Integrity Assurance System for Document Security Using Keccak and Quick Algorithm Response Code

by N/a N/a

Submission date: 20-Apr-2023 09:58AM (UTC+0700)
Submission ID: 2069937766
File name: Integrity_Assurance_System_for_Document_00_artikel.pdf (843.5K)
Word count: 3711
Character count: 19192

Integrity Assurance System for Document Security Using Keccak and Quick Algorithm Response Code

Faiz Muqorrir Kaaffah
Informatics Department
Universitas Islam Nusantara
Bandung, Indonesia
faiz@uninus.ac.id

Risnandy Maulana
Informatics Department
UIN Sunan Gunung Djati Bandung
Bandung, Indonesia
risnandy.maulana@gmail.com

Wildan Budiawan Zulfikar
Informatics Department
UIN Sunan Gunung Djati Bandung
Bandung, Indonesia
wildan.b@uinsgd.ac.id

Cepy Slamet
Informatics Department
UIN Sunan Gunung Djati Bandung
Bandung, Indonesia
cepy_lucky@uinsgd.ac.id

Nur Lukman
Informatics Department
UIN Sunan Gunung Djati Bandung
Bandung, Indonesia
n.lukman@uinsgd.ac.id

Ali Rahman
Informatics Department
UIN Sunan Gunung Djati Bandung
Bandung, Indonesia
ali@uinsgd.ac.id

***Abstract*— Still at risk from threats, public documents must have several security models, namely confidentiality, integrity, and availability. because many violations of public documents are carried out for various reasons that are utilized from one of these security models. art is needed to ensure such security, the method used to ensure that a document can be trusted can be confirmed who its owner is, its veracity, and ease of access. Maintain document security with the Hash Message Authentication Code method and use the Keccak algorithm which is sensitive if there are differences and changes in data. The sensitive nature of these methods and algorithms can guarantee the integrity of the documents. from all the processes above, this research aims to create an " Integrity Guarantee System" for documents that are distributed publicly.**

Keywords—***Security, Hash Message Authentication Code (HMAC), Keccak, Integrity Assurance System.***

I. Introduction

Crimes in documents are usually document falsification that can harm the victim, for example, a signature that is imitated or forged to take rights from the victim for the benefit of the perpetrator. There have been at least 151,750 cases of violations with the keyword “Document Forgery”. Cryptography, Steganography, and Watermarking techniques can be used to obtain security, confidentiality, privacy, and authenticity of data [1]. Cryptography encrypts messages and makes them in an unreadable form called a cipher. While steganography hides data in media such as text files, images, audio, video, etc, and hides the presence of messages in the media. The science of cryptography is the information protection technique to encrypt, store and secure data when transmitted, to prevent the reading of private information by intruders or the public [2].

Signatures have become a sign of legality in documents with a small level of risk before modern times. One of the functions of the signature on a document is as a sign that the document has been read and approved by the signer [3]. The signature requires carefulness and trust from the party receiving the document, in other words, if the recipient does not believe that the document was issued by the party that it should be, then the function of the signature as legality is no longer relevant.

With technology as it is now, this conventional signature system is very easy to imitate or forge, so the signature alone is not enough to fulfill the authentic nature of the document. This problem requires a system that can guarantee documents with unique marks, namely marks that are only owned by one document. So that if the mark is copied, the mark will not pass the identification or be declared a fake document. The cryptographic approach is carried out by *hashing* so that each registered document has a *message* and each unique *digest.*

II. Methodology

A. Data Protection

Data protection is one of the goals of implementing cyber security. The purpose of cybersecurity is to protect the confidentiality of information, the integrity of information from unauthorized changes, and to ensure the availability of information to users at the expected level of performance. Cybersecurity goals are very well defined by the CIA's triad model, which consists of Confidentiality, Integrity, and Availability. [4]. Here's the explanation:

1). Confidentiality: based on publications from the National Institute of Standard and Technologists (NIST) that “The term ' confidentiality ' means preserving authorized restrictions on access and disclosure, including means for protecting personal privacy and proprietary information." [5].

2). Integrity ensures that data is clean and reliable from changes intentionally or unintentionally by the system. [6].

3). Availability: availability is the main reason a system is reliable, information systems must be accessible to the user for the system to provide whatever results in the user demands.

In this modeling, each model has its own level of importance according to the conditions and needs of each system. An example of a Facebook that is more concerned with integrity and availability, IoT Health monitoring requires availability, and banking requires integrity and confidentiality to ensure that data is accessed by legitimate parties. but that does not mean the other models are not important, because Facebook still requires confidentiality. After all, not all data is uploaded publicly.

Data protection is a process of protecting that data is available, authenticated and access rights are maintained, following the CIA's Triad security model. Data security and privacy issues exist throughout the data lifecycle from

creation, transfer, use, sharing, storage, and archiving to destruction [7]. This protection is not only sought through the system side but refers to three sides, namely the policy, system, and user side.

B. Hash Message Authentication Code

Hashes still have vulnerabilities to dictionary attacks attack, rainbow table attacks, or brute force attacks. All these attacks have almost the same way of working, namely by guessing, gathering, and trying all the possibilities that exist to uncover confidential data. Let's assume a simple scenario where user A uses a password, say "12345" and the hash value is stored. Suppose User B also uses the same password "12345"', it will return the same value stored in the system for user A with password "12345". The adversary might use pre-computed tables (rainbow tables), generated by high-processing computers, to generate multiple hash value combinations. Using this table, adversaries can cross-reference stolen hash values (such as passwords) and perform reverse lookups to determine their original values [8]. To overcome this problem, additional data is needed which is inputted with the original data randomly and automatically by the system. This method of adding data for randomization is called Hash Message Authentication Code (HMAC) [9] [10], the technique to be used is technique salting or salting. The salting technique is used to prevent the calculation of dictionary attacks or rainbow table attacks.

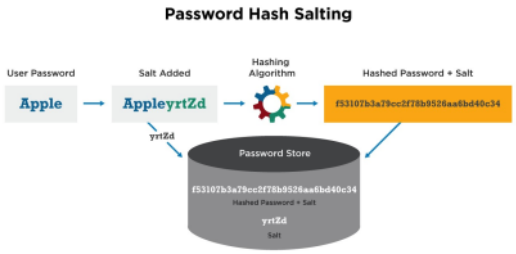

Fig. 1example of using the HMAC method.

Fig . 1 is an example of using the HMAC method with salting, Salt is a general term for data that completes password input to a one-way function that generates password verification data to make different instances of functions applied to the same input password produce different outputs. The practice of applying salt increases the cost of brute attack broad-scale force on multiple accounts using a pre-calculated set of values [11]. This salt function can be implemented in various ways by the designer and the implementation scheme and how this method is implemented must ensure security [12].

C. Keccak Algorithm

Keccak is the chosen algorithm to become the SHA3 algorithm after winning an open competition held by NIST with dozens of other algorithms as candidates. The keccak algorithm is the only algorithm that uses sponge construction from other candidate algorithms has arbitrary input and output that can be adjusted [13]. Keccak is a family of hash functions that are based on sponge construction and are used as building blocks for permutations of a set of 7 permutations [14]. Arbiter sponge function with fixed custom permutations, provides users with multiple functions from one fixed permutation, thus making implementation easier.

1). The Sponge Construction: In the context of cryptography, sponge functions provide a special way to generalize a hash function to a more general function whose output length is arbitrary. The sponge function creates a sponge construct, which is a simple iterative construct that constructs a variable-length input variable-length output function based on a fixed-length permutation (or transformation) [15]. The sponge construct is a simple iterative construct for constructing a function F with a variable input length and an arbitrary output length based on a fixed-length transformation or permutation f operating on a fixed number of bits b. Here b is called width. The sponge construction operates at the state b = r + c bit. The value of r is called the bitrate and the value of c is the capacity. First, all status bits are initialized to zero. The input message is filled in and cut into blocks of r bits. As depicted in Fig 2 the construction of the sponge takes place in two phases: an absorbing phase followed by a squeezing phase.

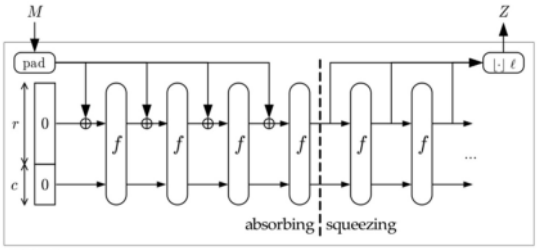

Fig . 2sponge construction.

- Absorption Phase: The r-bit input message block is XORed into the first r bits of state, inserted by application of the f function. When all message blocks are processed, the sponge construction goes into the squeeze phase.
- Squeeze Phase: The first r bits of state are returned as output blocks, interleaved with the application of the f function. The number of output blocks is chosen at will by the user.

The last c bit of state is never directly affected by the input block and is never outputted during the squeeze phase.

2). Keccak -f Permutations: Keccak -f has 7 fixed permutations represented by Keccak -f [b], where b = 25 × 2l and l ranging from 0 to 6. Keccak -f [b] is a permutation on Z_2^b where the bit number is from 0 to b -1. b is the width of the permutation.

Keccak permutation -f [b] is described as a sequence of operations on state a which is a three-dimensional array of GF(2) elements, namely a[5][5][w], with w = 2l. Expression a = [x][y][z] with x, y 5 and z w , indicates the bit in position (x, y, z). The mapping between bit s and bit a is s[w(5y + x) + z] = a[x][y][z]. Expressions in x and y coordinates must be taken in module 5 and expressions in z coordinates in w modules. Sometimes index [z] is omitted, both indexes [y][z] or all three indexes, implying that the statement is valid for all values of the omitted index.

Keccak -f [b] is a permutation iteration, consisting of a sequence of nr around R. It is indexed with nr from 0 to nr-1. Each round consists of 5 steps as follows:

$$R = \iota \circ \chi \circ \pi \circ \rho \circ \theta \quad (1)$$

$$\theta : a[x][y][z] \leftarrow a[x][y][z] + \sum_{y'=0}^{4} a[x-1][y'][z] + \sum_{y'=0}^{4} a[x+1][y'][z-1],$$

$$\rho : a[x][y][z] \leftarrow a[x][y]\left[z - \frac{(t+1)(t+2)}{2}\right], \quad (2)$$

with t satisfies $0 \le t < 24$ and $\begin{pmatrix} 0 & 1 \\ 2 & 3 \end{pmatrix}^t \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}$ in $GF(5)^{2x2}$, or $t = -1$ if $x = y = 0$

$$\pi : \quad a[x][y] \leftarrow a[x'][y'], \text{ with } \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix},$$

$$\chi : \quad a[x] \leftarrow a[x] = (a[x+1]+1)a[x+2],$$

$$\iota : \quad a \leftarrow a + \mathrm{RC}[i_r].$$

Addition and multiplication between terms are in GF. Except for the spin constant RC [i r], these spins are identical. The spin constant is given by (with the first index indicating the number of spins)

$$\mathrm{RC}[i_r][0][0][2^j - 1] = \mathrm{rc}[j + 7i_r] \quad \text{for all } 0 < j < \ell, \quad (3)$$

And every other "RC" [i_r][x][y][z] value is zero. The value of "rc" [t] "GF"(2) is defined as the output of the linear feedback shift register (LFSR):

$$\mathrm{rc}[t] = (x^t \bmod x^8 + x^6 + x^5 + x^4 + 1) \bmod x \text{ in } GF(2)[x]. \quad (4)$$

The number of rounds nr is determined by the width of the permutation, i.e.

$$n_r = 12 + 2\ell. \quad (5)$$

Quick Response Code (Qrcode) is the development of a *barcode* that has only one dimension to two dimensions by Denson Wave Corporation in Japan [16]. This development made major changes to the system, the form of the code, and the capacity of characters that can be stored. Barcodes can only store numbers and can only store twenty characters [17]. For full specifications, see the official website of the copyright or trademark holder of Denso Wave Corporation [18].

A very significant difference in barcodes to Qrcode refers to the added features and more complex structure that must be owned by Qrcode.

Structure Qrcode every Qrcode symbol must be built in a square module consisting of function patterns and an encoding region surrounded by a quiet zone border [19], as depicted in Fig . 3.

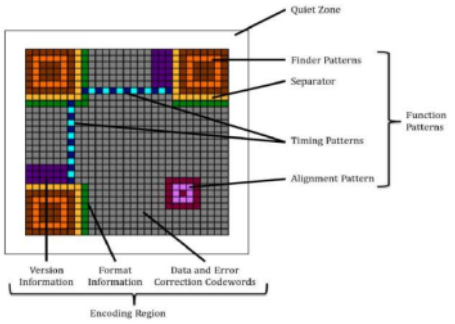

Fig. 3qrcode structure [19].

- Finder pattern: the detection pattern has a special position located in 3 corners, namely the top left, top right, and bottom left. This pattern is designed to ensure the direction of the QRcode pattern for decoding, this allows the QRcode to be scanned from any direction.
- Separator: is a white module that surrounds the pattern finder, so that it is separate from the encoding region.
- Timing pattern: time pattern consists of 1 × vertical and horizontal modules in black and white. This pattern serves to provide pattern density information, module coordinates, and version information area.
- Alignment pattern: alignment pattern is found in the symbols of Qrcode version 2 and higher, the position and number of patterns depend on the symbol version.
- Encoding region: the coding region contains format information, version information, data, and error correction codes. For format information, an array of one module should be backed up near the top left pattern finder, top right, bottom left and version information, a block area above the bottom left pattern finder, and 3×6 blocks to the left of the top right pattern finder.
- Quiet Zone is an empty (white) module that surrounds the entire Qrcode symbol to ensure that no misleading data occurs.

The system architecture is made to find out the efficient method to fulfill the system requirements. Combining HMAC method to guarantee the correctness of document data, using Keccak algorithm for hashing algorithm, and using Qrcode to store salt key. Judging from each function of the combination, there are several things that can complicate the falsification of the document, namely additional data or called salt. key is not stored in database. This can reduce the information that cybercriminals get if they break into databases.

III. Result and Discussion

Flowchart discussion that each major process consists of several units that have their respective roles and functions, in this section Table IV will describe what these units are.

TABLE I
FUNCTIONS OF THE UNIT SYSTEM

Unit	Destination
Making salt key	• Random word generation • Geberate qrcode .
import pdf,	• Converts an existing document into a template to modify. • Embed QRcode into pdf.
Hashing	• Hashing files _ _ pdf added with salt key.
Pdf to image	• Converts the last page of the file to an image.
Salt key return	• Degenerate QRcode.

Each unit is tested beforehand to find out how it is used and the problems that arise when the unit is installed. This can make it easier for us to diagnose if when all units are combined a problem arises, at least we already know which unit this problem occurs in. Unit testing is done by experimenting with the unit several times, with the following results:

4). Salt making unit key: is a unit test for making salt key and converting it to QRcode. This unit complies with FSR-01 and FSR-02.

5). Import units pdf: this test is done by uploading several files pdf with the different numbers of pages because in this unit there is a function for importing pdf with more than one-page conditions. Then the QRcode is listed on the last page. This unit complies with FSR-04.

6). Hashing units: hashing uses the HMAC method which converts the document into a string and then adds a salt key, which will generate a message with unique digests. This test scenario uses the same document but with salt different keys. The purpose of this scenario is to follow the advantages of the HMAC method, namely, even though the data or documents are the same, the results will be different because the salt the key added is different, so it still produces a message different digests. This unit complies with FSR-05.

7). pdf units to image: the expected test results are pages that have QRcode converted to image: png. The condition is that the QRcode is always on the last page. So the result is that the unit can issue an image from the last page.

8). Salt return unit key: this is the next step of the pdf unit to image, which scans the QRcode contained in the image and converts it into a string again. 3.12. This unit complies with FSR-07.

The purpose of the system created is to maintain the correctness of the data, therefore the result of the integration of all the functional units that build this application should be able to guarantee the correctness of the document. To find out the results, the author has made a scenario about how the document will be forged. This scenario indirectly explains how easy it is if previously only the signature is proof of the legality of a document. The following is Table V of test scenarios with several document conditions:

TABLE II
DOCUMENT TRUTH TEST SCENARIO

Condition	Results
Documents have been registered	Registered
Document not registered	Not Registered
QRcode copied from another document	Not Registered
One page missing document	Not Registered
Content or text changed	Not Registered
QRcode is not on the last page	Not Registered
The contents are the same but the document is regenerated.	Not Registered

This result occurs because the existing HMAC method in the PHP library retrieves the contents of the document along with its attributes and formats. So if a document has been registered but has tried to be changed (the contents are the same in the end), there will be a change in the attributes of the document. This causes the message and the resulting digest changes. The scenario is a cyber violation of the possible violations that can occur in the document.

IV. CONCLUSION

The use of a signature on a document is not sufficient as a benchmark for the authenticity of the document. Because the signature is only a marker for the owner of the signature. In addition, digital documents are very easy to change, fake, and even signatures can be taken and reproduced. Therefore, this document security system can anticipate these violations by using the HMAC method, the keccak algorithm, and QRcode. Experiments on possible violations that can occur have been carried out and the results answer the questions from the problem formulation, namely:

- The system managed to maintain the integrity of the document with the HMAC method using the Keccak Algorithm.
- The system can add random data that becomes a special mark on each document in the form of Qrcode,
- The system is in the form of a website that can be used at any time on any device that has a browser and internet connection.

This security should have become a minimum of the system of every organization or agency because long before technology like today, violations of digital document falsification have occurred. So that this research has minimized the possibility of this happening, at least on the document side.

The sensitive nature of the Keccak algorithm is very helpful in securing documents. Even if something changes a little, the output will be much different. This of course has fulfilled one of the CIA triads, namely Integrity. So that anyone who doubts the truth of the document will be answered with this system. Not only that, the trust between parties who share information through this document will increase, and if there is a problem this system can be proof of the truth of the document.

REFERENCES

[1] M. S. a. R. Euphrasia, "Data Security Through QR Code Encryption and Steganography," Advanced Computing: An International Journal (ACIJ), 2016.

[2] A. Sideris, T. Sanida and M. Dasygenis, "High Throughput Implementation of the Keccak Hash Function Using the Nios-II Processor," MDPI journals, vol. 8, 2020.

[3] R. A. Azdy, "TANDA TANGAN DIGITAL MENGGUNAKAN ALGORITME KECCAK DAN RSA," JNTETI, vol. 5, no. 3, pp. 184-191, Agustus 2016.

[4] H. Raad, "SECURITY," in FUNDAMENTALS OF IOT AND WEARABLE TECHNOLOGY DESIGN, ke1 ed., John Wiley & Sons, Inc, 2021, pp. 157-170.

[5] W. C. Barker, "Guidline for identifying an Information System as a National Security System," NIST Special Publication 800-59, Agustus 2003.

[6] E. Barker and W. C. Barker, "Recomendation for Key Management," NIST Special Publication 800-57, Mei 2019.

[7] Y. Shi, "Data Security and Privacy Protection in Public Cloud," Viterbi School of Engineering, University of Southern California., Los Angeles, CA., 2019.

[8] U. Rathod, M. Sonkar and B. R. Chandavarkar., "An Experimental Evaluation on the Dependency between One-Way Hash Functions and Salt," in 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, IEEE, 2020, pp. 1-7.

[9] K. Yasuda, ""Sandwich" Is Indeed Secure: How to Authenticate a Message with Just One Hashing," in Australasian Conference on Information Security and Privacy, Australia, 2007.

[10] M. Ichwan, M. Gustian and N. R. Nurjaman, "Implementasi Keyed-Hash Message Authentatication Code Pada Sistem Keamanan Rumah," MIND (Multimedia Artificial Intelligent Networking Database), 2016.

[11] IEEE Organization, "IEEE Standard Specification for Password-Based Public-Key Cryptographic Techniques," IEEE Std 1363.2-2008, pp. 1-140, 2009.

[12] A. Cui, M. Li, G. Qu and H. li, "A Guaranteed Secure Scan Design Based on Test Data Obfuscation by Cryptographic Hash," vol. 39, IEEE, 2020, pp. 4524-4536.

[13] N. Muhammad Asghar, R. Pulungan and M. Riasetiawan, "Data Integrity and Security using Keccak and Digital Signature Algorithm (DSA)," IJCCS (Indonesian Journal of Computing and Cybernetics Systems), vol. 13, 2019.

[14] G. Bertoni, J. Daemen, M. Peeters and G. V. Assche, "Keccak Specifications," 10 September 2009. [Online]. Available: https://keccak.team/specifications.html.

[15] G. Berrtoni, J. Daemen, M. Peeters and G. V. Assche, "Cryptographic Sponge Functions," 14 Januari 2011. [Online]. Available: https://keccak.team/sponge_duplex.html.

[16] A. F. Adri and A. F. Suni, "IMPLEMENTASI QR CODE VALIDATION PADA SISTEM INFORMASI SURAT PERINTAH PERJALANAN DINAS," Jurnal Sistem Komputer, vol. 10, no. 1, pp. 10-16, 2020.

[17] F. Masalha and N. Hirzallah., "A Student Attendance System Using QR Code," International Journal of Advance Computer Science and Applications, vol. 3, no. 5, pp. 75-79, 2014.

[18] Denso Wave Corporation, "QR code.com," [Online]. Available: https://www.qrcode.com/en/.

[19] S. Tiwari, "An Introduction to QR Code Technology," in 2016 International Conference on Information Technology (ICIT), Bhubaneswar, IEEE, 2016, pp. 39-44.

Integrity Assurance System for Document Security Using Keccak and Quick Algorithm Response Code

ORIGINALITY REPORT

29%	24%	21%	18%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.ru.nl Internet Source	4%
2	www.ijarse.com Internet Source	3%
3	keccak.noekeon.org Internet Source	3%
4	www.ijirset.com Internet Source	2%
5	hdl.handle.net Internet Source	2%
6	Urvesh Rathod, Meghna Sonkar, B. R. Chandavarkar. "An Experimental Evaluation on the Dependency between One-Way Hash Functions and Salt", 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020 Publication	2%
7	webpages.charlotte.edu Internet Source	2%

8	Sumit Tiwari. "An Introduction to QR Code Technology", 2016 International Conference on Information Technology (ICIT), 2016 Publication	1 %
9	Faiz Muqorrir Kaffah, Yana Aditia Gerhana, Ihsan Miftahul Huda, Ali Rahman, Khaerul Manaf, Beki Subaeki. "E-Mail Message Encryption Using Advanced Encryption Standard (AES) and Huffman Compression Engineering", 2020 6th International Conference on Wireless and Telematics (ICWT), 2020 Publication	1 %
10	www.journalijar.com Internet Source	1 %
11	www.researchgate.net Internet Source	1 %
12	dokumen.pub Internet Source	1 %
13	Mohamad Irfan, Pramadita Sielda Dewi, Wildan Budiawan Zulfikar, Cepy Slamet, Ichsan Taufik. "Sentiment Analysis as Assessment of the COVID-19 Social Assistance Pollemic using Random Forest Algorithm", 2022 8th International Conference on Wireless and Telematics (ICWT), 2022 Publication	1 %

14	link.springer.com Internet Source	1 %
15	pdfs.semanticscholar.org Internet Source	1 %
16	csrc.nist.gov Internet Source	1 %
17	www.semanticscholar.org Internet Source	1 %
18	Submitted to University of Glasgow Student Paper	<1 %
19	www.coursehero.com Internet Source	<1 %
20	Aijiao Cui, Mengyang Li, Gang Qu, Huawei Li. "A Guaranteed Secure Scan Design based on Test Data Obfuscation by Cryptographic Hash", IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2020 Publication	<1 %
21	Lecture Notes in Computer Science, 2015. Publication	<1 %

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		