Privacy-Preserving Inference on Embedded Devices

Elliott Jang[1,2], Seyda Nur Guzelhan[2], Lohit Daksha[2], Ajay Joshi[2]
Fairmont Preparatory Academy, 2200 W Sequoia Ave, Anaheim, CA 92801[1],
Boston University Integrated Circuits, Architectures, and Systems Group (ICAS), 8 St Mary's St, Boston, MA 02215[2]
elliottjang1@gmail.com[1]

Motivation

- Encrypted data is sent across the internet and decrypted in servers
- Many data centers risk adversaries **targeting data in the clear**
  - Each attack causes average of **$5.08 M** in damage in 2025[1]
- Threat to the field of Artificial Intelligence (AI), where sensitive data is processed.
- FHE[2] algorithms are **computationally intensive** and require strong server supports to do meaningful tasks.
  - Most IoT devices are **resource constrained** (limited by compute, memory and power consumption)
  - Can handle only very minimal FHE enabled tasks.

Aim: To investigate the **viability** of HE **logistic regression inference**[3] using the CKKS[4] scheme on **low-cost embedded devices**, in this case a Raspberry Pi, and push its limits to **determine the best working parameter set** for FHE.

Background

- **Homomorphic encryption (HE)**: enables **processing** over **encrypted data**; post-quantum secure[2]
  - **Cheon-Kim-Kim-Song (CKKS)**: HE scheme for use with real numbers[4]

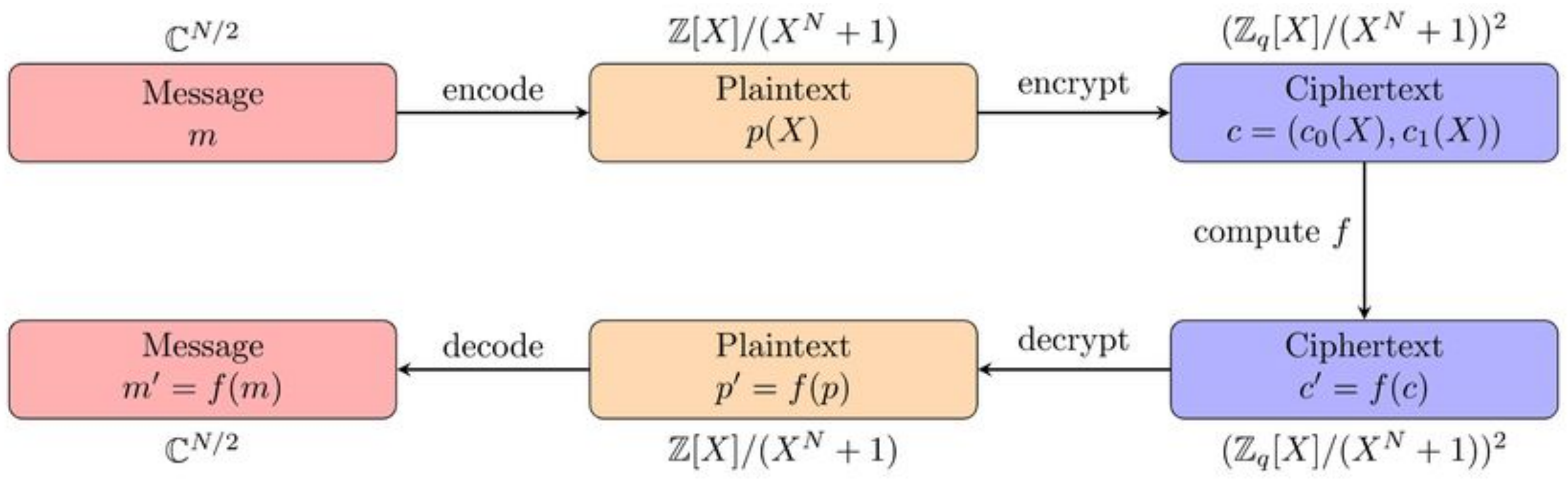
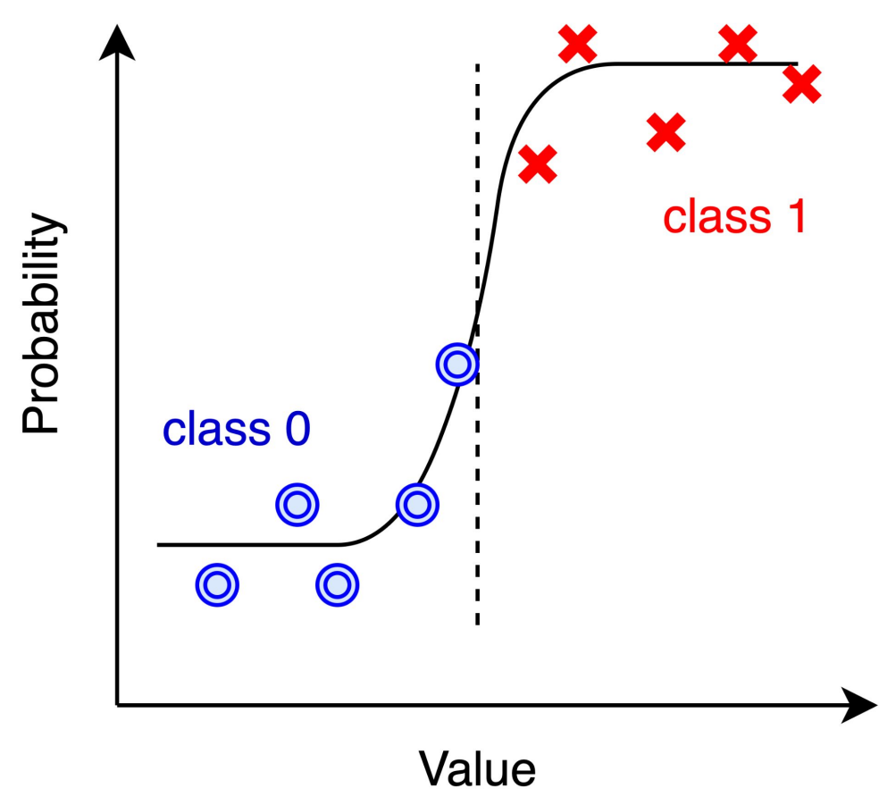


Figure 1: CKKS FHE algorithm steps[5]

- **Logistic Regression**: A machine learning technique for binary classification; critical for fields such as medicine.



- **Raspberry Pi (RPI)**: Single-Board Computer, a type of embedded device; typically used to create low-cost servers or internet of things devices.

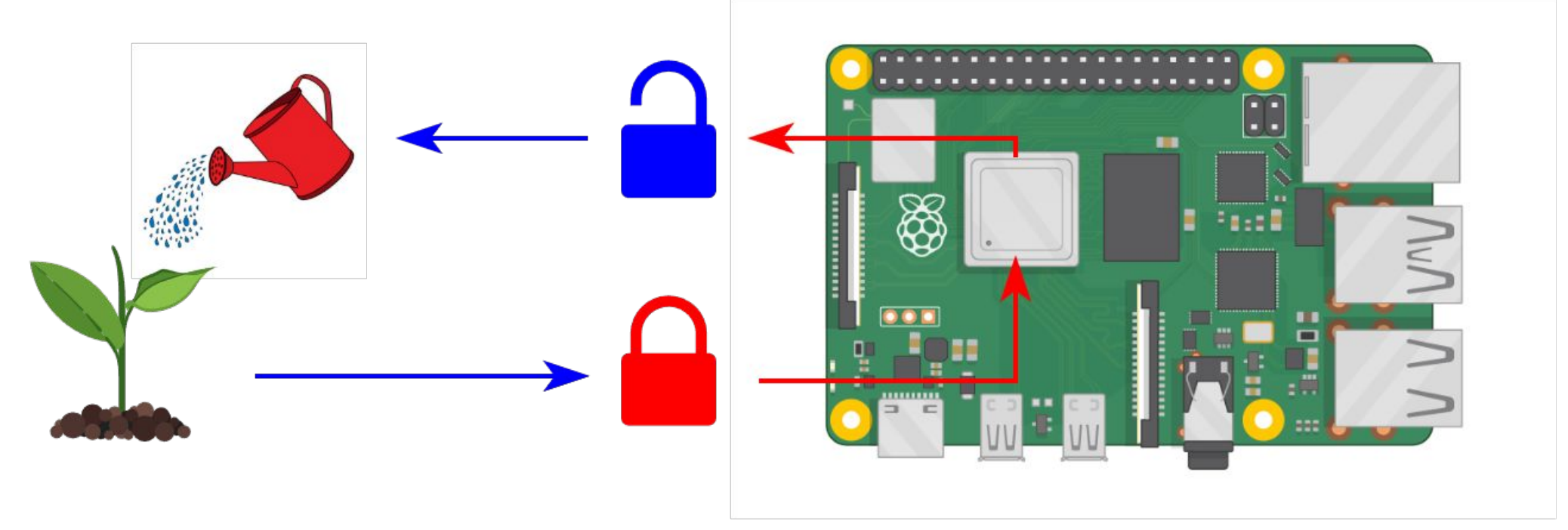


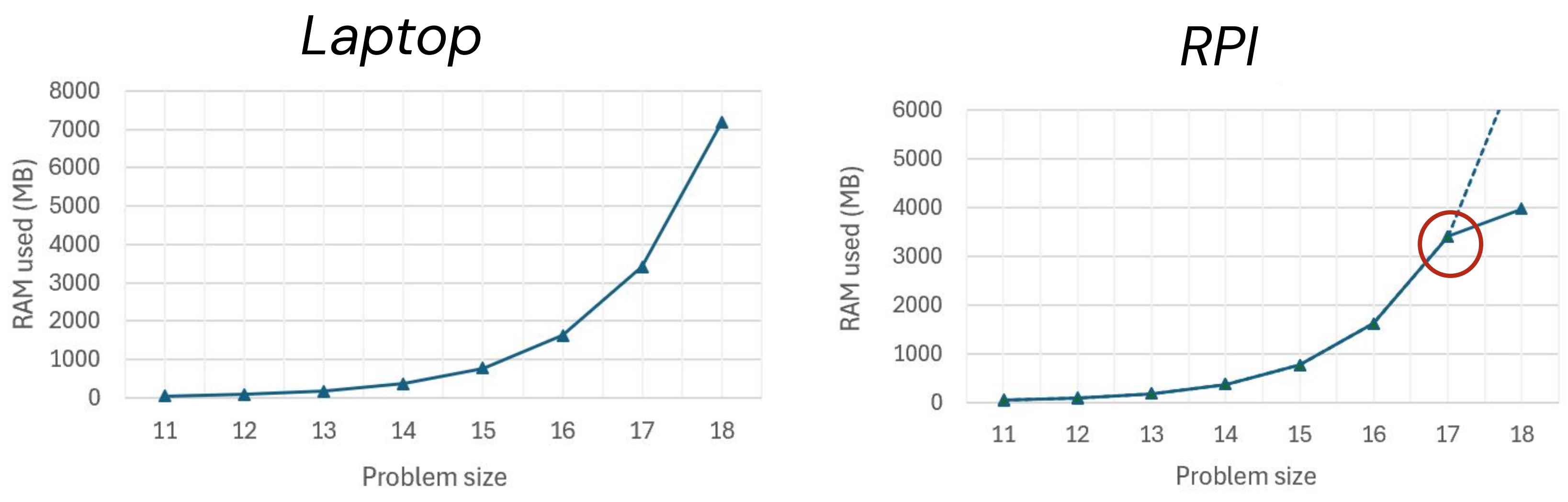
Figure 2: HE workflow with RPI as an embedded device

Methods

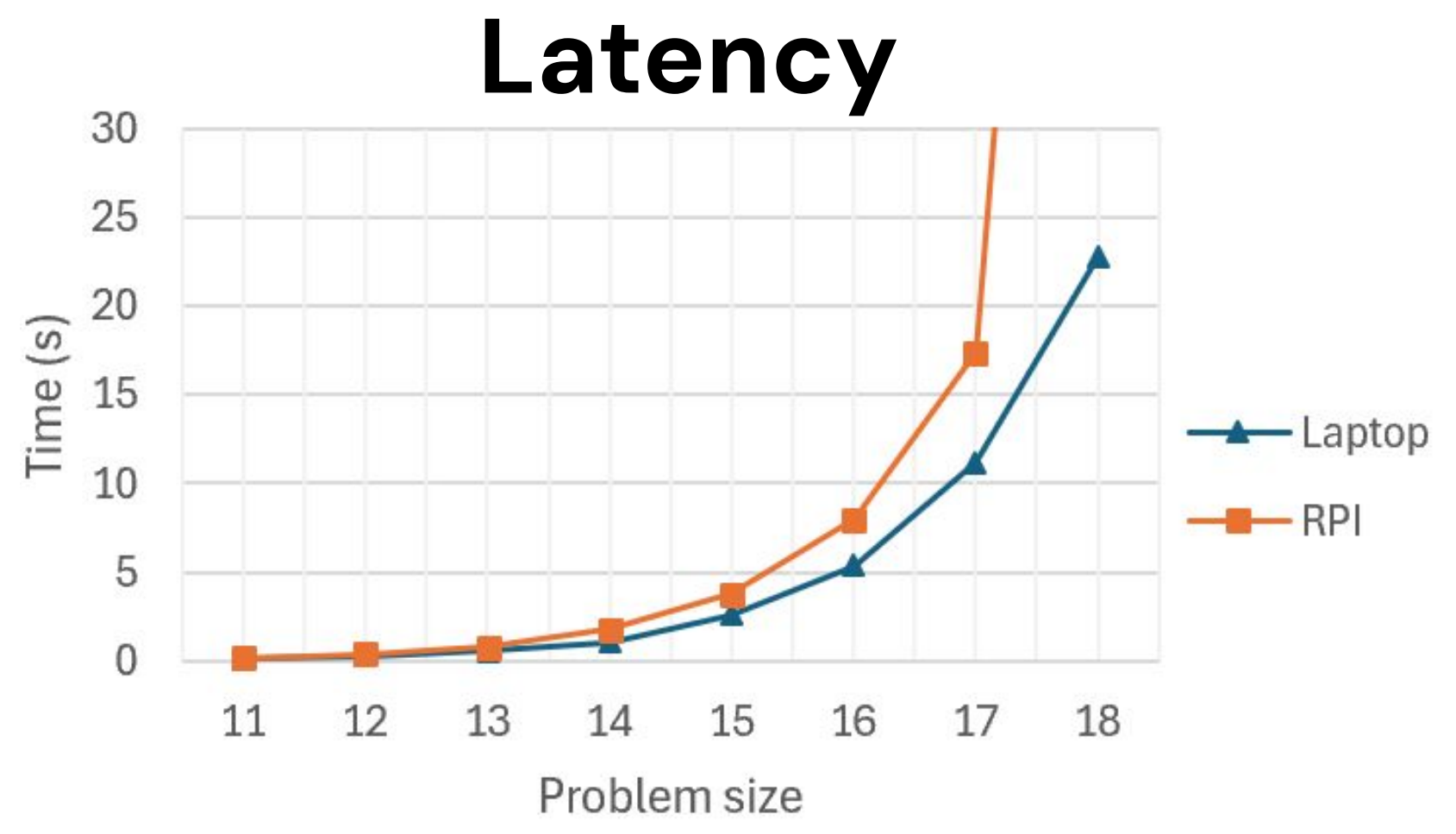
- **For Evaluation**: RPI 5; Quad-core CPU @ 1.5 GHz; 4GB RAM; Raspian OS Lite
- **For comparison**: Laptop; 8-core CPU @ 4GHz; 16GB RAM; Windows 11
- Created **C++ file** for encrypted **inference** using **OpenFHE library**[6], a leading CPU-based FHE library; **w/o bootstrapping**
- Used **std::chrono** and Linux **/usr/bin/time** command to measure **accuracy**, **time**, and **CPU** and **RAM** utilization

Results

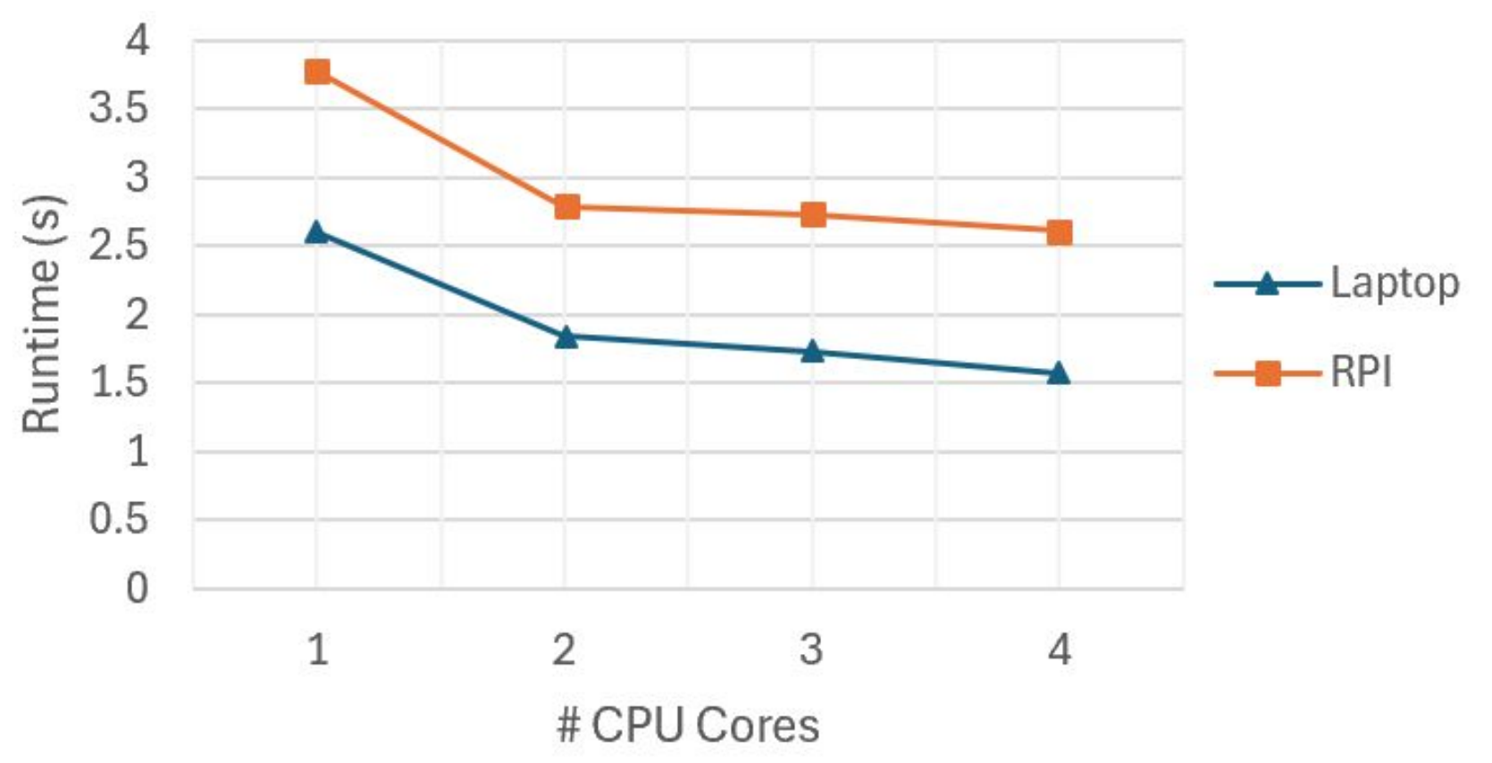
RAM Utilization



Latency



Runtime vs. # CPU cores



Averaged Metrics (N = 2^{15})

	Laptop	RPI	Diff
Inference time	2.77 s	3.82 s	**1.4x**
CPU Utilization	88.65%	97.23%	**9%**
RAM Utilization	714 MB	717 MB	**-**
Accuracy	71.88%	71.88%	**-**

Conclusion

- FHE logistic regression on a Raspberry Pi **may be viable** for use as cheap servers or several other privacy-preserving applications.
- Using parameters of **N = 2^{15}/2^{16}** help preserve 128-bit security while still using less RAM and maintaining optimal runtimes
- However, in cases where runtime is less of a priority (such as monitoring devices), **all runtimes are feasible**.
- Usage of multiple CPU cores (instead of only one) can open up further possibilities for applications.

Acknowledgements

I would like to thank Professor Ajay Joshi for giving me this opportunity, his valuable time to meet with us, and providing the resources and lab for this research. I would also like to thank my mentors, Seyda Nur Guzelhan and Lohit Daksha, for devoting so much of their time helping with debugging and answering technical questions.