УДК 004.056.523:004.931
DOI: https://doi.org/10.30837/IYF.IRTTPI.2024.403

ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ ІДЕНТИФІКАЦІЇ ОСОБИ ЗА КЛАВІАТУРНИМ ПОЧЕРКОМ З УРАХУВАННЯМ СИЛИ ТИСКУ НА КЛАВІШІ

Горелов Д.Ю., Терновий Я.І.
Науковий керівник – к.т.н., доц. Горелов Д.Ю.
Харківський національний університет радіоелектроніки,
студентський науковий гурток «Біометричні технології контролю доступу»
каф. КРіСТЗІ, м. Харків, Україна
e-mail: yaroslav.ternovyi@nure.ua

Using the "Queen Mary University Keystroke benchmark dataset" database and the Orange software, a study of the influence of time parameters, dynamics of changes in key pressure and their combinations on the accuracy of identification by keyboard handwriting was carried out. It has been experimentally confirmed that both for multi-class classification problems and for two-class classification problems based on key pressure signs, it is possible to obtain an identification accuracy 99 %.

Традиційні засоби аутентифікації зазвичай ґрунтуються на паролях або на перевірці індивідуальних особливостей суб'єкта (біометричних ознак). Перші дуже схильні до «людського фактору», а біометричні системи захисту також не позбавлені недоліків. Щоб об'єднати переваги згаданих технологій, можна використовувати таємні біометричні ознаки, які можуть бути засновані тільки на динамічних біометричних ознаках, наприклад, індивідуальному клавіатурному почерку суб'єкта в процесі набору парольної фрази. Недолік такого методу полягає у порівняно низькій надійності прийнятих рішень. Підвищити надійність розпізнавання суб'єктів за клавіатурним почерком можна за допомогою використання додаткових ознак, що реєструються спеціальними датчиками та характеризують динаміку набору тексту на клавіатурі, наприклад, враховувати тиск на клавіші в процесі набору парольної фрази.

На даний момент у вільному доступі розташовано більше 10 баз даних параметрів клавіатурного почерку. Проте тільки два датасети містять інформацію про динаміку зміни тиску на клавіші в процесі набору парольних фраз. І тільки один є у вільному доступі – Queen Mary University Keystroke benchmark dataset [1-2]. Цей датаест складається з двох датасетів. Перший містить параметри тиску на клавіші в процесі вводу паролю «.try4-mbs». Другий містить часові параметри вводу користувачами цього паролю.

В якості програмного засобу для проведення досліджень було використано інструмент інтелектуального аналізу даних Orange. В якості алго-

ритму класифікації використовувався метод випадкових лісів [3]. В якості інструменту оцінки точності класифікації користувачів використовувалась крос-валідація за 10 блоками [4].

Результати мультикласової класифікації користувачів датасету «Queen Mary University Keystroke benchmark dataset» за часовими ознаками клавіатурного почерку наведено на рис. 1. Як можна бачити, рівень FAR відповідає високому рівню ідентифікації, а рівень FRR – не відповідає.

Результати мультикласової класифікації користувачів датасету «Queen Mary University Keystroke benchmark dataset» за ознаками тиску на клавіші наведено на рис. 2. Як можна бачити, рівень помилки другого роду відповідає дуже високому рівню ідентифікації. Для 27 з 30 користувачів (90 %) значення FAR дорівнює нулю, тобто система абсолютно точно визначає зловмисника та блокує йому доступ до інформаційної системи. Для 3 користувачів рівень FAR склав 0.1 %,тобто для одного випадку з тисячі система надасть доступ зловмиснику. Графік гістограми значень помилки першого роду також ілюструє високу точність ідентифікації – для 24 користувачів (80 %) значення FRR становить менше 1 %, що відповідає високій точності, ще для 5 користувачів значення FRR лежить в зоні низької (або прийнятної) точності, і лише для одного користувача значення FRR більше 3 %, що відповідає неприйнятній точності ідентифікації.

Отже, для задач мультикласової класифікації ознаки тиску на клавіші є набагато більш інформативними, ніж часові параметри натискань на клавіші.

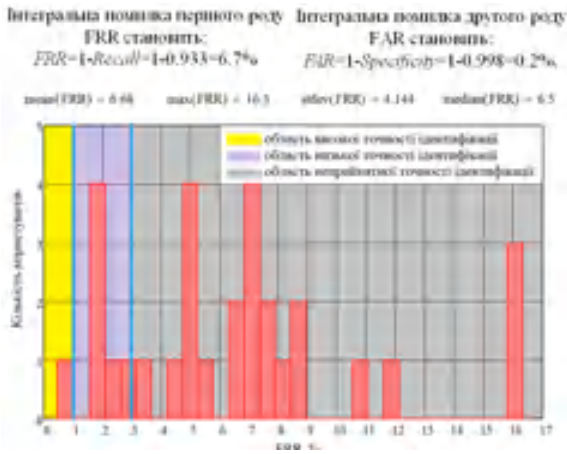


Рисунок 1

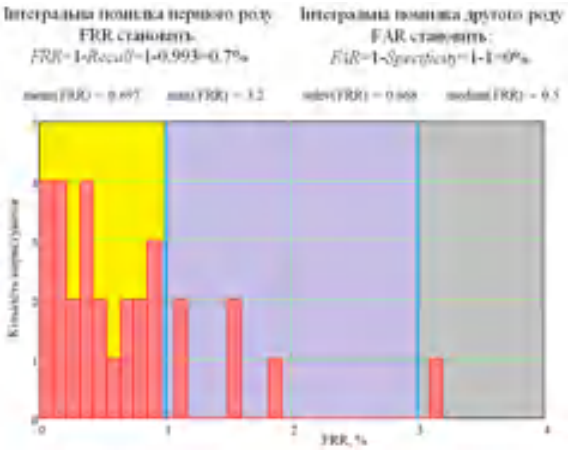


Рисунок 2

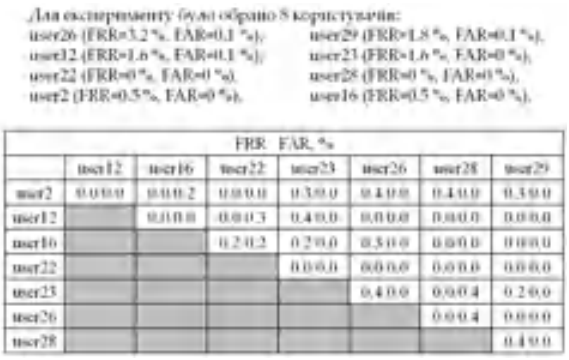
Наступним кроком є дослідження точності двокласної класифікації (є 2 класи: перший – зареєстрований користувач, другий – зловмисник) користувачів датасету «Queen Mary University Keystroke benchmark dataset» за ознаками тиску на клавіші. Для експерименту було обрано 8 користувачів, серед яких ті, в кого найнижча та найвища точність мультикласової класифікації.

Як можна бачити з рис. 3, для всіх можливих пар користувачів значення помилок першого та другого родів менше 0.5 відсотка, тобто

відповідають високій точності ідентифікації.

Отже, у випадку двокласової класифікації на основі ознак тиску на клавіші можна будувати основну систему контролю та управління доступом. В той час, як класичний клавіатурний почерк використовується лише, як допоміжна опція при основній системі ідентифікації.

Наступним кроком є дослідження точності ідентифікації на основі комбінованих ознак клавіатурного почерку: часових та тиску. Як можна бачити з рис. 4, отримані результати дуже цікаві.



Для експерименту було обрано 8 користувачів:
user26 (FRR=3.2 %, FAR=0.1 %);
user12 (FRR=1.6 %, FAR=0.1 %);
user22 (FRR=0 %, FAR=0 %);
user2 (FRR=0.5 %, FAR=0 %);
user29 (FRR=1.8 %, FAR=0.1 %);
user23 (FRR=1.6 %, FAR=0 %);
user28 (FRR=0 %, FAR=0 %);
user16 (FRR=0.5 %, FAR=0 %).

FRR, FAR, %							
	user12	user16	user22	user23	user26	user28	user29
user2	0.0 0.0	0.0 0.2	0.0 0.0	0.3 0.0	0.4 0.0	0.4 0.0	0.3 0.0
user12		0.0 0.0	0.0 0.3	0.4 0.0	0.0 0.0	0.0 0.0	0.0 0.0
user16			0.2 0.2	0.2 0.0	0.3 0.0	0.0 0.0	0.0 0.0
user22				0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0
user23					0.4 0.0	0.0 0.4	0.2 0.0
user26						0.0 0.4	0.0 0.0
user28							0.4 0.0

Рисунок 3

Комбінація ознак	FRR, %	FAR, %	max(FRR)	3-й квартиль(FRR)
Pressure	0.0	0.0	1.4	0.3
Pressure+ HoldTime	0.0	0.0	2.5	0.6
Pressure+ DownDownTime	0.0	0.0	2.2	0.3
Pressure+ UpDownTime	0.0	0.0	2.2	0.3
Pressure+ HoldTime+ DownDownTime	0.0	0.0	2.2	0.3
Pressure+ HoldTime+ UpDownTime	0.0	0.0	2.2	0.3
Pressure+ DownDownTime+ UpDownTime	0.0	0.0	2.2	0.3
Pressure+ HoldTime+ DownDownTime+ UpDownTime	0.0	0.0	2.8	0.3

Рисунок 4

По-перше, значення помилки першого роду для зменшеного датасету ознак тиску є меншими, ніж значення FRR для повного датасету. Цей ефект дуже нагадує помилку перенавчання, коли складна модель показує добрі результати на навчальній вибірці, але не працює на тестовій. Було проведено декілька експериментів, в яких змінювалась кількість дерев в алгоритмі Random Forest, а також використані класифікатори на основі k найближчіх сусідів та нейронна мережа з одним прихованим шаром. Результати усюди однакові: рівень помили FRR у зменшеного датасету менший. Отже, можна стверджувати про актуальність задачі пошуку та фільтрації шумових даних, отриманих з датчиків тиску.

По-друге, комбінація ознак тиску з будь-якими часовими ознаками клавіатурного почерку також призводить до погіршення точності ідентифікації.

Тут може бути два пояснення. Перше, це те, що невідомо чи відповідають в датасетах однакові номери спроб вводу пароля конкретній події, тобто вектор тиску та вектор часових параметрів відповідають одній і тій же спробі вводу паролю чи ні. Якщо ні – то отримані результати не мають цінності. Друге пояснення – це, знову ж, зашумлення даних. Слід очікувати, що більшій силі тиску відповідатиме більший час утримання клавіші. Можливі ці кореляційні зв’язки і зашумлюють дані. В будь-якому разі, це питання майбутніх досліджень та пошуків датасетів.

Оскільки питання впливу на точність ідентифікації використання комбінації ознак динаміки зміни тиску та часових параметрів не вирішено, рекомендувати використовувати ознаки тиску на клавіші в мультимодаль-

них біометричних системах неможна. Виключення становлять випадки, коли усі модальності використовуються окремо в залежності від сценарію роботи системи (наприклад, вхід в систему за паролем та одночасно моніторинг динаміки вводу паролю на клавіатурі, а далі прихований моніторинг за користувачем в процесі роботи на основі ознак тиску на клавіші).

Висновки. Враховуючи високу точність ідентифікації, ознаки тиску на клавіші в процесі роботи за клавіатурою можуть бути використані в системах виявлення потенційних внутрішніх порушників інформаційної безпеки – чим вища точність, тим точніше можна визначити діапазони зміни дослідних ознак клавіатурного почерку для кожного користувача.

Як відомо, здебільшого внутрішніми інсайдерами є звичайні співробітники, які вимушені з тієї чи іншої причини чинити протизаконні дії. Такі інсайдери часто характеризуються високим рівнем тривожності, схильні до інтрапунітивних реакцій (самообвинувачування, похмурість, злість, незадоволення, напруженість, сердитість), чим відрізняються від справжніх злочинців.

Для такої категорії порушників можливе проактивне (на ранніх етапах) виявлення потенційних, схильних до протиправних дій осіб шляхом контролю їхньої діяльності та оцінки психоемоційного стану.

Одним з технічних індикаторів, які можуть вказувати на наявність потенційної інсайдерської загрози, може виступати клавіатурний почерк у сукупності таких показників, як сила тиску на клавіші, динаміка введення, системні друкарські помилки та використання певних літер, символів та «гарячих» клавіш. Сукупність певних значень кожного з цих показників утворює досить індивідуальну картину, що відповідає конкретній людині у певному психоемоційному стані.

Список використаних джерел:

1. Loy C. C., Lim C .P., Lai W. K. Pressure-based typing biometrics user authentication using the fuzzy ARTMAP neural network // International Conference on Neural Information Processing (ICONIP), 2017.

2. Queen Mary University Keystroke benchmark dataset. URL: https://personal.ie.cuhk.edu.hk/~ccloy/downloads_keystroke100.html#:~:text=Keystroke100%20benchmark%20dataset%20is%20a,password%20%22try4%2Dmbs%22 (дата звернення: 20.12.2023).

3. Hastie T., Tibshirani R., Friedman J. Random Forests // The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. Springer-Verlag, 2009. Chapter 15. 746 p.

4. Cross-validation: evaluating estimator performance. URL: https://scikit-learn.org/stable/modules/cross_validation.html (дата звернення: 20.12.2023).

5. Дослідження інформативних параметрів диграфів клавіатурного почерку для задач ідентифікації користувачів комп'ютерних мереж / Д.Ю. Горелов, О.О. Іванова, О.В. Кокорін, Д.В. Маслій, О.В. Литвиненко // Радіотехніка: Всеукр. Міжвід. Наук.-техн. Зб. – 2020. – вип. 201. – с. 194 – 200.