Integrated Approach to User Authentication Based on Handwritten Signature

Evgeny Kostyuchenko, Egor Krivonosov, Alexander Shelupanov
Department of Complex Information Security and Electronic Computing Systems
Tomsk State University of Control Systems and Radioelectronics
Tomsk, Russia
key@keva.tusur.ru; egor-yrga@mail.ru; saa@tusur.ru

***Abstract*—A new approach to the integration of the results of several tools aimed at solving a single problem is considered. The approach is illustrated by the example of solving the authentication task for signature dynamics based on the naive Bayesian classifier and the neural network. The approach guarantees results not worse than any of the classifiers separately from the point of view of a monotonous combination of the probabilities of errors of the first and second kind. The obtained results can be applied in the construction of a multifactor authentication system.**

***Keywords*—naive Bayesian classifier, neural network, handwritten signature**

I. Introduction

There are many different methods of user authentication, one of such methods is biometric authentication. They can be separated into static methods and dynamic methods. Static biometric authentication includes using of fingerprint [1-3], iris [4-6], facial geometry [7-10], voice [11-13], hand geometry [14-16] and other physiological characteristics of a person. Examples of dynamic biometric authentication are using of keystroke dynamics [17-19] and signature dynamics [20]. More part of these approaches are intended to be used even when organizing user authentication for mobile devices.

However, such methods in most cases are either costly (typically, for static methods), or do not provide the required authentication accuracy for practical use (typically, fordynamic methods) [21]. There is a problem of increasing the accuracy of such systems. First way to solving this problem is the simultaneous use of several classifiers in the analysis of the characteristics of one biometric source. The second way is constructing a multifactor authentication system. However, simple "AND" scheme (the authentication is passed when all decision subsystems vote for the same user) can't be used for this purpose. The reason of this is a significant jump of the probabilities for the first kind error, payed for reduction of the probabilities for the second kind error, not allowing for practical using of such integration.

The problem can be solved by using a majority system, but in this case it is necessary to bring the number of decisive classifiers (or authentication factors) at least to three, which significantly increases the complexity of the system being developed. There arises the problem of developing an approach to integration that guarantees accuracy not worse than the best of the combined classifiers from the point of view of any possible monotonic combination of the probabilities for first and second kind errors.

II. The Integration of Several Approaches

The basic idea used for integrating several approaches is as follows. Suppose that we have a set (vector) of parameters *P* describing the set for authentication. In this case, the task of authentication can be reduced to the classification task with an illegal user. In this case a system containing n users classifies they to *n* + 1 class. Authentication is considered passed when the received class matches with the identifier. The work of one classifier can be represented as a mapping A_i of a parameters set *P* to a vector *U* with dimension *n* + 1. Vector *U* describes belonging a given set of parameters A_i to each of the users or none of them:

$$U_i=A_i(P) \tag{1}$$

In this case, as a rule, the class (user) is selected as the maximum component of the output vector *U* exceeding a certain threshold *T*. Note that the last class (none of the users) may not exist, but be defined as "none of the classes reached the minimum threshold".

To integrate classifiers, we apply the function *F* to combine the outputs U_i of *k* classifiers into a single output *U*. The resulting vector for decision can be defined as:

$$U=F(U_i \ldots U_i)=F(A_1(P)\ldots A_1(P)) \tag{2}$$

At the same time, it is obvious that there are some restrictions to integration function *F* and not every function can be used for this task:

- The function *F* must be monotonous function of any component of the vector U_i. In otherwise, we can have a

This work was supported by the Ministry of Education and Science of the Russian Federation within 1.3 federal program «Research and development in priority areas of scientific-technological complex of Russia for 2014-2020» (grant agreement № 14.577.21.0172 on October 27, 2015; identifier RFMEFI57715X0172).

mistake in the user's rank detection. For example, the user with the maximum rank in the output vector U_i (identifier of user that is authentication system solution) may cease to be such not under the influence of other classifiers, but only under the influence of one function F, which contradicts common sense.

- The integration function must contain weight values describing the influence of each of the classifiers $I_1 ... I_k$. We can detect settings of our integration function using changing of this coefficients for minimization of total classifier error.
- There should be a "neutral" value of each of the coefficients Ii, under which this classifier does not affect the integration as a whole. The example of such values is "1" for the multiplying integration functions or "0" for the sum integration functions. This restriction guarantees that there is a set of weights in which the complex degenerates into a separate classifier. As a consequence, integration classifier shows the final results no worse than this individual classifier from any of the possible points of view and any type of error calculation, because integration and individual classifiers are equal in case of using this "neutral" coefficient value.
- The final function implicitly contains the vector of thresholds T (possibly individually for each user), which is necessary for the principal possibility of making a decision in favor of each users. In case of every outputs are lesser than this threshold we have no legal class in our system and current user who is trying to pass authentication is an intruder.

In view of the foregoing, the final function can be described as:

$$U=F(I,\ T,\ U_i\ ...\ U_i\ )=F(I,\ T,\ A_1(P)\dots\ A_1(P)) \tag{3}$$

Setting up the integration system for the chosen function F is represented as the task of optimizing the final error in this case. Final error can be any fixed monotonous combination of the probabilities of errors of the first and second kind. Optimization result depends from the values of the parameters of the vectors I and T. This solution, based on the above restrictions, includes the results of each of the individual classifiers. This, as a consequence, should guarantee the results no worse than any of the individual classifiers, include the best.

III. Authentication Based on a Handwritten Signature as Two Classification Methods Integration Example

The first 8 harmonics of the Fourier decomposition of the pen's coordinates x, y, z, the pressure p and the slope angles α and θ are used as the parameters. Also velocities and accelerations of their change was used. Total count of parameters is 6 × 3 × 8 = 144 per signature.

The combination of two classifiers was considered. As approaches for integration the naive Bayesian classifier [22, 23] and the perceptron [24] were used. The integration functions are given in Table 1. In this case, functions 1 and 2 essentially show a degenerate classifier. Also should be noted that not all the above functions satisfy the above restrictions. This is done to practically confirm necessity of this restrictions. N - output of the neural network, B - output of the Bayesian classifier, α, β - weighting coefficients.

TABLE I. INTEGRATION FUNCTIONS

Function number	Function
1	$f(x)= \alpha^0\cdot N+\beta\cdot B\cdot 0$
2	$f(x)= \alpha\cdot N\cdot 0+ \beta^0\cdot B$
3	$f(x)=\alpha\cdot B+\beta\cdot N$
4	$f(x)=B\cdot\alpha\cdot N\cdot\beta$
5	$f(x)=B^\alpha*N^\beta$
6	$f(x)=lg(B^\alpha+N^\beta\ )$
7	$f(x)=sinh(B+\alpha)*sinh(N+\beta)$
8	$f(x)=sinh(B\cdot\alpha)+sinh(N\cdot\beta)$
9	$f(x)=sinh(B^\alpha\ )+sinh(N^\beta)$
10	$f(x)=tanh(B+\alpha)*tanh(N+\beta)$
11	$f(x)=tanh(b\cdot\alpha)+tanh(N\cdot\beta)$
12	$f(x)=tanh(B^\alpha\ )+tanh(N^\beta)$
13	$f(x)=\surd(B\cdot\alpha)+\surd(N\cdot\beta)$
14	$f(x)=\surd(B^\alpha\ )+\surd(N^\beta\ )$
15	$f(x)=sinh(B\cdot\alpha+N\cdot\beta)$
16	$f(x)=B^\alpha+N\cdot\beta$
17	$f(x)=lg(B^\alpha+N\cdot\beta)$
18	$f(x)=tanh(B+\alpha)\cdot tanh(N\cdot\beta)$
19	$f(x)=\surd(B^\alpha\ )+tanh(N\cdot\beta)$
20	$f(x)=tanh(B\cdot\alpha)+sinh(N\cdot\beta)$
21	$f(x)=tanh(B\cdot\alpha)\cdot sinh(N\cdot\beta)$
22	$f(x)=tanh(B+\alpha)+sinh(N\cdot\beta)$
23	$f(x)=tanh(B+\alpha)\cdot sinh(N\cdot\beta)$

IV. Experiment

At the time of the experiment, the database contains signatures from 8 users in the number of more than 1500 signatures. At the preliminary stage, 10 training cycles were done for all functions, after that 100 cycles for the top 10 and functions 1 and 2, for comparing. PFEK is an error of the first kind, PSEK is an error of the second kind, the PE is simply the probability of error, the CE is a linear combination of errors, with a 10-fold importance of the second kind of error. The results of 100 training cycles are presented in Table 2.

From the presented results it is clear that some integration functions outperform the results of each of the integrated systems separately. An additional check showed the statistical significance [25] of these differences except the functions number 5, 9, 10, 11, 12 (1 and 2 were not checked). On closer examination, it can be seen that all the remaining functions satisfy the above restrictions. This fact confirms the correctness of the assumption that these restrictions must be met for correct integration functions. A statistically significant improvement of the criterion of the total error (PE) due to the application of integration is also shown.

TABLE II. INTEGRATION FUNCTIONS

№	Function	PFEK	PSEK	PE	CE
1	$f(x)=\alpha^0\cdot N+\beta\cdot B\cdot 0$	0,0568	0,0148	0,0150	0,1413
2	$f(x)=\alpha\cdot N\cdot 0+\beta^0\cdot B$	0,0658	0,0050	0,0123	0,1098
3	$f(x)=\alpha\cdot B+\beta\cdot N$	0,0605	0,0046	0,0111	0,0999
4	$f(x)=sinh(B\cdot\alpha)+sinh(N\cdot\beta)$	0,0621	0,0042	0,0111	0,0998
5	$f(x)=tanh(B+\alpha)\cdot tanh(N+\beta)$	0,0598	0,0057	0,0118	0,1087
6	$f(x)=tanh(b\cdot\alpha)+tanh(N\cdot\beta)$	0,0588	0,0051	0,0111	0,0997
7	$f(x)=\sqrt{(B\cdot\alpha)}+\sqrt{(N\cdot\beta)}$	0,0454	0,0066	0,0109	0,0944
8	$f(x)=sinh(B\cdot\alpha+N\cdot\beta)$	0,0597	0,0048	0,0111	0,0991
9	$f(x)=tanh(B+\alpha)\cdot tanh(N\cdot\beta)$	0,0617	0,0054	0,0121	0,1085
10	$f(x)=tanh(B\cdot\alpha)+sinh(N\cdot\beta)$	0,0579	0,0047	0,0115	0,1016
11	$f(x)=tanh(B+\alpha)+sinh(N\cdot\beta)$	0,0594	0,0048	0,0114	0,1005
12	$f(x)=tanh(B+\alpha)\cdot sinh(N\cdot\beta)$	0,0624	0,0053	0,0119	0,1071

When we tried to compare this results with results of other authors, a problem of open dynamic signature database was detected. Other scientist work with image of the signature only or not provide a signature sources for result comparing. In this case we can only compare numerical results on different signature databases. This information is presented in Table 3.

TABLE III. COMPARISON OF THE RESULTS WITH THE RESULTS OF SIMILAR SYSTEMS OF IDENTIFICATION BY SIGNATURE

Article	Method	Accuracy	User count
[26]	Basic concepts of graph theory	94,25 %	27
[27]	Multi-section vector quantization	98%	330
[28]	Statistical analysis	97,75 %	200
[29]	Principal Component Analysis +Neural Network	89,475 %	200
[30]	Principal Component Analysis +Neural Network	93,1%	200
[31]	Neural Network	95%	10
Current, [23]	Neural Network+Bayesian	98,5%	8

It can be concluded that, on the one hand, the results are no worse than those for similar systems. On the other hand, only the work [22] can act as a direct analogue with comparing the volume of the database and the number of users in it. Works with an open database of examples for confirmation are not presented in open repositories such as [32]. The idea of including to the repository our own set data for future researchers is actual. It can be reliably asserted that the results obtained using the integration of the Bayesian classifier and the neural network are no worse than using separate classifiers on identical sets of data and no worse than using a separate neural network in comparison with the analogous work [31].

V. CONCLUSIONS

The approach to the integration of several decision-making apparatuses is theoretically and experimentally proved in the course of the work done. This is shown on the example of combining the neural network and the naive Bayesian classifier for solving the task of user authentication based on the dynamics of the signature. A statistically significant improvement of the criterion based on probability of error is shown. This makes it possible to speak about the applicability of the proposed approach. This integration approach allows to guarantee increasing the efficiency of the complex of several decision-making methods in comparison with any of them separately.

In the future, it is planned to test the efficiency of the proposed approach when constructing a multifactor authentication system in mobile devices and using examples from the UCI Machine Learning Repository [32].

ACKNOWLEDGMENT

This work was supported by the Ministry of Education and Science of the Russian Federation within 1.3 federal program «Research and development in priority areas of scientific-technological complex of Russia for 2014-2020» (grant agreement № 14.577.21.0172 on October 27, 2015; identifier RFMEFI57715X0172).

REFERENCES

[1] Ram A. Athira and T.S. Jyothis, “Reducing Vulnerability of a Fingerprint Authentication System”, In: Abawajy J., Mukherjea S., Thampi S., Ruiz-Martínez A. (eds) Security in Computing and Communications SSCC 2015. Communications in Computer and Information Science, vol 536, pp. 157-167, 2015, DOI: https://doi.org/10.1007/978-3-319-22915-7_15

[2] T. Chuan-Chin and N. Han-Foon, "Behavioral Fingerprint Authentication: The Next Future," ICBBT '17 Proceedings of the 9th International Conference on Bioinformatics and Biomedical Technology, pp. 1-5, 2017, DOI: https://doi.org/ 10.1145/3093293.3093296

[3] M. R. Zafar and M. A. Shah, "Fingerprint authentication and security risks in smart devices," 2016 22nd International Conference on Automation and Computing (ICAC), pp. 548-553, 2016, DOI: https://doi.org/ 10.1109/IConAC.2016.7604977

[4] N. Schmid, J. Zuo, F. Nicolo and H. Wechsler, ”Iris Quality Metrics for Adaptive Authentication”, In: Bowyer K., Burge M. (eds) Handbook of Iris Recognition. Advances in Computer Vision and Pattern Recognition, pp. 101-118, 2016, DOI: https://doi.org/10.1007/978-1-4471-6784-6_5

[5] S. Thavalengal, P. Bigioi, and P. M. Corcoran, "Iris authentication in handheld devices - considerations for constraint-free acquisition," IEEE Trans. Consumer Electronics, vol. 61, pp. 245-253, 2015, DOI: http://doi.org/10.1109/TCE.2015.7150600

[6] P. S. Vanthana and A. Muthukumar, "Iris authentication using Gray Level Co-occurrence Matrix and Hausdorff Dimension," 2015 International Conference on Computer Communication and Informatics (ICCCI), pp. 1-5, 2015, DOI: https://doi.org/10.1109/ICCCI.2015.7218133

[7] M. Chowdhury, J. Gao and R. Islam R., “Biometric Authentication Using Facial Recognition”, In: Deng R., Weng J., Ren K., Yegneswaran V. (eds) Security and Privacy in Communication Networks. SecureComm 2016. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering,, vol 198, pp, 287-295, 2017, DOI: https://doi.org/10.1007/978-3-319-59608-2_16

[8] C. Ding and D. Tao, "A Comprehensive Survey on Pose-Invariant Face Recognition," ACM TIST, vol. 7, pp. 37:1-37:42, 2016, DOI: http://doi.org/10.1145/2845089

[9] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 815-823, 2015, DOI: https://doi.org/10.1109/CVPR.2015.7298682

[10] T. Khadhraoui, S. Ktata, F. Benzarti, and H. Amiri, "Features Selection Based on Modified PSO Algorithm for 2D Face Recognition," 2016 13th International Conference on Computer Graphics, Imaging and Visualization (CGiV), pp. 99-104, 2016, DOI: https://doi.org/10.1109/CGiV.2016.28

[11] I.S. Kipyatkova and A.A. Karpov, “Variants of Deep Artificial Neural Networks for Speech Recognition Systems”, SPIIRAS Proceedings, vol. 49, no. 6, pp. 80-103, 2016, DOI: https://doi.org/10.156-22/sp.49.5

[12] R. C. Johnson, T. E. Boult, and W. J. Scheirer, "Voice authentication using short phrases: Examining accuracy, security and privacy issues," 2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS), pp. 1-8, 2013, DOI: https://doi.org/10.1109/BTAS.2013.6712713

[13] S. B. Sadkhan, B. K. Al-Shukur, and A. K. Mattar, "Biometric voice authentication auto-evaluation system," 2017 Annual Conference on New Trends in Information & Communications Technology Applications (NTICT), pp. 174-179, 2017, DOI: https://doi.org/10.1109/NTICT.2017.7976100

[14] R. Kumar, “Hand Image Biometric Based Personal Authentication System”,. In: Dey N., Santhi V. (eds) Intelligent Techniques in Signal Processing for Multimedia Security. Studies in Computational Intelligence, vol 660, pp. 201-226, 2016, DOI: https://doi.org/10.1007/978-3-319-44790-2_10

[15] E.-S. M. El-Alfy, "Automatic Identification Based on Hand Geometry and Probabilistic Neural Networks," 2012 5th International Conference on New Technologies, Mobility and Security (NTMS), pp. 1-5, 2012, DOI: https://doi.org/10.1109/NTMS.2012.6208758

[16] A. Gangopadhyay, O. Chatterjee, and A. Chatterjee, "Hand shape based biometric authentication system using radon transform and collaborative representation based classification," 2013 IEEE Second International Conference on Image Information Processing (ICIIP-2013), pp. 635-639, 2013, DOI: https://doi.org/10.1109/ICIIP.2013.6707672

[17] G. Jagadamba, S.P. Sharmila and T. Gouda,”A Secured Authentication System Using an Effective Keystroke Dynamics.”, In: Sridhar V., Sheshadri H., Padma M. (eds) Emerging Research in Electronics, Computer Science and Technology. Lecture Notes in Electrical Engineering, vol 248, pp. 453-460, 2016, DOI: https://doi.org/10.1007/978-81-322-1157-0_46

[18] A. Goodkind, D.-G. Brizan, and A. Rosenberg, "Improvements to keystroke-based authentication by adding linguistic context," 2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS), pp. 1-6, 2015, DOI: https://doi.org/10.1109/BTAS.2015.7358766

[19] J. Huang, D. Hou, S. Schuckers, and Z. Hou, "Effect of data size on performance of free-text keystroke authentication," IEEE International Conference on Identity, Security and Behavior Analysis (ISBA 2015), pp. 1-7, 2015, DOI: https://doi.org/10.1109/ISBA.2015.7126361

[20] S. Pal, U. Pal and M. Blumenstein, “Signature-Based Biometric Authentication”, In: Muda A., Choo YH., Abraham A., N. Srihari S. (eds) Computational Intelligence in Digital Forensics: Forensic Investigation and Applications. Studies in Computational Intelligence, vol 555, pp. 285-314, 2014, DOI: https://doi.org/10.1007/978-3-319-05885-6_13

[21] D. Bhattacharyya, R. Ranjan, F.A. Alisherov and M. Choi, “Biometric Authentication: A Review”, International Journal of u- and e- Service, Science and Technology, vol. 3, no. 2, pp. 13-28, 2009, URL: https://www.researchgate.net/publication/46189709_Biometric_Authentication_A_Review

[22] Irina Rish, “An empirical study of the naive Bayes classifier”, IJCAI Workshop on Empirical Methods in AI, 2001. URL: https://www.researchgate.net/publication/228845263_An_Empirical_Study_of_the_Naive_Bayes_Classifier

[23] E. Kostyuchenko, M. Gurakov, E. Krivonosov, M. Tomyshev, R. Mescheryakov and I. Hodashinskiy, “Integration of Bayesian Classifier and Perceptron for Problem Identification on Dynamics Signature Using a Genetic Algorithm for the Identification Threshold Selection”, in: Cheng L., Liu Q., Ronzhin A. (eds) Advances in Neural Networks – ISNN 2016. ISNN 2016. Lecture Notes in Computer Science, vol 9719, pp. 620-627, 2016, DOI: https://doi.org/10.1007/978-3-319-40663-3_71

[24] F. Rosenblatt, “Principles of neurodynamics; perceptrons and the theory of brain mechanisms”, Washington, Spartan Books, 1962.

[25] G. James, D. Witten, T. Hastie and R. Tibshirani, “An Introduction to Statistical Learning with Applications in R”, Springer-Verlag New York, 2013, DOI: https://doi.org/10.1007/978-1-4614-7138-7

[26] T. Fotak, M. Baa and P. Koruga, “Handwritten signature identification using basic concepts of graph theory”, WSEAS Transactions on Signal Processing., vol. 7, pp. 117-129, 2011

[27] M. Faundez-Zanuy and J. M. Pascual-Gaspar, "Efficient on-line signature recognition based on multi-section vector quantization," Pattern Analysis and Applications, vol. 14, pp. 37-45, 2010 , DOI: https://doi.org/10.1007/s10044-010-0176-8

[28] M.R. Nilchiyan, R.B. Yusof and Alavi S.E., “Statistical on-line signature verification using rotation-invariant dynamic descriptors”, In: The 10th Asian Control Conference 2015 (ASCC 2015), pp. 1-6, 2015, DOI: https://doi.org/10.1109/ASCC.2015.7244603

[29] W.A.W. Adnan, F.L. Malallah, S. Mumtazah and S. Yussof, “Online Handwritten Signature Recognition by Length Normalization using Up-Sampling and DownSampling”,. International Journal of Cyber-Security and Digital Forensics (IJCSDF), vol 4, pp. 302–313, 2015, DOI: https://doi.org/DOI10.17781/P001545

[30] V. Iranmanesh, S. M. S. Ahmad, W. A. W. Adnan, S. Yussof, O. A. Arigbabu, and F. L. Malallah, "Online Handwritten Signature Verification Using Neural Network Classifier Based on Principal Component Analysis," in TheScientificWorldJournal, 2014, DOI: https://doi.org/10.1155/2014/381469

[31] P. Babita, "Online Signature Recognition Using Neural Network," 2015, DOI: https://doi.org/10.4172/2332-0796.1000155

[32] M. Lichman ”UCI Machine Learning Repository”, Irvine, CA: University of California, School of Information and Computer Science, 2013, URL: http://archive.ics.uci.edu/ml