

Д.Ю. ГОРЕЛОВ, канд. техн. наук, О.О. ІВАНОВА, канд. техн. наук,
О.В. КОКОРИН, Д.В. МАСЛІЙ, О.В. ЛИТВИНЕНКО

ДОСЛІДЖЕННЯ ІНФОРМАТИВНИХ ПАРАМЕТРІВ ДИГРАФІВ КЛАВІАТУРНОГО ПОЧЕРКУ ДЛЯ ЗАДАЧ ІДЕНТИФІКАЦІЇ КОРИСТУВАЧІВ КОМП'ЮТЕРНИХ МЕРЕЖ

Одним з головних факторів, що визначають стан захищеності комп'ютерних інформаційних систем, є ефективність функціонування системи управління доступом, важливим елементом якої є захист від несанкціонованого доступу. Цей елемент захисту виконує свої функції за допомогою процедур ідентифікації і аутентифікації користувачів. Методи біометричної аутентифікації з огляду на невід'ємність біометричних характеристик від конкретної людини, неможливість відмови від авторства, точність і зручність здатні забезпечити більш надійну у порівнянні з іншими засобами перевірку особистості користувачів комп'ютерних мереж.

Важливе місце серед біометричних систем займають пристрої і програми, побудовані на аналізі динамічних параметрів особистості – аутентифікація за клавіатурним почерком та за динамікою курсору комп'ютерної миші. Враховуючи той факт, що стан людини постійно змінюється через різні зовнішні та внутрішні фактори впливу, що виникають в результаті втоми, стресів, хворобливих станів, нездужань тощо, клавіатурний почерк користувача безперервно змінюється. Відповідно, визначивши найбільш стабільні характеристики клавіатурного почерку, можна значно підвищити точність ідентифікації та аутентифікації користувачів комп'ютерних мереж.

1. Інформативні ознаки клавіатурного почерку

У задачі динамічної біометричної аутентифікації користувача за клавіатурним почерком основним етапом є аналіз і обробка первинних даних. Після даної операції вхідний потік інформації поділяється на характеристики, які відображають ті чи інші динамічні ознаки користувача, який проходить аутентифікацію. Далі дані ознаки дозволяють отримати ряд унікальних характеристик користувача.

Наразі можна виділити три класи інформативних параметрів клавіатурного почерку [1, 2].

1. Часові характеристики поодиноких подій клавіатури, наприклад, абсолютна тривалість утримання клавіші, абсолютна пауза перед клавішею, абсолютна пауза після клавіші, відношення величини паузи перед клавішею до тривалості утримання клавіші тощо.

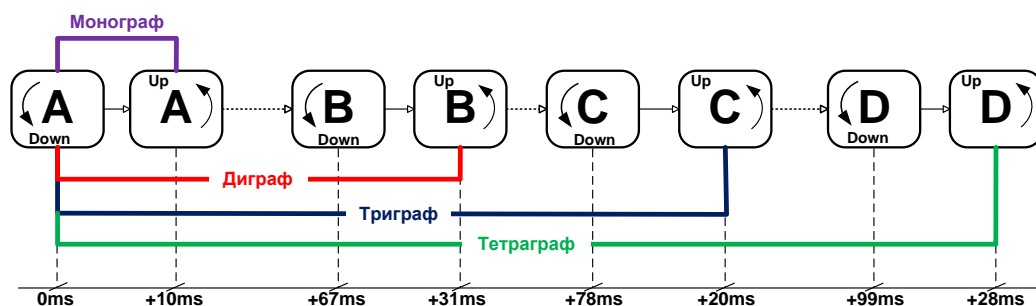


Рис. 1. N-графи клавіатури

2. Часові характеристики послідовних подій клавіатури (на рис. 1 комбінація з двох клавіш утворює диграф, комбінація з трьох клавіш – триграф, n послідовних подій клавіатури – n -граф), наприклад, абсолютне значення та розподіл паузи між відпусканням першої клавіші

та натисканням другої клавіші всіх диграфів у заданому тексті, абсолютне значення та розподіл часу між натисканням клавіш всіх диграфів у заданому тексті, абсолютне значення та розподіл часу між відпусканням клавіш всіх диграфів у заданому тексті тощо.

3. Інтегральні характеристики набору тексту, наприклад, швидкість набору символів, швидкість набору слів, ступінь аритмічності набору, кількість виправлень, кількість та особливості перекриття часів натискання клавіш, розподіл частот використання клавіш зміни регістра, пропуск певних букв у певних буквосполученнях/словах, друкарські помилки певних букв у певних буквосполученнях та словах тощо.

Клас інтегральних інформативних характеристик клавіатурного почерку в поєднанні з будь-яким з перших двох класів дає максимальну точність, проте вимагає значних витрат на розробку, впровадження та підтримку подібних систем.

Слід відзначити, що більшість авторів [3] серед основних факторів, що впливають на точність ідентифікації, відзначає наступні: основний алгоритм класифікації користувачів, кількість учасників експерименту з різною величиною досвіду роботи з клавіатурою, спосіб і організація введення даних і апаратна платформа, на базі якої проводиться тестування системи аутентифікації. При цьому такий найважливіший фактор, як інформативні ознаки клавіатурного почерку, найчастіше не враховується. В переважній більшості робіт просто використовуються абсолютні значення часів натискань окремих клавіш, а також пауз перед і після натискання. Тому відповідно до мети досліджень були проаналізовані наступні 19 інформативних ознак клавіатурного почерку, що включають в себе як параметри монографів, так і диграфів клавіатури (рис. 2):

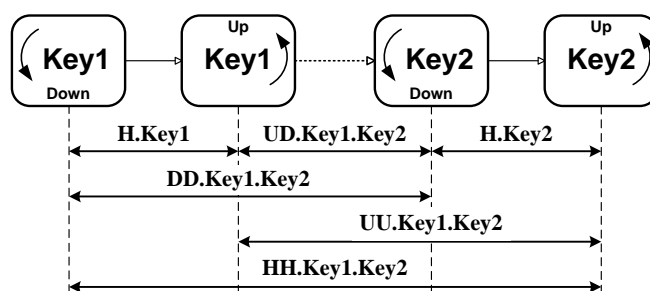


Рис. 2. Часові інтервали диграфів клавіатури

- 1) тривалість натискань клавіш «**H.Key**»;
- 2) тривалість пауз між натисканнями клавіш «**UD.Key1.Key2**»;
- 3) тривалість між натисканнями клавіш у диграфах «**DD.Key1.Key2**»;
- 4) тривалість між відпусканнями клавіш у диграфах «**UU.Key1.Key2**»;
- 5) тривалість диграфів «**HH.Key1.Key2**»;
- 6) відношення тривалості натискання першої клавіші диграфу до тривалості паузи між натисканнями «**H.Key1/UD.Key1.Key2**»;
- 7) відношення тривалості натискання першої клавіші диграфу до часу між натисканнями «**H.Key1/DD.Key1.Key2**»;
- 8) відношення тривалості натискання першої клавіші диграфу до часу між відпусканнями «**H.Key1/UU.Key1.Key2**»;
- 9) відношення тривалості натискання першої клавіші диграфу до тривалості диграфу «**H.Key1/HH.Key1.Key2**»;
- 10) відношення тривалості паузи до часу між натисканнями «**UD.Key1.Key2/DD.Key1.Key2**»;
- 11) відношення тривалості паузи до часу між відпусканнями «**UD.Key1.Key2/UU.Key1.Key2**»;
- 12) відношення тривалості паузи до тривалості диграфу «**UD.Key1.Key2/HH.Key1.Key2**»;
- 13) відношення тривалості натискання другої клавіші диграфу до часу паузи між натисканнями «**H.Key2/UD.Key1.Key2**»;
- 14) відношення тривалості натискання другої клавіші диграфу до часу між натисканнями «**H.Key2/DD.Key1.Key2**»;
- 15) відношення тривалості натискання другої клавіші диграфу до часу між відпусканнями «**H.Key2/UU.Key1.Key2**»;

- 16) відношення тривалості натискання другої клавіші диграфу до тривалості диграфу «H.Key2/HH.Key1.Key2»;
- 17) відношення часу між натисканнями клавіш у диграфі до часу між відпусканнями клавіш у диграфі «DD.Key1.Key2/UU.Key1.Key2»;
- 18) відношення часу між натисканнями клавіш у диграфі до тривалості диграфу «DD.Key1.Key2/ HH.Key1.Key2»;
- 19) відношення часу між відпусканнями клавіш у диграфі до тривалості диграфу «UU.Key1.Key2/ HH.Key1.Key2».

2. Keystroke Dynamics Benchmark Data Set

База даних «Keystroke Dynamics Benchmark Data Set» [4] опублікована у 2009 р. Датасет містить часові параметри вводу паролльної фрази «.tie5Roanl», яку набирали 51 користувач по 400 раз кожен. Дані представляють файл Excel format з 34 стовпчиками (рис. 3), три з яких це ідентифікатор користувача, номер сесії (від 1 до 8) та номер спроби (від 1 до 50) вводу. Останні 31 стовпчик – часові параметри (в секундах) вводу паролльної фрази (див. рис. 2): H.dot, DD.dot.t, UD.dot.t, H.t, DD.t.i, UD.t.i, H.i, DD.i.e, UD.i.e, H.e, DD.e.five, UD.e.five, H.five, DD.five.Shift.r, UD.five.Shift.r, H.Shift.r, DD.Shift.r.o, UD.Shift.r.o, H.o, DD.o.a, UD.o.a, H.a, DD.a.n, UD.a.n, H.n, DD.n.l, UD.n.l, H.l, DD.l.enter, UD.l.enter, H.enter.

	A	B	C	D	E	F	G	H	I	J	K	L
	subject	sessionIndex	rep	H.period	DD.period.t	UD.period.t	H.t	DD.t.i	UD.t.i	H.i	DD.i.e	UD.i.e
1	s002	1	1	0.1491	0.3979	0.2488	0.1069	0.1674	0.0605	0.1169	0.2212	0.1043
2	s002	1	2	0.1111	0.3451	0.234	0.0694	0.1283	0.0569	0.0908	0.1357	0.0449
3	s002	1	3	0.1328	0.2072	0.0744	0.0731	0.1291	0.056	0.0821	0.1542	0.0721
4	s002	1	4	0.1291	0.2515	0.1224	0.1059	0.2495	0.1436	0.104	0.2038	0.0998
5	s002	1	5	0.1249	0.2317	0.1068	0.0895	0.1676	0.0781	0.0903	0.1589	0.0686
6	s002	1	6	0.1394	0.2343	0.0949	0.0813	0.1299	0.0486	0.0744	0.1412	0.0668
7	s002	1	7	0.1064	0.2069	0.1005	0.0866	0.1368	0.0502	0.08	0.1407	0.0607
8	s002	1	8	0.0929	0.181	0.0881	0.0818	0.1378	0.056	0.0747	0.1367	0.062
9	s002	1	9	0.0966	0.1797	0.0831	0.0771	0.1296	0.0525	0.0839	0.1425	0.0586
10	s002	1	10	0.1093	0.1807	0.0714	0.0731	0.1457	0.0726	0.0766	0.1241	0.0475
11	s002	1	11	0.0887	0.166	0.0773	0.0876	0.156	0.0684	0.0839	0.1386	0.0547
12	s002	1	12	0.0911	0.1525	0.0614	0.0824	0.1516	0.0692	0.0731	0.1391	0.066
13	s002	1	13	0.1114	0.162	0.0506	0.09	0.1547	0.0647	0.0797	0.1349	0.0552
14	s002	1	14	0.0903	0.1871	0.0968	0.0805	0.1919	0.1114	0.0842	0.16	0.0758
15	s002	1	15	0.1169	0.2562	0.1393	0.0739	0.1549	0.081	0.0892	0.1462	0.057
16	s002	1	16	0.127	0.1839	0.0569	0.0911	0.1381	0.047	0.0895	0.1774	0.0879
17	s002	1	17	0.1016	0.1799	0.0783	0.0792	0.1434	0.0642	0.076	0.1412	0.0652
18	s002	1	18	0.1056	0.1755	0.0699	0.0781	0.1391	0.061	0.0898	0.1613	0.0715
19	s002	1	19	0.1177	0.2237	0.106	0.0837	0.188	0.1043	0.0919	0.1803	0.0884
20	s002	1	20	0.1027	0.1781	0.0754	0.0729	0.1418	0.0689	0.0792	0.1544	0.0752
21	s002	1	21	0.1016	0.1374	0.0358	0.0861	0.1629	0.0768	0.0774	0.1521	0.0747
22	s002	1	22	0.1072	0.2217	0.1145	0.0726	0.1349	0.0623	0.0768	0.1716	0.0948
23	s002	1	23	0.1243	0.1841	0.0598	0.0768	0.1568	0.08	0.085	0.1539	0.0689
24	s002	1	24	0.1241	0.2019	0.0778	0.0829	0.1745	0.0916	0.0734	0.131	0.0576
25	s002	1	25	0.1098	0.1567	0.0469	0.0768	0.1642	0.0874	0.0605	0.1637	0.1032

Рис. 3. База даних «Keystroke Dynamics Benchmark Data Set» у форматі xlsx

Цей датасет було обрано для аналізу інформативності часових ознак клавіатурного почерку, по-перше, в першу чергу через його відкритість. По-друге, через рівномірність та повноту дослідних даних (зазвичай датасети містять нерівномірні дані – кількість голосних букв більша за кількість приголосних, відповідно, і статистика голосних більш точна, ніж статистика приголосних). По-третє, через повноту дослідних даних – кожен з 31-го інформативного параметру було виміряно 400 разів, отже можна стверджувати про їх статистичну сталість.

З огляду на 19 обраних для аналізу інформативних ознак клавіатурного почерку дослідний датасет було переформатовано у наступний формат (див. табл. 1).

3. Схема експерименту

Класифікація користувачів за часовими параметрами клавіатурного почерку проводилась у програмному середовищі Orange. Схему експерименту у Orange наведено на рис. 4. У якості алгоритму класифікації користувачів використовувався метод Random forests.

Точність класифікації перевірялась за вбудованим у віджет «Test and Score» алгоритмом 10-fold cross-validation.

Таблиця 1

Номер стовпчика	Характеристика
1	ID користувача
2-12	Тривалості натискань клавіш « H.Key »
13-22	Тривалості пауз між натисканнями клавіш « UD.Key1.Key2 »
23-32	Тривалості між натисканнями клавіш у диграфах « DD.Key1.Key2 »
33-42	Тривалості між відпусканнями клавіш у диграфах « UU.Key1.Key2 », $UU.Key1.Key2 = UD.Key1.Key2 + H.Key2$
43-52	Тривалості диграфів « HH.Key1.Key2 », $HH.Key1.Key2 = DD.Key1.Key2 + H.Key2$
53-62	Відношення тривалості натискання першої клавіші диграфу до паузи між натисканнями « H.Key1/UD.Key1.Key2 »
63-72	Відношення тривалості натискання першої клавіші диграфу до часу між натисканнями « H.Key1/DD.Key1.Key2 »
73-82	Відношення тривалості натискання першої клавіші диграфу до часу між відпусканнями « H.Key1/UU.Key1.Key2 »
83-92	Відношення тривалості натискання першої клавіші диграфу до тривалості диграфу « H.Key1/HH.Key1.Key2 »
93-102	Відношення тривалості паузи до часу між натисканнями « UD.Key1.Key2/DD.Key1.Key2 »
103-112	Відношення тривалості паузи до часу між відпусканнями « UD.Key1.Key2/UU.Key1.Key2 »
113-122	Відношення тривалості паузи до тривалості диграфу « UD.Key1.Key2/HH.Key1.Key2 »
123-132	Відношення тривалості натискання другої клавіші диграфу до часу паузи між натисканнями « H.Key2/UD.Key1.Key2 »
133-142	Відношення тривалості натискання другої клавіші диграфу до часу між натисканнями « H.Key2/DD.Key1.Key2 »
143-152	Відношення тривалості натискання другої клавіші диграфу до часу між відпусканнями « H.Key2/UU.Key1.Key2 »
153-162	Відношення тривалості натискання другої клавіші диграфу до тривалості диграфу « H.Key2/HH.Key1.Key2 »
163-172	Відношення часу між натисканнями клавіш у диграфі до часу між відпусканнями клавіш у диграфі « DD.Key1.Key2/UU.Key1.Key2 »
173-182	Відношення часу між натисканнями клавіш у диграфі до тривалості диграфу « DD.Key1.Key2/ HH.Key1.Key2 »
183-192	Відношення часу між відпусканнями клавіш у диграфі до тривалості диграфу « UU.Key1.Key2/ HH.Key1.Key2 »

Результати класифікації користувачів за первинними 31 інформативними ознаками наведено на рис. 5. Як можна побачити, інтегральна точність класифікації складає 92.3 % (поле «СА»). Якщо дослідити отримані результати по кожному з користувачів, то можна відмітити як стовідсоткову точність розпізнавання користувача «s036», так і дуже погану точність розпізнавання користувача «s032» – 71 %.

Оскільки в дослідженні потрібно максимально позбавитись впливу випадкових факторів, то було прийнято рішення відібрати 20 користувачів з дослідного датасету, для яких точність класифікації максимальна (тобто вони володіють унікальним, добре розпізнаваним клавіату-

рним почерком). Це користувачі «s036», «s022», «s043», «s052», «s055», «s017», «s024», «s053», «s028», «s033», «s019», «s010», «s011», «s042», «s044», «s005», «s049», «s056», «s039», «s018». На рис. 6, 7 наведено результати класифікації 20 вибраних користувачів за первинними 31 інформативними ознаками. Як можна побачити, інтегральна точність складає 95,9 %, точність класифікації по кожному користувачеві – не менш 87,8 %.

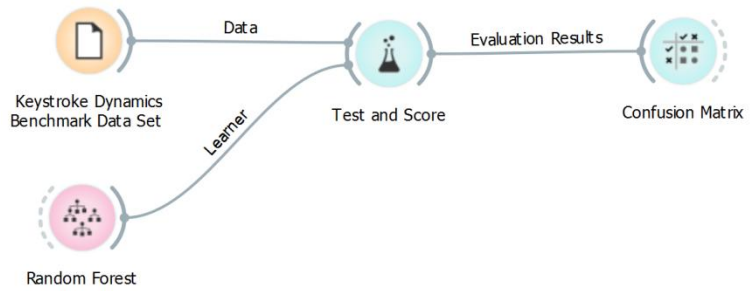


Рис. 4. Схема обчислювального експерименту у Orange

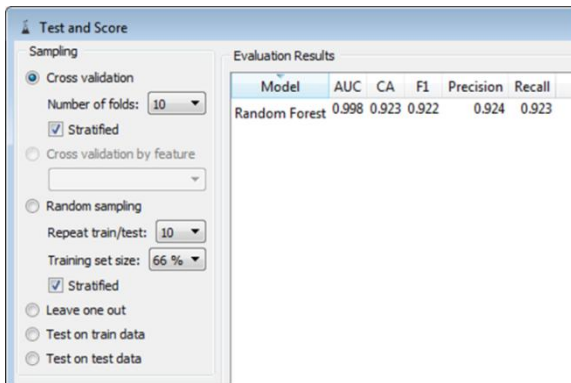


Рис. 5. Інтегральна точність класифікації 50 користувачів з дослідної бази даних за первинними 31 ознаками

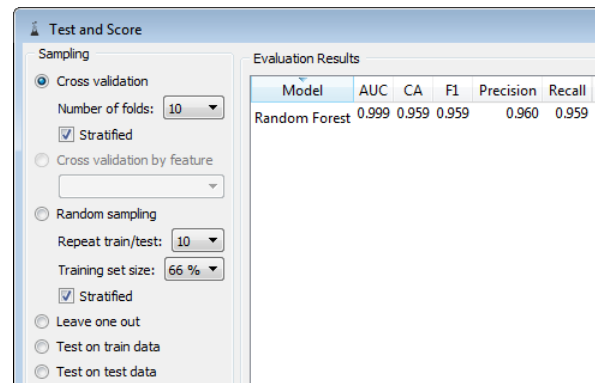


Рис. 6. Інтегральна точність класифікації 20 користувачів з дослідної бази даних за первинними 31 ознаками з добре розпізнаваним клавіатурним почерком

	Predicted																					
	s005	s010	s011	s017	s018	s019	s022	s024	s028	s033	s036	s039	s042	s043	s044	s049	s052	s053	s055	s056	Σ	
s005	96.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.2%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.8%	2.5%	0.0%	0.0%	0.0%	0.0%	400	
s010	0.5%	96.0%	0.0%	1.5%	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.8%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.0%	400
s011	0.2%	0.2%	91.2%	0.5%	2.2%	0.0%	0.0%	0.0%	0.0%	0.5%	0.0%	0.8%	0.0%	0.0%	0.0%	0.0%	0.0%	0.2%	0.2%	3.5%	400	
s017	0.2%	0.2%	0.0%	97.5%	0.0%	0.2%	0.0%	0.2%	0.5%	0.5%	0.0%	0.0%	0.0%	0.0%	0.5%	0.0%	0.0%	0.0%	0.0%	0.0%	400	
s018	0.8%	1.5%	1.2%	0.2%	87.8%	0.5%	0.0%	0.5%	1.0%	1.0%	0.0%	1.2%	0.8%	0.0%	0.0%	0.0%	0.0%	2.2%	0.0%	1.2%	400	
s019	0.8%	0.0%	0.0%	0.0%	0.0%	96.5%	0.5%	0.5%	0.2%	0.0%	0.0%	0.0%	0.0%	0.2%	0.0%	1.2%	0.0%	0.0%	0.0%	0.0%	400	
s022	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	98.5%	0.0%	0.0%	0.0%	1.2%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	400	
s024	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.5%	96.8%	1.0%	0.0%	0.0%	0.0%	0.0%	0.2%	1.2%	0.0%	0.0%	0.0%	0.0%	0.0%	400	
s028	0.2%	0.0%	0.0%	0.5%	0.0%	0.2%	0.0%	1.8%	96.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.8%	0.0%	0.0%	0.0%	0.0%	0.2%	400	
s033	1.5%	0.0%	0.0%	0.0%	0.0%	0.2%	0.0%	0.2%	0.0%	96.0%	0.0%	0.0%	0.0%	0.2%	0.0%	1.8%	0.0%	0.0%	0.0%	0.0%	400	
s036	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	99.8%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	400	
s039	2.2%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	95.5%	0.0%	0.0%	0.5%	0.5%	0.0%	0.0%	0.0%	0.0%	400	
s042	0.2%	0.5%	0.0%	0.0%	0.0%	0.0%	0.0%	0.5%	0.0%	0.0%	0.0%	0.0%	94.2%	0.2%	4.0%	0.2%	0.0%	0.0%	0.0%	0.0%	400	
s043	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.2%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	99.0%	0.2%	0.2%	0.0%	0.0%	0.0%	0.0%	400	
s044	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.5%	0.0%	0.0%	1.5%	96.8%	0.8%	0.0%	0.0%	0.0%	0.0%	400	
s049	1.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.5%	0.0%	0.0%	0.5%	0.8%	0.0%	0.0%	0.2%	0.0%	96.8%	0.0%	0.0%	0.0%	0.0%	400	
s052	0.0%	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.5%	0.0%	0.0%	0.5%	0.0%	0.0%	0.0%	98.8%	0.0%	0.0%	0.0%	400	
s053	0.0%	0.2%	0.5%	0.2%	0.8%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.2%	97.0%	0.0%	0.0%	1.0%	400	
s055	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.5%	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	95.2%	3.8%	400	
s056	0.8%	2.8%	0.0%	0.5%	0.0%	0.0%	0.0%	0.0%	0.8%	0.2%	0.0%	0.8%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.5%	92.8%	400	
Σ	420	407	372	407	364	392	402	406	401	402	409	397	382	408	419	416	396	398	388	414	8000	

Рис. 7. Матриця помилок класифікації 20 користувачів з дослідної бази даних за первинними 31 ознаками з добре розпізнаваним клавіатурним почерком

Результати класифікації 20 відібраних користувачів за кожним з 19 дослідних інформативних параметрів (через «СА» позначено інтегральну точність класифікації відповідно до рис. 5, 6):

$$\begin{aligned} CA[UD.Key1.Key2] &= 0.94; & CA[UU.Key1.Key2] &= 0.93; \\ CA[H.Key1/DD.Key1.Key2] &= 0.943; & CA[H.Key2/UD.Key1.Key2] &= 0.93; \\ CA[UD.Key1.Key2/DD.Key1.Key2] &= 0.942; & CA[DD.Key1.Key2] &= 0.921; \\ CA[UU.Key1.Key2/HH.Key1.Key2] &= 0.942; & CA[H.Key2/DD.Key1.Key2] &= 0.919; \\ CA[H.Key1/HH.Key1.Key2] &= 0.941; & CA[H.Key2/HH.Key1.Key2] &= 0.919; \\ CA[UD.Key1.Key2/HH.Key1.Key2] &= 0.941; & CA[DD.Key1.Key2/HH.Key1.Key2] &= 0.919; \\ CA[H.Key1/UU.Key1.Key2] &= 0.939; & CA[H.Key1] &= 0.915; \\ CA[H.Key2/UU.Key1.Key2] &= 0.939 & CA[HH.Key1.Key2] &= 0.91; \\ CA[UD.Key1.Key2/UU.Key1.Key2] &= 0.936 & CA[DD.Key1.Key2/UU.Key1.Key2] &= 0.86. \\ CA[H.Key1/UD.Key1.Key2] &= 0.934; \end{aligned}$$

Наведені результати дозволяють виділити чотири найінформативніших ознаки клавіатурного почерку (тобто такі, що забезпечують найвищу точність класифікації):

1) тривалість паузи між натисканнями клавіш (сценарій класифікації за 10 параметрами **UD.dot.t**, **UD.t.i**, **UD.i.e**, **UD.e.five**, **UD.five.Shift.r**, **UD.Shift.r.o**, **UD.o.a**, **UD.a.n**, **UD.n.l**, **UD.l.enter**);

2) відношення тривалості натискання першої клавіші диграфу до часу між натисканнями клавіш диграфу (сценарій класифікації за 10 параметрами **H.dot/DD.dot.t**, **H.t/DD.t.i**, **H.i/DD.i.e**, **H.e/DD.e.five**, **H.five/DD.five.Shift.r**, **H.Shift.r/DD.Shift.r.o**, **H.o/DD.o.a**, **H.a/DD.a.n**, **H.n/DD.n.l**, **H.l/DD.l.enter**);

3) відношення тривалості паузи між натисканнями клавіш диграфу до часу між натисканнями клавіш диграфу (сценарій класифікації за 10 параметрами **UD.dot.t/DD.dot.t**, **UD.t.i/DD.t.i**, **UD.i.e/DD.i.e**, **UD.e.five/DD.e.five**, **UD.five.Shift.r/DD.five.Shift.r**, **UD.Shift.r.o/DD.Shift.r.o**, **UD.o.a/DD.o.a**, **UD.a.n/DD.a.n**, **UD.n.l/DD.n.l**, **UD.l.enter/DD.l.enter**);

4) відношення часу між відпусканнями клавіш диграфу до тривалості диграфу (сценарій класифікації за 10 параметрами **UU.dot.t/HH.dot.t**, **UU.t.i/HH.t.i**, **UU.i.e/HH.i.e**, **UU.e.five/HH.e.five**, **UU.five.Shift.r/HH.five.Shift.r**, **UU.Shift.r.o/HH.Shift.r.o**, **UU.o.a/HH.o.a**, **UU.a.n/HH.a.n**, **UU.n.l/HH.n.l**, **UU.l.enter/HH.l.enter**).

Найменш інформативними є наступні інформативні ознаки клавіатурного почерку:

1) відношення часу між натисканнями клавіш диграфу до часу між відпусканнями клавіш у диграфі (сценарій класифікації за 10 параметрами **DD.dot.t/UU.dot.t**, **DD.t.i/UU.t.i**, **DD.i.e/UU.i.e**, **DD.e.five/UU.e.five**, **DD.five.Shift.r/UU.five.Shift.r**, **DD.Shift.r.o/UU.Shift.r.o**, **DD.o.a/UU.o.a**, **DD.a.n/UU.a.n**, **DD.n.l/UU.n.l**, **DD.l.enter/UU.l.enter**);

2) тривалості диграфів (сценарій класифікації за 10 параметрами **HH.dot.t**, **HH.t.i**, **HH.i.e**, **HH.e.five**, **HH.five.Shift.r**, **HH.Shift.r.o**, **HH.o.a**, **HH.a.n**, **HH.n.l**, **HH.l.enter**);

3) тривалості натискань клавіш (сценарій класифікації за 11 параметрами **H.dot**, **H.t**, **H.i**, **H.e**, **H.five**, **H.Shift.r**, **H.o**, **H.a**, **H.n**, **H.l**, **H.enter**).

Висновки

1. Інтегральна точність класифікації користувачів за 19 дослідними часовими ознаками клавіатурного почерку змінювалась у дуже широкому діапазоні: від 86 до 94,3 %, отже припущення про різну інформативність часових характеристик диграфів отримало експериментальне підтвердження.

2. Найінформативнішими для задачі класифікації користувачів за клавіатурним почерком є відношення часових параметрів, а не їх абсолютні значення. Перехід до аналізу відношень часових параметрів диграфів призводить до «нормалізації» закону розподілу дослідних ознак, тобто «випадковість» зміни параметрів клавіатурного почерку зменшується (подібний

висновок було отримано у [5] на основі досліджень гістограм часових параметрів диграфів клавіатури).

3. Найінформативнішими для задачі класифікації користувачів за клавіатурним почерком є відносні часові параметри диграфів, складовою котрих в тій чи іншій формі є час паузи між натисканнями клавіш у диграфах.

Для ілюстрації цього висновку на рис. 8 наведено матрицю помилок класифікації для набору з 30 інформативних ознак: **H.dot/DD.dot.t, H.t/DD.t.i, H.i/DD.i.e, H.e/DD.e.five, H.five/DD.five.Shift.r, H.Shift.r/DD.Shift.r.o, H.o/DD.o.a, H.a/DD.a.n, H.n/DD.n.l, H.l/DD.l.enter, H.dot/UU.dot.t, H.t/UU.t.i, H.i/UU.i.e, H.e/UU.e.five, H.five/UU.five.Shift.r, H.Shift.r/UU.Shift.r.o, H.o/UU.o.a, H.a/UU.a.n, H.n/UU.n.l, H.l/UU.l.enter, UD.dot.t/UU.dot.t, UD.t.i/UU.t.i, UD.i.e/UU.i.e, UD.e.five/UU.e.five, UD.five.Shift.r/UU.five.Shift.r, UD.Shift.r.o/UU.Shift.r.o, UD.o.a/UU.o.a, UD.a.n/UU.a.n, UD.n.l/UU.n.l, UD.l.enter/UU.l.enter.**

Як можна побачити, інтегральна точність класифікації складає 98,1 % (мінімальна – 95,2 %), що значно краще ніж 95,9 % (мінімальна – 87,8 %), що дає класифікація за 31 інформативною ознакою з первинного датасету (рис. 6, 7).

		Predicted																					
		s005	s010	s011	s017	s018	s019	s022	s024	s028	s033	s036	s039	s042	s043	s044	s049	s052	s053	s055	s056	Σ	
Actual	s005	97.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.2 %	0.0 %	0.0 %	0.0 %	2.8 %	0.0 %	0.0 %	0.0 %	0.0 %	400	
	s010	0.0 %	98.5 %	0.0 %	0.2 %	0.0 %	0.0 %	0.0 %	0.0 %	0.2 %	0.0 %	0.0 %	0.2 %	0.0 %	0.0 %	0.2 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.5 %	400
	s011	0.2 %	0.0 %	95.2 %	0.0 %	0.8 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	1.2 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.2 %	2.2 %	0.0 %	400
	s017	0.0 %	0.0 %	0.0 %	98.2 %	0.0 %	0.0 %	0.0 %	0.5 %	0.5 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.8 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	400
	s018	0.5 %	0.0 %	0.5 %	0.2 %	95.5 %	0.2 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	1.2 %	0.0 %	0.0 %	0.0 %	0.2 %	0.0 %	0.8 %	0.0 %	0.8 %	0.0 %	400
	s019	0.2 %	0.0 %	0.0 %	0.0 %	0.0 %	97.8 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	2.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	400
	s022	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	99.5 %	0.2 %	0.0 %	0.0 %	0.2 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	400
	s024	0.0 %	0.0 %	0.0 %	0.2 %	0.0 %	0.0 %	0.0 %	98.2 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	1.5 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	400
	s028	0.0 %	0.0 %	0.0 %	0.2 %	0.0 %	0.0 %	0.0 %	1.0 %	98.5 %	0.0 %	0.0 %	0.0 %	0.0 %	0.2 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	400
	s033	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	99.8 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.2 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	400
	s036	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	100.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	400
	s039	0.8 %	0.0 %	0.0 %	0.0 %	0.0 %	0.2 %	0.0 %	0.0 %	0.0 %	0.2 %	0.0 %	98.8 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	400
	s042	0.5 %	0.2 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.2 %	0.0 %	0.0 %	0.0 %	0.0 %	97.5 %	0.0 %	1.5 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	400
	s043	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.2 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	99.8 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	400
	s044	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.2 %	0.0 %	0.0 %	0.5 %	0.0 %	0.2 %	0.2 %	98.8 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	400
	s049	0.8 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.5 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	98.8 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	400
	s052	0.0 %	0.0 %	0.0 %	0.0 %	0.2 %	0.0 %	0.0 %	0.0 %	0.0 %	0.2 %	0.0 %	0.0 %	0.5 %	0.0 %	0.0 %	0.0 %	99.0 %	0.0 %	0.0 %	0.0 %	0.0 %	400
s053	0.0 %	0.2 %	0.0 %	0.0 %	0.5 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.2 %	98.8 %	0.0 %	0.2 %	0.0 %	400	
s055	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.2 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	95.2 %	4.5 %	400	
s056	0.2 %	0.5 %	0.0 %	0.2 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.8 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	1.2 %	97.0 %	400	
Σ	401	398	383	398	388	393	399	402	397	403	403	411	393	401	411	416	397	398	387	421	8000		

CA=0.981

Рис. 8. Матриця помилок класифікації 20 користувачів з дослідної бази даних за 30 найінформативнішими ознаками з добре розпізнаваним клавіатурним почерком

Список літератури:

1. Заяць В.М. Математичний опис системи розпізнавання користувача комп'ютера / В.М. Заяць, М.М.Заяць // Фізико-математичне моделювання та інформаційні технології. Львів, 2005. Вип. 1. С. 146 – 152.
2. Заяць В.М. Визначення пріоритету детермінованих ознак при побудові системи розпізнавання об'єктів / В.М.Заяць, О.Шокира // 36. пр. наук.-практ. конф. ЛДІНТУ імені В. Черновола "Математичне моделювання складних систем". 2007. С.135 – 137.
3. Сравнительный анализ перспективных технологий аутентификации пользователей ПК по клавиатурному почерку / В. А. Алексеев, Д. В. Маслий, Д. Ю. Горелов // Радиотехника. 2017. Вып. 189. С. 195 – 201.
4. Keystroke Dynamics Benchmark Data Set. URL: [http:// https://www.cs.cmu.edu/~keystroke](http://https://www.cs.cmu.edu/~keystroke) (дата звернення 25.02.2020).
5. Дослідження статистичних властивостей клавіатурного почерку для вирішення задач аутентифікації користувачів комп'ютерних мереж / Д. Ю. Горелов, В. О. Алексеев, В. М. Бублик, Д. В. Маслий // Радиотехника. 2017. Вып. 197. С. 78 – 85.

Харківський національний
університет радіоелектроніки

Надійшла до редколегії 05.04.2020