues for computing the data
 - Many services are invented due to the growth of the data
 - More service will enlarge the energy of the server_[4]
 - Transprecision computing is gaining attention for solving the issue



A demand for a new method, which preserves the privacy and lowers power consumption



Related works(1/2)

■ k-anonymity

- One privacy standard for anonymization
 - At least k number of tuples observed inside a q-block.

■ Identifier

- An attribute, which can detect a person individually
- Deleted

■ Quasi-identifier

- An attribute, which can detect a person by combining it with other attributes
- Anonymized

■ Sensitive attribute

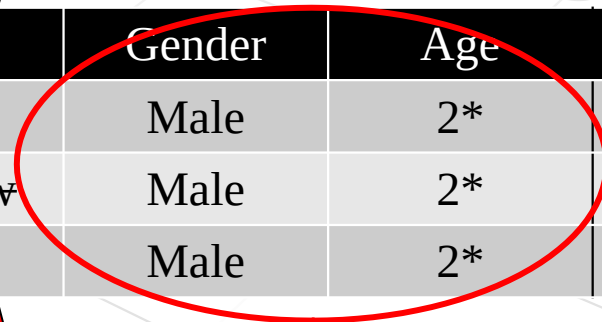
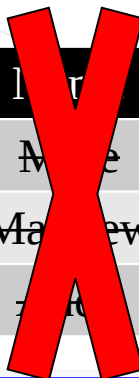
- An attribute necessary for the analyst
- Preserved

Identifier	Quasi-identifier		Sensitive
Name	Gender	Age	Disease
Mike	Male	25	Aids
Matthew	Male	28	Cancer
Andy	Male	29	Cancer



$k \geq 3$

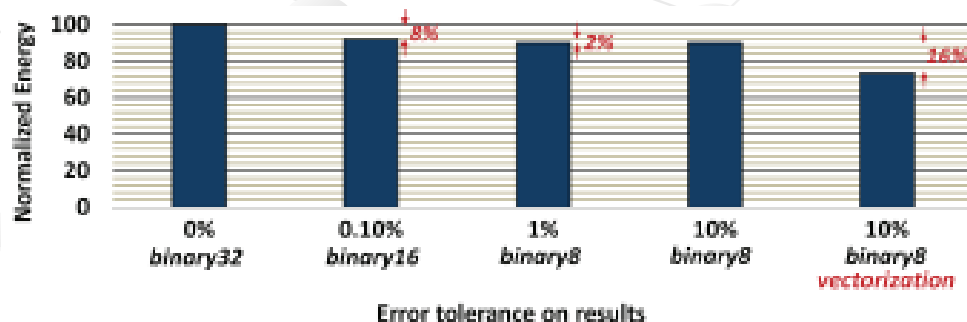
Name	Gender	Age	Disease
Mike	Male	2*	Aids
Matthew	Male	2*	Cancer
Andy	Male	2*	Cancer





Related works(2/2)

- Transprecision computing
 - One method of approximate computing
 - Precision variable computation
 - Enables to compute in the chosen precision
 - Open transprecision computing project (Oprecomp)
 - A project in the EU, which focus on the transprecision
 - CEA, IBM Zurich
 - By lowering the precision of the computing, the energy consumption of the computation will be reduced
 - 8~20% electricity reduction
 - Applications are simulated by using transprecision
 - k-nn, Mandelbrot-set





Connection

- Issues in the research fields

- Anonymization

- No consideration of energy in the computation

- Transprecision computing

- No consideration about privacy

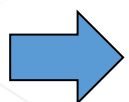
- Connection

- Transprecision computing

- Accepts the computation error to reduce energy for computation

- Anonymization

- Accepts the error to preserve privacy



A new privacy preserving method that fits the feature of the transprecision computing should be made



Proposed method

- Precision variable anonymization method supporting transprecision computing
 - The method has the parameters for level of anonymization and level of precision
 - Gives a trade-off between information loss and computational cost
- Steps
 - Use k-member clustering to group the data
 - Anonymize the exponent bit
 - Reduce the mantissa bit to chosen precision and anonymize the mantissa bit



K-member clustering

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



Anonymize the exponent bit

- Exponent bits
 - Change the exponent bit into the most appeared exponent in the cluster
 - If the exponent is larger than before change the mantissa to 0 and if smaller than before change the mantissa to 1

value	sign	exponent	mantissa	cluster
18.12	0	10000011	001...11000010	1
17.56	0	10000011	000...11100001	1
15.17	0	10000010	111...01010001	1



value	sign	exponent	mantissa	cluster
18.12	0	10000011	001...11000010	1
17.56	0	10000011	000...11100001	1
16	0	10000011	000...00000000	1



mantissa bit

- Mantissa bits
 - Change least number of bits into 0 chosen according to the precision
 - Change the mantissa bit into the most appeared bit in the cluster

value	sign	exponent	mantissa	cluster
18.12	0	10000011	001...11000010	1
17.56	0	10000011	000...11100001	1
16	0	10000011	000...00000000	1

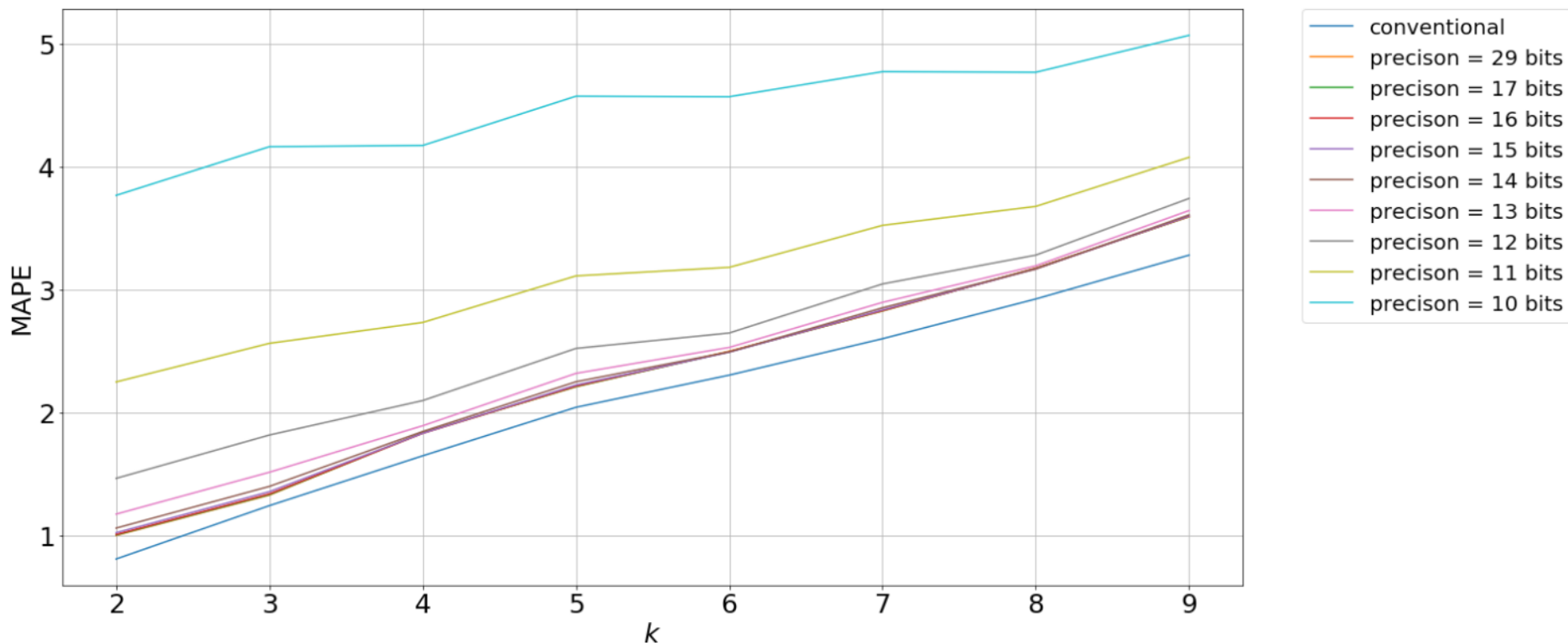


value	sign	exponent	mantissa	cluster
17.227	0	10000011	000...110000 <u>00</u>	1
17.227	0	10000011	000...110000 <u>00</u>	1
17.227	0	10000011	000...110000 <u>00</u>	1



Evaluation(1/3)

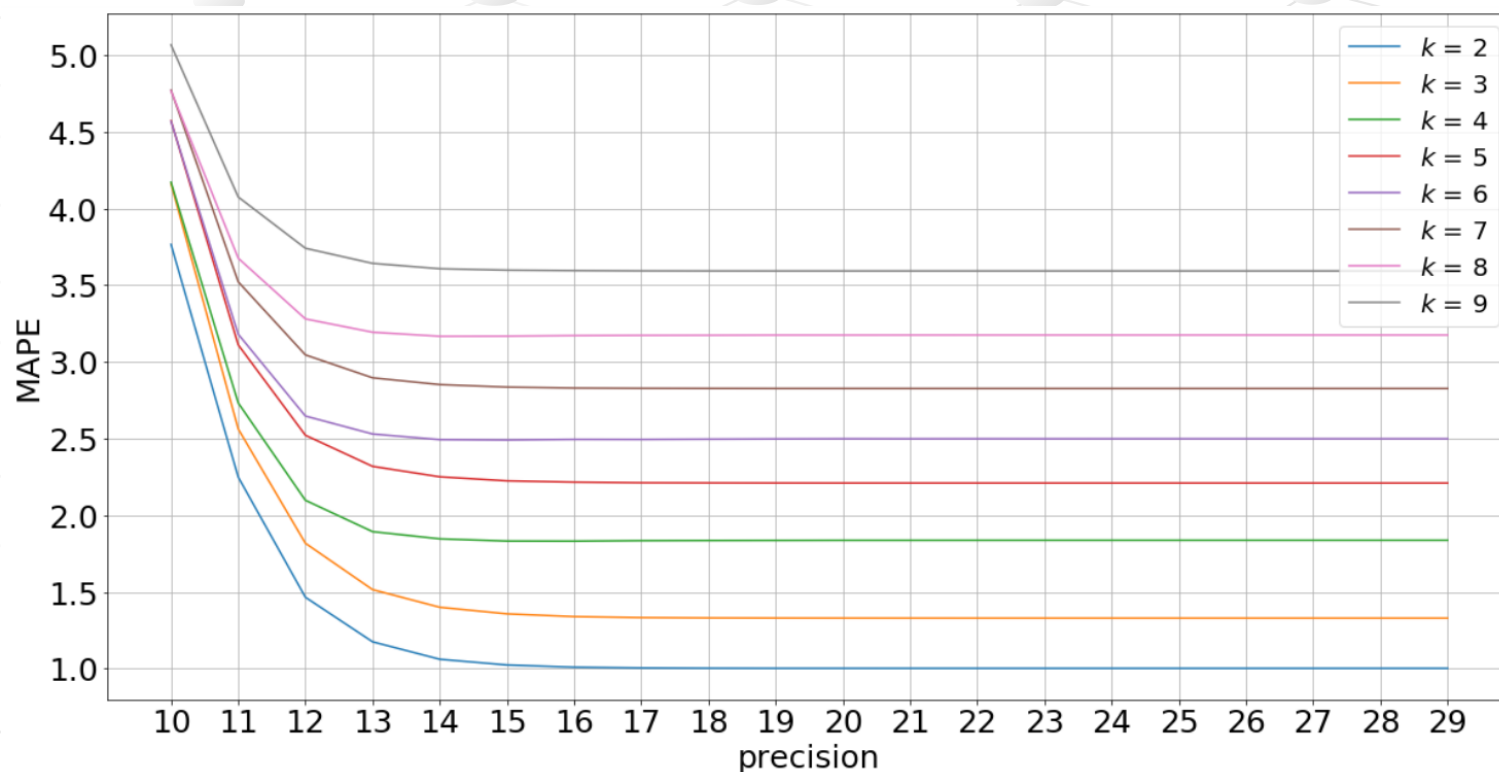
- The relationship between precision and MAPE
 - Proposed method gave less precision with higher MAPE
 - Less precision means less power consumption
 - Precision can be chosen by the anonymized data application





Evaluation(2/3)

- The relationship between k and MAPE
 - The MAPE of anonymization data only rises 0.14% compared to the conventional when the precision is 16bit
 - The larger the anonymity level is the error by precision lowers





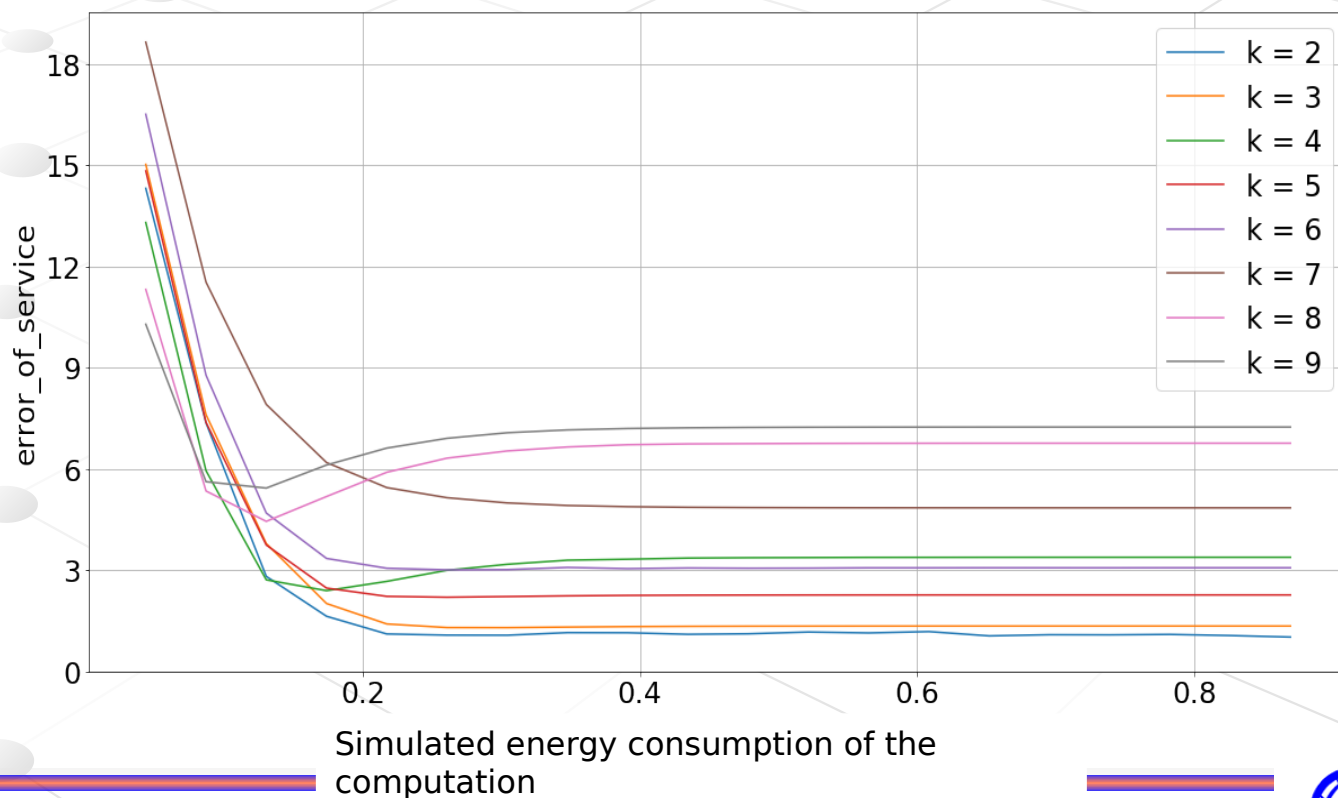
Evaluation(3/3)

- Evaluation of demand and response service and power consumption
- Demand and response service
 - The service curtails top 15% power users to 85% of the maximum power consumption
- Energy consumption simulation of the computation
 - Number of floating operation multiplied by the size of mantissa
 - The number of operations means the cycles of instruction
 - The size of mantissa = the length of the critical path



Evaluation(3/3)

- The relationship between the error of service and the simulated energy consumption
 - Energy consumption of the computation can be reduced to 15% when $k = 2$ and 18% when $k = 3, 4$





Conclusion

- Showed a connection between anonymization and transprecision computing
- Made an anonymization method, which has parameter of k and precision
 - Gave a trade-off between information loss and computational cost by deleting the mantissa bits
- Float32 anonymized data could be changed to float16 data with only 0.14% error increase
- Using this anonymization method the power consumption of the service can be reduced to 16% in average.



Thank you for your attention