Санкт-Петербургский государственный университет

Кафедра системного программирования

Разработка адаптивной системы защиты ML-моделей на основе мультиагентного подхода с гомоморфным шифрованием

Хокимзода Муборакшои Иноятулло, группа 24.М41-мм

Научный руководитель: к.ф.-м.н. Д.В. Луцив, доцент кафедры системного программирования
Консультант: В.А. Андриенко, старший преподователь кафедры системного программирования

Санкт-Петербург
2025

Введение и актульность

- Рост использования ML в банках (кредитный скоринг, выявление мошенничества)
- Проблема: ограниченные локальные ресурсы $\rightarrow$ аренда облачных серверов
- Риски утечки конфиденциальных данных
- Решение: полностью гомоморфное шифрование (FHE) и мультиагентный подход

Цель и задачи исследования

- **Цель:** Разработка ML-модели (логистическая регрессия) на зашифрованных данных для кредитного скоринга
- **Задачи:**
  - Изучить FHE и его применение в ML
  - Разработать мультиагентную архитектуру
  - Реализовать прототип с библиотекой TenSEAL
  - Сравнить производительность моделей на зашифрованных и незашифрованных данных

Теоретические основы

- Полностью гомоморфное шифрование (FHE):
  - Основы: операции сложения и умножения над зашифрованными данными
  - Схемы: BFV(Brakerski/Fan-Vercauteren), BGV(Fully Homomorphic Encryption without Bootstrapping), CKKS(Cheon-Kim-Kim-Song) (используется CKKS для вещественных чисел)
- Проблема сигмоиды: аппроксимация полиномом (минимаксная аппроксимация)
- Мультиагентные системы(MAS): автономность, координация, модульность

FHE

- **Что такое FHE?** Криптографическая техника, позволяющая выполнять вычисления (сложение и умножение) над зашифрованными данными без расшифровки.
- **История:** Впервые предложено Крейгом Джентри в 2009 году на основе решёток.
- **Математическая основа:**
  Для сообщений m_1, m_2 шифрование $\mathrm{Enc}(m)$ поддерживает:

$$\mathrm{Enc}(m_1) + \mathrm{Enc}(m_2) = \mathrm{Enc}(m_1 + m_2),$$

$$\mathrm{Enc}(m_1) \cdot \mathrm{Enc}(m_2) = \mathrm{Enc}(m_1 \cdot m_2)$$

CKKS

- Схема полностью гомоморфного шифрования (FHE), разработанная Чеонгом, Кимом, Кимом и Сонгом для работы с вещественными числами.
- Поддерживает приближённые вычисления над комплексными и вещественными числами.
- Кодирование данных в полиномы, шифрование с использованием решёток.
- Операции: сложение, умножение, масштабирование (для управления точностью).
- Bootstrapping для уменьшения шума, но с высокой вычислительной стоимостью.

Архитектура системы

- Мультиагентная система: 4 агента
  - Мониторинг: проверка данных на аномалии
  - Шифрование: CKKS с TenSEAL на клиенте
  - Передача: безопасная доставка данных
  - Анализ: логистическая регрессия на сервере
- Используемые технологии: TenSEAL, Mesa, PyTorch, Pandas, Scikit-learn

Архитектура системы

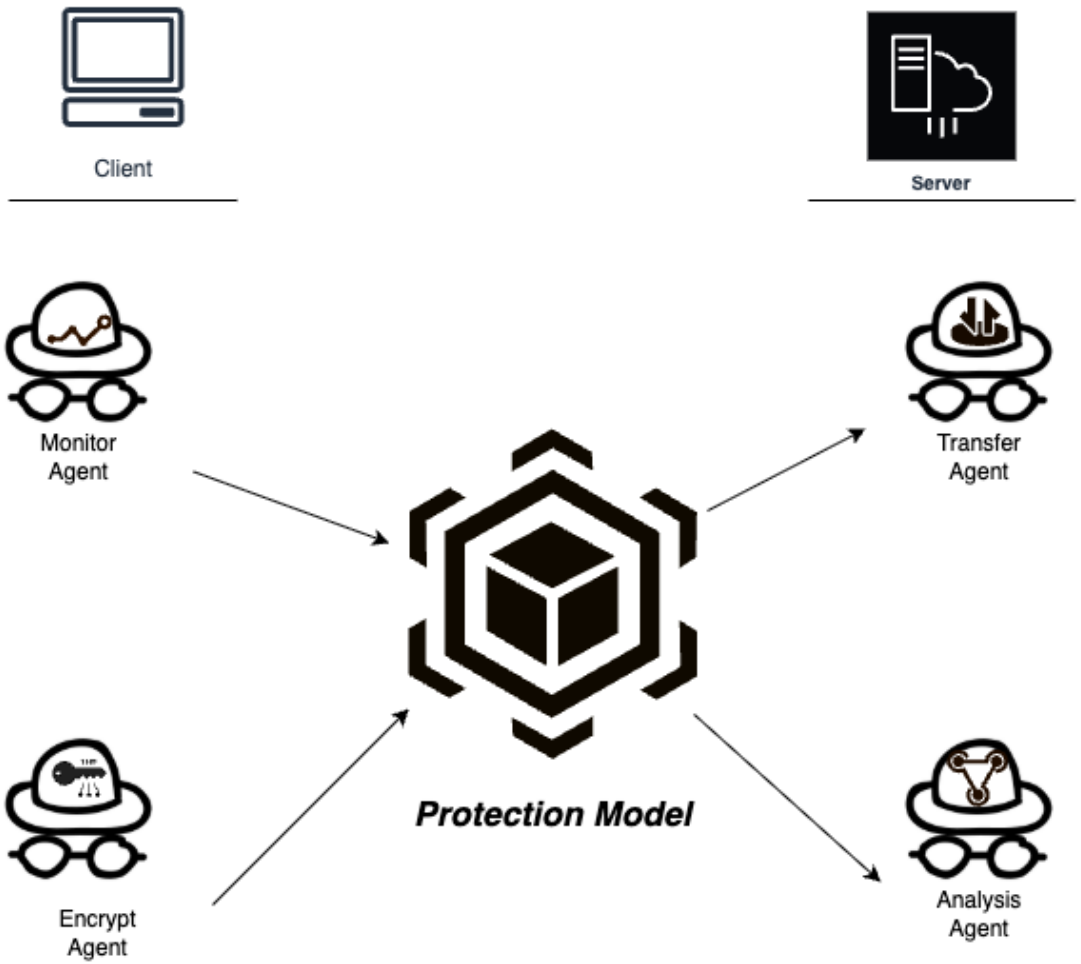

Рис.: Архитектура FHE + MAS

Реализация прототипа

- Датасет: «Credit Score Classification» (Kaggle)
- Предобработка: очистка, нормализация, логарифмирование
- Шифрование: CKKS с параметрами $N = 8192/16384$, $\Delta = 2^{21}$
- Обучение: PyTorch, логистическая регрессия (50 эпох, SGD)
- Тестирование: сравнение предсказаний на зашифрованных/незашифрованных данных

Результаты тестирования

Таблица: Сравнение производительности моделей

Конфигурация	**Точность**	**F1-мера**
Незашифрованные данные	0.3961	0.1816
Зашифрованные данные ($N = 8192$)	0.3800	0.0606
Зашифрованные данные ($N = 16384$)	0.6400	0.4375

- Выводы: FHE снижает точность на 1–5%, увеличивает время выполнения в 10–100 раз
- Преимущества: полная конфиденциальность данных

Заключение и перспективы

- Достижения: разработана система, обеспечивающая конфиденциальность с приемлемой точностью
- Ограничения: высокая вычислительная сложность FHE
- Перспективы:
  - Оптимизация алгоритмов шифрования
  - Аппаратное ускорение или облачное
  - Интеграция с дифференциальной приватностью

Источники

- MESA Documentation. [Электронный ресурс]. URL: https://mesa.readthedocs.io/stable/getting-started.html
- Microsoft SEAL Documentation. [Электронный ресурс]. URL: https://github.com/Microsoft/SEAL
- TenSEAL Documentation. [Электронный ресурс]. URL: https://github.com/OpenMined/TenSEAL?tab=readme-ov-file
- Kahya A. Machine Learning over Encrypted Data With Fully Homomorphic Encyption. – Master of Science, METU. 2022;