Validation of the Authenticity of Printed Documents (June 2023)

António Ramos[1]

[1]Universidade de Aveiro, Aveiro, Portugal

Corresponding author: António Ramos (e-mail: ajframos@ua.pt).

ABSTRACT Knowledge and information are essential for a country's well functioning and the information can be about politics, health, the military and nuclear. Being that, inevitably, there are enemies to any country logically these will try their best to gain access to the information that's classified. In order to ensure that the high-classified information is kept safe and doesn't fall into the wrong hands, there must be a method put in use to guarantee their safety. The following thesis speaks on the issue and the potential solutions for classified documents that, after being printed and handed to a third party either credentialed or not allowed to see the document, lose their credibility therefore the veracity of the document cannot be proven.

INDEX TERMS Authenticity, Barcodes, Digital Documents, Electronic Documents, Integrity, Line Segments, Printed documents, Security, Watermark

I. INTRODUCTION

Each country has its own secrets and valuable information, such as military, political, or health-related, that they do not want to be disclosed to third parties or enemies. As Francis Bacon said, "Knowledge is power" [1], which demonstrates that there are people who have an insatiable desire to uncover the unknown about the enemy, hence the theft, destruction, and access to classified materials. Therefore, these materials require special protection to prevent leaks. This protection assumes that the organization responsible for ensuring the security of the country's classified materials has knowledge of the possession of the classified material, the geographical location of the material, where it should be located, and who has or had access to it. As mentioned earlier, it is necessary to prevent unauthorized access by third parties and enable the investigation of incidents or security breaches. This article addresses the issue and possible solutions for classified documents that, after being printed at a control station, can be presented to inappropriate third parties or accredited third parties who have no way of validating the information in the document.

To solve the problem of access by third parties, methods of visible and invisible watermarking have been studied to validate the integrity of the document. Additionally, a method of document validation based on metadata and document information has been chosen for development and testing.

This model assumes that the original documents are in PDF format, containing letters at specific positions and not bitmaps (image format). Validation will only be possible after scanning without OCR.

The metadata or metadata of the document consists of a set of letters obtained at the intersection of various randomly drawn line segments invisibly on the first page of the original PDF. This information is stored in protected databases in access-controlled buildings, where the visualization of document verification is only possible if the application has access to the database.

If the application successfully accesses the visualization, it will display the letters mentioned in the metadata at the position of the scanned document, thus ensuring the validation of the document's authenticity.

II. Algorithm for Marking and Analysis of Printed Documents

The proposed algorithm is divided into processing a document to generate and put the mark, and verifying the mark on document to analyze if the information is the same as the original.

A. Processing Documents

The logic for processing a document is described below:

1. Choose the Document to process;
2. Verify the chosen document. If it has been processed, it is necessary to choose another one. There is a distinction between processed and unprocessed files with an extension created based on the original file name "filename_watermark_dd_mm_yy_hh_m_ss". If it doesn't exist, proceed to insert the file's metadata into the database;
3. Obtain characters and positions from the Document. The algorithm will open a console to execute a Java

command to read the PDF file and extract its positions and characters;

4. Create points for line segments and calculate the intersection point. Using the 9 predefined positions, calculate the 3 new positions (r, l, b) with a range of integer numbers, where each x and y value for each position will be randomly chosen;
5. Create the barcode containing an identifier of the document, which is inserted in an internal or external database;
6. Insert the barcode into the duplicated file of the original, resulting in a new file. This file must be accepted or rejected by a user.

B. Verification of Printed Documents

The verification of printed documents consists at:

1. Choose the file to examine, which can be in electronic or digital PDF format. In both cases, the document is converted into an image for the placement of the letters present in the database;
2. Read the barcode to decode the information contained within it. If the file's identifier (ID) is in the database and the watermark has been accepted or rejected, certain metadata is returned to the user. The user performs a quick verification of a set of information about the document, such as the user who printed it and the current status;
3. Verify the document's integrity. If the verification fails or if the user wants to proceed regardless, the process involves comparing the letters in the document with those in the database. This is done by converting the PDF document into an image, placing the letters and their respective points, and then converting it back to PDF. In this approach, it was found that multiple intersection points could be inserted within the same letter, but with different values. This made it difficult to visualize the document, so duplicates were subsequently eliminated, resulting in multiple letters in the same area being displayed as a single letter, which ultimately led to the removal of duplicates.

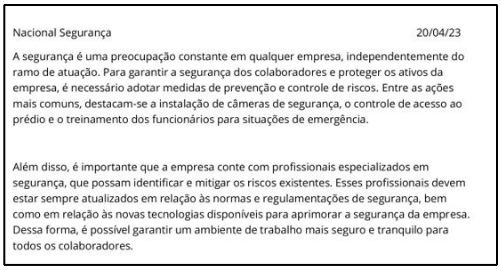
III. Tests

To test the algorithm, a random document is created with the following text (Figure 1). The algorithm processes the document inserting the barcode 39 (header) and barcode 128 (footer), due to the dimensions of the file, it can't show all the document.

A possibility of the integrity resulted by the process is shown in Figure 2. The text is in Portuguese, but it also works with English, what matters is that the letters are codded in utf-8 and Latin.

For testing the integrity of the document, the word "importante" is replaced, where "n" is associated in database, to "necessário" (Figure 3 - black arrow) demonstrating that the algorithm can detect replacing in that zone.

For testing the alteration of the text by removing one paragraph (Figure 4), resulting in changing of the points in y axis.

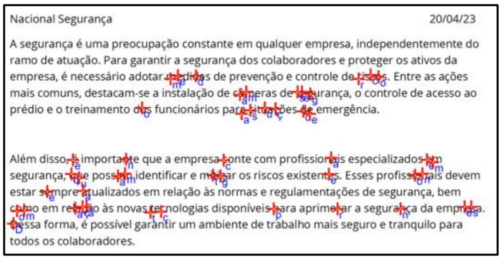


Nacional Segurança 20/04/23

A segurança é uma preocupação constante em qualquer empresa, independentemente do ramo de atuação. Para garantir a segurança dos colaboradores e proteger os ativos da empresa, é necessário adotar medidas de prevenção e controle de riscos. Entre as ações mais comuns, destacam-se a instalação de câmeras de segurança, o controle de acesso ao prédio e o treinamento dos funcionários para situações de emergência.

Além disso, é importante que a empresa conte com profissionais especializados em segurança, que possam identificar e mitigar os riscos existentes. Esses profissionais devem estar sempre atualizados em relação às normas e regulamentações de segurança, bem como em relação às novas tecnologias disponíveis para aprimorar a segurança da empresa. Dessa forma, é possível garantir um ambiente de trabalho mais seguro e tranquilo para todos os colaboradores.

Figure 1 - Original Document

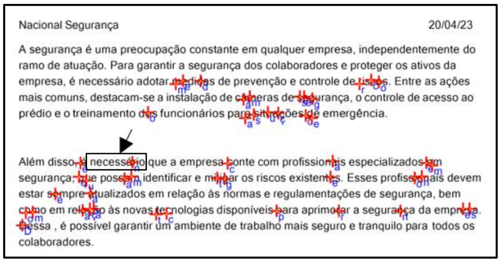


Nacional Segurança 20/04/23

A segurança é uma preocupação constante em qualquer empresa, independentemente do ramo de atuação. Para garantir a segurança dos colaboradores e proteger os ativos da empresa, é necessário adotar medidas de prevenção e controle de riscos. Entre as ações mais comuns, destacam-se a instalação de câmeras de segurança, o controle de acesso ao prédio e o treinamento dos funcionários para situações de emergência.

Além disso, é importante que a empresa conte com profissionais especializados em segurança, que possam identificar e mitigar os riscos existentes. Esses profissionais devem estar sempre atualizados em relação às normas e regulamentações de segurança, bem como em relação às novas tecnologias disponíveis para aprimorar a segurança da empresa. Dessa forma, é possível garantir um ambiente de trabalho mais seguro e tranquilo para todos os colaboradores.

Figure 2 - Integrity

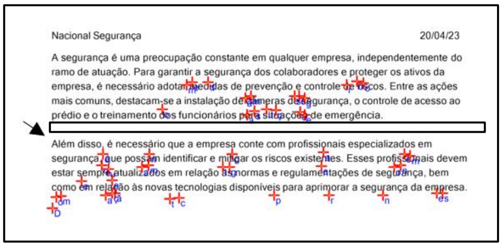


Nacional Segurança 20/04/23

A segurança é uma preocupação constante em qualquer empresa, independentemente do ramo de atuação. Para garantir a segurança dos colaboradores e proteger os ativos da empresa, é necessário adotar medidas de prevenção e controle de riscos. Entre as ações mais comuns, destacam-se a instalação de câmeras de segurança, o controle de acesso ao prédio e o treinamento dos funcionários para situações de emergência.

Além disso, é necessário que a empresa conte com profissionais especializados em segurança, que possam identificar e mitigar os riscos existentes. Esses profissionais devem estar sempre atualizados em relação às normas e regulamentações de segurança, bem como em relação às novas tecnologias disponíveis para aprimorar a segurança da empresa. Dessa , é possível garantir um ambiente de trabalho mais seguro e tranquilo para todos os colaboradores.

Figure 3 - Replace Word



Nacional Segurança 20/04/23

A segurança é uma preocupação constante em qualquer empresa, independentemente do ramo de atuação. Para garantir a segurança dos colaboradores e proteger os ativos da empresa, é necessário adotar medidas de prevenção e controle de riscos. Entre as ações mais comuns, destacam-se a instalação de câmeras de segurança, o controle de acesso ao prédio e o treinamento dos funcionários para situações de emergência.

Além disso, é necessário que a empresa conte com profissionais especializados em segurança, que possam identificar e mitigar os riscos existentes. Esses profissionais devem estar sempre atualizados em relação às normas e regulamentações de segurança, bem como em relação às novas tecnologias disponíveis para aprimorar a segurança da empresa.

Figure 4 - Remove Paragraph

IV. Conclusion

The algorithm developed satisfies the initial objectives of the creation of a barcode to quickly validate a document and create a watermark capable of guaranteeing that document is secure against alteration. The algorithm only works for the first page of large documents, if need it, can be upgraded by inserting barcodes and reading all the file, with the limitation of an increase in processing time.

V. REFERENCES

[1] "Conhecimento é Poder - Significado, origem e autores." https://conceitosdomundo.pt/conhecimento-e-poder/ (accessed Apr. 31, 2023).

ANTÓNIO RAMOS was born in Viseu, Portugal at 1998. In 2020 graduated in Computer Science at IPV, Viseu, Portugal. From 2020 to 2023 – master's degree on Computer Science.