Quantitative analysis of PIN choices: a contribution to the establishment of authentication requirements

José Carlos Carvalho[1], Maria José Magalhães[1], Vítor J. Sá[1,2], Sérgio Tenreiro de Magalhães[1,2]

[1] Faculty of Social Sciences, Catholic University of Portugal, Braga, Portugal
[2] ALGORITMI Research Center, University of Minho, Braga/Guimarães, Portugal
jc.carvalho28@gmail.com
stmagalhaes@braga.ucp.pt
mjmagalhaes@braga.ucp.pt
vitor.sa@braga.ucp.pt

Abstract: The authentication using a PIN number remains one of the most used ways to enter a system (mobile phone, ATM, etc.). Many people seem to dislike this form of authentication because they simply despise their use, placing unsafe PINs just because they have to put some. A large percentage of PINs use the combination 1234, use only one digit (example: 1111), or use the central line of the numerical keypad (2580). On the other hand there is some understanding because it is proven that remember strong passwords is a difficult task for humans, and the tendency is to choose the simplest ones.

This research had a sample of 306 participants and aimed to understand the preferred choice of the participants in relation to the number of digits used for a PIN number (a choice between four and/or six digits) and realized the amount of times that each of the available digits was used.

To this end it was developed a web-based tool for entering the data. This application was intended only to the data collection process, being the information processed further. Through this application, the user was asked to enter four and/or six-digit PINs. The method does not raise any doubt on the participants, which were informed about the anonymity and confidentiality of the data, and never they were asked to identify themselves. Participants were asked to use the PINs that they normally use in other contexts.

With the analysis of the data it was possible to understand the distribution of digits per position in a PIN, check which digits is more/less used in each position, and check which digit is more/less used regardless of its position. Among the conclusions it appears that the layout of the numeric keypad of the system influence the PIN choice.

Keywords: PIN, digits, security, authentication, system, keypad

1. Introduction

The omnipresence of interaction processes with computer systems gained increasing relevance today evoking the need for security mechanisms through user authentication via a secret key (Costa et al, 2005).

This concept of secret code or password is reflected in very vulnerable systems despite its wide use by users, as is the case of home banking (Neto & Bellinetti, 2008).

According to studies by Sasse and colleagues (2001), it is scientifically proven that the operation process with regard to remember strong passwords is an impossible task for humans and, therefore, there is always a tendency to choose the simplest ones (Sasse et al, 2001).

The authentication by a password is usually associated with low cost processes and can guarantee acceptable security levels since their use is well done (Pavezi, 2007). As this is not the case most of the time, and as this is the security mode associated with many information systems, the performance of an organization can be greatly affected, since the sharing of a secret between the user and the system may lead to the existence of numerous vulnerabilities (Magalhães, 2009).

In addition to the password paradox, in which passwords need to be both easy to remember, therefore simple, and secure, therefore complex, computer users have not yet completely realized 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. Even among those that have technical knowledge, the need for password security is underestimated (Magalhaes et al, 2006).

To address this problem it becomes indispensable to invest in a culture of security and provide training to computer system users (from an early age), so that these issues related to the incorrect use of passwords do not proliferate (Silva et al, 2007), and think about solutions that combine these with other innovative forms of authentication (Sá, 2013).

In this article we focus on the particular case of the use of Personal Identification Number (PIN), a numeric password, because the authentication using a PIN remains one of the most used ways to enter a system (mobile phone, ATM, etc.). Many people seem to dislike this form of authentication because they simply despise their use, placing unsafe PINs just because they have to put some. A large percentage of PINs use the combination 1234, use only one digit (example: 1111), or use the central line of the numerical keypad (2580). With this in mind, has come up the idea of trying to figure out which is the most common use people make of PINs, in order to contribute to the establishment of some authentication requirements suggesting "weak" passwords that should not be used.

Therefore, in the next section we present the methodology used for the data collection, in the following section we analyze the obtained data and, finally, we draw some conclusions.

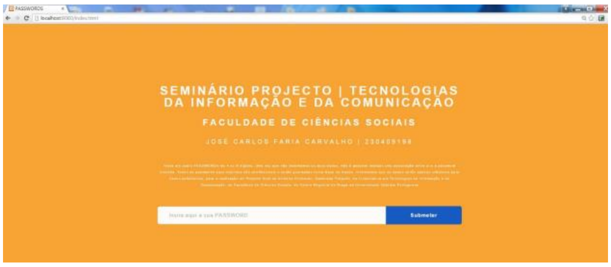
2. Methodology

This research had a sample of 306 participants and aimed to understand the preferred choice of the participants in relation to the number of digits used for a PIN number (a choice between four and/or six digits) and realized the amount of times that each of the available digits was used.

To this end it was developed a web-based tool for entering the data. This application was intended only to the data collection process, being the information processed further.

Through this application, the user was asked to enter four and/or six-digit PINs. The method does not raise any doubt on the participants, which were informed about the anonymity and confidentiality of the data, and never they were asked to identify themselves. Participants were asked to use the PINs that they normally use in other contexts.

The participants only had to open the application and enter the PIN they normally use in situations where an authentication to access a computer-based system is needed (Figure 1).



SEMINÁRIO PROJECTO | TECNOLOGIAS DA INFORMAÇÃO E DA COMUNICAÇÃO

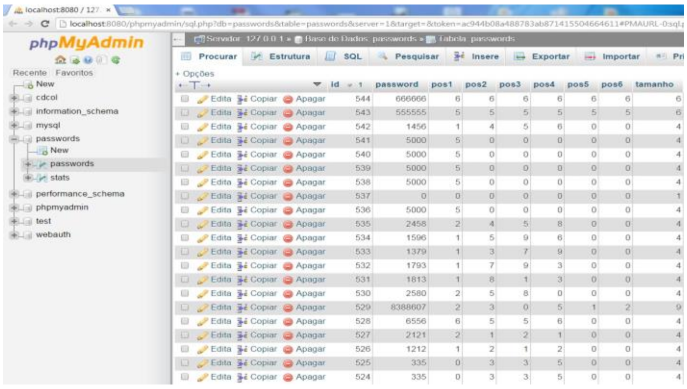
FACULDADE DE CIÊNCIAS SOCIAIS

JOSÉ CARLOS FARIA CARVALHO | 230409198

Insira aqui a sua PASSWORD

Figure 1: Web application for data collection

The mechanism for automatic data collection consisted in a simple web application programmed in PHP and the data were managed using MySQL, the free but sufficiently powerful and widely used database management system (Figure 2).



	id	password	pos1	pos2	pos3	pos4	pos5	pos6	tamanho
☐ Edita Copiar Apagar	544	666666	6	6	6	6	6	6	6
☐ Edita Copiar Apagar	543	555555	5	5	5	5	5	5	6
☐ Edita Copiar Apagar	542	1456	1	4	5	6	0	0	4
☐ Edita Copiar Apagar	541	5000	5	0	0	0	0	0	4
☐ Edita Copiar Apagar	540	5000	5	0	0	0	0	0	4
☐ Edita Copiar Apagar	539	5000	5	0	0	0	0	0	4
☐ Edita Copiar Apagar	538	5000	5	0	0	0	0	0	4
☐ Edita Copiar Apagar	537	0	0	0	0	0	0	0	1
☐ Edita Copiar Apagar	536	5000	5	0	0	0	0	0	4
☐ Edita Copiar Apagar	535	2458	2	4	5	8	0	0	4
☐ Edita Copiar Apagar	534	1596	1	5	9	6	0	0	4
☐ Edita Copiar Apagar	533	1379	1	3	7	9	0	0	4
☐ Edita Copiar Apagar	532	1793	1	7	9	3	0	0	4
☐ Edita Copiar Apagar	531	1813	1	8	1	3	0	0	4
☐ Edita Copiar Apagar	530	2580	2	5	8	0	0	0	4
☐ Edita Copiar Apagar	529	8388607	2	3	0	5	1	2	9
☐ Edita Copiar Apagar	528	6556	6	5	5	6	0	0	4
☐ Edita Copiar Apagar	527	2121	2	1	2	1	0	0	4
☐ Edita Copiar Apagar	526	1212	1	2	1	2	0	0	4
☐ Edita Copiar Apagar	525	335	0	3	3	5	0	0	4
☐ Edita Copiar Apagar	524	335	0	3	3	5	0	0	4

Figure 2: Database management system

3. Data analysis

In the process of data analysis it was made the division between users who entered one four-digit PIN and, beyond that, the users who also included a six-digit PIN.

Regarding the frequency of occurrence of the various digits, for PINs with 4 digits, the most used is digit 5. Curiously, in a PIN pad 5 is the digit that is in the center of the keyboard. The least used digits (regardless of its position) are 6 and 7 followed by 3 and 4 (Figure 3).

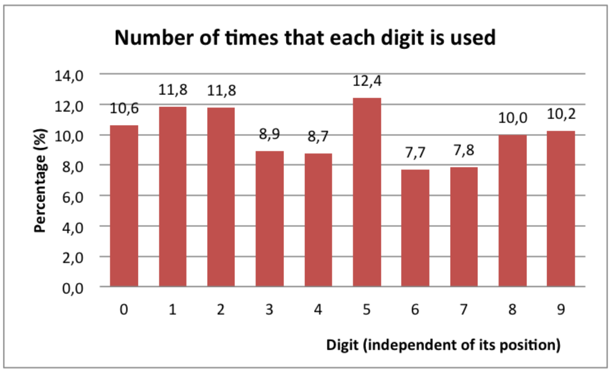


Figure 3: Number of times that each digit is used (4 digits)

For PINs with 6 digits, the most used are the digits 0 and 1. The least used digits (regardless of its position) are 7 and 8 (Figure 4).

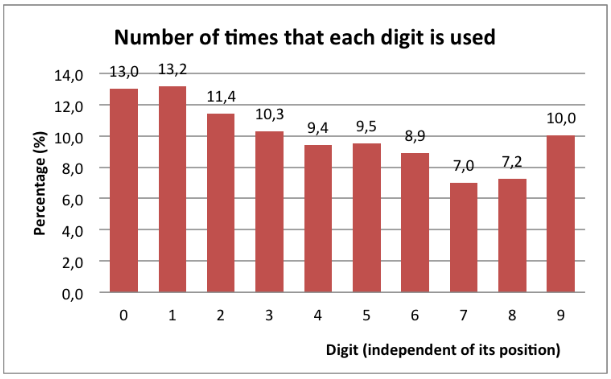


Figure 4: Number of times that each digit is used (6 digits)

Regarding the use of digits in the four possible positions, the digit 1 is the most used in the first position, with a percentage of 21.6%. The least used in the first position is the digit 3, with a percentage of only 4.6%.
In the second position the most used is the digit 5, with 17%, and the least used is the digit 7 with 5.9%.
In the third position the most used digit is the 3 with 12.1%, and the least used digit is the 4 with 7.2%.
In the fourth position the most used is the digit 2, with 12.7%, and the least used is the digit 3 with 7.8%

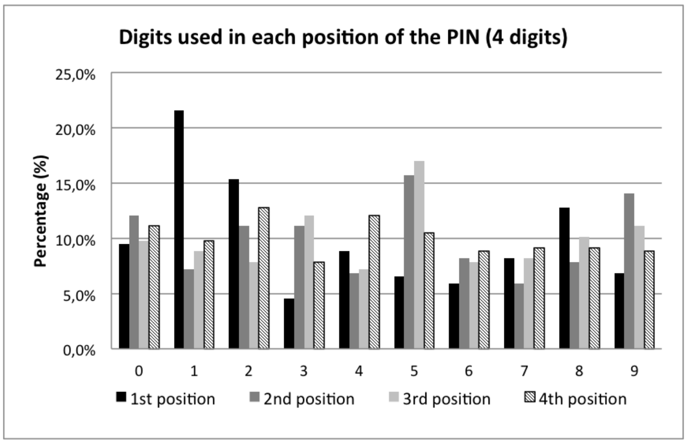


Figure 5: Digits used in each position of the PIN (4 digits)

Regarding the use of digits in the possible positions of a six-digit PIN, the results are the following.
In the first position the most used is the digit 1, with 23.6%, and the least used is the digit 5 with 3.7%.

In the second position the most used is the digit 0, with 20.9%, and the least used is the digit 7 with 3.1%.
In the second position the most used is the digit 0, with 17.8%, and the least used is the digit 8 with 4.7%.

In the fourth position the most used is the digit 1, with 17.3%, and the least used is the digit 7 with 4.2%
In the second position the most used is the digit 5, with 14.7%, and the least used is the digit 1 with 5.8%.
In the second position the most used is the digit 6, with 15.7%, and the least used is the digit 8 with 6.3%.
Globally, the most used digits are the 0, with 20.9%, and the 1, with 23.9%. The least used digits are the 5, with 3.7%, and the 7, with 3.1%.

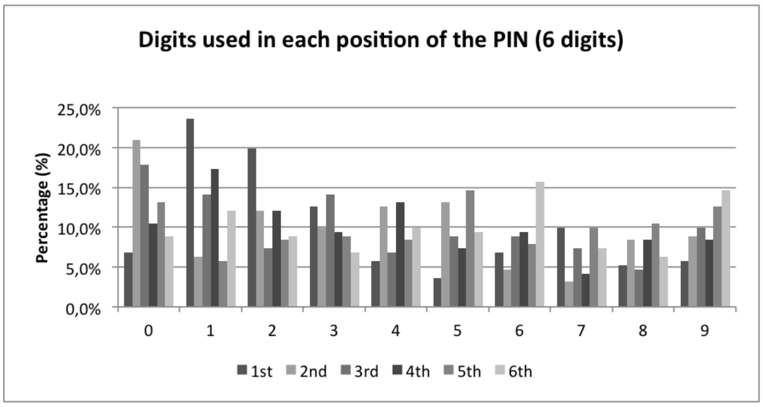


Figure 6: Digits used in each position of the PIN (6 digits)

One of the problems of PIN quality is the use of repeated digits. That occurs in 45.4% of the PINs with four digits and in 56.5% of the PINs with six digits.
Worse than having some repeated digits is repeat them in full. As mentioned before, some people use just one digit repeated in all the positions. The collected data prove that. In the PINs with four digits 8.8% have the same digit in all positions, being the combination with the digits 1 and 7 the most used (Table 1 and Figure 7).

Table 1: Absolute and relative frequency of PINs with all digits repeated (4 digits)

PIN	Absolut Frequency	Relative Frequency	%
0000	3	0,1250	12,50%
1111	4	0,1667	16,67%
2222	2	0,0833	8,33%
3333	3	0,1250	12,50%
4444	2	0,0833	8,33%
5555	2	0,0833	8,33%
6666	2	0,0833	8,33%
7777	4	0,1667	16,67%
8888	2	0,0833	8,33%
9999	3	0,1250	12,50%

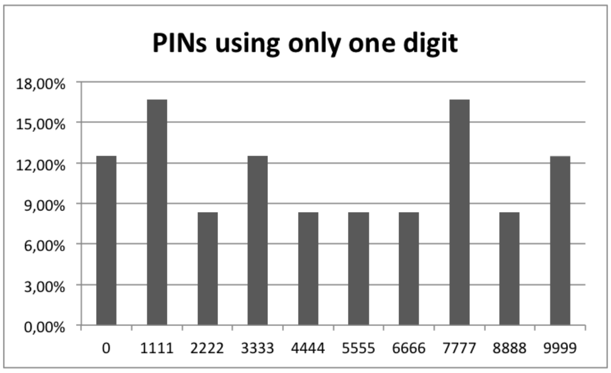


Figure 7: Frequency of PINs with all digits repeated (4 digits)

Curiously, perhaps because it is a larger sequence, the 6-digit PINs don't have combinations with 6 equal digits. About the use of sequential numbers, other misuse of PINs, found quite often, 2.94% of 4-digit PINs are sequences of numbers from which 66.67% correspond to the sequence 1234 (Table 2 and Figure 8).

Table 2: Absolut and relative frequency of PINs using a sequence of numbers (4 digits)

PIN	Absolut Frequency	Relative Frequency	%
0123	1	0,111	11,11%
1234	6	0,667	66,67%
5678	2	0,222	22,22%

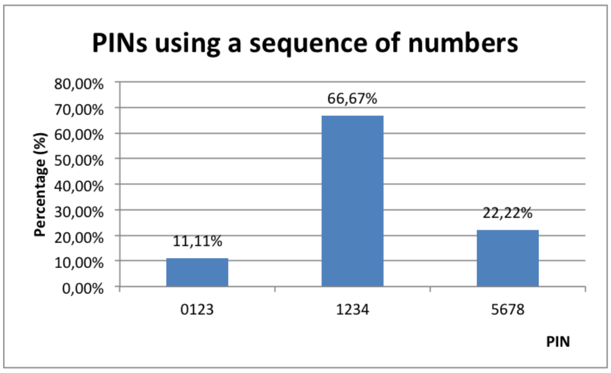


Figure 8: PINs using a sequence of numbers (4 digits)

Regarding 6-digit PINs, 5.24% of the PINs are sequences of numbers from which 90% correspond to the sequence 123456 (Table 3 and Figure 9).

Table 3: Absolut and relative frequency of PINs using a sequence of numbers (6 digits)

PIN	Absolut Frequency	Relative Frequency	%
123456	9	0,9	90,00%
345678	1	0,1	10,00%

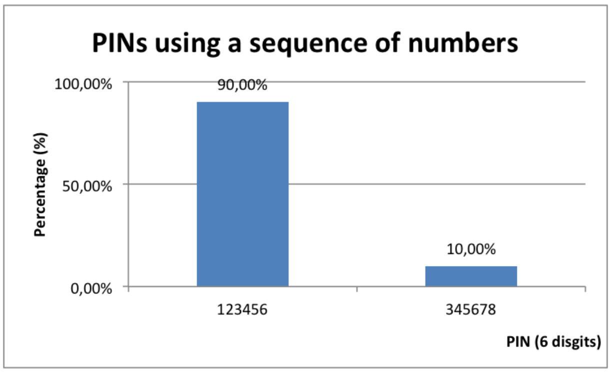


Figure 9: PINs using a sequence of numbers (6 digits)

4. Conclusion

With the advancement of technology there are very effective access control techniques, so the security problems result most of the time from human error. As a historical example, this was the cause when the Allies managed to break the security of the Enigma machine from the Nazis (McCallion, 2014).

The ideas we have about things are just guesses until they are studied rigorously. That's what we wanted to profess in this article regarding the PIN types used by the common people.

We know that there is a misuse of passwords, including the special case of the numerical ones (PINs). Having this kind of passwords a very widespread usage, it is important to understand the mistakes that are often made to define some authentication requirements. It was this contribution we wanted to give to the area, being this study a preliminary work.

In this research it must be pointed out the size of the sample that reached 306 elements. From these, 191 subjects inserted a six-digit password.

In addition to the information presented in the previous section, the collected data appear to induce that the numeric keypad of the system can influence the password choice, which encourages us to do in the future a further study to detect possible patterns and relationships between variables.

Acknowledgements

This work has been supported by FCT – Fundação para a Ciência e Tecnologia within the Project Scope UID/CEC/00319/2013.

References

Costa, C. R. D. N., Yared, G. F., Rodrigues, R. N., Yabu-Uti, J. B., Violaro, F., & Ling, L. L. (2005). Autenticação Biométrica via Dinâmica da Digitação em Teclados Numéricos. In XXII Simpósio Brasileiro de Telecomunicações–SBrT'05.

McCallion, J. (2014). "Enigma interview: hackers exploiting weak passwords is nothing new", *Computing*. Retrieved from http://www.pcpro.co.uk/computing/1000098/enigma-interview-hackers-exploiting-weak-passwords-is-nothing-new.

Magalhães, S., Revett, K. & Santos, H. (2006). Generation of Authentication Strings From Graphic Keys. *International Journal of Computer Science and Network Security*, Vol. 6, No. 3, pp. 240-246.

Magalhães, S. T. (2009). Estudo de viabilidade da utilização de tecnologias biométricas comportamentais na autenticação do cidadão perante os serviços electrónicos do Estado. Doctoral Thesis, University of Minho.

Neto, M. F., & Bellinetti, G. (2008). "A assinatura digital como prova de autoria do documento eletrônico", *Revista Em tempo*, Vol 7, No. 7.

Pavezi, R. S., de Macedo, D. D. J., Andrade, R., & von Wangenheim, A. (2009). Dinâmica da digitação aplicada a ambientes WEB. Laboratório de telemedicina na Universidade Federal de Santa Catarina (UFSC), Brasil.

Sá, V. J. (2013). Dinâmica gestual com condutividade da pele : uma abordagem multimodal para autenticação biométrica, Doctoral Thesis, University of Minho.

Sasse, M. A., Brostoff, S. & Weirich, D. (2001). "Transforming the 'Weakest Link' — a Human/Computer Interaction Approach to Usable and Effective Security", *BT Technology Journal*, Vol 19, No. 3, pp 122-131.

Silva, D. R. P., Stein L. M. (2007). "Segurança da informação: uma reflexão sobre o componente humano", *Ciências & Cognição*, Vol. 10, pp 46-53