

УДК 004.931+004.04+004.056

**ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ ІДЕНТИФІКАЦІЇ ОСОБИ
ЗА КЛАВІАТУРНИМ ПОЧЕРКОМ ДЛЯ СЦЕНАРІЮ
ВИКОРИСТАННЯ БІЛІНГВАЛЬНИХ ПАРОЛІВ. ЧАСТИНА 2**

Пономаренко В.О., Кривцун С.Ю.

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

Харківський національний університет радіоелектроніки, каф. КРiCTЗi
студентський науковий гурток «Біометричні технології контролю доступу»
e-mail: serhii.kryvtsun@nure.ua

Using the "Bilingual Keystroke Dynamics Dataset" database and the Orange software, a study of the influence of a set of passphrases in native and non-native languages on the accuracy of keystroke identification was carried out. It has been experimentally confirmed that the transition to two-class classification allows not only to significantly increase the accuracy of identification, but also illustrates the advantage of using passwords entered in English over passwords entered in their native language for users with unformed keystroke.

Результати мультикласової класифікації користувачів датасетів «BKSD Phrase-English» та «BKSD Phrase-Arabic» наведено на рис. 5 та рис. 6. Як можна бачити для англійської мови лише для 5-ти користувачів (5 %) точність класифікації є високою ($FRR < 1\%$) і для 14-ти користувачів (40 %) точність класифікації є середньою ($1\% < FRR < 3\%$). В той же час для арабської мови висока точність класифікації спостерігається для 8-ми користувачів (23 %) і для 14-ти користувачів (40 %) точність класифікації є середньою. Але, якщо прийнятний рівень FRR обрати рівним 5 % (в 5 спробах зі 100 зареєстрованого користувача буде ідентифіковано як зловмисника), то картина змінюється: 28 користувачів (80 %) для датасету «BKSD Phrase-English» і 27 користувачів (77 %) для датасету «BKSD Phrase-Arabic». Отже, набирання паролю рідною мовою не підвищує точність ідентифікації.

Аналізуючи рис. 6, можна зробити наступні висновки: 1) в датасетах присутні користувачі з добре сформованим клавіатурним почерком на арабській мові – користувачі S67, S81 та S85 – жовтий колір; 2) в датасетах присутні користувачі з добре сформованим клавіатурним почерком на англійській мові – користувачі S73, S76 та S84 – синій колір; 3) в датасетах присутні користувачі з добре сформованим клавіатурним почерком на обох мовах – користувачі S75 та S79 – помаранчевий колір; 4) в датасетах присутні користувачі з майже не сформованим клавіатурним почерком на обох мовах – користувачі S70, S102 та S105 – червоний колір.

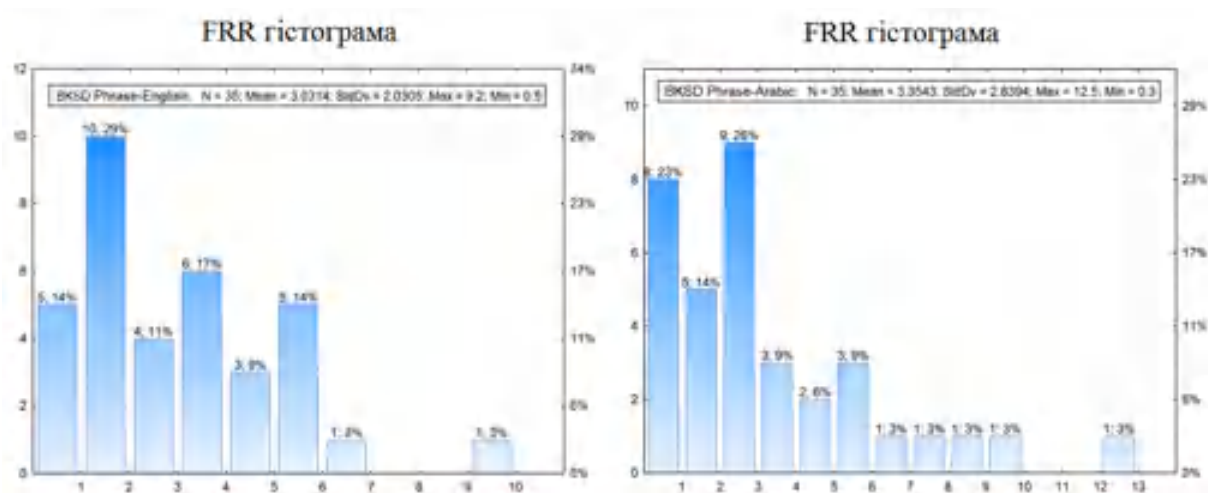


Рисунок 5

User	FRRen, %	FRRar, %	User	FRRen, %	FRRar, %
S65	3.8	3	S86	4.4	8.2
S66	5.1	2.3	S87	1.8	0.8
S67	1.8	0.5	S88	1.3	1
S68	2	1.5	S107	1.3	5.5
S69	5.1	4.1	S108	2.8	5
S70	9.2	5.1	S111	5.2	3.9
S73	0.5	1.5	S112	3.5	0.8
S74	4	2.6	S122	4.5	3
S75	0.5	0.8	s93	3.5	1.8
S76	0.8	2	s94	1.5	6.1
S78	1.8	2.1	s95	1.2	12.5
S79	0.8	0.3	s96	1.5	2.8
S80	5.2	3.7	s100	3.5	8
S81	1.3	0.8	s101	2.3	2.3
S82	2.1	3	s102	6.8	5.5
S83	4.3	3.3	s103	3.8	1.2
S84	0.8	2.3	s105	6	9.8
S85	2.1	0.3			

Рисунок 6

Підвищити точність ідентифікації можна за рахунок переходу до двокласової класифікації, тобто сценарію, на комп'ютері зареєстровано лише одного користувача, а всі інші автоматично відносяться до класу «зловмисник». Для експерименту було відібрані користувачі, що було позначено кольорами на рис. 6.

Як можна бачити, кількість зелених комірок, тобто випадків, коли FRR та FAR менше 1-го відсотка (високий рівень точності) для обох датасетів однакова – 52 або 94.5 %. При цьому точність класифікації користувачів з добре сформованим клявіатурним почерком (жовтий, синій та помаранчевий кольори) однаково висока для обох сценаріїв набору

парольної фрази.

FRR/FAR, %	S70	S73	S75	S76	S79	S81	S84	S85	S102	S105	
S67	0.3/0.0	0.0/0.3	0.3/0.0	0.0/0.0	0.0/0.3	0.0/0.0	0.0/0.0	0.0/0.0	0.8/2.0	1.5/0.8	
S70		0.0/0.0	0.5/0.3	0.0/0.0	0.0/0.0	0.0/0.0	0.0/0.3	0.0/0.0	0.0/0.2	0.0/0.5	
S73			0.0/0.0	0.0/0.0	0.3/0.3	0.0/0.0	0.0/0.0	0.3/0.3	0.5/0.0	0.5/0.5	
S75				0.0/0.3	0.0/0.0	0.0/0.0	0.0/0.0	0.0/0.0	0.5/0.5	0.3/0.3	
S76					0.0/0.0	0.0/0.0	0.0/0.0	0.0/0.0	0.0/0.2	0.0/0.0	
S79						0.8/0.0	0.0/0.0	0.0/0.0	0.5/0.2	0.8/0.3	
S81							0.0/0.0	0.0/0.0	0.0/1.2	0.7/1.3	
S84		BKSD Phrase-English							0.0/0.0	0.0/0.0	0.0/0.3
S85									0.5/0.7	0.5/0.8	
S102										0.0/0.2	

FRR/FAR, %	S70	S73	S75	S76	S79	S81	S84	S85	S102	S105	
S67	0.0/0.3	0.0/0.5	0.0/0.0	0.0/0.3	0.0/0.0	0.0/0.0	0.0/0.0	0.0/0.0	0.3/0.2	0.0/0.0	
S70		0.0/0.3	0.0/0.0	0.3/0.3	0.3/0.0	0.0/0.0	0.0/0.3	0.3/0.0	0.0/0.05	0.5/1.0	
S73			0.3/0.0	0.0/0.0	0.0/0.0	0.0/0.0	0.3/0.0	0.0/0.0	0.3/0.2	0.3/0.0	
S75				0.0/0.3	0.0/0.0	0.0/0.0	0.0/0.0	0.0/0.0	0.0/0.0	0.0/0.0	
S76					0.0/0.0	0.3/0.0	0.3/0.3	0.0/0.3	0.3/0.2	0.3/0.0	
S79						0.0/0.0	0.3/0.0	0.0/0.3	0.3/0.0	0.5/0.0	
S81							0.3/0.5	0.0/0.0	1.8/1.5	0.3/1.0	
S84		BKSD Phrase-Arabic							0.0/0.0	0.5/1.5	0.3/0.8
S85									0.0/0.0	0.0/0.3	
S102										0.0/0.8	

Рисунок 7

В той же час точність класифікації користувачів з не сформованим клавіатурним почерком для сценарію набору парольної фрази рідною мовою вище за рахунок пари S102-S105. Для підтвердження або спростування цієї тези було проведемо додаткові дослідження двокласової класифікації подібних користувачів – на рис. 6 вони позначені сірим кольором – див. рис. 8. Як можна бачити, кількість зелених комірок, тобто випадків, коли FRR та FAR менше 1-го відсотка (високий рівень точності) більша для датасету «BKSD Phrase-English». Крім того для даного датасету немає жодної пари користувачів з рівнем FRR / FAR більшим 3 %. В той же час для датасету «BKSD Phrase-Arabic» таких пар дві (S83-S86 та S100-S122). Отже, для користувачів з несформованим клавіатурним почерком для підвищення точності ідентифікації можна рекомендувати використовувати парольні фрази на англійській мові.

Останнім кроком в дослідженнях є оцінка точності ідентифікації за умови використання в парольній фразі слів на обох мовах. Для цього з датасету «BKSD Password-Arabic» було взято 10 стовпчиків, що описують часові параметри вводу слова «актор» арабською, а з датасету «BKSD Password-English» також було взято 10 стовпчиків, що описують часові параметри вводу слова «бандит» англійською.

FRR/FAR, %	S66	S69	S83	S86	S95	S100	S111	S122
S65	0,3/0,0	0,0/0,0	0,5/0,8	0,5/1,3	0,0/0,0	0,5/0,3	0,0/0,0	0,5/0,8
S66		1,5/1,5	0,0/0,3	0,3/0,0	0,0/0,0	0,8/0,0	0,5/0,3	0,0/0,5
S69			0,3/0,0	0,0/0,0	0,5/0,0	0,0/0,0	1,8/1,3	0,0/0,0
S83				2,6/1,8	0,0/0,0	1,3/1,0	0,0/0,0	0,0/0,0
S86					0,0/0,0	0,3/1,0	0,0/0,0	0,0/0,0
S95		BKSD Phrase-English				0,0/0,0	0,8/1,0	0,0/0,0
S100							0,3/0,0	0,3/0,5
S111								0,0/0,5

FRR/FAR, %	S66	S69	S83	S86	S95	S100	S111	S122
S65	0,2/0,5	1,8/1,3	0,7/0,8	0,5/0,8	0,0/0,3	0,5/0,5	0,2/0,5	0,2/0,5
S66		0,3/1,0	0,0/0,3	0,0/0,3	0,8/0,5	0,0/0,0	0,0/0,8	0,0/0,3
S69			0,0/0,3	0,0/0,3	1,5/0,8	0,3/0,3	0,0/1,0	0,0/1,5
S83				0,0/5,1	0,0/0,0	0,0/1,0	0,0/0,3	0,3/0,8
S86					0,3/0,3	2,8/2,0	0,0/0,0	1,0/1,3
S95		BKSD Phrase-Arabic				0,3/0,0	0,0/0,3	0,3/0,3
S100								0,0/0,0
S111								0,0/0,0

Рисунок 8

Інтегральна помилка першого роду FRR становить:

$$FRR = 1 - Recall = 1 - 0.943 = 5.7\%$$

Інтегральна помилка другого роду FAR становить:

$$FAR = 1 - Specificity = 1 - 0.999 = 0.1\%$$

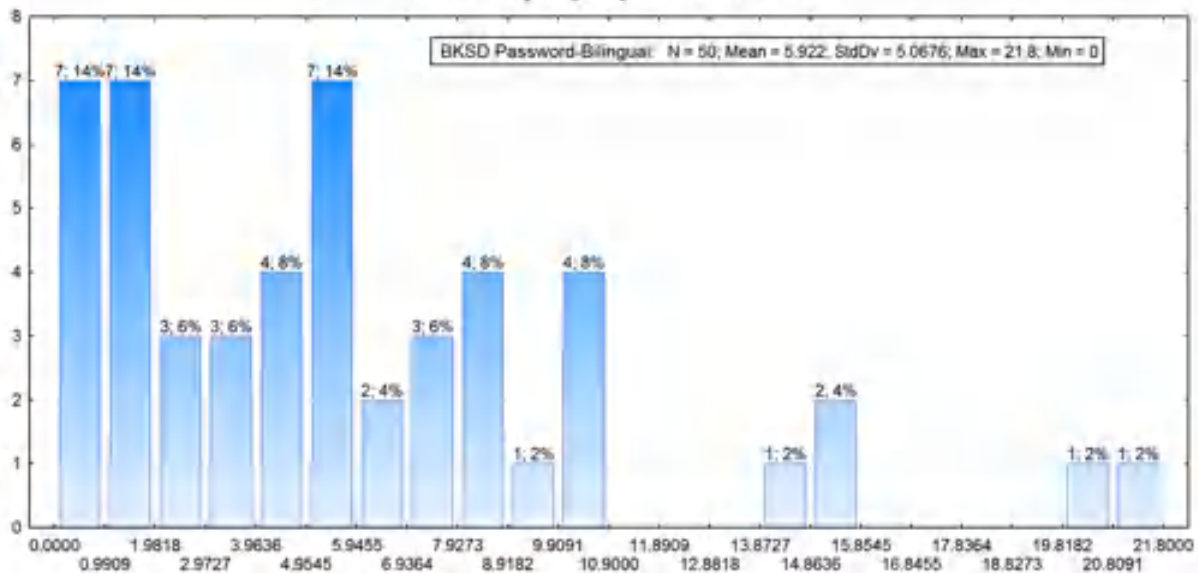


Рисунок 9

Як можна бачити з рис. 9, кількість користувачів, для яких точність класифікації зафіксована рівнем $FRR \leq 5\%$ дорівнює 25 або 50% від загальної кількості, що значно краще ніж результати мультикласової класифікації за пароліними фразами. Отже, перехід до білінгвальних паролів має сенс для підвищення точності ідентифікації у разі використання коротких пароліних фраз.