Robust Visible Digital Stamp for Instant Documents Authentication and Verification

Hassan Al-Maksousy(1), Hussein Abdulhussein(2)
(1) Iraqi Prime Minister Office, Baghdad, Iraq
(2) University of Technology, Baghdad, Iraq
E-mail: (1) halma002@odu.edu, (2) 50179@uotechnology.edu.iq

Abstract. Confidential documents, especially the ones issued by educational institutions, large companies, government departments, military organizations, and intelligence agencies are facing great challenges of being used, copied, and redistributed without authorization such as being used for forgery cases. This puts great threats to businesses and any legislature, executive, and judiciary forms in any community. In this paper, a novel model called robust visible digital stamp, RVDS, is introduced to secure and protect important / sensitive documents with the aid of low-power computers that do not require any type of network connections. The proposed model in this paper converts the document’s information into a coded form of a customized quick response code, QR code. This code is then bound to the document for the verification process. This model uses a combination of keys for encryption and the authentication check process. The model does not need network connection, which makes it an off-line process that guarding users from the risk of losing their privacy.

Keywords: classified documents, documents authentications, QR code, network-independent authentication, cryptography, information encryption, data transmission

1. Introduction
Nowadays technological progress has made the flexibility to make the counterfeited documents much easier. Documents forgery is a challenging problem for any printed document [1]. Securing the important documents by adding combination of materials-based security features is possible. In many cases, counterfeiting is inexpensive and easy to be done, it only requires time for fraudsters to copy or manipulate these features. People may change information or put false ones into documents to create forged documents for example: government publications, real estate documents, graduation certificates, and identity documents. Documents forgery may help people getting unfair employment opportunities. In addition, some communities may be under a serious security threat if an unauthorized access to important security related documents is granted, and sensitive or life threatening details are leaked [2]. Usually some forms of forgery can be detected or prevented by using a conventional ways like stamps and signatures as from of authentication methods. However, in case of professional forgery where the forger has knowledge about the name of the document issuer and the way the documents are signed, detecting the forgery or any alteration is going to be harder [3, 4]. In addition to that, changing the designs, printing materials, and adding ink based features are used widely to prevent documents manipulation. Other methods like network authentication and verification systems will be affected by many factors such as software vulnerabilities and service availability time. In addition, using network...
dependent systems may affect users’ privacy in addition to the cost of the network infrastructure or services providing. In this paper, a novel model called robust visible digital stamp, RVDS, is developed to make important / sensitive documents secured and protected with the aid of any computer that does not require network connections nor shared databases. The proposed model in this paper encrypts the document’s information into a customized QR code which is then attached to the document for the verification process. This model uses a private one-time usage key for the encryption process and a public key for the verification process. The self-authenticated document will include both the document content and the QR Code, which makes it easier for the receiving party to verify the document as the QR Code holds all the document information, but in an encrypted form. What makes this method novel is that the process of converting the document contents into a code does not require databases or third party solutions involved to provide encryption / encoding keys. The model does not use any network-dependent processes, which makes it an off-line and secure process that helps maintaining users’ information privacy. The proposed method in this paper provides an efficient approach in the data embedding algorithm to secure the documents content in the QR code form while keeping the QR code valid and readable. Different from the other methods, the proposed algorithm in this paper encrypts the documents with very low error and high capacity QR code which means higher amount of data being transferred and protected.

2. Related Work and History
In 1994, a team at Denso Wave invented and developed a 2D barcode called quick response code, QR Code, to help in the manufacturing process [5]. At that time, it was capable to only hold limited number of alphanumeric characters. It was then developed to pack as much information as possible. In 2000, QR Code became very popular in Japan. It was then approved as an international standard by ISO [6]. QR code is a fast and cost effective way of data transmission. It is commonly used on daily bases. Any QR code reader can easily decode the documents information that is embedded in. A paper-based document authentication is introduced in [7], yet their method did not add encryption to the original text. This leads to privacy / security breaches when delivering the embedded documents in QR code. Therefore, documents content have to be encrypted before embedding into the QR code. In addition to that, optical character recognition, OCR, is used to generate a hashed value, but it ignored right to left languages and other languages that are not supported by OCR. A paper-based documents authentication was proposed as a solution for encrypted authentication [8, 9]. However, their service model requires having a third party to handle verifying documents or credentials to be used by the algorithm. In [10], a system is used to store information in a central database. Meaning a non-offline solution is used while the original text is not encrypted nor compressed. Therefore, it cannot maintain data privacy nor can handle a lot of information. In addition to data / information privacy, the amount of data may affect the size of the used QR code. A color document authentication based on using QR code and embedding digital watermarking was implemented [11] where storing a signature requires multi QR codes. Their approach uses the color map of the document, which makes the size of the document image affects the size of the signature. In many cases, QR code can be tampered in order to change the documents related information. Dual anti-counterfeiting was introduced [12] to verify the legitimacy of users and documents copyrights by encoding encrypted user information in addition to embedding digital watermark, which they will be extracted and decrypted later to be compared together for document verification.

In this paper, the proposed method, RVDS, does not require having third party nor having any intermediate services between the sender and the receiver. In addition, the resulted QR code has almost always constant size. Also, in this paper, the verification process only involves scanning the QR code and extracting the original content as the QR code includes the digital signature and the document contents in compressed and encrypted form.
3. Method Framework and Motivation
The motivation behind designing the proposed method is to create a digital stamp that is printed within the actual hard copy document which makes authentication and verification are done without the help of any type of network related operations nor having any prior knowledge (also known as zero knowledge) about the document information. The proposed RVDS method staged into two parts, authentication stage, and verification stage. The stages are explained as in the following:

3.1. Authentication Stage
This stage is implemented at the documents issuer side at which the document is created, printed, and then stamped with QR code, such as governments’ agencies, universities, and institutions require having their documents protected. This stage includes four main steps as shown in figure (1) and detailed as in the following:

![Authentication Stage Diagram](image)

Figure 1: Authentication Stage (red: user entry, orange: machine controlled entry, yellow: intermediate output, green: stage output.)

3.1.1. Data Compression: It involves changing how data is represented so that the compressed results take less space to be stored, which means less time to process and transmit the information. In this step, documents information such as the main text for up to 450 words per page is compressed to reduce the actual data size that efficiently represents the entered document information. Such step can be done using any simple text optimized lossless compression algorithms. For this method, dictionary-based text reduction for limited common or popular
words is used, where each used text language has its own table. Other than that, Lempel-Ziv-Markov chain algorithm (LZMA) is used for text compression.

3.1.2. Data Encryption: It translates data into another configuration, where it can be only accessed and read by the parties having the password that is used in the encryption (decryption key). Securing important and sensitive data can be done in many methods, of which encryption is the most effective and popular one. In this step, the compressed text data will be encrypted using a unique encryption key. The encryption key is a one-time password because it consists of a unique combination of pre-shared key, machine signature (or identification number of the machine where the encryption process is being executed), and the time stamp (epoch time at which the encryption process is initiated). The result of this process will be an encrypted form of data consists of the compressed text and the machine signature. The encryption process can be mathematically represented by the following equation:

\[ Y = E(K, X) \]  

Where:
- \( E \) is the encryption algorithm
- \( K \) is the encryption key
- \( X \) is the text representation in compressed form

3.1.3. Data Signature: Data signatures are like fingerprints in electronic forms. In this step, the encrypted data are signed. However, the encrypted data has an arbitrary size. In order to get the data in a fixed size form, data mapping is implemented. This can be done by calculating the hash of the encrypted data. Any available hash function can be used depending on the complexity level; however, for accuracy and simplicity, the hash type that is used in this method is SHA-3 with standard size of 256 bit. Then RSA private key is used to encrypt the hash itself, which means signing the encrypted data. Equation (2) represents the mathematical representation of this step. 2048-bit RSA key size is recommended to be used until 2030 [13].

\[ S = E_{RSA}(Hash, Prk) \]  

Where:
- \( E_{RSA} \): is the signing/encryption algorithm
- \( Hash \): is the SHA-3 hash of the encrypted data
- \( Prk \): is the RSA private key

3.1.4. QR Code Generation: The QR code in this method has the encrypted data and the data signature embedded in. Meaning, the result of equation (1) is appended to result from equation (2). Together they form one text line. In addition to that, encryption key will be masked into the same text line. The final text line will be compressed using lossless data compression called Lempel-Ziv-Markov chain algorithm (LZMA) as 7-Zip. This makes decrypting the document data much faster at the receiver party and it does not require network access for the encryption key. The compressed information is then converted to QR code form using Zebra Crossing, ZXing, library. This QR code is printed anywhere on the document as required by the issuer. The QR code in this method is considered as a digital stamp that replaces the traditional ink based stamps.
3.2. Verification and Decryption Stage

This stage is implemented at the documents receiver side at which the document is checked to be verified for authenticity purposes. This stage is usually done at the time where the received documents have QR code printed on them and submitted for verification such as degree certificates submitted for job applications. This stage includes two main steps: verification, and decryption step. These steps are detailed as in the following:

![Diagram of Verification Step]

3.2.1. Verification: In this step, the verification process is done by extracting encrypted data and the signature from the scanned QR code. Then RSA public key is used to decrypt the hash value from the extracted signature. The resulted hash should represent the original encrypted data. If the resulted hash is the same as the hash for the extracted encrypted data, the document is verified and has not been manipulated in any means. The flow chart of this step is shown in figure (2). To sort this step mathematically, equations (3) and (4) are applied to the extracted QR content (encrypted data and the signature only). If they yield the same results ($H_1 = H_2$), the document is considered verified and valid.

$$H_1 = \text{Hash}(E\text{Data})$$  \hspace{1cm} (3)

Where:

- \text{Hash}: \text{is the SHA-3 hash function}
- \text{EData}: \text{is the extracted encrypted data}
\[ H_2 = D_{RSA}(\text{Sig}, Puk) \]  

Where:
- \( D_{RSA} \): is the decryption algorithm
- \( \text{Sig} \): is the extracted signature
- \( Puk \): is the RSA public key

### 3.2.2. Data Decryption:
After the document is being processed in the verification step, a second process, data decryption, is done. This process requires having the original encryption key used to decrypt the original entered data. In other words, the data was encrypted with one-time password (called encryption combinations) which is a combination of pre-shared key, machine signature, and time stamp. However, these codes are embedded into the QR code in masked forms. So, in addition to the extracted encrypted data and the signature, encryption combinations are extracted and unmasked to be used as key for the decryption process. The decryption process yields the compressed data that should match the original entered data at the issuer party. The flow chart of this step is shown in figure (3).

Figure 3: Data Decryption Step (red: user entry, yellow: intermediate output, green: stage output.)

### 4. Experimental Results
The algorithm is developed using C# programming language and .NET framework. Since the majority of the institutions are using word processor such as Microsoft Word to issue the documents, the prototype is designed to insert the text and the QR code into their Microsoft Word template. Experiments done at the documents issuer side were all conducted by a machine with specifications of: Intel Celeron N3350 2.4 GHz CPU, 2GB RAM, 32GB eMMC flash memory, and 32-bit Windows 10 operating system. On the recipient side, the machine however can be either a low-power computer as in the issuer side or any commercially available off-the-shelf (COTS) devices such as mobile devices with Android operating system (the program can be cross compiled into different platforms). Also, two languages were used during the test, English and Arabic. Figure (4) shows the graphical user interface (GUI) of the proposed method while being used in authentication stage.
Iraq's GDP growth slowed to 1.1% in 2017, a marked decline compared to the previous two years as domestic consumption and investment fell because of civil violence and a sluggish oil market. The Iraqi Government received its third tranche of funding from its 2016 Stand-By Arrangement (SBA) with the IMF in August 2017, which is intended to stabilize its finances by encouraging improved fiscal management, needed economic reform, and expenditure reduction. Additionally, in late 2017 Iraq received more than $1.4 billion in financing from international lenders, part of which was generated by issuing a $1 billion bond for reconstruction and rehabilitation in areas liberated from ISIL. Investment and key sector diversification are crucial components to Iraq’s long-term economic development and require a strengthened business climate with enhanced legal and regulatory oversight to bolster private-sector engagement. The overall standard of living depends on global oil prices, the central government passage of major policy reforms, a stable security environment post-ISIS, and the resolution of civil discord with the Kurdish Regional Government (KRG).
In this application, the only required pieces of information are the clerk name (sometimes can be replaced with the pre-shared key) and the document content. The result will be as QR code form of the encrypted document information posted on the document as shown in figure (5). For statistical and future comparison purposes, the QR capacity (number of words per page), language type, and the size of printed QR code are calculated for each document in the experiments. It can be noticed in table (1), where one language is used, that the size of the QR code is almost always the same and it goes up to 4.2x4.2 cm. However, using multi languages at the same time may increase the QR code size up to 5x5 cm, shown in table (2).

| Capacity (words/page) | QR Size (cm²) | Capacity (words/page) | QR Size (cm²) |
|-----------------------|--------------|-----------------------|--------------|
| 100                   | 2.4²         | 100                   | 2.5²         |
| 200                   | 2.5²         | 200                   | 2.7²         |
| 300                   | 3.2²         | 300                   | 3.6²         |
| 400                   | 3.9²         | 400                   | 4.1²         |
| 450                   | 4.2²         | 450                   | 4.9²         |
| 500                   | N/A          | 500                   | N/A          |

Table 1: One language (Arabic or English)  
Table 2: Two languages (Arabic & English)

5. Protection Evaluation
The proposed method has a very high accuracy in distinguishing between tampered and genuine documents. It is a significant contribution in documents tamper detection since it uses hash functions to represent all the document information. Thus, any minor alteration even if it is done by one letter, the hash followed by the QR code would all be different than the original ones. This paper also helps in forensics document analysis. Figure (6) shows some results of the original and forged sentence detected by RVDS.

![QR Code](image1.png)  
**The quick brown fox jumps over the lazy dog**  
(a) Original Sentence

![QR Code](image2.png)  
**The quick brown foxes ↔ jumps over the lazy dog**  
(b) Forged Sentence

Figure 6: Genuine and Tampered Sentences: the difference between them is letters (ex) which are added to word (fox) as an intentional modification.

6. Conclusion
Documents authentication and verification are very important. There are many cases where documents can be altered, leaked, or accessed without proper authorization. Thus, RVDS is proposed as a solution to secure and protect important documents. This method is robust and very fast, it does not require having network access which makes it mobile and offline solution. In addition, one of the main contributions in this paper, the QR code (the stamp) has almost always the same size no matter how the data changes. This method creates digital authentication stamp
that is printed within the actual document. Verification is done instantly with zero knowledge about the document information. This method does not require having any form of network related operations nor involving any third party solutions to provide encryption / decryption keys transfers.

References
[1] Garain U, Halder B. On automatic authenticity verification of printed security documents. In: Proceedings - 6th Indian Conference on Computer Vision, Graphics and Image Processing, ICVGIP 2008; 2008. p. 706--713.
[2] Abeyratne R. Security Issues. In: Strategic Issues in Air Transport. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012. p. 165--278.
[3] Lowe R, Walske C, Royal Philatelic Society (Great Britain). The work of Jean De Sperati II : including previously unlisted forgeries. Royal Philatelic Society; 2001.
[4] Verma R, Rao DS. Offline Signature Verification and Identification Using Angle Feature and Pixel Density Feature And Both Method Together; 2013. 4.
[5] DENSO CORPORATION TCRII. Two-dimensional Code; 1994. Patent JP2938338B2.
[6] ISO. ISO/IEC 18004:2000 - Information technology - Automatic identification and data capture techniques - Bar code symbology - QR Code. ISO Standards. 2000;2000:122.
[7] Warasart M, Kuacharoen P. Paper-based Document Authentication using Digital Signature and QR Code. 4TH International Conference on Computer Engineering and Technology. 2012;40(January):94--98.
[8] Li CM, Hu P, Lan WC. AuthPaper: Protecting paper-based documents and credentials using Authenticated 2D barcodes. In: IEEE International Conference on Communications. vol. 2015-Septe. Institute of Electrical and Electronics Engineers Inc.; 2015. p. 7400--7406.
[9] Ahmed HA, Jang JW. Document certificate authentication system using digitally signed QR code tag. In: ACM International Conference Proceeding Series. Association for Computing Machinery; 2018. ,
[10] Singh A, S Pavith R. Degree Certificate Authentication using QR Code and Smartphone. International Journal of Computer Applications. 2015 jun;120(16):38--43.
[11] Mohsin Arkah Z, Alzubaidi L, Ali AA, Abdulameer AT. Digital color documents authentication using QR code based on digital watermarking. In: Advances in Intelligent Systems and Computing. vol. 940. Springer Verlag; 2020. p. 1093--1101.
[12] Xun Y, Li Z, Zhong X, Li S, Su J, Zhang K. Dual Anti-counterfeiting of QR Code Based on Information Encryption and Digital Watermarking. In: Lecture Notes in Electrical Engineering. vol. 543. Springer Verlag; 2019. p. 187--196.
[13] Barker E. Recommendation for Key Management – Part 1: General. NIST Special Publication 800-57. 2016;p. 1--142.