European Research Studies Journal
Volume XXVI, Issue 4, 2023

pp. 718-726

Securing Digital Copies of the Documents to Ensure Documents' Integrity

Submitted 02/10/23, 1st revision 10/10/23, 2nd revision 20/11/23, accepted 30/11/23

Grzegorz Szyjewski[1]

Abstract:

Purpose: *This article addresses the critical issue of securing the electronic documentation flow, focusing on maintaining document content integrity without compromising ease of use in everyday processes involving documents.*
Design/Methodology/Approach: *The proposed electronic document verification method is versatile and applicable to both electronic and printed documents. The described approach has been meticulously designed and rigorously tested within a certification environment, where the need for confirming users' skills and knowledge is paramount.*
Findings: *In the contemporary landscape dominated by the prevalence of electronic markets and document flow, the potential susceptibility of electronic documentation to forgery often goes unnoticed. The transition to electronic formats renders documentation more vulnerable to unauthorized modifications, leaving no discernible traces of tampering. The advancement of Information Technology (IT) techniques further facilitates the facile alteration of electronic documents, irrespective of whether they are in the form of scanned images stored as bitmap files or in Portable Document Format (PDF). Although commonly perceived as tamper-resistant, these formats are not immune to manual document file manipulations, presenting an avenue for fraudulent activities.*
Practical Implications: *Established techniques such as Private Key Infrastructure (PKI) or cryptographic hash functions for generating document checksums can be employed for this purpose. While effective in preserving content integrity, these methods pose a significant challenge in terms of usability for the average computer user. The study demonstrates the effectiveness of the QR codes usage in simplifying the verification process while maintaining credibility of the document.*
Originality/value: *The article presents a detailed description of the approach, focusing on secure QR codes with extended unique identifiers for efficient verification. It further examines the implementation of this approach within the author's ICT system within International Computer Driving License (ICDL) Certification Program in Poland.*

Keywords: *Electronic documents integrity, credibility, document digital alternatives, certificate verification, online application.*

JEL codes: *M11, M15, O32.*

Paper type: *Research article.*

[1]*Ph.D., University of Szczecin, Institute of Management, grzegorz.szyjewski@usz.edu.pl;*

1. Introduction

In the contemporary landscape dominated by the prevalence of electronic markets and document flow, the potential susceptibility of electronic documentation to forgery often goes unnoticed. The transition to electronic formats renders documentation more vulnerable to unauthorized modifications, leaving no discernible traces of tampering.

The advancement of Information Technology (IT) techniques further facilitates the facile alteration of electronic documents, irrespective of whether they are in the form of scanned images stored as bitmap files or in Portable Document Format (PDF). Although commonly perceived as tamper-resistant, these formats are not immune to manual document file manipulations, presenting an avenue for fraudulent activities.

The evolving landscape of Artificial Intelligence (AI) technology exacerbates the risk of undetectable electronic document modifications. Consequently, it becomes imperative to augment electronic documents with robust verification mechanisms that ensure document integrity.

Established techniques such as Private Key Infrastructure (PKI) or cryptographic hash functions for generating document checksums can be employed for this purpose (Putro, 2017; Knott, 1975; Bellare *et al.*, 1996). While effective in preserving content integrity, these methods pose a significant challenge in terms of usability for the average computer user.

This article addresses the critical issue of securing the electronic documentation flow, focusing on maintaining document content integrity without compromising ease of use in everyday processes involving documents. The proposed electronic document verification method is versatile and applicable to both electronic and printed documents. The described approach has been meticulously designed and rigorously tested within a certification environment, where the need for confirming users' skills and knowledge is paramount.

2. Literature Review

Electronic documents are commonly regarded as editable files that can be accessed and modified using software applications such as text editors or spreadsheets. The COVID-19 pandemic, which necessitated widespread isolation measures, prompted numerous organizations to transition from traditional paper-based documentation to electronic formats (Pranggono and Arabo, 2021).

A prevalent format for electronic documents during this period was the scanned reproduction of physical documents (Buckland, 1997). Typically, these scanned documents included personal signatures, manually placed on the paper prior to scanning (Turro, 2008).

Even as the pandemic subsides, the electronic document format generated through scanning has maintained its popularity. Many institutions continue to accept documents in this electronic form, submitted through electronic communication channels.

Scanned document can be considered an electronic document (Goodrum *et al.*, 2020). An electronic document is essentially any form of information that is stored and processed in a digital format, and a scanned document is a digital representation of a physical document. When the document is scanned it is converted from a hard copy (paper document) into a digital format, usually an image file such as a JPEG or PDF.

This digital version can be stored, transmitted, and manipulated electronically. While the scanned document itself is a digital file, it's important to note that the content within the document is essentially an image, and the text within it is not inherently machine-readable (Stromer *et al.*, 2018). The distinction between different types of electronic documents becomes important in the context of document management, security, and interoperability.

For example, text in a scanned document may not be easily searchable or editable without additional processing such as optical character recognition (OCR), which converts the image-based text into machine-readable text (Gupta, 2006; Hsu *et al.*, 2022).

A scanned document falls under the category of electronic documents; however, its functionality and usability can fluctuate based on the chosen format and the application of additional technologies (Kristine *et al.*, 2016). Furthermore, it should not be deemed a secure form of documentation.

Despite the inherent difficulty in modifying the text of a scanned document, thereby preventing direct editing, it remains susceptible to alterations that compromise its integrity (Vimalachandran *et al.*, 2016). It is crucial to acknowledge that modifications to the image of a scanned document can be made without leaving discernible signs of tampering. Consequently, a signature affixed to the document may appear authentic, yet the document's content could have been altered subsequent to its endorsement by the signatory (Dugler, 2021).

There are many articles that highlight the diverse applications and challenges associated with scanned documents, ranging from their use in historical research to practical issues in healthcare and scientific research (Hegghammer, 2022; Hsu *et al.*, 2022). The use of OCR technology and the selection of appropriate scanning technology are crucial factors in these domains.

Unfortunately only few of them focus on the authentication factors. The integrity and security of scanned documents are important concerns, particularly in fields

where the authenticity and confidentiality of information are crucial. One of the primary concerns with scanned documents is ensuring their authenticity. This involves verifying that the document has not been altered or tampered with since it was scanned.

Techniques like digital watermarking, cryptographic signatures, and timestamps can be used to ensure the authenticity of scanned documents (Mayr *et al.*, 2022; Lax *et al.*, 2015). To protect the confidentiality of scanned documents, especially those containing sensitive or personal information, encryption is essential. Encrypting the scanned documents ensures that they cannot be read or altered by unauthorized individuals during storage or transmission.

Regrettably, these techniques are not readily accessible and might not be applicable in the routine processing of everyday documents. Employing such intricate methods of verification would significantly augment the time required during the processing of documents (Boudrez, 2007).

For certain types of documents, such as legal, medical, or governmental records, there are often specific regulations governing how they must be scanned, stored, and protected (Hyla and Fabisiak, 2020). Ensuring compliance with these regulations is crucial for maintaining the integrity and security of the documents.

This represents merely a fraction of the entire electronic document environment, where regulations have been established and implemented (Scott and Williams, 2009; KAWAKAMI *et al.*, 2019). In the majority of instances, electronic documents are not safeguarded and cannot be verified through straightforward methods.

In summary, while scanned documents offer convenience and accessibility, ensuring their integrity and security requires a combination of technological solutions and strict management practices. This includes using advanced scanning and encryption technologies, which are not commonly accessible and used in everyday life. Conversely, in numerous instances, such an elevated level of security may not be necessary.

This is particularly relevant in scenarios where arrangements have been established through face-to-face interactions or alternative communication methods, and the document merely serves as a legal affirmation of these agreements. Additionally, in certain situations, an electronic document is not exclusively utilized between two individuals. Consequently, the verification of its authenticity should be accessible to all parties potentially involved in the process.

3. Research Environment

Qualification certificates serve as formal documentation that substantiates an individual's competence in a specific field or domain. Typically, these certificates

are conferred upon the successful completion of recognized training programs, courses, or examinations. They facilitate expedited assessments by employers to ascertain the suitability of an applicant for a particular role, resulting in time and resource savings during the hiring process. Historically, printed qualification certificates have been employed as a means of validating and documenting individuals' qualifications.

The tangible nature of paper certificates furnishes a physical and visually identifiable record of an individual's accomplishments. Moreover, printed certificates are easily transportable, presentable, and verifiable, serving as concrete proof of qualifications when seeking employment. However, the advent of digital credentials and electronic verification systems has instigated a transition towards digital alternatives to printed certificates. Digital certificates offer advantages in terms of streamlined management and accessibility.

An indispensable facet of any certification process is the certificate's credibility. Certificate credibility pertains to the degree of trust and recognition accorded to a qualification certificate by pertinent authorities or employers. While furnishing a digital replica of a printed certificate is generally accepted, it poses challenges in terms of authenticity verification.

Scanned or photographed copies are susceptible to manipulation or alteration through digital editing tools, potentially compromising the document's credibility.n conclusion, the verification of certificate credibility constitutes a pivotal aspect of the certification process. Various factors contribute to certificate credibility, with the assurance of document authenticity and the establishment of trust remaining significant challenges.

Furthermore, certificate issuers must provide a straightforward document verification mechanism for an unspecified number of individuals who may encounter the certificate subsequent to its issuance. Consequently, this environment has emerged as an ideal testing ground for a novel approach, wherein an electronic document, even when printed as a hard copy, may undergo easy authentication verification.

4. Methods and Results

This study aims to tackle the challenge of ensuring the secure delivery of electronic certificates that can be easily authenticated in both digital and hardcopy formats. The research suggests the incorporation of secure QR codes containing extended unique identifiers onto the certificates (Szyjewski and Fabisiak 2018).

These QR codes establish a link to the issuer's digital verification platform, combining physical and digital elements. When the QR code is scanned using a mobile device or application, it directs users to an official verification website where

the provided information can be verified, affirming the document's legitimacy. This process offers a straightforward and practical solution for verifying both electronic copies and printed copies of the certificates.

The subsequent sections of this article furnish a comprehensive description of the proposed methodology and the verification process. To implement and validate this approach, a proprietary IT system was created. The efficacy of this method was demonstrated through its application in the Polish segment of a large certification program, involving the replacement of all paper certificates with their electronic counterparts.

The primary aim of this strategy was to increase the flexibility of issuing certificates while preserving their credibility and simplifying the verification process. The subsequent chapter details the method for issuing and verifying documents within the adopted approach.

Figure 1. *Document verification code area.*

Source: *Own elaboration.*

In this instance, the document is a single user certificate, containing information about the owner, issue date, completed modules, and specifics like exam dates and the syllabus version used during the test. Instead of creating a physical file for the certificate, it is stored as a new database record. This method improves data storage, vital for the system's large scale.

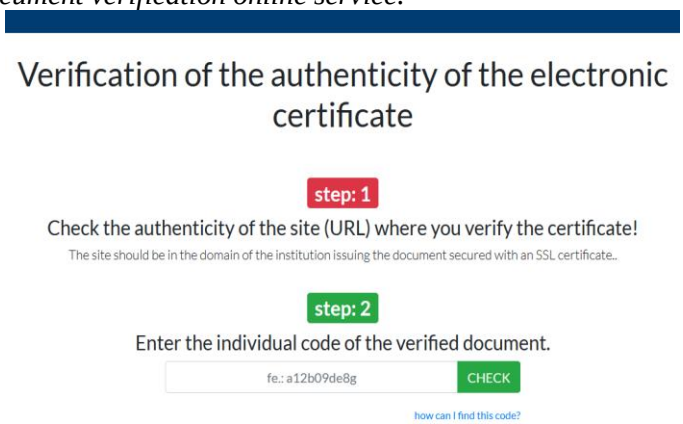
Certificates are generated as PDFs, a widely recognized electronic format, with raster graphics at a minimum of 300 dpi. After creation, recipients get a printed copy, bypassing the certification authority in the printing process. This method permits an electronic document to be printed without losing validity.

The key feature of this approach is the simple verification of the certificate, whether in electronic or printed form. Verification is achieved using a QR code and an individual key, with a dedicated online service by the authority. The first verification method involves scanning the QR code on the certificate. QR codes, scannable with camera-equipped devices like smartphones, link to a URL embedded with a certificate code in the issuer's database.

Loading this URL opens the authority's verification page, secured with an SSL certificate. The URL service authenticates all electronic certificates issued by the Polish authority involved in this research. On the verification page, users input the certificate's unique code and click "verify."

The second method caters to those without camera devices or unfamiliar with QR codes. It involves visiting the official URL and manually entering the certificate code. Both methods display the original certificate, assuring the data's authenticity from the issuer's trusted service. Any discrepancies in the document should alert to potential forgery. The document's format—digital, printed, or scanned—does not affect its credibility, as it can still be verified online. The image illustrates online verification system inputs, that allows any form of document validation.

Figure 2. *Document verification online service.*



Source: *Own elaboration*

This approach, initiated in 2020, initially targeted popular document types in the certification program. After a year, including minor updates, it gained acceptance among certificate holders and presenters, and even national government entities. The approach, supported by EU funds and national administration, issued numerous certificates for digital tool proficiency. A year later, it expanded to all issuer certificates, including internal organizational ones, phasing out standard printed certificates.

Adopting this method brought organizational benefits like cost and resource reduction. It also improved certificate recognition, allowing candidates to display desired competencies. The documents, easily verified, are accepted by employers and other stakeholders evaluating users' competencies. Issuing high-resolution electronic documents enables unlimited authentic copies.

This approach, tailored to electronic document issuer needs, was tested in the described environment. From document production to verification, it aimed to

validate the concept in electronic document flow. The online verification application and QR code stamping module, part of the author's system in this research, showed the approach's successful real-world application.

5. Conclusion

This paper addresses the challenge of ensuring the authenticity and integrity of electronic documents, particularly in the context of the increased vulnerability to forgery in the digital age. It highlights the susceptibility of electronic documentation, including scanned images and PDFs, to unauthorized modifications that leave no trace of tampering.

The advancement of IT and AI technologies further escalates this risk, necessitating robust verification mechanisms. The article reviews the transition to electronic documents during the COVID-19 pandemic and the continued popularity of scanned documents as a prevalent electronic format. It emphasizes the distinction between various electronic document types and their management, security, and interoperability challenges. Despite their widespread use, the paper notes the lack of emphasis on the authentication of scanned documents in existing literature.

The study presents a novel approach for securing electronic documentation flow, aiming to maintain document content integrity without sacrificing usability. This involves using secure QR codes linked to a digital verification platform, applicable to both electronic and printed documents. The method was tested in a certification environment, demonstrating its effectiveness in ensuring the credibility of qualification certificates.

The paper describes the implementation of this method in a large-scale certification program in Poland, which transitioned from paper to electronic certificates in 2020. The adoption of this approach yielded significant organizational benefits, including cost and resource savings, and improved recognition and verification of certificates.

It demonstrates a successful real-world application of combining physical and digital elements in document verification, tailored to the needs of electronic document issuers. The study's findings suggest that this method could be broadly applicable in various contexts where electronic document authenticity is crucial.

References:

Alpi, K.M., Brown Jr, J.C., Neel, J.A., Grindem, C.B., Linder, K.E., Harper, J.B. 2016. Scanning technology selection impacts acceptability and usefulness of image-rich content. Journal of the Medical Library Association, JMLA, 104(1), 15.

Bellare, M., Canetti, R., Krawczyk, H. 1996. Keying hash functions for message authentication. In Advances in Cryptology—CRYPTO'96: 16th Annual

International Cryptology Conference Santa Barbara, California, USA August 18-22, 1996 Proceedings 16, (pp. 1-15). Springer Berlin Heidelberg.

Boudrez, F. 2007. Digital signatures and electronic records. Archival Science, 7(2), 179-193.

Buckland, M.K. 1997. What is a "document"? Journal of the American society for information science, 48(9), 804-809.

Dülger, M.V. 2021. Electronic Document as the Subject of Document Fraud Crimes (Belgede Sahtecilik Suçlarının Konusu Olarak Elektronik Belge). Available at SSRN 3792220.

Goodrum, H., Roberts, K., Bernstam, E.V. 2020. Automatic classification of scanned electronic health record documents. International journal of medical informatics, 144, 104302.

Gupta, G., Niranjan, S., Shrivastava, A., Sinha, R.M.K. 2006. Document layout analysis and classification and its application in OCR. In 2006 10th IEEE International Enterprise Distributed Object Computing Conference Workshops (EDOCW'06), 58-58. IEEE.

Hegghammer, T. 2022. OCR with Tesseract, Amazon Textract, and Google Document AI: a benchmarking experiment. Journal of Computational Social Science, 5(1), 861-882.

Hsu, E., Malagaris, I., Kuo, Y.F., Sultana, R., Roberts, K. 2022. Deep learning-based NLP data pipeline for EHR-scanned document information extraction. JAMIA open, 5(2).

Hyla, T., Fabisiak, L. 2020. Measuring cyber security awareness within groups of medical professionals in Poland. Hawaii International Conference on System Sciences. DOI:10.24251/hicss.2020.473.

Kawakami, T., Nagase, K., Yokoi, Y., Sai, Y., Murayama, T. 2019. Improvement of Informed Consent Document Management in Clinical Trials Using an Electronic Medical Record System. Rinsho yakuri/Japanese Journal of Clinical Pharmacology and Therapeutics, 50(3), 81-86.

Knott, G.D. 1975. Hashing functions. The Computer Journal, 18(3), 265-278.

Lax, G., Buccafurri, F., Caminiti, G. 2015. Digital document signing: Vulnerabilities and solutions. Information Security Journal: A Global Perspective, 24(1-3), 1-14.

Mayr, L., Schardong, F., Custódio, R. 2022. Simplifying Electronic Document Digital Signatures. arXiv preprint arXiv:2208.03951.

Pranggono, B., Arabo, A. 2021. COVID-19 pandemic cybersecurity issues. Internet Technology Letters, 4(2), e247.

Putro, P.A.W. 2017. Physical document validation with perceptual hash. In 2017 3rd International Conference on Science in Information Technology (ICSITech) (pp. 582-587). IEEE.

Scott, P., Williams, P. 2009. Deploying electronic document management to improve access to hospital medical records. Journal of Management & Marketing in Healthcare, 2(2), 151-160.

Stromer, D., Christlein, V., Martindale, C., Zippert, P., Haltenberger, E., Hausotte, T., Maier, A. 2018. Browsing through sealed historical manuscripts by using 3-D computed tomography with low-brilliance X-ray sources. Scientific reports, 8(1), 15335.

Szyjewski, G., Fabisiak, L. 2018. A study on existing and actually used capabilities of mobile phones technologies. Procedia Computer Science, 126, 1627-1636.

Turró, M.R. 2008. Are PDF documents accessible? Information Technology and Libraries, 27(3), 25-43.

Vimalachandran, P., Wang, H., Zhang, Y., Heyward, B., Whittaker, F. 2016. Ensuring data integrity in electronic health records: A quality health care implication. In 2016 International Conference on Orange Technologies (ICOT), 20-27. IEEE.