A New Algorithm of Automatic Complex Password Generator Employing Genetic Algorithm

Sura Jasim Mohammed
Information Technology College, University of Babylon
programmer_sura@yahoo.com

Abstract

Due to the occurred increasing in information sharing, internet popularization, E-commerce transactions, and data transferring, security and authenticity become an important and necessary subject. In this paper an automated schema was proposed to generate a strong and complex password which is based on entering initial data such as text (meaningful and simple information or not), with the concept of encoding it, then employing the Genetic Algorithm by using its operations crossover and mutation to generated different data from the entered one. The generated password is non-guessable and can be used in many and different applications and internet services like social networks, secured system, distributed systems, and online services. The proposed password generator achieved diffusion, randomness, and confusions, which are very necessary, required and targeted in the resulted password, in addition to the notice that the length of the generated password differs from the length of initial data, and any simple changing and modification in the initial data produces more and clear modification in the generated password. The proposed work was done using visual basic programing language.

Keywords*:* Security, Authentication, Genetic Algorithm, Password Generator, Crossover, Mutation.

الخلاصة

نتيجة للتزايد الحاصل بمشاركة المعلومات, عمومية الانترنت, معاملات التجارة الإلكترونية, وتناقل البيانات, لذا اصبحت الامنية والموثوقية موضوع مهم وضروري. في هذا البحث تم اقتراح خطة الية لتوليد كلمة سر قوية ومعقدة بالاعتماد على ادخال بيانات اولية مثلا نص (معلومات ذات معنى وبسيطة او لا), مع مفهوم تشفيرها, ثم توظيف الخوارزمية الجينية وذلك باستخدام عملياتها الارتباط والطفرة لتوليد بيانات مختلفة عن الاخرى المدخلة. ان كلمة السر المتولدة لا يمكن تخمينها وممكن استخدامها بعدة ومختلف الخدمات وتطبيقات الانترنت مثل الشبكات الاجتماعية, الانظمة المؤمنة, الانظمة الموزعة, وخدمات عبر الانترنت. ان مولد كلمة السر المقترح يحقق الانتشار، العشوائية، والارتباك, والذين هم ضروريين ومطلوبين ومستهدفين بكلمة السر الناتجة, اضافة الى ملاحظة ان كلمة السر المتولدة يختلف عن طول البيانات الاولية, واي تغيير وتحديث بسيط في البيانات الاولية ينتج تحديث كبير وواضح بكلمة السر المتولدة. انجز العمل المقترح باستخدام لغة البرمجة فيجوال بيسك.

الكلمات المفتاحية: امنية، توثيق، الخوارزمية الجينية، مولد كلمات السر ، طفرة.

1- Introduction

During recent years and yet now the terms of security, data integrity, authenticity are very concern, whereas in order to authenticate users of any service that is worked online, the authentication of password is used [Mathew *et al.*, 2013]. There are several attempts for replacing authentication of a password, like basing on biometrics (eye, fingerprint, palm print, etc.), authentication using multi factor, and tokens. One schemas of data security ensuring is encryption operation for converting the data from its original form to non-cleared one based on a secret key [Stallings & W., 2014]. There are three categories of user authentication techniques, knowledge base, object base, biometric base, the first class use the username and password which they are textual and can be stolen or forgotten [Bhanushali *et al.*, 2015], the second one based on additional something of someone in addition to the textual information [Robert *et al.,* 2011], while the third class is based on human's biometrics like iris, palm, vein palm, signature, face, etc. [Bhanushali *et al.,* 2015].

One of the search algorithms that its search is adaptive heuristic is Genetic Algorithms (GA), the evolutionary concepts of genetics and natural selection is based in GA [Mathews, 2010]. Genetic Algorithms was conducted by John Holland, which is applied in optimization problems, searching and finding solutions [Poornima & Girish, 2015].

Genetic Algorithms contains the following operations [Gondro & Kinghorn, 2007]:

- Generating a random population.
- Selection two parents.
- Crossover.
- Mutation.
- Replacement.

The contents of chromosome's genes are coding in a certain way that is suitable with the values of the solutions, so there are many coding methods such as binary, real, and decimal coding. For selection operation, there are several strategies like randomly binary selection or triple selection, the fitness value which is attached with each chromosome is based in selection operation. In crossover operation, there are three types, they are single point, two points multi points based crossover. The strategy of mutation aims to alter in the contents of the children chromosomes, so there are many mutation types [Mathews, 2010].

There were two main categories of issues that are associated with text password usage, they are:

1. User related issues.
2. Online Service related issues.

The first class includes the following notes [Fatma & Chris, 2017]:

- Users need many passwords for internet services, this may lead multiple accounts based on the usage of the same password.
- Password that was chosen is often guessable or has a meaning, such as user's birthday, name, etc.
- When the user forced to change its password, he does often make minimal changes to his current password, like by adding a serial no.

The second class includes the following [Eric et. al., 2009]:

- Enforcing with the policy of complex password, this state is occurred in many sites, for example few character's number is required in the contents of the password, or there is ability or disability including a special characters.
- Some cites of internet services force the user to modify his password in regular times.

2. Related Works

For password generation, there were many works and attempts to generate complex password, the inherent nature in GA provided desired disorder, one application of the data communication that is secure was achieved using chaos theory with entropy [Mohammad et. al., 2012]. Many attempts of password authentication were analyzed in [Cormac & Paul, 2012] using tokens, biometrics, and authentication based on multi factor. In [Poornima & Girish, 2014] the researchers attempted to employing the randomness of crossover and mutation in barcode generation that is based for authentication process. The issues that are related with the text password usage were listed and cleared, they were classified as user or online service related [Dinei et. al., 2014]. In [Poornima & Girish, 2015] data authentication was proposed, it was distributed among many servers of authentication, a distributed three dimensional password (text, graphic, barcode) based authentication were used. In [Fatma & Chris, 2017] an autopass was generated for addressing the issues of forced password modifications, pre specific password, data of user or sever specific configuration. In the proposed schema of password generator, GA is used due to the randomness of its two operations (crossover and mutation), the produced password has the properties of diffusion, randomness, and confusions.

3. Designing

In this paper an automated schema for password generating using the concept of Genetic Algorithm, the initial data that was entered firstly was text as implemented in the experiments, but it also can be an image or registered audio. In the used scenario the password is generated based on entering initial text that is a set of symbols, it can be clear information about the user such as the surname. Figure (1) indicates the block diagram of the proposed scenario.

Firstly, the initial text is entered, it may be meaningful or not, it is spliced into two parts by putting the first symbol of the initial text in the first part and the last symbol in the second part, then reverse the strategy of splitting by putting the second symbol of the initial text in the second part and the symbol that is previous to the last one in the first part, and so on. If the length of initial text is even, this means that it can be spliced into two equals parts, but if the number of its symbols is odd, a random character must be added to a certain part in order to make two parts with the same length. Note that the space symbol is not canceled.

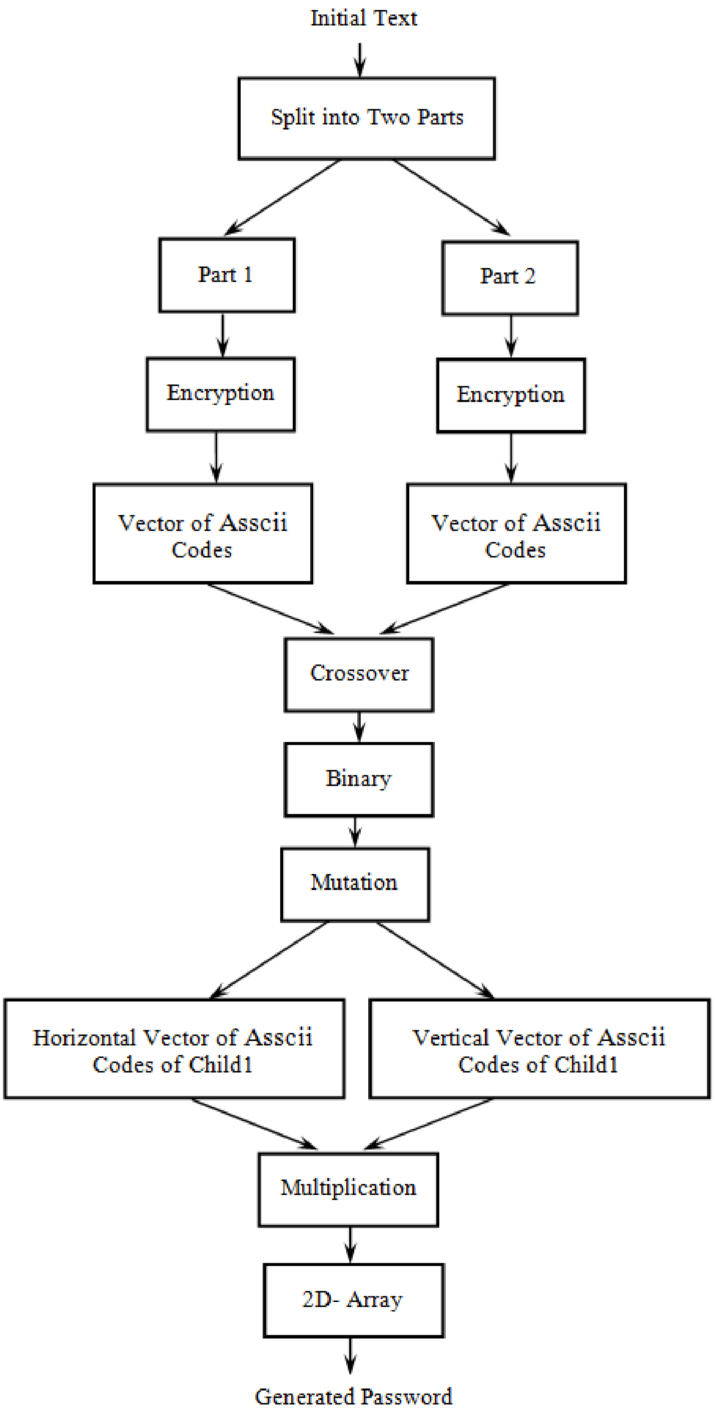

Figure (1) Text Based Password Generation.

Now each part is encrypted separately using one of the encryption methods, after that each symbol of those two encrypted parts is converted to its corresponding number based on the ASSCII Code Table. Now the two operations of genetic algorithm (crossover and mutation) are employed, two types of crossover operation were applied to the numeric codes of the two encrypted parts, they are 2points and multiple points, one type of them is selected based on generating a random number and finding the mod result by 2, if it is 0 then choosing 2points crossover, otherwise choosing multiple points crossover. Figure 2 shows example of crossover types.

Then executing the mutation operation of the genetic algorithm, but now the contents of the two vectors of the resulted children are numeric, so if the mutation is done by swapping between two genes of the chromosomes of two parts, the resulted modification is very little. Thus before applying the mutation operation, each parent chromosome is converted to the binary form, and then in the mutation operation each bit has 0 is swapped with 1, and each bit has 1 is swapped with 0, this strategy ensure more modification in the genes than those of the children that are resulted from crossover operation.

Now the content of the two chromosomes is binary, each gene of them is converted to the corresponding number depending on the ASSCII Code Table. The first chromosome is represented in a horizontal vector, while the second one is represented in a vertical vector, then multiply those two vectors to get a two dimensional array, that will be converted to a 1D-vector by putting each row besides one, finally the elements of this vector are converted to their corresponding symbols based on the ASSCII Code Table to result the generated complex and strong password.

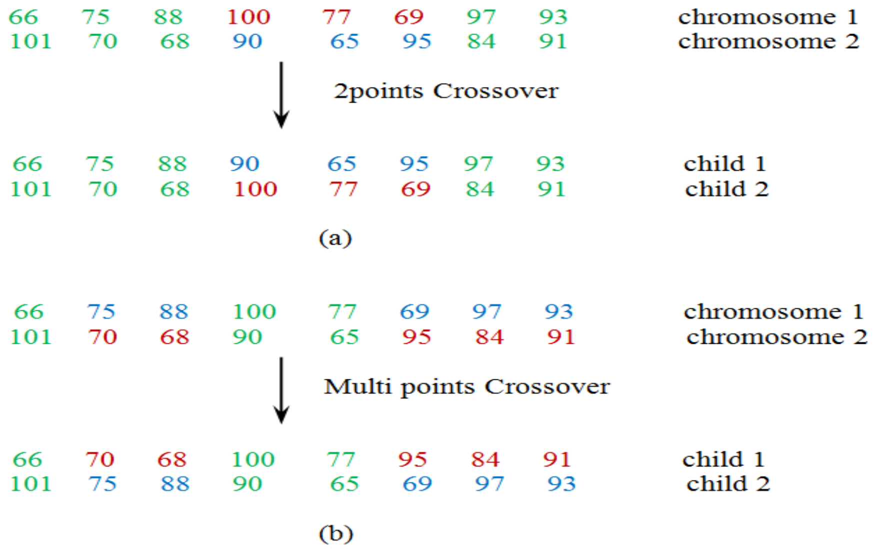
66 75 88 100 77 69 97 93 chromosome 1
101 70 68 90 65 95 84 91 chromosome 2
2points Crossover
66 75 88 90 65 95 97 93 child 1
101 70 68 100 77 69 84 91 child 2
(a)
66 75 88 100 77 69 97 93 chromosome 1
101 70 68 90 65 95 84 91 chromosome 2
Multi points Crossover
66 70 68 100 77 95 84 91 child 1
101 75 88 90 65 69 97 93 child 2
(b)

Figure (2) Crossover Types.
(a): Two Points. (b): Multipoint.

4. Experimental Results

The proposed password generation scenario was implemented as follows:
In the first step, the initial text is entered, which is divided into two parts, and encrypt every part, the RSA encryption method was used here, then translating the symbols of each encrypted part to their corresponding numbers, which are implement the ASSCII codes.

The next step is applying the genetic algorithm spatially the main two operations (crossover and mutation), one type of crossover is selected between 2points and multi points types. Then translated the contents of each resulted chromosome of crossover into the binary form, to apply the mutation operation on all their genes. The finally step is multiplying the two vectors of children after converting them from binary to decimal form and represent one of them in a vector horizontally and the other vertically, and a 2D array will be get. The rows of the resulted array are transferred in a 1D vector, and its numerical element will be converted to their corresponding symbols, that represent the generated password.

For example if the initial text is "DEPT", it is divided into two parts, part1 is "DP" and part2 is "ET", after applying the RSA encoding method to these two parts, the result of RSA is 87, 19 for part1, and 62, 28 for part2. Now in the crossover the resulted two parts become 87, 28 and 62, 19. Before do mutation in the two new parts, they are converted to the binary form as 01011111, 00011100 for part1 and 00111110, 00010011 for part2, in the mutation operation they become as 10100000, 11100011 and 11000001, 11101100 respectively, they are 160, 227 and 193, 236 in decimal form. Because the ASSCII table has 128 entries, the result of mod operation by 128 is considered, so the result become 32, 99 (horizontal vector) and 65, 108 (vertical vector) respectively. Now multiply the two vectors, and the result is 2D array which will be converted into 1D vector, it will be as 2080, 3456, 6435, and 10692, after compute the mod by 128, the result is 32, 0, 35, and 68. Because the entries (0 .. 31) of ASSCII table are symbols of controlling the keyboard and sended and stored information, add 32 to the result in the range (0 .. 31), so the result is 32, 32, 35, 68, which represents ␣␣#D, it is the generated password, whereas ␣ is space. In this example notice that the length of generated password is equal to that's of initial text, only in this case when the length is 4 symbols.

Table-1 shows samples of experimental result of the proposed password generation schema for different length of the initial text. In [Poornima & Girish, 2014] the length of generated barcode is equal to the initial generated digit random number, while the proposed schema generates different length of password than its of initial text. In [[Mohammad et. al., 2012],[Poornima & Girish, 2015] one type of crossover was based, while in the proposed system the crossover type is chosen dynamically without the aid of the user, so that the type of recombination influences the result of generated password. Many factors can make more modification in the generated password even if only one of them is occurred, these factors are: the prime parameters of RSA encoder, the used type of crossover, the length of initial text, the random symbols that was appended to the initial text in the case of odd length, the strategy of splitting the initial text to the two parts, randomness of GA, completely mutation operation, all these factors lead to generate strong and complex password that is undetectable and it achieves diffusion, randomness, and confusions.

Table 1. Samples of the Experiments Results

Initial Data	Generated Password
Software	2A!658.?"/ht7d+R
Car 18	O5yw9;ta*
Info. Tech.	X@7g>|kH/49-:sD\%yr)lC)'f8{rW].#b^P3
1st Ex.	8,0eGln4|5&v*'7m

5. Discussion and Conclusions

In order to utility of randomness of genetic algorithm that is founded in crossover and mutation operations, GA was employed in this paper for proposing a password generator, it achieves the requirements of a produced password, they are diffusion, randomness, and confusions. Whereas the generator is based on entering the initial text from the user, so nay simple change in it causes wide modification in the generated password, also the same initial text can leads to different generated passwords when changing in the prime numbers of RSA parameters. In addition to the fact that the initial data can be clear and meaningful, while the produced password is non- guessable and surely has no any meaning. Also the number of symbols that were resulted by the generator is different from it's of the initial text. Encoding the initial text before applying the genetic algorithm instead of working with original initial text achieves the security concept. The generated password cannot be remembered, so that the application of such password is as a master password, which is saved in a file after it was generated and it is used in user specific configuration data for access the server, also it can be distributed among the users who share information and data, it can also be used for barcode generation for commercial field. The proposed system belongs to the first class of user authentication (knowledge base), but the idea of this paper is that the user does not need to operate the proposed algorithm in each logging time, the password is generated when the user decides to change it, and he does not need to remember it, but it is stored in a file, so that the logging time is not influenced.

6. References

Bhanushali, A. B. Mange, H. Vyas, H. Bhanushali, and P. Bhogle, 2015, Comparision of Graphical Password Authentication Techniques, International Journal of Computer Application, Vol. 116, No. 1, pp. 11-14, 2015.

Cormac Herley, and Paul C. Van Oorschot, 2012, A Research Agenda Acknowledging the Persistence of Passwords, IEEE Security & Privacy, Vol. 10, No. 1, pp. 28-36, 2012.

Dinei Florencio, Cormac Herley, and Paul C. Van Oorschot, 2014, Password Portfolios and the Finite-Effort User: Sustainably Managing Large Numbers of Accounts, In Proceedings of the 23rd USENIX Security Symposium, San Diego, CA, USA, pp. 575-590, 2014.

Eric Cole, Ronald Krutz, and James W. Conley, 2009, Network Security, Indiana, USA: Wiley, 2009.

Fatma Al Maqbali, and Chris J. Mitchell, 2017, AutoPass: An Automatic Password Generator, arXiv:1703.01959v2 [cs. CR], pp. 1-20, 2017.

Gondro C., Kinghorn BP, 2007, A Simple Genetic Algorithm Multiple Sequence Alignment, Genetics and Molecular Research, Vol. 6, No. 4, pp. 964-982, 2007.

Mathews J., 2010, The Use of Genetic Algorithm in Cryptanalysis, Cryptologia, Vol. 17, pp. 187-201, 2010.

Mohammad Sazzadual Hoque, Abdul Mukit, and Abu Naser Bikas, 2012, An Implementation of Intrusion Detection System Using Genetic Algorithm, International Journal of Network Security & Its Applications (IJNSA), Vol. 4, No. 2, 2012.

Mathew, G., and S. Thomas, 2013, A Novel Multifactor Authentication System Ensuring Usability and Security, Preprint, Vol. 2, pp. 21-30, 2013.

Poornima G. Naik, and Girish R. Naik, 2014, Secure Barcode Authentication Using Genetic Algorithm, IOSR Journal of Computer Engineering (IOSR-JCE), Vol. 16, No. 2, pp. 134-142, 2014.

Poornima G. Naik, and Girish R. Naik, 2015, A Framework for Secure 3D Password Using Genetic Algorithm, International Journal of Advanced Research in Computer Science and Management Studies, Vol. 3, No. 1, 2015.

Robert Biddle, Mohammad Mannan, Paul C. van Oorschot, and Tara Whalen, 2011, User Study, Analysis, and Usable Security of Password Based on Digital Objects, IEEE Transactions on Information Forensics and Security, Vol. 6, No. 3, pp. 970-979, 2011.

Stallings, W., 2014, Cryptography and Network Security Principles and Practice (5 th ed.), Boston: Pearson, 2014.