Modelling user behaviour characteristics for identification

Daniela Chudá, Maroš Majerčík

Abstract: *In the e-learning process is identification and authentication of user very important. This paper proposes a new form of passive authentication of user during his work with computer, which is based on behavioural biometrics. We use high level behaviour of the user for his identification, which is based on sensible action made in computer system. We have designed, implemented and verified a system for passive authentication of user, which is based on monitoring the events generated by operating system in response to user activity within the scope of arbitrary program.*

Key words: *User behaviour characteristics, identification, authentication, monitoring of user, behavioural biometrics.*

INTRODUCTION

Checking the identity of the user in e-learning systems is designed only to identification and authentication of the user at the beginning of e-learning process. The users can share the identity in systems by means of sharing the username and the password. The introduction of additional authentication techniques by modelling user behaviour characteristic provides an added level of security.

This paper is organized as follows: Section 1 describes the related works and Section 2. is about the our solution of modelling user behaviour, proposed algorithm, its implementation and Section 3 is about the testing and evaluations and the final section presents conclusions and suggestions for future work.

1. RELATED WORK

It was shown [3] that the same neurophysiological factors that shape our writing activities, and make it unique, are also responsible for the rhythm and dynamics of writing on the keyboard. It is a sign that it is for each individual characteristic and hardly imitable. For this reason, the model of keystroke dynamics provides of interesting opportunities in the field of electronic systems. We analyze previous work [5] on the topic keystroke dynamics, we analyze the type of used measures, the evaluation methods and the data used for testing. Some authors uses the statistics evaluation methods [2], [13] and some authors used as evaluation method the neural networks [18], [19]. The measure for keystroke dynamics are digraphs, trigraphs, flight time and dwell time. Mouse dynamics can be used for user identification [1] too. For this characteristic are use the mouse actions in the categories: mouse-move general mouse movement, drag-and-drop the action starts with mouse button down, movement, and then mouse button up, point-and-click mouse movement followed by a click or a double click, and silence no movement. This is a user behaviour on lower level, which is monitored the keystroke dynamics and mouse movement.

We can model user behaviour on higher level, which is based on his action in computer system. We analyze previous work to identified user behaviour patterns. Interaction between user and program can be analyze on the several levels of the abstraction [10], [11], from the lower level where are physical events to the higher level like windows system events, applications level,… The monitoring user work style and identifying different episode from the users actions [16] should by used for intelligent adaptive user interface.

2. OUR SOLUTION

We present the system for passive authentication of user, which is based on monitoring the events generated by operating system in response to user activity within the scope of arbitrary program. Our solution is based on analyze of the research [10], [11] and [16] and we propose the monitoring user work style without the time. Our

system process and evaluate the events in several steps, see Fig. 1. We have implemented the modules for the capture events generated by the operating system using the Windows API libraries, a module for processing these events, a module for creating a user model, and finally evaluation module.

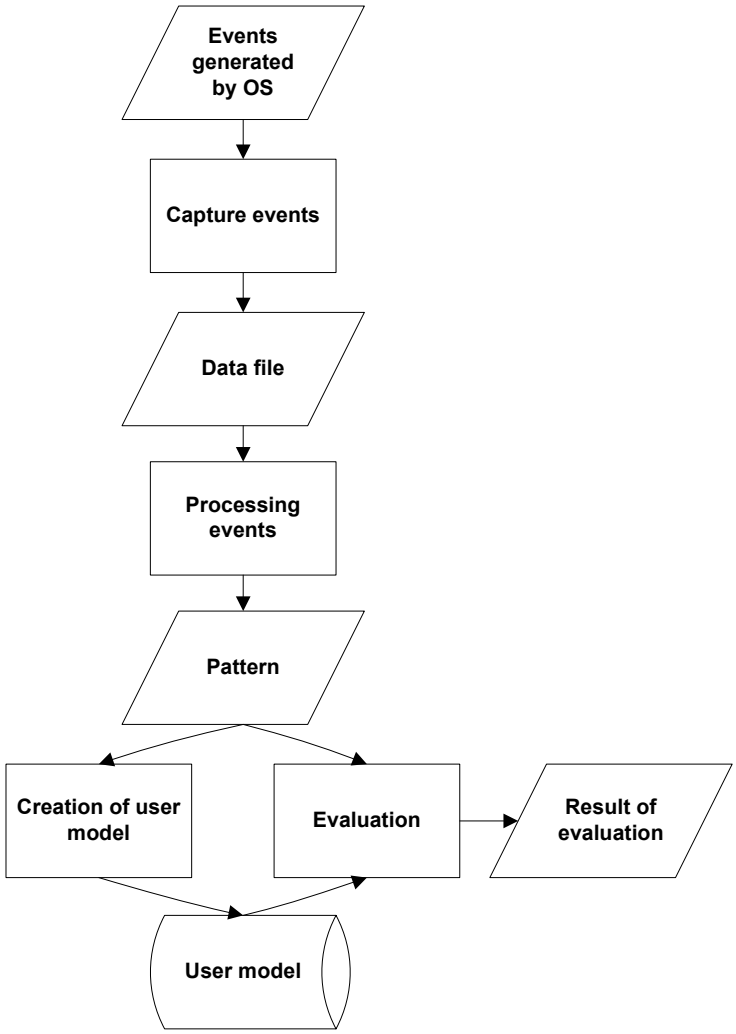

Fig.1. The diagram of propose system for modelling user behaviour characteristics for identification

We propose 3 comparative algorithms.

We propose comparative algorithm which is based on percentage of appearance events in patterns and events on the all user models. The winner is user model which is closest to the compatible model (our algorithm 2) [17].

$$winner = Max\{ each\ m\ in\ M : NEPEMm/NEP*100\% \} \quad (1)$$

Legend of (1):

NEPEM number of events in pattern which is equal to the model m,
NEP number of all events in pattern,
M set of all users models.

The model is created from user events generated by the operating system in response to user actions. Events are stored in the model as a unique set of events together with their frequency, which determines how often the event is generated in the

work of a particular user. Model also includes basic information about the user who created it in addition to the events.

The module for capture events generated by the OS, logging module is implemented like a applications which generated a data file with system events. In the table 1 we can see the list of attributes monitoring events.

The systems events for monitoring:

- events applicable for higher level of user monitoring:
  - Accelerator event (A),
  - Control event (C),
  - Key event (K),
  - Language change event (LC),
  - Menu action event (MeA),
  - Menu select event (MS),
  - Sys command event (SC),
  - Window activated event (WA),
  - Window created event (WC),
  - Window destroyed event (WD),
- events applicable for lower level of user monitoring:
  - Key event (K),
  - Mouse action event (MoA),
  - Mouse move event (MM),
  - Mouse wheel event (MW).

Table 1: The list of attributes monitored events

Event / Attribute	A	C	K	LC	MeA	MS	MoA	MM	MW	SC	WA	WC	WD
Event type	x	x	x	x	x	x	x	x	x	x	x	x	x
Name of the process	x	x	x	x	x	x	x	x	x	x	x	x	x
Time	x	x	x	x	x	x	x	x	x	x	x	x	x
ID component	x	x	-	-	x	x	-	-	-	-	-	-	-
Tags	-	-	x	-	-	x	x	x	x	-	-	-	-
Key	-	-	x	-	-	-	-	-	-	-	-	-	-
Allocation of keys.	-	-	-	x	-	-	-	-	-	-	-	-	-
Coordinates	-	-	-	-	-	-	x	x	x	-	-	-	-
Identifier	x	x	-	-	x	-	-	-	-	x	-	-	-

3. EVALUATION

We are using for testing and evaluating [17] the data from 22 users, what is 471 hours of active work with computer in operating system. The data were first filtered to include only higher level events. Subsequently was all files distributed in the samples, while the limit for the distribution we used the number of events.

Characteristics for evaluation of model:

- FRR, system do not authenticated authorized user (FRR - False Rejection Rate, it also uses the term False Alarm Rate), error type 1 - a legitimate user is rejected,
- FAR, system authenticated impostor as an authorized user (FAR - False Acceptance Rate, also used the term Impostor Pass Rate), error type 2 - an impostor is accepted as a legitimate user."

For evaluating the our algorithm 2 we choice the threshold in range 50-80%. All calculated values FAR and FRR are on figure 2 and 3.

We managed to obtain sufficient amount of data, but not all users to provide sufficient amount of data to create a stable model of the user. This resulted in high levels of FRR in the number of users 11,13,18,20.

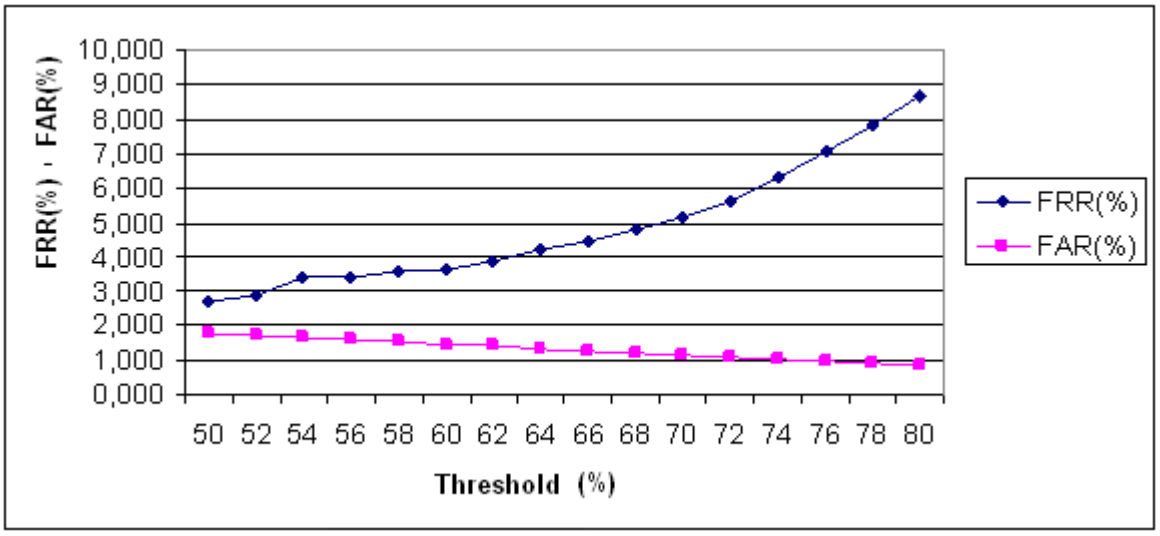

Fig.2. Calculated values FRR and FAR for threshold 50% - 80%, algorithm 2

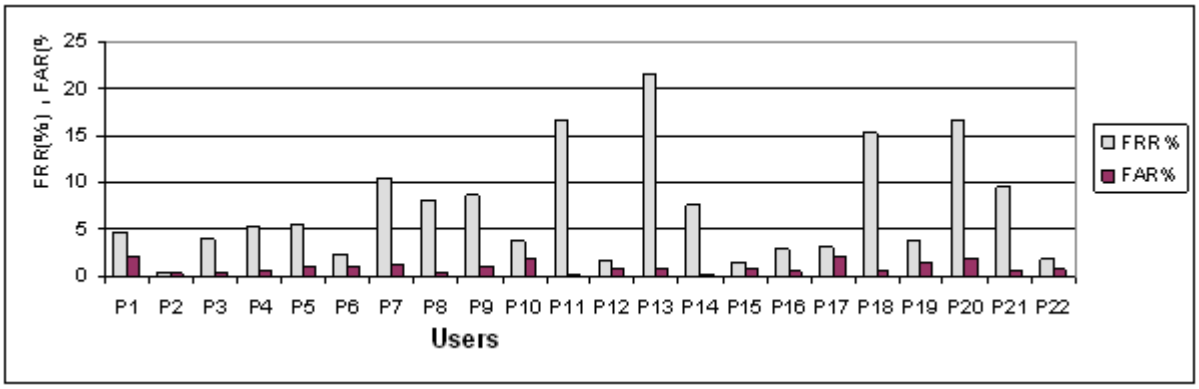

Fig.3. Calculated values FRR and FAR for users, threshold is 74%, pattern 300 events, algorithm 2

The results of our test shows that the algorithm 2 is capable of when the sample size of 100 events with relatively high accuracy to identify the user when the values of FRR and FAR 3.077% 2.289%. Please note that 100 events will generate an average user in the active work time for 1-2 minutes.

The constrains of our systems are in knowledgeable mimicking of user behaviour in systems and in changing of user environment - operating systems, software.

One of the main benefits of this work is to confirm the ability to identify the user based on his behavior at a higher level. For verification of the proposed solution has been shown that using high-level model of the user is able, with relatively high precision (FRR = 3.66% and FAR = 1.47%) to determine whether the computer works authorized user or an attacker. Important role in identifying and selecting appropriate play features (in our case, events) on the basis of which it is possible to distinguish the behaviour of multiple users.

Another benefit of high-level modelling in comparison with the modelling at the lower level is to improve the reaction time of the security system. In the case of writing on the keyboard this time around about 10 minutes [9], and we must take into account the fact that the user must use the keyboard at work. A similar situation is also in the case of work with the mouse, where you can calculate response time moving beyond

the 13 minute [1], that we know with reasonable accuracy to identify the user. Again it is a condition that the user must work with the mouse. The advantage of the solution proposed in this work is the fact that the data are obtained simultaneously from both the input devices, allowing to reduce the response time to a level of 1-2 minutes (depending on the intensity of the work of the user) while maintaining acceptable levels FRR = 3.077% and FAR = 2.289%.

CONCLUSIONS AND FUTURE WORK

We propose a method for passive authentication of user when use the computer systems during e-learning process. Similar models can be employed in e-learning environments [12]. Module was implemented for the capture events generated by the operating system using the Windows API libraries, a module for processing these events, a module for creating a user model, and finally evaluation module. To verify the identity of the user in the evaluation module, we designed and tested 3 methods of comparison samples of compliance with the model. Verification of the proposed solution we make to set of 22 users, from which we obtained a total of 736 hours of work for a computer, which is about 1.14 * 106 high-level events. The effectiveness of different methods of comparison, we verify the calculated values FRR (number of authorized user rejection expressed in percentages) and FAR (the number of penetrations attacker as a percentage) for all 22 users. For best results we achieved user authentication with comparing, with which we are in the size of the sample 300 events measured value FRR = 3.66% and FAR = 1.47%, with 300 events in the active user will generate work for about 3-6 minutes.

For the above limitations and benefits that better results could be achieved by combining the two levels (upper and lower) and thus created a model of behavior, which would use the advantages of both approaches. The connection could it be implemented in several ways: combination of different models keystroke dynamics, mouse movements and users behaviour on higher level using systems.

ACKNOWLEDGEMENT

This work was partially supported by the Scientific Grant Agency of the Slovak Republic, grant No. VEGA 1/0508/09.

REFERENCES

[1] Ahmed A. E. A., Traore I. A New Biometric Technology Based on Mouse Dynamics, IEEE Transactions on Dependable and Secure Computing, July-September 2007,vol. 4, no. 3, pp. 165-179.

[2] Bergadano, F., Gunetti, D., and Picardi, C. User authentication through keystroke dynamics. ACM Trans. Inf. Syst. Secur. 5, 4, Nov. 2002, 367-397.

[3] Bryan, W. L., Halter, N. Studies in the physiology and psychology of the telegraphic language. In: Gardener, E. H., Gardner, J. K.: The psychology of skill: Three studies. NY Time Co., NY 1973. s. 35-44. As consistent with: Obaidad, M. S., Sadoun, B.: Keystroke dynamics based authentication. In: Jain, A. K., Bolle, R., Pankanti, S.: Biometrics: Personal identification in networked society. IBM T. J. Watson Research Center, Yorktown Heights, NY, 1998, s.213-230

[4] Chandra, A. and Calderon, T. Challenges and constraints to the diffusion of biometrics in information systems. Commun. ACM 48, 12 , Dec. 2005, 101-106.

[5] Chudá, D., Ďurfina, M. Multifactor Authentication based on keystroke dynamics In: Compsystech 2009, Rousse, Bulgaria, 2009.

[6] Costa, P. C. G., Barbará, D., Laskey, K. B. et al. DTB Project: A Behavioral Model for Detecting Insider Threats. International Conference on Intelligence Analysis, 2005,

[7] Klein, D. V. Foiling the cracker: A survey of, and improvements to, password security. In Proceedings of the 2nd USENIX Security Workshop, pages 5–14, August 1990.

[8] Enhanced Online Banking Security - Zero Touch Multi-Factor Authentication. Entrust, Inc.

[9] Gunetti, D. and Picardi, C. Keystroke analysis of free text. ACM Trans. Inf. Syst. Secur. 8, 3 , Aug. 2005, 312-347.

[10] Hilbert, D. M. and Redmiles, D. F. Extracting usability information from user interface events. ACM Comput. Surv. 32, 4 , Dec. 2000, 384-421.

[11] Hilbert, D.M., Robbins, J.E., and Redmiles, D.F. Supporting Ongoing User Involvement in Development via Expectation-Driven Event Monitoring. Tech Report UCI-ICS-97-19, Department of Information and Computer Science, University of California, Irvine, April 1997.

[12] Ivanović, M., Pribela, I., Vesin, B., Budimac, Z: Multifunctional environment for e-learning purposes, Journal of Mathematics, novi Sad, vol.38, br. 2, str. 153-170, 2008.

[13] Joyce, R. and Gupta, G. 1990. Identity authentication based on keystroke latencies. Commun. ACM 33, 2, Feb. 1990, 168-176. URL: http://doi.acm.org/10.1145/75577.75582

[14] Lammers, A. and Langenfeld, S.: Identity Authentication Based on Keystroke Latencies Using Neural Networks. J. Comput. Small Coll. 6, 5 (Apr. 1991), 48-51.

[15] Lee, J., Choi, S., and Moon, B.: An evolutionary keystroke authentication based on ellipsoidal hypothesis space. In: Proceedings of the 9th Annual Conference on Genetic and Evolutionary Computation (London, England, July 07 - 11, 2007). GECCO '07. ACM, New York, NY, 2090-2097.

[16] Liu, J., Wong, C. K., Hui, K. K.: An adaptive user interface based on personalized learning. In: IEEE Intelligent Systems, Mar.-Apr. 2003, s. 52-57.

[17] Majerčík M.: Model používateľa pre jeho identifikáciu (), In: diploma work, Faculty of Informatics and Information Technologies, Slovak university of technology in Bratislava, may 2009

[18] Obaidat, M. S., Sadoun, B.: An Evaluation Simulation Study of Neural Network Paradigm for Computer Users Identification. Information Sciences Journal-Applications, Elsevier, November 1997, vol. 102, no. 1-4, pp. 239-258.

[19] Obaidat, M.S., Macchairolo, D.T.: An On-line Neural Network System for Computer Access Security. IEEE Trans. Industrial Electronics, April 1993, Vol. 40, N. 2, pp. 235-241.

[20] Schneier, B.: Sensible Authentication. Queue 1, 10 (Feb. 2004), 74-78.

[21] Urnphress, D., and Williams, G.: Identity verification through keyboard characteristics. Int. J. Man-Machine Studies 23, 3 (Sept. 1985), 263-273

ABOUT THE AUTHORS

Mgr. Daniela Chudá, PhD., Institute of Informatics and Software Engineering, Faculty of Informatics and Information Technology, Slovak University of Technology, Ilkovičova 3, 842 16 Bratislava 5, Slovak Republic, E-mail: chuda@fiit.stuba.sk

Ing. Maroš Majerčík, graduated at Faculty of Informatics and Information Technology, Slovak University of Technology in Bratislava, Ilkovičova 3, 842 16 Bratislava 5, Slovak Republic, E-mail: maros.majercik@gmail.com