Development and Analysis of a Cryptographic Algorithm for Genetically Optimized Wireless Sensor

Abstract. This article proposes a methodology for obtaining a cryptographic algorithm, optimized for wireless sensor networks, through genetic algorithm. With the objective of increasing the level of security, computational efficiency and highlighting the energy consumption, considering that the autonomy of the wireless sensor devices is directly influenced by this factor. In aptitude function of genetic algorithm, were used metrics of algorithm runtime, maximum deviation and irregular, space occupied in memory and correlation coefficient (a new proposed metric), in order to find a safe and fast algorithm. The results obtained through computational simulations show the efficiency of the proposed methodology, in terms of processing time, coefficient of correlation and occupation of memory.

Streszczenie. W tym artykule zaproponowano metodologię uzyskiwania algorytmu kryptograficznego, zoptymalizowanego dla bezprzewodowych sieci czujników, za pomocą algorytmu genetycznego. W celu zwiększenia poziomu bezpieczeństwa, wydajności obliczeniowej i podkreślzenia zużycia energii, biorąc pod uwagę fakt, że ten czynnik ma bezpośredni wpływ na autonomię bezprzewodowych czujników. W funkcji udzielenia algorytmu genetycznego wykorzystano metryki czasu pracy algorytmu, maksymalnego odchylenia i nieregularności, miejsca zajmowanego w pamięci i współczynnika korelacji (nowa proponowana metryka), aby znaleźć bezpieczny i szybki algorytm. Wyniki uzyskane za pomocą symulacji obliczeniowych pokazują efektywność proponowanej metodologii, pod względem czasu przetwarzania, współczynnika korelacji i zajęcia pamięci.

Keywords: Cryptographic Algorithm, Genetic Programming, Wireless Sensor Networks

Słowa kluczowe: Algorytm kryptograficzny, Programowanie genetyczne, Sieci czujników bezprzewodowych

Introduction

The wireless sensor networks consist of nodes interconnected by wireless channels with sensing, control and actuation capacity. The data obtained and transmitted by the wireless sensor nodes in an industrial environment, especially when the sensor nodes are installed outdoors, must be confidential. For this purpose, are available in the literature various techniques that aim to maintain the confidentiality of the information transmitted. Among them, there is an approach that employs the increase of data security transmitted in wireless sensor networks, it is the use of cryptography techniques, according to [1].

Crypographic security is responsible for keeping the information confidential, preventing the system from being invaded. By allowing the inclusion of false information in the system, an invasion can lead to even greater failures in the entire monitoring and control process, according to [1].

The most widely used standard in wireless sensor networks is the IEEE 802.15.4 [2], created in parallel with the Network Protocol ZigBee [3]. The standard IEEE 802.15.4 defines the physical layer and the Medium Access Control (MAC) layer, as well as the recommended encryption technique for system security, the Advanced Encryption Standard (AES) [4].

The AES-128 is an efficient technique in terms of data processing and security. There are some reports of security breaches in partial versions of the algorithm, which does not mean that it is unsafe. However, there are studies indicating that it can be broken in a timely manner in the coming decades [5, 6].

It is important to emphasize that the energy efficiency of the nodes of the wireless sensor networks is more critical than in designed systems in which it does not need an autonomy of months and even years. In this context, electronic devices of low power and complexity are used, such as 8-bit and low memory microprocessors [7, 8].

There are papers in the literature that demonstrate that it is possible to create cryptographic algorithms with acceptable performance and security characteristics using genetic programming (GP) [9, 10, 11]. However, none of these papers presented results focusing on devices with only 8-bit processing capacity, taking into account both energy efficiency, occupied memory space and security, or a more robust comparative study, including cryptographic algorithms in effect.

To demonstrate the feasibility of using the GP technique in the development of cryptographic algorithms, two classical encryption algorithms Advanced Encryption Standard, XXTEA, and the CRYSEED algorithm presented in the paper [12]. The results presented in this work demonstrate that the use of GA can create efficient cryptographic algorithms, being a significant advance for the diffusion of artificial intelligence techniques.

Cryptographic Algorithms

This section presents two cryptographic algorithms which were used in this work to compare performance of the proposed algorithm.

The Advanced Encryption Standard (AES), is the chosen algorithm by the US government for the protection of data traffickedby digital means [13]. AES can use keys of 128, 192 and 256 bits. The algorithm considers groups of 16 bytes that form one state, which in other words is simply a 4 × 4 array with each element being a byte. This algorithm can be divided into four basic operations:

- AddRoundKey: Each byte of a state is combined with a byte of the key through the XOR (OR Exclusive) bit by bit.
- SubBytes: A nonlinear substitution where each byte is replaced by another according to a table lookup.
- ShiftRows: The bytes of each line undergo a shift cyclic by a certain amount of steps.
- MixColumns: Each column of state a(x) is combined according to a known matrix c(x) 4 × 4, which multiplied by state columns returns a new column for the resulting

Xiankleber C. BENJAMIM1, Felipe O. S. GAMA1, Rodrigo S. SEMENTE2,Elmer R. L. VILLAREAL2, Andrés O. SALAZAR3

Università Federal do Rio Grande do Norte (1), Universidade Federal Rural do Semi-Árido (2)

doi:10.15199/48.2018.08.21

PRZEGŁAD ELEKTROTECHNICZNY, ISSN 0033-2097, R. 94 NR 8/2018 79
state $b(x)$. The operation is reversible, simply by multiplying the columns of state $b(x)$ by a matrix $d(x)$, the result will be state $a(x)$.

Given these four operations, the algorithm can be divided into four stages, where the third step is usually repeated 20 times:

1. Key expansion based on key scheduling;
2. Start round: AddRoundKey;
3. Rounds:
   - SubBytes;
   - ShiftRows;
   - MixColumns;
   - AddRoundKey
4. Final round (without column mixing):
   - SubBytes;
   - ShiftRows;
   - MixColumns;

In the work [14] he developed the block encryption, Tiny Encryption Algorithm (TEA), which was notable for its simplicity and few lines of code, but that soon presented a security flaw that made the maximum size of its equal key 126 bits, making it so easy to break it such as a hash table or cryptographic scatter, tables substitutions created by spreading functions.

In 1997 the TEA was improved with an extension, thus generating Block TEA, or as it is better known the XTEA (eXtended TEA), which keys with a size of 128 bits [15]. Due to other security issues, the algorithm was restructured a second time, creating the XXTEA [16].

The XTEA is based on the classical Feistel network, varying only the Feistel function, as can be observed in Figure 1. XOR and modular sum operations are represented, respectively, by the square and the circle in Figure 1. The left and right shift operations are also used, but with a fixed number of offsets, 4 bits and 5 bits respectively.

XXTEA is an unbalanced Feistel network, whose input block is divided into $N$ parts, each part being called $X_R$. So, the encryption of each die depends both his predecessor and successor neighbors, as observed in Figure 2, which presents the Feistel function that depends on both the $X_{R-1}$ and $X_{R+1}$. Also used is $K_R$, which applies a left-hand cyclic shift of 2 bits each round.

A fault in the XXTEA has already been found, with key breaking through differential cryptanalysis, performing a number of queries of only $2^{10}$ [19]. This method is a little difficult to apply in practice in an RSSF, since you have to choose the correct plain text for comparison. The data generated by a node of the RSSF are obtained with each measurement, they are not data that can be known by an attacker, being difficult to carry out this attack. Even so, this break can mean the existence of other faults, which may indicate the non-recommendation of its use, even in the Wireless sensor network.

Genetic Programming

Based on natural evolution, several computational program solution methods have been developed, such as genetic algorithm, evolutionary programming, evolutionary strategies and genetic programming. The genetic algorithm allows the creation of programs to solve a problem, created automatically by the computer using a procedure inspired by evolution [20]. This program is usually represented by a tree structure, where each node and a function, where it receives an input from the son and returns an output to the parent, and the root node returns the final output of the program.

Genetic programming can be represented by a tree structure, where each node is a function, which receives an input from each child and returns an output to the parent, and where the root node returns the final output of the program. Thus PG follows the following steps:

- Initial population: it is the initial step of the genetic algorithm is the creation of an initial population of individuals. In this work the initial population is created in a random way.
- Fitness Function: it is the evaluation of the performance of each individual of the population, measuring the how much the result of that species is better compared to the others.
- Selection: it select the individuals with better aptitude from the previous population to generate the population. These individuals may pass through reproduction, crossover or mutation.
- Reproduction: choose the individuals with better aptitude from the previous population to generate the next population. These individuals may pass through reproduction, crossover or mutation.
The following are the changes in the library:

- Crossover: each gene can be exchanged between two individuals, according to a uniform probability, generating two sons.
- Mutation: small probability of one or more random genes being exchanged for any other, generating a new individual slightly modified.
- New population: from the processes of reproduction, crossover and mutation a new population is generated.
- Satisfy Stop Condition: stop criteria are checked, which can contain: maximum number of iterations, so that the program can stop (mandatory condition); rate of population, stopping the program if it is not able to vary the population further.
  - If the stopping criterion is satisfied, the aptitude test must be performed, returning the best individual of the current population.
  - If it is not satisfied you should return to the second step.

The execution time was used for the fitness function in genetic programming. In this work we used the coefficient of correlation in Fitness Function of genetic programming, where this metric is a measure of similarity between two variables [21, 22, 23]. If the two variables are an image and its encryption, then they are in perfect correlation (that is, coefficient of correlation equals one) if they are highly dependent (identical). In this situation, the encryption process failed to hide the details of the original image. If the correlation coefficient is zero, then the original image and the encrypted image are totally different, that is, the encrypted image has no recognizable feature and is highly uncorrelated from the original image. If the coefficient of correlation equals \(-1\), this means that the encrypted image is the negative of the original image. The parameters used in the genetic programming were: a population of 200 individuals, 8192 test samples with 100 generations. Two weights were used for variable balancing of the fitness formula, varied ten times for each, thus creating the equivalent of 100 individual tests. A total of four test batteries were made, totaling 400 samples.

The evaluation function used in this phase follows as defined by:

\[
(1) \text{Fit} = \left( \frac{T_N}{100} + \left( \Delta R \frac{\text{MIDR}}{\text{MIDR}_B} + \Delta C \frac{\text{CC}_S}{\text{CC}} \right) \right),
\]

where \(\Delta R\) and \(\Delta C\) are the variation weights of the MIDR and coefficient of correlation parameters, \(\text{CC}_S\) is the lowest correlation coefficient between the algorithms studied for each image, and \(\text{MIDR}_B\) is the Maximum and Irregular Deviation Ratio of best individual. This variation has as main objective to find the relation between the two metrics, a balance of difficult balance. It was noticed that small variations (with difference of 0.1 between the weights) did not present significant changes in the results, so variations were used in the order of 10\(^{-N}\), with \(N\) being the execution index of each of the genetic programming individually.

### Results Obtained

In this section presents algorithm obtained through computational simulation. In the development of the algorithm the Maximum and Irregular Deviation Ratio, and Correlation Coefficient were considered, the algorithm obtained through the proposed methodology of this work was named CRYSEED2. The execution time was used for the fitness function in genetic programming. The calculation of the occupied memory was calculated after the final results, because a reliable calculation of the occupied memory through the structures of trees created by genetic programming is not feasible. To compensate for this problem, we used a 50-node number limitation on the tree. Another factor that also helps to compen-

---

**Fig. 3.** Diagram of execution of a genetic programming.

- Crossover: each gene can be exchanged between two individuals, according to a uniform probability, generating two sons.
- Mutation: small probability of one or more random genes being exchanged for any other, generating a new individual slightly modified.
- New population: from the processes of reproduction, crossover and mutation a new population is generated.
- Satisfy Stop Condition: stop criteria are checked, which can contain: maximum number of iterations, so that the program can stop (mandatory condition); rate of population, stopping the program if it is not able to vary the population further.
  - If the stopping criterion is satisfied, the aptitude test must be performed, returning the best individual of the current population.
  - If it is not satisfied you should return to the second step.

The Figure 3 presents the order of execution of these steps, presented the probabilities of crossover \(P_{\text{cross}}\), probability of reproduction \(P_{\text{reprod}}\) and probability of mutation \(P_{\text{mut}}\). Some genetic programming do not use the reproduction step, since the crossover is similar in some respects to the first one, changing genes between two individuals [20].
sate for this limitation is the use of runtime optimization due to the relationship of small algorithms to runtime. Also in this section a performance analysis is performed of the algorithm obtained.

The sets of functions used in this work were based on the functions used in the work [24]. Randomly generated constants were also added. The evaluation function used in this phase is given by:

\[
(2) \quad \text{Fit} = \left(0,2T_N + \frac{0.4DM}{DM_M} + \frac{0.4DI_M}{DI}\right).
\]

After 20 interactions, the PG converged to an algorithm with a tree function \((\text{add(shift a}_{10} a_{12})(\text{xor a}_{5} a_{2}))\). In Algorithms 1 and 2 they present the pseudocodes of the encryptor and decryptor CRYSEED2, respectively.

**Algorithm 1 Encryption Pseudocode CRYSEED2**

1: for \( t = 0 \) to number of rounds do
2:   \( v_1(j) = v_2(j) \land \text{round key} \)
3:   end for
4: for \( k = 0 \) to 16 do
5:   for \( j = 0 \) to 16 do
6:     \( v_2(j) = v_1((k + j)\%8) \)
7:   end for
8: end for
9: \( T_1 = v_2(10) \ll v_2(12) \)
10: \( T_2 = (v_2(5) \land v_2(2)) \)
11: \( v_1(k) = (T_1 + T_2) \land v_2(0) \)
12: end for
13: for \( j = 0 \) to 8 do
14:   \( v_2(j) = v_1(j) \)
15: end for
16: end for

**Algorithm 2 Pseudocode of decryptor CRYSEED2**

1: for \( t = 0 \) to number of rounds do
2:   for \( j = 0 \) to 16 do
3:     \( v_1(j) = v_2(j) \land \text{round key} \)
4:   end for
5: for \( i = 1 \) to 16 do
6:   for \( j = 0 \) to 16 do
7:     \( v_2(j) = v_1((k + j)\%8) \)
8:   end for
9: \( T_1 = v_2(10) \gg v_2(12) \)
10: \( T_2 = (v_2(5) \land v_2(2)) \)
11: \( v_1(k) = (T_1 - T_2) \land v_2(0) \)
12: end for
13: for \( j = 0 \) to 8 do
14:   \( v_2(j) = v_1(j) \)
15: end for
16: end for

Table 1 presents the execution time of the obtained algorithm and the reference algorithms. Comparing the results presented in Table 1, one can verify that the algorithm proposed in this work has a processing time reduction of 28%.

In particular, it can be seen that CRYSEED2 has a processing time reduction of 84% over XXTEA. These times demonstrate the low computational cost of the proposed algorithm when compared with the reference algorithms, thus increasing the energy efficiency of the wireless sensor devices. It is important to note that the tests were performed on a computer with an Intel Core i5-2520M 2.5 GHz processor, with 8 GB of RAM DDR3-1333 SDRAM, Windows 7 Ultimate 64-bit operating system.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Encryption Time (μs)</th>
<th>Decryption Time (μs)</th>
<th>Total Time (μs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AES</td>
<td>0.708</td>
<td>0.706</td>
<td>1.414</td>
</tr>
<tr>
<td>XXTEA</td>
<td>1.1097</td>
<td>1.097</td>
<td>2.204</td>
</tr>
<tr>
<td>CRYSEED2</td>
<td>0.2718</td>
<td>0.22</td>
<td>0.4918</td>
</tr>
<tr>
<td>CRYSEED2</td>
<td>0.175</td>
<td>0.175</td>
<td>0.351</td>
</tr>
</tbody>
</table>

The level of quality of an encryption depends on how much greater the difference between the original data and the encrypted data, and even if it spreads the data in an irregular way, therefore a good cryptographic algorithm must be able to generate large values of \( D_M \) (maximum deviation) and small \( D_I \) values (irregular deviation). In this context, a performance analysis of the proposed algorithm through the MIDR, where the MIDR is given by:

\[
(3) \quad \text{MIDR} = \left(D_M \times \frac{D_I}{D_M}ight),
\]

where \( D_M \) and \( D_I \) are the maximum and irregular deviations of the best individual.

The images used for the comparative test among the algorithms were: lena, the most used image in scientific articles; cameraman, used for having shades of gray and light and dark areas; mandrill, an image with high contrast of high frequency colors; testpat, an image with squares of various sizes of various shades of gray and in the center the image lena. These images are shown in Figure 4. All images were obtained from The USC-SIPI Image Database [25]. Tests with other images were also performed to obtain more significant results from the proposed technique, these images are illustrated in Figure 5. The results of this encryption were used to feed the adjust the importance of each metric in the quality assessment cryptographic. In the figure below we can observe the images used in encryption.

Table 2 shows the results obtained through simulations of the performance of the proposed algorithm and the reference algorithms, in terms of MIDR, each value of the images refers to the Maximum and Irregular Deviation Deviation of each algorithm related to the figures. It is observed that CRYSEED 2 had a considerable gain over other reference algorithms such as AES and XXTEA. It is observed that this gain of CRYSEED2 in the MIDR occurs through its encryption of the proposed algorithm, showing that the encrypted algorithm is superior to its predecessors comparing the means MIDR obtained by the reference algorithms, thus becoming the best algorithm analyzed in this work demonstrating the metric Coefficient of Correlation applied was better in the cryptographic analysis of the images.

Coefficient of Correlation (CC) is a measure of similarity between two variables [22, 23, 21]. If the two variables are an image and its encryption, then they are in perfect correlation (that is, CC equals one) if they are highly dependent (iden-
Coefficient of Correlation

The standardized CC is given by:

\[ CC = \frac{CC_M}{CC_G} \times 100 \]

Where CCN is the normalized value in percentage and CCM is the lowest correlation coefficient between the algorithms studied for each image.

The results obtained in the computational simulations with the objective of measuring the correlation coefficient can be seen in Table 3. It should be noted that the results obtained with the images lena, black, arial and mandrill were the better results. The proposed technique obtained good gains in the correlation coefficient showing that the analyzed images are different from the originals, since CC is negative. So it shows the efficiency of the proposed technique.

$CTABLE 3. COEFFICIENT OF CORRELATION$

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>lena</th>
<th>cman</th>
<th>mandrill</th>
<th>testpat</th>
<th>black</th>
<th>arial</th>
<th>random</th>
<th>earth</th>
</tr>
</thead>
<tbody>
<tr>
<td>AES</td>
<td>0.14</td>
<td>-0.03</td>
<td>0.005</td>
<td>-0.005</td>
<td>-0.02</td>
<td>-0.03</td>
<td>-0.07</td>
<td>-0.02</td>
</tr>
<tr>
<td>XXTEA</td>
<td>0.011</td>
<td>-0.002</td>
<td>-0.001</td>
<td>-0.002</td>
<td>-0.002</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.001</td>
</tr>
<tr>
<td>CRYSEED1</td>
<td>0.18</td>
<td>-0.18</td>
<td>-0.17</td>
<td>-0.18</td>
<td>-0.03</td>
<td>-0.01</td>
<td>-0.05</td>
<td>-0.05</td>
</tr>
<tr>
<td>CRYSEED2</td>
<td>-0.018</td>
<td>-0.017</td>
<td>-0.017</td>
<td>-0.008</td>
<td>-0.028</td>
<td>-0.003</td>
<td>-0.01</td>
<td>-0.008</td>
</tr>
</tbody>
</table>

Fig. 5. Standard images: (a) black; (b) arial; (c) random (d) earth.

The Table 4 presents the performance of the algorithms analyzed in terms of occupied memory, in which one can analyze the gain of the proposed algorithm, specifically if we compare CRYSEED2 in relation to AES.

$TABLE 4. CRYPTOGRAPHIC OCCUPIED MEMORY$

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Occupied memory (Bytes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AES</td>
<td>9,437</td>
</tr>
<tr>
<td>XXTEA</td>
<td>1,570</td>
</tr>
<tr>
<td>CRYSEED1</td>
<td>1,078</td>
</tr>
<tr>
<td>CRYSEED2</td>
<td>1,078</td>
</tr>
</tbody>
</table>

Conclusions

The genetic programming methodology was applied and improved for the development of cryptographic algorithm for 8-bit devices, widely used in wireless sensor networks. The CRYSEED2 algorithm was automatically developed from the use of the lil-GP library, adapted to find an efficient cryptographic algorithm for 8-bit devices with high-quality encryption. Also shown performed comparative tests between the algorithm proposed in relation to widely used algorithms literature, in which the following comparison metrics were used: runtime and Maximum and Irregular Deviation Ratio, space occupied in memory and correlation coefficient. The obtained results demonstrate that CRYSEED2 is competitive in comparison to the other presented algorithms. Making possible the insertion of the proposed algorithm in wireless sensor networks, since CRYSEED 2 has low computational complexity and good cryptographic performance. Finally, these results help to prove that genetic programming can create efficient cryptographic algorithms and is a significant advance in the diffusion of artificial intelligence techniques.

Authors: Msc. Xiankleber C. Benjamim, Brazil, E-mail: xiankleber@dca.ufrn.br; Msc. Felipe O. S. Gama, Brazil, E-mail: felipe.gama@dca.ufrn.br; Dr. Rodrigo S. Semente, Brazil, E-mail: rodrigo.semente@ufes.edu.br; Dr. Elmer R. L. Villarrea, Brazil, E-mail: Dr. André O. Salazar, Brazil, andres@dca.ufrn.br

REFERENCES


