

Удосконалення методу багатофакторної аутентифікації шляхом додавання контекстних факторів поведінкової біометрії на основі мережі DP-GAN

Вінницький національний технічний університет

Анотація. Робота присвячена дослідженню та розробці методу формування контекстних поведінкових біометричних факторів з використанням диференційно-приватної генеративної мережі DP-GAN. Запропонований підхід спрямований на підвищення рівня захищеності сучасних систем шляхом створення синтетичних біометричних даних, які зберігають статистичні властивості реальних зразків, але не містять персональної інформації. У роботі проведено аналіз існуючих методів генерації та захисту поведінкових ознак, розглянуто принципи диференційної приватності та архітектуру GAN-мереж.

Розроблено програмний засіб, що реалізує DP-GAN для генерації синтетичних поведінкових шаблонів з контролем рівня приватності (ϵ , δ) та оцінюванням якості за допомогою FID-метрики. Практична частина включає створення моделі генератора та дискримінатора, реалізацію DP-SGD-механізму за допомогою бібліотеки Opacus, а також формування набору зразків для експериментального аналізу. Отримані результати підтверджують можливість використання DP-GAN як інструменту для генерації безпечних біометричних даних, придатних для досліджень, тестування та потенційної інтеграції у системи багатофакторної аутентифікації.

Розробка має практичне значення у сферах інформаційної безпеки, захисту персональних даних та проектування систем, що працюють із чутливою біометричною інформацією.

Ключові слова: DP-GAN, диференційна приватність, поведінкова біометрія, генеративні моделі, синтетичні дані, інформаційна безпека.

Abstract. The work thesis is devoted to the development and investigation of a method for generating contextual behavioral biometric factors using a differentially private generative adversarial network (DP-GAN). The proposed approach aims to enhance the security of modern systems by producing synthetic biometric data that preserve the statistical characteristics of real samples while containing no personal or identifiable information. The study analyzes existing methods for behavioral data modeling, explores the principles of differential privacy, and examines the architecture of GAN-based generative models.

A software tool was developed that implements a DP-GAN for generating synthetic behavioral patterns with controllable privacy parameters (ϵ , δ) and quality evaluation using the FID metric. The practical implementation includes the construction of generator and discriminator models, the integration of the DP-SGD mechanism via the Opacus library, and the creation of a dataset of synthetic samples for experimental analysis. The obtained results confirm the feasibility of using DP-GAN as a tool for generating privacy-preserving biometric data suitable for research, testing, and potential integration into multi-factor authentication systems.

The development has practical value in the fields of cybersecurity, personal data protection, and the design of systems that operate with sensitive biometric information.

Keywords: DP-GAN, differential privacy, behavioral biometrics, generative models, synthetic data, cybersecurity.

Вступ

Розвиток сучасних інформаційних систем потребує нових підходів до обробки та захисту даних, зокрема поведінкових біометричних ознак. Через зростання вимог до приватності та відсутність достатніх реальних біометричних наборів актуальною стає задача генерації синтетичних даних, які зберігають статистичні властивості, але не розкривають інформацію про конкретних користувачів. Одним із перспективних рішень є використання диференційно-приватних генеративних моделей DP-GAN.

Метою роботи є розробка методу генерації синтетичних поведінкових біометричних факторів з використанням DP-GAN для демонстрації можливості їх застосування у дослідницьких та безпекових системах.

Об'єкт дослідження — процес формування поведінкових біометричних даних.

Предмет дослідження — методи генерації синтетичних біометричних ознак із забезпеченням диференційної приватності.

Наукова новизна полягає у поєднанні генеративної моделі GAN з механізмом DP-SGD для отримання безпечних синтетичних біометричних зразків.

Результати дослідження

1. Аналіз методів генерації біометричних даних

На початковому етапі було проведено аналіз існуючих методів генерації та моделювання поведінкових біометричних ознак. Було встановлено, що класичні генеративні моделі GAN здатні відтворювати окремі елементи реальних даних, що створює ризик витоку персональної інформації. Це особливо критично для біометричних даних, які належать до категорії чутливої інформації. Тому в роботі обґрунтовано необхідність застосування механізму диференційної приватності, який не дозволяє моделі запам'ятовувати конкретні приклади. Для цього досліджено алгоритм DP-SGD, що додає шум у градієнти під час навчання та обмежує їхню норму. Поєднання GAN та DP-SGD було визначено як оптимальний підхід для безпечної генерації синтетичних біометричних зразків.

2. Розробка моделі DP-GAN

На основі отриманих теоретичних результатів створено модель DP-GAN, яка складається з двох нейронних мереж:

- Генератора, що створює синтетичні дані на основі випадкового латентного вектора;
- Дискримінатора, який навчається відрізняти реальні дані від згенерованих.

До дискримінатора застосовано механізм DP-SGD, що включає:

- нормування градієнтів (clipping);
- додавання випадкового шуму;
- обчислення параметра приватності ϵ для кожної епохи.

Такий підхід дозволяє формувати статистично подібні поведінкові ознаки, але без можливості відновлення реальних даних.

3. Програмна реалізація

Система реалізована мовою Python із використанням бібліотек PyTorch та Opacus. Програмний комплекс включає:

- модуль завантаження й підготовки даних (dataset);
- модулі генератора та дискримінатора (models);
- навчальний модуль із застосуванням DP-SGD (train.py);
- модуль оцінювання якості згенерованих даних за допомогою FID;
- систему збереження результатів.

У рамках роботи реалізовано:

- збереження сітки з 16 зображень (4×4) для наглядної демонстрації;
- генерацію 100 окремих синтетичних зразків, які моделюють поведінкові патерни;
- автоматичний підрахунок параметра приватності ϵ ;
- формування журналу тренування та оцінки якості.

4. Експериментальне тестування

У ході тестування було проведено серію запусків моделі, які продемонстрували:

- стабільну роботу DP-GAN при застосуванні приватності;
- очікуване погіршення якості зображень зі зменшенням ϵ , що є нормальним для DP-моделей;
- можливість контролю якості генерації за показником FID;
- формування повноцінного набору з 100 синтетичних прикладів, які підтверджують працездатність моделі.

Результати довели, що DP-GAN здатна формувати приватні синтетичні дані навіть без доступу до реальних біометричних записів.

5. Практичне значення

Створені синтетичні біометричні дані можуть бути використані для:

- тестування алгоритмів поведінкової аналітики;
- навчання моделей без ризику витоку приватної інформації;
- створення демонстраційних або дослідницьких датасетів;
- оцінки ефективності методів генерації та приватності;
- попередніх досліджень у сфері багатофакторної аутентифікації.

DP-GAN дозволяє працювати з чутливими даними без роботи з реальними користувачами, що значно розширює можливості досліджень і спрощує їх юридичний та етичний супровід.

Висновки

У роботі розроблено метод генерації синтетичних поведінкових біометричних даних на основі DP-GAN. Модель дозволяє створювати дані, які зберігають статистичні властивості реальних зразків, але не розкривають інформацію про користувачів завдяки застосуванню DP-SGD.

Програмна реалізація забезпечує тренування моделі, формування синтетичних зображень та оцінку їх якості. Експерименти підтвердили стабільність генерації та можливість регулювання рівня приватності.

Отримані результати демонструють практичну цінність підходу для створення безпечних біометричних наборів і подальших досліджень у сфері поведінкової аналітики та інформаційної безпеки.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Ахрамович В. М. Ідентифікація й аутентифікація, керування доступом // Сучасний захист інформації. – 2016. №4. – С. 47–51.
2. Бурячок В.Л., Гришук Р.В., Хорошко В.О. Політика інформаційної безпеки. – 2014.
3. Єсін В.І., Безпека інформаційних систем і технологій. – 2013.
4. Goodfellow I. Generative Adversarial Networks. – 2014.
5. Xie et al. Differentially Private GAN. – 2018.
6. *Opacus: Differential Privacy Library for PyTorch.* – Meta AI.

Гончарук Олександр Михайлович – студент групи КІТС-24м, факультет менеджменту і інформаційної безпеки, Вінницький національний технічний університет, Вінниця, e-mail: gsashatop@gmail.com

Науковий керівник: Грицак Анатолій Васильович – доцент каф. Менеджменту та безпеки інформаційних систем, Вінницький національний технічний університет, м. Вінниця, e-mail: grytsak.a.v@gmail.com

Honcharuk Oleksandr M. – student of the KITS-24m group, Faculty of Management and Information Security, Vinnytsia National Technical University, Vinnytsia, e-mail: gsashatop@gmail.com

Supervisor: Hrytsak Anatolii Vasyliovych – docent of the Department of Management and Security of Information Systems, Vinnytsia National Technical University, Vinnytsia, e-mail: grytsak.a.v@gmail.com