K

Keystroke Dynamics and Graphical Authentication Systems

Sérgio Tenreiro de Magalhães
University of Minho, Portugal

Henrique M. D. Santos
University of Minho, Portugal

Leonel Duarte dos Santos
University of Minho, Portugal

Kenneth Revett
University of Westminster, UK

INTRODUCTION

In information systems, authentication involves, traditionally, sharing a secret with the authenticating entity and presenting it whenever a confirmation of the user's identity is needed. In the digital era, that secret is commonly a user name and password pair and/or, sometimes, a biometric feature. Both present difficulties of different kinds once the traditional user name and password are no longer enough to protect these infrastructures, the privacy of those who use it, and the confidentiality of the information, having known vulnerabilities, and the second has many issues related to ethical and social implications of its use (Magalhães & Santos, 2005).

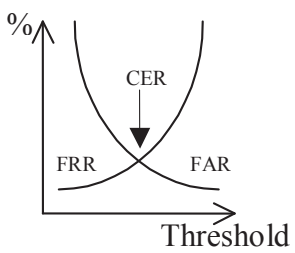
Password vulnerabilities come from their misuse that, in turn, results from the fact that they need to be both easy to remember, therefore simple, and secure, therefore complex. Consequently, it is virtually impossible to come up with a good password (Wiedenbeck, Waters, Birget, Brodskiy, & Memon, 2005). On the other hand, once users realize the need for securing their authentication secrets, even fairly good passwords become a threat when the security policies (if at all existing) fail to be implemented. The results of an inquiry made by the authors in 2004 to 60 IT professionals show that, even among those that have technical knowledge, the need for password security is underestimated (Magalhães, Revett, & Santos, 2006). This is probably one of the reasons why the governments increased their investment in biometric technologies after the terrorist attack of 9/11 (International Biometric Group [IBG], 2003).

The use of biometric technologies to increase the security of a system has become a widely discussed subject, but while governments and corporations are pressing for a wider integration of these technologies with common security systems (like passports or identity cards), human rights associations are concerned with the ethical and social implications of their use. This situation creates a challenge to find biometric algorithms that are less intrusive, easier to use, and more accurate.

The precision of a biometric technology is measured by its false-acceptance rate (FAR), which measures the permeability of the algorithm to attacks; its false-rejection rate (FRR), which measures the resistance of the algorithm to accept a legitimate user; and its crossover error rate (CER), the point of intersection of the FAR curve with the FRR curve that indicates the level of usability of the technology (Figure 1). For a biometric technology to be usable on a stand-alone base, its CER must be under 1%. As an algorithm becomes more demanding, its FAR is lower and its FRR is higher. Usually the administrator of the system can define a threshold and decide what the average FAR and FRR of the applied algorithm will be according to the need for security, which depends on the risk evaluation and the value of what is protected; also, the threshold can be, in theory, defined by an intrusion detection system (software designed to identify situations of attack to the system).

Establishing the error rates of a biometric technology is a complex problem. Studies have been made to normalize

Figure 1. Crossover error rate



their evaluation, but the fact is that the results are strongly dependent on the number of individuals involved in the process and, what is worst, on who is chosen. This means that, even with a large amount of data collected, the results can be very different if we change the evaluated group. The lack of trust in the precision evaluation methodologies and values is one of the reasons why the human rights associations are opposing the generalization of use of biometric technologies and their acceptance as standards for authentication procedures (Privacy International, Statewatch, & European Digital Rights, 2004). Even so, in an inquiry made by Epaynews (http://www.epaynews.com), 36% of users stated that they would prefer to use biometric authentication when using credit cards, a value only comparable to the use of personal identification numbers (PINs) and much higher than the 9% of authentication obtained by signature.

Considering all the advantages and disadvantages of biometric procedures, it seems that the only way is to allow the user a choice. Being so, the traditional password systems must be enhanced both in the biometrical way and in another completely different way. On the biometric component we propose keystroke dynamics, a biometrical authentication algorithm that tries do define a user's typing pattern and then verifies in each log-in attempt if the pattern existing in the way the password was typed matches the user's known pattern; it is the only biometric technology that can be used with the existing log-in and password systems without requiring any extra hardware. On the nonbiometric component, we propose the use of a graphical authentication system, a log-in system that verifies the user's knowledge of specific images or parts of images to grant or deny successful log-in, because it has been proven that it provides a wider key space and because it can be used to generate complex secret strings from simple passgraphs (the user's secret code to access a system protected by a graphical authentication system, constituted by a sequence of points where the user must click in order to obtain a successful log-in).

BACKGROUND

Keystroke Dynamics

As in many other problems, there have been two different approaches to the challenge of finding an algorithm for keystroke dynamics that minimizes the CER: machine learning and deterministic algorithms.

Among the solutions based on machine learning, we can find the work presented by Ord and Furnell (2000) that tested this technology with a 14-person group to study the viability of applying it to the simple use of PINs typed on a numeric pad. Unfortunately, the results suggest that, for large-scale use, the technology is not feasible. Deterministic algorithms have been applied to keystroke dynamics since the late '70s. In 1980, Gaines et al. (1980) presented a report on the study of the typing patterns of seven professional typists. The small number of volunteers and the fact that the algorithm is deducted from their data and not tested for other people later results in lower confidence in the FAR and FRR values presented. However, the method used to establish a pattern was a breakthrough: the study of the time spent to type the same two letters (digraph) when together in the text. Since then, many algorithms based on algebra and on probability and statistics have been presented. Joyce and Gupta presented in 1990 an algorithm to calculate a value that represents the distance between acquired keystroke latency times and correspondent times previously stored. In 1997, Monrose and Rubin used the Euclidean distance and probabilistic calculations based on the assumption that the latency times for one digraph exhibits a normal distribution. Later in 2000, they also presented an algorithm for identification based on the similarity models of Bayes, and in 2001 they presented an algorithm that uses polynomials and vector spaces to generate complex passwords from a simple one using the keystroke pattern (Monrose et al., 2001).

In 2005, Magalhães, Revett, and Santos presented an improvement of the Joyce and Gupta algorithm and tested it with 170.391 attacks to 143 patterns, obtaining a 0% FAR with an FRR of 26%, and an estimated CER below 5%.

Graphical Authentication Systems

A graphical authentication system is a log-in system that verifies the user's knowledge of specific images or parts of images to grant or deny successful log-in. Greg Blonder (1996) was the first to describe graphical passwords, presenting in a United States patent a system that would allow users to choose a picture, the number of regions to be clicked, and their sizes and positions. Since then, many variations of this system were presented and images have gained their way into the authentication processes.

Among the most popular graphical authentication systems, we find Passfaces™ from the Passfaces Corporation (2005), a commercial system where the user chooses a previously selected face from a set of faces and repeats this process for different faces in different sets for a defined number of times. However, being popular does not imply being secure, and a study of the users' choices demonstrated that they are, in some cases, similar for all users. For instance, 10% of the passwords of males could have been guessed with only two attempts (Davies, Monrose, & Reiter, 2004).

The déjà vu scheme involves a matrix of m images in a set, where n images are part of the user's portfolio, previously chosen from a set of proposed images. The user must identify those n images to log in.

The draw-a-secret (DAS) scheme is a graphical authentication system with an approach completely different. In DAS, the user draws something over a grid that becomes the

authentication secret. This system has been implemented with success on PDAs (personal digital assistants) and further studies will be made to analyse the users' choices and acceptance (Jermyn, Mayer, Monrose, Reiter, & Rubin, 1999).

In the visual identification protocol (VIP) several possibilities were created. From a set of 10 predefined images the user chooses 4 placed on the same position and typed in the same order (VIP1), or placed in random positions (VIP2). VIP3 is a process where four of the eight images existing in the user's portfolio are displayed along with 12 distractors, and the user must identify them in no particular order. The studies showed that the most common errors associated with VIP1 and VIP2 were related to bad sequences, where the identified images are correct but selected in the wrong order, and in VIP3 most of the errors were due to the wrong identification of the images, for instance, any flower being considered as the chosen flower (de Angeli, Coventry, Johnson, & Coutts, 2003).

In 2006, Magalhães et al. presented a graphical authentication system that included letters and numbers in images with the objective of allowing PDA users to click their user names in an easy way. From that system they discovered that the selection of the image and the rules that control the choice of passgraphs are critical factors in the success of the implementation of this kind of system. In particular, they found that users have a common tendency to choose the first available images, and that the use of images with corners and the existence of letters placed in a row create serious vulnerabilities to the system. Eyes are also a common choice and should be avoided. Therefore, the results suggest the use of images without corners, like nature images, cut in a round form. If the choice of keeping the letters is made, they must be placed in a random way throughout the images. Another dangerous tendency is the use of passgraphs constituted by regions placed in the same row or in the same column, therefore the system must reject the choice of passgraphs that meet this criteria, forcing the users to navigate inside the image by demanding the use of at least two different rows and two different columns.

ENHANCEMENT OF LOG-IN AND PASSWORD SYSTEMS

Since most of the existing systems trust passwords to provide access control and considering that passwords are not enough, we propose the enhancement of this process by adding a new module to the authentication system. This module gives two options to the user: a password with biometric control (keystroke dynamics) or a passgraph. If the user chooses the password system, he or she will be prompted (Figure 2) to enter the password several times in order to establish a pattern (this is called the enrollment process), and the everyday authentication process is, from the user's point of view, exactly the same as it was before the introduction of our module.

Each time that a user enters a password for authentication, the window captures both the characters stroked and the times between successive actions (pressing a key, releasing the key, pressing another key, and so on). The module verifies if the sequence of times matches the stored pattern (locally or in a portable device, like a smart card) and if (and only if) it does, the sequence of characters is sent to the original password authentication system that will verify correctness and allow or deny access to the user. Therefore, we have introduced another layer of security (biometrics) without any extra effort or equipment.

If the user chooses to use a passgraph, avoiding the biometrics component, he or she will have to choose several positions in an image. These positions, clicked in the same sequence, will be the secret access code of that user. Figure 3 shows a possible authentication window with a place to choose one of several possible images (in this case, the choice Mozart is the one that is active) and a place to enter the user name (in this case, the student's identification number). Nevertheless, this image would not be a good choice since it is not compliant with the best procedures in image selection for authentication, as described before.

Each time a user enters a passgraph, the sequence of clicks is transformed by a unidirectional function into a complex string that is passed into the original password field of the hosting system. In this way, we have obtained a simple and easy-to-memorize way of having extremely complex passwords.

As a last remark, one should notice that passgraph-area technology is very vulnerable to eavesdropping and, therefore, are more suitable for access made in private environ-

Figure 2. Keystroke-dynamics enrollment window

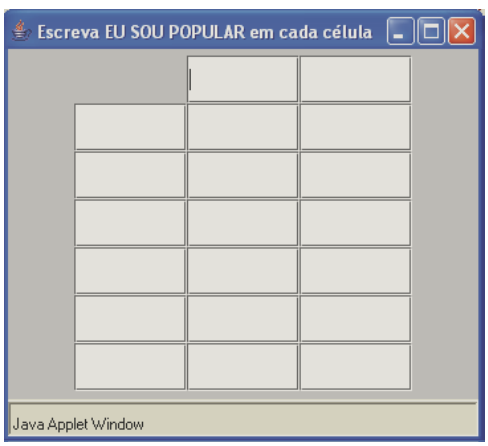
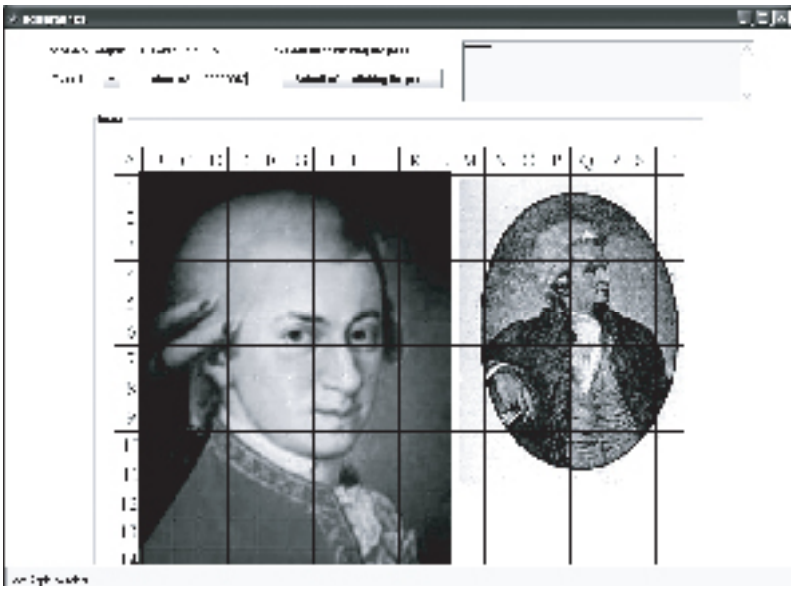


Figure 3. Passgraph authentication window



ments or on small portable devices, like smart phones or PDAs.

FUTURE TRENDS

Future work in this field will focus on improving the algorithms for keystroke dynamics, namely by combining the recent results provided by the artificial intelligence systems with the existing statistical algorithms.

Concerning passgraphs, studies are needed to improve the algorithms that convert the sequence of clicks into a sequence of characters and to improve the quality of the guidelines for image selection. This technology can also be integrated with other information security technologies in order to maximize their potential. On the other hand, artificial intelligence techniques can also be used to understand further more the use of secret codes in order to improve the quality of the proposed systems.

CONCLUSION

In conclusion, we can say that the technology has achieved a way to overcome, at least to a certain point, the entropy generated by users that continues to proceed in a way that is not the most efficient concerning security. Assuming human behaviour as a fact, ways were found to achieve best practices in security (like complex passwords) from the normal and traditional practices of users.

We have verified that keystroke dynamics and graphical authentication systems can, when used together, improve the security of the traditional log-in and password systems without adding significant complexity to their use and avoiding the ethical problems generated by biometrics when they are presented not as a choice but as an imposition. In fact, not only do these systems not present any ethical problems (when used together and leaving to the user the choice of which one to use), they can even provide a good use of the digital authentication processes by allowing those that cannot read or write (and therefore cannot use a password) to use the system, a matter especially relevant in the third-world countries that are now embracing new technologies, for instance, in electoral processes.

REFERENCES

Blonder, G. E. (1996). *Graphical password.*

Davies, D., Monrose, F., & Reiter, M. K. (2004). *On user choice in graphical password schemes.* Paper presented at the 13th USENIX Security Symposium.

de Angeli, A., Coventry, L., Johnson, G. I., & Coutts, M. (2003). Usability and user authentication: Pictorial passwords vs. PIN. In P. T. McCabe (Ed.), *Contemporary ergonomics 2003* (pp. 253-258). London: Taylor & Francis.

Gaines, R., et al. (1980). *Authentication by keystroke timing: Some preliminary results* (Rand Report No. R-256-NSF). Rand Corp.

International Biometric Group (IBG). (2003). *The biometric industry: One year after 9/11*. Retrieved November 2004 from http://www.biometricgroup.com/reports/public/reports/9-11.html

Jermyn, I., Mayer, A., Monrose, F., Reiter, M. K., & Rubin, A. (1999). *The design and analysis of graphical passwords*. Paper presented at the Eighth USENIX Security Symposium, Washington.

Joyce, R., & Gupta, G. (1990). Identity authorization based on keystroke latencies. *Communications of the ACM,* 33(2), 168-176.

Magalhães, S. T., Revett, K., & Santos, H. D. (2005). *Password secured sites: Stepping forward with keystroke dynamics.* Paper presented at the IEEE International Conference on Next Generation Web Services Practices (NweSP'05), Los Alamitos, CA.

Magalhães, S. T., Revett, K., & Santos, H. D. (2006). *Critical aspects in authentication graphic keys.* Paper presented at the International Conference on I-Warfare and Security, MD.

Magalhães, S. T., & Santos, H. D. (2005). An improved statistical keystroke dynamics algorithm. In *Proceedings of the IADIS MCCSIS 2005*.

Monrose, F., & Rubin, A. D. (1997). Authentication via keystroke dynamics. In *Proceedings of the Fourth ACM Conference on Computer and Communication Security*, Zurich, Switzerland.

Monrose, F., & Rubin, A. D. (2000). Keystroke dynamics as a biometric for authentication. *Future Generation Computing Systems (FGCS) Journal: Security on the Web.*

Monrose, F., et al. (2001). Password hardening based on keystroke dynamics. *International Journal of Information Security*.

Ord, T., & Furnell, S. M. (2000). User authentication for keypad-based devices using keystroke analysis. In *Proceedings of the Second International Network Conference: INC 2000*, Plymouth, United Kingdom.

Passfaces Corporation. (2005). *The science behind Passfaces.* Retrieved September 2005 from http://www.passfaces.com

Privacy International, Statewatch, & European Digital Rights. (2004). *An open letter to the ICAO: A second report on "Towards an International Infrastructure for Surveillance of Movement."* Retrieved from http://www.privacyinternational.org

Wiedenbeck, S., Waters, J., Birget, J. C., Brodskiy, A., & Memon, N. (2005, July 25-27). *Authentication using graphical passwords: Basic results.* Paper presented at Human-Computer Interaction International (HCII 2005), Las Vegas, NV.

KEY TERMS

Authentication: It is the process of verifying the identity alleged by a user who tries to gain access to a system.

Collaborative Biometric Technology: It is a biometric authentication technology that requires the user's voluntary and intended participation in the process. It opposes the stealth biometric technologies that can be used without the user's consent.

Crossover Error Rate (CER): Authentication algorithms need to simultaneously minimize permeability to intruders and maximize the comfort level, therefore they have to be both demanding and permissive. This contradiction is the base for the optimisation problem in authentication algorithms, and the measure of success for the overall precision of an algorithm and its usability is the CER, the value obtained at the threshold that provides the same false-acceptance rate and false-rejection rate.

False-Acceptance Rate (FAR): This rate is a measure of the permeability of an authentication algorithm. It is calculated by dividing the number of the intruder's successful log-in attempts by the total number of the intruder's log-in attempts.

False-Rejection Rate (FRR): This rate is a measure of the comfort level of an authentication algorithm. It is calculated by dividing the number of unsuccessful attempts made by legitimate users by the total number of legitimate log-in attempts.

Graphical Authentication System: It is a log-in system that verifies the user's knowledge of specific images or parts of images to grant or deny successful log-in.

Identification: It is the process of discovering the identity of a user who tries to gain access to a system. It differs from authentication because in the identification process, no identity is proposed to the system, while in authentication, an identity is proposed and the system will only verify if that identity is plausible.

Keystroke Dynamics: It is a biometrical authentication algorithm that tries do define a user's typing pattern and then verifies in each log-in attempt if the pattern existing in the way the password was typed matches the user's known pattern. Another application of keystroke dynamics, at least in theory, is the permanent monitoring of the user's typing pattern in order to permanently verify if the user that is typing is the legitimate owner of the system's account being used.

Passgraph: It is the user's secret code to access a system protected by a graphical authentication system. It is constituted by a sequence of points the user must click in order to obtain a successful log-in.

Stealth Biometric Technology: It is a biometric authentication technology that can be used without the user's consent. It opposes the collaborative biometric technologies that require the user's voluntary and intended participation in the process.

Threshold: It is the variable that defines the level of tolerance of an algorithm. It can be set to a more demanding value, raising the false-rejection rate and lowering the false-acceptance rate, or it can be set to a less demanding value, lowering the false-rejection rate and raising the false-acceptance rate.