Journal of Computing and Information Technology - CIT 23, 2015, 4, 317–328
doi:10.2498/cit.1002639

Building Secure and Fast Cryptographic Hash Functions Using Programmable Cellular Automata

Alaa Eddine Belfedhal and Kamel Mohamed Faraoun
Computer Sciences Department, Djillali Liabes University of Sidi Bel Abbes, Algeria

Cryptographic hash functions have recently brought an exceptional research interest. With the increasing number of attacks against the widely used functions as MD5, SHA-1 and RIPEMD, the need to consider new hash functions design and conception strategies becomes crucial. In this paper, we propose a fast and efficient hash function using programmable cellular automata that are very suitable for cryptographic applications due to their chaotic and complex behavior derived from simple rules interaction. The proposed function is evaluated using several statistical tests, while obtained results demonstrate very admissible cryptographic properties such as confusion/diffusion capability and high sensitivity to input changes. Furthermore, the hashing scheme can be easily implemented through software or hardware, so it provides very competitive running performances.

Keywords: cryptographic hash function, programmable cellular automata, cellular automata rule, pseudo-randomness, avalanche effect

1. Introduction

Cryptographic hash functions are considered as key components of almost all cryptographic protocols, and of many security applications. Their common usage scenarios range from the reduction of the amount of data to be signed, to timestamping or checking a file's integrity and enabling user's authentication across a network. Hash functions compute a *digest* of a given message which is a short and fixed-length bitstring. For a given message with an arbitrary length, the message digest, or *hash value*, is considered as its fingerprint (i.e., a unique and compact representation of a message). Unlike all other cryptographic algorithms, hash functions do not need a key to compute the message's hash, except in some specific implementations named keyed hash functions that are used for specific signature scenarios. The use of hash functions in cryptography is manifold: hash functions are an essential part of digital signature schemes and message authentication codes (MACs). Hash functions are also widely used for other cryptographic applications, e.g., for storing of password hashes or key derivation [1].

From a structural point of view, cryptographic hash functions may be categorized into three main classes, based on the nature of the operations performed in their internal compression functions: the first class functions use a block cipher as compression functions [2] and [3], while the second class includes hash functions based on difficult mathematical problems, such as discrete logarithm problem [4], factorization problem [5] and the problem of finding cycles in expander graphs [6].

Another recently emerging class covers dedicated hash functions designed especially to achieve optimal speed and reliability [7]. Among such dedicated functions, some new schemes are based on the use of chaotic maps that exhibit chaotic and complex behaviors with high sensitivity to initial conditions' variations. Examples of such maps include the logistic map proposed to build a hashing schema in [8], the tent map used in [9] and the Kolmogorov systems used in [10]. In addition, cellular automata constitute another class of dynamical systems that have been extensively used to design dedicated hash functions. Cellular automata (CA) are very appropriate to design hash functions with a low hardware and software complexity because of

their logical operation attributes [7], their inherent parallelism and extreme sensitivity to initial conditions alterations. Hence, CA-based hash functions are able to achieve very high performance speeds [10]. The main approaches using CA to construct hash functions are discussed in detail in related works section.

In the present work, we propose a novel cryptographic hash function using programmable CA conjointly with the Davies-Mayer construction [11] used to define compression function, when the Merkle-Damgard variant construction is adopted as a domain extension algorithm. The proposed scheme is shown experimentally to be robust against the main hashing attacks, and evaluated with respect to the strict avalanche criterion and the pseudorandomness of the hashing output considered for large scale streams. Obtained results are compared to several existing hash functions. The rest of the paper is organized as follows: Section 2 introduces relevant background about cryptographic hash functions, CA preliminaries with related works. Section 3 details the proposed hash function construction, while Section 4 presents the performed experiments with corresponding obtained results. Finally, conclusions are drawn in Section 5.

2. Cryptographic Hash Functions and Cellular Automata

In the following, we introduce the basic necessary definitions related to cryptographic hash functions and to cellular automata preliminaries. We briefly describe the main works with relevance to the construction of CA-based hash functions.

2.1. Cryptographic Hash Functions

Formally speaking, a hash function H is a deterministic and efficient algorithm that maps an arbitrary length binary message M to some fixed length (typically 128, 160, 256 or 512 bits) fingerprint h [1].

$$\begin{aligned} H : \{0,1\}^* &\rightarrow \{0,1\}^n \\ M &\rightarrow h = H(M) \end{aligned} \quad (1)$$

In order to be considered cryptographic, a hash function should resist the three main attacks: the pre-image attack, the second pre-image attack and the collision attack. The pre-image attack consists for a given hash value h to find the corresponding message M such that $H(M) = h$. The second pre-image attack succeeds if for a given message M and a given hash value $h = H(M)$ we can find a message $M \neq M'$ such that $H(M') = h$. Finally, the collision attack aims to find two arbitrary distinct messages M and M' such that $H(M) = H(M')$.

Using a brute-force attack, pre-images and 2^{nd} pre-images attacks succeed deterministically after 2^n call to the function H, when a collision is very leaky to succeed using only $2^{n/2}$ call to H according to the birthday paradox theorem. It is usually the goal in the design of a cryptographic hash functions that no attacks perform better than the brute-force attack.

In general, it is not always possible to get a formal proof of resistance to such attacks. But, in contrast, some statistical properties are easily verifiable to show that a given hash function has a good cryptographic level of security. A hash function that behaves like a pseudorandom function and satisfies the avalanche effect is generally considered to be secure and can be used safely for cryptographic purposes. The avalanche effect reflects the sensitivity of the hash function to elementary changes in the hashed message: a little change in the input message (flipping one single bit) produces a significant change of the output (the final hash). Such statistical properties are easy to check, and then enable a fast and acceptable evaluation tool to validate the hash functions design.

Cryptographic hash functions have generally two main independent components: the mode of operation (a domain extender algorithm) and the compression function. Most popular hash functions are based on iterating a compression function that processes a fixed number of bits. The message to be hashed is split into blocks of a certain length where the last block is possibly padded with extra bits.

Let $h : \{0,1\}^n x \{0,1\}^L \rightarrow \{0,1\}^n$ denote a compression function, where n and L are positive integers, and let $M = m_0|m_1|\ldots|m_k$ be the message to be hashed, where $|m_i| = L$ for $0 \leq i \leq k$. The hash value is then defined to be h_k, where $h_i = h(h_{i-1}, m_i)$ defines the chaining variables.

A hash function can be either keyed or non-keyed, depending on the corresponding intended use. For non-keyed functions addressed in the present work, a fixed initialization vector IV is used to define the value h_0. If the message M to be hashed cannot be split into blocks of equal length n, (if the last block consists of less than n bits), then a collision-free padding rule should be used [12].

It has been shown that attack on a given hash function implies similar attack on the corresponding, used compression function. So in order to show the hash function's resistance to the collision, it suffices to show that the property is verified for the compression function. When iterating a compression function that provides collision resistance property, we can achieve a global collision resistance, and guarantee a perfect avalanche effect satisfaction by the constructed hash function.

2.2. Cellular Automata Preliminaries

Cellular automata are dynamical systems in which space and time are discrete. They consist of collections of cells organized in a grid, when each cell has a corresponding current state. The states of the cells evolve over time, depending on their current states and the states of the neighboring cells, according to a local and identical interaction rule in the case of uniform CAs, or different interaction rules in the case of non-uniform CAs [13].

CA were originally used by von Neumann [14] while he was studying self-reproducing systems and then popularized by Wolfram's substantial work in this area [15]. Wolfram observed that, based on simple rules, very complex behaviors can be obtained. He pioneered the investigation of CA as mathematical models for self-organizing statistical systems and suggested the use of elementary CAs, which are simple 1-dimensional linearly connected array of n cells, usually referred to as 3-neighborhood CAs. Each cell in the array takes a discrete state s equal to 0 or 1. If a configuration of the CA at a time step t is defined by the binary vector C^t, and the i^{th} cell's state is denoted by $(C^t)_i$, then the transition function f is used to determine the next state of each cell from a neighborhood's corresponding configuration. Cell's states are updated in parallel with respect to each other, using the transition function in each time step. The next state of a cell at time $t+1$ is only influenced by its own state and the states of its left and right neighbors at time t. The configuration C^{t+1} at time $t+1$ can be computed by:

$$(C^{t+1})_i = f((C^t)_{i-1}, (C^t)_i, (C^t)_{i+1}), \quad \forall i = 0 \ldots n-1 \tag{2}$$

where $(C^t)_{i-1}$, $(C^t)_i$ and $(C^t)_{i+1}$ are the states of the left neighbor, self and right neighbor of the i^{th} cell at time t. A cellular automaton with 2 states and a neighborhood's radius equal to 1 (3 cells) has $2^3 = 8$ possible neighborhood's configurations. If the set of all possible configurations is expressed using a truth table, the decimal equivalent of its sequence output is referred to as a "Rule" [15], and by the way the transition rules length is equal to 8 and a total of $2^8 = 256$ CA local rules can be used. Table 1 illustrates an example of two elementary rules defined by the corresponding truth tables.

If the same rule applies to all cells in a CA, the CA is named uniform or regular CA, whereas if different rules apply to different cells, it is

Neighborhood's configuration at time t	**Value of the central cell at time t+1**					
	(Rule 30)	**(Rule 2)**	**(Rule 154)**	**(Rule 207)**	**(Rule 255)**	**(Rule 133)**
111	0	0	1	1	1	1
110	0	0	0	1	1	0
101	0	0	0	0	1	0
100	1	0	1	0	1	0
011	1	0	1	1	1	0
010	1	1	1	1	1	0
000	0	0	0	1	1	1

Table 1. Truth table of some arbitrary selected elementary CA transition rules.

named a hybrid (or non-uniform) CA. A programmable CA (PCA) is a hybrid CA controlled by a number of signals such that different rules can be generated.

Cellular automata have several properties that favor their use as basis for the design of hash functions. Their chaotic, complex and unpredictable behavior of some transition rules enables their effective use to design safe and reliable hash functions. The simplicity of their implementations and their parallel nature makes them suitable as a basis for fast compression functions.

2.3. Cellular Automata for Hash Functions: Related Works

Cellular automata have been widely used recently to construct cryptographic primitives. They have been used for the construction of symmetric cryptosystems, public key cryptosystems, secret sharing schemas and hash functions. The first cryptographic application of CA initiated by Wolfram [16] describes a stream-based cipher using the elementary CA rule 30. The CA was used as a pseudorandom numbers generator to produce statistically good sequences used to cipher plain data by the Vernam ciphering model. CAs were also used for block-ciphers constructions using reversible and irreversible rules [17], and also to build public-key cryptosystems by exploiting two-dimensional CAs reversibility problem [18].

In [19], Damgard was the first to propose a hash function based on CAs, he used Wolfram's pseudorandom bit generation's scheme to design a compression function, but his proposal was cryptanalysed by Daemen et al. in [20] who, in turn, proposed a framework of collision free hash functions based on CA, named CellHash. The same authors proposed later an improved version named SubHash in [21]. Both CellHash and SubHash are hardware-oriented so making extremely high speed possible, but unfortunately, the two schemas were cryptanalysed later in [22].

Another CA-based hash function has been proposed by Mihaljevic, et al. in [23], where they describe a family of fast dedicated one-way hash functions using linear CA over GF(q). The proposal is an extension of the bit's oriented hashing proposed earlier in [24], enhanced by employing the approved model of iterative hash function with compression and output's functions: the compression function is one of the Davies-Meyer types employing CA, when the output function was a key generator based on CA over GF(q). In [25] Dasgupta et al. proposed a CA based scheme for message authentication by investigating a particular class of non-group CA that can be employed to generate an efficient message authentication function. Two-dimensional CA was also proposed as the base for construction of hash functions in [26].

More recently, Jun-Cheol proposed a one-way hash function using linear and nonlinear CAs [7]. Norziana et al. proposed an alternative hash function based on CA rules 30, 134 and Omega-Flip Network in [10], then proposed another hash function in [13] named STITCH-256 using balanced CA elementary rules like rules: 29, 39 and 27, and diffusion functions to implement the compression function.

3. The Proposed Hashing Scheme

In the following, we present the proposed new CA-based cryptographic hash function. The function follows iterative hash model inspired by the Damgard's one presented in [19], and uses two internal functions namely: a compression function C and a transformation function T. The former employs programmable CA with 4 rules (30, 90, 105 and 150), while the latter T uses a hybrid cellular automaton with transition rules 30 and 105. Both message blocks and hash value are 256 bit binary strings.

3.1. The Mode of Operation

In the proposed hashing scheme, we use a variant of the Merkle-Damgard construction. The proposed algorithm requires a padding function (in such a way that the length of the message becomes a multiple of 256 bit) for which the last 64 bits encode the length of the message M to be hashed. The padded message is then divided into blocks M_i $(i = 1 \ldots k)$ with $|M_i| = 256$ bits. The algorithm requires also a fixed initialization vector: IV$\in \{0,1\}^{256}$

The compression function C is then iterated for k times. C takes as inputs a block message M_i

and a chaining variable h_{i-1}, to produce a 256 bit string as output. This output string is XORed with the output of a transformation function T applied to the block message M_i to form the next chaining variable h_i. Figure 1 illustrates a pictorial representation of the proposed hashing scheme.

3.2. The Compression and Transformation Functions

To construct the compression function, we use an elementary programmable CA (PCA) with 4 rules, namely the rules: 30, 90, 105, 150 (As many works have studied the properties of different CA rules, we use only the rules that have been proven to have good pseudorandom properties).

We firstly define a function F that takes as input a block message M_i and a chaining variable h_{i-1}, to produce an output string f on 256 bits as follows:

$$F : \{0,1\}^{256} \times \{0,1\}^{256} \rightarrow \{0,1\}^{256}$$
$$(h_{i-1}, M_i) \rightarrow F(h_{i-1}, M_i) = f_i \qquad (3)$$

The function F is defined like the following: a PCA of 256 bit length is initially loaded with the message block M_i. Additional 256 bits from chaining variable are doubled to form a 512 bit string S (i.e. $S = h_{i-1}|h_{i-1}$). The bits from S are used with the number of the current iteration to control the rule configuration of the individual CA cells. Each 2 bits of S control one cell rule (the bits j and $j+1$ control the rule of the cell number $j/2$) conjointly with the remainder of the iteration number modulo 4. The control logic of the proposed PCA is described in Table 2.

The CA is then iterated for n times. The value of the parameter n defining the number of iterations may be fixed according to the desired ratio of speed/reliability performances. The output of F is defined by the final PCA state.

The compression function C is then defined using the function F as follows:

$$C : \{0,1\}^{256} \times \{0,1\}^{256} \rightarrow \{0,1\}^{256}$$
$$(f_i, h_{i-1}) \rightarrow C(f_i, h_{i-1}) = c_i = f_i \oplus h_{i-1} \qquad (4)$$

The transformation T takes as input a block message of 265 bit length and produces an output

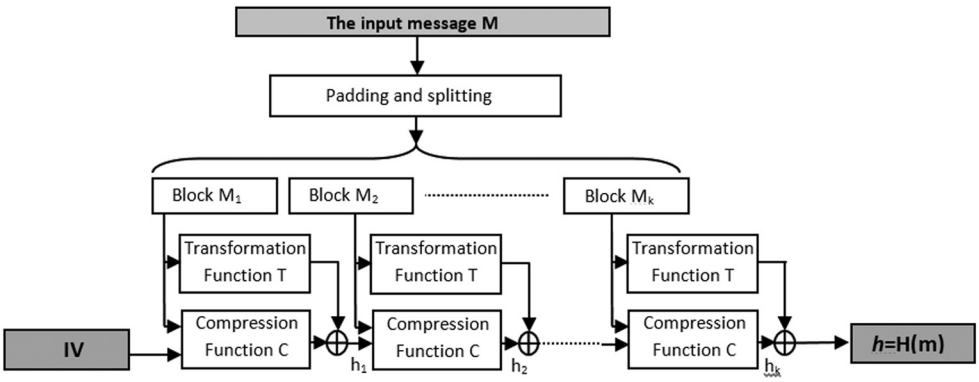

Figure 1. Pictorial representation of the proposed schema.

2 bits from M_i	i mod 4			
	0	1	2	3
00	Rule 30	Rule 90	Rule 105	Rule 150
01	Rule 90	Rule 30	Rule 150	Rule 105
10	Rule 105	Rule 150	Rule 30	Rule 90
11	Rule 150	Rule 105	Rule 90	Rule 30

Table 2. Definition of the control logic for the proposed PCA.

string t of 256 bit that is XORed with the compression function output.

The transformation function is defined as follows: A hybrid CA with rules 30 and 105 (rules are applied in alternation i.e. 30, 105, 30....) is initially loaded with the message block M_i. The CA is iterated 50 times to obtain an intermediate configuration, then iterated another 256 times to form a 256 by 256 bit square. The diagonal of this square is then taken as output of T. Figure 2 illustrates pictorial specification of the transformation T operations mechanism.

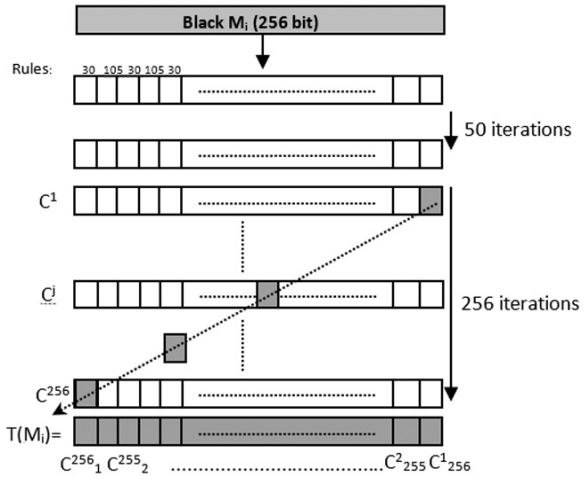

Figure 2. Pictorial description of the transformation function T.

The transformation function T is used to obtain the next chaining variable h_i by:

$$\begin{aligned} h_i &= c_i \oplus T(M_i) \\ &= f_i \oplus h_{i-1} \oplus T(M_i) \\ &= F\left(h_{i-1}, M_i\right) \oplus h_{i-1} \oplus T(M_i) \end{aligned} \quad (5)$$

Using the described scheme, the proposed hash function can take any message M of arbitrary length as input, decompose it into consecutive blocks M_i, and produce the corresponding hash value $H(M)$. The proposed function is benchmarked with respect to several performance tests with several experiments illustrated in the following section.

4. Performances Evaluations and Obtained Results

As explained above, it is not always possible to formally prove the resistance of a given hash function to common attacks, but since a Merkle-Damgard construction is used as a domain extension mechanism, the proof can be reduced to the resistance of the proposed compression function. In addition, several statistical properties can be checked in order to show that the function provides good cryptographic level of security. Pseudorandom behavior and avalanche effect are generally considered as good security's indicators of a hash function. In this section, we perform several statistical experiments on the proposed hashing scheme. We also show that best computational performances can be achieved by the proposed scheme with respect to existing models.

4.1. Security Analysis

The best assistance available about security of a particular hash function is the complexity of the most efficient applicable known attack, which gives an upper bound on its security. An attack of complexity 2^n requires approximately 2^n operations, each being an appropriate unit of work. According to [35], it is possible to relate the security of $H()$ to the security of h and g according to the following theorem:

Theorem 1 [35]: Let H be an iterated hash function with Merkle-Damgard construction, then preimage and collision attacks on H (where an attacker can choose IV freely) have roughly the same complexity as the corresponding attacks on h and g.

The function g is an optional transformation used in a final step to map the n-bit chaining variable to an n'-bit result $g(Hm)$; g is often the identity mapping $g(H_m) = H_m$. In the present work the function g is trivially considered as the identity function, while the compression function is handled by the function F composed by the two transformations C and T. Accordingly, the security of the proposed hash function is relied on the resistance of the compression function h. The function h is applied on each sub-block M_i and transforms it using the two functions C and T according to equation (4). The transformation steps consist of the following:

– A non-linear mapping of M_i using the transformation T with a non-uniform PCA;

– A non-linear compression of M_i using the function C with a PCA evolution mechanism that maps the input vector into another vector of the n-dimensional binary space.

– A bit-by-bit addition between the outputs of C and T to compute the final value of F.

As a result, the following facts imply the security of the proposed compression function:

1. Cellular automata have chaotic characteristic that maps any nonzero state into a nonzero state which belongs to the sequence of all possible different 2^n-1 nonzero n-dimensional vectors in such manner that the expected Hamming distance between the current state and the next one is $n/2$.

2. A high non linearity is performed on each sub-block Mi due to the dynamic characteristic of the nonlinear PCA with rules 30, 90, 105 and 150.

3. The computational infeasibility of reconstructing pervious configurations of a given nonlinear PCA.

The facts 1-3 imply that the proposed compression function is a cryptographically secure one-way function. Hence, according to Theorem 1, and since the iterating mechanism is performed according to a Merkle-Damgard construction, the proposed hash function is cryptographically secure.

4.2. The Avalanche and Strict Avalanche Criteria

Avalanche effect is a desirable property for cryptographic hash functions that tries to reflect the idea of high-nonlinearity [27]: a little change in the input (flipping one single bit) produces a significant change of the output (approximately half of the bits are flipped). Formally, if a function F has the avalanche effect, then the Hamming distance between its output on a random input binary string x and the output obtained when randomly changing one bit of x should be, on average, half of the output size [28]. This effect tries to abstract the intuitive idea of high nonlinearity: very small difference in the input must produce high changes in the output, hence an avalanche of changes.

Mathematically, $F : \{0,1\}^m \rightarrow \{0,1\}^n$ has the avalanche effect if it holds the following:

$$\forall x, y \in \{0,1\}^m : Hamming(x,y) = 1 \implies average(Hamming(F(x),F(y))) = \frac{n}{2} \quad (6)$$

When Hamming (x,y) denotes the Hamming distance between the two n-bits blocks x and y. Figure 3 shows the results of the avalanche effect test performed on the proposed hash function, using a set of 10 000 pairs of arbitrary messages M_i and M_i' such that Hamming $(M_i, M_i') = 1$. Obtained results show that the hamming distances between the hash values (i.e. Hamming $(H(M_i), H(M_i'))$) are concentrated around the value 128, which indicates that the hash function has a good avalanche effect.

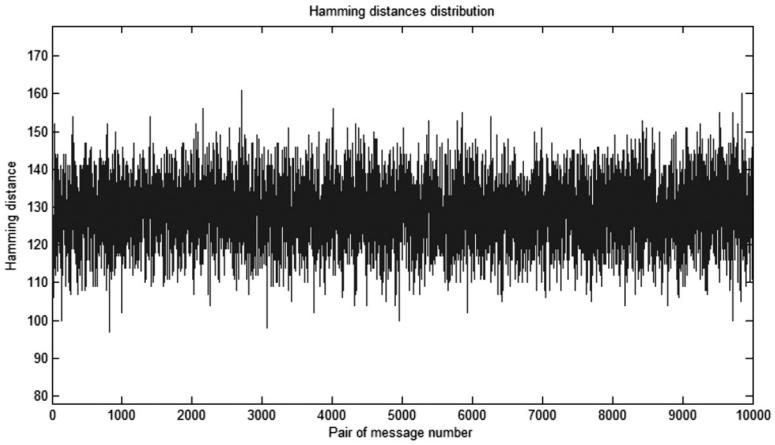

Figure 3. Obtained distribution of Hamming distances between hashes of random messages and hashes of their one-bit-flipped alterations.

Another, more accurate and demanding nonlinearity measurement is the so called strict avalanche criterion [29] which, in particular, implies the avalanche effect. The strict avalanche criterion (SAC) is a more demanding property that was originally presented by as a generalization of the avalanche effect to measure the quantity of confusion and diffusion in substitution boxes (s-boxes). Formally, a function F is said to satisfy the SAC if, whenever a single input bit is flipped, each of the output bits must change with a probability of one half. This implies that the distribution of hamming distances between outputs that have similar inputs (differing in one bit) should follow a binomial distribution. Mathematically, the SAC is described by [28]:

$$\forall x, y \in \{0,1\}^m : Hamming(x,y) = 1 \Rightarrow Hamming(F(x),F(y)) \approx B\left(\frac{1}{2}, n\right) \quad (7)$$

where $B(1/2,n)$ denotes binomial distribution with parameters $1/2$ and n. This definition also tries to abstract the more general concept of independence of the output from the input. An ideal hash function F will resemble a perfect random function where inputs and outputs are statistically unrelated [30].

In order to evaluate the proposed hash function with respect to the SAC, the following experiment has been conducted: the 256 elements of an integer's vector V (each one corresponding to a bit position of the hash function's output) are initialized firstly to 0. A set of 1000 random messages with arbitrary length are then generated, and their corresponding hash values $H(M)$ are computed. For each one of these messages, only one bit is randomly flipped getting a new message M', that is also hashed to obtain a new hash value $H(M')$. The Hamming distance Hamming $(H(M),H(M'))$ is calculated, and the result is used to determine the element of V to be incremented (i.e. V[Hamming $(H(M),H(M'))$] $++$). This operation is repeated 1000 times for each message.

Finally, the values of the vector V elements are divided by the total number of performed experiments (equal to $1000 * 1000 = 10^6$) in order to perform a normalization of the computed distribution. The obtained values represent the distribution of Hamming distances that is compared to the binomial distribution as plotted in Figure 4.

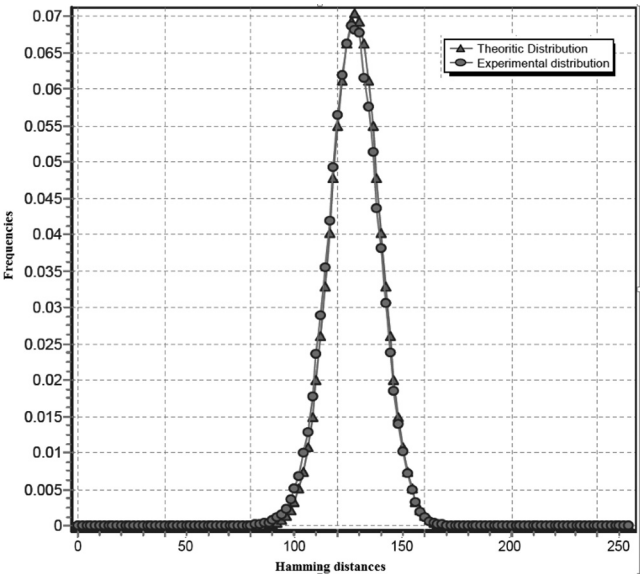

Figure 4. Experimental Hamming distances distribution vs. theoretical binomial distribution.

The distance between the observed distribution of the Hamming distances and their theoretical distribution under perfect Strict Avalanche Criterion hypothesis $B(1/2,n)$, has also been measured by means of a chi-square χ^2 goodness-of-fit test [30]. The chi-square measurement is a statistical test that allows verifying the adequacy of a data set to a probability distribution by using the following formula [27]:

$$\chi^2 = \sum_{i=0}^{256} \frac{(Observed\_frequency_i - Expected\_frequency)^2}{Expected\_frequency} \quad (8)$$

Using the distribution values obtained in the experiment described above, the value of the χ^2 measurement has been computed and found to be equal to 0,005465. Using the probability $\alpha = 0.01$ as our critical threshold, the hypothesis of equivalence between the two distributions is accepted if the χ^2 value is less than the quantile $\chi_{255,0.01} = 310.45$. It is clear that the obtained χ^2 is negligible with respect to the quantile value, and consequently, the null hypothesis is accepted and the Hamming distribution of the proposed hash function is then following a binomial distribution $B\left(\frac{1}{2}, 256\right)$. Results of the SAC test show that the hash function provides good avalanche effect criterion, which is one of the most important features of secure cryptographic hash functions.

4.3. Randomness Statistical Tests

Cryptographic primitives and especially hash functions should act like pseudorandom functions to avoid statistical attacks; therefore the output of a secure hash function must be statistically indistinguishable from the output of a random function. We performed several statistical randomness tests on the output of the proposed hash function in order to show that it provides best randomness properties.

In the performed experiment, the hash function has been used as a pseudonumber generator to create a data stream of 10Mb. The stream is generated using a counter mode scheme applied using the hash function on an initial random integer seed S and then calculating the values $H(S), H(S+1), \ldots, H(S+327680)$ each one on 256bit. The resulting outputs are finally concatenated to form a data stream.

The produced stream is analyzed statistically using both Diehard [31] and ENT [32] statistical tests batteries, and then obtained results are averaged and reported in Tables 3 and 4.

It is clear from presented results that the binary stream generated by the hash function has successfully passed all DIEHARD and ENT tests. We can conclude that the function has a good

Test Name	P-value	Interpretation
Birthday Spacing	0.451552	Pass
Overlapping 5-permutation	0.654729	Pass
Rank test for 31x31 binary matrices	0.64841	Pass
Rank test for 32x32 binary matrices	0.894523	Pass
Rank test for 6x8 binary matrices	0.562742	Pass
BITSTREAM TEST	0.329124	Pass
OPSO Test	0.42485	Pass
OQSO Test	0.642714	Pass
DNA Test	0.42839	Pass
Count the 1s in a Stream of Bytes	0.69275	Pass
Count the 1s in Specific Bytes	0.39258	Pass
Parking Lot Test	0.512293	Pass
Minimum Distance Test	0.64942	Pass
Random Spheres Test	0.35862	Pass
The Squeeze Test	0.43175	Pass
Overlapping Sums Test	0.58441	Pass
Runs Up and Down Test	0.74349	Pass
The Craps Test	0. 83457	Pass

Table 3. Results of the DIEHARD tests battery applied on designed hash function's output.

Test Name	Value	Norm
Entropy	7.997982	8.0
Optimum compression	0.000003	0.0
Arithmetic mean value of data bytes	127.5586	127.5 = random
Monte Carlo value for π	3.140865524	π
Serial correlation coefficient	0.000347	totally uncorrelated = 0.0

Table 4. Results of the ENT tests battery applied on the designed hash function's output.

pseudorandom behavior and can as a result be considered statistically indistinguishable from random function, which is a principal characteristic of a secure hashing scheme.

4.4. Comparison With Existing CA-based Hash Functions

In order to illustrate the advantages of the proposed hash function with respect to existing CA-based ones, a comparative study has been performed with respect to several criteria including sensitivity to elementary alterations, strict avalanche criterion, and randomness. The proposed function has been compared to CellHash [20], SubHash [21], STITCH-256 [13] and the function proposed initially by Damgard in [19]. Performances and speed comparison are provided in the next section.

Table 5 illustrates several comparison result using the chi-square measurement to evaluate the strict avalanche criterion, the averaged p-value of the DIEHARD test, the averaged entropy from the ENT test and the average runtime performances. The functions were implemented in software while performance's experiments were performed using an Intel Core i5 (2.5 GHz) microprocessor platform. It is clear from the obtained results that the proposed function achieves very competitive runtime performances with equivalent randomness and sensitivity results. Figure 5 illustrates a comparison of the experimental Hamming distributions between the proposed function and the mentioned CA-based ones.

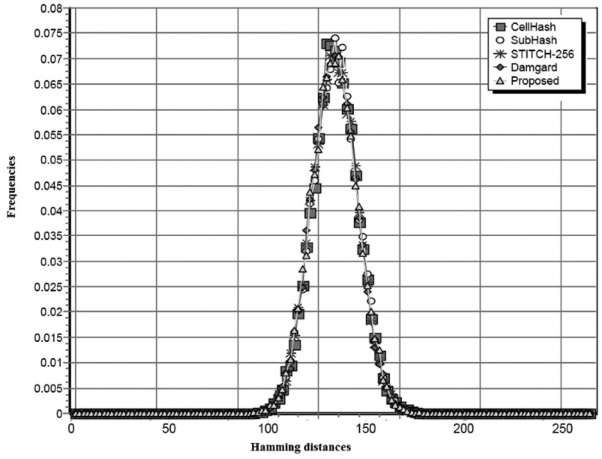

Figure 5. Experimental Hamming's distribution compared with respect to some known CA-based hash functions.

4.5. Performance Analysis of the Designed Function

Cellular automata uses simple binary operations. Therefore CA-based primitives can achieve very high speed in both hardware and software implementations. In the present work, the proposed hash function has been implemented using Microsoft Visual C++, while performance's experiments were performed using an Intel Core i5 (2.5 GHz) microprocessor platform. Table 6 shows a comparison between the hash function and the widely used ones implemented by the Crypto++ library using Microsoft Visual C++ [33]. It is clear that the designed function achieves very competitive speeds with respect to other standards, and we assume that it can achieve much better rates if hardware implementation is used.

Hash function	**Averaged DIEHARD p-value**	**Averaged ENT Entropy**	**Averaged Chi-square Value**	**Speed (Mb/s)**
CellHash [20]	0.54368722	7.989825	0,07782	90
SubHash [21]	0.55567534	7.991735	0,006581	105
STITCH-256 [13]	0.53996212	7.998321	0,005387	113
Damgard [19]	0.48216211	7.998212	0,13556	125
Proposed function	0.55307924	7.997982	0,005465	138

Table 5. Security performances compared with respect to some known CA-based hash functions.

Hash function	Speed (MB/s)
MD5	406
SHA-1	158
SHA-256	143
SHA-512	82
RIPEMD-128	240
RIPEMD-256	195
RIPEMD-320	104
Whirlpool	80
Proposed function (256 bit)	**138**

Table 6. Speed performances compared with respect to some known hash functions.

5. Conclusions

In this paper, we propose a cryptographic hash function based on cellular automata. The proposed function uses two internals, namely: a compression function based on programmable cellular automata controlled by the chaining variable bits, and a transformation function construct from a hybrid cellular automaton with rules 30 and 105. The proposed scheme has been experimentally analyzed with respect to several statistical tests, and the obtained results show that the proposed function provides good cryptographic properties such as pseudo-random behavior and sensitivity to the input changes. In addition, the function is simple, fast, and can be easily implemented through software or hardware. Performance evaluations show that extremely optimal performances are achieved by the function with respect to existing standards, and we presume that better performances can be obtained if hardware implementation is adopted due to the inherent parallelism of cellular automata.

References

[1] M. NAOR, M. YUNG, Universal one-way hash functions and their cryptographic applications. In *Proceedings of the twenty-first Annual ACM Symposium on the Theory of Computing*, (1989, February) pp. 33–43. ACM.

[2] B. PRENEEL, R. GOVAERTS, J. VANDEWALLE, Hash Functions Based on Block Ciphers: A Synthetic Approach. In *Crypto '93*, **773** of LNCS, (1993) pp. 368–378. Springer-Verlag.

[3] J. BLACK, P. ROGAWAY, T. SHRIMPTON, Black-Box Analysis of the Block-Cipher-Based Hash-Function Constructions from PGV. In *Crypto'02*, **2442** of LNCS, (2002) pp. 320–335. Springer-Verlag.

[4] J. BUCHMANN, S. PAULUS, A One Way Function Based on Ideal Arithmetic in Number Fields. In *Crypto '97*, **1294** of LNCS, (1997) pp. 385–394. Springer-Verlag.

[5] S. CONTINI, A. LENSTRA, R. STEINFELD, VSH, an Effcient and Provable Collision – Resistant Hash Function. In *Eurocrypt '06*, **4004** of LNCS, (2006) pp. 165–182. Springer-Verlag.

[6] D. CHARLES, K. LAUTER, E. GOREN, Cryptographic Hash Functions from Expander Graphs. *Journal of Cryptology*, **22**(1), (2007) 93–113.

[7] J-C. JEON, Analysis of Hash Functions and Cellular Automata Based Schemes. *International Journal of Security and Its Applications*, **7**(3), (May, 2013).

[8] M. MAQABLEH, A. SAMSUDIN, M. ALIA, New Hash Function Based on Chaos Theory (CHA-1). *International Journal of Computer Science and Network Security*, **8**(2), (2008) 20–27.

[9] X. YI, Hash Function Based on Chaotic Tent Maps. *IEEE Transactions on Express Briefs*, **52**(6), (2005) 354–357.

[10] J. NORZIANA, M. RAMLAN, R. Z. MUHAMMAD, I. U. NUR, A. Z. ZURIATI, A New Cryptographic Hash Function Based on Cellular Automata Rules 30, 134 and Omega-Flip Network. *International Proceedings of Computer Science & Information Tech.*, **27**, (2012) 163.

[11] A. MENEZES, P. OORSCHOT, S. VANSTONE, *Handbook of Applied Cryptography, chapter Hash Functions and Data Integrity*. CRC Press, 1996, pp. 321–384.

[12] NIST., Federal Information Processing Standard (FIPS) Publication 180-2, Secure Hash Standard (SHS), U.S. Doc/NIST, 2002. Available from `http://csrc.nist.gov/publications/fips/fips180-2/fips180-2.pdf`

[13] N. JAMIL, R. MAHMOOD, M. R. Z'ABA, N. I. UDZIR, Z. A. ZUKARNAEN, A New Cryptographic Hash Function Based on Cellular Automata Rules 30. *134 and Omega-Flip Network International Conference on Information and Computer Networks (ICICN 2012) IPCSIT*, **27** (2012). IACSIT Press, Singapore.

[14] J. V. NEUMANN, *The World of Physics: A Small Library of the Literature of Physics from Antiquity to the Present*, chapter *The General and Logical Theory of Automata*. Simon and Schuster, New York, 1987, 606–607.

[15] S. WOLFRAM, *A New Kind of Science*. Wolfram Media, 2002.

[16] S. WOLFRAM, Cryptography with Cellular Automata in Advances in Cryptology. *Crypto'85 Proceedings. LNCS 218*, (1985) pp. 429–432. Springer.

[17] Z. CHAI, Z. CAO, Y. ZHOU, Encryption Based on Reversible Second-Order Cellular Automata. *ISPA Workshops, LNCS 3759*, (2005) pp. 350–358.

[18] A. CLARRIDGE, K. SALOMAA, A Cryptosystem Based on the Composition of Reversible Cellular Automata. *LATA 2009, LNCS 5457*, (2009) pp. 314–325.

[19] I. DAMGARD, A Design Principle for Hash Functions. In *Crypto '89, volume 435 of LNCS*, (1989) pp. 416–427. Springer-Verlag.

[20] J. DAEMEN, R. GOVAERTS, J. VANDEWALLE, A Framework for the Design of One-Way Hash Functions Including Cryptanalysis of Damgard's One-Way Function Based on a Cellular Automaton. In *Asiacrypt '91, volume 739 of LNCS*, (1991) pp. 82–96. Springer-Verlag.

[21] J. DAEMEN, R. GOVAERTS, J. VANDEWALLE, A hardware design model for cryptographic algorithms, Computer Security – ESORICS 92. *Proc. Second European Symposium on Research in Computer Security, LNCS 648*, (1992) pp. 419–434. Springer-Verlag.

[22] D. CHANG, Preimage Attacks on CellHash, SubHash and Strengthened Versions of CellHash and SubHash. *Cryptology ePrint Archive, Report 2006/412.*, (2006). (eprint.iacr.org/2006/412)

[23] M. MIHALJEVIC, Y. ZHENG, H. IMAI, A Fast Cryptographic Hash Function Based on Linear Cellular Automata over GF(q). *Special Section on Cryptography and Information Security. IEICE TRANS. FUNDAMENTALS*, **E82**(1), January 1999.

[24] M. MIHALJEVIC, Y. ZHENG, H. IMAI, A Cellular Automaton Based Fast One-Way Hash Function Suitable for Hardware Implementation. *Proceedings of PKC'98, LNCS 1431*, (1998) pp. 217–233.

[25] P. DASGUPTA, S. CHATTOPADHYAY, I. SENGUPTA, Theory and Application of Non-group Cellular Automata for Message Authentication. *Journal of Systems Architecture*, **47**(55), (2001) pp. 383–404.

[26] S. HIROSE, S. YOSHIDA, A one-way hash function based on a two-dimensional cellular automaton. *The 20th Symposium on Information Theory and Its Applications (SITA97). Proc.* **1**, (Dec. 1997) pp. 213–216. Matsuyama, Japan.

[27] A. F. WEBSTER, S. E. TAVARES On the design of s-boxes, Lecture notes in computer sciences. *218 on Advances in cryptology – CRYPTO 85*, (1985) pp. 523–534. Springer-Verlag, New York, USA.

[28] J. C. H. CASTROA, J. M. SIERRAB, A. SEZNECA, A. IZQUIERDOA, A. RIBAGORDAA, The strict avalanche criterion randomness test. *Mathematics and Computers in Simulation*, **68**, (2005) pp. 1–7.

[29] R. FORRE, The strict avalanche criterion: spectral properties of booleans functions and an extended definition. Advances in cryptology, in: *Crypto'88, Lecture Notes in Computer Science*, **403**, (S. GOLDWASSER, Ed.), (1990) pp. 450–468. Springer-Verlag.

[30] A. DOGANAKSOY, B. EGE, O. KOÇAK, F. SULAK, Cryptographic Randomness Testing of Block Ciphers and Hash Functions. *IACR Cryptology ePrint Archive 2010*, **564** (2010).

[31] G. MARSAGLIA, *Diehard Battery of Tests of Randomness*, 1995. `http://www.stat.fsu.edu/pub/diehard/`

[32] J. WALKER, *ENT A Pseudorandom Number Sequence Test Program*, 2008. `http://www.fourmilab.ch/random/`

[33] W. DAI, *Crypto++*, 2013. `http://www.cryptopp.com/`

[34] R. MERKLE, One way hash functions and DES. *Advances in cryptology – CRYPTO 89, Lecture Notes in Computer Science*, **435**, (1990) pp. 428–446.

[35] L. KNUDSEN, B. PRENEEL, Fast and secure hashing based on codes. *Advances in cryptology – CRYPTO 97, Lecture Notes in Computer Science*, **1294**, (1997) pp. 485–498.

Received: February, 2015
Revised: July, 2015
Accepted: August, 2015

Contact addresses:
Alaa Eddine Belfedhal
Computer Sciences Department
Djillali Liabes University of Sidi Bel Abbes
Algeria
e-mail: `belfedhal.alaa@gmail.com`

Kamel Mohamed Faraoun
Computer Sciences Department
Djillali Liabes University of Sidi Bel Abbes
Algeria
e-mail: `kamel_mh@yahoo.fr`

ALAA EDDINE BELFEDHAL is currently a PhD student at the Computer Science Department of Djillali Liabes University, Algeria, and an assistant teacher at Mascara University, Algeria. He received his Engineering and Master's degrees in computer science from Djillali Liabes University in 2008 and 2011 respectively. His research interests include cryptographic primitives and information security.

KAMEL MOHAMED FARAOUN received his Master's Degree in computer science from the Djilali Liabes University of Sidi-Bel-Abbes, Algeria in 2002, his Ph.D Degree in computer science, in 2006, and his HDR Degree in computer science and intelligent systems, in 2009. His current research areas include computer security systems, cryptography, genetic algorithms, cellular automata, and information theory. Dr. Faraoun is currently a teacher at the Computer Sciences Department of Djilali Liabes University, he teaches information theory and cryptography.