Метричний аналіз та обчислювальна стійкість цільових словників паролів

УДК 004.056.52 (519.1)

Сергій Бабич[1], Андрій Сидор[2], Петро Голуб[3]

Національний університет водного господарства та природокористування, [1]s.v.babych@nuwm.edu.ua, [2]a.i.sydor@nuwm.edu.ua, [3]p.p.holub@nuwm.edu.ua

Дослідження спрямовані на перетворення хаотичної генерацію варіантів у керований і обчислювально ефективний процес [1] синтезу ефективних цільових словників. Аби дослідження тримати в межах актуальності, необхідно оцінювати ефективність перетворення даних з «відомостей про сутність» (вхідні дані у форматі OSINT – дослідження) у подальші токени (стадія 1), паролі (стадія 2) та їх об'єднання – безпосередньо словник (стадія 3).

Центральним об'єктом запропонованої моделі є парольний токен — мінімальна семантично-персональна одиниця, вилучена з OSINT-профілю [2]. Наукова новизна підходу полягає у відмові від розуміння токена як лінгвістичної морфеми (наприклад, у алгоритмах BPE чи SentencePiece) на користь суб'єктивно значущої сутності (ім'я, дата, місце). Для структурування OSINT-даних запропоновано семикласову таксономію (T1–T7) [3], що включає ідентифікатори, часові, реляційні, просторові, тематичні, числові та контекстуальні дані. Описавши об'єкт оцінки ефективності, варто перейти до компоненти дослідження, що стосується вже саме оцінки ефективності.

Аналітичний огляд досліджень у сфері оцінки ефективності словників паролів демонструє поступовий перехід від класичних методів до складних моделей, що враховують як структурні характеристики паролів, так і контекстні дані користувачів. Розвиток цієї тематики відбувся у працях, де були запропоновані моделі TarGuess, які інтегрують відкриті дані (OSINT), включно з персональними атрибутами та інформацією із соціальних мереж, для побудови цільових словників. Ефективність таких словників значно зростає завдяки використанню реальних цифрових слідів користувачів, що підтверджує важливість ймовірнісного та статистичного аналізу у вимірюванні їх стійкості. Водночас ці роботи піднімають питання етики та приватності, адже використання персональних даних у словниках може створювати додаткові ризики [4]. Трендом останніх років стало застосування нейромережевих моделей: запропоновано PassGAN, який використовує генеративні змагальні мережі для створення словників. Цей підхід дозволяє моделювати розподіл паролів без явних правил, що робить словники більш адаптивними та здатними відображати реальні патерни користувачів. Метричний аналіз у цьому випадку включає оцінку ентропії та варіативності згенерованих словників, що дозволяє порівнювати їх ефективність із класичними методами [5].

Підсумувавши дослідження інших авторів щодо оцінки ефективності словників паролів загалом та виокремивши необхідне для нашого дослідження, що стосується цільових словників, варто зауважити, що останнє десятиліття ознаменувалося переходом від простої частотної статистики до складних моделей представлення знань, де базові показники: CR (Coverage Rate) та DSR

(Dictionary Size Ratio) залишаються основними для оцінки компактності та влучності. Але варто згадати і використання математичної близькості за метрики Дамерау-Левенштейна [6], котра стала стандартом для моделювання транспозицій (перестановок сусідніх знаків), що відображає механічні помилки користувачів. Щодо порівняння моделей, звернемо увагу, що класичні ймовірнісні граматики (PCFG) розбивають пароль на сегменти [L][D][S], але ігнорують семантичний зв'язок. Нейромережеві підходи, такі як PassGAN та PASSLLM, демонструють здатність вивчати логіку витоків, проте в цільових атаках вони у 10 000 разів повільніші за комбінаторні методи та часто страждають від перенавчання (overfitting) [7]. Ідеї Representational Learning (PLR), в своєю чергою, використовують латентний простір паролів для генерації навколо опорних точок (pivots), проте вимагають значних обчислювальних ресурсів.

Як висновок, варто звернути увагу, що в межах даного представлення – окреслено бачення команди щодо оцінки ефективності сформованих словників, при формуванні яких буде використано запропоновану методику СРП з роботи [1]. Але, вже можна виокремити те, що перехід до інтегрованої моделі оцінки, що поєднує OSINT-профілювання, семантичну токенізацію та метод перманентної декомпозиції, дозволяє перетворити аудит парольної політики з хаотичного підбору на науково обґрунтовану процедуру. Запропонований підхід не лише підвищує точність генерації при цільовому використанні, а й забезпечує обчислювальну стійкість систем захисту в умовах зростання цифрового сліду користувачів.

1. С. В. Бабич, «Інформаційна технологія складання розкладу занять згідно перманентної декомпозиції», дисертація кандидата технічних наук, Хмельницький національний університет, Хмельницький, Україна, 2023.
2. M. Bazzell, Open Source Intelligence Techniques: Resources for Searching and Analyzing Online Information, 9th ed. IntelTechniques, 2022.
3. B. Rader, “Targeted password cracking with OSINT data,” Purdue School of Engineering & Technology, IUPUI, 2023.
4. M. Weir, S. Aggarwal, B. de Medeiros, and B. Glodek, “Password cracking using probabilistic context-free grammars,” in Proc. IEEE Symp. Security and Privacy (SP), Oakland, CA, USA, 2009, pp. 391–405.
5. Y. Xie, J. Wang, and J. Yan, “TarGuess II: A statistical framework for targeted password guessing,” in Proc. ACM Conf. Computer and Communications Security (CCS), London, UK, 2020, pp. 1–14.
6. F. J. Damerau, “A technique for computer detection and correction of spelling errors,” Communications of the ACM, vol. 7, no. 3, pp. 171–176, 1964.
7. B. Hitaj, P. Gasti, G. Ateniese, and F. Perez-Cruz, “PassGAN: A deep learning approach for password guessing,” in Proc. Int. Conf. Applied Cryptography and Network Security (ACNS), Bogota, Colombia, 2019, pp. 217–237.