РусКрипто

XXVII НАУЧНО-ПРАКТИЧЕСКАЯ КОНФЕРЕНЦИЯ

Квантово-безопасное федеративное глубокое обучение

РусКрипто

Обзор на основе работы:

Авторы: Kfir Sulimany и др.
Лаборатория: MIT Research Laboratory of Electronics

Quantum-secure multiparty deep learning

Kfir Sulimany*[1], Sri Krishna Vadlamani[1], Ryan Hamerly[1,2], Prahlad Iyengar[1], and Dirk Englund[1]
[1]Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, USA
[2]Physics & Informatics Laboratories, NTT Research, Inc., Sunnyvale, CA, USA

Автор обзора: Воробей Сергей С.
Заместитель начальника отдела лицензирования и сертификации ООО «КуРэйт»

QRATE

1. Предпосылки развития в РФ

РусКрипто

Слайд из презентации Сорокина Павла Юрьевича, Первый заместитель Министра энергетики Российской Федерации

Тензорные вычислители

IVA TECHNOLOGIES

Модуль

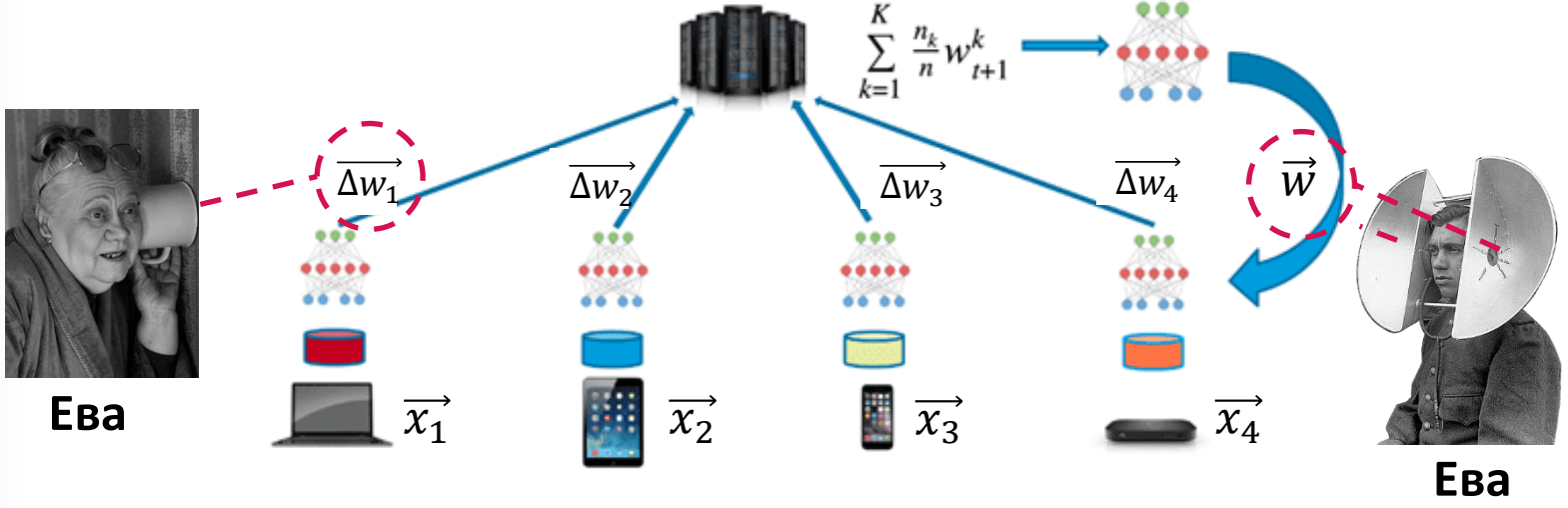
УБИ. 218	Угроза раскрытия информации о модели машинного обучения	конф.
УБИ. 219	Угроза хищения обучающих данных	конф.
УБИ. 220	Угроза нарушения функционирования («обхода») средств, реализующих технологии искусственного интеллекта	конф.
УБИ. 221	Угроза модификации модели машинного обучения путем искажения («отравления») обучающих данных	цел.
УБИ. 222	Угроза подмены модели машинного обучения	конф. цел.

2. Конфиденциальность федеративного глубокого машинного обучения

РусКрипто

Вычиление $f(\vec{x}, \vec{w}) = \overrightarrow{\Delta w}$, не раскрывая аргументов:

- $\vec{x}$ – приватные данные клиента,
- $\vec{w}$ – веса модели сервера,
- $\overrightarrow{\Delta w}$ – прожуточные градиенты



3. Недостатки гомоморфного шифрования (homomorphic encryption - HE)

РусКрипто

$$Enc(f(x)) = f(Enc(x))$$

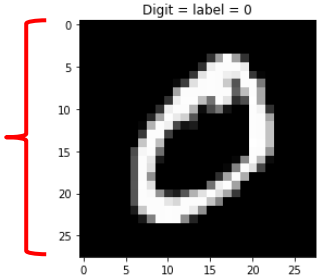
Недостатки гомоморфного шифрования:

- Вычислительная сложность
- Уязвимо к кв.вычислениям

$O(n^k)$ – сложность, где

n – размер входных данных:

- $\vec{x} \in \mathbb{R}^N$ – локальный вектор данных клиента
- $\vec{w} = vec(\boldsymbol{W}), \boldsymbol{W} \in \mathbb{R}^{N \times M}$ – вектор когерентных состояний весов. MNIST: $N = 784$ (28×28 пикселей), M — число нейронов в слое
- k – сложность ($P(2 \leq k \leq 3) \approx 1$)

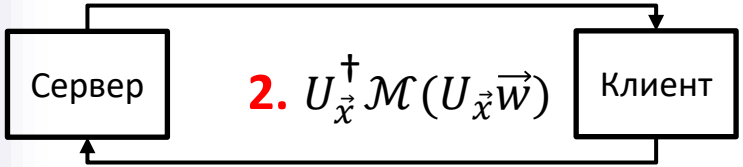


$\vec{x} = vec(\boldsymbol{Fig})$

4. Архитектура решения

$$\underbrace{\vec{w}}_{\text{сервер}} \rightarrow \underbrace{U_{\vec{x}} \rightarrow \mathcal{M}(U_{\vec{x}}\vec{w}) \rightarrow U_{\vec{x}}^{\dagger} \rightarrow \rho_v}_{\text{клиент}}$$

1. $\boldsymbol{W}_i \rightarrow \overrightarrow{w_i} = \{|\alpha_0\rangle, \dots\}$ – Когерентные состояния весов

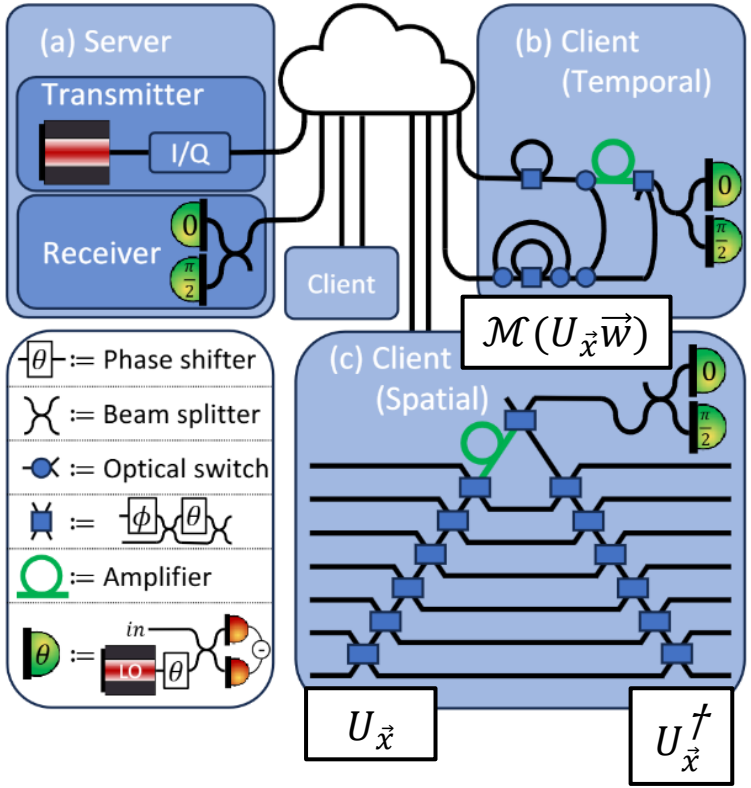


3. ρ_v – Верификационное состояние

$\alpha_j = \frac{\sqrt{\mu}\boldsymbol{W}_{ij}}{\|\boldsymbol{W}\|_{RMS}}$, где

μ – ср. число фотонов

$$\vec{x} \rightarrow \hat{x} = \frac{\vec{x}}{\|\vec{x}\|}$$



5. Операции на клиенте

$$\vec{w} \to \underbrace{U_{\vec{x}} \to \mathcal{M}(U_{\vec{x}}\vec{w})} \to U_{\vec{x}}^{\dagger} \to \rho_v$$

1 Измерение

$$U_{\vec{x}}\vec{w} = (\vec{w} \cdot \hat{x}, v_2, \dots, v_N) = (\alpha, v_2, \dots, v_N)$$

Клиент измеряет обе квадратуры ($\hat{X}$, $\hat{P}$) когерентного состояния $|\alpha\rangle$ с помощью гомодинного детектирования. $|\alpha\rangle$ имеет дисперсию в *1 SNU (Shot Noise Unit)*

$$\left\langle (\Delta\hat{X})^2 \right\rangle = \left\langle (\Delta\hat{P})^2 \right\rangle = 1\ SNU$$

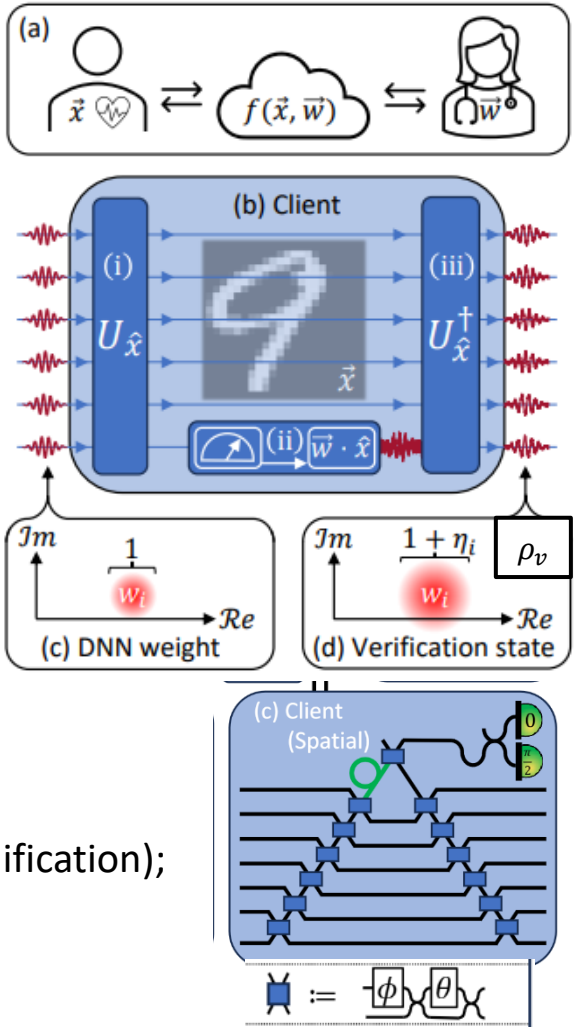
2 Прямое распространение

$$\mathcal{M}(U_{\vec{x}}\vec{w}) = (\tilde{\alpha}, v_2, \dots, v_N)$$

Операция $\mathcal{M}$ обобщается через $\mathcal{G}$, которая включает:

- Усиление с коэффициентом G (phase-insensitive amplification);
- Разделение луча с соотношением $1 - \frac{1}{G} : \frac{1}{G}$;

Где $\tilde{\alpha}$ – новое состояние в результирующей моде



6. Операции на клиенте

$$\vec{w} \to U_{\vec{x}} \to \mathcal{M}(U_{\vec{x}}\vec{w}) \to \underbrace{U_{\vec{x}}^{\dagger} \to \rho_v}$$

Прямое распространение

После измерения $U_{\vec{x}}^{\dagger}$ добавленный шум распределяется по всем модам верификационного состояния ρ_v, для i-й моды шум:

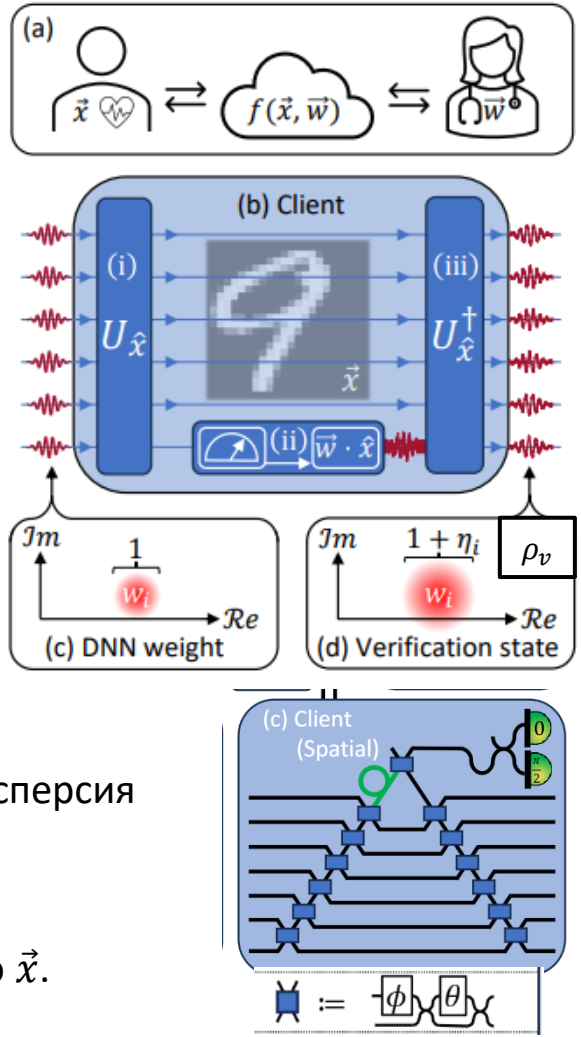
$$\eta_i = (2 - \frac{2}{G})|\hat{x}_i|^2$$

$$\rho_v = U_{\vec{x}}^{\dagger}\mathcal{M}(U_{\vec{x}}\vec{w})$$

Среднее значение ρ_v совпадает с $\vec{w}$: $\langle\tilde{\alpha}\rangle = \vec{w}\cdot\hat{x}$, но дисперсия каждой моды увеличивается на η_i:

$$\sigma^2 = 1 + \left(2 - \frac{2}{G}\right)|\hat{x}_i|^2 = 1 + \eta,$$

где η – избыточный шум, что маскирует информацию о $\vec{x}$.



7. Утечка w_i, оценка границы Холево

РусКрипто

Операция $\mathcal{M}$ добавляя квантовый шум:

- Защищает $\vec{w}$ сервера, так как клиент измеряет только проекцию $\vec{w} \cdot \hat{x}$, а остальной шум маскирует исходные веса.

$$I_{w_i} \leq \chi = S(\rho_{AB}) - S(\rho_{B|A}) \rightarrow I_{w_i} \leq g(v_1) + g(v_2) - g(v_3)$$

$$g(v) = \left(\frac{v+1}{2}\right)\log_2\left(\frac{v+1}{2}\right) - \left(\frac{v-1}{2}\right)\log_2\left(\frac{v-1}{2}\right)$$

$$a = 2\mu + 1, b = 2\mu + 1 + \eta_i, c = \sqrt{4\mu^2 + 2\mu + 1}, z = \sqrt{(a+b)^2 - 4c^2}$$

$$v_{1,2} = \frac{1}{2}(z \pm [b-a]), v_3 = a - \frac{c^2}{b+1}$$

8. Утечка x_i, оценка границы Крамера-Рао

РусКрипто

Операция $\mathcal{M}$ добавляя квантовый шум:

- Защищает $\vec{x}$ клиента, т.к. обратно высылается верификационное состояние и только в дисперсии содержится информация о данных. Сервер видит только шум η_i, зависящий от $|\hat{x}_i|^2$:

Квантовая граница Крамера-Рао (QCRB) – оценка границы дисперсии случайной квантовой величины.

$$Var(\hat{x}_i) \geq \frac{1}{M\mathcal{F}_R[\rho_x]}, \text{ где } M - \text{число измерений, } \mathcal{F}_R[\rho_x] = \frac{4|\hat{x}_i|^2\left(2-\frac{2}{G}\right)^2}{\sigma_i^4}$$

$$I_{x_i} = \frac{1}{2}\log_2\left(1+\frac{|\hat{x}_i|^2}{Var(\hat{x}_i)}\right)$$

$$I_{x_i} \leq \frac{1}{2}\log_2\left(1+k\cdot\frac{8M(G-1)^2|\hat{x}_i|^4}{G^2\sigma_i^4}\right), \qquad k=2$$

(следствие из границы Крамера-Рао)

9. Компромисс точности и безопасности

РусКрипто

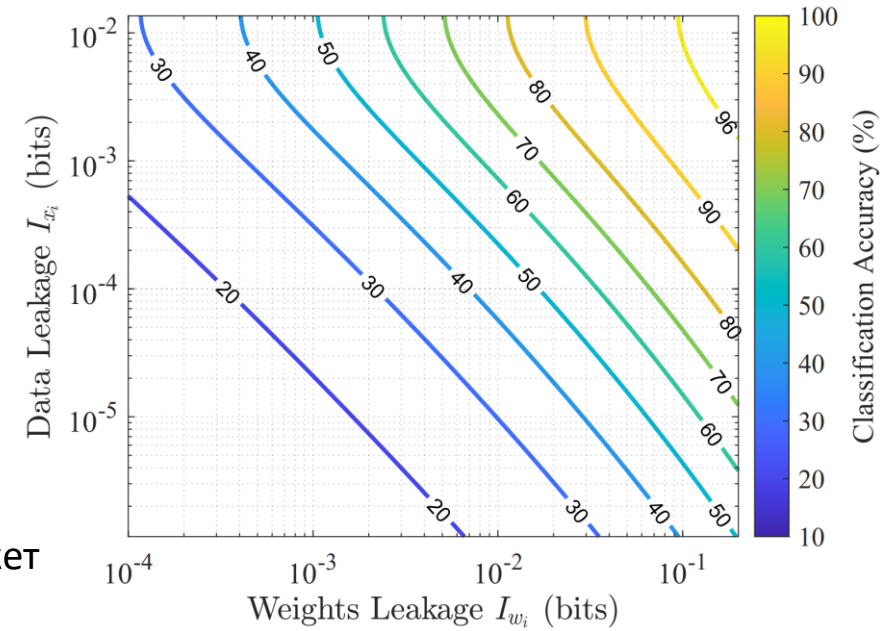
Естественный компромисс: уменьшение μ снижает I_{w_i}, но требует роста G, увеличивая I_{x_i}.

$$I_{w_i} < 0.1 \text{ (для } \mu = 4)$$

$$I_{x_i} < 0.01 \text{ (для } G = 3)$$

Современные стандарты квантования, необходимые для обеспечения минимальной точности для DNN, соответствуют 1 биту*.

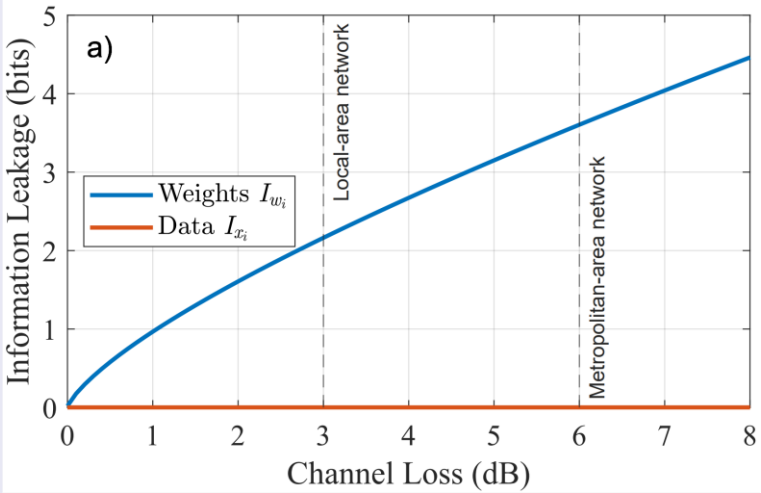
Даже зная утечку, атакующий не сможет восстановить модель без доступа к обучающим данным.

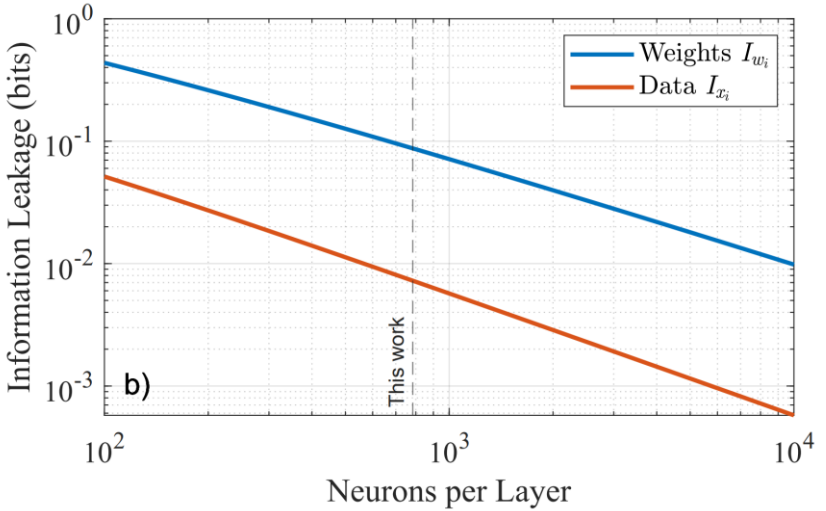


* – Amir Gholami, Sehoon Kim, Zhen Dong, Zhewei Yao,Michael W Mahoney, and Kurt Keutzer. A survey of quan-tization methods for efficient neural network inference. InLow-Power Computer Vision, pages 291–326. Chapman andHall/CRC, 2022.

10. Зависимость утечки от параметров

РусКрипто





Потери в каналах. При 6 дБ – типичных для локальных сетей – I_{w_i} растет до 4 бит, но I_{x_i} не зависит от потерь. С увеличением числа нейронов утечка падает: шум распределяется по модам. Для больших DNN, с тысячами нейронов, безопасность только возрастает.

РусКрипто

СПАСИБО
ЗА ВНИМАНИЕ

QRATE