

ДИНАМІЧНА АУТЕНТИФІКАЦІЯ ЗА КЛАВІАТУРНИМ ПОЧЕРКОМ З ВИКОРИСТАННЯМ КРИТЕРІЮ МАННА-ВІТНІ

Маслій Д.В.

Науковий керівник – к.т.н., доц. Горелов Д.Ю.

Харківський національний університет радіоелектроніки
(61166, Харків, пр. Науки, 14, каф. Комп'ютерної радіоінженерії
і систем технічного захисту інформації (КРiCTЗІ),
тел. (057) 702-13-06, email: masliy93@gmail.com.

In work the algorithm of keystroke authentication based on the Mann-Whitney U -test is developed. The algorithms of formation of the user profile and its authentication has been designed.

Аналіз літератури в області біометричних систем контролю доступу за клавіатурним почерком показує, що найбільш поширеними є методи класифікації клавіатурного почерку на основі параметричних статистичних підходів. Ці методи порівняння параметрів розподілів припускають, що дослідник заздалегідь володіє фундаментальною інформацією – йому відомий вид закону розподілу ймовірностей, найчастіше нормальний закон, що дозволяє звести задачу розпізнавання до перевірки гіпотез про подібність таких характеристик як середнє, медіана і стандартне відхилення. У тих випадках, коли припущення про гіпотетичний закон розподілу ймовірностей не є переконливими, слід застосовувати непараметричні статистичні методи.

Для нівелювання впливу ергономіки клавіатури було прийнято в якості інформаційних характеристик клавіатурного почерку використовувати чотири часові інтервали: T_1 – тривалість натискання 20 найважливіших букв англійської мови E, A, R, I, O, T, N, S, H, D, L, C, U, M, W, F, G, Y, P, B; T_2 – час паузи в 20 найважливіших біграм англійської мови IN, TH, TI, ON, AN, HE, AT, ER, RE, ND, HA, EN, TO, IT, OU, EA, HI, IS, OR, TE; T_3 – тривалість в 20 найважливіших біграм англійської мови IN, TH, TI, ON, AN, HE, AT, ER, RE, ND, HA, EN, TO, IT, OU, EA, HI, IS, OR, TE; T_4 – тривалість 20 вживаних слів англійської мови FOR, AND, THE, IS, IT, YOU, HAVE, OF, BE, TO, THAT, HE, SHE, THIS, THEY, WILL, I, ALL, A, HIM.

В експерименті приймали участь клавіатурні почерки 4-х чоловік. Для кожного користувача за результатами введеного тексту № 1 було сформовано по 20 векторів параметрів $\vec{T}_1 \div \vec{T}_4$. Цю сукупність з 80-ти векторів можна вважати початковим біоеталоном користувача:

$$ET = \left\{ \begin{array}{l} \vec{T}_1^E, \vec{T}_1^A, \vec{T}_1^R, \dots, \vec{T}_1^B \\ \vec{T}_2^{IN}, \vec{T}_2^{TH}, \vec{T}_2^{TI}, \dots, \vec{T}_2^{TE} \\ \vec{T}_3^{IN}, \vec{T}_3^{TH}, \vec{T}_3^{TI}, \dots, \vec{T}_3^{TE} \\ \vec{T}_4^{FOR}, \vec{T}_4^{AND}, \vec{T}_4^{THE}, \dots, \vec{T}_4^{HIM} \end{array} \right\}. \quad (1)$$

Далі за результатами введеного тексту № 2 для кожного користувача за допомогою непараметричного критерію Манна-Вітні для кожного з 20-ти векторів множини $\{\vec{T}_1^E, \vec{T}_1^A, \vec{T}_1^R, \dots, \vec{T}_1^B\}$, що відповідає часовому параметру T_1 , було перевірено гіпотезу про відсутність відмінностей між еталонним вектором \vec{T}_1^X та дослідним $\vec{T}_1^{X_{досл}}$. Якщо більше 12 і більше (60 % і більше) гіпотез з 20 було прийнято, то вважалось, що множина $\{\vec{T}_1^E, \vec{T}_1^A, \vec{T}_1^R, \dots, \vec{T}_1^B\}$ є біоеталоном. Якщо кількість прийнятих гіпотез було меншою 12, то користувачеві пропонувалось знову пройти процес формування біоеталону. Для формування повного біоеталону аналогічні розрахунки було проведено для множин $\{\vec{T}_2^{IN}, \vec{T}_2^{TH}, \vec{T}_2^{TI}, \dots, \vec{T}_2^{TE}\}$, $\{\vec{T}_3^{IN}, \vec{T}_3^{TH}, \vec{T}_3^{TI}, \dots, \vec{T}_3^{TE}\}$ та $\{\vec{T}_4^{FOR}, \vec{T}_4^{AND}, \vec{T}_4^{THE}, \dots, \vec{T}_4^{HIM}\}$.

На *першому* кроці аутентифікації дослідного користувача розраховувались 80 векторів:

$$ET = \left\{ \begin{array}{l} \vec{T}_{1exp}^E, \vec{T}_{1exp}^A, \vec{T}_{1exp}^R, \dots, \vec{T}_{1exp}^B \\ \vec{T}_{2exp}^{IN}, \vec{T}_{2exp}^{TH}, \vec{T}_{2exp}^{TI}, \dots, \vec{T}_{2exp}^{TE} \\ \vec{T}_{3exp}^{IN}, \vec{T}_{3exp}^{TH}, \vec{T}_{3exp}^{TI}, \dots, \vec{T}_{3exp}^{TE} \\ \vec{T}_{4exp}^{FOR}, \vec{T}_{4exp}^{AND}, \vec{T}_{4exp}^{THE}, \dots, \vec{T}_{4exp}^{HIM} \end{array} \right\}. \quad (2)$$

На *другому* кроці аутентифікації за допомогою непараметричного критерію Манна-Вітні для кожного параметру T_i перевіряються 20 гіпотез про відсутність відмінностей між векторами $\vec{T}_i^{X_{exp}}$ та біоеталоном:

$$H_0^{iX}: \vec{T}_{iexp}^X = \vec{T}_i^X. \quad (3)$$

Якщо більше 12 і більше з 20 гіпотез було прийнято, то приймалось загальне рішення про відсутність відмінностей між вектором аутентифікації та еталоном за параметром T_i .

На *третьому* кроці аутентифікації приймалось загальне рішення про аутентифікацію: якщо для трьох і більше з чотирьох параметрів T_i прийнято позитивні рішення про відсутність відмінностей, то висувається рішення про позитивну аутентифікацію. В іншому випадку система приймає рішення про негативну аутентифікацію.

За результатами проведених досліджень загальна помилка надання системою доступу неавторизованому користувачеві та заборони доступу авторизованому користувачеві склала 12 %.

Перелік джерел посилання: 1. Aliksieiev Vasyi, Elena Sharapova, Olena Ivanova, Gorelov Denis, Synytsia Yuliia. Web-Based Application to Collect and Analyze Users Data for Keystroke Biometric Authentication. In Proceedings of the First IEEE Ukraine Conference on Electrical and Computer Engineering (UKRCON). Pages 917-922, 2017. 2. Vasyi Aliksieiev, Aleksey Strelnitskiy, Dmitry Gavva, Denis Gorelov, Yuliia Synytsia. Studying of keystroke dynamics statistical properties for biometric user authentication. Proceedings of 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET), Pages 559-563, 2018.