DOI: https://doi.org/10.25140/2411-5363-2025-2(40)-294-302
УДК 004.03:004.9

Сергій Васильович Коваленко[1], Олексій Васильович Красножон[2]

[1]аспірант кафедри інформаційних та комп'ютерних систем, група АСД-122-23
Національний університет «Чернігівська політехніка» (Чернігів, Україна)
E-mail: kovalenkoptp10@gmail.com. **ORCID:** https://orcid.org/0009-0003-6993-010X

[2]кандидат технічних наук, доцент кафедри інформаційних та комп'ютерних систем
Національний університет «Чернігівська політехніка» (Чернігів, Україна)
E-mail: wingcommander2011@gmail.com. **ORCID:** https://orcid.org/0000-0003-2500-254X
ResearcherID: G-4623-2014. **Scopus Author ID:** 57190377188

МЕТОДИ ПЕРВИННОЇ ОБРОБКИ ДАНИХ У СИСТЕМАХ ІДЕНТИФІКАЦІЇ КОРИСТУВАЧІВ НА ОСНОВІ КЛАВІАТУРНОГО ПОЧЕРКУ

Запропонована стаття є оглядовою, її присвячено дослідженню існуючих методів ідентифікації користувачів на основі клавіатурного почерку, зокрема таких, як методи динамічного аналізу тривалості натискання клавіш, статистичні моделі розпізнавання шаблонів набору тексту та нейромережеві підходи до класифікації користувачів.

Методи динамічного аналізу тривалості натискання клавіш базуються на вимірюванні часових характеристик набору тексту, таких як середній інтервал між натисканнями та утриманням клавіш. Ці параметри є унікальними для кожного користувача та можуть використовуватися для автентифікації.

Статистичні моделі розпізнавання шаблонів набору тексту дозволяють аналізувати комбінації символів, які вводяться з певною швидкістю та послідовністю. Використання таких моделей дає змогу виявляти закономірності у введенні тексту, що допомагає у верифікації користувача.

Нейромережеві підходи до класифікації користувачів використовують глибоке навчання для обробки великих обсягів даних про клавіатурний почерк. Ці методи дозволяють створювати складні багаторівневі моделі для аналізу поведінкових характеристик набору тексту, що підвищує точність ідентифікації.

Використання методів динамічного аналізу, статистичних моделей і нейромережевих алгоритмів може значно підвищити рівень безпеки користувачів, знизити ризики несанкціонованого доступу та забезпечити надійну ідентифікацію в інформаційних системах.

Ключові слова: *клавіатурний почерк; біометрична автентифікація; ідентифікація користувачів; динамічний аналіз затримок; статистичні моделі розпізнавання; нейромережеві алгоритми; поведінкова автентифікація; машинне навчання; інформаційна безпека; первинна обробка даних.*

Рис.: 1. Бібл.: 13.

Актуальність теми дослідження. Сучасні інформаційні технології постійно вдосконалюють методи ідентифікації користувачів, оскільки традиційні паролі та інші статичні методи автентифікації не завжди забезпечують належний рівень безпеки. Одним із перспективних напрямів у цій сфері є використання клавіатурного почерку, який дозволяє аналізувати індивідуальні особливості набору тексту для підтвердження особи користувача.

Зростання інтересу до цієї технології зумовлено низкою факторів, зокрема:

– паролі часто піддаються атакам, забуваються або передаються третім особам, що створює ризик несанкціонованого доступу;

– на відміну від разової перевірки пароля при вході, аналіз клавіатурного почерку дозволяє здійснювати контроль користувача протягом усього сеансу роботи із системою;

– методи ідентифікації за клавіатурним почерком можуть бути інтегровані в банківські системи, корпоративні мережі, освітні платформи тощо для підвищення безпеки.

Згідно з дослідженнями в галузі біометричної автентифікації, методи аналізу клавіатурного почерку дозволяють досягти високого рівня точності при мінімальних витратах на впровадження, оскільки не потребують додаткового обладнання. Водночас проблема первинної обробки даних у таких системах залишається актуальною через необхідність ефективної фільтрації, нормалізації та аналізу великої кількості параметрів.

Таким чином, розвиток інформаційних систем первинної обробки даних у системах ідентифікації користувачів на основі клавіатурного почерку є важливою науковою та практичною задачею, вирішення якої сприятиме підвищенню рівня безпеки та зниженню ризиків несанкціонованого доступу до інформаційних ресурсів.

Постановка проблеми. Основними проблемами, які виникають при розробці та впровадженні систем автентифікації на основі клавіатурного почерку, є:

1. Низька точність у порівнянні з іншими біометричними методами: клавіатурна біометрія поступається технологіям розпізнавання обличчя, відбитків пальців або райдужної оболонки ока за рівнем ідентифікації користувача.

2. Вплив зовнішніх факторів: фізичний стан користувача (втома, емоційний стан) або зміна пристрою введення можуть суттєво впливати на характеристики набору тексту, що ускладнює коректну автентифікацію.

3. Необхідність адаптації до змін у поведінці користувача: шаблони клавіатурного почерку можуть змінюватися з часом, тому система повинна забезпечувати механізми оновлення еталонних даних для мінімізації відхилення легітимних користувачів.

4. Обмеженість стандартних наборів даних: у порівнянні з іншими біометричними методами, є небагато відкритих наборів даних для навчання моделей, що ускладнює розробку ефективних алгоритмів машинного навчання.

5. Ризики атак та шахрайств: хоча клавіатурна біометрія є складнішою для підробки, існують методи імітації почерку, які можуть бути використані зловмисниками для обходу систем безпеки [1].

На рис. 1 представлено класифікацію існуючих методів автентифікації, включаючи біометричні технології, серед яких клавіатурна біометрія посідає окреме місце в поведінковій автентифікації [1].

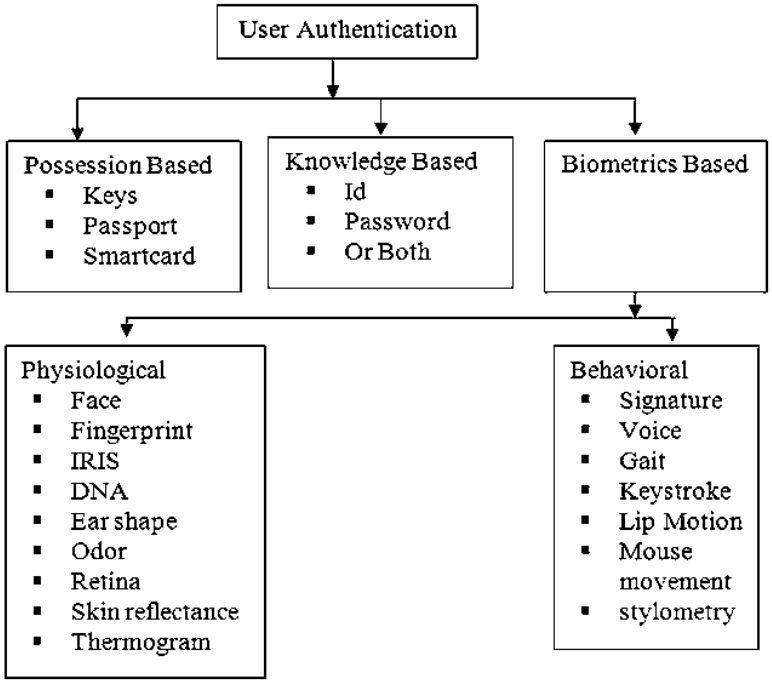

Рис. 1. Онтологія різних методів автентифікації

Джерело: [1].

Центральною проблемою, що розглядається в цьому дослідженні, є забезпечення надійної первинної обробки даних клавіатурного почерку, яка дозволяє зменшити вплив зовнішніх факторів, підвищити точність автентифікації та забезпечити стабільність характеристик у динаміці поведінки користувача. Саме від якості попередньої обробки даних залежить ефективність подальшого застосування алгоритмів розпізнавання та ідентифікації в реальних умовах експлуатації.

Аналіз останніх досліджень і публікацій показує, що методи ідентифікації користувачів за клавіатурним почерком, зокрема методи динамічного аналізу затримок натискання клавіш, статистичні моделі розпізнавання шаблонів набору тексту та нейромережеві підходи до класифікації користувачів, демонструють значний прогрес у забезпеченні високої точності автентифікації. Ці методи мають велику перспективу для покращення безпеки в інформаційних системах і можуть значно знизити ризики несанкціонованого доступу [2].

Методи динамічного аналізу затримок натискання клавіш орієнтуються на вимірювання часового інтервалу між натисканнями клавіш, а також на аналіз середнього часу натискання клавіш і періодів між натисканнями та утриманням клавіші. Оскільки ці параметри є індивідуальними для кожного користувача, вони можуть бути використані для автентифікації. Зокрема, час між натисканнями клавіш дозволяє визначити стиль набору тексту конкретного користувача, що є унікальним для кожної особи та допомагає ідентифікувати користувача серед інших. Ці методи забезпечують високу чутливість та специфічність для розпізнавання користувачів [3].

Статистичні моделі розпізнавання шаблонів набору тексту зосереджуються на виявленні регулярних залежностей у введенні тексту, таких як швидкість набору, послідовність натискання клавіш та комбінації символів. Статистичні методи аналізують середні значення та стандартні відхилення часів між натисканнями для побудови профілю користувача. Це дозволяє виявляти приховані закономірності в поведінці набору тексту, що покращує точність автентифікації. Останні дослідження показують, що ці методи продовжують використовуватись у багатьох системах, але вони поступаються більш сучасним методам, таким як нейронні мережі, за точністю та гнучкістю [3].

Нейромережеві методи, зокрема методи глибокого навчання, здатні обробляти великі обсяги даних про клавіатурний почерк, створюючи складні моделі для класифікації користувачів. Вони забезпечують високу точність, оскільки враховують різноманітні фактори, такі як тиск на клавіші, тривалість натискання та інші варіації в поведінці користувача. Ці підходи також можуть адаптуватися до змін у стилі набору тексту, що дає можливість здійснювати точну автентифікацію навіть при змінах у поведінці користувача з часом. Вони є одними з найбільш точних підходів для ідентифікації користувачів у системах безпеки [3].

Виділення недосліджених частин загальної проблеми. Незважаючи на значний прогрес у розвитку систем автентифікації на основі клавіатурного почерку, існує низка невирішених питань, які потребують додаткового дослідження. Однією з основних проблем є недостатня точність ідентифікації в порівнянні з іншими біометричними методами, такими як відбитки пальців або розпізнавання обличчя.

Ще одним важливим аспектом є адаптація алгоритмів до змін у поведінці користувача. Наразі більшість систем використовує статичні моделі, які не враховують природні зміни в ритмі друку, спричинені втомою, стресом або використанням різних пристроїв введення. Це може призводити до збільшення рівня помилкових відмов легітимним користувачам.

Мета статті. Метою цієї статті є дослідження існуючих методів ідентифікації користувачів на основі клавіатурного почерку, аналіз їх ефективності та визначення напрямів розвитку інформаційних систем первинної обробки даних. Дослідження охоплює основні алгоритми, що використовуються для аналізу динамічних характеристик набору тексту, а також перспективи застосування машинного навчання та адаптивних моделей у цій сфері. Особлива увага приділяється проблемам забезпечення високої точності, адаптації до змін у поведінці користувачів та можливим загрозам безпеки, пов’язаним із підробкою клавіатурного почерку.

Виклад основного матеріалу. Методи динамічного аналізу затримок натискання клавіш є одним з основних підходів для автентифікації користувачів за допомогою клавіатурного почерку. Вони базуються на вимірюванні різних часових характеристик під час набору тексту, зокрема середнього часу натискання клавіші, інтервалів між натисканнями, а також тривалості утримання клавіші. Це дозволяє створити унікальний профіль поведінки кожного користувача, що є основою для подальшої автентифікації [4].

Один з основних аспектів динамічного аналізу полягає в вимірюванні часу між натисканнями клавіш, який має тенденцію варіюватися від користувача до користувача. Динамічний аналіз включає не тільки безпосередній час натискання, але й час, що проходить між натисканнями сусідніх клавіш, а також час між натисканням і відпусканням клавіші [5]. Такі параметри є індивідуальними для кожного користувача і створюють унікальний «візерунок» або шаблон, який можна використовувати для ідентифікації. Наприклад, швидкий або повільний набір тексту, часті помилки чи корекції під час набору можуть бути відображені в часових інтервалах між натисканнями клавіш. Таким чином, навіть якщо користувач змінює інтерфейс або пристрій введення, його стиль набору тексту може залишатися достатньо стабільним для ідентифікації [6]

Крім того, важливе значення мають такі характеристики, як час утримання клавіші. Цей параметр є індикатором того, наскільки довго користувач утримує певну клавішу, що також може варіюватися між людьми залежно від їхнього стилю набору тексту. Наприклад, людина, яка швидко набирає текст, може утримувати клавіші на коротший період, тоді як користувач, що набирає текст повільніше, може утримувати клавіші довше. Це робить методи динамічного аналізу дуже корисними для автентифікації в реальному часі, оскільки вони дозволяють відстежувати навіть мінімальні зміни в поведінці користувача.

Методи динамічного аналізу також можуть бути використані для виявлення аномалій у поведінці користувачів. Наприклад, якщо певний користувач раптово змінює стиль набору тексту (наприклад, його інтервали між натисканнями клавіш різко змінюються або він починає набирати текст набагато повільніше), це може свідчити про спробу несанкціонованого доступу або атаку, що змушує систему активувати додаткові заходи безпеки. Застосування алгоритмів машинного навчання та статистичних моделей для аналізу цих змін дозволяє підвищити точність системи автентифікації та знизити рівень помилкових відмов [6].

Однією з переваг методів динамічного аналізу є їхня здатність адаптуватися до змін у поведінці користувачів з часом. Наприклад, користувач може змінювати свою швидкість набору тексту через фізичні або психологічні зміни, такі як втома, стрес або зміна робочого середовища. Динамічні методи можуть бути налаштовані таким чином, щоб враховувати ці зміни, а отже, зберігати точність автентифікації навіть в умовах змін у поведінці користувача.

Однак, попри свою високу ефективність, методи динамічного аналізу мають і певні недоліки. Один із них – це вплив різноманітних факторів, таких як використання різних пристроїв (наприклад, клавіатур на різних типах пристроїв), що можуть змінювати інтервали натискання клавіш та тривалість утримання. Іншим обмеженням є потенційна зміна поведінки користувачів через фізичні фактори (наприклад, травми пальців або зміни в фізичному стані), що може призвести до невірної ідентифікації [6].

Статистичні моделі розпізнавання шаблонів набору тексту є інструментом для ідентифікації та автентифікації користувачів на основі їх індивідуального стилю введення. Ці моделі ґрунтуються на математичних та статистичних методах для виявлення регулярних залежностей у поведінці користувача під час набору тексту, що дозволяє створювати унікальні профілі для кожної особи. Вони можуть бути використані для автентифікації користувачів, запобігання шахрайству, а також для забезпечення безпеки в онлайн-системах.

Статистичні моделі для розпізнавання шаблонів набору тексту зазвичай включають такі компоненти, як тривалість натискання клавіші, інтервали між натисканнями клавіш, а також стиль набору тексту в цілому. Основний принцип цих моделей полягає в тому, щоб за допомогою різних алгоритмів і статистичних методів аналізувати та класифікувати ці шаблони. Однією з найбільш поширених методик є використання статистичних класифікаторів, таких як методи на основі ймовірнісних моделей (Naive Bayes), лінійні або нелінійні класифікатори, а також нейронні мережі [7].

Одним із ключових аспектів статистичних моделей є аналіз ймовірностей певних подій, наприклад, ймовірність натискання певної клавіші в конкретний момент часу або ймовірність появи певного інтервалу між натисканнями клавіш. Ці ймовірності потім використовуються для побудови профілю користувача, що дозволяє ідентифікувати його на основі шаблонів набору тексту, які він генерує. Застосування ймовірнісних моделей дає змогу системам автентифікації враховувати статистичні коливання в поведінці користувачів, наприклад, через фізіологічні або психологічні зміни, які можуть вплинути на стиль набору тексту.

Важливим етапом у розпізнаванні шаблонів набору тексту є попередня обробка даних. Це включає фільтрацію шуму, усунення аномальних даних та нормалізацію інтервалів часу між натисканнями клавіш. Завдяки цьому можна отримати більш точні й стабільні характеристики для кожного користувача, що полегшує подальшу класифікацію та зменшує ймовірність помилкових спрацьовувань системи автентифікації [7]. Крім того, перед обробкою даних важливо розуміти, як різні фактори, такі як стрес, втома або використання різних пристроїв введення, можуть вплинути на стиль набору тексту, що може змінити статистичні характеристики шаблонів.

Одним із підходів до побудови статистичних моделей є використання марковських моделей для моделювання послідовностей натискань клавіш. Марковські моделі можуть бути використані для моделювання ймовірностей переходу від одного натискання клавіші до іншого, що дозволяє аналізувати ідентифікаційні шаблони набору тексту на основі послідовностей подій. Вони можуть враховувати як поточний стан, так і історичні дані, що дозволяє моделі адаптуватися до змін у поведінці користувачів із часом [8].

З іншого боку, статистичні моделі можуть стикатися із проблемами, пов’язаними з адаптацією до змін у поведінці користувачів. Наприклад, зміна пристрою введення (перехід із ноутбука на смартфон) може значно змінити шаблони набору тексту. Для вирішення цих проблем необхідно використовувати гібридні моделі, що комбінують статистичні методи з іншими технологіями, такими як аналіз контексту чи поведінкові аналізи, що дозволяє зменшити вплив таких змін на точність системи [8].

Крім того, важливим аспектом є ймовірність появи помилкових спрацювань або високої чутливості до змін у поведінці користувача. Тому багато сучасних систем використовують багаторівневі методи автентифікації, що поєднують статистичні моделі з іншими біометричними характеристиками, такими як розпізнавання обличчя або відбитків пальців. Це дозволяє досягти вищої точності при збереженні надійності та стійкості до атак [8].

Для підвищення точності розпізнавання шаблонів набору тексту використовуються методи машинного навчання, зокрема глибокі нейронні мережі (Deep Neural Networks, DNNs), вони дозволяють будувати складні багаторівневі представлення даних, що дає змогу ефективно класифікувати користувачів на основі їхніх унікальних характеристик. Нейронні мережі здатні автоматично виділяти важливі риси з великих наборів даних і пристосовуватися до нових шаблонів поведінки користувачів. Це дозволяє знижувати ймовірність помилкових відмов та підвищувати точність системи навіть за умов великих змін у стилі набору тексту користувачем [9].

Нейромережеві підходи до класифікації користувачів є одними з найсучасніших та ефективних методів у сфері біометричної автентифікації, зокрема для класифікації користувачів на основі їхнього стилю набору тексту. Нейронні мережі, завдяки своїй здатності навчатися з великих обсягів даних і виявляти складні, нелінійні зв'язки, використовуються для побудови моделей, які можуть автоматично класифікувати користувачів на основі їхніх індивідуальних шаблонів введення. Застосування нейромереж дозволяє значно підвищити точність і стійкість системи, зменшуючи ймовірність помилкових відмов і помилкових спрацьовувань при автентифікації користувачів [9].

Основні переваги нейронних мереж у класифікації користувачів полягають у їхній здатності працювати з великими обсягами даних, їх адаптивності до змін у поведінці користувачів та здатності виділяти значущі риси навіть із неструктурованих даних, таких як інтервали між натисканнями клавіш або тривалість натискання. Крім того, нейромережі можуть обробляти складні залежності між характеристиками, що дозволяє враховувати всі можливі аспекти поведінки користувача, які можуть варіюватися від фізіологічних змін до впливу стресових ситуацій або зміни пристроїв введення [9].

Для класифікації користувачів на основі стилю набору тексту використовуються різні архітектури нейронних мереж, серед яких найбільш популярними є багатошарові персептрони (Multi-Layer Perceptron або MLP), згорткові нейронні мережі (Convolutional Neural Network або CNN) і рекурентні нейронні мережі (Recurrent neural Network або RNN). Кожен із цих підходів має свої особливості і переваги залежно від типу даних і поставленої задачі [10].

Рекурентні нейронні мережі використовуються для роботи із послідовностями даних, де важливий контекст та порядок подій. Оскільки набір тексту є послідовним процесом, де кожне натискання клавіші може залежати від попередніх, RNN є дуже підхожим для задач класифікації користувачів за допомогою їхнього стилю набору тексту. Використовуючи внутрішні стани для збереження контексту, ці мережі можуть ефективно обробляти послідовності натискання клавіш і створювати точні профілі користувачів [10].

Згорткові нейронні мережі часто використовуються для роботи із зображеннями, але вони також можуть бути ефективно застосовані для аналізу послідовностей даних, таких як шаблони набору тексту. Використовуючи операції згортки, CNN можуть автоматично виявляти важливі особливості в послідовностях натискання клавіш, що дозволяє підвищити точність класифікації. У контексті класифікації користувачів це означає, що мережа може ефективно розпізнавати шаблони введення, які є характерними для конкретного користувача, навіть при варіаціях його поведінки [11].

Багатошарові персептрони є одними з найпростіших видів нейронних мереж, але вони можуть бути дуже ефективними при класифікації користувачів за допомогою шаблонів набору тексту. Вони складаються із вхідного, одного або декількох прихованих і вихідного шарів, що дозволяє моделі навчатися від простих до більш складних ознак. Для задачі класифікації користувачів такий тип мережі може використовувати характеристики натискання клавіші, такі як інтервал між натисканнями, тривалість натискання та інші параметри, що допомагає створити унікальні профілі користувачів [12].

Процес навчання нейронних мереж для класифікації користувачів передбачає використання великих наборів даних, що містять інформацію про шаблони набору тексту різних користувачів. Дані, зібрані під час введення тексту, зазвичай включають інформацію про інтервали між натисканнями клавіші, тривалість натискання, швидкість введення та інші характеристик. Під час навчання нейронна мережа вивчає ці дані та створює модель, яка може класифікувати нові дані, отримуючи точніші результати щодо того, до якого користувача належить той чи інший набір шаблонів [12].

Нейронні мережі, на відміну від традиційних статистичних методів, здатні працювати із неструктурованими даними й автоматично визначати важливі характеристики без потреби в ручному відборі ознак. Це дозволяє зменшити людський фактор і підвищити точність класифікації, навіть коли користувач змінює свій стиль набору тексту або використовує різні пристрої введення [13].

Незважаючи на високий потенціал штучних нейронних мереж для класифікації користувачів, існують деякі виклики, пов'язані з їх застосуванням у реальних умовах. Одним із головних викликів є проблема адаптації до змін у поведінці користувача. Зміни у пристроях введення, психологічному стані користувача або навіть зміна середовища можуть значно вплинути на шаблони набору тексту. Для вирішення цієї проблеми часто використовуються методи перенавчання або комбінування різних типів мереж, що дозволяє адаптувати модель до нових умов.

Іншим важливим аспектом є проблема загальної варіативності серед користувачів. Не всі користувачі мають однаковий стиль набору тексту, і нейронні мережі повинні бути здатні враховувати цю різноманітність для досягнення високої точності класифікації. Це може вимагати додаткових налаштувань і параметризації моделей для забезпечення більшої гнучкості при обробці даних [13].

Висновки. У статті було розглянуто основні методи ідентифікації користувачів на основі клавіатурного почерку, зокрема динамічний аналіз затримок натискання клавіш, статистичні моделі розпізнавання шаблонів набору тексту та нейромережеві підходи до класифікації користувачів. Проведений аналіз дозволяє визначити переваги та обмеження кожного з методів, а саме:

– методи динамічного аналізу затримок натискання клавіш забезпечують простоту впровадження та не потребують додаткового обладнання, проте їхня точність може знижуватися під впливом зовнішніх факторів, таких як зміна пристрою введення чи фізичний стан користувача;

– статистичні моделі розпізнавання шаблонів набору тексту дозволяють аналізувати послідовності символів та їх характеристики, що підвищує стійкість до випадкових змін у почерку, але вимагає збору значного обсягу даних для коректного навчання;

– нейромережеві підходи демонструють високу точність розпізнавання користувачів завдяки здатності враховувати нелінійні залежності у даних, проте їх реалізація потребує значних обчислювальних ресурсів та якісних наборів даних для навчання.

Зважаючи на все зазначене вище, використання комбінованих методів, що поєднують динамічний аналіз затримок із машинним навчанням, є найбільш перспективним напрямком розвитку систем ідентифікації користувачів на основі клавіатурного почерку. Це дозволить підвищити рівень безпеки інформаційних систем та мінімізувати ризик від несанкціонованого доступу.

Список використаних джерел

1. Shadman, R., Wahab, A. A., Manno, M., Lukaszewski, M., Hou, D., & Hussain, F. (2025). Keystroke dynamics: Concepts, techniques, and applications. *ACM Computing Surveys*. https://www.semanticscholar.org/reader/826113b73fa96eb1d6be90e66ae688c06d086109.

2. Shanmugapriya, D., & Padmavathi, G. (2010). A Survey of Biometric Keystroke Dynamics: Approaches, *International Journal of Computer Science and Information Security,* 5(1) https://www.researchgate.net/publication/41171720_A_Survey_of_Biometric_keystroke_Dynamics_Approaches_Security_and_Challenges.

3. Rashik Shadman Ahmed Anu Wahab Michael Manno Matthew Lukaszewski Daqing Hou Faraz Hussain (2024). *Keystroke Dynamics: Concepts, Techniques, and Applications* https://arxiv.org/pdf/2303.04605.

4. Teh, P. S., Teoh, A. B. J., & Yue, S. (2013). A survey of keystroke dynamics biometrics. *The Scientific World Journal*, *2013*, 1–24. https://doi.org/10.1155/2013/408280.

5. Puvirajasingam, K., & Sangeetha, D. (2013). Biometric Authentication using Keystroke Dynamics: A Survey. *IJREAT International Journal of Research in Engineering & Advanced Technology*, 1(5). http://ijreat.org/Papers%202013/Issue5/IJREATV1I5079.pdf.

6. Gaines, R. & Lisowski, William & Press, S. & Shapiro, Norman. (1980). Authentication by Keystroke Timing: Some Preliminary Results. 52.

7. Migdal, D., & Rosenberger, C. (2019). Statistical modeling of keystroke dynamics samples for the generation of synthetic datasets. *Future Generation Computer Systems*, *100*, 907–920. https://doi.org/10.1016/j.future.2019.03.056.

8. Revett, K., Gorunescu, F., Gorunescu, M., Ene, M., Tenreiro de Magalhães, S., & Santos, H. (2007). A machine learning approach to keystroke dynamics based user authentication. *International Journal of Electronic Security and Digital Forensics, 1*(1), 55–70. https://www.researchgate.net/publication/216864662_A_machine_learning_approach_to_keystroke_dynamics_based_user_authentication.

9. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press. https://www.deeplearningbook.org.

10. Lu Xiaofeng, Zhang Shengfei, Yi Shengwei: Continuous authentication by free-text keystroke based on CNN plus RNN (2018). https://www.sciencedirect.com/science/article/pii/S1877050919302935.

11. Elgallad, E. A., Ouarda, W., & M., A. (2019). Dense hand-cnn: A novel CNN architecture based on later fusion of neural and wavelet features for identity recognition. *International Journal of Advanced Computer Science and Applications*, *10*(6). https://doi.org/10.14569/ijacsa.2019.0100647

12. Harun, N., Dlay, S. S., & Woo, W. L. (2010). Performance of keystroke biometrics authentication system using Multilayer Perceptron neural network (MLP NN). У *2010 7th international symposium on communication systems, networks & digital signal processing (CSNDSP 2010)*. IEEE. https://doi.org/10.1109/csndsp16145.2010.5580334.

13. Monrose, F., & Rubin, A. D. (2000). Keystroke dynamics as a biometric for authentication. *Future Generation Computer Systems*, *16*(4), 351–359. https://doi.org/10.1016/s0167-739x(99)00059-x.

References

1. Shadman, R., Wahab, A. A., Manno, M., Lukaszewski, M., Hou, D., & Hussain, F. (2025). Keystroke dynamics: Concepts, techniques, and applications. *ACM Computing Surveys*. https://www.semanticscholar.org/reader/826113b73fa96eb1d6be90e66ae688c06d086109.

2. Shanmugapriya, D., & Padmavathi, G. (2010). A Survey of Biometric Keystroke Dynamics: Approaches, *International Journal of Computer Science and Information Security,* 5(1) https://www.researchgate.net/publication/41171720_A_Survey_of_Biometric_keystroke_Dynamics_Approaches_Security_and_Challenges.

3. Rashik Shadman Ahmed Anu Wahab Michael Manno Matthew Lukaszewski Daqing Hou Faraz Hussain (2024). *Keystroke Dynamics: Concepts, Techniques, and Applications* https://arxiv.org/pdf/2303.04605.

4. Teh, P. S., Teoh, A. B. J., & Yue, S. (2013). A survey of keystroke dynamics biometrics. *The Scientific World Journal*, *2013*, 1–24. https://doi.org/10.1155/2013/408280.

5. Puvirajasingam, K., & Sangeetha, D. (2013). Biometric Authentication using Keystroke Dynamics: A Survey. *IJREAT International Journal of Research in Engineering & Advanced Technology*, 1(5). http://ijreat.org/Papers%202013/Issue5/IJREATV1I5079.pdf.

6. Gaines, R. & Lisowski, William & Press, S. & Shapiro, Norman. (1980). Authentication by Keystroke Timing: Some Preliminary Results. 52.

7. Migdal, D., & Rosenberger, C. (2019). Statistical modeling of keystroke dynamics samples for the generation of synthetic datasets. *Future Generation Computer Systems*, *100*, 907–920. https://doi.org/10.1016/j.future.2019.03.056.

8. Revett, K., Gorunescu, F., Gorunescu, M., Ene, M., Tenreiro de Magalhães, S., & Santos, H. (2007). A machine learning approach to keystroke dynamics based user authentication. *International Journal of Electronic Security and Digital Forensics, 1*(1), 55–70. https://www.researchgate.net/publication/216864662_A_machine_learning_approach_to_keystroke_dynamics_based_user_authentication.

9. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press. https://www.deeplearningbook.org.

10. Lu Xiaofeng, Zhang Shengfei, Yi Shengwei: Continuous authentication by free-text keystroke based on CNN plus RNN (2018). https://www.sciencedirect.com/science/article/pii/S1877050919302935.

11. Elgallad, E. A., Ouarda, W., & M., A. (2019). Dense hand-cnn: A novel CNN architecture based on later fusion of neural and wavelet features for identity recognition. *International Journal of Advanced Computer Science and Applications*, *10*(6). https://doi.org/10.14569/ijacsa.2019.0100647

12. Harun, N., Dlay, S. S., & Woo, W. L. (2010). Performance of keystroke biometrics authentication system using Multilayer Perceptron neural network (MLP NN). У *2010 7th international symposium on communication systems, networks & digital signal processing (CSNDSP 2010)*. IEEE. https://doi.org/10.1109/csndsp16145.2010.5580334.

13. Monrose, F., & Rubin, A. D. (2000). Keystroke dynamics as a biometric for authentication. *Future Generation Computer Systems*, *16*(4), 351–359. https://doi.org/10.1016/s0167-739x(99)00059-x.

Отримано 08.04.2025

UDC 004.03:004.9

Sergiy Kovalenko[1], Oleksii Krasnozhon[2]

[1]PhD student of the Department of Information and Computer Systems, group ASD-122-23
Chernihiv Polytechnic National University (Chernihiv, Ukraine)
E-mail: kovalenkoptp10@gmail.com. **ORCID:** https://orcid.org/0009-0003-6993-010X

[2]PhD in Technical Sciences, associate professor of the Department of Information and Computer Systems
Chernihiv Polytechnic National University (Chernihiv, Ukraine)
E-mail: wingcommander2011@gmail.com. **ORCID:** https://orcid.org/0000-0003-2500-254X
ResearcherID: G-4623-2014. **Scopus Author ID:** 57190377188

STUDY OF PRIMARY DATA PROCESSING METHODS IN USER IDENTIFICATION SYSTEMS BASED ON KEYSTROKE DYNAMICS

Solving the problem of user identification of computer systems and networks is very relevant, as it allows to increase the level of security, especially when accessing critical data. There are both traditional identification methods (passwords, biometrics) and more modern ones.

This review article examines existing methods of user identification based on the dynamics of keystrokes. In particular, such methods as dynamic analysis of keystroke times, statistical models for recognizing printing patterns, and neural network approaches to user identification are considered.

Dynamic analysis of keystroke times is based on measuring the time characteristics of printing, including the average interval between keystrokes and the duration of holding down keys. These parameters are unique for each user and can be used for authentication.

Statistical models for recognizing printing patterns analyze sequences of characters typed at a certain speed and in a certain order. These models help to detect individual text input patterns, which helps to verify users.

Approaches based on the use of artificial neural networks use deep learning to process large sets of statistical data on keystrokes. These methods allow the development of complex, multi-layered models for analyzing behavioral characteristics of typing, increasing the accuracy of identification.

Thus, we can conclude that none of the methods considered in the article is exhaustive and cannot guarantee 100% correct user identification results. Only by integrating dynamic analysis, statistical modeling and neural network algorithms can we significantly improve authentication security, reducing the risk of unauthorized access and ensuring reliable user identification in information systems.

Key words*: keystroke dynamics, biometric authentication, user identification, dynamic timing analysis, statistical recognition models, neural network algorithms, behavioral authentication, machine learning, information security, primary data processing.*

Figures: 1. References: 13.

Коваленко С. В., Красножон О. В. Методи первинної обробки даних у системах ідентифікації користувачів на основі клавіатурного почерка. *Технічні науки та технології.* 2025. № 2(40). С. 294-302. DOI: https://doi.org/10.25140/2411-5363-2025-2(40)-294-302.