

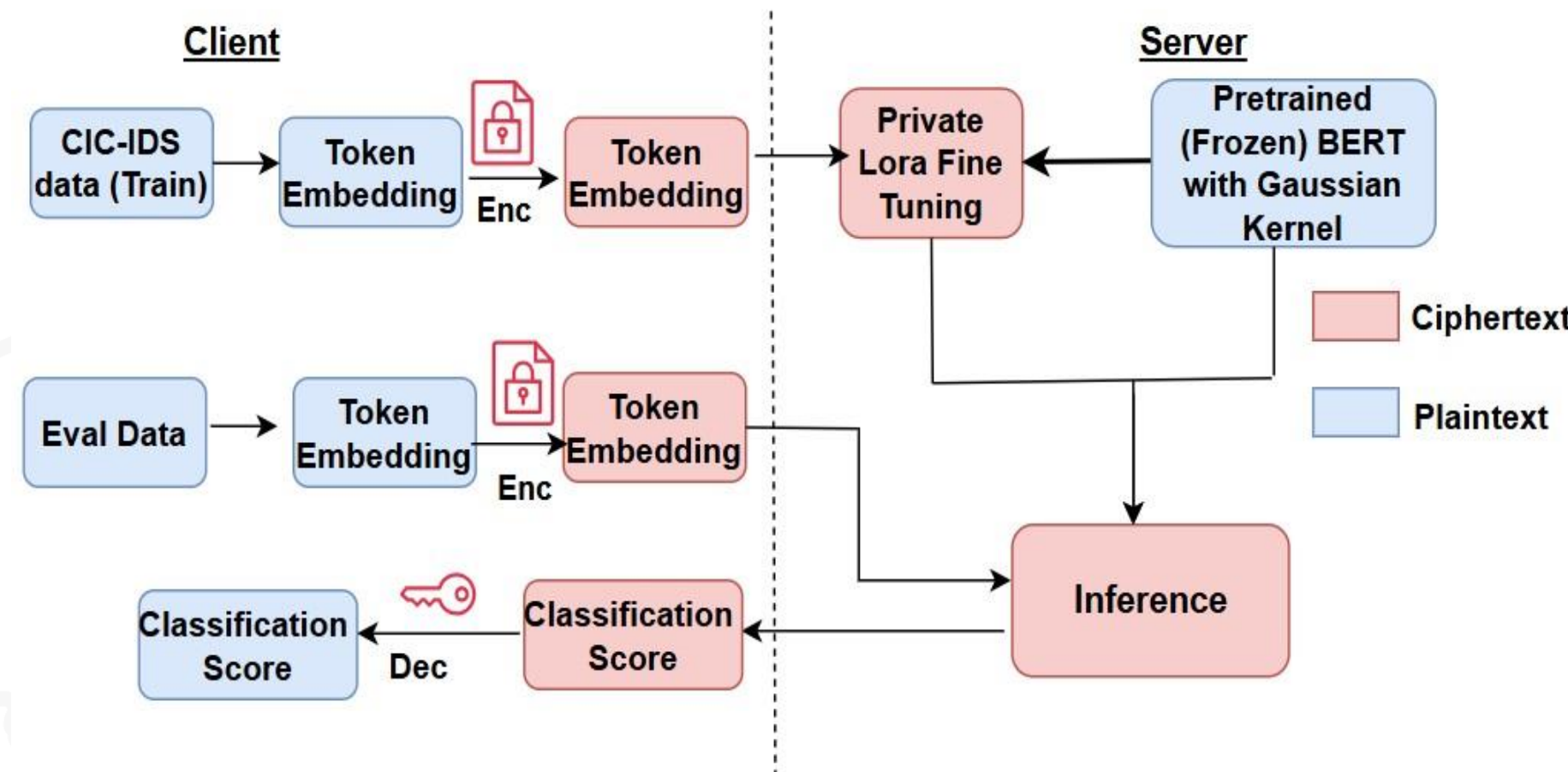
Raushan K Pandit, Ram Dantu, Kirill Morozov
University of North Texas, Denton, TX

Motivation

Network traffic logs may have sensitive information such as IP addresses, payload signatures, and internal communication parameters. Training Intrusion Detection System (IDS) models directly on raw data exposes critical privacy risks, and organizations may be reluctant to share such data for model training in collaborative scenarios.

- **Traditional IDS:** Sensitive information and model weights at risk.
- **HE-LoRA (Our):** Fine-tuning and inference on ciphertexts via Homomorphic Encryption (HE)-based LoRA fine-tuning of BERT, a pre-trained language model. **Privacy is Preserved.**

Proposed Architecture of LLM Inference



CIC-IDS-2017 Dataset and Tokenizer

- **Dataset:** CIC-IDS-2017, Friday-DDoS subset — 225,745 samples, 79 features, binary labels (benign / DDoS).
- **Feature selection:** Top 20 features picked using mutual information.
- **Tokenization (IDSFeatureTokenizer —our custom approach):**
 - Normalize each feature via RobustScaler (quantile range 5–95%)
 - Clip to $[-4, +4]$
 - Quantize into 256 uniform bins
 - Map to token ID in range $[5, 260]$; special tokens [CLS], [SEP], [PAD], [MASK] occupy IDs 0–4 (vocab size = 261)
- **Output sequence:** [CLS] $f_0 f_1 \dots f_{19}$ [SEP] [PAD]... padded to length 128.

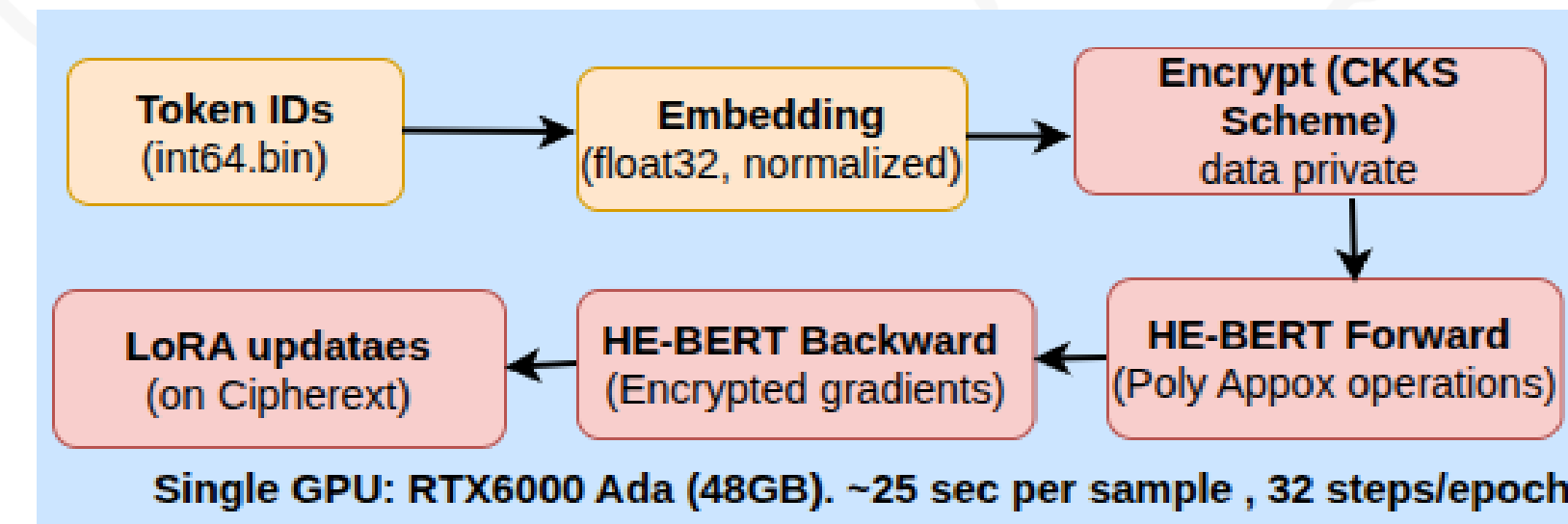
HE-LoRA Pipeline

- Pre-trained weights (W_0) are converted to HE-friendly format (.msg).
- LoRA adapters A,B are trained directly on ciphertexts — already HE-friendly.

- Applied to query, key, value, and output projection matrices W_Q, W_K, W_V, W_{out} :

$$W^* = W_0 + \Delta W, \Delta W = AB \quad | \quad \text{Where, } A \in \mathbb{R}^{d \times r}, B \in \mathbb{R}^{r \times d}, r=8 \ll d=768$$

- Only A,B updated during HE fine-tuning — <3 % of total parameters, while " W_0 " remains frozen.

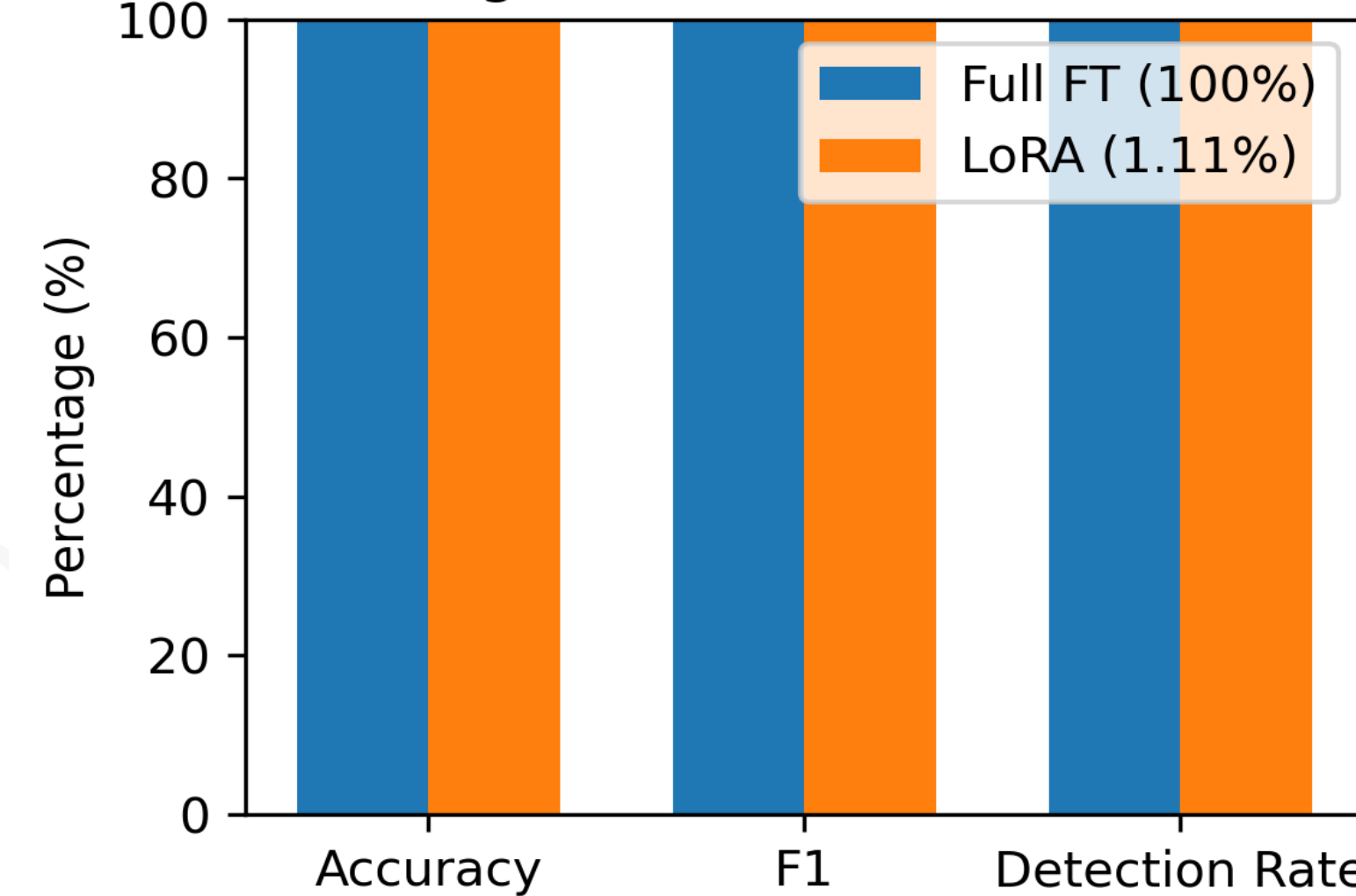


Crammed BERT Model

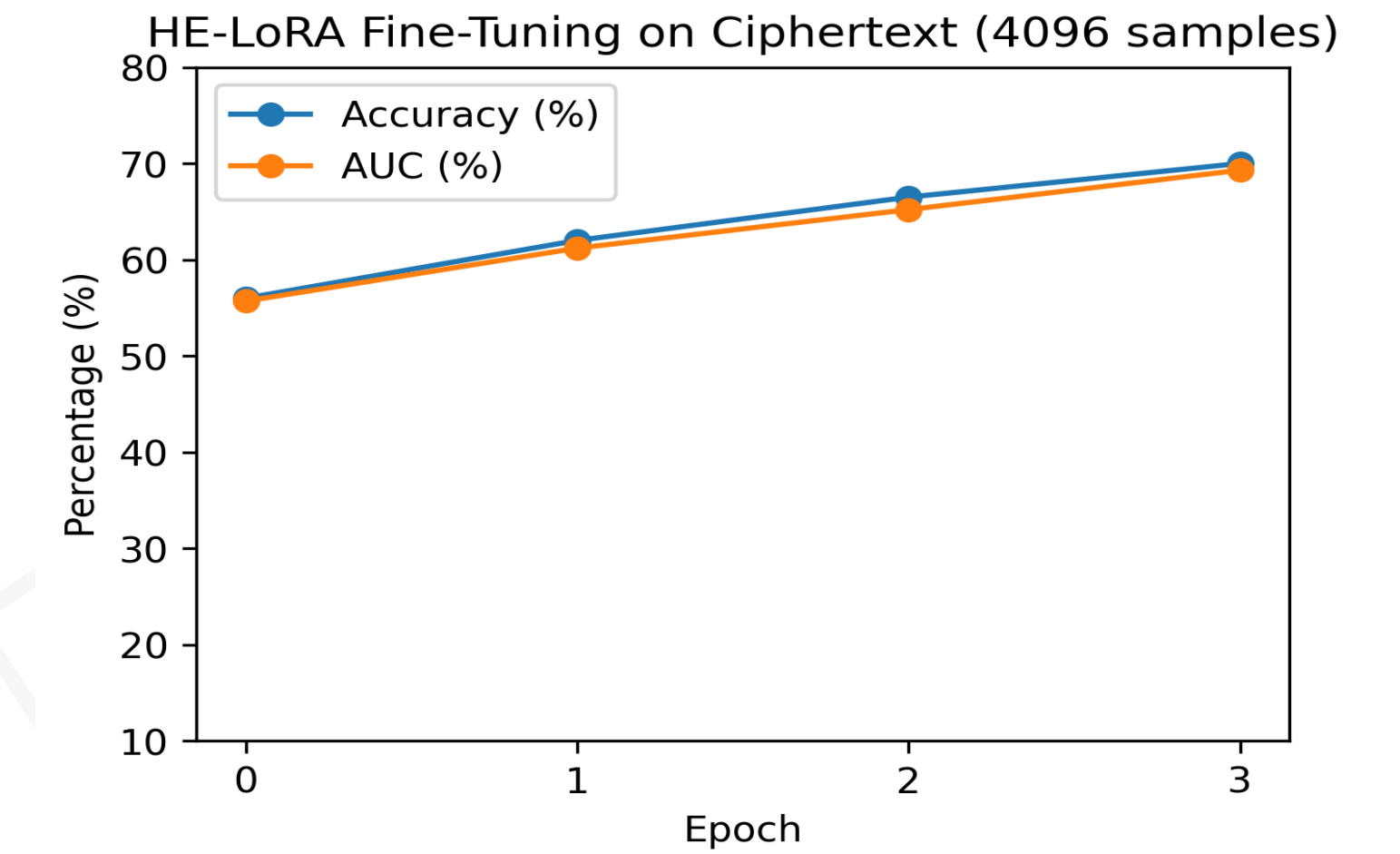
- Employed 2-layer CrammedBERT — hidden size 768, ~37M parameters.
- Embedding layer maps input tokens (vocab size 261 for IDS) to 768-dimensional vector with positional encoding and layer normalization.
- Softmax replaced with exponential function for HE compatibility.
- FFN expands $768 \rightarrow 3,072$ dimensions via ReLUGLU, projects back to 768.

Results

Full Fine-Tuning vs LoRA Performance on Plaintext



- LoRA fine-tuning (1.11% trainable parameters, 180,596 samples) achieves 99.85% accuracy — outperforming full fine-tuning (99.73%) with 90× fewer parameters.
- Therefore, LoRA is the preferred strategy for HE setting — fewer trainable parameters directly reduce encrypted computation overhead.



- **HE-LoRA fine-tuning:** 4,096 samples (2.27% of full set), 4 epochs, batch size 128, 32 steps/epoch, and took ~53 minutes on single NVIDIA RTX 6000 Ada GPU.
- Evaluated on 200 samples, the model achieves 70% accuracy on ciphertext under HE-based classification.

Contributions

- **Custom tokenizer:** Since CIC-IDS 2017 is a numerical tabular dataset incompatible with standard NLP tokenizers, we designed a custom tokenizer that converts numerical network features into BERT-compatible token sequences via robust scaling and uniform quantization.
- **First, HE-IDS:** To the best of our knowledge, this is the first work to apply HE-based LoRA fine-tuning to IDS on the CIC-IDS dataset.
- **Single GPU:** Adapted the HE-LLM from its original 8-GPU [D. Rho] setup to run on a single NVIDIA RTX 6000 Ada GPU, reducing the hardware requirements for privacy-preserving LLM fine-tuning on ciphertext.

Key Finding

- HE-LLM polynomial approximations, calibrated for NLP tasks, fail for tabular IDS features whose intermediate activations exceed the valid polynomial range — causing value overflow and accuracy degradation.
- The accuracy gap between plaintext (99.85%) and HE (70%) is due to CKKS noise & approximate arithmetic nature causing activations functions to exceed the approximation range. IDS-specific polynomial recalibration for suitable activation ranges is a key direction for future work.

References

Donghwan Rho et al. 2025. Encryption-Friendly LLM Architecture. *arXiv:2410.02486*.