Improved umac algorithm with crypto-code mceliece's scheme

Serhii Yevseiev
Doctor of Technical Science, Senior Research
Department of Cyber Security and Information Technology
Simon Kuznets Kharkiv National University of Economics
ORCID: http://orcid.org/0000-0003-1647-6444

Alla Havrylova
Senior Lecturer
Department of Cyber Security and Information Technology
Simon Kuznets Kharkiv National University of Economics
ORCID: https://orcid.org/0000-0002-2015-8927

Abstract. *The article discusses the mathematical model of the mini-version of the UMAC hashing algorithm, using various cryptoalgorithms as a pseudo-random substrate when generating key data, presents the results of studies of the considered mini-versions. A new approach is proposed to ensure the cryptographic stability of this hashing algorithm based on Mac-Elis crypto-code constructions on modified elliptic codes. This approach will provide the required level of cryptographic security in post-quantum cryptography.*

Keywords: *Mac-Elis crypto-code, UMAC hashing algorithm, El-Gamal cryptosystems*

Introduction

The development of decentralized systems is closely related to the use of cryptographic data processing tools. As a rule, such systems use asymmetric algorithms on elliptic curves (El-Gamal cryptosystems, digital signature ECDSA), the SHA-2 algorithm (version SHA-256) is used as the hash functions. The main advantages of such systems in the digital economy are: anonymity of transfers, the minimum cost of making a profit, the minimum commission from transactions, security and reliability. However, in 2018, it was decentralized circulation and cryptocurrency storage systems that took the first place in the number of cyber-attacks. In addition, the increase in demand for this type of banking activity causes certain difficulties in the operation (computational capabilities) of such systems (increase in electronic clients of systems, improvement of verification protocols, appearance of smart contracts). All this requires a revision of the cryptoalgorithms used in decentralized systems, and first of all, hashing algorithms.

Material and method

To ensure the required levels of speed and security in the protocols of decentralized systems it is proposed to use the winning algorithm of the NESSIE

competition, based on universal hash algorithms. The features of such hashing algorithms are their theoretical predictability of the collision probability and the uniform distribution of the entire power of the algorithm profiles on the collision set [1]. Consider the main characteristics of this algorithm on the example of its mini-version, which allows conducting research on the entire set of hash codes. This approach is used by the scientists of the school Gorbenko I. D. to assess the strength of block-symmetric ciphers

Reduced model UMAC (mini-UMAC). UMAC message authentication code generation scheme uses several transformation layers in its structure, including a block symmetric cipher (AES cipher recommended for use) [1].

It was shown above that the scheme of forming UMAC codes consists of the following layers:

- three-level universal hashing for generating hash-codes $Y = Hash(K, M, Taglen)$;
- cryptographic transformations using a block symmetric cipher to form a pseudo-random substrate $Pad = PDF(K, Nonce, Taglen)$;
- $Tag = UMAC(K, M, Nonce, Taglen) = Y \oplus Pad$.

Consider each layer of the UMAC message authentication code generation scheme for scaling.

We construct a mini-version of the three-level universal hashing without changing the structure of algebraic transformations by simply reducing the dimension of the blocks and the processed data eight times.

The corresponding length of the hash code Y_{mini} of the reduced model of the first layer will be a multiple of 4 bits; we will form its value by combining (concatenating) four sequences Y_{miniL3_i}

$$Y_{mini} = Y_{miniL3_1} \| Y_{miniL3_2} \| Y_{miniL3_3} \| Y_{miniL3_4},$$

where Y_{miniL3_i} – is the result of multi-level hashing of the message of the reduced model of the first layer mini-UMAC.

Consider the process of forming a single block Y_{miniL3_i} (we will not perform the second level of hashing in the reduced model):

$$Y_{miniL3_i} = Y_{miniL3} = Hash_{miniL3}\left(K_{miniL3_1}, K_{miniL3_2}, Hash_{miniL1}\left(K_{miniL1}, M_{mini}\right)\right),$$

where K_{miniL1}, K_{miniL3_1}, K_{miniL3_2} – are the mini-UMAC key sequences;

$Hash_{miniL1}$ и $Hash_{miniL3}$ – are reduced versions of the first and third levels hashing, respectively.

At the first level, a 32-bit string array M_{mini} is transformed by a function $NH(K_{L1}, M_i)$. This string is the result of first-level hashing $Y_{miniL1} = NH_{mini}(K_{miniL1}, M_{mini})$.

The value of the function $NH_{mini}(K_{miniL1}, M_{mini})$ is calculated according to the following rule. The information block M_{mini} is divided into eight four-bit sub-blocks.

$$M_{mini} = M_{mini_1} \| M_{mini_2} \| \ldots \| M_{mini_8}.$$

Similarly, a key sequence K_{L1} is represented as a sequence of eight four-bit sub-blocks:

$$K_{mini\,L1} = K_{mini\,L1_0} \| K_{mini\,L1_1} \| ... \| K_{mini\,L1_8}.$$

After that (taking the initial state $Hash_{L1} = 0$) the following operations are performed:

$$Hash_{mini\,L1} = Hash_{mini\,L1} +_8 ((M_{mini_0} +_4 K_{mini\,L1_0}) \times_8 (M_{mini_4} +_4 K_{mini\,L1_4})),$$

$$Hash_{mini\,L1} = Hash_{mini\,L1} +_8 ((M_{mini_1} +_4 K_{mini\,L1_1}) \times_8 (M_{mini_5} +_4 K_{mini\,L1_5})),$$

$$Hash_{mini\,L1} = Hash_{mini\,L1} +_8 ((M_{mini_2} +_4 K_{mini\,L1_2}) \times_8 (M_{mini_6} +_4 K_{mini\,L1_6})),$$

$$Hash_{mini\,L1} = Hash_{mini\,L1} +_8 ((M_{mini_3} +_4 K_{mini\,L1_3}) \times_8 (M_{mini_7} +_4 K_{mini\,L1_7})),$$

where $+_8$, $+_4$ – addition operations modulo 28 and 24, respectively;

$\times_8$ – multiplication operation modulo 28.

As a result of calculations, an eight-bit value is formed $Y_{mini\,L1} = Hash_{mini\,L1}$.

The third level of hashing converts the eight-bit data $Y_{mini\,L1}$ sent to its input into a hash code $Y_{mini\,L3}$ of length 4 bits. The key sequences $K_{mini\,L3_1}$ and $K_{mini\,L3_2}$ are the lengths of 16 and 4 bits, respectively.

Hashable data $Hash_{mini\,L1}$ and key sequence $K_{mini\,L3_1}$ are evenly divided into four blocks, each of which is represented as an integer $Y_{mini\,L2_i}$ and $K_{mini\,L3_{1i}}$, $i = 1,2,...,4$.

The hash value $Y_{mini\,L3}$ is calculated as follows:

$$Y_{mini\,L3} = \left(\left(\left(\sum_{i=1}^{4} Y_{mini\,L2_i} K_{mini\,L3_{1i}} \right) \bmod(17) \right) \bmod(2^4) \right) xor(K_{mini\,L3_2}),$$

where $(x)xor(y)$ – is the "exclusive OR" operation on the values x and y.

The mini-version of the final transform for generating the mini-UMAC message authentication codes consists of modulo 2 values Y_{mini} and Pad_{mini}: $Tag_{mini} = Y_{mini} \oplus Pad_{mini}$.

Thus, the scaling of the applied transformations on the respective layers of the formation of the message authentication codes makes it possible to construct a reduced model of UMAC, experimentally investigate the collision properties of the generated images (codes). The scaling factor in the development of the UMAC mini-model is chosen so that the length of the generated hash-codes Y, pseudo-random substrates Pad and message authentication codes $Tag = Y \oplus Pad$ is equal to the length of the mini-block block of the AES symmetric block cipher [2], i.e. 16 bits. The choice of such a scaling factor allows, on the one hand, preserving the algebraic structure of the basic transformations of the UMAC algorithm, including the AES algorithm included in its scheme, on the other hand, it makes it possible to conduct experimental studies using the methods of statistical testing of hypotheses and mathematical statistics, considering a limited set of elements Y, Pad and $Tag = Y \oplus Pad$ and the corresponding results for estimating the number of collisions as a sample from the general population.

Let us consider the method of statistical estimation of the collision properties of the formed elements (we denote them for simplicity $h(x)$), consider the basic conditions and limitations when conducting experiments.

Methods of statistical studies of collision properties

Experimental studies of the collision properties of UMAC message authentication codes will be carried out along the appropriate transformation layers:

1. At the first stage, collisional properties of the mini-version of universal hashing are investigated. To do this, it is necessary to confirm in the course of the experiment theoretical estimates of the number of collisions generated by the hash-codes Y_{mini};

2. At the second stage, collisional properties of pseudorandom substrates are investigated based Pad_{mini} on the analysis of the properties of the reduced Baby-Rijndael cipher model;

3. At the third stage, collisional properties of message authentication codes generated using mini-UMAC $Tag_{mini} = Y_{mini} \oplus Pad_{mini}$ are investigated. This is the most important part of the research, since it will allow us to answer the question about preserving the properties of universal hashing after applying the layer of cryptographic information transformation.

Estimation of the number of collisions of the generated elements is carried out taking into account the collision properties of universal hashing, which allows to confirm the hypothesis about the preservation of collision properties of universal hashing at all stages of the formation of message authentication codes mini-UMAC.

The idea of universal hashing is to define such a set of elements of a finite set H of hash-functions $h: A \to B$, $|A| = a$, $|B| = b$, so that the random selection of a function $h \in H$ would provide a low collision probability, i.e. for any different inputs x_1 and x_2 the probability that $h(x_1) = h(x_2)$ (the probability of collision, collision) cannot exceed some predetermined value ε:

$$P_{кол} = P(h(x_1) = h(x_2)) \le \varepsilon,$$

moreover, the probability of a collision can be calculated as

$$P_{кол} = \frac{\delta_H(x_1, x_2)}{|H|},$$

where $\delta_H(x_1, x_2)$ – the number of such hash-functions in H for which the values $x_1, x_2 \in A$, $x_1 \uparrow x_2$ cause a collision, i.e. $h(x_1) = h(x_2)$.

We give two definitions of universal hashing [1].

1. Let $0 < \varepsilon < 1$. H is ε – a universal hash-class (abbreviated $\varepsilon - U(H, A, B)$,) if for two different elements $x_1, x_2 \in A$ there are no more than $|H| \cdot \varepsilon$ functions $f \in H$ such that $h(x_1) = h(x_2)$, if $\delta_H(x_1, x_2) \le \varepsilon |H|$ for all, $x_1, x_2 \in A$, $x_1 \uparrow x_2$.

2. Let $0 < \varepsilon < 1$. H is ε – a strictly universal hash-class (abbreviated $\varepsilon - SU(H, A, B)$), if the following conditions are true:

– for everyone $x_1 \in A$ and for everyone $y_1 \in B$, $|\{h \in H : h(x_1) = y_1\}| = |H| / |B|$;

– for everyone $x_1, x_2 \in A$, $x_1 \uparrow x_2$, and for everyone, $y_1, y_2 \in B$, $\left|\{h \in H : h(x_1) = y_1, h(x_2) = y_2\}\right| \le \varepsilon |H|$.

The definition of a universal class of hash-functions is equivalent to the definition of an authentication code generation algorithm, in which the number of different authentication code generation rules (number of keys) for which there is a collision (matching authentication codes) for two arbitrary input sequences is limited. The number of such keys cannot exceed the value $P_{кол} \cdot |H|$, where $P_{кол}$ – is the probability of a collision, $|H|$ – is the number of all rules (keys).

The definition of a strictly universal class of hash functions is equivalent to the definition of such an algorithm for the formation of authentication codes, in which the following rules will be executed:

1. The number of rules for the formation of the authentication code (the number of keys) for which the value of the authentication code does not change for an arbitrary input sequence is limited. The number of such keys cannot exceed the value $|H|/|B|$, where $|H|$ – is the number of all keys, $|B|$ – is the number of possible states of the authentication code;

2. The number of rules for the formation of an authentication code (the number of keys) for which the corresponding values of the authentication code for two arbitrary input sequences do not change is limited. The number of such keys cannot exceed the value $P_{кол}|H|$, where $P_{кол}$ – is the probability of a collision, $|H|$ – is the number of all keys.

The probability of collision of authentication codes in a scheme with strictly universal hashing is defined as $P_{кол} \le \varepsilon$.

The basis of the proposed method of statistical research of collision properties of the formed elements $h(x)$ is an empirical estimate of the maxima of the number of keys (hashing rules) for which:

1. For arbitrary $x_1, x_2 \in A$, $x_1 \uparrow x_2$ equality holds

$$h(x_1) = h(x_2). \tag{1}$$

2. For arbitrary $x_1 \in A$ and $y_1 \in B$ equality

$$h(x_1) = y_1. \tag{2}$$

3. For arbitrary $x_1, x_2 \in A$, $x_1 \uparrow x_2$ and $y_1, y_2 \in B$ equalities hold.

$$h(x_1) = y_1, h(x_2) = y_2. \tag{3}$$

The evaluation by the first criterion corresponds to the verification of the fulfillment of the condition for the universal class of hash functions, the evaluation by the second and third criterion – the conditions for the strictly universal class of hash-functions.

We introduce the following notation:

$$n_1(x_1, x_2) = \left|\{h \in H : h(x_1) = h(x_2)\}\right|,\ x_1, x_2 \in A,\ x_1 \uparrow x_2;$$

$$n_2(x_1, y_1) = \left|\{h \in H : h(x_1) = y_1\}\right|,\ x_1 \in A,\ y_1 \in B;$$

$$n_3(x_1, x_2, y_1, y_2) = \left|\{h \in H : h(x_1) = y_1, h(x_2) = y_2\}\right|,\ x_1, x_2 \in A,\ x_1 \uparrow x_2,\ y_1, y_2 \in B.$$

The first indicator $n_1(x_1, x_2)$ characterizes the number of hashing rules for which for given ones $x_1, x_2 \in A$, $x_1 \uparrow x_2$ equality (1) holds, i.e. the number of keys for which there is a collision (hash codes match) for two input sequences x_1 and x_2.

The second indicator $n_2(x_1, y_1)$ characterizes the number of hashing rules for which $x_1 \in A$, $y_1 \in B$ equality (2) is fulfilled for given ones, i.e. the number of keys for which the hash code value does not change for the input sequence.

The third indicator $n_3(x_1, x_2, y_1, y_2)$ characterizes the number of hashing rules for which $x_1, x_2 \in A$, $x_1 \uparrow x_2$, $y_1, y_2 \in B$ equality (3) is fulfilled for given, i.e. the number of keys for which the two input sequences x_1 and x_2 the corresponding hash code values y_1 and y_2 do not change.

Since the number of keys for which equalities (1), (2) and (3) can hold, should not exceed the corresponding values $P_{кол} \cdot |H|$, $|H|/|B|$ and $P_{кол}|H|/|B|$ we will estimate the maximum number of such keys for each of the considered set of elements.

Thus, the technique allows to evaluate the statistical characteristics of the maxima of these quantities, and then compare the results with the number $P_{кол} \cdot H$ (for the first criterion), with the number $|H|/|B|$ (for the second criterion) and the number $P_{кол} \cdot H$ (for the third criterion).

As statistical indicators for assessing the collision properties, which are carried out experimental studies are used:

- mathematical expectations $m(n_1)$, $m(n_2)$ and $m(n_3)$ the maximum number of hashing rules, under which equalities (1), (2) and (3) are satisfied, respectively;
- variances $D(n_1)$, $D(n_2)$ and $D(n_3)$, characterizing the dispersion of the values of the number of hashing rules for which equalities (1), (2) and (3) are satisfied, relative to their mathematical expectations $m(n_1)$, $m(n_2)$ and $m(n_3)$, respectively.

Evaluation of collision properties by the above criteria is carried out in the average sense. In other words, when setting up an experiment, a limited set of elements $x_1, x_2 \in A$, $x_1 \uparrow x_2$ and the corresponding hash-images $y_1, y_2 \in B$ are used, considering the corresponding results as a sample from the general population.

The natural estimate for the expectation m of a random variable X is the arithmetic average of its observed values X_i (or statistical average) [1]:

$$\tilde{m} = \frac{1}{N}\sum_{i=1}^{N} X_i,$$

where N – the number of realizations of the random variable X.

The variance estimate of the random variable X is determined by the expression

$$\tilde{D} = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \tilde{m})^2.$$

By virtue of the central limit theorem of probability theory, for large values of the number of realizations N, the arithmetic average will have a distribution close

to the normal law with the expectation $m[\bar{m}] \approx \bar{m}$ and standard deviation $\sigma[\bar{m}] \approx \frac{\sigma}{\sqrt{N}}$, where σ – standard deviation of the estimated parameter.

In this case, the probability that the estimate deviates from its expected value by less than ε (by confidence probability) is equal to [1]:

$$P(|\bar{m} - m| < \varepsilon) \approx 2\Phi\left(\frac{\varepsilon}{\sigma[\bar{m}]}\right),$$

where $\Phi(x)$ – Laplace function, determined by the expression:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt. \quad (4)$$

Thus, when conducting experimental studies of collisional properties, methods of statistical testing of hypotheses and mathematical statistics are used.

1. From the general population of the random variable X, a sample is formed as follows:

– for the average statistical estimate of the expectation $m(n_1)$ and variance $D(n_1)$, the maximum at which the equality holds is used as a random variable $n_1(x_1, x_2)$ therefore, a sample of the volume $N: x_1, \ldots, x_n$ is formed by selecting N sets, each of which contains M pairs of elements $x_1, x_2 \in A$, $x_1 \uparrow x_2$ and evaluates $n_1(x_1, x_2)$ those, the total volume of the formed pairs of elements $x_1, x_2 \in A$, $x_1 \uparrow x_2$ will be $N \times M$;

– for the average estimate $m(n_2)$ and $D(n_2)$ the random variable is the maximum $n_2(x_1, y_1)$ at which equality holds $y_1 = h(x_1)$, therefore, the sample of the volume $N: x_1, \ldots, x_n$ is formed by selecting N sets, each of which contains M pairs of elements $x_1 \in A$, $y_1 \in B$ and is estimated $n_2(x_1, y_1)$. The total volume of the formed pairs of elements $x_1 \in A$, $y_1 \in B$ will be $N \times M$;

– for the average estimate $m(n_3)$ and $D(n_3)$ the random variable is the maximum $n_3(x_1, x_2, y_1, y_2)$ at which the equalities are fulfilled $y_1 = h(x_1)$ and $y_2 = h(x_2)$, therefore, the sample of the volume $N: x_1, \ldots, x_n$ is formed by selecting N sets, each of which contains M quadruple elements $x_1, x_2 \in A$, $x_1 \uparrow x_2$, $y_1, y_2 \in B$ and $n_3(x_1, x_2, y_1, y_2)$ is estimated, the total the volume of fours formed will be $N \times M$.

2. In experimental studies of collision hashing properties, the arithmetic average of the observed maximum values n_i and variance is estimated $\bar{D}(n_i)$, $i = 1,2,3$.

3. The reliability of the average estimates obtained is justified as follows. The accuracy is fixed ε and the value of the Laplace function is calculated, which, in accordance with expression (4), gives the corresponding confidence probabilities:

$$P(|\bar{m}(n_i) - m(n_i)| < \varepsilon) \approx 2\Phi\left(\frac{\varepsilon}{\sigma[\bar{m}(n_i)]}\right), \quad \sigma[\bar{m}(n_i)] \approx \frac{\sqrt{\bar{D}(n_i)}}{\sqrt{N}}.$$

In the reverse formulation of the problem, i.e. for a fixed confidence probability P_ε with a sample size of N, the confidence interval is defined as follows:

$$\bar{m}(n_i) - t_\rho \cdot \sigma[\bar{m}(n_i)] < m(n_i) < \bar{m}(n_i) + t_\rho \cdot \sigma[\bar{m}(n_i)], \quad (5)$$

where t_ρ – root of the equation $2\Phi(t_\rho) = P_\beta$.

Thus, the technique, using reduced models of individual transformation layers, on the basis of an estimate of the collision distribution of the images being formed, makes it possible to experimentally investigate the collision properties of message authentication codes.

Experimental results

Using the developed miniature UMAC model (mini-UMAC) and the method of statistical study of the collision properties of message authentication codes, we will experimentally estimate the distribution of the number of collisions (collisions) of the images being formed.

Since in the UMAC scheme discussed above, the first layer (during the formation of the hash-code Y_{mini}) uses the family of universal hashing functions that were studied in detail in [1–7], we will conduct statistical studies only on the second layer (during the formation of a pseudo-random substrate Pad_{mini}) and at the final stage of formation authentication codes (after performing the summation $Tag_{mini} = Y_{mini} \oplus Pad_{mini}$). It is at these stages, according to our assumption, that the universality properties of the generated authentication codes are violated.

To form a miniature UMAC model (mini-UMAC), we use:

1-st layer:

Let us denote the procedure of encrypting a data block T of the length of a Blocklen-byte using the secret key K of the length of the Keylen-byte as a function $Enchiper(K,T)$. Then the procedure for the formation of a pseudo-random key sequence $K' = KDF(K, Index, Numbyte)$ can be represented as the following iterative (for all $n = 1,2,...,n$) transformations:

$$T_i = Index \| i,$$
$$K'_i = Enchiper(K,T),$$
$$K' = K'_1 \| K'_2 \| ... \| K'_n,$$

where $n = \left\lceil \frac{Numbyte}{Blocklen} \right\rceil$.

The generated sequence of pseudo-random key data bits K' has a length of Numbyte-bytes, a multiple of the length of the block Blocklen-bytes.

$$n = \left\lceil \frac{Numbyte}{Blocklen} \right\rceil = \frac{1024 + 16 \times 3}{32} = \frac{1072}{32} = 33{,}5 \approx 33 \Rightarrow i = 1,2,...,33, T_i = Index \| i$$

For the first layer $Index$ =1, => T_i :

T1 = 1 ||1 = 00000001
T2 = 1 ||2 = 0000000100000010
T3 = 1 ||3 = 0000000100000011
T18 = 1 ||18 = 0000000100010010
T19 = 1 ||19 = 0000000100010011
T20 = 1 ||20 = 0000000100010100

T4 = 1 ||4 = 0000000100000100
T5 = 1 ||5 = 0000000100000101
T6 = 1 ||6 = 0000000100000110
T7 = 1 ||7 = 0000000100000111
T8 = 1 ||8 = 0000000100001000
T9 = 1 ||9 = 0000000100001001
T10 = 1 ||10 = 0000000100001010
T11 = 1 ||11 = 0000000100001011
T12 = 1 ||12 = 0000000100001100
T13 = 1 ||13 = 0000000100001101
T14 = 1 ||14 = 0000000100001110
T15 = 1 ||15 = 0000000100001111
T16 = 1 ||16 = 0000000100010000
T17 = 1 ||17 = 0000000100010001
T21 = 1 ||21 = 0000000100010101
T22 = 1 ||22 = 0000000100101110
T23 = 1 ||23 = 0000000100101111
T24 = 1 ||24 = 0000000100110000
T25 = 1 ||25 = 0000000100111001
T26 = 1 ||26 = 0000000100111010
T27 = 1 ||27 = 0000000100111011
T28 = 1 ||28 = 0000000100111100
T29 = 1 ||29 = 0000000100111101
T30 = 1 ||30 = 0000000100111110
T31 = 1 ||31 = 0000000100111111
T32 = 1 ||32 = 0000000100100000
T33 = 1 ||33 = 0000000100100001

$$K' = K'_1 \| K'_2 \| \ldots \| K'_n$$

2-nd layer:

$$n = \left\lceil \frac{Numbyte}{Blocklen} \right\rceil = \frac{24 \times 4}{32} = \frac{96}{32} = 3 \Rightarrow i = 1,2,3,$$

$$T_i = Index \| i$$

For the second layer $Index$ =2, => T_i:

T1 = 2 ||1 = 0000001000000001 =>K1
T2 = 2 ||2 = 0000001000000010 =>K2
T3 = 2 ||3 = 0000001000000011 =>K3
KL2 = K1 || K2 || K3

3-rd layer:

$$n = \left\lceil \frac{Numbyte}{Blocklen} \right\rceil = \frac{64 \times 4}{32 \times 4} = 2 \Rightarrow i = 1,2$$

$$T_i = Index \| i$$

For the third layer $Index$ =3, => T_i:
T1 = 3 ||1 = 0000001000000001 =>K1
T2 = 3 ||2 = 0000001000000010 =>K2
KL3 = K1 || K2

When conducting statistical studies of the collision properties of the values formed Pad_{mini} and Tag_{mini} for each experiment, mathematical expectations $m(n_1)$, $m(n_2)$ and $m(n_3)$, variances $D(n_1)$, $D(n_2)$ and $D(n_3)$, as well as for fixed accuracy ε =0,1, the corresponding confidence probabilities were calculated $P(|\tilde{m}(n_i) - m(n_i)| < \varepsilon)$. Studies were conducted on the sample, the volume N = 100, for the formation of

each element of the sample was calculated by a maximum of a set of M = 1000 tuples of elements. Thus, the total volume of the formed sets was N × M = 105.

The results of experimental studies are summarized in Table 1.

When examining the collision properties of authentication codes generated using the mini version of the AES cipher, the number of keys for which equality $h(x_1) = h(x_2)$ is performed was zero for all tests, i.e. $n_1(x_1, x_2) = 0$ in all N = 100 experiments. This result is explained by the following property. The AES cipher (like its mini version) implements a bijective mapping of a set of plaintext into a set of cipher programs, i.e. for a fixed key, the generated cipher texts corresponding to different plaintext will be different. The experimental research conducted by the first criterion entered was precisely to count the number of keys in which there is a collision (collision) of two cipher texts corresponding to two different plaintext, which is impossible by definition of a bijective cipher. In this regard, the statistical data on the first criterion for the mini version of the AES cipher in the table 1 are not listed as non-informative.

Table 1. – The results of experimental studies of the collision properties of authentication codes generated using mini-AES and mini-UMAC, mini MASH-1, MASH-2 and mini-UMAC

	mini-AES, Pad_{mini}	mini-UMAC, Tag_{mini}	MASH-1	MASH-2
$\bar{m}(n_1)$	-	4.23	41.42	0
$\bar{D}(n_1)$	-	0.18	42.74	0
$P_\varepsilon = P(\lvert\bar{m}(n_1) - m(n_1)\rvert < \varepsilon)$	-	0.98	0.98	≈ 1
$\bar{m}(n_2)$	6.68	4.78	3.99	1
$\bar{D}(n_2)$	0.42	0.42	0.01	0
$P_\varepsilon = P(\lvert\bar{m}(n_2) - m(n_2)\rvert < \varepsilon)$	0.88	0.88	0.99	≈ 1
$\bar{m}(n_3)$	0.19	5.31	0.26	0.31
$\bar{D}(n_3)$	0.15	0.24	0.21	0.22
$P_\varepsilon = P(\lvert\bar{m}(n_3) - m(n_3)\rvert < \varepsilon)$	0.99	0.96	0.97	0.97

The analysis of the data in Table 1 suggests that the results obtained are adequate and that they correspond to the statistical properties of the entire population of data. For fixed accuracy = 0,1, high confidence values were obtained, which indicates the validity and reliability of the experimental results obtained.

Let us analyze the results of statistical studies of the collision properties of message authentication codes, compare the obtained results of average estimates of mathematical expectations $m(n_1)$, $m(n_2)$ and $m(n_3)$ the number of hashing rules

for which equalities (1), (2) and (3) are satisfied, respectively, with theoretical estimates: number $P_{кол.т} \cdot |H|$ (for the first criteria), with a number $|H|/|B|$ (for the second criterion) and a number $P_{кол.т} \cdot H$ (for the third criterion).

Consider the first criterion by which we estimate the number of hashing rules for which there is a collision (coincidence of authentication codes) for two arbitrary input sequences. In accordance with theoretical estimates, this value is bounded above by a number $P_{кол.т} \cdot |H|$. Let us specify this (theoretical) estimate for authentication codes generated using mini-AES and mini-UMAC.

The power of the key set for mini-AES and mini-UMAC is $|H| = 2^{16}$, the power of the set of generated authentication codes is also $|B| = 2^{16}$. If we use the upper estimate of the probability of collisions as the inverse of the power of the generated authentication codes $P_{кол.т} = 2^{-16}$, we get $n_1(x_1, x_2) \le P_{кол.т} \cdot |H| = 1$. For the mini version of the AES cipher, this condition is met (justified by the bijectivity of the encryption transform), but the collisional properties of mini-UMAC are significantly inferior to this upper theoretical estimate. In fact, the number of collisions is more than four times higher than the theoretical limit and this position is confirmed with a high confidence level $P_д = P(|\bar{m}(n_1) - m(n_1)| < 0{,}1) > 0{,}98$.

Consider the second criterion by which the hashing rules are evaluated, for which for an arbitrary input sequence the value of the authentication code does not change. According to theoretical estimates, this value for authentication codes generated using mini-AES and mini-UMAC is bounded above by a number $|H|/|B| = 1$. The experimental results obtained indicate that the collisional properties of authentication codes generated using mini-AES and mini-UMAC do not satisfy the second criterion, the number of keys for which the value of the authentication code does not change several times the arbitrary input sequence exceeds the theoretical estimate for universal hashing.

In accordance with the third criterion, the number of hashing rules is estimated, for which, for two arbitrary input sequences, the corresponding values of the authentication code do not change. The theoretical estimate of this value for universal hashing is bounded above by a number $P_{кол.т} \cdot |H|$, which, using the upper estimate of the probability of collisions $P_{кол.т} = 2^{-16}$, gives $n_3(x_1, x_2, y_1, y_2) \le P_{кол.т} \cdot |H| = 1$. The values given in table 2 indicate that the collisional properties of authentication codes generated using mini-AES satisfy the third criterion. At the same time, the number of mini-UMAC keys, for which the corresponding authentication code values do not change for two arbitrary input sequences, is more than five times higher than the upper theoretical estimate.

Thus, to ensure the universality of the UMAC hash code, it is necessary to "replace" the substrate based on the block symmetric AES cipher with a crypto-resistant sequence, which ensures the crypto-resistance of the entire hashing system. In [3, 4], as a "replacement", it is proposed to use the MASH-1 and MASH-2 hashing algorithms, which provide the required level of cryptographic strength. Evaluation of cryptographic strength is given in (Table 2).

However, the results presented in table 1 show that the implementation of the key hashing scheme based on the MASH-1 algorithm, while changing the values of the initialization vector by the secret key, does not allow for high collision properties. The number of collisions that occur significantly higher than the upper theoretical boundaries, both in the first and second criteria, therefore, this design is not a universal and, especially, strictly universal hashing scheme.

Table 2. – Results of experimental studies of the statistical security of hashing algorithms using the NIST STS package

Generator	The number of tests where the test passed > 99%	The number of tests where the test passed > 96%
MASH-1	101 (53%)	188 (99%)
MASH-2	126 (67%)	189 (100%)
MASH(EC)	141 (74%)	189 (100%)
HMAC-SHA-256	134 (71%)	187 (98%)
EMAC	138 (73%)	189 (100%)
RIPEMD-160	129 (68%)	189 (100%)
UMAC+MASH-2	173 (91%)	189 (100%)

This result was obtained with a high confidence level for high accuracy. Thus, for the first criterion, the confidence interval was (confidence 0,98), for the second criterion (confidence 0,99), for the third (confidence 0,97). The considered key hashing scheme based on the MASH-1 algorithm, when the initialization vector values change with a secret key, satisfies only the third criterion.

Using key hashing based on the MASH-2 algorithm when changing the values of the initialization vector with a secret key, on the contrary, provides high collision characteristics of universal hashing. So, their use as a "substrate" significantly reduces the rate of conversion and the formation of a hash code. A promising direction is the use of crypto-code structures proposed in [5]. (Fig. 1) shows a block diagram of the hashing algorithm, taking into account the use, *Pad* as a substrate, of the Mac-Alice crypto-code construct on elliptic (EC) modified elliptical (shortened / extended) codes.

This approach ensures the preservation of the properties of the universal hashing algorithm, as well as the required level of security, efficiency and reliability in the formation of the hash code.

Conclusion

Thus, the results of statistical studies of the collision properties of hash codes based on mini-AES and mini-UMAC, as well as MASH-1 and MASH-2, allow us to state:

- the cryptographic layer of the formation of message authentication codes (mini-AES) satisfies the properties of universal hashing, the probability of a

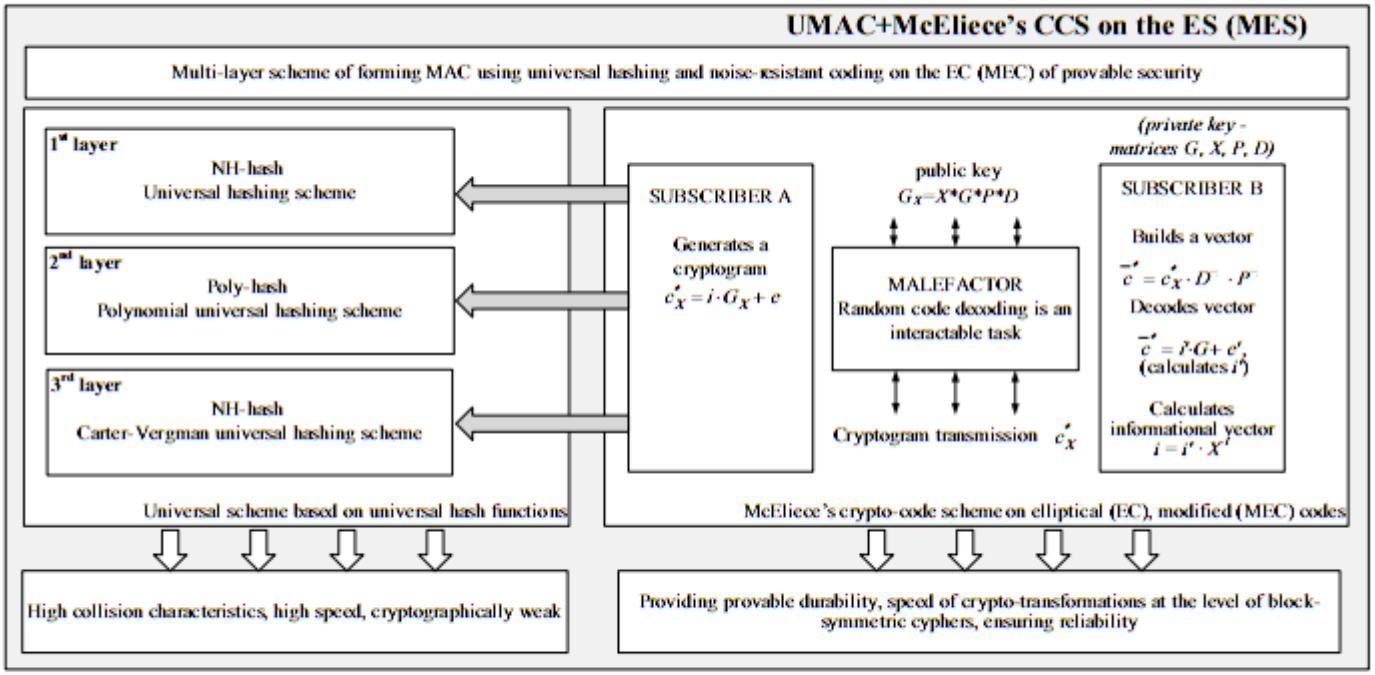

Figure 1 – The block diagram of the UMAC algorithm with Mc-Alice's KKK on the EC

collision of generated hash images does not exceed a predetermined value. However, this transformation layer does not satisfy the properties of strictly universal hashing;

- the result of the formation of message authentication codes using the mini-UMAC scheme does not satisfy the properties of both universal hashing and, especially, the properties of strictly universal hashing. This is explained by the fact that the scheme with simple sum modulo two (XOR) of two universal hash results does not always preserve the properties of universal hashing;
- the use of the MASH-1 algorithm as a hash-code does not satisfy the requirements for security and speed, the MASH-2 algorithm does not satisfy the speed of transformations, which does not allow their practical use in decentralized cryptocurrency systems.

The proposed application of the Mac-Alice crypto-code construct on elliptic (EC), modified elliptical (shortened / elongated) codes (MEC) retains the universality property on the first two layers of the UMAC algorithm, provides the required level of security, efficiency and reliability when generating the message hash. Thus, this algorithm allows increasing the levels of basic hashing functions in the blockchain technology in Bitcoin protocols under the conditions of using a full-scale quantum computer.

References:

1. Yevseiev S.P., Iohov O.Y. ta Korol O. H. Heshuvannia daniih v informaciyniih systemah. Monografia. – Kharkiv, Ukraiina: Vyd. KhNEU, 2013.
2. Yevseiev S., Ostapov S. ta Koroliov R. Vykorystannia mini-versiy dlia ocinky stiykosti blokovo-symetrychnyh shyfriv, Naukovo-tehnichnyi zurnal "Bezpeka informacii", – Tom 23. – № 2. – P. 100–108. 2017.
3. Parhuc L., Korol O., i Yevseiev S. Razrabotka modeli i metodaa kaskadnogo formirovaniya MAC s ispolzovaniiem modulyarniih preobraxovaniiy, Zahyst informacii: naukovo-tehnichnyi zurnal, – Tom 15. – № 3. – P. 186–196. 2013.
4. Yevseiev S., i Korol O. Metod kaskadnogo formirovaniia MAC-kodov na osnove modulyarnyh preobrazovaniy, Izvestiia Vysshiih tehnicheskih uchebniih zavedeniy Azerbaydzana, – № 1(89). – P. 71–78. 2014.
5. Yevseiev S., Korol O., and Kots H. Construction of hybrid security systems based on the crypto-code structures and flawed codes, Vostochno-yevropeyskiy zurnal peredovyh tehnologiy, 4/9(88). 2017. – P. 4–20.