Neural Network Approach for Secure Hash-Based Color Image Authentication and Analysis

Yakup Kutlu
Department of Computer Engineering Mustafa Kemal University, Hatay, Turkey

Apdullah Yayik
Turkish Land Forces, Hatay, Turkey

This work is supported by 8702 numbered Mustafa Kemal University Science Research Project.

Abstract
Various neural network based image hash functions have been proposed, in related works. However, statistical and security analysis must validate usage of them in military applications. This paper proposes a new neural network based hash function approach for color image authentication. Proposed hash function re-sizes input color image to a constant size, then generates hash values using neural network one-way property and non-linear techniques, for 3 dimensions respectively. As a result security and performance analysis are performed and satisfying results are achieved. The proposed system is robust, useful and secure.

Keywords: image authentication, cryptology, hash function, statistical and security analysis.

1. Introduction

Cryptology is a master of science in mathematics and computer engineering that deals with the most required topics in 21th century like, information and cyber security. Cryptography uses mathematical methods for information security. Information security is now a necessary component of commercial implementations, military communications and also social media applications. It can be many threats and attacks that can be made to these networks by people with malicious intent. Cyber-terrorists, crackers, hackers, so-called 'script kiddies' and industrial spies are all masters in the manipulation of information systems [1]. Current communication techniques, using computers, routers or switches connected through huge networks, make all data even more vulnerable for these threats. Cryptography is by far the most significant part of communication security [2]. Any cryptosystem must require confidentiality, authentication, integrity and non-repudiation. Authentication relates to the identification of two parties entering into communication, while integrity addresses the unauthorized modification of an element inserted into the system [3]. Up to date, there have been a large number of studies intended to advance robust and secure cryptosystems [2–11] and use them in military communications. It is a novel and growing technique to perform the non-linear property of neural network to create hash function.

Hash functions convert major definitions to minor values. As an input any message can be used, and as an output fixed length hexadecimal value is produced. Most popular hash functions are MD-5 [12], SHA-2 [13] that is published in 2002 by the NIST U.S FIPS (Federal Information Processing Standard), SHA-3 [14] that is based on an instance of the KECCAK algorithm that NIST selected as the winner of the Cryptographic Hash Algorithm Competition in 2013.

Hash functions can be used for data integrity and digital signature. Digital signature signs data in order to confirm the integrity of data and identity of sender. Hash function is the digest of message which is attached to original message. Any modifying in original message makes hash function disabled.

In other words; hash function is an information generating process from any message using mathematical methods. Generated digital information is called message digest. Recycling or reverse engineering of hash function must be impossible, so hash function must not inspire anyone about the original message.

Also, it must be impossible to predict different messages whose hash values are the same. Cryptography needs functions like this because they are able to provide safety communications [15].
Hash functions are also used in Network and Internet Security. Domain or local controlled PC user's passwords are saved in system file server manager as its hash value, so administrator or super users of WAN (Wide Area Network) or LAN (Local Network Area) is not able to see user's original passwords. Also any malicious accessed to server database cannot capture user's passwords.
In related works, there have been lots of studies to advance robust machine learning based hash-functions and use them in communications up to now [4,9,10,16–23]. Near past and recently there is relatively much interest in using neural networks for cryptography[19]. Ref.[23] Triggered our research team for this study, statistical analysis for sensitivity of SHA-2 secure hash algorithm and neural based hash function are nearly same, so it can be said that neural network will be used in cryptology in near future.

Survey of hash-based image authentication studies

A number of researchers have dedicated themselves to advance high statically and security performance image hash functions.
There are many different approaches for image hash function algorithms. Jin et al. [24] proposed Radon Transform based image fingerprinting (hashing). Monga and Evans [25] extracted vital image features using wavelet-based feature detection algorithm in order to advance image hashing system. Swaminathan et al. [26] introduced rotation invariance of Fourier-Mellin transform and controlled randomization based image hashing algorithm. Wu et al. [27] proposed print-scan resistant image hashing algorithm based on the Radon and Wavelet Transform. Chuan Qin et al. [28] proposed image hashing method in discrete Fourier transform domain for image authentication. Monga and Mıhçak developed image hashing algorithm based on dimensionality reduction technique called non-negative matrix factorization (NMF). Sengupta et al. [29] developed image hash based authentication for thermal images. Fawad et al. [30] advanced randomized pixel modulation and wavelet transform Based image authentication scheme.
In last two decades neural network based hash function is studied by some researchers [16–19,21–23,31,32]. Common feature of these neural network based related works is considering just gray scale images or texts (Only ref [32] performed color image hashing with insufficient security analysis) that ignores color images' hue and saturation features and restricts their distinct capacities.
In this paper, secure and robust neural image hash function color image, which considers all features of color image is proposed, for novelty. Then, many experiments are performed to validate its security and statistical requirements.
The rest of this paper is organized as follows. Section 2 describes proposed image hash function. Section 3 presents experimental results and performance analysis. Finally conclusion is given in Section 4.

Materials and Methods

Process of neural network based secure image hash function diagram is shown in Figure 1.

Proposed neural network based hash function

In the proposed hash function, artificial neural network in Figure 2 is used which has three layers that carries out ideal hash functions confusion, diffusion and compression properties.
Let the inputs and outputs of layers be

$$K = (K_0 K_1 K_2 .... K_{512}),$$
$$C = (C_0 C_1 C_3 ..... C_8),$$
$$D = (D_0 D_1 D_3 ..... D_8), \quad (1)$$
$$H = (H_0 H_1 H_2 .... H_{32}),$$

and let transfer functions be f_1, f_2 and f_3, let weight and biases be, w_1, w_2, w_3, b_1, b_2 and b_3 so neural network can be defines as;

$$\begin{aligned} H &= f_3(w_3 * D + b_3) \\ &= f_3(w_3 f_2(w_2 C + b_2) + b_3 \\ &= f_3(w_3 f_2(w_2 f_1(w_1 K + b_1) + b_2 + b_3 \end{aligned} \quad (2)$$

Here w_1 is of 30 x 512 size, w_2 of 32 x 30 size, b_2 of 30x1 size and b_3 of 32x1 size, f_1, f_2 and f_3 are sigmoid functions.

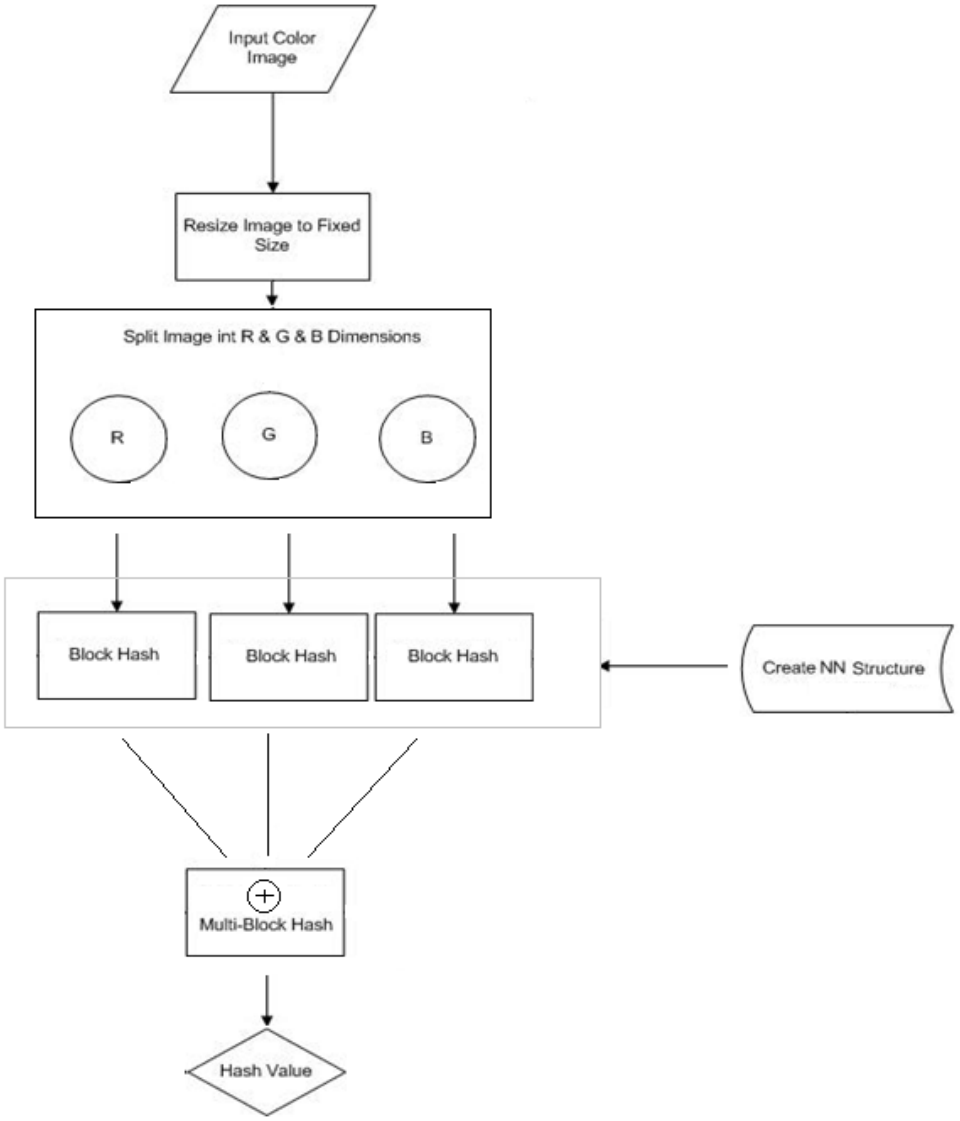


Figure 1. Neural Network Based Hash Value Generator Diagram

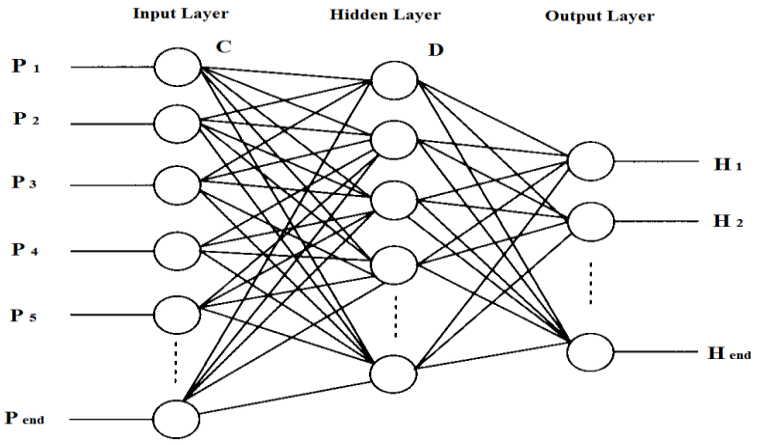


Figure 2. Neural Network Structure

Multi Block Hash Algorithm

Multi-block hash algorithm consists of several block hash model and XOR operator combinations. Color image is splited into blocks according to dimensions and H_i values are calculated. In other words RGB format color image is separated as component of R, G and B dimension blocks.

Color image is splited into blocks according to dimensions and H_i values are calculated. In other words let image be Img; these blocks are calculated as;

$$BlockR = \mathrm{Im}\, g(:,:,1)$$
$$BlockG = \mathrm{Im}\, g(:,:,2) \quad (3)$$
$$BlockB = \mathrm{Im}\, g(:,:,3)$$

Then neural network structures are created for each block. After block hash algorithm apply XOR operation for combination and hash value.

Block Hash model

Neural network based hash function is depicted in Figure 3. Neural Network consists of 1 (one) hidden layer, 512-30-32 neuron structure and sigmoid activation functions.

$$H_r = H_{r1} \oplus H_{r2} \oplus H_{r3} \ldots . H_i$$
$$H_b = H_{b1} \oplus H_{b2} \oplus H_{b3} \ldots . H_{bi} \quad (4)$$
$$H_g = H_{g1} \oplus H_{g2} \oplus H_{g3} \ldots . H_{gi}$$

Each dimensions are passed through the block hash and 32 *x* 512 sized H_i is performed. XOR values of H_i consecutive rows are calculated in order to obtain 1 *x* 512 binary H value (1).

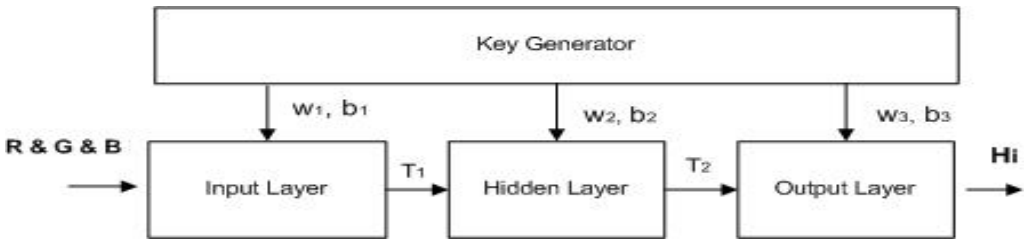


Figure 3. Block Hash Model

The block hash encodes 32 *x* 512 bits to 1 *x* 512 bits. These application are performed three times for R, G and B dimensions, respectively.

Following, these blocks are encoded with combined several block hash and XOR operators that is called multi-hash block depicted in Figure 4 and (2).

$$H_{rgb} = H_r \oplus H_g \oplus H_b \quad (5)$$

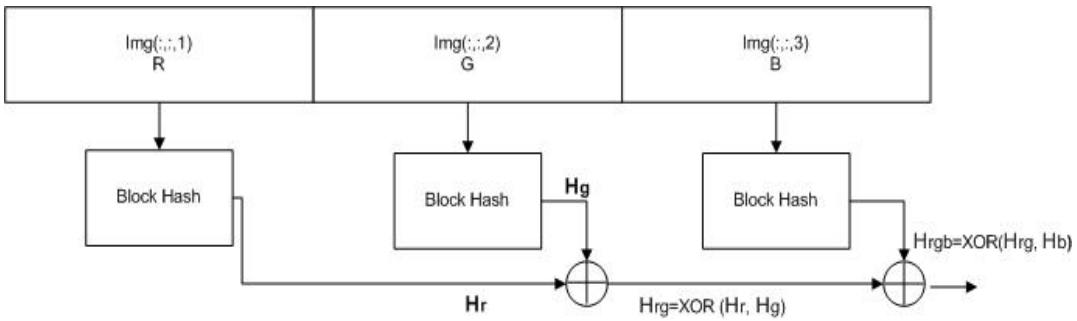


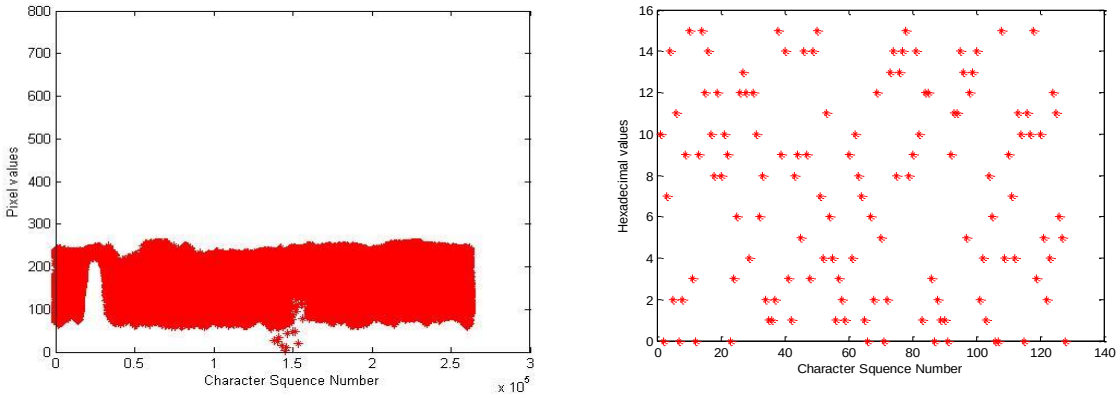
Figure 4. Multi Block Hash Model

Performance Analysis

In this section, whether proposed hash function satisfies statistical and security requirements or not is analyzed. So that, statistical distribution, diffusion and confusion, collision resistance and meet-in-the-middle analysis are performed.

Statistical Distribution of Hash Value

Security of hash function is directly proportional with uniform distribution of hash value. While, genuine image is localized in minimal area and hexadecimal hash value is localized uniformly, as seen in Figure 5.



(a) (b)

Figure 5. Distribution of (a) original image pixel values, (b) hash values **Hata! Başvuru kaynağı bulunamadı.** shows binary hash values of LENA image with different changes. In Table 3 sensitivity of hexadecimal hash values of LENA image with different changes are shown.

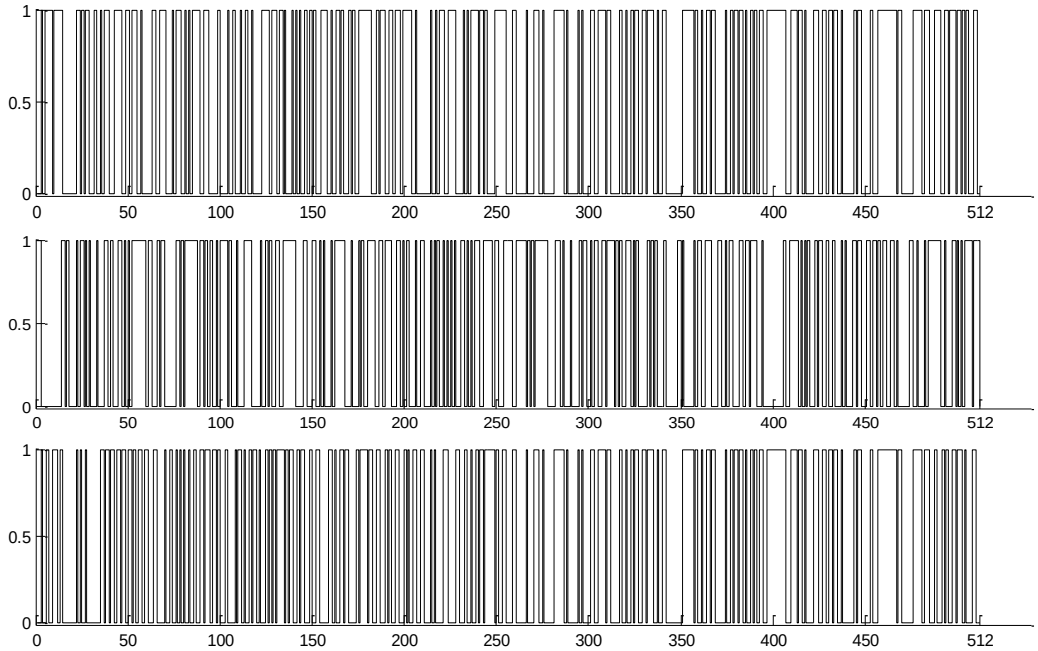


Figure 6. 512 bit binary hash values of LENA image with least difference

Statistical Analysis of Diffusion and Confusion

Binary format of hash value consists of only 0 and 1 bits, while hexadecimal hash value consists of 16 different characters. Because of these changes in binary hash value must be nearly 50% (as shown in Table 1), in contrast changes in hexadecimal value must be nearly 100%, for each modification. Otherwise diffusion property does not satisfy. In order to control binary and hexadecimal hash value changes following steps are applied:

Algorithm of Statistical Analysis of Diffusion and Confusion

for all Q

1: Calculate original image binary and hexadecimal hash value

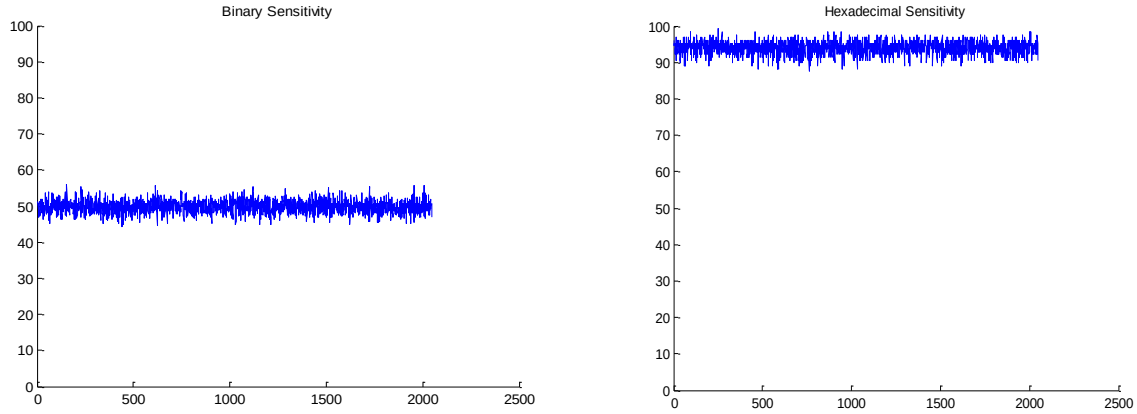
2: Modify value of image 10 pixels, randomly.

3: Calculate modified image's binary and hexadecimal hash value.

4: Compare and find differences between original and modified image's binary and hexadecimal hash values.

end for

Figure 7, binary and hexadecimal sensitivity of hash value is presented. As it is mentioned binary sensitivity is nearly 50% that satisfies diffusion of hash value. Also, almost 100% hexadecimal sensitivity means that the algorithm has very high ability of robustness.



(a) (b)

Figure 7. Sensitivity of (a) Binary and (b) Hexadecimal Hash Value

Statistical parameters for binary sensitivity are defined below: Mean number of bits changed:

$$\overline{B} = \frac{1}{N}\sum_{i=1}^{N} B_i \tag{6}$$

Standard deviation of changed bits:

$$\Delta B = \sqrt{\frac{1}{N-1}\sum_{1}^{N}(B_i - \overline{B})^2} \tag{7}$$

Mean changed probability:

$$P = (\overline{B}/512)\times 100\% \tag{8}$$

Standard deviation of probability:

$$\Delta P = \sqrt{\frac{1}{N-1}\sum_{1}^{N}(B_i/128 - P)^2}\times 100\% \tag{9}$$

where;

$N =$ Character number of binary hash value (512)

Through tests $Q = 256,512,1024,2048$ performed and results are listed in Table 1.

Table 1. Statistics of the number of bit changed

	***Q*=256**	***Q*=512**	***Q*=1024**	***Q*=2048**	***Mean***
B min	210	229	228	227	224
B max	280	280	290	286	284
B	254.6	254.87	254.87	255.21	255
ΔB	8.57	8.37	8.80	8.70	8.61
P min	41.06	44.72	44.53	44.33	43.66
P max	54.68	54.68	56.64	55.85	55.46
P %	49.72	49.77	49.77	49.84	49.78
ΔP	4.92	4.92	4.92	4.93	4.92

One-Way Property

Neural Network's most important and engaging ability that makes them convenient for implementations is their generalization capability that is their ability to produce logical outputs when they are simulated with inputs not formerly met. If targets' size y_k is so different from inputs' size x_k it is difficult to compute target from input, while it is easy to compute input from target. Due to this property neural network can be used in hash functions [21]. Parallel implementation is a significant property of neural networks. Each layer is paralleled. So, they can implement specific functionality independently. According to this, ANN are available for data progressing Neural networks has ability to make relationship using training function with non-linear and complicated values. Confusion is a particular ability caused by the nonlinear topology of artificial neural networks. This ability makes the output depend on the input in a nonlinear and complex type. It defers that, a bit of output is in relation with all the bits of the input in a complex way. Due to this, it is so hard to calculate the complete input. The confusion ability of artificial neural networks makes them indispensable material for hash function designing.

Collision Resistance Analyze

After generating hash value, experimenters must make sure that each bit of original image effects hash value, in other words hash value is fully depended on original image. Otherwise single bit change in image do not affects hash value that means vital information security vulnerability.

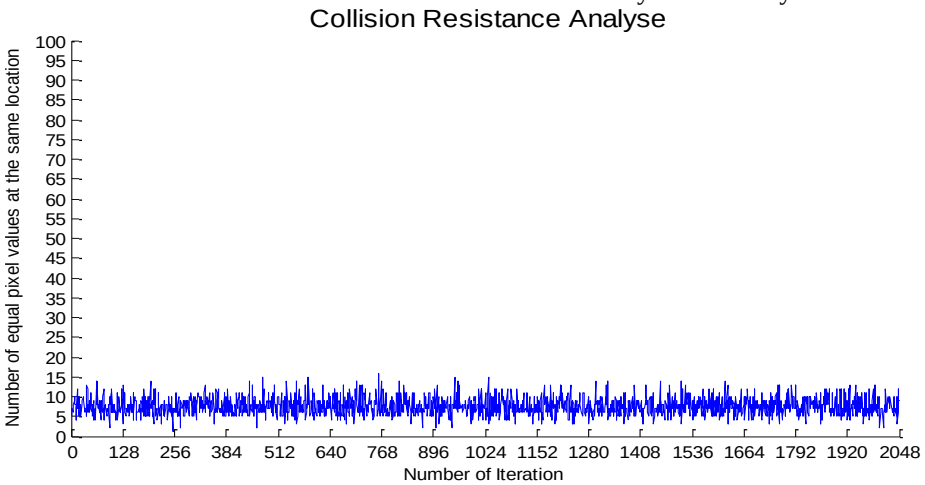


Figure 8. Collision Resistance Analyze

So in this paper, collision resistance analyze is performed Q times as follows:
Graph of dispersion of the number of collision hits is illustrated in Figure 8.

Algorithm of Collision Resistance Analyze

for all Q

1: Generate hash value of original image (described in section 3.2) and store in ASCII format.
2: Randomly modify least bits in original image
3: Generate hash value of new modified version of original image and store in ASCII format
4: Compare hash values generated in 1 and 3. Find and count same ASCII values at the same location.(3)

end for

Maximum, minimum, mean values and significance level are listed in Table 2.**Hata! Başvuru kaynağı bulunamadı.**

Table 2. Collision Resistance Statistical Parameters

Iteration	*Maximum*	*Minimum*	*Mean*	*Standard Error*
256	15	1	7.39	2.31
512	16	2	7.45	2.31
1024	15	1	7.59	2.27
2048	16	1	7.52	2.20

Meet-In-The-Middle Attack

Meet-in-the middle attack is a technique of cryptanalysis against a block cipher introduced in 1977 [33]. It is a passive attack; it may allow the attacker to read messages without authorization, but against most cryptosystems it does not allow to alter messages or send [34]. The attacker must be able to calculate possible values of the same intermediate variable (the middle) in two independent ways, starting either from the original or from the hash value. The attacker calculates some possible values each way and compares the results. If original image is $M = (M_0 M_1 M_2 ..... M_{n-1} M_n)$ its hash value $= H$. Expected image found using meet-in-the-middle attack is $M' = (M_0 M_1 M_2 ..... M_{n-1} M_n')$ its hash value $= H$. In other words attack process is replacing M_n with M_n'. But attacker cannot create M_n' that is not in relation with hash value described in previous sections.

Table 3. 128 Bit Hash Values Of Original Lena Image And Its Modified Versions

Input	Hash Value	Binary Sensitivity (%)
Original Lena	068756D11AE72B598F9953C0CFE66AA114A7CF3A210ABCB7581895E6E92E9D6D8ED579B27ADC0FBDA2812416AFE8E5BD9BBF9056146E2E742F16FC8BB85505CF	0
10-bit randomly changed Lena	661AE1EA248B16D4BC42235E0EB7684A5DFD37A874EE7D8B03A1B973FDF7B687FA90DAA3FBB7FE0255A9022C15A7F8BA2CFDBB4B63EFD39BB6092F3F2FD804DE	51.36
10-bit randomly changed Lena	5F2E4D4C6860C6D2326029A6FCD6A26652A4FBC2FC22D9E7AB67B7775F3051A345EE3E8A400116803B19D330D9A4681622F9D97ECA62F2322C5B774CC3548F40	50.97
10-bit randomly changed Lena	C0182D04FE3799ABEECA8F422ADC6AE2D2211D8A16A86E86AB33D15342086AC052328650DEAEF88C9F288661401BB9DF6FDD141ADB36C0E730B20D1978FF15CB	51.36
10-bit randomly changed Lena	A25E1913BD31BFEE88EA8A116FFE4F9682203C9E31897F83CD55A54213186B962163E523EFBDDE88D91CC210213AB9CE7EFD150BDB06F3EDAB0B8480F366AE50	51.17
10-bit randomly changed Lena	6E2E6D6C6860C6D232602994DCD6A26652B4FBC2FC22D976AB67A7775F2050B154FF2E8A400507903B1FD330D9A4681622F9F97ECA62F2322C5B665C83541F40	50.97

Conclusion

Artificial Neural Network based secure hash algorithm is presented and analyzed here. Proposed algorithm is efficient to require diffusion and confusion properties due to neural network information transfer process inspired from real biological systems. Analysis and experiments explained in this paper revealed that hash function satisfies sensitivity, minimum collision hit requirements and powerful against attacks like meet-in-the-middle.

In Figure 5, uniform distribution of hexadecimal hash value against local distribution of original image means high randomness that requires confusion. In Figure 7, nearly 50% difference of binary format of hash value means high sensitivity that requires diffusion. But Figure 7 is not sufficient only by itself. In order to correlate Figure 7, statistically approaches are shown in Table 1. When looping sensitivity testing process as Q times, average 254, minimum 224, maximum 284 bits of 512 bits differs with minor standard deviation (8. 61) and minimum 43. 66%, maximum 55. 46%, average 49. 78% of 512 bits differs with minor standard deviation (4. 92). These results satisfy sensitivity of neural network based hash function.

Calculation of same ASCII hash values at the same location that is called collision resistance is performed as Q times. Figure 8 illustrates collision resistance when Q =2048. When looping collision resistance testing process as Q times, average 7 bit same ASCII values are found at the same location that can be ignored due to minority. So these results satisfy collision resistance of neural network based hash function.

As a result; this system can be used in communication applications especially in military.

References

[1] Sivatha S S, Geetha S and Kannan A 2012 Expert Systems with Applications Decision tree based light weight intrusion detection using a wrapper approach *Expert Syst. Appl.* **39** 129–41

[2] Arvandi M, Wu S, Sadeghian A, Melek W W and Woungang I 2006 Symmetric Cipher Design Using Recurrent Neural Networks *Int. Jt. Conf. Neural Networks* 2039–46

[3] Sağıroğlu S and Özkaya N 2007 Neural Solutions for Information Security *J. Polytech.* **10** 21–5

[4] Lian S, Liu Z, Ren Z and Wang H 2006 Hash function based on chaotic neural networks *IEEE Int. Symp. Circuits Syst.* 4

[5] Li Y, Xiao D, Deng S, Han Q and Zhou G 2011 Parallel Hash function construction based on chaotic maps with changeable parameters *Neural Comput. Appl.* **20** 1305–12

[6] Desai V, Patil R and Rao D 2012 Using Layer Recurrent Neural Network to Generate Pseudo Random Number Sequences *Int. J. Comput. Sci. Issues* **9** 324–34

[7] Hughes J M and College K 2007 Pseudo-random Number Generation Using Binary Recurrent Neural Networks

[8] Ruttor A 2006 *Neural Synchronization and Cryptography* (Bayerischen Julius-Maximilians-Universitat at Wurzburg PhD Thesis)

[9] Yayık A and Kutlu Y 2013 *Neural Network Based Cryptology Applications* (Mustafa Kemal University Deparment of Informatics M.S.Thesis)

[10] Yayık A and Kutlu Y 2013 Neural Network Based Hash Function for Text *International Conferance on Cryptology and Information Security* pp 257–62

[11] Yayık A and Kutlu Y 2013 Improving Pseudo Random Number Generator Using Artificial Neural Networks *Signal Processing and Communication Conferance*

[12] R. Rivest: 1992 The MD5 Message Digest Algorithm. *MIT Lab. Comput. Sci. RSA Data Secur. Inc*

[13] NIST 2002 Secure Hash Standard *FIBS 180-2 Publ.*

[14] NIST 2014 SHA-3 Standard: Permutation-Based Hash and Extendable-Output Functions *FIBS 202 Publ.*

[15] Soyalıç S 2005 *Cryptographic Hash Functions And Its Applications* (Erciyes University Natural and Applied Sciences M. S. Thesis)

[16] Xiao D, Liao X and Wang Y 2009 Parallel keyed hash function construction based on chaotic neural network *Neurocomputing J.* **72** 2288–96

[17] Shiguo Lian, Sun J and Wang Z 2006 Secure hash function based on neural network *Neurocomputing* **69** 2346–50

[18] Huang Z 2011 A more secure parallel keyed hash function based on chaotic neural network *Commun. Nonlinear Sci. Numer. Simul.* **16** 3245–56

[19] Lian S, Sun J and Wang Z 2007 One-way Hash Function Based on Neural Network *CoRR*

[20] Yee L P and Silva L C De 2002 Application of MultiLayer Perceptron Network as a one-way hash function *Proc. 2002 Int. Jt. Conf. Neural Networks. IJCNN'02 (Cat. No.02CH37290)* 1459–62

[21] Desai V 2013 Image Hash using Neural Networks *Int. J. Comput. Appl. (0975* **63** 12–8

[22] Yang Q, Gao T, Fan L and Gu Q 2009 Analysis of One-way Alterable Length Hash Function Based on Cell Neural Network *2009 Fifth Int. Conf. Inf. Assur. Secur.* 391–5

[23] Sumangala G, Kulkarni V R, Sali S and Apte S 2011 Performance Analayis of Sha-2 Algorithm With And Without using Artificial Neural Networks *World J. Sci. Technol.* **1** 12–20

[24] Seo J S, Haitsma J, Kalker T and Yoo C D 2004 A robust image fingerprinting system using the Radon transform *Signal Process. Image Commun.* **19** 325–39

[25] Monga V, Evans B L and Member S 2006 Perceptual Image Hashing Via Feature Points : Performance Evaluation and Tradeoffs *IEEE Trans. IMAGE Process.* **15** 3453–66

[26] Swaminathan A, Member S, Mao Y and Wu M 2006 Robust and Secure Image Hashing *IEEE Trans. Inf. FORENSICS Secur.* **1** 215–30

[27] Wu D, Zhou X and Niu X 2009 A novel image hash algorithm resistant to print–scan *Signal Processing* **89** 2415–24

[28] Qin C, Chang C-C and Tsou P-L 2013 Robust image hashing using non-uniform sampling in discrete Fourier domain *Digit. Signal Process.* **23** 578–85

[29] Sengupta M, Mandal P, Das T and Dey A 2013 A Novel Hash based Technique for Thermal Image Authentication *Procedia Technol.* **10** 147–56

[30] Ahmed F, Siyal M Y and Uddin V 2010 A secure and robust hash-based scheme for image authentication *Signal Processing* **90** 1456–70

[31] Reyhani S Z and Mahdavi M 2007 User Authentication Using Neural Network in Smart Home Networks *Int. J. Smart Home* **1** 147–54

[32] Engineering I 2014 Choatic Neural Network Based Hashing Algorith For Image *Int. J. Adv. Res. Electr. Instrum. Eng.* **3** 6690–7

[33] Diffie W and Hellman M E 1977 Special Feature Exhaustive Cryptanalysis of the NBS Data Encryption Standard *Computer (Long. Beach. Calif).* **10** 74–84

[34] Vanstone S A ., A.J.Menezes; and P.C.Oorshot 1996 *Handbook of Applied Cryptography* (Boca Raton: CRC Press)