Paper-based Document Authentication using Digital Signature and QR Code

Maykin Warasart and Pramote Kuacharoen [+]

Department of Computer Science, Graduate School of Applied Statistics

National Institute of Development Administration

118 Serithai Rd. Bangkapi, Bangkok 10240 Thailand

Abstract. There are still needs for paper-based documents in certain circumstances where electronic documents cannot efficiently replace them. For example, documents issued by the government such as birth certificates, driver licenses, and passports must be paper-based. With advanced scanning and printing technologies, paper-based document fraud can easily be conducted without significant high cost. In this paper, an implementation of paper-based document authentication is presented. The integrity of the text message and the author of the document can be verified with the use of a digital signature and QR code. The proposed method can be automatic or semi-automatic. It is semi-automatic when the OCR is not accurate and it requires the user to visually compare the text message on the paper and the one obtained from the QR code; however, this method does provide convenience for the user in dealing with a large amount of documents.

Keywords: paper-based document, digital signature, QR code, authentication

1. Introduction

The trend of paperless office or e-Government has grown rapidly. However, no matter how the vision of a paperless office has been brought to reality, there are needs for certain types of work that are still needed to be communicated in a form of paper documents. For example, government-issued and legal documents, whether it is a birth certificate, a driving license, a passport, an insurance document, or a contract, need to remain in a paper form [1].

With the technological advancement of digital printing and scanning that can be obtained at low costs but has very high efficiency and quality, criminals can easily produce counterfeit documents for defrauding. This makes it difficult to differentiate the counterfeit documents from the authentic documents. By misusing modern equipment whether it is a scanner, a printer or a plotter, it is regarded as a threatened danger to the nation's society and economy [2].

Authenticating paper-based documents usually requires a specialist. This solely depends on a capability of the expert who can verify to the documents. There are a lot of organizations which are set up to work in the area of forensic science in many countries. This type of work utilizes special equipment such as a UV lamp, a magnifying glass, or an infrared inspector to help with the inspection. In practice, it is difficult for the organization which has a large amount of documents needed to be inspected quickly. For instance, a bank has to work with checks, bills of exchange, and receipts. Due to the forensic science work that must follow a certain procedure in accordance to the law, it may take a significant amount of time. The procedure may involve sending the documents needed to be verified to the police, having the specialist verify the document, and so on.

[+] Corresponding author. Tel.: + 66-2727-3085; fax: +66-2374-4061.
E-mail address: pramote@as.nida.ac.th.

This paper presents the process of authenticating paper-based documents that can be used quite conveniently, quickly, and semi-automatically, by applying digital signature and QR code. This enables the verification of the documents without depending on any special institute such as the forensic science center.

2. BACKGROUND AND RELATED WORK

This section provides background information related to this paper such as cryptographic hash function, digital signature, and 2D barcodes.

2.1. Cryptographic Hash Function

A hash function maps a variable-length message into a fixed length hash value or message digest without using any key. The hash function needed for security applications is referred to as a cryptographic hash function which is computationally infeasible to find either a message that maps to a pre-specified hash value or two messages that map to the same hash value. In other words, a cryptographic hash function must have the one-way property and the collision-resistant property. With the aforementioned properties, a cryptographic hash value is used to determine whether or not the corresponding message has been modified [3]. However, the hash value must be protected.

2.2. Digital Signature

A digital signature is a bit pattern that depends on the message being signed and uses some information unique to the signer [4]. The message M is fed into a cryptographic hash function resulting in a hash value h or a message digest. The hash value h which depends on the message M is encrypted using the signer's private key producing the signature.

To verify whether or not the digital signature is valid, the result hash value from the message M' is compared to the value from decrypting the signature using the signer's public key. If both values are identical, the owner of the public key is the author of the message. Otherwise, the signature is invalid.

Digital Signature Standard (DSS) [5] includes three techniques, namely; the Digital Signature Algorithm (DSA), the RSA digital signature algorithm [6], and the Elliptic Curve Digital Signature Algorithm (ECDSA) [7]. The security of the digital signature depends on the cryptographic hash function and the public key cryptographic algorithm. For breaking a digital signature, an attacker may create a fraudulent digital signature by creating a new message for an existing digital signature which is an attack on the cryptographic hash function or by constructing a fraudulent digital signature for a given message which is an attack on the public key cryptographic algorithm. The hash function must be collision resistant and the public key algorithm must be strong against attacks. The approved techniques are considered secure. It is computationally infeasible to forge a digital signature.

The digital signature provides authentication and non-repudiation. Therefore, if the signature is valid, the author of the message cannot deny creating the message.

2.3. 2D Barcodes

The research for storing more data in barcodes led to the development of 2D barcodes (two-dimensional barcodes) that can store large amount of data in a small area to support information distribution and detection without accessing the database. When selecting and using 2D barcodes, one must consider the following factors: a) application usage b) standard c) implementation d) the data needed to be encoded in barcodes and e) barcode printing. Each barcode type is a standard that defines the printed symbol and how a device such as a barcode scanner reads and decodes the printed symbol. Common 2D barcodes are Data Matrix, PDF417, Maxi Code, and QR Code [8][9].

A QR code is a 2D barcode which consists of a black square pattern on white background. The QR code contains information in the vertical direction as well as the horizontal direction. The data capacity can be the maximum of 7,089 numeric characters, 4,296 alphanumeric characters, or 2,953 bytes. QR codes use the Reed-Solomon error correction which can detect and correct multiple errors. QR codes can be read by QR scanners or mobile phones with a camera. The snapshots of QR codes taken by mobiles phones or scanned by scanner usually are not perfectly aligned causing the image to be distorted. However, algorithms for

correcting distorted images exist. Ohbuchi et al. [10] show new algorithms and implementations for reorganizing QR codes in mobile phones. Sun et al. [11] present an algorithm for analyzing and correcting the distorted QR code image.

3. DESIGN OF PAPER-BASED DOCUMENT AUTHENTICATION

In this section, the design of paper-based document authentication using digital signature and QR code is presented. By using the proposed method, the authenticity of the document can be verified.

3.1. Sender Process

For this process, the message and the corresponding verification code in a form of QR code are printed on paper. After the message is composed, its hash value h generated. Then, the hash value h is encrypted with private key of a sender resulting in the digital signature on the message M. After that, the message M and the digital signature are combined and compressed to reduce size so that they can be stored in a QR code. The compressed message and signature are fed into a QR code generator. After the QR code has been created, message M together with the QR code are printed on to paper and sent to a receiver. The simplified process of sender is shown in Figure 2 (a).

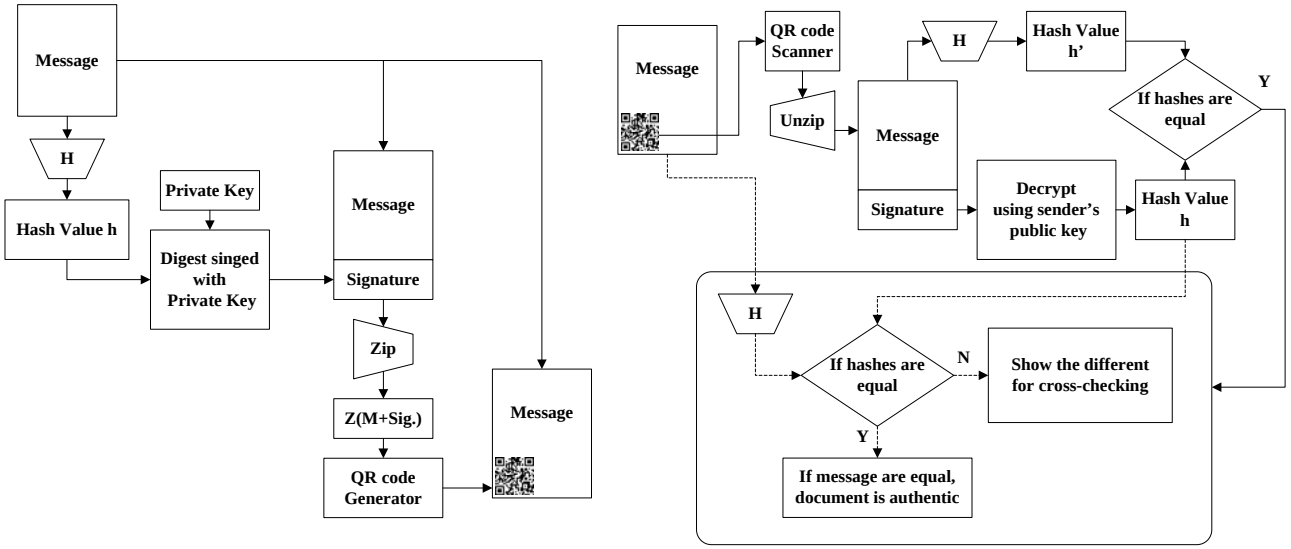

Fig. 2: The simplified processes of the sender and the receiver

3.2. Receiver Process

When a receiver obtains the document from a sender, he/she may verify the authenticity of document by scanning the document and processing the image as illustrated in Figure 2 (b). The verification process starts with checking the integrity of the information stored in the QR code. The information in the QR code consists of the message and the signature of the sender on the message, which are compressed. After scanning the QR code and uncompressing the data, the signature can be verified by comparing the hash value of the obtained message and the value from decrypting the signature using the sender's public key. If both values are identical, the signature is valid.

To validate the printed message, the Optical Character Recognition (OCR) must be employed. The hash value of the message obtained from the OCR is computed and compared with the hash value obtained from the message in the QR code. If they are identical, the printed message is authentic. However, if they are different, it cannot be concluded that the printed message has been modified. Further human review must be conducted. This is because the OCR may not be 100% accurate. The message obtained from the QR code can be shown next to the printed message which can be visually inspected.

4. SECURITY ANALYSIS

This section describes the security analysis of the proposed method. With the use of the digital signature, the recipient can be ensured that if the digital signature is valid, the message was created by the sender and was not altered in transit. The proposed method has three security properties as described below.

4.1. Authentication

The hash value of the original message is encrypted with the sender's private key which is only known by the sender. Therefore, a valid signature implies that the message was created by the sender. The corresponding public key is used to verify the signature. If the signature is invalid, the recipient cannot authenticate the message.

4.2. Integrity

The cryptographic hash value serves as a digital fingerprint of the message. If the message has been altered, the hash value is most likely different since it is infeasible to find another message which has the same hash value. However, the cryptographic hash value must be protected. Otherwise, an attacker is able to modify the message and regenerate the corresponding hash value. In the process of creating the digital signature, the cryptographic hash value is encrypted with the sender's private key. It is infeasible for an attacker to modify the message and signature in such a way that it is valid without the knowledge of the sender's private key. Hence, the integrity of the message is preserved.

4.3. Non-Repudiation

The private key and the public key are mathematically related. Information encrypted with the private key can only be decrypted with the corresponding public key. Since the sender signed the message with the private key and a valid signature is verified using the sender's public key. The sender cannot deny having signed the message.

5. IMPLEMENTATION AND EXPERIMENTAL RESULTS

A prototype system is implemented to verify the proposed method. The program is developed using Java programming language. Java Cryptography Architecture (JCA) provides cryptographic functionality such as digital signature and digital certificate. Java Keytool was used to generate 1024-bit RSA public/private key pairs and certificates. The digital certificate complies with X.509 standard. After the certificate is created, it can be exported and distributed. In a real life situation, the digital certificate should be issued by a reputable Certificate Authority (CA).

Since a majority of documents in organizations are created using a word processor such as Microsoft Word, the prototype allows the message to be composed in Microsoft Word. Apache POI, the Java API for Microsoft Documents, provides an API for extracting the message from the document. After obtaining the message, the digital signature is created using SHA-256 and RSA algorithm.

The message and the digital signature are then compressed to reduce size using Java APIs. It is necessary to compress the data because QR code has limited capacity. The compressed information is stored in a QR code using ZXing library. The next step is to use iText® PDF library to format the document which contains the message and a QR code. Finally, the document is ready to be printed on paper.

The receiver verifies the authenticity of a document by scanning the printed document and inputting the result to the verification program. In this experiment, the resolution for scanning is set to 300 dpi. The scanned image is divided into two parts, namely, the message and the QR code. ZXing library is used to decode the QR code. Upon successful decoding, the compressed information is obtained. By uncompressing it, the message and the digital signature are obtained. The signature is verified by comparing the hash value of the message and the hash value from the digital signature. If the verification fails, the paper is not authenticated. If the signature is valid, it implies that the information in the QR code is authenticated.

The more challenging task is to verify that the printed message has not been modified. The message part of the scanned image is passed through an OCR and its hash value is compared with the one calculated from

the message in the QR code. If both hash values are identical, the printed message is authenticated. Otherwise, the program displays both the scanned message and the message in the QR code. This requires human review to decide whether or not the printed paper is counterfeit. This is because the accuracy of OCR may not be 100% accurate or the scanned image is not clear. The program also highlights where the results of OCR and the message from the QR code are different. This facilitates the process of inspection by humans.

The table below shows the number of words in the documents and the number of mistakes of the OCRs. The Asprise OCR makes approximately 6 to 11 mistakes per 100 words whereas Adobe Acrobat OCR and Microsoft OneNote OCR have much better performances. The mistakes of both commercial OCRs are mostly from quotes and spaces. If quotes and spaces were corrected, most documents could be automatically verified.

Document No	Number of Words	Asprise OCR v. 4	Adobe Acrobat	Microsoft OneNote
1	246	20	0	2
2	140	13	2	4
3	227	18	0	9
4	145	16	1	5
5	235	15	4	10

6. CONCLUSION

Authenticity of paper-based documents can be achieved by using digital signatures and QR codes without accessing the database. The verification process can be done automatically if the OCR is accurate. Otherwise, human inspection is required. Even with this semi-automatic process, this proposed method facilitates the verification process. The inspector can see the differences between the printed message and the message in the QR code.

7. References

[1] R.L. van Renesse, “Paper-based document security–A Review,” in European Conf. on Security and Detection, 1997.

[2] U. Garain and B. Halder, “On automatic authenticity verification of printed security documents,” 6th Indian Conf. on Comput. Vision, Graphics & Image Process., 2008, pp. 706-713.

[3] M. Singh and D. Garg, “Choosing best hashing strategies and hash functions,” Int. Advance Computing Conf., 2009, pp. 50 – 55.

[4] P. Kuacharoen, “Design and analysis of methods for signing electronic documents using mobile phones,” Int. Conf. on Comput. Applicat. and Network Security, 2011, pp. 154-158.

[5] Digital Signature Standard (DSS), FIPS PUB 186-3, 2009.

[6] RSA Cryptography Standard, PKCS #1 v2.1, 2002.

[7] Public Key Cryptography for the Financial Services Industry, The Elliptic Curve Digital Signature Algorithm (ECDSA), ANSI X9.62-2005.

[8] J. Z. Gao, “Understanding 2D-barcode technology and applications in M-commerce – design and implementation of a 2D barcode processing solution,” in Proc. Int. Conf. on Comput. Software and Applicat., 2007, pp. 49 – 56.

[9] H. I. Hahn, J. K. Joung, “Implementation of algorithm to decode two-dimensional barcode PDF-417,” Int. Conf. on Signal Processing, 2002, pp. 1791-1794.

[10] E. Ohbuchi et al., “Barcode readers using the camera device in mobile phones,” in Proc. Int. Conf. on Cyberworlds, 2004, pp. 260- 265.

[11] A. Sun et al., “The QR-code reorganization in illegible snapshots taken by mobile phones," in Proc. Int. Conf. on Computational Sci. and its Applicat., 2007, pp.532-538.