One-way Function Based on Modified Cellular Automata in the Diffie-Hellman Algorithm for Big Data Exchange Tasks through Open Space

Volodymyr Shevchenko a, Georgi Dimitrov b (✉), Denis Berestov a, Pepa Petrova b, Igor Sinitcyn c, Eugenia Kovatcheva b, Ivan Garvanov b, Iva Kostadinova b

a Taras Shevchenko National University of Kyiv, Kyiv, Ukraine
https://pst.knu.ua/
b University of Library Studies and Information Technologies, Sofia, Bulgaria, https://www.unibit.bg/
c Institute of Software Systems, National Academy of Sciences of Ukraine
http://www.isofts.kiev.ua

ABSTRACT:
The article deals with ways to quickly change passwords in information exchange through open space. It suggests an improvement of the Diffie-Hellman algorithm by creating a one-way function on the basis of cellular automata with an extended set of rules. The authors have expanded the rules of the game of life towards definition of the rules of birth rate and life extension, control of the radius of intra-population interaction, rules of death from the age of cells, multi-component (multi-population) system of cells. The created algorithm on the basis of a cellular automaton is used to create keys for safe information transfer. Depending on the needs of encryption, the algorithm can be enhanced by using variable parameters of the cell field and cell behaviour, which will allow to regulate the speed and reliability of encryption. The implementation is in Python and MatLab, which allows to compare results and change the modelling environment when changing the features of the task.

ARTICLE INFO:
RECEIVED: 07 JUNE 2020
REVISED: 23 AUG 2019
ONLINE: 22 SEP 2020

KEYWORDS:
Diffie-Hellman algorithm, cellular automata, one-way function

Creative Commons BY-NC 4.0

✉ E-mail: g.dimitrov@unibit.bg
Introduction

Modern life involves the exchange of information that is transmitted much faster and in larger quantities than it used to be. This was made possible by the emergence of computer networks. But with the advent of a large network of information, there are also offenders who want to get or change the information from the network. In the context of Big Data, these may be, for example, cases of interference with bank transactions and transfer of particularly large data between companies.\(^1\), \(^2\) The number of incidents of cybernetic attacks in the world is growing at least twice as fast as the increase in global GDP,\(^3\), \(^4\), \(^5\), \(^6\) which is now almost linearly related to the growth of the digital devices market. At the same time, the success rate of password cracking is between 20 and 40% of the total number of passwords found by hackers.\(^7\), \(^8\) A highly skilled hacker\(^7\), \(^8\) can break 38% of passwords in 4 hours, 62% in a week, and 89% in 5 weeks. These statistics concern passwords created by people; artificially generated passwords are more reliable. To prevent interference with the transmission of information, it is necessary to encrypt it. At the same time, the encryption keys need to be constantly changed with two basic rules in mind: the generated password must be created without human intervention and must be random (pseudo-random), passwords should be changed as often as possible.

It is important to note that usually the generation of encryption keys takes place in an open network environment using special methods. One of such methods is the Diffie-Hellman method\(^9\), \(^10\) which is designed to generate shared private keys using public channels. The basis of this type of key generation is the use of one-way arithmetic functions\(^11\) with the commutability property. An intruder can decrypt the information without knowing the key only by finding the reverse function. Since the calculated power of computers used to find reverse functions is increasing rapidly every year, increasing resistance to finding reverse one-way functions is an actual task.

Increase of computational power in combination with social engineering methods create a danger of finding the inverse function to the known one-way functions.\(^10\), \(^11\) One of the ways to overcome this situation is to widen the range of one-sided functions and expand the parameters which will vary the behaviour of such functions. Cellular automata can be used as a new type of one-way functions. For example, in works\(^12\) cellular automata were used to generate pseudo-random numbers in cryptographic systems. This experience can be used, but, first, it does not directly concern one-way functions, and secondly, standard cell automata, which were proposed by von Neumann, were used in the research. Such cellular automata are guided by a small number of parameters, and in their turn have a small number of combinations among themselves. The number of possible combinations of input parameters is not large enough for research and further selection of optimal one-way functions. At the same time, a deeper study of the history of creation and general properties of cellular automata can provide additional ideas for improving the quality of unilateral functions on the basis of cellular automatons.
The problem is that in general, the existing examples of cellular automata are quite predictable in their structure. As in the previous example, in these automata, the behaviour is controlled by a small number of parameters, so that diversity is created for the choice of input parameters. Also because of this the following problem appears: cellular automata generated in this way are not able to exist without repeats and statics for a sufficiently long period of time. This creates a contradiction between the need for a variety of one-way functions for the Diffie-Hellman algorithm and the uniformity of the rules, as well as a narrow set of control parameters of the classic models of cellular automata such as "Game of Life" by John Conway.

Purpose of work. To improve the quality of the process of creating encryption keys for the transmission of Big Data information through open communication channels within the Diffie-Hellman algorithm using a one-way function based on complex cellular automata.

For the first time, the concept of cellular automaton was introduced by John von Neumann to denote models of self-replicating structures. Most of the structures, he investigated were one-dimensional and two-dimensional. The greatest attention was paid to automata in the 60s and 70s, when this topic was studied by such scientists as Edgar Codd, Gordon Moore, Stanislav Ulam. In 1970 the most cellular automaton was created and described - "Game of Life" by John Conway. He became interested in the problem proposed by Neumann, who tried to create a hypothetical machine that can reproduce itself. Neumann managed to create a mathematical model of such a machine with very complex rules. Conway simplified Neumann's ideas and created the "Life Games" rules.  

The "Games of Life" model was popular with both Conway's colleagues and ordinary users. The model and its further variations influenced different sections of computer science, mathematics and physics. This is evidenced by the many different computer implementations of this game in different fields of knowledge. The following studies are clear proof of this opinion: the study of population growth dynamics in different conditions, in particular, in the case of the city area increase models, the "predator-prey" model, the model of traffic in one lane. In the demographic dynamics model, Rosana Motta Jaeflile and Patricia Nunes da Silva used an approach similar to the predator-prey model. Such a model represents a two-dimensional cellular automata in which neighbouring cells are calculated by the von Neumann rules. This model can be useful for creating a more complex model of cell behaviour, but the disadvantage of such an approach is that it does not take into account the ability of creatures to move in any direction, which causes the loss of 4 diagonal neighbours.

In his research Conway considers one of 2025 possible variants of cellular automata of this type. This variant of initial conditions creates a stable automata in comparison with its analogues, i.e. complete extinction or complete filling of the field is unlikely. Anyway in some interval of time the model comes to a condition when there are either static or cyclic sets of figures. It makes the classical automata "Game of Life" inconsistent to create rather unpredictable structures.
of cells. Therefore, it can only be used as a basis for a more complex model of interaction and development of cells. Another disadvantage is that in the cellular automaton "Game of Life", as in most others, the interaction radius is always \( R = 1 \). This imposes restrictions on the possible number of combinations of initial parameters of cell behaviour.

In our work, the rules of "Game of Life" are supplemented by the rules of interaction of cells, which will become the tools for subsequent creation of an unpredictable one-way function for the Diffie-Hellman algorithm.

1. Problem Setting

Key points:

1. It is necessary to create a method that allows you to determine new passwords, as often as needed, using only the open channel of information exchange.

2. Before the algorithm starts working, there is at least one secret exchange of information about the type, properties, parameters of a one-way function and the algorithm of agent interaction when determining a new password.

3. It is reasonable to take the Diffie-Hellman algorithm as the basic one and modify it.

Definition of the operation procedure of a one-way function on the basis of cellular automats.

First, we decompose the task into 2 components:

Methods of transforming the result of work of the cellular automata into something that can be perceived as the result of a one-way function (hashing).

Modification of cellular automata properties, for variety of variants of the function behaviour.

Properties of the function:

- Unilaterality. That is, the possibility of obtaining the function value based on information about the argument \( y = f(x) \) and at the same time, the impossibility to obtain the argument value based on the function value, to be more precise, the absence of the inverse function, would provide the argument finding for the given function value \( x = f^{-1}(y) \).

- Commutative. If we denote the action of the function by the sign "\( \times \)", the property of the group operations commutation means: \( x1 \times x2 = x2 \times x1 \).

On the example of the function, it can be:

\[
\begin{align*}
f(x1 \times x2) &= f(x2 \times x1), \\
f(x1 + x2) &= f(x2 + x1). \\
f(x1, x2) &= f(x2, x1)
\end{align*}
\]

The variety of parameters and behavioural variants of the one-sided function on the basis of cellular automata should extend the set of rules of the classic "Game of Life."
2. General idea to improve the Diffie-Hellman algorithm using a one-way function based on cellular automata

The input field for the cellular automata can be transmitted in encrypted form by agents during a secret communication session or by steganographic methods. For example, the picture from which a cell field is then formed can be on some website; during a secret meeting, the agents must determine where to take the picture and by what rules the input cell field is formed. After that, the agents find the picture by following some link in the Internet and form identical cell fields.

In some cases, it is even possible not to hide, what exactly the picture is used as the input one, because in this case it will be almost impossible to determine the reverse function by the results of intermediate functions.

The result of the cellular automata is what picture of division of cells that contain life (1) and that do not contain life (0). As a result of the improvement of the "Game of Life" rules, it is expected that each cell may contain not only binary information, but also code numbers describing the state of the cell in a wider range of values. But this does not change anything fundamentally about the further use of these results. Intermediate pictures of the agents are sent to each other as is. The final result in the form of a "Game of Life" picture can be used to create a common code either directly (as all values of the game field) or as a hash function.

The hash function. If it is decided to use a hash function, it is proposed to find a certain hash function on the basis of the values located in individual cells, which converts the result of the picture of “Game of Life” into a standardized set of numbers, which are proposed to obtain the following methods.

Arithmetic and logical operations performed by certain rules for numbers that are stored in final field: in separate rows or columns, in individual cells, selected by template in a pseudo-random order, in certain areas of the painting that may or may not intersect.

In more serious conditions, a hash function can be created based on the current Ukrainian standards. In our case, the construction of a hash function is not the main subject of research, so for the transparency of the results as a test hash function were used sums of values in the columns of the matrix of the final cell field. But during numerical testing, even such a simple hash function for the matrix of picture 89x89 did not give a single repetition of values in 40 000 steps of the “Game of Life.”

3. Improving the rules of birth and death process

As a basis we take the cellular machine “Game of Life,” which uses only constant parameters of movement and development. We improve the rules of birth rate and life extension. To calculate the number of neighbours, use Moore's rule, that is, each cell will have 8 neighbours.

Each cell can take values of either zero or one. Therefore, the rules concerning the life and death conditions of the cell at stage $t + 1$ will be presented as
a Boolean function of nine variables, where the first change corresponds to the state of the selected cell, and the remaining 8 are the states of neighbouring cells. The cell becomes alive if the expression is one:

\[
s_{t+1} = (1 - s_t) \land (d_l \leq n) \land (n \leq d_r) \lor s_t \land (a_l \leq n) \land (n \leq a_r),
\]

where \(n\) is the number of neighboring cells,
\(s_t\) - cell state at iteration \(t\),
\(d_l\) is the minimum number of neighbors for the birth of life in an empty cell,
\(d_r\) is the maximum number of neighbors for the birth of life in an empty cell,
\(a_l\) is the minimum number of neighbors to prolong a cell's life,
\(a_r\) is the maximum number of neighbors to prolong a cell's life.

When the minimum number of neighbours for birth of life in an empty cell is equal to the minimum number of neighbours for life extension of a cell and the maximum number of neighbours for birth of life in an empty cell is equal to the maximum number of neighbours for life extension of a cell \(d_l = a_l\) \& \(d_r = a_r\), formula (1) can be reduced:

\[
s_{t+1} = (d_l \leq n) \land (n \leq d_r)
\]

Attempts to program the given dependencies have unexpectedly shown additional advantages of simulation (computer) modelling. So, in notations of MatLab and Python expression (1) can be presented more concisely, than Boolean function:

\[
s_{t+1} = (1 - s_t) \ast (d_l \leq n \leq d_r) + s_t \ast (a_l \leq n \leq a_r)
\]

Now the classic rules for "The Game of Life" can be represented as follows:

\[
s_{t+1} = (1 - s_t) \ast (3 \leq n \leq 3) + s_t \ast (2 \leq n \leq 3)
\]

Workability of dependencies (1) has been checked on an example of modelling “Games of Life” in which parameters have been changed \(d_l = 1\), \(d_r = 7\), \(a_l = 1\), \(a_r = 8\) (Fig. 1):

![Game Of Life Step 7](image)

**Figure 1:** Iteration in the modified "Game of Life".
The initial position of the live points in the field is randomly selected using a pseudo-random number generator, or, as mentioned above, the initial field can also be a piece of the originally selected picture.

4. Introduction of an intra-population interaction radius rule

The radius of intra-population interaction was also introduced to complicate the model behavior and increase the number of possible parameter combinations. This parameter is responsible for which cells around the selected cell influence its state in the next iteration. In order to find the number of neighbouring cells affecting the cell at an arbitrary radius, the formula was derived:

\[ n = (2R + 1)^2 - 1, \]

where \( R \) is the radius of intra-population interaction.

The viability of a cellular automaton with a variable radius of interaction was tested using the MatLab programming language on the example of modelling a previously used model with the following parameters \( d_l = 1, d_r = 7, a_l = 1, a_r = 8 \) and with a radius of interaction \( R = 2 \) (Fig. 2).

Figure 2: "Game of Life" with a modified radius R=2.

5. Introducing the rule of death from old age

In the real world, any creature can't live forever. On the one hand, it may be related to the old age of the creature, on the other hand, it may be related to the exhaustiveness of resources located in each cell of the field. In our model we introduce the rule of death from static that will work the fuse from static in the model. To implement the new model, we propose the following rules:

- with each iteration, the age of the cell increases by one,
- if a cell is older than the max_age specified, it dies.

The model of age-related death rule (creation or news) is implemented as follows (Fig. 3). It is important to say that the empty field is black, the older the living cell is, the brighter it is.
6. Improvement of rules for the implementation of a multi-population model

One of the most famous multi-population models is the "predator prey" model. Above we have noted the imperfections of the model, which are associated with ignoring the radius of interaction of different creatures and the closed space.

To solve this problem, we first split the model into two: the so-called "rabbits" live in the first (victims), the second one is "wolves" (predators). Each population has its own separate independent survival parameters: radius of interaction $R$, parameters of birth $d_l, d_r$ and life extension $a_l, a_r$. This allows more flexibility in model management.

Also, the model has the rules of interaction of the two populations directly:

Rule 1. The wolf eats the rabbit. If a wolf and a rabbit get to the same cell, only the wolf remains in this cell (the rabbit disappears - it is eaten). Consider that rabbits have an interaction radius of $R = 1$ and wolves $R = 1$, although these parameters can change in any value depending on the type of creation or other circumstances. This is necessary to correct the model to a truly realistic state.

Rule 2. The life of every creature in the world depends, one way or another, on the availability of food in its interaction radius. Therefore, let us assume that if the wolf does not have the right amount of rabbits in its interaction radius, it dies from "hunger". The amount of necessary food in this rule corresponds to the parameter $food_{min}$.

A similar rule can be entered for rabbits, but for the unborn picture we enter the condition that there is always enough food in the rabbit population. This rule is necessary in order to partially equalize the forces of rabbits and wolves. The implementation of the model with two rules working (eating rabbits by wolves and starving) to compare with the implementation with only the first rule will work with the same population parameters as before (Fig. 4). In a model with the entered rabbit eating rule, the field on which the model is implemented is white, the predator population cells are red, and the victim cells are blue.
7. Testing a modified Diffie-Hellman Algorithm

The algorithm was tested in the integrated MATLAB environment. As a one-way function, the cellular automata "Game of Life" was used on the 89x89 field with such extended rules: the rule of birth rate and life extension, closed space rule, the rule of the radius of intra-population interaction.

On the basis of a random image from open ordinary graphic pictures (Fig. 5), the agents have formed one for themselves on all primary field "Games of Life" (Fig. 6).

Figure 4: "Game of Life" with starving wolves and eating rabbits by wolves.

Figure 5: Picture for generating the input field of "Game of Life".

Figure 6: Generated input field of "Game of Life" size 89x89.
After that each of the agents invented his secret word (3728 and 5307), which corresponds to the number of steps in the game, found his intermediate picture and sent it to the second agent (Fig. 7 a, b). The code words corresponded to the number of steps the program had to pass from the received picture. In the first agent it was 3728 + 5307 = 9035 steps. In the other 5307 + 3728 = 9035 steps. So, both agents in different ways received a single code picture 9035 (Fig. 7 c).

![Figure 7: Pictures of "Games of Life" 89x89 in the process of applying secret words. a) intermediate picture to agent 1; b) intermediate picture to agent 2; c) final picture of both agents.]

Then the resulting code picture can be used depending on the established requirements for protection directly in binary form, directly in hexadecimal form or as a hash (Fig. 8).

![Figure 8: The general code based on the hash picture.]

242
For the test was used a regular household computer Dell Inspiron, Intel Core 3 Gen.8 Processor with a clock speed of 2.1 - 2.3 MHz, RAM 8 Gb, SSD 128 Gb. Test results: 40000 steps of the "Game of Life" for picture 89x89 was counted in 25 minutes 30 seconds.

At the same time, the repeatability of the pictures was tested for time consumption. Since each picture needed to be compared with each of the 32,000 other pictures, the comparison of pictures was replaced by a comparison of hashes of these pictures to save memory. In this case there is a danger that some different pictures will have the same hash. But this is corrected by additional checking of pictures with the same hash. But in several implementations among 40000 images, no two had the same hash, so no two images had the same hash. In fact, the number of steps "Game of Life" was limited by the memory capacity of the computer. In this case there is a danger that some different pictures will have the same hash. But this is corrected by additional checking of pictures with the same hash. But in several implementations among 40000 images, no two had the same hash, so no two images had the same hash. In fact, the number of steps "Game of Life" was limited by the memory capacity of the computer. If we reduce the size of the picture, the memory limitation will not be so influential.

Conclusions

1. The work has an improved Diffie-Hellman algorithm by creating a one-way function on the basis of cellular automata with an extended set of rules.

2. The author extended the rules of the "life game" for the first time in the directions of defining the rules of birth rate and life extension, controlling the radius of intra-population interaction, the rules of death from the age of cells, and the multi-component (multi-population) system of cells.

3. The created algorithm on the basis of a cellular automaton can be used to create keys for safe information transfer.

4. Depending on users' needs, code security can be increased by using variable parameters of the cell field and cell behavior, which also allows to regulate the encryption speed.

5. Since the development of the hash function is not a research goal, but only one of the tools, the sum of the values in the columns of the "Game of Life" matrix has been used as a hash function for the transparency of the results. But even such a simple hash function during the numerical testing did not give a single repetition of values for 40 000 steps of the "Game of Life".

6. The direction of further research is to improve the "Game of Life" model and create methods for effective selection of parameters of one-way functions.

Acknowledgements

This work is supported by the National Science Program “Information and Communication Technologies for Unified Digital Market in Science, Education and
Security.” Part of the research equipment was supported by PPNIP-2020-03/09.03.2020 “Extraction, processing and analysis of parametric data from external devices” and company FADATA.

References


Modified Cellular Automata in the Diffie-Hellman Algorithm for Data Exchange

About the Authors

Volodymyr V. Shevchenko is a bachelor student in the Software System and Technologies Department of Taras Shevchenko National University of Kyiv. Research interests: computer science, mathematical modelling, software development, software development technologies, information security.


Denis S. Berestov, PhD, Assistant, Department of Software Systems and Technology, Faculty of Information Technology Taras Shevchenko National University of Kyiv. Scientific interests: methods and tools data science, embedded system, industrial internet of things, ERP systems, cybersecurity, software and information security for national defence. Scopus Author ID: 57195138493.


Igor P. Sinitcyn, Dr.Sc., Senior Researcher, Head of Department at the Institute of Software Systems of the National Academy of Sciences of Ukraine. Chief Designer of a number of Information-analytical systems for managing defence resources of the Armed Forces of Ukraine. Research interests: computer science, mathematical modelling, software development, operations research, defence planning, defence resources management, program-targeted approach, software development technologies, e-government.


Ivan Garvanov, Professor, DSc, Vice-Rector and a Head of Department “Information Systems and Technologies” in the University of Library Studies and Information Technologies (ULSIT), Sofia, Bulgaria. He is the author/co-author of over 250 scientific publications. His research interests involve information and communication technologies, radar signal and data processing, image processing, and navigation systems.

Iva Kostadinova, PhD, Chief assist. prof. at the University of Library Study and Information Technologies. She is a lecturer of the Department “Information Systems and Technology,” Faculty of Information Sciences at ULSIT. Research interests in the fields of e-learning, distance learning, on-line communications, ICT in education, educational technology.