Keystroke Dynamics

P018 - term project, 2001

Petr Švenda <xsvenda@fi.muni.cz>
Masaryk University

Technology overview

History:

The original technology was derived from the idea of identifying a sender of Morse code using a telegraphy key known as the *"fist of the sender"*, whereby operators could identify senders transmitting a message by the rhythm, pace and syncopation of the signal taps. During World War II, the Army Signal Core identified that an individual keying rhythm on a telegraph key was quite unique. In the early-‘80s the National Science Foundation and the National Bureau of Standards in the United States conducted studies establishing that typing patterns contain unique characteristics that can be identified.

Technology details:

Keystroke dynamics are one of behavioral biometrics and look at the way a person types at a keyboard. Specifically, keyboard dynamics measures two distinct variables: *dwell time*, which is the amount of time you hold down a particular key and *flight time*, which is the amount of time it takes a person to move between keys. Keyboard dynamics systems can measure one's keyboard input up to 1000 times per second. Keystroke dynamics requires, as most biometrics technologies, a *reference template*. This involves several sessions of a person using a keystroke dynamic system so that the system can construct or build the reference template by detecting one's typing rhythms.
There are some other characteristics, which can be also used, such as *typing error frequency*, *frequency of using characters* etc. Typing characteristics can be extracted from a *structured text* (login, first and last name, password ...) same as from an *unstructured (free) text*, but results of verification/recognition are now much better for structured text.

Extraction (classification) methods:

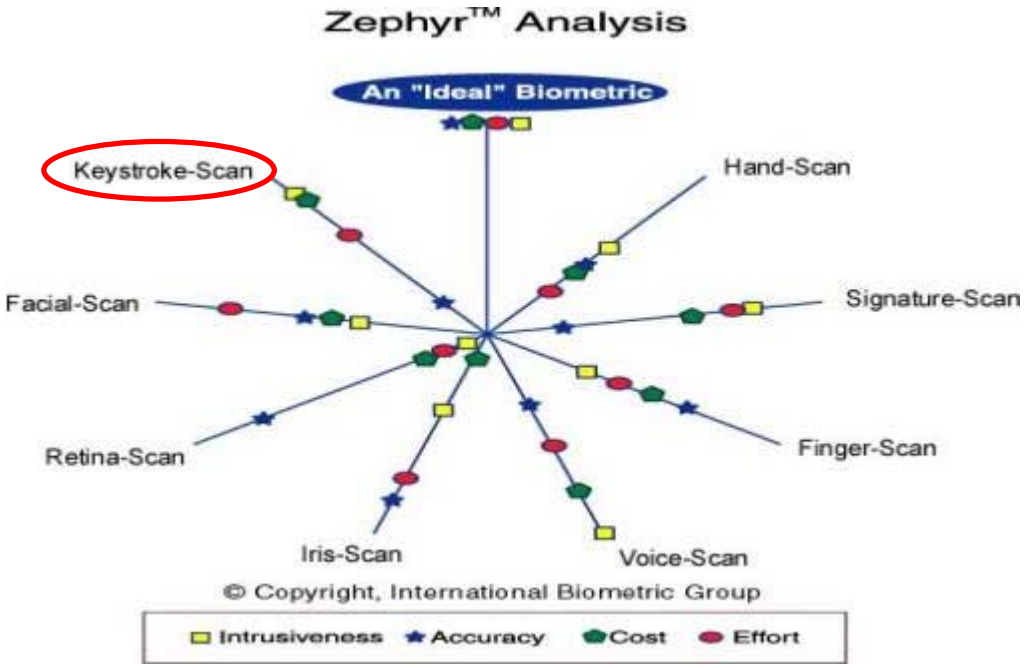
Reference template is constructed by using some possible extraction techniques, which calculate person typing characteristic. The same techniques are used during verification/recognition process. Different techniques give the different quality description of typing rhythm, what influent success of verification. During recognition, when database of user‘s reference template is large, some clustering methods can be used, such as typing speed for decrease in amount of templates used for verification. F. Monrose and A. Rubin in [1] extend previous research and report about success of Euclidean Distance Measure, Non-Weighted Probability and Weighted Probability Measure used as extraction (classification) techniques. Reference templates were created from few (exact number not specified in report) sentences typed by users and contain timing variables of most common features (th, he, nd, re, in, ing, are, is, ...). This were represented by n-dimensional feature vector same as tested text and compared using follows classification techniques:
Euclidean Distance Measure – simply Euclidean distance between vectors
Non-Weighted Probability – assign higher probabilities to timing variables of feature, which are more close to mean of this feature and lower probabilities to these are further away.
Weighted Probability Measure – also different probabilities are assigned to single features because they are come from larger sample set or have relatively higher frequency in written language (er, th, re have greater weights than qu or ts).

S. Cho, Ch. Han, D. H. Han, H.Kim in [2] present results of using neural networks as an extraction (classification) technique. *Multilayer perceptron networks* using back-propagation for learning give very good results on small databases, but neural networks has fundamental limitation. Each time a new user is introduced into database, the network must be retrained. In situations, where turnover of users is high, the down time associated with retraining can be significant.
Reference templates were created from 7 characters length password. User types this password 150 to 400 times during a period of several days, and the last 75 timing vectors were used for reference template. The remaining timing vectors were used to train the network. If any of its elements was larger than the upper 10%, however, the vector was classified as an outlier and discarded. Depending on the user, 6 to 50% of the training vectors were discarded.

	Acceptance		
	Known vs. Known	Free vs. Known	Free vs. Free
Euclidean Distance Measure	79.1 %	45.0 %	21.3 %
Non-weighted Probability	85.6 %	48.9 %	21.5 %
Weighted Probability Measure	90.7 %	44.1 %	23.0%
Neural Networks	FRR = 1% when FAR = 0%		

Known vs. Known – both compared characteristics are extracted from structured text (but different)
Free vs. Known – one characteristic is extracted from structured and next from unstructured (free) text
Free vs. Free – both compared characteristics are extracted from unstructured text

Compare with other biometrics:
source: Zephyr Analysis [3]



Advantages of technology:
1. The ultimate goal is ability to continually checking the identity of a person as they type at a keyboard.
2. Neither enrolment nor verification affect the regular work flow because the user would be tapping needed text anyway. Easy to use for example with login and password during logon process.

3. Unlike other biometrics system, keystroke dynamics is almost free. The only hardware required is the keyboard.
4. Time to training of users is minimal and ease of use is very high.
5. Public acceptability is very high. There are no prejudices such in case of criminal pattern in fingerprint verification or discomfort such as retina pattern scanning.

ad 1.: The ideal scenario for using this feature is accessing highly restricted documents or executing tasks in environment, where user must be „alert" all the times (for example air traffic control). Keystroke dynamics can be used here to detect user's uncharacteristic typing rhythm (changed by drowsiness, fatigue, ...) and can notify third parties.

ad 1., 2.: If actual characters besides timing information are also used for analyses, privacy of users can be violated when this data are not used only for keystroke dynamic analysis. Over that there is no obvious reason to do this, passive monitoring will only work in environments, where users have no expectation of privacy.

Disadvantages of technology:

1. Keystroke dynamics are non-static biometrics same as for example voice. Experimental data suggest that median inter-key latency of expert typists is approximately 96 ms, while that of novice typists is near 825 ms. This can change quite fast during time, also one-hand typing (due to injury, ...), etc. can influent typing rhythm.
2. Low accuracy. Keystroke dynamics is one less unique biometrics.
3. Small commercial widespread of technology.

Applications:

Keystroke dynamics can be used very well in cooperation with other authentication methods, especially with login and password (structured text), which gain good security results. Now only one company, Net Nanny, works on commercial release of their product BioPassword.

There are many potentially areas for this technology, especially for its low cost and feature of continually checking. Limitations are mainly non-consist typists.

F. Monrose in [2] believes that keystroke dynamics can be theoretical used as possible attack to PGP, because random seed collected during key generation is calculated from user's typing. This can be weakness, if users typing characteristics are known. Monrose also reports, that there can be some differences between left-handed and right-handed users, but he has only small part of left-handed users in testing group to give some useful results.

References:

[1] F. Monrose, A. Rubin, Authentication via Keystroke Dynamics
http://avirubin.com/keystroke.ps

[2] S. Cho, Ch. Han, D. H. Han, H.Kim,
Web based Keystroke Dynamics Identity – Verification using Neural Network
http://icec.net/eclib/papers/specialissue/OCEC-S6/New-Sungzoon%20Cho%20Paper.html

[3] Zephyr Analysis, International Biometric Group
http://www.biometricgroup.com

BioPassword® - Net Nanny Software Inc.

http://www.biopassword.com/

Now, only one company provides commercial product using keystroke dynamics. The set of mathematical algorithms used for extraction of characteristics is called BioPassword and was developed between 1979 – 1985 in Standfort Research Institute and patented. All rights are now owned by Net Nanny Software Inc.
Their first product based on BioPassword is alternatively logon screen for Windows NT. The system access is allowed not only if login and password information match, but when user‘s keystroke dynamics match too.

Product description:

- Alternative logon screen for Windows NT.
- Combination of password and keystroke dynamics security, the user will be required to enter his/her login and password to gain access to computer.
- The user‘s electronics signature record (reference template) is developed when the user enrolls in the system by typing login or password, providing 15-18 samples.
  This procedure takes about 3 minutes.
- Reference template is stored in encrypted mode.
- System Administrator can:
  - add or remove users
  - program system to allow only certain number of access attempts, after which the computer will be locked
  - set length of non-use period before re-entry of password is required
  - reset the PC after it has been locked-up by an unauthorized user access attempt
  - adjust the tolerance spectrum of keystroke dynamic‘s measurement with respect to „tightness or looseness“

BioPassword Demo:

I tested only this demo which only show the possibilities of full version.
First, system administrator must create his profile typing 10 entries (First Name, Last Name, Password), after successful login, he can set :

- **security level** (1-10) – tolerance to accept users keystroke dynamics
- **number of rejected logs before „possible imposter“** – notify third party, ...
- **number of rejected logs before program shutdown** – lock computer, ...
- **number of valid entries for registering a signature** (6-10 recommended) - number of repeating login entries during registration of new user

After entering Log In entries, BioPassword Demo shows message which contain information about password match and dynamic signature match. Access is allowed only if both match.
Only 30 users profile can be stored (in full version unlimited).