The University Of Queensland AUSTRALIA

33RD USENIX Security Symposium

usenix THE ADVANCED COMPUTING SYSTEMS ASSOCIATION

Being Transparent is Merely the Beginning: Enforcing Purpose Limitation with Polynomial Approximation

Shuofeng Liu, Zihan Wang, Minhui Xue, Long Wang, Yuanchao Zhang, and Guangdong Bai

The University Of Queensland AUSTRALIA

CSIRO

ANT GROUP

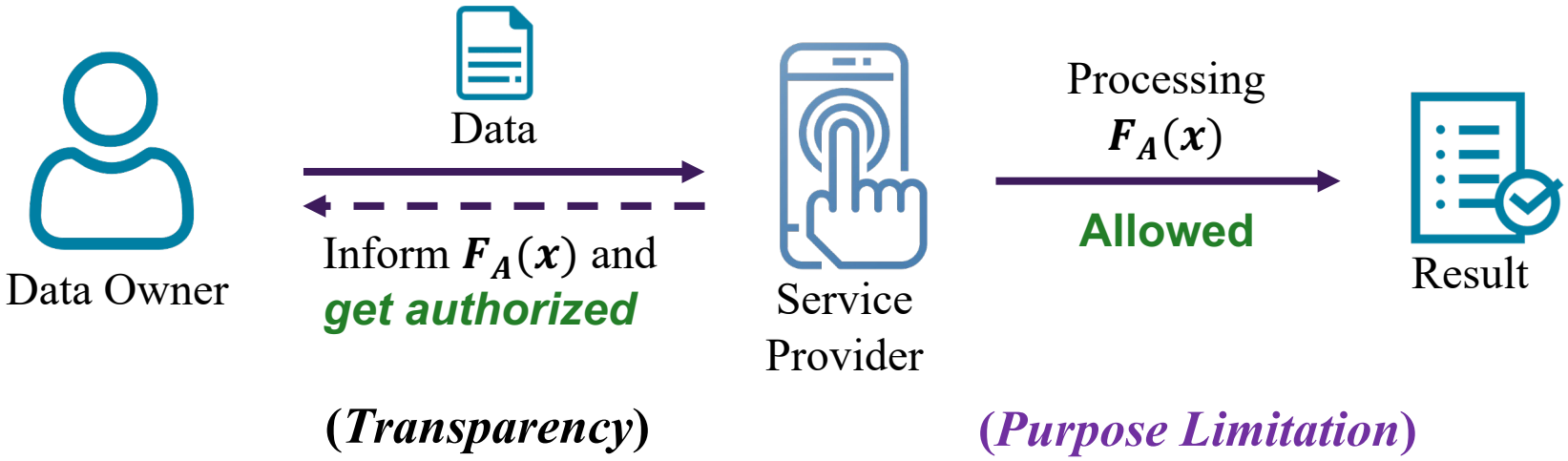
Privacy Regulations

- Personal Data Protection
  - Data Transparency

    Individuals should be fully informed about how their personal data is being collected, used, shared, and stored.
  - Purpose Limitation Principle

Purpose Limitation Principle

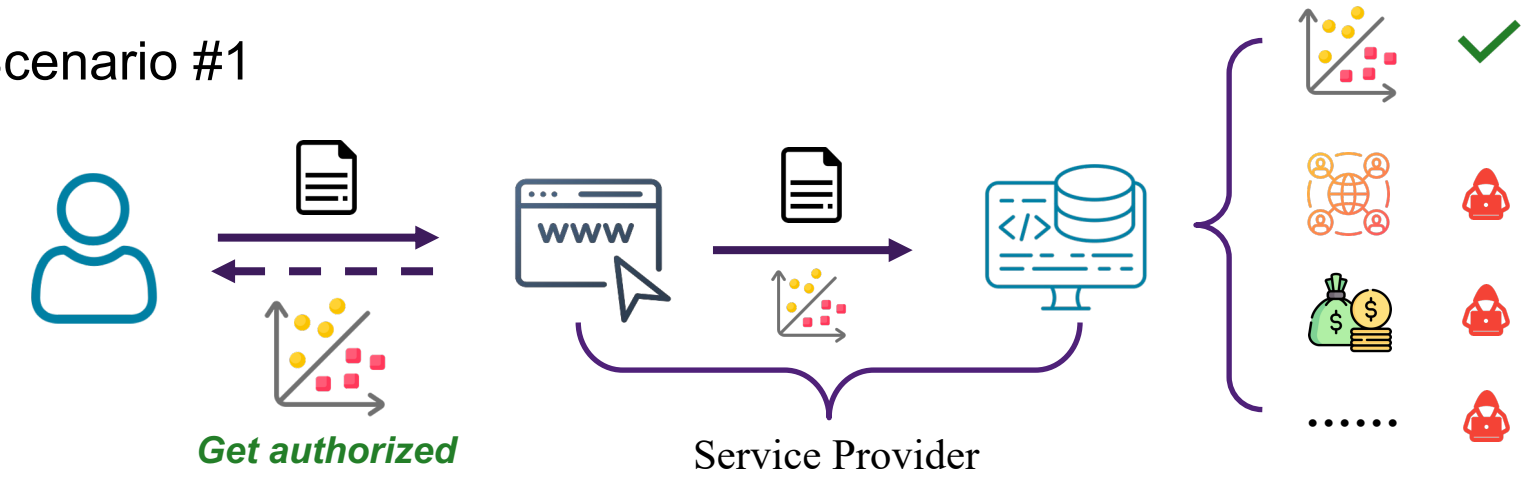
- Personal data should be collected for specified, explicit, and legitimate purposes and not further processed in a manner that is incompatible with those purposes.



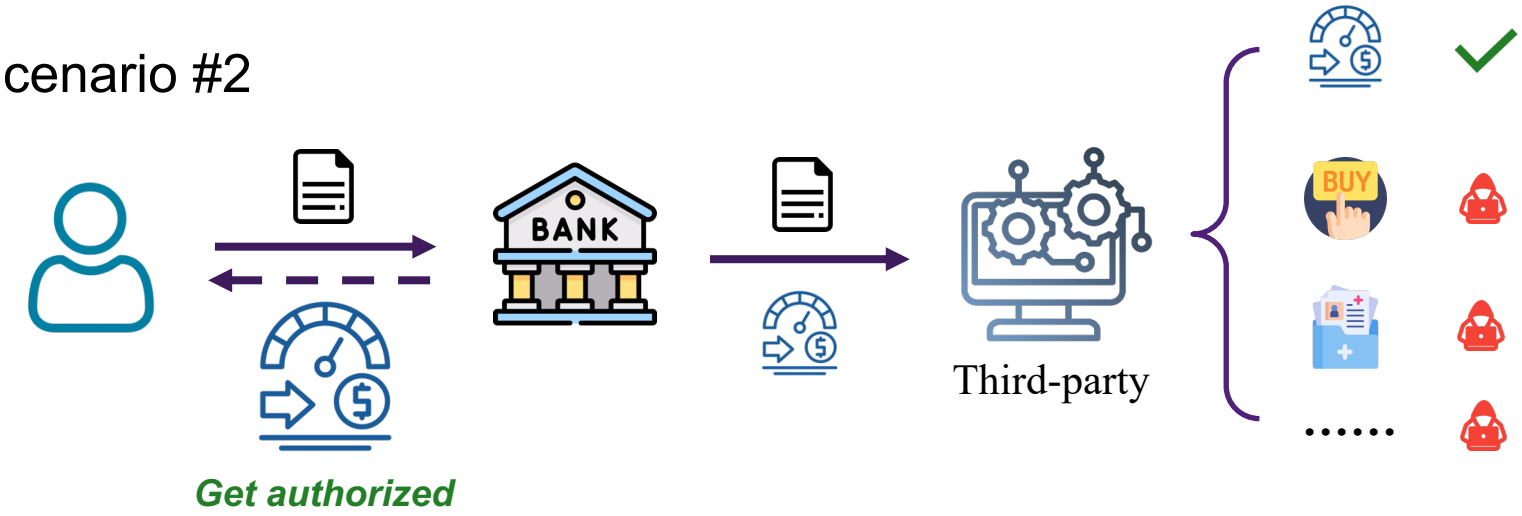
Only allowed for authorized Algorithm $F_A(x)$, but not for other unauthorized algorithms.

Practical Scenarios & Challenges

❑ Scenario #1

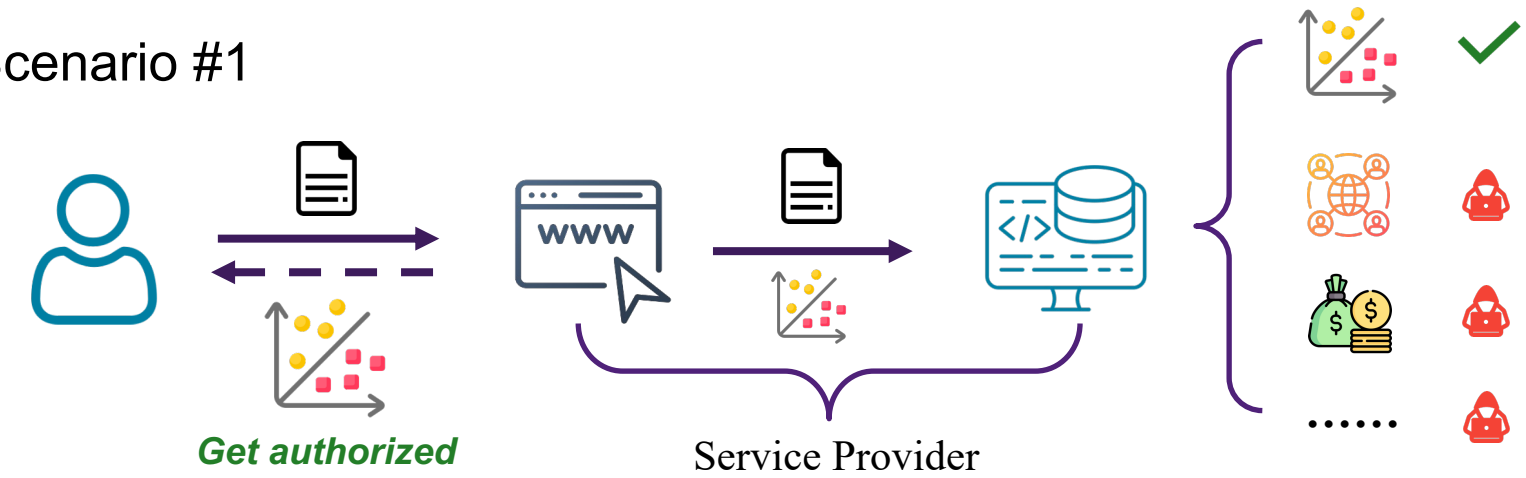


❑ Scenario #2

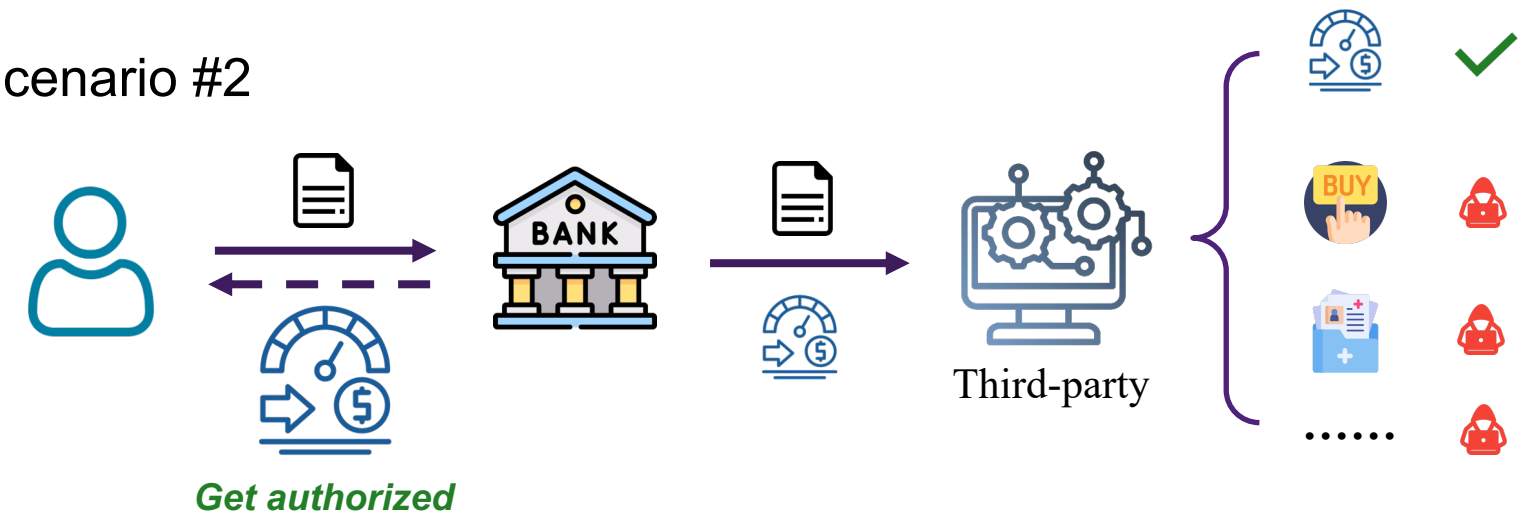


Practical Scenarios & Challenges

❑ Scenario #1



❑ Scenario #2



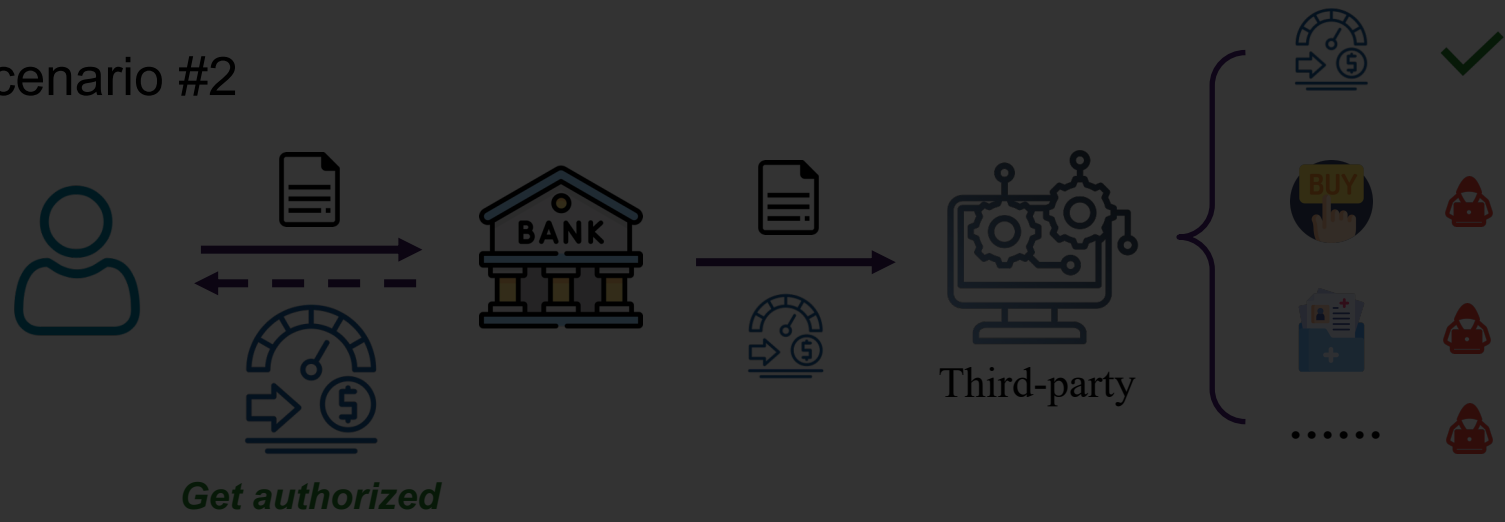
- Transparency alone is insufficient to protect user privacy.
- When data reaches the backend or third parties, it places the data owner in an ***unfair position***, where they lose control over their data

Practical Scenarios & Challenges

- Scenario #1

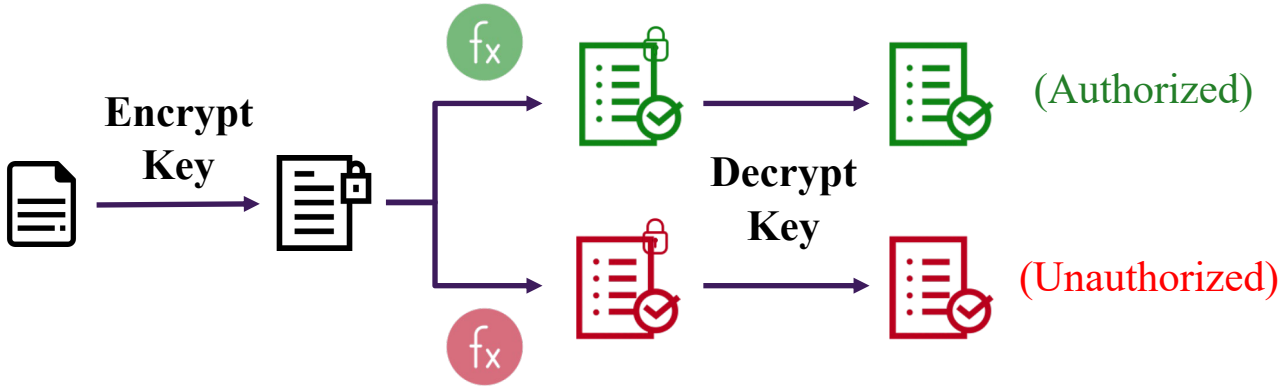
Enforcing the Purpose Limitation principle is of urgent significance!

- Scenario #2



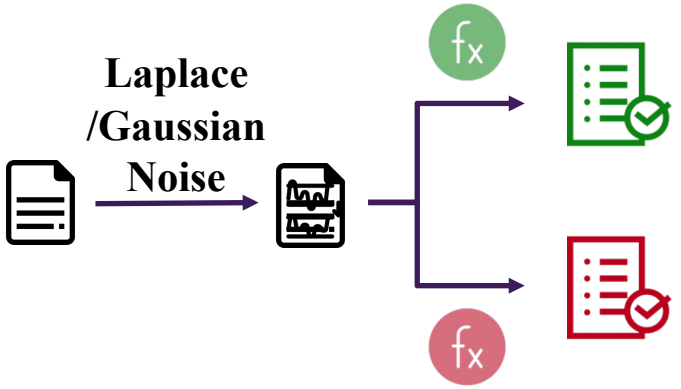
Existing Solutions

❑ Fully Homomorphic Encryption (FHE) [1]



- ❖ Asymmetric encryption systems increase **computational complexity**.
- ❖ Data can still **undergo unauthorized computations** while encrypted.
- ❖ The result can be decrypted using the **decrypt key** to determine if unauthorized operations have been performed.

❑ Differential Privacy (DP) [2]

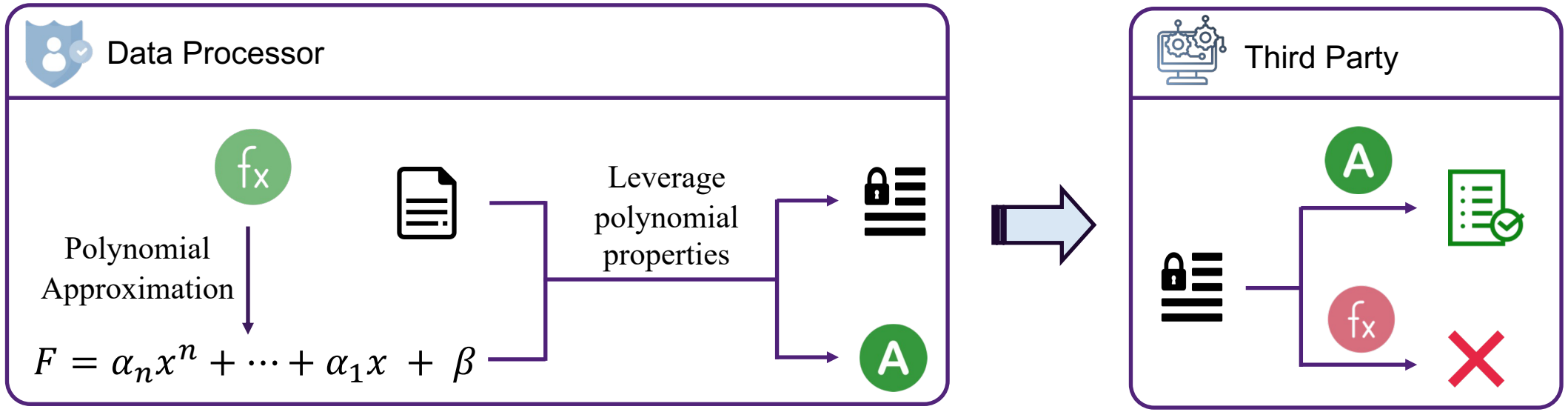


- ❖ To achieve privacy protection, a substantial amount of noise should be added, leading to a **reduction in the accuracy** of the calculation results.
- ❖ Not only preventing unauthorized algorithms from producing correct results but may also **affect authorized algorithms**.

Our Solution: AlgoSpec

Algorithm Specificity (AlgoSpec)

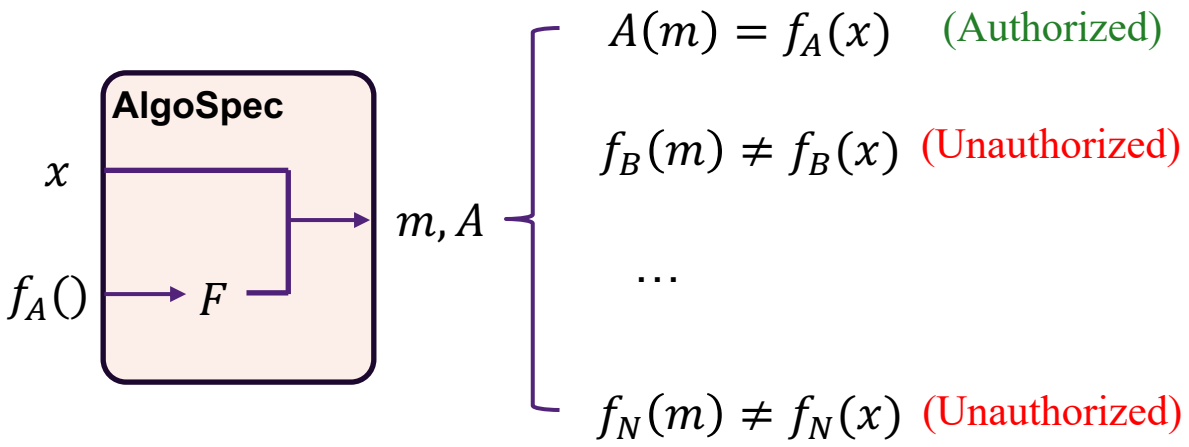
- Main Idea



- A is the substitution of fx
- contains the information of both and fx
- AlgoSpec ensures that only authorized algorithm fx can be calculated to get correct result by using polynomial approximation and its properties.

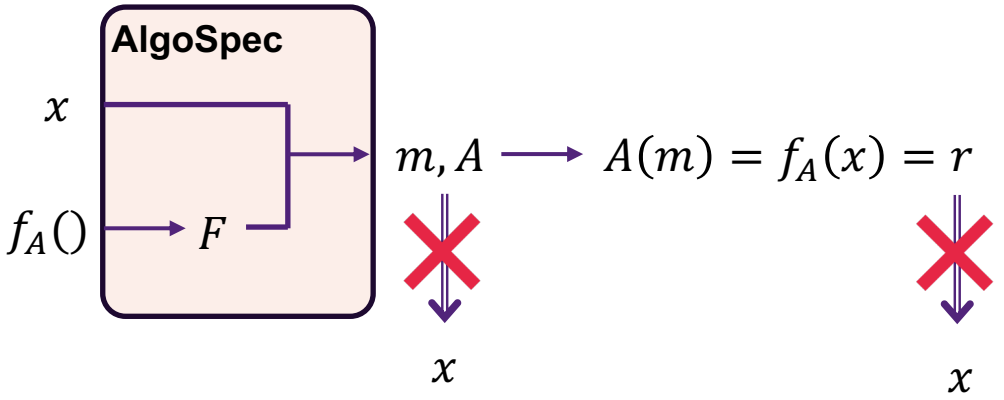
Two Goals of AlgoSpec

❑ *Goal #1*. Algorithm specificity



❖ Only executing the authorized algorithm can obtain correct results.

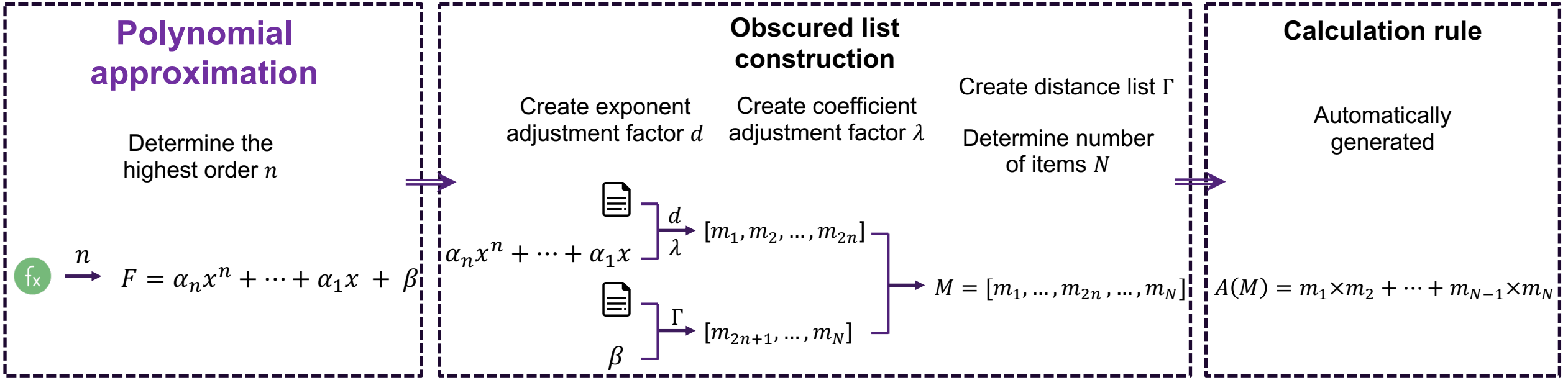
❑ *Goal #2*. One-way obfuscation



❖ Given m and A, the polynomial time adversary (PTA) cannot recover the original data.

❖ Assuming the original algorithm is complicated, given the final result r, the PTA cannot recover the original data

Approach

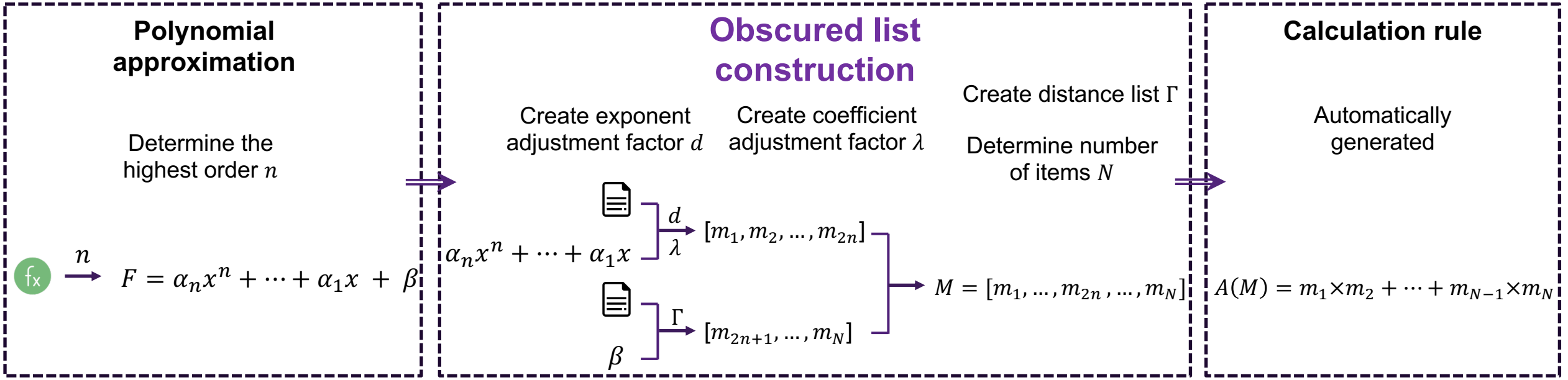


Determine the highest order n using ***discrete logarithm problem & Diffie-Hellman problem*** [3, 4]. ⟶ *n is indistinguishable*

Use the polynomial fitting method, a typical machine learning strategy, to do polynomial approximation with n.

Uniqueness of the polynomial with respect to the degrees for polynomial approximation [5]. ⟶ *Achieve Goal #1 algorithm specificity*

Approach



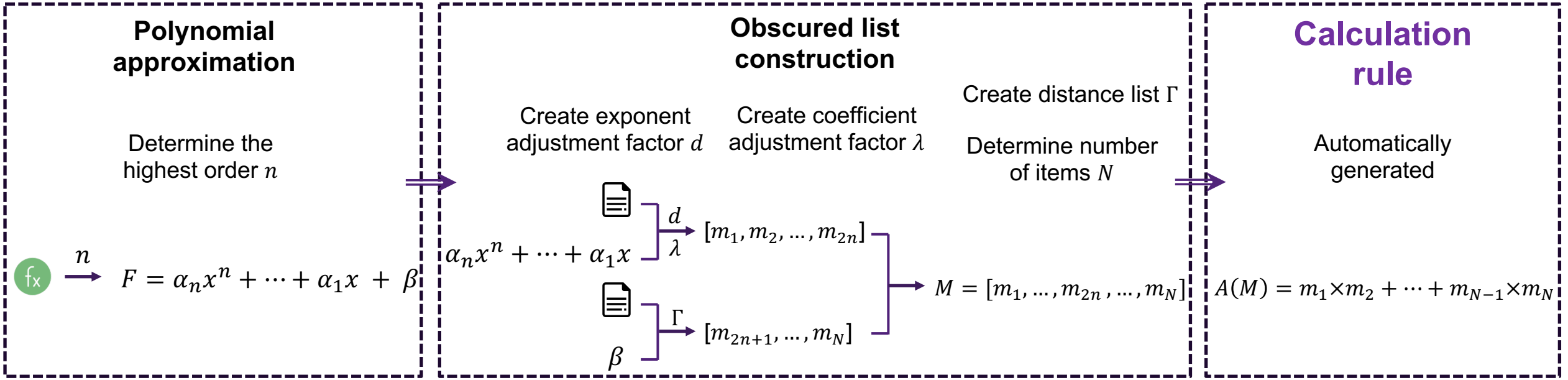
$m_1 = \frac{\alpha_n}{\lambda_n} x^{n-d_n}$ $m_2 = \lambda_n x^{d_n}$ … $m_{2n-1} = \frac{\alpha_1}{\lambda_1} x^{1-d_1}$ $m_{2n} = \lambda_1 x^{d_1}$ ⟶ *Processing of non-constant terms*

pattern: A pattern refers to the distribution of the distance $\left\lfloor \log \frac{m_i}{m_{i+1}} \right\rfloor$, where i is odd.

m_{2n+1} m_{2n+2} … m_{N-3} m_{N-2} follows the patterns, and randomly select m_{N-1}, then $m_N = \frac{\beta - (m_{2n+1} \times m_{2n+2} + \cdots + m_{N-3} \times m_{N-2})}{m_{N-1}}$

Set $N \gg 2n$ to ensure the highest order n is hidden ⟶ *Achieve Goal #2 one-way obfuscation*

Approach



$$m_{2n-1} = \frac{\alpha_1}{\lambda_1} x^{1-d_1} \quad m_{2n} = \lambda_1 x^{d_1} \Rightarrow m_{2n-1} \times m_{2n} = \frac{\alpha_1}{\lambda_1} x^{1-d_1} \times \lambda_n x^{d_1} = \alpha_1 x^1$$

$$m_N = \frac{\beta - (m_{2n+1} \times m_{2n+2} + \cdots + m_{N-3} \times m_{N-2})}{m_{N-1}} \Rightarrow \beta = m_{2n+1} \times m_{2n+2} + \cdots + m_{N-1} \times m_N$$

$$\underbrace{\boldsymbol{m_1 \times m_2 + \cdots + m_{N-1} \times m_N}}_{\text{Calculation rule}} = \underbrace{\boldsymbol{\alpha_n x^n + \alpha_{n-1} x^{n-1} + \cdots + \alpha_1 x + \beta}}_{\text{Polynomial}} = \boldsymbol{F_x(x)}$$

Original Algorithm

Ensure the accuracy of computation

Security Analysis

- **Main theorem** (*Unrecoverability of original data*)

  Under a CPA, the original data is not recoverable by analyzing obscured list or solving the polynomial by obtaining the highest degree.

- **Theorem 1** (*Confidentiality of the highest order*)

  Under a CPA, the highest order is not recoverable either from the obscured list or the initialization step.

- **Theorem 2** (*Indistinguishability of obscured list*)

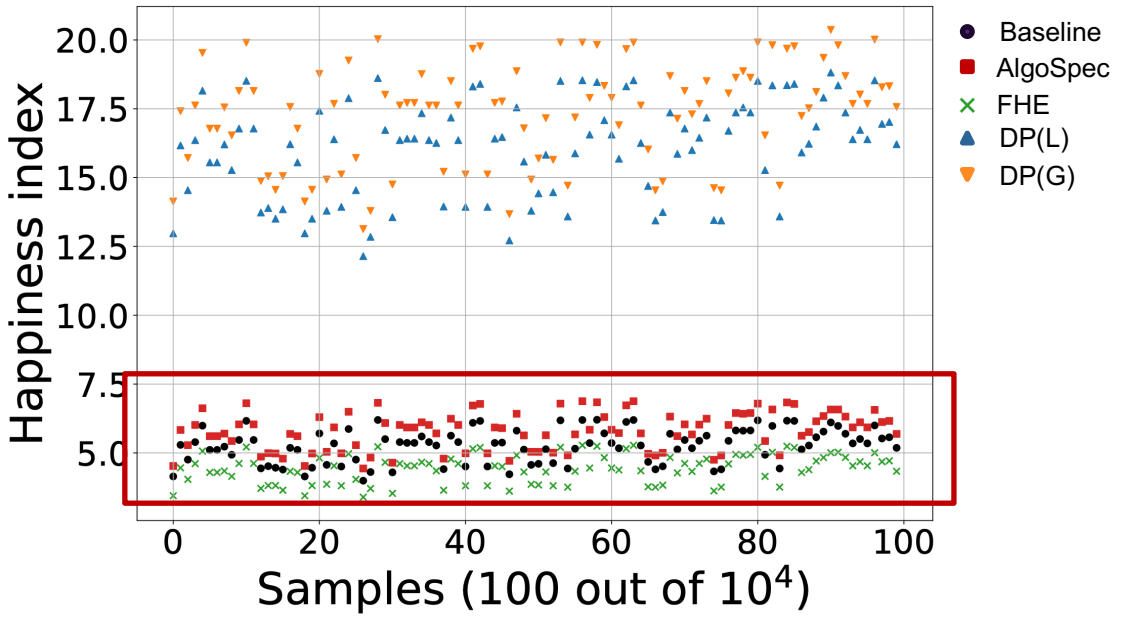
  Given M_1 and M_2 that are generated from different degrees n_1 and n_2, respectively, M_1 and M_2 are indistinguishable when given a large N.

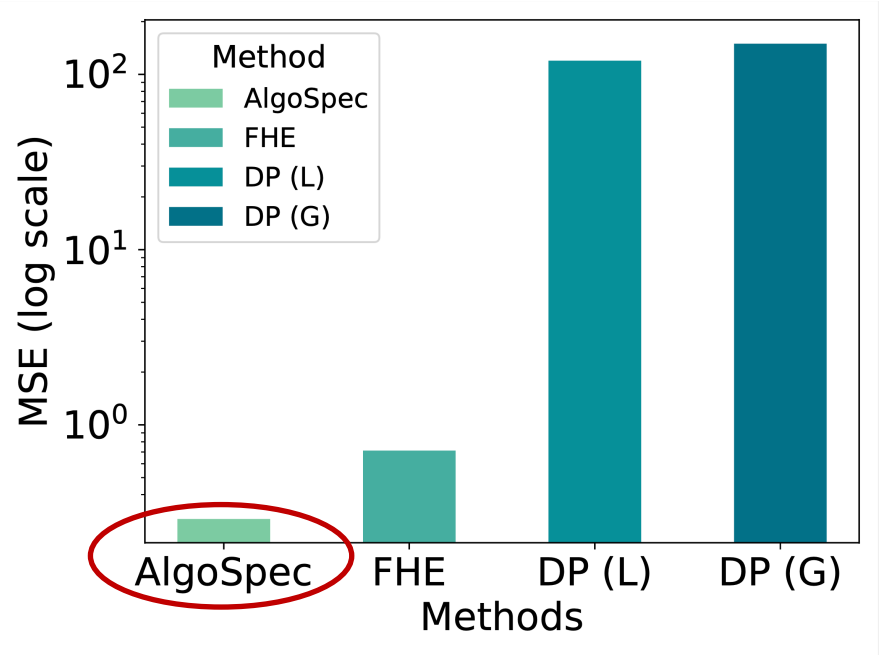
We leverage *Decisional Diffie-Hellman problem* [4], *Discrete logarithm problem* [3], *Chosen-plaintext attack* [6], *Indistinguishability* [7] to construct **Code-based Game Playing** to prove the three theorems showing the security of our AlgoSpec.

Experimental Evaluation

❑ Benchmarking

❑ Apply AlgoSpec to entropy method



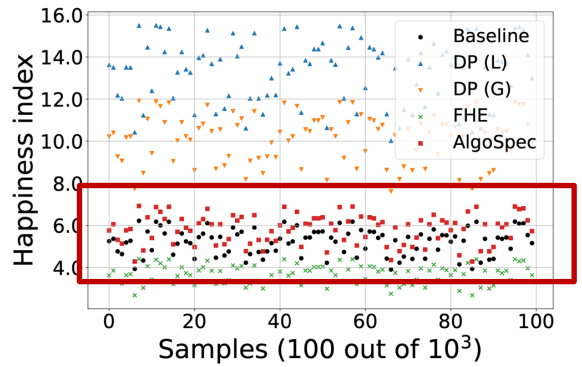


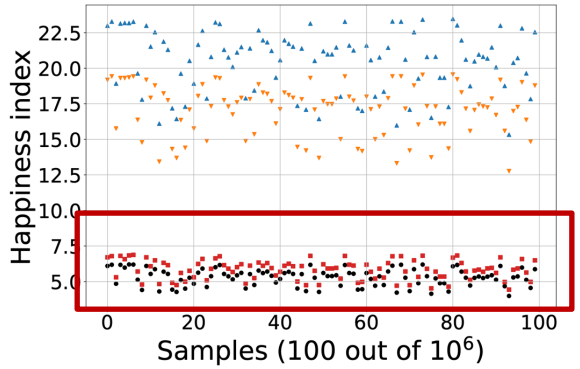
- ❖ The results obtained using AlgoSpec most closely align with those calculated by the baseline.
- ❖ AlgoSpec achieves the lowest MSE of 0.289 compared with other methods.

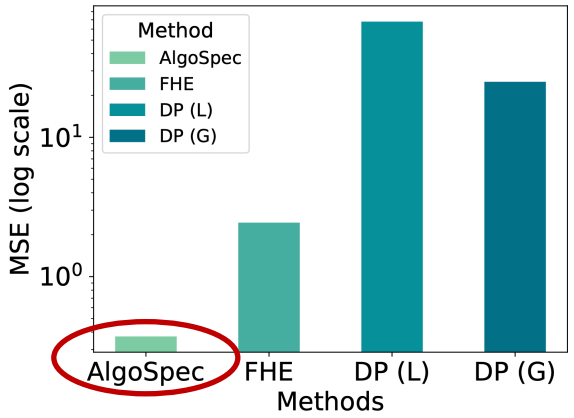
Experimental Evaluation

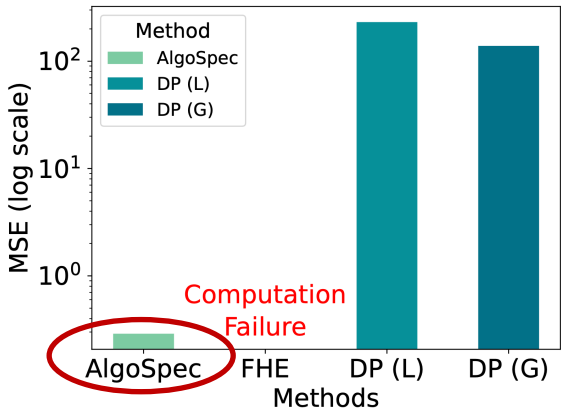
❑ Scalability

❑ Settings: Dataset from the scale of 10^3 to 10^6.









- The results obtained by AlgoSpec are the closest to those calculated by the baseline across all scales of dataset.
- AlgoSpec achieves the lowest MSE compared with other methods across all settings.
- FHE encounters computation failure when the scale of dataset is very large, while AlgoSpec maintains the high accuracy.
- The loss in accuracy of AlgoSpec is significantly smaller than DP and FHE.

Experimental Evaluation

❑ Scalability

Efficiency of different scale datasets from 10^3 to 10^6 on entropy method

	10^4			10^3			10^5			10^6		
	PT	ET	TT	PT	ET	TT	PT	ET	TT	PT	ET	TT
Baseline	0.000	0.235	0.235	0.000	0.028	0.028	0.000	2.168	2.168	0.000	20.793	20.793
DP (L)	1.007	0.218	1.225	0.097	0.025	0.122	10.194	2.136	12.330	101.904	21.684	123.588
DP (G)	1.030	0.219	1.249	0.128	0.023	0.151	10.137	2.141	12.278	104.117	21.016	125.133
FHE	0.134	487.663	487.797	0.015	5.717	5.732	-	-	-	-	-	-
AlgoSpec	36.378	1.069	37.447	3.601	0.109	3.710	365.054	10.409	376.463	3675.673	102.339	3778.012

PT: Processing Time **ET**: Execution Time **TT**: Total Time

❖ AlgoSpec outperforms FHE across all scales of dataset, in terms of total computation time.

Conclusion

- We make a preliminary attempt on the algorithm level to achieve purpose limitation principle.
- AlgoSpec achieves *Goal #1* algorithm specificity primarily by leveraging polynomial approximation and its nature of the uniqueness results of a certain polynomial.
- AlgoSpec achieves *Goal #2* one-way obfuscation by constructing a hidden highest order of the polynomial using the *Discrete logarithm problem* and *Diffie-Hellman problem*.
- Compared to data transparency, executing purpose limitation principle faces more challenges, so we hope that more studies could look into this domain.

Reference

1. Craig Gentry. Fully homomorphic encryption using ideal lattices. *In Proceedings of the 41st Annual ACM symposium on Theory of Computing (STOC)*, pages 169–178, 2009.
2. Cynthia Dwork. Differential privacy. *In 33rd International Colloquium on Automata, Languages and Programming (ICALP)*, pages 1–12, 2006.
3. Smart, Nigel P. "The discrete logarithm problem on elliptic curves of trace one." *Journal of cryptology* 12 (1999): 193-19.
4. Boneh, Dan. "The decision diffie-hellman problem." *International algorithmic number theory symposium*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998.
5. Daniele Funaro. Polynomial approximation of differential equations, volume 8. *Springer Science & Business Media*, 2008.
6. Peng, Xiang, Hengzheng Wei, and Peng Zhang. "Chosen-plaintext attack on lensless double-random phase encoding in the Fresnel domain." *Optics letters* 31.22 (2006): 3261-3263.
7. Tore Vincent Carstens, Ehsan Ebrahimi, Gelo Noel Tabia, and Dominique Unruh. Relationships between quantum ind-cpa notions. *In Theory of Cryptography Conference (TCC),* pages 240–272, 2021

Thank you!

Email: shuofeng.liu@uq.edu.au