Сергей В. Запечников
Национальный исследовательский ядерный университет «МИФИ»,
Каширское ш., 31, Москва, 115409, Россия
e-mail: SVZapechnikov@mephi.ru, http://orcid.org/0000-0002-7975-6040

КОНФИДЕНЦИАЛЬНОЕ МАШИННОЕ ОБУЧЕНИЕ НА ОСНОВЕ ТРЕХСТОРОННИХ ПРОТОКОЛОВ БЕЗОПАСНЫХ ВЫЧИСЛЕНИЙ*

DOI: http://dx.doi.org/10.26583/bit.2022.1.04

Аннотация. Статья посвящена анализу систем конфиденциального машинного обучения, основанных на концепции безопасных трехсторонних вычислений. После общих сведений о постановках задач безопасных многосторонних вычислений и конфиденциального машинного обучения проводится обзор существующих систем конфиденциального машинного обучения и перспектив их развития. Анализ работ ведущих зарубежных исследовательских коллективов позволяет выделить ряд критериев, существенных для оценки систем конфиденциального машинного обучения на основе многосторонних протоколов безопасных вычислений. Проводится сравнительная оценка систем конфиденциального машинного обучения по выделенной системе критериев. Дальнейшим предметом рассмотрения являются только системы на основе трехсторонних протоколов безопасных вычислений. Основное внимание уделяется алгоритмическим аспектам организации таких систем, реализованных в них методам и протоколам защиты информации. Рассматриваются системы, стойкие к различным типам противника, как основанные на универсальных модулях безопасных двусторонних вычислений, так и специализированные, предназначенные для обеспечения конфиденциальности конкретных методов машинного обучения, таких как нейронные сети. Подробно рассматриваются примеры прототипов таких систем. Основываясь на результатах проведенного анализа, формулируются выводы о перспективах развития систем конфиденциального машинного обучения, ставятся задачи продолжения исследований.

Ключевые слова: конфиденциальное машинное обучение, безопасные многосторонние вычисления, схемы разделения секрета, гомоморфное шифрование.

Для цитирования: ЗАПЕЧНИКОВ, Сергей В. КОНФИДЕНЦИАЛЬНОЕ МАШИННОЕ ОБУЧЕНИЕ НА ОСНОВЕ ТРЕХСТОРОННИХ ПРОТОКОЛОВ БЕЗОПАСНЫХ ВЫЧИСЛЕНИЙ. Безопасность информационных технологий, [S.l.], т. 29, № 1, с. 30–43, 2022. ISSN 2074-7136. URL: https://bit.mephi.ru/index.php/bit/article/view/1400. DOI: http://dx.doi.org/10.26583/bit.2022.1.04.

***Благодарности.** Работа выполнена при поддержке Министерства науки и высшего образования РФ (проект государственного задания № 0723-2020-0036).*

Sergey V. Zapechnikov
National Research Nuclear University MEPhI (Moscow Engineering Physics Institute),
Kashirskoe shosse, 31, Moscow, 115409, Russia
e-mail: SVZapechnikov@mephi.ru, http://orcid.org/0000-0002-7975-6040

Privacy-preserving machine learning based on secure three-party computations*

DOI: http://dx.doi.org/10.26583/bit.2022.1.04

Abstract. The paper is devoted to the analysis of privacy-preserving machine learning systems based on the concept of secure three-party computations. After general information about the purposes of secure multi-party computations and privacy-preserving machine learning, an overview of existing privacy-preserving machine learning systems and perspectives for their development is offered. An analysis of the work of leading foreign research teams allows to identify several criteria essential for evaluating privacy-preserving machine learning systems based on multi-party secure computations. A comparative analysis of privacy-preserving machine learning systems is carried out according to a dedicated system of criteria. The further

subject of consideration is only systems based on three-party secure computations. The main attention is paid to the algorithmic aspects of the organization of such systems, the methods and protocols of information security implemented in them. Systems secure to various types of adversary are considered, both based on universal modules of secure two-party computations, and specialized ones designed to ensure the privacy of specific machine learning methods, such as neural networks. Examples of prototypes of such systems are considered in detail. Based on the results of the analysis, conclusions are made about the prospects for developing privacy-preserving machine learning systems, and the tasks of future research are described.

Keywords: privacy-preserving machine learning, secure multi-party computations, secret sharing scheme, homomorphic encryption.

For citation: ZAPECHNIKOV, Sergey V. Privacy-preserving machine learning based on secure three-party computations. IT Security (Russia), [S.l.], v. 29, n. 1, p. 30–43, 2022. ISSN 2074-7136. URL: https://bit.mephi.ru/index.php/bit/article/view/1400. DOI: http://dx.doi.org/10.26583/bit.2022.1.04.

***Acknowledgement.** This work was supported by the Ministry of Science and Higher Education of the Russian Federation (state assignment project No. 0723-2020-0036).*

Введение

Безопасные многосторонние вычисления (БМВ) – одно из важнейших направлений развития современной криптографии. Напомним постановку задачи БМВ [1]. Рассматривается многосторонний криптографический протокол, в котором каждый из участников имеет свой индивидуальный секрет. Требуется вычислить заданную функцию, аргументами которой являются эти секреты, так чтобы результат вычислений был известен всем участникам группы, но сами секреты не были разглашены участниками протокола ни друг другу, ни какой-либо третьей стороне. А именно, пусть участники криптографического протокола $P_1, P_2, \dots, P_n$ имеют конфиденциальные входные данные $x_1, x_2, \dots, x_n$ соответственно. В результате выполнения протокола ими совместно должна быть вычислена функция вида $y = f(x_1, x_2, \dots, x_n)$, при этом протокол должен обладать следующими двумя свойствами:

- *корректностью*: каждый из участников $P_1, P_2, \dots, P_n$ получает y;
- *приватностью* (конфиденциальностью): никому из участников либо третьих лиц не разглашается никакая дополнительная информация, кроме той, которую они знали до начала выполнения протокола.

Частным случаем БМВ можно считать задачу конфиденциального машинного обучения (КМО). Целью КМО является обеспечение конфиденциальности данных каждого из участников системы машинного обучения в условиях, когда лица, предоставляющие обучающую выборку на этапе обучения модели (training) либо запросы к модели на этапе ее эксплуатации (inference) и ожидающие получения ответов на свои запросы (клиенты), дистанционно взаимодействуют с провайдером, способным выполнять вычисления с помощью этой модели (сервером). Задача КМО может решаться при помощи БМВ с разным числом участников. Настоящая статья посвящена преимущественно исследованию случая, когда КМО реализуется на основе трехсторонних протоколов безопасных вычислений, т.е. в приведенной выше постановке задачи n=3.

Предлагаемая вниманию читателя статья является продолжением исследования, посвященного КМО на основе двусторонних протоколов безопасных вычислений [2].

1. Системы КМО и перспективы их развития

Как показывает анализ научной литературы, в настоящее время теоретические и прикладные исследования в сфере разработки и реализации систем КМО выполняет не менее 10 научных коллективов, рассредоточенных по всему миру. В связи с высоким

темпом научных исследований в области систем КМО рассматривались только системы, созданные за последние три года (2019–2021 гг.). Далее приведем краткие сведения о работах каждого из коллективов.

1. *Коллектив международного исследовательского подразделения корпорации Microsoft*. Усилия коллектива сосредоточены на создании систем КМО двухуровневой архитектуры, в которых клиентские компоненты позволяют интерпретировать описания моделей машинного обучения, выполненные с помощью средств библиотеки TensorFlow во внутреннее представление, а серверные компоненты – автоматически исполнять протоколы БМВ, реализующие вычисления, при помощи модулей с набором универсальных двухсторонних и трехсторонних протоколов безопасных вычислений.

Основные работы коллектива:

- система SecureNN (2019 г.) [3];
- система EzPC (2019 г.) [4];
- система CrypTFlow (2020 г.) [5];
- система CrypTFlow2 (inference, 2020) [6].

2. *Исследовательская группа Дармштадтского технического университета* (ФРГ). Основное направление работы коллектива в области систем КМО – реализация универсальных средств исполнения двусторонних протоколов безопасных вычислений на основе сочетания представления вычисляемых функций в виде арифметических, булевых и искаженных схем (garbled circuits), которые могут использоваться в виде готового ядра при создании отдельных приложений, включая федеративное обучение, обработку медицинских изображений методами машинного обучения и пр.

Основные работы коллектива:

- модуль ABY (2015 г.) [7];
- система MP2ML (2020 г.) [8];
- модуль ABY 2.0 (2020 г.) [9];
- система FLGuard (2021 г.) [10].

3. *Исследовательская группа Калифорнийского университета в Беркли* (UC Berkeley, США). Коллектив работает в области создания систем КМО для получения ответов на запросы, содержащие конфиденциальную информацию, к уже обученным моделям на основе двусторонних протоколов безопасных вычислений с усиленными свойствами, включая самую «сильную» модель нарушителя – модель злоумышленного клиента.

Основные работы коллектива:

- система Delphi (2018 г.) [11];
- экспериментальные системы и прототипы Visor, Bost, Cerebro (2019–2021 гг.) [12];
- система Muse (2021 г.) [13].

4. *Исследовательская группа Индийского института наук в Бангалоре* (Indian Institute of Science, Bangalor). Деятельность научной группы сосредоточена на создании систем КМО преимущественно для глубоких нейронных сетей на основе четырехсторонних протоколов безопасных вычислений с возможностью реализации некоторых систем и на трехсторонних протоколах.

Основные работы коллектива:

- Trident (2020 г.) [14];
- FLASH (2020 г.) [15];
- Blaze (2020 г.) [16];

- SWIFT (2021 г.) [17];
- Tetrad (2021 г.) [18].

Совместно с Дармштадтским университетом члены исследовательской группы участвовали в разработке модуля ABY 2.0 [9].

5. *Группа исследователей из компании Facebook и Visa Research*. Деятельность членов команды сосредоточена на создании универсального модуля для трехсторонних протоколов безопасных вычислений на основе сочетания арифметических, булевых и искаженных схем, а также создания прикладных систем КМО на его основе. В настоящее время основное внимание уделяется протоколам и системам конфиденциальной кластеризации.

Основные работы коллектива:

- SecureML (2017 г.) [19];
- ABY^3 (Arithmetic-Binary-Yao) framework (2018 г.) [20];
- K-means clustering (2020 г.) [21].

6. *Исследовательская группа Принстонского университета* (США) занимается разработкой систем КМО на основе трехсторонних протоколов безопасных вычислений со все более строгими моделями нарушителей.

Основные работы коллектива:

- SecureNN (2019 г., совместно с Microsoft) [3];
- FALCON (2021 г.) [22];
- Ponytail (2012–2021 гг.) [23].

7. *Международная исследовательская группа Национального института индустриальных наук и технологий Японии, корпорации NTT и университета Санкт-Галлен* (Швейцария). Имеются сведения об одной разработке этого коллектива – системе КМО Adam для глубоких нейросетей, поддерживающая расширенную по сравнению с известными функциональность при обучении и применении нейросетей [24]. Система основана на трехсторонних протоколах безопасных вычислений.

8. *Исследовательская группа Массачусетского технологического института* (США). Имеются сведения об одной разработке этой группы – системе Gazelle (2018 г.) [25] на основе двусторонних протоколов. В настоящее время отдельные идеи этой разработки используются в более новых системах КМО, а сама система Gazelle представляет исторический интерес.

9. *Исследовательская группа университета Аальто* (Финляндия). Имеются сведения об одной разработке этой группы – системе MiniONN (2017 г.) [26], которая представляет лишь исторический интерес, поскольку уступает более новым системам КМО по всем основным показателям.

10. Исследовательская группа Парижского университета (Франция). Имеются сведения об одной разработке этой группы – системе AriaNN [27], которая также представляет лишь исторический интерес, поскольку уступает более новым системам КМО по всем основным показателям.

2. Критерии оценки систем КМО

Анализ работ исследовательских коллективов позволяет выделить ряд критериев, существенных для оценки разработанных и реализованных систем КМО на основе протоколов БМВ. Далее охарактеризуем их подробнее.

1. *Количество сторон в протоколах БМВ*, реализующих функциональность систем КМО:

1.1) двусторонние;

1.2) трехсторонние;

1.3) четырехсторонние.

Некоторые системы КМО позволяют реализовывать функциональность посредством протоколов с разным числом участников. В то же время систем с количеством участников вычислений более четырех в ходе настоящего исследования обнаружено не было.

2. *Криптографические примитивы*, используемые для реализации системы:

2.1) схемы разделения секрета;

2.2) искаженные схемы (garbled circuits);

2.3) схемы гомоморфного шифрования.

Для большинства систем характерно сочетание двух или даже всех трех перечисленных типов криптографических примитивов, хотя есть попытки построить системы КМО, используя только один вид примитивов, но они, как правило, обладают ограниченной функциональностью.

3. *Модель нарушителя*, в предположении о которой разрабатывалась система КМО и в которой обеспечивается ее криптографическая стойкость:

3.1) получестный нарушитель;

3.2) злоумышленный нарушитель.

Подавляющее большинство систем, основанных на двусторонних протоколах безопасных вычислений, обеспечивают стойкость к получестному нарушителю, в то время как абсолютное большинство систем, основанных на трех- и четырехсторонних протоколах обеспечивает стойкость как к получестному, так и к злоумышленному нарушителю.

4. *Поддержка стадий жизненного цикла машинного обучения*:

4.1) обучение моделей (training);

4.2) применение моделей для получения прогнозных ответов на запросы пользователей (inference).

Стадия обучения моделей является многократно (иногда на несколько порядков величины) более трудоемкой, чем их применение. В то же время обучение модели – относительно нечасто выполняемая операция по сравнению с последующим применением обученной модели. Как показывает анализ литературы, большая часть систем КМО в настоящее время поддерживает лишь стадию применения уже обученных моделей, в то же время ряд систем поддерживают обе стадии.

5. *Поддержка методов машинного обучения*:

5.1) элементарных статистических и логических методов машинного обучения (линейная регрессия, логистическая регрессия, кластеризация, решающие деревья);

5.2) полносвязных нейронных сетей;

5.3) глубоких нейронных сетей;

5.4) специальных приемов обучения и применения нейронных сетей для повышения точности прогнозирования, производительности, сходимости параметров сети и т.п., таких как пакетная нормализация, оптимизация по методу Adam и др.

Как показывает анализ, среди систем КМО преобладают разработки для обеспечения безопасности глубоких нейронных сетей, прежде всего, сверточных. Нарастает количество работ, посвященных обеспечению конфиденциальности при использовании специальных приемов обучения нейросетей, часто используемых на практике.

6. *Пригодность для использования в различных коммуникационных архитектурах*:

6.1) локальных компьютерных сетях (LAN);

6.2) глобальных компьютерных сетях (WAN).

Системы КМО, которые реализуются посредством протоколов БМВ с большой коммуникационной сложностью, а также со сбалансированными требованиями к

коммуникационным и вычислительным ресурсам участников значительно лучше подходят для локальных сетей. В то же время системы КМО, предназначенные для глобальных сетей, должны минимизировать коммуникационные требования к участникам за счет более высоких вычислительных требований.

7. *Объем массивов данных*, использованных для апробации систем КМО:

7.1) массивы данных относительно малого объема;

7.2) «большие данные».

Многие системы КМО, которые показывают хорошие результаты при экспериментах на относительно малых массивах данных (например, датасет MNIST, часто используемый в качестве эталона для апробации алгоритмов классификации), могут оказаться непрактичными из-за неприемлемо большого времени работы на массивах, представляющих практический интерес. В связи с этим большое значение имеет апробация экспериментальных систем КМО на массивах данных объема, сопоставимого с тем, который будет встречаться при практическом использовании (например, таких как известный эталон CIFAR-10).

8. *Архитектуры нейросетей*, для которых апробированы системы КМО:

8.1) относительно простые нейросети с небольшим количеством слоев (например, LeNet, AlexNet и т.п.);

8.2) глубокие нейросети с числом слоев порядка 50–200 (например, VGG-16, ResNet, DenseNet).

Практический интерес представляют такие системы КМО, которые могут эффективно работать с глубокими нейросетями, получившими наибольшее практическое применение.

3. Сравнительная оценка систем КМО на основе многосторонних протоколов безопасных вычислений

Проанализированные в ходе настоящего исследования системы КМО могут быть охарактеризованы по критериям, перечисленным в предыдущем разделе. Результаты оценки приведены в табл. 1.

Таблица 1. Результаты сравнительной оценки систем КМО

№ п/п	Системы КМО	Критерии оценки																			
		1			2			3		4		5				6		7		8	
		1.1	1.2	1.3	2.1	2.2	2.3	3.1	3.2	4.1	4.2	5.1	5.2	5.3	5.4	6.1	6.2	7.1	7.2	8.1	8.2
1	SecureNN		+	+/-	+			+	+/-	+	+		+	+	+/-	+	+	+		+	
2	EzPC	+			+	+		+			+	+	+			+	+	+	+	+	
3	CrypTFlow	+	+		+			+	+		+			+		+		+	+	+	+
4	CrypTFlow2	+			+			+			+			+		+		+	+	+	+
5	ABY	+			+	+		+		+	+	+				+		+		+	
6	Delphi	+			+	+	+	+			+		+			+		+	+	+	+/-
7	Muse	+			+	+	+	+	+		+		+			+		+	+	+	
8	Trident			+	+	+		+	+	+	+		+			+	+	+		+	
9	FLASH			+	+			+	+		+		+	+		+	+	+	+/-	+	
10	Blaze		+		+	+		+	+	+/-	+	+	+			+	+	+		+	
11	Tetrad			+	+	+		+	+	+	+		+			+	+	+	+	+	+/-
12	SecureML	+			+	+	+	+		+	+	+	+			+	+/-	+		+	
13	ABY[3]		+		+	+		+	+	+	+	+				+	+/-	+		+	
14	Falcon		+		+			+	+	+	+		+	+	+	+	+	+	+	+	+
15	Gazelle	+			+	+	+	+			+		+			+		+	+	+	
16	MiniONN	+			+	+	+	+			+	+	+			+		+	+	+	
17	Система [24]		+		+			+	+	+	+			+	+	+	+	+	+	+	+

Нумерация критериев оценки соответствует введенной в п. 2 настоящей статьи. Условные обозначения: «+» – соответствие критерию (наличие свойства), «-» – несоответствие критерию (отсутствие свойства), «+/-» – частичное соответствие критерию (наличие части свойств).

Эти результаты свидетельствуют о том, что в качестве основного классификационного признака систем КМО целесообразно использовать количество сторон в протоколах БМВ, реализующих функциональность систем КМО. Этот признак позволяет вполне определенно разделить все системы на три класса – системы, основанные на двусторонних, трехсторонних и четырехсторонних протоколах.

Дальнейшим предметом рассмотрения в настоящей статье являются системы КМО на основе трехсторонних протоколов безопасных вычислений.

4. Системы КМО на основе трехсторонних протоколов безопасных вычислений

Основное внимание будем уделять алгоритмическим аспектам организации систем, а также методам и протоколам защиты данных в них.

Модуль ABY³. Модуль ABY^3 [20] задуман и реализован как виртуальный процессор, выполняющий набор базовых операций для трехсторонних протоколов безопасных вычислений над целыми числами. Основная идея заключается в использовании трех форм разделения секрета: арифметического, булева и Яо-разделения, вычислений с использованием соответствующих схем и переключений между ними для выбора наиболее производительного протокола вычислений. Модуль реализован на языке C++. Среди участников протокола допускается наличие не более одного нарушителя: получестного либо злоумышленного.

Внешне идеи модуля ABY^3 выглядят аналогично идеям, заложенным в основу модуля ABY [9], рассмотренного в [2], однако криптографические протоколы сильно отличаются в связи с тем, что разделение секрета здесь трехстороннее.

Для арифметического разделения секретного числа $x \in \mathbb{Z}_{2^k}$ (здесь принято, что k=64) между тремя участниками выбирается три случайных числа $x_1, x_2, x_3 \in \mathbb{Z}_{2^k}$ таких, что $x = x_1 + x_2 + x_3$. Доли секрета распределяются между участниками парами: $\{(x_1, x_2), (x_2, x_3), (x_3, x_1)\}$, где i-й участник протокола хранит i-ю пару долей секрета.

Определяется ряд базовых операций над разделенными секретами: сложение, умножение, разделения на доли нулевого секрета, разделения на доли случайного секрета, сборка секрета из долей, разделение секрета на доли.

Булево разделение секрета определяется как частный случай арифметического при k=1, а вместо операций сложения, вычитания и умножения используются операции $\oplus, \wedge$.

Наиболее оригинальной частью модуля ABY^3 можно считать специальную трехстороннюю схему Яо-разделения секрета для использования в трехстороннем варианте искаженной схемы [28].

Модуль ABY^3 поддерживает следующий набор трехсторонних протоколов безопасных вычислений:

- умножение целых чисел с фиксированной запятой: $z = xy$, где $x, y \in \mathbb{Z}_{2^k}$;
- скалярное умножение двух векторов целых чисел с фиксированной запятой: $\vec{z} = \vec{x} \cdot \vec{y} = \sum_{i=1}^{n} x_i y_i$, где $x, y \in \left(\mathbb{Z}_{2^k}\right)^n$;
- конвертирование долей между различными формами разделения: арифметической, булевой, Яо-разделением;
- умножение разделенного целого числа на разделенный бит;

- вычисление кусочно-полиномиальной функции: пусть $f_1, \dots, f_m$ – многочлены с общеизвестными коэффициентами, и $-\infty = c_0 < c_1 < \cdots < c_{m-1} < c_m = \infty$, такие что

$$f(x) = \begin{cases} f_1(x), x < c_1 \\ f_2(x), c_1 \le x < c_2 \\ \dots\dots\dots \\ f_m(x), c_{m-1} \le x \end{cases},$$

– протокол позволяет вычислить функцию $f(x) = \sum_i b_i f_i(x)$, где $b_1, \dots, b_m \in \{0,1\}$ – вектор разделенных секретных битов таких, что $b_i = 1$ тогда и только тогда, когда $c_{i-1} < x \le c_i$.

Авторы апробировали модуль ABY[3] для обучения и применения линейной регрессии, логистической регрессии, полносвязной трехслойной нейронной сети с функциями активации ReLU, а также сверточной нейронной сети с двумя слоями свертки. При применении обученной сети авторами получено приемлемое время вычислений (6–10 мс) и объем передаваемых данных (порядка 5 МБ). Следует, однако, отметить, что нейросети и датасеты, на которых проводились эксперименты, слишком упрощены по сравнению с моделями, представляющими практический интерес. Исследование влияния протоколов на точность предсказаний моделей не проводилось.

Система CrypTFlow. CrypTFlow представляет собой систему КМО, предназначенную для использования на стадии применения обученных моделей для получения прогнозных ответов на запросы пользователей [5]. Она конвертирует исходный код модели, описанный программистом на языке библиотеки TensorFlow в протоколы БМВ без необходимости для программиста вникать в детали криптографических конструкций.

Архитектура системы CrypTFlow – двухуровневая и включает в себя три компонента. Компонент уровня фронтенд – модуль Athos, который транслирует код библиотеки TensorFlow во внутреннее представление системы. Компоненты уровня бэкенд – модули Porthos и Aramis, которые транслируют функции, записанные модулем Athos на языке внутреннего представления системы CrypTFlow, в протоколы БМВ, стойкие в модели получестного нарушителя.

Модуль Porthos обеспечивает сборку из криптографических примитивов трехстороннего протокола безопасных вычислений. Встроенные в модуль примитивы реализуют функциональность линейных и нелинейных слоев сверточных нейронных сетей. К первому типу относится слой свертки, для чего используется трехсторонний протокол умножения матриц. Ко второму типу относятся функции активации ReLU и Maxpool, для чего используются протоколы безопасного вычисления старшего бита целого числа и конвертации долей секрета.

Модуль Aramis конвертирует любой протокол БМВ, стойкий в модели получестного нарушителя, в протокол, стойкий к злоумышленному нарушителю, с использованием аппаратных функций защиты, которые обеспечиваются процессором. Созданная авторами системы CrypTFlow реализация опирается на функции Intel SGX, однако возможно использование аналогичных функций других процессорных архитектур, например, ARM TrustZone.

В качестве бэкенда может также использоваться модуль ABY [7], который обеспечивает сборку двустороннего протокола безопасных вычислений, реализующего функциональность, описанную на языке внутреннего представления модулем Athos.

Экспериментально продемонстрирована работоспособность системы CrypTFlow на сверточных сетях ResNet50 и DenseNet121 на тестовом датасете ImageNet. Среднее время получения клиентом ответа на свой запрос составило около 30 с при использовании модулей, обеспечивающих стойкость в модели получестного нарушителя, и около 2 мин – в модели злоумышленного нарушителя. Суммарный объем передаваемых в протоколе данных около 7–10 ГБ.

Таким образом, система CrypTFlow может считаться вполне практичной по критериям времени выполнения и точности прогнозирования при двусторонних и трехсторонних вычислениях прогнозных ответов на запросы пользователей по обученной нейросети.

Система SecureNN. SecureNN – это система КМО, поддерживающая трех- и четырехсторонние вычисления при обучении и применении глубоких нейронных сетей [3]. Криптографическая стойкость обеспечивается в модели получестного противника.

В основе системы лежат новые протоколы безопасных вычислений для различных блоков нейросетей:

- умножения матриц;
- вычисления функции ReLU (rectified linear units);
- пулинга по максимальному значению (maxpool);
- пакетной нормализации.

Эти блоки позволяют конструировать трех- и четырехсторонние протоколы, стойкие в теоретико-информационном смысле, для обучения и предсказаний с использованием глубоких нейросетей, в том числе сверточных. Ни одна из сторон протокола не имеет полного доступа к обрабатываемым в протоколе данным. Однако количество участников, обладающих долями входных и выходных данных, в общем случае может быть меньше количества участников, выполняющих вычисления. Цель – построить протоколы для вычисления линейных и нелинейных функций так, чтобы они могли легко комбинироваться между собой.

Скорость вычислений значительно повышается из-за отказа от искаженных схем при вычислениях нелинейных функций. Традиционный подход состоял в использовании арифметических схем для вычисления линейных функций, применяя тройки Бивера (Beaver’s triplets) и гомоморфное шифрование, а также булевых схем для вычисления нелинейных функций, используя искаженные схемы. Для совмещения двух типов вычислений необходима также конверсия арифметических схем в булевы и обратно, которая требует немалых вычислительных затрат.

Основные криптографические протоколы системы Secure NN следующие:

- трехсторонний протокол умножения матриц, составленных из элементов конечного поля $\mathbb{Z}_{2^k}$, который может быть преобразован в четырехсторонний протокол;
- протокол конфиденциального сравнения разделенного на доли числа x с числом r, который позволяет участникам получить ответ 1, если $x > r$, и 0 в противном случае;
- протокол вычисления старшего бита (MSB) разделенного на доли числа x;
- протокол вычисления функции ReLU(x);
- протокол вычисления производной функции ReLU’(x);
- протокол целочисленного деления разделенных на доли чисел;
- протокол нормализации множества разделенных на доли чисел;
- протокол пулинга по максимальному значению для множества разделенных на доли чисел.

Доказательства криптографической стойкости всех перечисленных протоколов проведены в модели универсальной компонуемости (UC-security).

Система Falcon. Falcon – система КМО на основе трехсторонних протоколов безопасных вычислений, предназначенная для использования на стадиях обучения и применения глубоких нейросетей, поддерживающая, в отличие от ранее известных систем КМО, операцию пакетной нормализации входных данных [22]. Пакетная нормализация

играет существенную роль в обучении нейросетей, позволяя ускорить обучение и улучшить сходимость алгоритмов.

Система Falcon обеспечивает стойкость в модели злоумышленного нарушителя, предполагая, что большинство участников протокола являются честными (т.е. допускается не более одного получестного либо злоумышленного нарушителя). В случае обнаружения воздействия злоумышленного нарушителя выполнение протокола прерывается (англ. security with abort).

Система Falcon основана на идеях SecureNN [3] и ABY^3 [20], используя их в комбинации для повышения производительности. Однако есть и существенные отличия от этих систем. Для обеспечения стойкости к злоумышленному нарушителю вместо (2,2)-пороговой используется (2,3)-пороговая СРС, что приводит к очень существенным изменениям в криптографических примитивах и протоколах.

В системе Falcon реализованы следующие базовые криптографические конструкции, используемые в качестве примитивов:

- вычисление линейных комбинаций разделенных секретов;
- умножение разделенных секретов;
- матричное умножение и вычисление сверток (операция кросс-корреляции) разделенных секретов;
- восстановление секретов из долей;
- протокол выбора долей одного из двух разделенных секретов x или y в зависимости от значения бита выбора c;
- вычисление XOR-суммы разделенного секрета с публично известным битом;
- вычисление долей числового значения вида $(-1)^{\beta} \cdot x$ из долей секретов x и β.

На их основе реализованы следующие криптографические протоколы:

- конфиденциального сравнения разделенного секрета $x \in \mathbb{Z}_p$ с открытым общеизвестным числом r;
- вычисления бита переноса при сложении долей двух и трех разделенных секретов;
- вычисления нелинейных функций активации слоев нейросети ReLU(a) и ее производной DReLU(a) с разделенным секретом a;
- вычисления функции пулинга по максимальному значению $Maxpool(a_1, a_2, \dots, a_n)$ над разделенными секретами $a_1, a_2, \dots, a_n \in \mathbb{Z}_L$, а также производной этой функции;
- вычисления функции целочисленного деления двух разделенных секретов a/b, $a, b \in \mathbb{Z}_L$;
- вычисления функции пакетной нормализации разделенных секретов $a_1, a_2, \dots, a_m \in \mathbb{Z}_L$, где m – размер пакета.

Все протоколы – трехсторонние. Для каждого протокола доказаны теоремы об их криптографической стойкости в модели злоумышленного нарушителя.

Система Attrapadung, Hamada, Ikarashi и др. В [24] описана система КМО, не имеющая собственного наименования, предназначенная для конфиденциального обучения и применения глубоких нейросетей. Она поддерживает достаточно развитые функции и операции, характерные для современных нейросетевых моделей, такие как адаптивная оценка моментов (Adam) и вычисление функций многоклассовой логистической регрессии (softmax), не прибегая к аппроксимациям.

Криптографическая стойкость протоколов обеспечивается в предположении о наличии не более одного получестного или злоумышленного нарушителя.

В системе определены три типа данных:

- двоичные величины – элементы кольца $\mathbb{Z}_2$;
- l-битные целые числа со знаком и без знака;
- l-битные рациональные числа со знаком и без знака.

Для разделения секретных величин используется три вида СРС:

- (2,3)-пороговые СРС над $\mathbb{Z}_p$;
- (2,3)-пороговые СРС над $\mathbb{Z}_2$;
- простое аддитивное разделение секрета над $\mathbb{Z}_p$.

Определяются операции конвертации долей секретов из одной формы представления в другую и восстановления секрета.

Поддерживается следующий набор базовых операций над разделенными секретами:

- протокол деления разделенного секрета на открытый общеизвестный делитель;
- протокол вычисления долей секрета, обратного на множестве рациональных чисел к заданному разделенному секрету;
- протокол деления разделенного секрета на разделенный секретный делитель;
- протокол вычисления $\sqrt{x}$ и $\frac{1}{\sqrt{x}}$ для разделенного секрета x;
- протокол вычисления e^x для разделенного секрета x.

Протоколы могут быть адаптированы для использования на множестве целых чисел со знаком.

Авторы апробировали систему на широко известных глубоких нейросетях AlexNet и VGG16. Эксперименты показали, что быстродействие системы превышает систему Falcon от 10 до 40 раз [22] при показателях точности 70–75% на массиве данных CIFAR-10.

Заключение

В ходе работы выполнено поисковое исследование и проведен обзор существующих систем КМО, которые реализованы преимущественно в виде прототипов и лабораторных образцов. Главное внимание уделено принципам и технологиям реализации систем КМО на основе трехсторонних протоколов безопасных вычислений. Проведен обзор архитектур и математического обеспечения наиболее известных систем КМО, основанных на трехсторонних протоколах безопасных вычислений.

Показано, что основными алгоритмическими инструментами при создании систем КМО на основе трехсторонних протоколов безопасных вычислений служат СРС. В системах КМО используются три вида СРС: арифметическое, булево и Яо-разделения, которые позволяют выполнять безопасные вычисления с разделенными секретами для функций, представленных в форме арифметических, булевых либо искаженных схем соответственно.

Анализ систем КМО на основе трехсторонних протоколов безопасных вычислений позволяет выделить две ведущих линии их развития, сходных с теми, которые были обнаружены при анализе систем КМО на основе двусторонних протоколов безопасных вычислений:

- системы с ядром на основе универсальных программных модулей с набором базовых операций, реализующих произвольную функциональность протокола безопасных вычислений, ограниченную лишь сложностью выполнения протокола (например, модули ABY[3]);
- специализированные КМО с набором криптографических примитивов, оптимизированным для достижения высоких показателей производительности и точности

вычислений, но предназначенных для сравнительно узкого круга методов машинного обучения (например, система CrypTFlow).

Продолжением исследования будет анализ систем КМО на основе четырехсторонних протоколов безопасных вычислений.

СПИСОК ЛИТЕРАТУРЫ:

1. Evans D., Kolesnikov V., Rosulek M. A pragmatic introduction to secure multi-party computation. – 182 p. URL: https://securecomputation.org/docs/pragmaticmpc.pdf (дата обращения: 10.01.2022).
2. Запечников С.В., Щербаков А.Ю. Конфиденциальное машинное обучение на основе двусторонних протоколов безопасных вычислений. Безопасность информационных технологий, [S.I.], т. 28, № 4, 2021, с. 39–51. DOI: http://dx.doi.org/10.26583/bit.2021.4.03.
3. Wagh, S. SecureNN: Efficient and private neural network training. Cryptology ePrint Archive. 2018. – 24 p. URL: https://eprint.iacr.org/2018/442 (дата обращения: 10.01.2022).
4. Chandran N., Gupta D., Rastogi A., Sharma R., Tripathi S. EzPC: Programmable and Efficient Secure Two-Party Computation for Machine Learning. 2019 IEEE European Symposium on Security and Privacy (EuroS&P), Stockholm, Sweden. 2019, p. 496–511. DOI: http://dx.doi.org/10.1109/EuroSP.2019.00043.
5. Kumar E. et al. CrypTFlow: Secure TensorFlow Inference. arXiv preprint. 2020. – 18 p. URL: https://arxiv.org/pdf/1909.07814v2.pdf (дата обращения: 10.01.2022).
6. Rathee D. et al. CrypTFlow2: Practical 2-Party Secure Inference. arXiv preprint. 2020. – 18 p. URL: https://arxiv.org/pdf/2010.06457.pdf (дата обращения: 10.01.2022).
7. Patra A. ABY2.0: Improved mixed-protocol secure two-party computation. A. Patra, T. Schneider, A. Suresh et al. URL: https://ia.cr/2020/1225 (дата обращения: 10.01.2022).
8. Boemer F. MP2ML: a mixed-protocol machine learning framework for private inference. ARES '20: Proceedings of the 15th International Conference on Availability, Reliability and Security. 2020, p. 1–10. DOI: http://dx.doi.org/10.1145/3407023.3407045. URL: https://dl.acm.org/doi/abs/10.1145/3407023.3407045 (дата обращения: 10.01.2022)
9. Demmler D. ABY – a framework for efficient mixed-protocol secure two-party computation. D. Demmler, T. Schneider, M. Zohner. 22nd Network and Distributed System Security Symposium (NDSS'15), Internet Society, San Diego, CA, USA, February 8–11, 2015. URL: https://encrypto.de/papers/DSZ15.pdf (дата обращения: 10.01.2022).
10. Thien Duc Nguyen, Phillip Rieger, Hossein Yalame, Helen Möllering, Hossein Fereidooni, Samuel Marchal, Markus Miettinen, Azalia Mirhoseini, Ahmad-Reza Sadeghi, Thomas Schneider, and Shaza Zeitouni. FLGUARD: Secure and private federated learning, Jan 6, 2021. URL: https://ia.cr/2021/025 (дата обращения: 10.01.2022).
11. Mishra P. Delphi: A Cryptographic Inference Service for Neural Networks. P. Mishra, R. Lehmkuhl, A. Srinivasan et al. Proc. of USENIX Security 2020 (USENIX Security Symposium). URL: https://www.usenix.org/system/files/sec20spring_mishra_prepub.pdf (дата обращения: 10.01.2022).
12. Raluca Ada Popa homepage: Research. URL: https://people.eecs.berkeley.edu/~raluca/#Research (дата обращения: 10.01.2022).
13. Lehmkuhl R. Muse: Secure Inference Resilient to Malicious Clients. R. Lehmkuhl, P. Mishra, A. Srinivasan et al. Proc. of USENIX Security 2021 (USENIX Security Symposium). URL: https://people.eecs.berkeley.edu/~raluca/MUSEcamera.pdf (дата обращения: 10.01.2022).
14. Rachuri R. Trident: Efficient 4PC framework for privacy preserving machine learning. Cryptology ePrint Archive. 2019. – 26 p. URL: https://eprint.iacr.org/2019/1315 (дата обращения: 10.01.2022).
15. Byali M. FLASH: Fast and robust framework for privacy-preserving machine. Cryptology ePrint Archive. 2019. – 29 p. URL: https://eprint.iacr.org/2019/1365 (дата обращения: 10.01.2022).
16. Patra A. BLAZE: Blazing Fast Privacy-Preserving Machine Learning. Cryptology ePrint Archive. 2020. – 28 p. URL: https://eprint.iacr.org/2020/042.pdf (дата обращения: 10.01.2022).
17. Koti N. SWIFT: Super-fast and Robust Privacy-Preserving Machine Learning. Cryptology ePrint Archive. 2020. – 36 p. URL: https://eprint.iacr.org/2020/592.pdf (дата обращения: 10.01.2022).
18. Koti N. Tetrad: Actively Secure 4PC for Secure Training and Inference. Cryptology ePrint Archive. 2021. – 31 p. URL: https://eprint.iacr.org/2021/755.pdf (дата обращения: 10.01.2022).
19. Mohassel P. SecureML: A system for scalable privacy-preserving machine learning. Cryptology ePrint Archive. 2017. – 38 p. URL: https://eprint.iacr.org/2017/396 (дата обращения: 10.01.2022).
20. Mohassel P. ABY^3: A mixed protocol framework for machine learning. Cryptology ePrint Archive. 2018. – 40 p. URL: https://eprint.iacr.org/2018/403 (дата обращения: 10.01.2022).

21. Mohassel P. Practical privacy-preserving k-means clustering. Cryptology ePrint Archive. 2019. – 30 p. URL: https://eprint.iacr.org/2019/1158 (дата обращения: 10.01.2022).
22. Wagh S. Falcon: Honest-Majority Maliciously Secure Framework for Private Deep Learning. Proc. of Privacy Enhancing Technologies Symposium (PETS), June 2021, p. 1–21. URL: https://arxiv.org/pdf/2004.02229.pdf (дата обращения: 10.01.2022).
23. Sameer W. New directions in efficient privacy-preserving machine learning. Ph. D. Theses. Princeton university. 2020. – 203 p. URL: https://dataspace.princeton.edu/bitstream/88435/dsp01s7526g34f/1/Wagh_princeton_0181D_13320.pdf (дата обращения: 10.01.2022).
24. Attrapadung N. Adam in Private: Secure and Fast Training of Deep Neural Networks with Adaptive Moment Estimation. Cryptology ePrint Archive. 2021. – 24 p. URL: https://eprint.iacr.org/2021/736.pdf (дата обращения: 10.01.2022).
25. Juvekar C. GAZELLE: A Low Latency Framework for Secure Neural Network Inference. Cryptology ePrint Archive. 2021. – 17 p. URL: https://eprint.iacr.org/2018/073.pdf (дата обращения: 10.01.2022).
26. Liu J. Oblivious Neural Network Predictions via MiniONN transformations. Cryptology ePrint Archive. 2017. – 13 p. URL: https://eprint.iacr.org/2017/452.pdf (дата обращения: 10.01.2022).
27. Ryffel T. AriaNN: Low-Interaction Privacy-Preserving Deep Learning via Function Secret Sharing. Preprint. URL: https://arxiv.org/pdf/2006.04593.pdf (дата обращения: 10.01.2022).
28. Mohassel P. Fast and secure three-party computation: The garbled circuit approach. Cryptology ePrint Archive. 2015. – 18 p. URL: https://eprint.iacr.org/2015/931 (дата обращения: 10.01.2022).

REFERENCES:

[1] Evans D., Kolesnikov V., Rosulek M. A pragmatic introduction to secure multi-party computation. – 182 p. URL: https://securecomputation.org/docs/pragmaticmpc.pdf (accessed: 10.01.2022).
[2] Zapechnikov S., Shcherbakov A. Privcy-preserving machine learning based on secure two-party computations. IT Security (Russia), [S.I.], vol. 28, no. 4, 2021, p. 39–51. DOI: http://dx.doi.org/10.26583/bit.2021.4.03 (in Russian).
[3] Wagh, S. SecureNN: Efficient and private neural network training. Cryptology ePrint Archive. 2018. – 24 p. URL: https://eprint.iacr.org/2018/442 (accessed: 10.01.2022).
[4] Chandran N., Gupta D., Rastogi A., Sharma R., Tripathi S. EzPC: Programmable and Efficient Secure Two-Party Computation for Machine Learning. 2019 IEEE European Symposium on Security and Privacy (EuroS&P), Stockholm, Sweden. 2019, p. 496–511. DOI: http://dx.doi.org/10.1109/EuroSP.2019.00043.
[5] Kumar E. et al. CrypTFlow: Secure TensorFlow Inference. arXiv preprint. 2020. – 18 p. URL: https://arxiv.org/pdf/1909.07814v2.pdf (accessed: 10.01.2022).
[6] Rathee D. et al. CrypTFlow2: Practical 2-Party Secure Inference. arXiv preprint. 2020. – 18 p. URL: https://arxiv.org/pdf/2010.06457.pdf (accessed: 10.01.2022).
[7] Patra A. ABY2.0: Improved mixed-protocol secure two-party computation. A. Patra, T. Schneider, A. Suresh et al. URL: https://ia.cr/2020/1225 (accessed: 10.01.2022).
[8] Boemer F. MP2ML: a mixed-protocol machine learning framework for private inference. ARES '20: Proceedings of the 15th International Conference on Availability, Reliability and Security. 2020, p. 1–10. DOI: http://dx.doi.org/10.1145/3407023.3407045. URL: https://dl.acm.org/doi/abs/10.1145/3407023.3407045 (accessed: 10.01.2022)
[9] Demmler D. ABY – a framework for efficient mixed-protocol secure two-party computation. D. Demmler, T. Schneider, M. Zohner. 22nd Network and Distributed System Security Symposium (NDSS'15), Internet Society, San Diego, CA, USA, February 8–11, 2015. URL: https://encrypto.de/papers/DSZ15.pdf (accessed: 10.01.2022).
[10] Thien Duc Nguyen, Phillip Rieger, Hossein Yalame, Helen Möllering, Hossein Fereidooni, Samuel Marchal, Markus Miettinen, Azalia Mirhoseini, Ahmad-Reza Sadeghi, Thomas Schneider, and Shaza Zeitouni. FLGUARD: Secure and private federated learning, Jan 6, 2021. URL: https://ia.cr/2021/025 (accessed: 10.01.2022).
[11] Mishra P. Delphi: A Cryptographic Inference Service for Neural Networks. P. Mishra, R. Lehmkuhl, A. Srinivasan et al. Proc. of USENIX Security 2020 (USENIX Security Symposium). URL: https://www.usenix.org/system/files/sec20spring_mishra_prepub.pdf (accessed: 10.01.2022).
[12] Raluca Ada Popa homepage: Research. URL: https://people.eecs.berkeley.edu/~raluca/#Research (accessed: 10.01.2022).
[13] Lehmkuhl R. Muse: Secure Inference Resilient to Malicious Clients. R. Lehmkuhl, P. Mishra, A. Srinivasan et al. Proc. of USENIX Security 2021 (USENIX Security Symposium). URL: https://people.eecs.berkeley.edu/~raluca/MUSEcamera.pdf (accessed: 10.01.2022).

[14] Rachuri R. Trident: Efficient 4PC framework for privacy preserving machine learning. Cryptology ePrint Archive. 2019. – 26 p. URL: https://eprint.iacr.org/2019/1315 (accessed: 10.01.2022).
[15] Byali M. FLASH: Fast and robust framework for privacy-preserving machine learning. Cryptology ePrint Archive. 2019. – 29 p. URL: https://eprint.iacr.org/2019/1365 (accessed: 10.01.2022).
[16] Patra A. BLAZE: Blazing Fast Privacy-Preserving Machine Learning. Cryptology ePrint Archive. 2020. – 28 p. URL: https://eprint.iacr.org/2020/042.pdf (accessed: 10.01.2022).
[17] Koti N. SWIFT: Super-fast and Robust Privacy-Preserving Machine Learning. Cryptology ePrint Archive. 2020. – 36 p. URL: https://eprint.iacr.org/2020/592.pdf (accessed: 10.01.2022).
[18] Koti N. Tetrad: Actively Secure 4PC for Secure Training and Inference. Cryptology ePrint Archive. 2021. – 31 p. URL: https://eprint.iacr.org/2021/755.pdf (accessed: 10.01.2022).
[19] Mohassel P. SecureML: A system for scalable privacy-preserving machine learning. Cryptology ePrint Archive. 2017. – 38 p. URL: https://eprint.iacr.org/2017/396 (accessed: 10.01.2022).
[20] Mohasse P. ABY^3: A mixed protocol framework for machine learning. Cryptology ePrint Archive. 2018. – 40 p. URL: https://eprint.iacr.org/2018/403 (accessed: 10.01.2022).
[21] Mohassel P. Practical privacy-preserving k-means clustering. Cryptology ePrint Archive. 2019. – 30 p. URL: https://eprint.iacr.org/2019/1158 (accessed: 10.01.2022).
[22] Wagh S. Falcon: Honest-Majority Maliciously Secure Framework for Private Deep Learning. Proc. of Privacy Enhancing Technologies Symposium (PETS), June 2021, p. 1–21. URL: https://arxiv.org/pdf/2004.02229.pdf (accessed: 10.01.2022).
[23] Sameer W. New directions in efficient privacy-preserving machine learning. Ph. D. Theses. Princeton university. 2020. – 203 p. URL: https://dataspace.princeton.edu/bitstream/88435/dsp01s7526g34f/1/Wagh_princeton_0181D_13320.pdf (accessed: 10.01.2022).
[24] Attrapadung N. Adam in Private: Secure and Fast Training of Deep Neural Networks with Adaptive Moment Estimation. Cryptology ePrint Archive. 2021. – 24 p. URL: https://eprint.iacr.org/2021/736.pdf (accessed: 10.01.2022).
[25] Juvekar C. GAZELLE: A Low Latency Framework for Secure Neural Network Inference. Cryptology ePrint Archive. 2021. – 17 p. URL: https://eprint.iacr.org/2018/073.pdf (accessed: 10.01.2022).
[26] Liu J. Oblivious Neural Network Predictions via MiniONN transformations. Cryptology ePrint Archive. 2017. – 13 p. URL: https://eprint.iacr.org/2017/452.pdf (accessed: 10.01.2022).
[27] Ryffel T. AriaNN: Low-Interaction Privacy-Preserving Deep Learning via Function Secret Sharing. Preprint. URL: https://arxiv.org/pdf/2006.04593.pdf (accessed: 10.01.2022).
[28] Mohassel P. Fast and secure three-party computation: The garbled circuit approach. Cryptology ePrint Archive. 2015. – 18 p. URL: https://eprint.iacr.org/2015/931 (accessed: 10.01.2022).

Поступила в редакцию – 11 января 2022 г. Окончательный вариант – 27 января 2022 г.
Received – January 11, 2022. The final version – January 27, 2022.