ISSN No. 0976-5697

Volume 8, No. 5, May – June 2017

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Presskey- A Keystrokes Dynamics Based Authentication System

Dharmendra Singh[1] , Bhawnesh Jaggi[1] , Himanshu Nayyar[1] , Amit Kumar[1]

[1] Department Of Computer Science
A.R.S.D. College (University of Delhi)
New Delhi, India

Abstract: Biometric systems have been applied to improve and enhance the security of various systems. These systems analyses physiological or behavioural features obtained from the users in order to perform authentication. Biometric features should ideally meet a number of requirements, including permanence i.e. the analysed biometric feature will not change over time. However, recent studies have shown that this is not the case for several biometric modalities. Adaptive biometric systems deal with this issue by adapting the user model over time. In this paper some algorithms for adaptive biometrics have been investigated and suggestions have been made for the use of soft biometrics while taking considerations of various performance and other related issues.

Keywords: Keystroke; Unique Identity; Soft Biometrics; Security; Classification.

I. INTRODUCTION

Presskey is a keystroke dynamics based authentication system which recognizes users based on their typing rhythm. Keystrokes dynamics seems to be the ultimate solution to every authentication problem we have ever had. It is unique, reliable, uncopiable and even adaptive to changes in the typing patterns of a person. Even after all these benefits over normal biometric solutions which are prone to counterfeit, it is not popular. The technology has been around for quite some time now and a considerable amount of research has been done in the field, still it seems to remain unused. Several studies already emphasized that the combination of individual techniques in ensembles may lead to more accurate and stable decision models.

The major drawback of physical and biological biometrics is that they can be forged although it needs a lot of effort. With the advancement of technology it has become very easy to counterfeit fingerprint and iris scans. Further, the growth of medicine has facilitated disguising one’s biological features. Therefore, these methods can no longer be relied upon for recognizing identity thefts.

Even bluffing a voice recognition system or a signature based authentication system is no longer tough; we came across such frauds in the recent time. At this point of time we need a security system that is dynamic in nature which cannot be faked or copied by any other person. Here comes Presskey, a keystrokes dynamics based authentication system which resembles some needed features.

II. LITERATURE REVIEW AND BACKGROUND STUDY

The current technological advancements provided a number of services to society for the betterment and ease, particularly owing to the Internet-based applications. However, at the same time, this situation came up with increased data exposure, giving a new momentum to concern regarding identity theft [5]. With fast growing technological advancement we need a system with enhanced security. An alternative approach is the use of biometrics that takes into account the consideration of physiological/ behavioral of masses [2].

For this purpose, we analysed various aspects of the system and proposed a model that enhances the security of such systems. Like other behavioral biometrics such as signatures, typing pattern is unique to a person and also the same for him. Presskey intends to use this property to provide a security mechanism which is almost unbreachable by an impostor.

A brief review of related work:

A. Trajectory Mining for Keystroke Dynamics Authentication [1].

This work focuses on strengthening the already used username and password mechanism via introducing additional measures in the form of keystroke dynamics. Through careful Keystroke Dynamic Analysis (KDA) particular features are selected for profile generation and generated profile is then used for authenticating users.

It uses username keystroke data for generating profile of users. Particular letters in the username works as a point and they transformed to a trajectory path [1]. After that trajectory profile is generated and based on trajectory dissimilarity users authentication will be done [10]. Although typo behavior of a user can be different at different situations but still it can be used for enhancing security in real systems with careful and sufficient data gathering for profile generation.

B. Emphasizing typing signature in keystroke dynamics using immune algorithms [5].

This work uses one-class classification approach coupled with immune algorithms for identification purposes in keystroke dynamics. The key here is a deep analysis and through understanding of data that helps in preprocessing and extraction of more refined features; after that rank transformation is applied to improve the recognition [5].

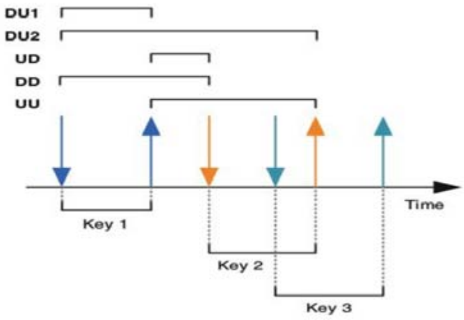

Figure 1. Feature extraction [14].

HERE:

1. DU1: Time difference between the instants in which a key is pressed and released. This feature represents the time that the key keeps pressed and is also named by some authors as dwell time.

2. DU2: Time difference between the instants in which a key is pressed and the next key is released.

3. UD: Time difference between the instants in which a key is released and the next is pressed. This feature is also known as flight time.

4. DD: Time difference between the instants in which a key is pressed and the next key is pressed.

5. UU: Time difference between the instants in which a key is released and the next key is released.

Based on typing rthym and extracted features it generates a profile of typing signature(characteristics of typo bahavior). Rank transformation might be helpful in managing the dimensionality that is unavoidable in keystroke dynamics [5].

C. Keystrokes dynamics based user authentication using long and free text strings from various input devices [7].

This analyses the critical issues of Keystroke Dynamics Authentication (KDA) based on long and free text strings, with variety of input devices, instead of short predefined texts. To do so it works around three main questions, that whether [7]:

- The authentication performance depends on the type of input device.
- The length of the text affects the authentication performances.
- The authentication algorithms are appropriate for certain conditions (input device and text lengths).

Foe getting the answer of the above mentioned questions they built 12 One-class classifier; although multi-class classifier comes up with better results but use of multi-class classifier would be impractical in the application scenario. Keyboard as input device works well with various methods; however, requires customization when used with other input devices. The size of both reference and test keystrokes affects the performance considerably. If only one factor is to be increased, then, increasing the length of test keystrokes reduces the authentication error to a greater degree than increasing the reference keystrokes. Error rate in authentication gets minimal when used with appropriate measures.

D. Continuous keystroke dynamics: A different perspective towards biometric evaluation [8].

This describes a way to evaluate a biometric continuous keystroke dynamics system: a system that will continuously monitor the typing behaviour of a user and will determine if the current user is at present the genuine one or not, so that the system can be locked if a different user is detected. The performance evaluation standards for static and continuous keystroke dynamics systems vary vastly due to its complexity and size. The static system relies on the count of wrong decisions made by the user but on the other hand the continuous system needs to work faster and with accurate results to identify the impostor quickly [8]. Detection of imposter as soon as possible with minimum keystrokes increases the trust and performance of the system [8].

In the case of continuous authentication systems every user is provided with a trust count which is initially set to 100 and will ever cross 100, this value can decrease and increase according to the user's typing patterns and behaviours. On the other hand, in a static authentication system the user is evaluated on the factors like "False match rate" and "False non match rate", these two factors help in calculation of "Equal error rate" which determines the genuineness of the user at the moment it is checked.

E. Hybrid Model with Fusion Approach to Enhance the Efficiency of Keystroke Dynamics Authentication [9].

This proposed a hybrid model for keystroke dynamic authentication with four fusion approach. Database works as a base to extract features for generation of template, which is compact form of keystroke feature data. Hybrid model based on combination of Gaussian probability density function (GPDF) and Support Vector Machine (SVM) will convert test features into scores [9]. Finally apply four fusion rules on hybrid model to fusing GPDF and SVM scores to improve the final result. The primary focus was on using two separate functions of keystroke dynamics authentication systems and combines them to come up with better and accurate results.

F. A Secured Authentication System Using an Effective Keystroke Dynamics [12].

This work shows that an authentication system that is based on effective Adaptive Learning Classification (ALC) algorithm, where a self-threshold for each user was decided based on user input. Training and testing data lead to an average false reject rate of 10.00 % and the average false accept rate of 0.0025 % [15].

Raw measurements available from each keyboard can be recorded to determine dwell time (the time a key is pressed) or flight time (the time between key-down and the next key down and the time between key up and the next key up) [11]. After recording, data are processed through the algorithm, which serves the primary pattern for future comparison and analysis. Various methodologies are as follows: Hold key timings, Ant colony optimization technique, Keystroke timings, Virtual Key Force [3], Telling human and bot apart [4].

III. METHODOLOGY

After examining existing tools and going through some experimental studies, we discovered that, the performance of keystrokes dynamics based biometrics are lower as compared to other forms of authentication due to inter class variability pertaining to computer users which can be accounted for by a way of typing which is different when they are nervous or angry or even sad. The different state of mind pertaining to these variable emotions results in a different typing rhythm. At such a stage the system would fail because the user won't be authenticated with temporal changes in typing style.

The distinctive feature of the user's typing behavior but not limited to only these, are:

-the pressure that is exerted on each key
-the position adopted by hands when user types
-functional relationship between fingers and keys
-the sound generated by keystroke
-the vibration generated by keystroke
-the sequence used to perform an action
-quantity of errors committed when writing and methods to correct them
-the use of special keys
-keystroke speed
-time interval that remains on a pressed a key
- time interval between pressing a key and then another.

- ***Breachability concern due to similarity pattern in human typo behaviour:***

Looking at the vast application areas of such an authentication system, we surveyed by asking people, belonging to different occupations, and browsed the internet for various statistics about the possible drawbacks or the possible methods to cheat such a system. After a careful evaluation and survey, it came out that according to statistics 10 people out of every 100 have approximately the same typing signatures; close enough to fool such a system. Furthermore, if an imposter was to only use mouse or an on-screen keyboard to carry out its malicious tasks, even then the system would fail.

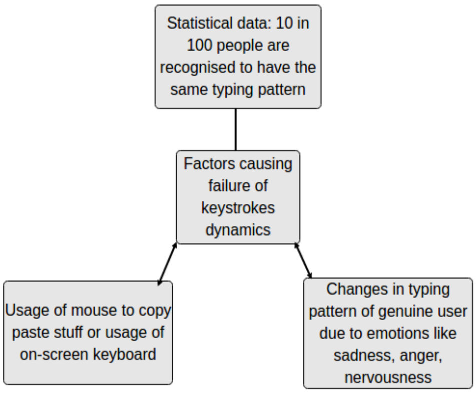
Statistical data: 10 in 100 people are recognised to have the same typing pattern
Factors causing failure of keystrokes dynamics
Usage of mouse to copy paste stuff or usage of on-screen keyboard
Changes in typing pattern of genuine user due to emotions like sadness, anger, nervousness

Figure 2. Factors affecting keystroke typing.

Owing to these problems, Presskey cannot be as promising as it guarantees to be. To overcome these problems we have the concept of soft biometrics. Soft biometrics are the characteristics that provide some information about the individual but lack the distinctiveness and permanence to sufficiently differentiate any two individuals. If recognition of these characteristics supplements the primary authentication system, then it will allow a refinement of the search of the genuine users amongst imposters. The soft biometrics we intend to implement are facial recognition including gender recognition and clothes color recognition in case employees wear uniforms, mouse dynamics to prevent mouse usage, and finally a google authenticator for the purpose of unlocking a locked down mechanism.

Even though every 10 in 100 people have the same typing signature, a combination of such soft biometrics along with keystrokes dynamics provides us with an unbreakable authentication system. When typing patterns would be complemented by mouse dynamics and facial recognition, the complete biometric signature of the person would be unique.

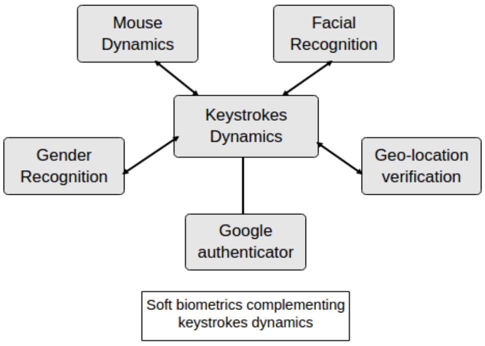

Figure 3. Soft-Biometrics

- ***Design and Authentication criteria for such System***

The table given below summarizes all the methods; criteria for measurements; parameters for evaluation as well as the evaluation metrics to be used in building a system employing keystroke dynamics.

Table 1. Criteria/Parameter for Authentication [12]

Sr. No.	Type of Keystroke Dynamic Authentications	Static and continuous
1	Measurement criteria for static systems	Count of wrong decisions made by the user
2	Measurement criteria for continuous systems	Identify the impostor quickly
3	Parameters for evaluation of the keystroke data	Dwell time and Flight time
4	Analysis patterns for logged keystroke data	Hold key timings, Ant colony optimization technique, Keystroke timings, Standard and measure, Bio password, Telling human and bot apart, Virtual key force, Gaussian probability density function, Support Vector Machine
5	Best metrics to use for best results	Single feature - PR (Press 1, Release 1) Multi feature - PR (Press 1, Release 1) + RP (Press 2, Release 1)
6	Soft biometrics to be used alongside Keystroke Dynamics	Mouse dynamics, Facial recognition, and Google authenticator
7	Different Key Instants	Figure given above
8	5 Components of a Keystroke Dynamics Authentication Systems	Data acquisition, Features extraction, Classification and matching, Decision, and Adaptation

9	Evaluation metrics	(i) False Rejection Rate (FRR): percentage of genuine users rejected. (ii)False Acceptance Rate (FAR): percentage of impostors accepted. (iii) Equal Error Rate (EER): Point on the ROC curve where FRR is equal to FAR, and (iv) Half-Total Error Rate (HTER)

the system we intend to design would employ machine learning and neural network to make it adaptive and also perform big data analysis, on the huge amounts of data gathered and back it up weekly on to the cloud. To avoid the system being heavier due to numerious computational work, the algorithmic computations would also be carried out on the cloud. Therefore, building a robust system would require tapping into technologies from almost all fields.

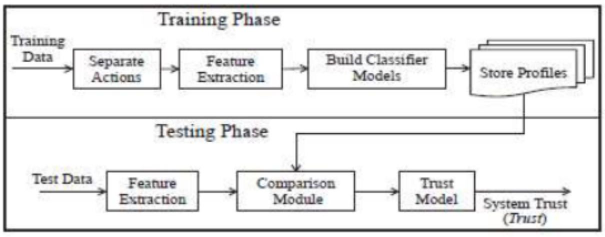

Figure 4. Training/ Testing Phase in machine learning

Keystroke dynamics includes four modules i.e. keyboard monitoring, features extraction, classifier algorithm, database. Basically, each user will be required to type some training data upon installation of Presskey. This training will generate a unique profile corresponding to the user depending on his typing pattern and this profile will be used to recognize him the the next time he uses his system. The following diagram depicts it clearly.

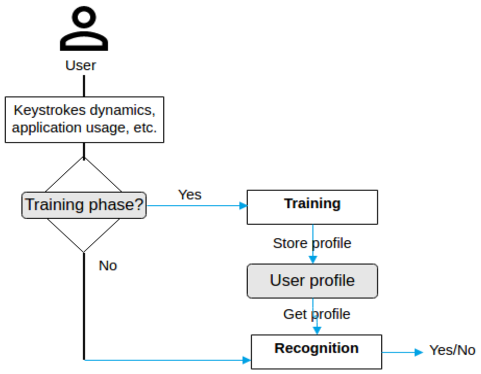

Figure 5. Profile Storing and Recognition Flow

- ***Steps and work flow for the system***

The following points presents a step by step go through, through the entire system i.e. mentioning at each step what happens next and to which component the user is directed to on reaching that stage.

1) On setting up Presskey in the system, the user would be prompted for typing some sample data. This would be part of training the system for user's typing patterns. This initial training data would be used to recognize the user for a few initial times. The system would forever train for recognizing the user's typing patterns as a background process. Each new set of training data would update the existent data.

2) Similar training exercise for the system would also exist for recognizing the mouse-using-behaviour i.e. mouse dynamics as well as for recognizing the user's face from various angles, if a facial recognizer is also implemented. If Facial recognition carried out, it should be done after a certain time interval continuously.

3) The Presskey security system though would be securing a standalone system but would not be limited to personal systems. Each user's training data would be weekly backed up on cloud, so that if a user uses some other system having Presskey installed, then upon entering some unique ID and password, his training data would be fetched from the cloud and user would be authenticated according to his typing pattern.

4) This raises the point that each user's training data would be saved on the cloud against some unique ID and password. Even if some imposter uses it to fetch the training data and tries to login, the system, within a minute of its usage, would recognize the counterfeit and lock itself.

5) The Presskey security system would be developed in such a way that it recognizes the user while it is entering its username or login ID or anything along with a password, which is used as the first level of authentication, i.e. a user would be checked for as soon as it starts entering its credentials.

6) If at any stage Presskey evaluates a user to be an imposter, then it must lock down the application it is protecting, after giving a certain amount of warnings, if required. This can be made customizable according to the user. If it wants to protect a high security application then the application may get locked without any warnings in cases of imposter recognition otherwise for low security applications 2-3 levels of warnings may be provided.

For employing warning mechanism for a low security system, the following criteria would be used: firstly, the suspected imposter would be again repeatedly checked for its typing pattern for a certain amount of time (say 5 minutes) initially; if he is again suspected to be an imposter, then as a second warning and check he would be asked to type out some training data. This data would be the same data the user would have typed during the installation of Presskey, because this data is recognizant of the user's first typing signature. If he is again suspected, then he would be asked to re-login with its private credentials and again checked. If again he is suspected to be a wrong user, then Presskey would finally lock down the application it is protecting.

7) Now to unlock such a system, the user may be required to scan his fingerprint on a connected device. The device application would be internally connected to the Presskey system, which on recognizing the fingerprint would unlock the application which was previously locked down.

To unlock an application in cases of imposter recognition, if the system has to resolve to physical biometrics only, then the idea of building an E-biometric system is rendered pointless. But possibly there exists no other way to resolve such a problem rather than using methods such as sending an OTP to the actual user's phone, receiving a fingerprint scan from its phone, sending an authentication mail, etc. Amongst all these *false positives* the best solution would be to employ Google authenticator.

A brief on algorithms that are helpful to solve the authentication problem in KDA:

1) Immune algorithms: Artificial immune systems (AIS) are a class of computationally intelligent systems inspired by the principles and processes of the vertebrate immune system. The algorithms are typically modeled after the immune system's characteristics of learning and memory for use in problem-solving. Immune algorithms can be classified into two categories as: Positive selection and Negative selection algorithms. Study shows that negative selection algorithms are the most used in intrusion detection [13]. The algorithms are:

Negative selection:

- V-detector [14]
- CRNS [15]

Positive selection:

- Self-detector [16]

A biometric based authentication system can be evaluated using either a genuine test or an impostor test, described as follows: - The genuine test (or False Rejection Rate (FRR)): the percentage of valid inputs which are incorrectly rejected. It is denoted by the number of incorrectly rejected attempts divided by the total attempts of legitimate users who try to access the system. - The impostor test (or False Acceptance Rate (FAR)): the percentage of invalid inputs which are incorrectly accepted. It is denoted by the number of incorrectly accepted attempts divided by the total attempts of impostors who try to access the system.

2) Ensemble of adaptive algorithms: Ensemble is defined as a method in which classification is performed on the outputs from many base classifiers [6]. There are several ensemble approaches:

- Majority voting, bagging, boosting [17]
- Stacking [18]
- K-nearest neighbor

In the majority voting several base classifiers performs classification on the input. If base classifiers returns positive result then it is treated as genuine user otherwise as impostor. In stacking, an additional classifier will receive the classification results from the base classifiers as input and then return the final classification result [6].

IV. NOVELTY/BENEFITS:

Biometric authentication is individual characteristic that's why it will be almost impossible by an imposter to penetrate the system. Keystroke dynamics based authentication verifies user from their typing pattern. To authenticate user based on their typing samples, it is required to find out the resemblance of a typing samples of user regardless of the text typed.

Keystrokes dynamics are a part of behavioural biometrics and are unique to a person to a large extent. If they are complemented with soft biometrics then the resulting authentication system becomes unbreachable.

V. CHALLENGES AND OPEN ISSUES

Such an authentication system in terms of existence is very limited; even those that are using it have not reviewed a positive result. Moreover the companies already using it; are deploying a system only making use of keystrokes dynamics. None of the company deployed a system complemented with soft biometrics.

Keystrokes dynamics are not enough by itself. They have to be used in conjunction with something because a keystrokes dynamics only system may be easier to breach. Plus such a system requires heavy software engineering in a sense that it should be able to continuously check for user authentication as the user types, and at the same time perform the necessary computations without slowing down the entire system. Also the data generated in the process would be too huge to be handled by any standalone system. It would be required to be backed up at least weekly, if not daily, on the cloud.

VI. CONCLUSION

Since the onset of the technological era and the boom of internet there have been identity crises, wherein people have been using fraudulent methods to fake identities. Today, there exists no authentication system which cannot be misleaded. For example, consider the following problems:

- Phone patterns can be snooped upon and so can be email passwords.
- Even 2-step verification can be cracked by stealing the phone of the user.
- Furthermore, there exists enough media depicting on how to copy fingerprint and iris scans.

Even with all the technological advancements the world has made, we haven't been able to come up with a cheap software based authentication system which is enough in itself by all means. More research is needs to be carried out to make such a dynamic authentication system that is efficient and cost effective.

Today, if employees leave their workstations unattended for lunch or some other purpose, within minutes of fraudulent conduct by some imposter they can be accused of being national terrorists. In such a crisis, the need of the hour is a self sufficient, adaptive, and impenetrable security system. Such a system is can be Presskey supplemented with facial recognition and mouse dynamics.

VII. REFERENCES

[1] K. Wangsuk, T.A. Amornkul, "Trajectory Mining for Keystroke Dynamics Authentication," Procedia Computer Science, Volume 24, 2013, Elsevier Press, pp 175-183, DOI: 10.1016/j.procs.2013.10.041

[2] A. K. Jain, R. Bolle, and S. Pankanti (editors), Biometrics: Personal Identification in a Networked Society, Kluwer Academic Publishers, 1999.

[3] D. Shanmugapriya, G. Padmavathi, "Virtual Key Force – A New Feature for Keystroke," J. International Journal of Engineering Science and Technology, Vol. 3 No. 10 October 2011, pp. 738-743.

[4] D. Stefan, D. Yao, "Keystroke-Dynamics Authentication against Synthetic.Forgeries," Proc 6th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom), IEEE Press, Oct 2010, DOI: 10.4108/icst.collaboratecom.2010.16

[5] P.H. Pisani, A.C. Lorena, "Emphasizing typing signature in keystroke dynamics using immune algorithms," Applied Soft Computing Volume 34, September 2015, pp 178–193, Elsevier Press, DOI: 10.1016/j.asoc.2015.05.008

[6] P.H. Pisani, A.C. Lorena, Andre C.P.L.F. de Carvalho, “Ensemble of adaptive algorithms for keystroke dynamics,” Brazilian Conference on Intelligence Systems IV, Nov 2015, DOI: 10.1109/BRACIS.2015.29

[7] P. Kang, S. Cho, “Keystrokes dynamics based user authentication using long and free text strings from various input devices,” Information Sciences 308:72-93, July 2015, DOI: 10.1016/j.ins.2014.08.070

[8] P. Bours, “Continuous keystroke dynamics: A different perspective towards biometric evaluation,” information security technical report xxx (2012), Elsevier Press, DOI: 10.1016/j.istr.2012.02.001

[9] R. Thanganayagam, A. Thangadurai, “Hybrid Model with Fusion Approach to Enhance the Efficiency of Keystroke Dynamics Authentication,” Proc 3rd International Conference on Advanced Computing, Networking and Informatics, Oct 2015, SIST vol 43, Springer Press, pp 85-96, DOI: 10.1007/978-81-322-2538-6_10

[10] P. Laurinen, P. Siirtola, J. Röning, “Efficient Algorithm for Calculating Similarity Between Trajectories Containing an Increasing Dimension,” Proc 24th IASTED international conference on Artificial intelligence and applications, Feb 2006, ACTA Press, pp 392-399.

[11] P.H. Pisani, A.C. Lorena, “A systematic review on keystroke dynamics,” J. Braz. Comput. Soc. 19(4) (2013) 573–587.

[12] G. Jagadamba, S.P. Sharmila, T. Gouda, “A Secured Authentication System Using an Effective Keystroke Dynamics,” Proc Emerging Research in Electronics, Computer Science and Technology, 2013. LNEE vol 248 Springer press, pp 453-460, DOI: 10.1007/978-81-322-1157-0_46

[13] S.X. Wu, W. Banzhaf, “The use of computational intelligence in intrusion detection systems: a review,” Appl. Soft Computing, vol 10 (1) 2010, Elsevier press, pp 1–35. https://doi.org/10.1016/j.asoc.2009.06.019

[14] Z. Ji, D. Dasgupta, “Revisiting negative selection algorithms,” Evol. Computing 15 (2) 2007, pp. 223–251.

[15] Z. Ji, D. Dasgupta, “Real-valued negative selection algorithm with variable-sized detectors,” Proc GECCO, LNCS vol 3102, Springer press, 2004, pp. 287–298, DOI: 10.1007/978-3-540-24854-5_30

[16] T. Stibor, J. Timmis, “Is negative selection appropriate for anomaly detection,” ACM GECCO 2005, pp. 321–328.

[17] T. G. Dietterich, “Ensemble methods in machine learning,” Proc 1st International Workshop on Multiple Classifier Systems, Springer-Verlag, 2000, pp. 1–15.

[18] L. I. Kuncheva, Combining Pattern Classifiers: Methods and Algorithms. Wiley-Interscience, 2004.