ДОСЛІДЖЕННЯ МОЖЛИВОСТЕЙ ВИКОРИСТАННЯ КУРСОРНОГО ПОЧЕРКУ ДЛЯ ВИРІШЕННЯ ЗАДАЧ БІОМЕТРИЧНОЇ ІДЕНТИФІКАЦІЇ

Литвиненко О.В.
Науковий керівник – к.т.н., доц. Горелов Д.Ю.
Харківський національний університет радіоелектроніки
(61166, Харків, пр. Науки,14, навчально-наукова лабораторія «Систем технічного захисту інформації (відеоспостереження, охоронні сигналізації і контроль доступу)», тел. (057) 702-14-78.

The objective of the work is explore the possibility of using mouse clickstream in computer networks' user identification tasks.

Останнім часом велика увага приділяється методам біометричної ідентифікації особистості за динамікою підсвідомих рухів рук. Мова йде про виявлену стабільність відпрацьованих рухових навичок людини – клавіатурний та курсорний почерки – і можливості її розпізнавання за цією ознакою. Висока ефективність клавіатурного почерку в рішенні вищезазначених задач вже доказана великою кількістю опублікованих робіт. Курсорний почерк менш вивчений і досі актуальними є питання: чи є даний засіб ідентифікації самостійним, як, наприклад, клавіатурний почерк?

Для проведення досліджень був обраний датасет «The Balabit Mouse Challenge Data Set», який був створений для пошуку алгоритмів та методів ідентифікації користувачів за динамікою комп'ютерної миші. Датасет складається з двох частин: навчального датасету та тестового, який в свою чергу містить як позитивні (легальні), так і негативні (незаконні) сесії. Також до датасету сторонніми дослідниками розроблено пакет програм, що дозволяє розраховувати будь-які, цікаві для конкретного дослідника, динамічні параметри системи «користувач-миша».

На першому кроці досліджень усі дані дослідного датасету було розділено на три типи дії миші: movement (MM), point click (PC), drag-and-drop (DD), бо особливості роботи з комп'ютерною мишею описують як характеристики взаємодії з маніпулятором на рівні інтерфейсу (переміщення курсору), так і взаємодії з маніпулятором як з фізичним об'єктом (тривалість натискання клавіш миші) і такий розподіл дозволить визначити які параметри, а саме динаміка переміщення курсору чи натискання клавіш є більш стабільними,а, отже, і оптимальними для використання в задачах ідентифікації.

На другому кроці з огляду на літературні джерела дослідні датасети було проріджено з метою залишити тільки дії з кількістю індивідуальних подій миші більше 40. Саме цей поріг найчастіше вказують дослідники у своїх роботах. Дії миші з меншою кількістю індивідуальних подій мають випадковий характер з значно погіршують точність класифікації, якщо включені до аналізу.

На третьому кроці було сформовано 37 інформативних параметрів курсорного почерку: середні, мінімальні, максимальні та середньоквадратичні значення горизонтальної, вертикальної, загальної, кутової швидкостей переміщення курсору, прискорення, ривка та кривизни траєкторії переміщення курсору; тип дії (PC чи DD); довжина траєкторії; довжина відрізка між двома кінцевими точками траєкторії; прямолінійність траєкторії ; час виконання дії; напрямок дії (один з восьми квадрантів, якому належить кут між віссю абсцис та відрізком, що з'єднує дві кінцеві точки траєкторії); найбільша відстань від траєкторії до відрізка, що з'єднує дві кінцеві точки траєкторії; відношення тривалості проходження перших десяти відсотків траєкторії переміщення курсору до часу виконання дії; швидкість проходження перших десяти відсотків довжини траєкторії.

На четвертому кроці за допомогою програмного засобу Orange було проведено класифікацію користувачів датасету за 37 інформативними параметрами курсорного почерку за допомогою методу опорних векторів (SVM) з гаусовим ядром та радіальною базовою функцією; методу випадкового лісу (RF) з 20 деревами прийняття рішень; методу k найближчих сусідів (k-NN) з k=50; радіальної нейронної мережі (NN) з одним скритим шаром та 100 нейронами в ньому. Основні отримані результати:

1) найінформативнішими для дослідних класифікаторів є DD дії;

2) алгоритми k найближчих сусідів та штучних нейронних мереж забезпечили найменший рівень FRR;

3) алгоритми Random Forest та штучних нейронних мереж забезпечили найменший рівень FAR;

4) мінімальний рівень ймовірності заборони доступу зареєстрованим користувачам – 0.045, мінімальний рівень ймовірності заборони доступу зареєстрованим користувачам – 0.0678;

5) курсорний почерк можна використовувати як самостійний засіб ідентифікації. Хоча, слід зазначити, для досягнення малих значень помилок FAR та FRR треба дуже великі об'єми датасетів або організовувати спеціальні сценарії отримання даних, що неможливо.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Balabit Mouse Dynamics Challenge Data Set. Режим доступу: https://github.com/balabit/Mouse-Dynamics-Challenge.

2. Teng Hu, Weina Niu, Xiaosong Zhang, Xiaolei Liu, Jiazhong Lu, Yuan Liu, "An Insider Threat Detection Approach Based on Mouse Dynamics and Deep Learning", Security and Communication Networks, 2019.

3. C. Shen, Z. Cai, X. Guan, Y. Du, and R. A. Maxion, "User authentication through mouse dynamics," IEEE Transactions on Information Forensics and Security, vol. 8, no. 1, pp. 16–30, 2013.