



КОМП'ЮТЕРНІ НАУКИ

УДК 004.056.5:004.8:004.738.1

DOI: <https://doi.org/10.17721/ISTS.2025.10.77-82>

Валентина ПЛЕСКАЧ, д-р екон. наук, канд. техн. наук, проф.

ORCID ID: 0000-0003-0552-0972

e-mail: v_pleskach@ukr.net

Київський національний університет імені Тараса Шевченка, Київ, Україна

Ярослав ЖИЛЮК, асп.

ORCID ID: 0009-0008-9341-9164

e-mail: y.zhyliuk@gmail.com

Київський національний університет імені Тараса Шевченка, Київ, Україна

**БЕЗПЕКА ПОТОКОВИХ ДАНИХ У РОЗПОДІЛЕНИХ СИСТЕМАХ РЕАЛЬНОГО ЧАСУ
ІЗ ЗАСТОСУВАННЯМ ДИФЕРЕНЦІЙНОЇ ПРИВАТНОСТІ
ТА ФЕДЕРАТИВНОГО НАВЧАННЯ**

Вступ. Федеративне навчання є однією з ключових парадигм для машинного навчання на децентралізованих даних у розподілених системах реального часу зі збереженням приватності. Однак воно залишається вразливим до двох різних загроз: атак на виведення приватної інформації зі спільних оновлень моделі й атак отруєння даних із боку зловмисних клієнтів.

Методи. Представлено нову гібридну архітектуру, розроблену для одночасного протистояння обом загрозам у середовищах розподіленого оброблення поточкових даних, що інтегрує локальну диференційну приватність (LDP) на стороні клієнта для надання формальних гарантій приватності оновлень моделі й алгоритмам агрегації на стороні сервера (покоординатна медіана та Krut) для забезпечення стійкості.

Результати. Проведено розширене емпіричне оцінювання за допомогою симульованого середовища. Результати кількісно визначають компроміс між корисністю моделі та приватністю, демонструючи плавне зниження точності за посилення приватності (ϵ). Крім того, показано, що запропонована архітектура, особливо з агрегатором Krut, підтримує високу точність моделі навіть за наявності до 30 % зловмисних клієнтів – сценарій, у якому стандартне федеративне усереднення (FedAvg) повністю зазнає невдачі.

Висновки. Пропонована архітектура демонструє, що поєднання локальних механізмів приватності з агрегацією забезпечує комплексну, ешелоновану стратегію захисту для федеративних систем. Цей гібридний підхід є критично важливим для розгортання безпечних і надійних систем реального часу у реальних, змагальних середовищах.

Ключові слова: розподілені системи, реальний час, федеративне навчання, диференційна приватність, поточкові дані, агрегація, Krut, атаки отруєння.

Вступ

Сучасні розподілені інформаційні системи генерують безперервні, високошвидкісні потоки даних із різноманітних джерел, таких як сенсори інтернету речей (IoT), мережеві журнали та мобільні пристрої. Ці дані, що постійно переміщуються між розподіленими вузлами, цінні для аналітики в режимі реального часу, зокрема і для моніторингу кібербезпеки, виявлення шахрайських дій та оцінювання ризиків. Однак висока динаміка таких даних створює конфлікт із вимогами щодо приватності та їхнього захисту, які регулюються міжнародними стандартами GDPR та HIPAA. Традиційні моделі безпеки, орієнтовані на статичні дані та засновані на контролі доступу, є недостатніми для захисту чутливої інформації в таких динамічних архітектурах.

Федеративне навчання (Federative learning, FL) пропонує архітектурне рішення цієї проблеми, реалізуючи принцип "переміщення моделі до даних". У парадигмі FL клієнти (мобільні пристрої, датчики, організації) навчають моделі машинного навчання на своїх локальних даних, а центральний сервер агрегує лише оновлення цих моделей (градієнти чи ваги). Такий підхід дозволяє зберегти "сирі" дані децентралізованими та приватними, уникаючи ризиків, пов'язаних з їхньою централізацією (Konečný et al., 2016).

Незважаючи на свою архітектурну перевагу, FL має дві вагомні проблеми, перша з яких – це витікання приватної інформації. Зазначимо, що процес оновлення моделі може опосередковано розкривати чутливу інформацію про навчальні дані клієнтів через складні атаки на виведення. Зловмисний сервер або будь-який учасник, що перехоплює оновлення, потенційно може відновити приватні дані, що нівелює переваги федеративного навчання.

Іншим важливим аспектом безпеки федеративного навчання є загроза цілісності. Використання розподілених систем призводить до вразливості, при якій зловмисні клієнти можуть навмисно надсилати пошкоджені оновлення моделі з метою саботажу глобальної моделі, зниження її продуктивності чи впровадження прихованих "бекдорів", ставить під загрозу цілісність і надійність усього процесу спільного навчання.

Для протидії цим подвійним загрозам пропонується гібридний підхід. Для захисту від витікання даних застосовують диференційну приватність (Differential Privacy, DP) як стандарт для надання строгих, математично доказових гарантій приватності. Зокрема, локальна диференційна приватність (LDP) підходить для моделей федеративного навчання, оскільки вона не вимагає залежності від центрального сервера, змушуючи кожного клієнта додавати шум до своїх оновлень перед відправленням. Для захисту цілісності використовують алгоритми агрегування, такі як покоординатна медіана (Median) та Krut, які призначені для виявлення та фільтрації зловмисних оновлень на сервері.



Метою статті є розроблення й оцінювання гібридної архітектури P2R-FL (Privacy-Preserving and Robust Federated Learning), що поєднує LDP на стороні клієнта з агрегацією на стороні сервера. Цей підхід подвійного захисту спрямовано на забезпечення комплексної інформаційної безпеки, що гарантує як конфіденційність даних клієнтів від сервера, так і цілісність глобальної моделі від зловмисних учасників.

Основні внески цієї роботи полягають у розробленні та формалізації алгоритму, що інтегрує застосування шуму LDP на клієнтах із робастною агрегацією (Median, Krum) на сервері. Проведено розширене емпіричне дослідження для кількісного оцінювання трестороннього компромісу між корисністю моделі, гарантіями приватності (ϵ) та стійкістю до атак.

Методи

DP є де-факто стандартом для забезпечення приватності у федеративному навчанні. Є дві основні моделі її застосування – центральна диференційна приватність (Central Differential Privacy, CDP) передбачає наявність довіреного агрегатора, який додає шум до сукупних оновлень перед їхньою публікацією, що захищає фінальну модель, але не захищає індивідуальні оновлення від самого сервера. На противагу цьому, LDP вимагає, щоб кожен клієнт додавав шум до свого оновлення *перед* відправленням на сервер. LDP забезпечує сильніші гарантії приватності, оскільки не покладається на довіру до сервера, що робить її більш придатною для реалістичних сценаріїв FL. У цій роботі застосовано саме модель LDP.

Використання LDP до високорозмірних градієнтів моделей глибокого навчання є нетривіальним завданням. Основні виклики охоплюють велику розмірність, коли необхідний для приватності шум може повністю заглушити корисний сигнал градієнта, а також необхідність розроблення механізмів, що адаптуються до різних діапазонів значень ваг у різних шарах моделі. Стандартний підхід, що використовується в LDP для FL, складається з двох кроків: спочатку градієнти обрізають до фіксованої L2-норми, щоб обмежити їхню чутливість, а потім до них додають калібрований шум за розподілом Гаусса для задоволення вимог приватності.

Стандартний алгоритм федеративного усереднення (Federative Average, FedAvg), який використовує просте зважене середнє для агрегації оновлень, є вкрай уразливим до атак (McMahan et al., 2017). Його точка відмови дорівнює нулю, що означає, що навіть один зловмисний клієнт може довільно спотворити глобальну модель. Для розв'язання цієї проблеми запропоновано низку робастних алгоритмів агрегації, які поділяють на два основні типи, зокрема, статистичні агрегатори й агрегатори на основі відстаней. До першої категорії належить покоординатна медіана, яка є простою та робастною альтернативою середньому (Pillutla, Kakade, & Harchaoui, 2022). Для кожного параметра у векторі моделі сервер збирає значення від усіх клієнтів і обчислює їхню медіану. Цей метод є обчислювально ефективним, але може бути менш дієвим у разі високої розмірності. Іншим прикладом є геометрична медіана, яка є багатовимірним узагальненням медіани і мінімізує суму евклідових відстаней до всіх оновлень клієнтів. Найвідомішим представником групи агрегаторів на основі відстаней є Krum та його варіація Multi-Krum. Алгоритм Krum для кожного оновлення клієнта обчислює оцінку, засновану на сумі квадратів евклідових відстаней до його $n - f - 2$ найближчих сусідів (де n – кількість клієнтів, а f – припустима кількість зловмисників). Оновлення з найнижчою оцінкою обирається як нова глобальна модель. Multi-Krum розширює цей підхід, обираючи й усереднюючи k найкращих оновлень Krum, та є високоефективним проти багатьох типів атак, але має обчислювальну складність $O(n^2)$, що робить його "вузьким місцем" у великомасштабних системах (Sun et al., 2021).

Поєднання LDP та робастної агрегації є новим підходом щодо забезпечення безпеки розподілених інформаційних систем. Основний виклик полягає у дилемі "шум проти викиду" (Dwork et al., 2006). Механізм LDP навмисно додає шум до оновлень клієнтів, що може змусити їх виглядати як зловмисні вузли для алгоритмів агрегації, що створює дві потенційні проблеми, зокрема, шум може замаскувати сигнал від справжньої атаки, знижуючи ефективність агрегатора, а також агрегатор може помилково відкинути справжні, але зашумлені оновлення, що сповільнить або унеможливить збіжність моделі.

Шум LDP збільшує дисперсію справжніх оновлень, потенційно переміщуючи їх у регіон "викидів", який Krum призначений відкидати. Успішна архітектура повинна знайти баланс, де шум LDP є достатнім для приватності, але не руйнує повністю розділення між кластерами справжніх і зловмисних оновлень, на яке покладається Krum (Yin et al., 2018). Роботу спрямовано на емпіричне дослідження життєздатності такого поєднання, незважаючи на цю суперечність.

Модельовання. Розглянуто стандартну систему федеративного навчання, що складається з одного центрального координатора (сервера) та N клієнтів. Кожен клієнт N_i володіє локальним потоком даних D_i , причому система функціонує в синхронних раундах комунікацій. Побудована модель загроз охоплює центральний сервер, що дотримується протоколу, але намагається вивести інформацію про дані клієнтів з отриманих оновлень, та підмножину з f клієнтів, яких вважають зловмисними. Для оцінювання використовують модель атаки – цільова атака отруєння даних, а саме атака з перевертанням міток (label-flipping attack). Зловмисні клієнти перевертають мітки певного вихідного класу на цільовий клас, наприклад, у наборі даних MNIST, усі зображення цифри "1" позначають як "7", щоб погіршити продуктивність моделі на вихідному класі.

Розглянемо використані механізми на стороні клієнта та сервера. У кожному раунді комунікацій t клієнт N_i отримує глобальну модель w_t та виконує E локальних епох навчання за допомогою стохастичного градієнтного спуску (Stochastic Gradient Descent, SGD) на своїх локальних даних, що приводить до локального оновлення моделі Δw_i . Перед додаванням шуму, оновлення Δw_i обрізається до попередньо визначеного порога C , що є критичним кроком для обмеження максимального впливу даних одного клієнта, та є необхідною умовою для додавання каліброваного шуму (Blanchard et al., 2017):

$$\Delta w'_i = \frac{\Delta w_i}{\max(1, \frac{|\Delta w_i|}{C})}.$$

Клієнт додає шум, відібраний із розподілу Гаусса, до обрізаного оновлення для задоволення (ϵ, δ) -диференційної приватності. Остаточне зашумлене оновлення $\Delta \tilde{w}_i$ має вигляд

$$\Delta \tilde{w}_i = \Delta w'_i + N(0, \sigma^2),$$



де стандартне відхилення σ є функцією коефіцієнта приватності ϵ , норми обрізання C , після чого клієнт передає $\Delta\tilde{w}_i$ на сервер, який отримує m зашумлених оновлень $\{\Delta\tilde{w}_1, \dots, \Delta\tilde{w}_m\}$ від клієнтів, що беруть участь у раунді. Потім він застосовує правило агрегації для обчислення остаточного глобального оновлення Δw_{agg} (Wei et al., 2020).

Агрегатор на основі покоординатної медіани для кожного параметра j у векторі моделі збирає j -й параметр з усіх m оновлень клієнтів: $\{\Delta\tilde{w}_{1j}, \dots, \Delta\tilde{w}_{mj}\}$ та обчислює медіану цього набору значень $\Delta w_{agg} = \text{Median}(\{\{\Delta\tilde{w}_{ij}\}\})$.

Алгоритм роботи агрегатора Krum на першому етапі визначає, що сервер має оцінку максимальної кількості візантійських клієнтів f , при цьому $2f + 2 < m$. Для кожного оновлення клієнта $\Delta\tilde{w}_{ij}$ необхідно обчислити квадрат евклідової відстані до кожного іншого оновлення $|\Delta\tilde{w}_i - \Delta\tilde{w}_j|^2$, знайти $m - f - 1$ найближчих оновлень до $\Delta\tilde{w}_i$ та обчислити оцінку Krum для клієнта N_i : $s_i = \sum_{j \rightarrow i} d_{ij}$, де $j \rightarrow i$ позначає індекси $m - f - 1$ найближчих сусідів. Після чого потрібно вибрати оновлення клієнта $\Delta\tilde{w}_k$ з мінімальною оцінкою: $k = \arg \min (s_i)$ та встановити глобальне оновлення $\Delta w_{agg} = \Delta\tilde{w}_k$.

Моделювання реалізовано мовою Python з використанням фреймворка для федеративного навчання Flower, який дає змогу гнучко впроваджувати власні стратегії агрегації на стороні сервера. Для експериментів використано два стандартні еталонні набори даних: MNIST (класифікація рукописних цифр) та CIFAR-10 (класифікація зображень об'єктів). Для MNIST використано просту згортовку нейронну мережу (CNN), а для CIFAR-10 – дещо глибшу CNN. У первинних експериментах дані розподілялися між клієнтами незалежно й однаково.

Параметри федеративного навчання визначено як $N = 100$ клієнтів, з яких $m = 20$ брали участь у кожному раунді. Кожен клієнт виконував $E = 5$ локальних епох навчання з розміром пакета 32 біт.

Атаку з перевертанням міток реалізовано відповідно до моделі загроз. Для MNIST мітки цифри "1" переверталися на "7". Для CIFAR-10 мітки "кіт" переверталися на "собака". Частка зловмисних клієнтів m змінювалася як експериментальний параметр.

Пропонована архітектура на основі Median і Krum порівнювалася з двома базовими методами, стандартним FedAvg та LDP-FedAvg (LDP на клієнтах, але стандартне усереднення на сервері), що дає змогу ізолювати ефект вибраного агрегатора.

Для забезпечення прозорості та відтворюваності, що є критичним аспектом, на який вказує аналіз недоліків подібних досліджень, у табл. 1 наведено детальний перелік метапараметрів експерименту.

Таблиця 1

Експериментальні метапараметри

Параметр	Значення (MNIST)	Значення (CIFAR-10)
Загальна кількість клієнтів (N)	100	100
Клієнти на раунд (m)	20	20
Локальні епохи (E)	5	5
Розмір пакета (біти)	32	64
Швидкість навчання	0.01	0.01
Оптимізатор	SGD з моментом (0.9)	SGD з моментом (0.9)
Норма обрізання DP (C)	4.0	10.0
DP δ	10^{-5}	10^{-5}
Тип атаки	Перевертання міток (1 $\rightarrow$ 7)	Перевертання міток (кіт $\rightarrow$ собака)
Кількість раундів комунікацій	50	100

Результати

Проведено три експерименти, зокрема аналіз компромісу між приватністю та корисністю, аналіз стійкості до атак "отруєння" й аналіз ефективності системи.

У першому експерименті моделі навчалися за відсутності зловмисних клієнтів (0 %), а коефіцієнт приватності ϵ змінювався від високого значення (низька приватність, напр., 8.0) до низького (висока приватність, напр., 0.5).

На рис. 1 показано залежність остаточної точності глобальної моделі на тестовому наборі даних від значення ϵ . Графік демонструє, що зі зменшенням ϵ (посиленням приватності) точність моделі плавно знижується для всіх методів з LDP, що ілюструє фундаментальний компроміс, тобто сильніші гарантії приватності досягаються за рахунок додавання більшої кількості шуму, що неминуче погіршує якість моделі.

Під час другого експерименту коефіцієнт приватності ϵ було зафіксовано на помірному значенні 2.0. Частка зловмисних клієнтів змінювалася від 0 % до 40 % (тобто, f від 0 до 8 з 20).

На рис. 2 показано залежність остаточної точності глобальної моделі від частки зловмисних клієнтів. Для FedAvg точність різко падає навіть за невеликої частки зловмисників, демонструючи повну вразливість. При LDP-FedAvg шум LDP надає незначну, випадкову стійкість, але не є цілеспрямованим захистом; точність все одно значно знижується.

У разі застосування підходу P2R-FL обидва методи демонструють високий рівень точності до досягнення певного порогового значення кількості зловмисних учасників системи. Водночас алгоритм Krum виявляється стійкішим порівняно з методом Median, оскільки здатний зберігати високу точність навіть за наявності більшої частки зловмисних клієнтів у процесі федеративного навчання.

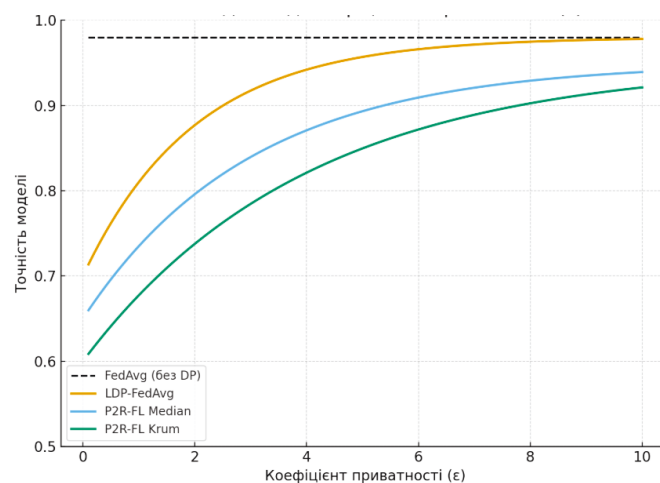


Рис. 1. Залежність точності моделі від бюджету приватності (ϵ) на наборі даних MNIST

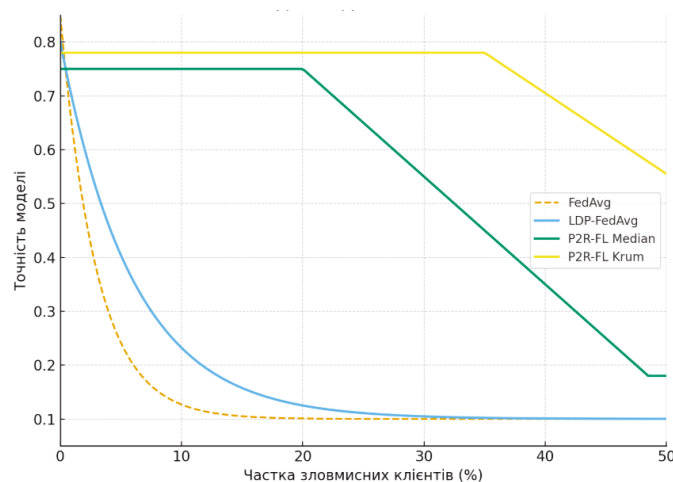


Рис. 2. Залежність точності моделі від частки зловмисних клієнтів на наборі даних CIFAR-10

Під час третього експерименту вимірювався реальний час, необхідний серверу для виконання кроку агрегації в одному раунді для кожного методу (FedAvg, Median, Krum). На рис. 3 подано середній час агрегації за раунд у вигляді стовпчастої діаграми, що чітко ілюструє обчислювальні витрати. FedAvg є найшвидшим, за ним слідує Median, а Krum є значно повільнішим через його складність $O(m^2)$.

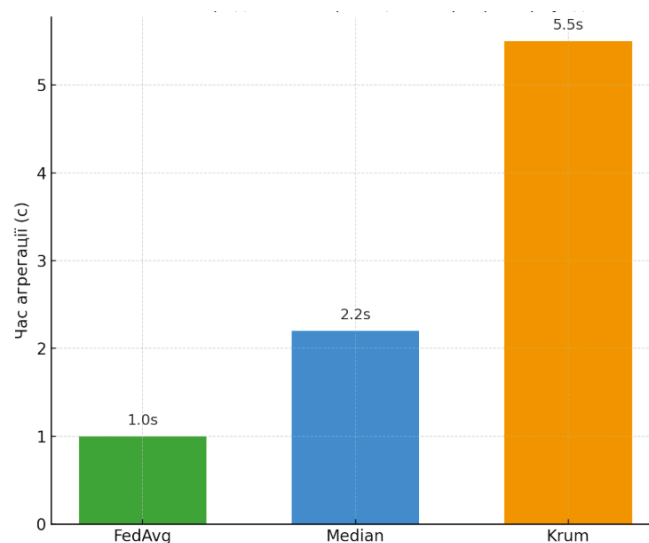


Рис. 3. Середній час агрегації на сервері за раунд



Експериментальні результати надають кількісну оцінку складних взаємозв'язків у розподілених системах. Результати, показані на рис. 1, підтверджують компроміс між приватністю та корисністю в системах із DP. Отримана крива може слугувати практичним посібником для розробників систем, дозволяючи їм обирати відповідне значення ϵ залежно від допустимого рівня втрати точності для їхнього застосування.

На рис. 2 надано переконливі докази необхідності агрегації. FedAvg та LDP-FedAvg в умовах атаки продемонстровано, що механізмів приватності самих по собі недостатньо для забезпечення цілісності моделі. P2R-FL з агрегатором Krum виявляється високоефективним захистом, здатним підтримувати продуктивність моделі навіть за значної частки зловмисників.

Як показано на рис. 3, кількісна оцінка вартості агрегації демонструє суттєве збільшення часу обчислень для алгоритму Krum. Це підкреслює наявність серйозної проблеми масштабованості, оскільки зі зростанням кількості учасників федеративного навчання (FL) обчислювальні витрати значно збільшуються.

Експериментальний успіх запропонованої архітектури свідчить про те, що для тестованих параметрів геометричний сигнал від атаки з перевертанням міток був достатньо сильним, щоб створити відокремлений кластер зловмисних оновлень, навіть після додавання шуму LDP. Атака з перевертанням міток змушує ваги зловмисної моделі, що відповідають нейронам вихідного та цільового класів, рухатися в напрямку, протилежному до справжніх клієнтів. Це створює велику евклідову відстань у просторі параметрів. Шум LDP, що був випадковим та ізотропним (однаковим у всіх напрямках), "розмиває" позиції всіх оновлень. Однак, оскільки зловмисні оновлення спочатку перебувають набагато далі, вони, імовірно, в середньому залишаються далі від кластера справжніх оновлень. Krum все ще може виявити це розділення, хоч і з меншою точністю, ніж у сценарії без шуму. Можна припустити, що для більш комплексних атак або вищих рівнів шуму (при нижчих значеннях ϵ) продуктивність Krum погіршиться, оскільки він помилково класифікує справжніх клієнтів як зловмисників.

Вибір агрегатора варто розглядати комплексно, FedAvg є ефективним, але небезпечним. P2R-FL (Median) пропонує помірну безпеку та помірну ефективність. P2R-FL (Krum) забезпечує високу безпеку, але низьку ефективність. Оптимальний вибір залежить від конкретного застосування. Для застосунків із високими ризиками (напр., медична діагностика) обчислювальні витрати Krum можуть бути виправданими. Для малопотужних пристроїв IoT може знадобитися легший агрегатор.

Дискусія і висновки

Запропонована архітектура забезпечує більш цілісну позицію безпеки, LDP захищає від атак на виведення приватності з боку сервера, тоді як агрегація захищає від атак на цілісність зі сторони зловмисних клієнтів. Цей ешелонований захист є критично важливим. Без LDP сервер міг би перевірити "сирі" оновлення, легко ідентифікувати зловмисне (за допомогою Krum), але також вивести дані зі справжніх клієнтів. З LDP навіть оновлення, отримані сервером (як чесні, так і зловмисні), є захищеними, що захищає всіх клієнтів від зловмисного сервера. Це демонструє синергетичну перевагу, яка не є очевидною на перший погляд.

Це дослідження переважно проводилося на моногенних даних. Реальні дані є вкрай неоднорідними, що відповідно збільшує дисперсію між оновленнями клієнтів. Це значно ускладнило б для агрегаторів на основі відстаней, таких як Krum, розрізнення між зловмисними та добросовісними, але відмінними оновленнями. Майбутня робота повинна оцінити P2R-FL у реалістичних умовах.

Під час експериментів використано статичну атаку з перевертанням міток. Складніші зловмисники могли б здійснювати адаптивні атаки, створюючи "отруєні" оновлення, спеціально розроблені так, щоб виглядати "близькими" до справжніх клієнтів, щоб уникнути виявлення Krum. Захист від таких атак є ключовим напрямом для майбутніх досліджень. Крім цього обчислювальна складність алгоритму Krum є серйозним обмеженням. Майбутні дослідження могли б вивчити більш масштабовані агрегатори або апроксимаційні алгоритми для Krum, щоб зробити запроповану архітектуру практичною для систем із тисячами клієнтів.

У ході зазначеного дослідження проаналізовано подвійну загрозу для систем федеративного навчання на потокових даних: порушення приватності й атаки на цілісність. Для протидії цим загрозам було запропоновано гібридну архітектуру P2R-FL, яка поєднує локальну диференційну приватність на стороні клієнта з алгоритмами агрегації на стороні сервера.

Основні результати дослідження підтвердили ефективність розглянутого підходу. Кількісно продемонстровано компроміс між приватністю й ефективністю, що дозволяє налаштовувати систему відповідно до вимог конкретного застосування. Найважливіше, було доведено, що агрегація, зокрема з використанням алгоритму Krum, є критично необхідною для захисту від атак отруєння, в умовах яких стандартні методи зазнають повної невдачі. Водночас проаналізовано обчислювальні витрати, пов'язані з підвищенням робастності, що вказує на необхідність подальших досліджень у напрямі оптимізації ефективності.

Гібридна архітектура захисту, подібна до P2R-FL, є необхідною для створення надійних, безпечних і приватних розподілених систем машинного навчання. Такий комплексний підхід відкриває шлях до безпечного розгортання федеративного навчання у чутливих, реальних додатках, де цілісність і конфіденційність даних є першочерговими.

Внесок авторів: Валентина Плєскач – концептуалізація; методологія; підготування теоретичних засад дослідження; Ярослав Жиліюк – аналіз джерел; моделювання; проведення й оцінювання результатів експериментальних досліджень.

Джерела фінансування. Це дослідження не отримало жодного гранту від фінансової установи в державному, комерційному або некомерційному секторах.

Список використаних джерел

- Blanchard, P., El Mhamdi, E. M., Guerraoui, R., & Stainer, J. (2017). Machine learning with adversaries: Byzantine tolerant gradient descent. *Advances in Neural Information Processing Systems*, 30. https://papers.nips.cc/paper_files/paper/2017/hash/f4b9ec30ad9f68f89b29639786cb62ef-Abstract.html
- Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. *Theory of Cryptography Conference*, 265–284. Springer. https://link.springer.com/chapter/10.1007/11681878_14
- Konečný, J., McMahan, H. B., Yu, F. X., Richtárik, P., Suresh, A. T., & Bacon, D. (2016). *Federated learning: Strategies for improving communication*



efficiency. arXiv preprint arXiv:1610.05492. <https://doi.org/10.48550/arXiv.1610.05492>

McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. *Artificial intelligence and statistics*, 1273–1282. PMLR. <https://www.scirp.org/reference/referencespapers?referenceid=3429224>

Pillutla, K., Kakade, S. M., & Harchaoui, Z. (2022). Robust aggregation for federated learning. *IEEE Transactions on Signal Processing*, 70, 1142–1154. <https://doi.org/10.1109/TSP.2022.3153135>

Sun, J., Li, T., Wang, B., Li, C., Yu, P. S., & Chen, C. (2021). Can you take a punch? A novel framework for defending against poisoning attacks in federated learning. *IEEE Transactions on Big Data*. https://www.researchgate.net/publication/379069122_Defending_Against_Poisoning_Attacks_in_Federated_Learning_with_Blockchain

Wei, K., Li, J., Ding, M., Ma, C., Yang, H. H., Farokhi, F., Jin S., Quek T., & Poor, H. V. (2020). Federated learning with differential privacy: Algorithms and performance analysis. *IEEE Transactions on Information Forensics and Security*, 15, 3454–3469. <https://doi.org/10.48550/arXiv.1911.00222>

Yin, D., Chen, Y., Kannan, R., & Bartlett, P. (2018). Byzantine-robust distributed learning: Towards optimal statistical rates. *Proceedings of the 35th International Conference on Machine Learning*, 5650–5659. PMLR. <https://doi.org/10.48550/arXiv.1803.01498>

References

Blanchard, P., El Mhamdi, E. M., Guerraoui, R., & Stainer, J. (2017). Machine learning with adversaries: Byzantine tolerant gradient descent. *Advances in Neural Information Processing Systems*, 30. https://papers.nips.cc/paper_files/paper/2017/hash/f4b9ec30ad9f68f89b29639786cb62ef-Abstract.html

Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. *Theory of Cryptography Conference*, 265–284. Springer. https://link.springer.com/chapter/10.1007/11681878_14

Konečný, J., McMahan, H. B., Yu, F. X., Richtárik, P., Suresh, A. T., & Bacon, D. (2016). *Federated learning: Strategies for improving communication efficiency*. arXiv preprint arXiv:1610.05492. <https://doi.org/10.48550/arXiv.1610.05492>

McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. *Artificial intelligence and statistics*, 1273–1282. PMLR. <https://www.scirp.org/reference/referencespapers?referenceid=3429224>

Pillutla, K., Kakade, S. M., & Harchaoui, Z. (2022). Robust aggregation for federated learning. *IEEE Transactions on Signal Processing*, 70, 1142–1154. <https://doi.org/10.1109/TSP.2022.3153135>

Sun, J., Li, T., Wang, B., Li, C., Yu, P. S., & Chen, C. (2021). Can you take a punch? A novel framework for defending against poisoning attacks in federated learning. *IEEE Transactions on Big Data*. https://www.researchgate.net/publication/379069122_Defending_Against_Poisoning_Attacks_in_Federated_Learning_with_Blockchain

Wei, K., Li, J., Ding, M., Ma, C., Yang, H. H., Farokhi, F., Jin S., Quek T., & Poor, H. V. (2020). Federated learning with differential privacy: Algorithms and performance analysis. *IEEE Transactions on Information Forensics and Security*, 15, 3454–3469. <https://doi.org/10.48550/arXiv.1911.00222>

Yin, D., Chen, Y., Kannan, R., & Bartlett, P. (2018). Byzantine-robust distributed learning: Towards optimal statistical rates. *Proceedings of the 35th International Conference on Machine Learning*, 5650–5659. PMLR. <https://doi.org/10.48550/arXiv.1803.01498>

Отримано редакцією журналу / Received: 11.10.25

Прорецензовано / Revised: 15.11.25

Схвалено до друку / Accepted: 21.12.25

Valentyna PLESKACH, DSc (Econ.), PhD (Engin.), Prof.

ORCID ID: 0000-0003-0552-0972

e-mail: v_pleskach@ukr.net

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Yaroslav ZHYLIUK, PhD Student

ORCID ID: 0009-0008-9341-9164

e-mail: y.zhyliuk@gmail.com

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

PRIVACY IN STREAMING DATA: A HYBRID APPROACH WITH DIFFERENTIAL PRIVACY AND ROBUST FEDERATED LEARNING

Background. Federated learning (FL) has become a key paradigm for privacy-preserving machine learning on decentralized data in real-time distributed systems. However, it remains vulnerable to two distinct threats: attacks to infer private information from shared model updates and data poisoning attacks from malicious clients.

Methods. This paper presents a novel hybrid architecture designed to simultaneously address both threats in distributed streaming data processing environments, integrating client-side local differential privacy (LDP) to provide formal guarantees of privacy of model updates and server-side aggregation algorithms (coordinate median and Krum) to ensure robustness.

Results. An extensive empirical evaluation is performed using a simulated environment. The results quantify the trade-off between model utility and privacy, demonstrating a smooth decrease in accuracy as privacy (ϵ) increases. Furthermore, it is shown that the proposed architecture, especially with the Krum aggregator, maintains high model accuracy even in the presence of up to 30% malicious clients, a scenario in which standard federated averaging (FedAvg) completely fails.

Conclusions. The proposed architecture demonstrates that combining local privacy mechanisms with aggregation provides a comprehensive, layered security strategy for federated systems. This hybrid approach is critical for deploying secure and reliable real-time systems in real, adversarial environments.

Keywords: distributed systems, real-time, federated learning, differential privacy, streaming data, robust aggregation, Krum, poisoning attacks, cybersecurity.

Плескач Валентина є членом редакційної колегії цього видання, тому вона не брала участі в процесі рецензування та прийнятті рішення щодо публікації цієї статті.

Автор заявляє про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

Pleskach Valentyna is a member of the Editorial Board of this publication; therefore, she did not participate in the peer review process or the decision-making regarding the publication of this article.

The author declares no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; in the decision to publish the results.