УДК 621.369:534

В.А. АЛЕКСЕЕВ, Д.В. МАСЛИЙ, Д.Ю. ГОРЕЛОВ, канд. техн. наук

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ПЕРСПЕКТИВНЫХ ТЕХНОЛОГИЙ АУТЕНТИФИКАЦИИ ПОЛЬЗОВАТЕЛЕЙ ПК ПО КЛАВИАТУРНОМУ ПОЧЕРКУ

Введение

В настоящий момент защита персональных данных пользователей приобретает первоочередное значение. Стремительный рост и развитие огромного числа сетевых приложений для выполнения различного рода операций от социальной коммуникации и развлечений (социальные сети, игровые серверы) до проведения крупных платежных транзакций (системы онлайн банкинга, крупные торговые онлайн площадки, личные порталы абонентов) требует комплексных подходов по обеспечению безопасности на всех этапах их реализации. Потенциальный ущерб от кражи персональной информации может исчисляться колоссальными материальными издержками. В этой связи надежность процедуры аутентификации является необходимым обстоятельством, гарантирующим сохранность конфиденциальных данных.

Аутентификация пользователей является одним из наиболее важных и сложных аспектов обеспечения контроля несанкционированного доступа к ресурсам компьютерной системы. Она представляет собой процесс, с помощью которого система проверяет, имеет ли пользователь законное право на доступ к ней. Традиционно выделяют три основных подхода для аутентификации пользователя. К ним относятся [1]:

- методы, основанные на владении дополнительными программно-аппаратными средствами для идентификации пользователя системой (смарт-карты, документы, удостоверяющие личность, ключи для генерации электронных цифровых подписей, аппаратные ключи, USB-токены);
- методы, основанные на уникальном знании пользователей (персональные идентификаторы, логины, пароли, их совместное использование);
- биометрические технологии аутентификации (статические и динамические).

На рис. 1 приведена онтология различных режимов аутентификации, включая биометрические поведенческие модели.

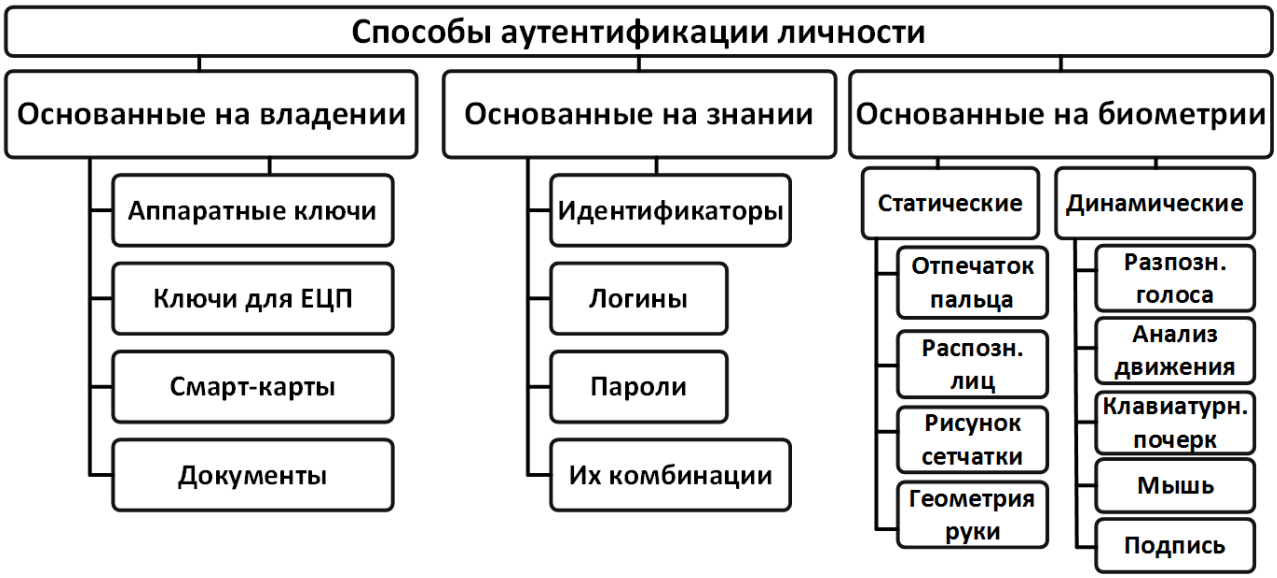

Рис. 1

Наибольший интерес для исследователей представляют биометрические параметры пользователей, которые в отличие от традиционных паролей, достаточно трудно воспроизвести и они не могут быть утеряны, переданы третьим лицам, украдены или забыты.

Биометрические характеристики пользователя условно можно разделить на две большие группы: физиологические (статические) и поведенческие (динамические) [2]. Первые – присущи человеку от момента рождения и неизменны на протяжении всей жизни. К ним относятся отпечатки пальцев, распознавание характерных черт лица, геометрии руки, сетчатки глаза и т.д.

Биометрические технологии, основанные на поведенческих моделях пользователя, могут включать: распознавание голоса человека, детекцию и изучение характера движения, анализ подписи, анализ характера работы пользователя с интерфейсом, взаимодействие пользователя с компьютерной мышью, анализ клавиатурного почерка. Данные характеристики все чаще используются в качестве полноценной технологии контроля доступа к ресурсам системы или дополнительной меры, способной повысить безопасность прохождения процедуры аутентификации. На основании собранных и проанализированных данных создаются уникальные профили пользователей, которые в последующем будут участвовать в прохождении процедуры аутентификации. Низкая стоимость реализации и последующего сопровождения такого рода решений, отсутствие необходимости создания дополнительного оборудования, легкость интеграции с существующими системами безопасности привели к бурному росту исследований в данной области [3].

Еще одним преимуществом использования биометрических характеристик на основе поведенческих моделей является то обстоятельство, что аутентификация может производиться на протяжении всего сеанса работы пользователя – так называемая «непрерывная аутентификация» – в отличие от разовой парольной проверки во время первого входа в систему. Это может предотвратить несанкционированное вмешательство в пользовательскую сессию уже после того, как был осуществлен начальный вход в систему.

Анализ и сравнение методов, основанных на изучении характеристик клавиатурного почерка пользователя

Впервые вопросы использования индивидуальных особенностей работы с клавиатурой были рассмотрены в середине 70-х годов прошлого века [4]. За последних два десятилетия исследователи использовали различные методы, подходы и алгоритмы для сбора и обработки необходимых персональных данных, их представления, классификации и оценки эффективности для изучения возможности построения систем аутентификации на базе обширного парка устройств (персональных компьютеров и ноутбуков с механическими клавиатурами (МК), мобильных платформ (МП) с сенсорными дисплеями (СД), специализированных устройств ввода (СУВ) [5]. При этом наиболее распространенными параметрами для анализа являются временные характеристики (см. рис. 1) нажатия клавиш: время удержания клавиши (ВУК) и время между нажатиями соседних клавиш (ВСК), которые могут быть измерены с точностью до единиц миллисекунд. На рис. 2 два последовательных события клавиатуры образуют диграф, три события – триграф, n следующих друг за другом событий клавиатуры – n-граф [6].

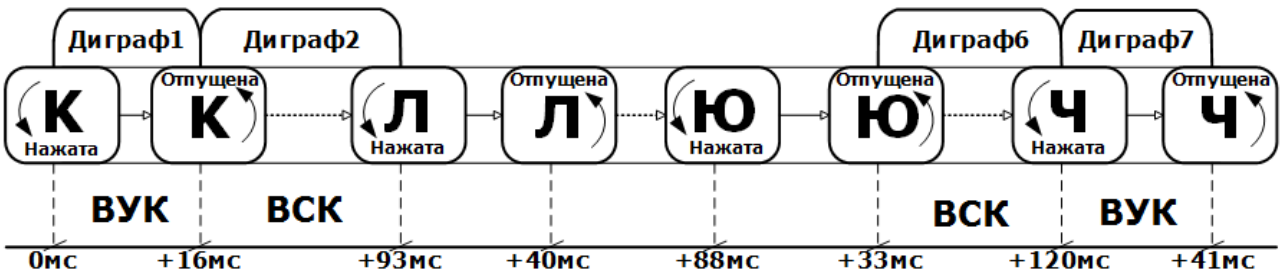

Рис. 2

Дополнительными пространственными параметрами, извлекаемыми из данных о нажатиях клавиш, могут быть величина давления, положение пальцев на сенсорном экране, частота ошибок при наборе текста, скорость набора и функции коррекции текста [7].

Основными показателями качества работы системы биометрической аутентификации личности являются ошибки трех видов, выраженные в процентном соотношении. К ним относятся: FRR (False Reject Rate) – ошибка первого рода, которая определяется как вероятность ошибочного отказа законному пользователю; FAR (False Accept Rate) – ошибка второго рода, которая определяется как вероятность допуска незарегистрированного пользователя. Общая оценка системы описывается при помощи равного уровня ошибок EER (Equal Error Rates), при котором FAR и FRR равны. Более низкий показатель EER указывает на лучшую эффективность системы аутентификации. Другие критерии оценки системы включают в себя технологичность, производительность и удобство использования.

Одной из важнейших частей системы аутентификации, основанной на анализе какой-либо поведенческой модели, является базовый алгоритм обработки полученных данных. Анализ литературы в данной предметной области показывает, что в более ранних работах большинство методов классификации представляли собой вероятностно-статистические подходы, однако, в настоящее время исследователи сосредоточились на изучении и апробации подходов по классификации параметров клавиатурного почерка, в основе которых лежат современные методы машинного обучения.

В рамках вероятностно-статистических подходов обработки полученных данных чаще всего применяются [8]:

- математическое ожидание и среднее квадратичное отклонение;
- алгоритмы, основанные на вычислении оценок сходства между объектами, например, метод ближайших соседей k-NN (k-nearest neighbors algorithm);
- классификаторы, основанные на измерении геометрических расстояний – Евклидово расстояние (Euclidean distance), расстояние Махаланобиса (Mahalanobis distance), Манхэттенское расстояние (Manhattan distance), расстояние Хемминга (Hamming distance);
- методы, основанные на величине меры энтропии (неупорядоченности) системы;
- алгоритмы динамической трансформации временной шкалы (dynamic time warping);
- скрытые марковские модели (hidden Markov model);
- байесовские классификаторы (Bayes classifier);
- критерии проверки гипотез Колмогорова и Смирнова (Kolmogorov, Smirnov criterions);
- методы дисперсионного анализа ANOVA (Analysis of variance).

Методы машинного обучения включают в себя [9]:

- искусственные нейронные сети ANN (artificial networks);
- деревья принятия решений (decision trees);
- решения на базе элементов нечеткой логики (fuzzy logic);
- эволюционное моделирование (evolutionary computation);
- методы опорных векторов (support vector machines).

Архитектура искусственных нейронных сетей может быть представлена в виде:

- многослойного перцептрона MLP (multilayer perceptron);
- сети радиально-базисных функций RBFN (radial basis function network);
- обучаемого векторного квантования LVQ (learning vector quantization);
- самоорганизующейся карты Кохонена SOM (self-organizing map) или SOFM (self-organizing feature map).

В качестве основы эволюционного моделирования применяются:

- генетические алгоритмы GA (genetic algorithms);
- метод роя частиц PSO (particle swarm optimization);
- муравьиный алгоритм ACO (ant colony optimization algorithms).

В таблице представлены результаты сравнения различных подходов анализа характеристик клавиатурного почерка, использующих наиболее распространенные методы машинного обучения и вероятностно-статистические модели.

Год	Участники эксперимента	Анализируемые параметры	Метод классификации	Вид и длина вводимого текста	Устройство	Результаты, %
Методы, основанные на расчете среднего значения и дисперсии (Mean and STD)						
2005	205	ВСК	*Mean and STD*	длинный	МК	FAR:0.5 FRR:5
2009	30	ВСК, ВУК	*Mean and STD*	текст	МП	EER:13
2009	1254	ВСК, ВУК	*Mean and STD*	короткий	МК	FAR:16 FRR:1
2012	51	ВСК, ВУК	*Mean and STD*	короткий	МК	EER:8.4
2013	152	ВСК, ВУК, давление	*Mean and STD*	цифровой	МП	FAR:4.19 FRR:4.59
Подходы, основанные на методе k ближайших соседей (k-NN)						
2002	7	ВСК, ВУК,	*k-NN*	цифровой	СУВ	EER:78-99
2008	10	давление	*k-NN*	цифровой	ТС	EER:1.00
2010	120	ВСК, ВУК	*k-NN*	короткий	МК	EER:1.00
2010	100	ВСК, ВУК	*k-NN*	текст	МК	EER:2.7
2010	30	ВСК, ВУК	*k-NN*	цифровой	МК	EER:0.5
2013	40	ВСК, ВУК	*k-NN*	длинный	МК	EER:6.1
Методы, основанные на измерении геометрических расстояний (Euclidean distance)						
2007	21	ВСК, ВУК	*Euclidean distance*	короткий	МК	EER:3.8
2008	30	ВСК, ВУК давление	*Euclidean distance*	цифровой	СУВ	FAR:15 FRR:0 EER:10
2009	16	ВСК, ВУК	*Euclidean distance*	короткий	МК	EER:4.28
2010	100	ВСК, ВУК	*Euclidean distance*	текст	МК	EER:2.7
2010	189	ВСК, ВУК	*Euclidean distance*	длинный	МК	FAR:0.01 FRR:3
2011	51	ВСК	*Euclidean distance*	длинный	МК	EER:0.84
2011	20	ВСК	*Euclidean distance*	длинный	МК	FAR:2 FRR:4
Подходы, основанные на величине энтропии системы (Entropy)						
2005	31	ВСК	*Entropy*	длинный	МК	FAR: 1.99 FRR: 2.42
2005	205	ВСК	*Entropy*	длинный	МК	FAR: 0.5 FRR: 5
2009	21	ВСК, ВУК	*Entropy*	длинный	МК	FAR: 0.14 FRR: 1.59
2011	50	ВСК	*Entropy*	длинный	МК	EER: 10
2011	186	ВСК, ВУК	*Entropy*	длинный	МК	FAR: 1.65 FRR: 2.75

Продолжение таблицы

Год	Участники эксперимента	Анализируемые параметры	Метод классификации	Вид и длина вводимого текста	Устройство	Результаты, %
Подходы, основанные на других статистических методах, а также их комбинациях						
1990	26	ВСК	Baysian, Minimum Distance	короткий	МК	FAR: 2.8 FRR: 8.1
2004	41	ВСК, ВУК	Gaussian mixture modeling	короткий	МК	FAR: 4.3 FRR: 4.8 EER: 4.4
2005	9	ВСК, ВУК, давление,	ANOVA	цифровой	МК	EER: 2.4
2006	100	ВСК, ВУК, давление,	Dynamic time warping	цифровой	МК	EER: 1.4
2006	20	ВСК, ВУК	Hidden Markov model	цифровой	МК	EER: 3.6
2006	20	ВСК, ВУК	Euclidian, Mahalanobis	цифровой	ТС	FAR: 0 FRR: 2.5
2009	25	ВСК, ВУК	Gauss, Parzen, K-NN, K-mein	короткий	МК	EER: 1.00
2009	100	ВСК, ВУК	Bayesian, Euclidean, Hamming	короткий	МК	EER: 6.96
2010	51	ВСК, ВУК	Manhattan distance	короткий	МК	EER: 7.16
2010	35	ВСК	Kolmogorov-Smirnov	длинный	МК	EER: 7.16
2011	100	ВСК, ВУК	Gaussian PDF	короткий	МК	EER: 1.401
2011	55	ВСК, ВУК	Spearmans's foot rule distance	длинный	МК	FAR: 2.02 FRR: 1.84
2011	33	ВСК, ВУК	Naive Bayesian	длинный	МК	EER: 1.72
2013	152	ВСК, ВУК, давление, датчики	k-mean	цифровой	МП	FAR: 4.19 FRR: 4.59
2013	10	ВСК, ВУК, датчики	Bayesian	цифровой	ТС	FAR: 0.02 FRR: 0.018
2014	30	ВСК, ВУК	SMD, SED	цифровой	МК	EER: 26
Методы, основанные на деревьях принятия решений (Random forest decision tree, RFDT)						
2010	21	ВСК, ВУК,	RFDT	длинный	МК	FAR: 3.47 FRR: 0 EER: 1.73
2010	28	ВСК, ВУК,	RFDT	цифровой	МК	FAR: 0.03 FRR: 1.51 EER: 1
Методы, основанные на искусственных нейронных сетях (ANN)						
2007	100	ВСК, ВУК,	ANN	короткий	МК	FAR: 1 FRR: 8
2010	25	ВСК, ВУК, давление	ANN	цифровой	МК	FAR: 4.12 FRR: 5.55

Продолжение таблицы

Год	Участники эксперимента	Анализируемые параметры	Метод классификации	Вид и длина вводимого текста	Устройство	Результаты, %
Подходы, основанные на методе опорных векторов (SVM)						
2007	24	ВСК, ВУК	SVM	длинный	МК	FAR: 0.76 FRR: 0.81 EER: 1.57
2007	61	ВСК, ВУК	SVM	короткий	–	FAR: 14.5 FRR: 1.78
2007	5	давление	SVM	цифровой	МК	FAR: 0.95 FRR: 5.6
2011	117	ВСК, ВУК	SVM	короткий	МК	EER: 11.8
2014	30	ВСК, ВУК, давление	SVM	цифровой	СД	EER: 2.8
Подходы, основанные на других менее распространенных методах машинного обучения						
2005	43	ВСК, ВУК	Decision trees, Monte Carlo	короткий	МК	FAR: 0.88 FRR: 9.62
2005	53	ВСК, ВУК	Fuzzy ARTMAP	короткий	МК	FAR: 0.87 FRR: 4.4
2007	30	ВСК, ВУК	Sequence alignment algorithms	короткий	МК	FAR:0.2 FRR: 0.2 EER: 0.4
2014	42	ВСК, ВУК, давление, датчики	Naive, Bayesian	короткий	МП	EER: 12.9
2015	42	ВСК, ВУК, давление, датчики	Two-class	короткий	МП	EER: 3

Выводы

1. Анализируя приведенные в таблице данные, можно сделать вывод, что конечная точность полученных результатов определяется рядом факторов, главными среди которых являются: основной алгоритм классификации полученных данных, количество участников эксперимента с различной величиной опыта работы с клавиатурой, способ и организация ввода данных и аппаратная платформа, на базе которой производится тестирование системы аутентификации.

2. Использование современных методов машинного обучения дает возможность получить более высокие результаты при аутентификации в сравнении с вероятностно-статистическими, но требует от системы большей вычислительной сложности, что зачастую не оправдано и может быть реализовано лишь частично. Максимальные результаты показывают методы, основанные на деревьях принятия решений и искусственных нейронных сетях.

3. Использование комбинаций вероятностно-статистических подходов анализа особенностей клавиатурного почерка в рамках одного алгоритма позволяет повысить уровень точности системы в целом.

4. Использование дополнительных пространственных параметров клавиатурного почерка (давление, координаты пальцев на сенсорных дисплеях мобильных устройств) дает значительный прирост в показателях точности в сравнении со стандартными механическими

клавиатурами за счет выделения дополнительных информативных признаков анализируемого почерка субъекта.

5. Существенно повысить качество систем аутентификации пользователей ПК можно путем перехода к комплексным моделям, которые учитывают, например, следующие характеристики:

- данные, которые позволят однозначно идентифицировать пользователя (уникальный идентификатор, пароль, цифровые подписи оборудования и т.д.);
- информационный почерк пользователя – клавиатурный почерк и динамику системы «пользователь – мышь»;
- активность пользователя в рамках операционной системы (средний процент использования центрального процессора, средний объем занимаемой памяти, тип наиболее
  часто открываемых файлов и т.д.);
- сетевая активность пользователя (наиболее часто используемые сетевые сервисы и приложения, тип активности пользователя в сети и т.д.);
- программно-аппаратные изменения в конфигурации ПК (установка нового программного обеспечения, установка или замена внешних устройств и т.д.).

Список литературы: 1. *Брюхомицкий, Ю.А.* Исследование биометрических систем динамической аутентификации пользователей ПК по рукописному и клавиатурному почеркам : учеб.-метод. пособие / Ю.А Брюхомицкий, М.Н. Казарин. – Таганрог, 2004. 2. *Kyle O.Bailey, James S. Okolica, Gilbert Peterson*, "User identification and authentication using multi-modal behavioral biometrics," Computers and Security journal, vol.43, pp.77-89, June 2014. 3. *Roman V.Yampolskiy, Venu Govindaraju*. Behavioural biometrics: a survey and classification // International Journal of Biometrics, vol.1, pp.81-113, Novemberv 2008.4. *Gaines, R., Lisowski, W., Press, S. and Shapiro, N.* Authentication by keystroke timing: some preliminary results // Technical Report Rand Rep. R-2560-NSF, RAND Corporation, p. 51, 1980. 5. *Spillane, R.* Keyboard apparatus for personal identifica-tion. IBM Technical Disclosure Bulltin, 17(11), 1975. 6. *Sim, T. and Janakiraman, R.* Are digraphs good for free-text keystroke dynamics? // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–6, 2007. 7. *Umphress, D., Williams, G.* Identity verification through keyboard characteristics // International Journal of Man-Machine Studies, vol. 23(3), pp. 263–273, 1985. 8. *Balagani, K.S., Vir V. Phoha, Ray, A. and Phoha, S.* On the discriminability of keystroke feature vectors used in fixed text keystroke authentication // Pattern Recognition Letters, vol. 32(7), pp. 1070–1080, 2011. 9. *Md Liakat Ali, John Monaco, Charles Tappert, Meikang Qiu.* Keystroke Biometric Systems for User Authentication // Journal of Signal Processing Systems, vol. 86(2) pp.175–190, March 2017.

Харьковский национальный
университет радиоэлектроники

Поступила в редколлегию 25.03.2017