A SURVEY ON GENETIC ALGORITHM BASED CRYPTOGRAPHY

[1]Marilyn Fanyo, [2]Tanuja Kumari Sharma
[1]Post Graduate Student, [2]Assistant Professor
[1]Departement of Computer Science and Engineering,
[1]AP Goyal Shimla University, Shimla (HP), India

Abstract: Since the beginning of the twentieth century, data security has become a very important subject for safe communication and correspondence. Starting from military and diplomatic level security, encryption has reached several areas and applications requiring data transmission, among which: banking, industry, e-commerce, video broadcasting, computer networks and personal computers, telecommunications and so forth. In all activities data has taken a very central role. With that, rises the need of having it protected in the most secure manner, to ensure it is used only by rightful users with rightful intents. However as technology evolves, security threats also increase and become more sophisticated. Therefore, traditional cryptography schemes taken alone are less suitable to ensure high levels of security. There is need to use evolutionary computing method that can mitigate the vulnerability of data. Genetic algorithms (GA) are among such modern techniques that derive from artificial intelligence and can be used to solve difficult problems. This paper is a survey report on various genetic-algorithm-based cryptography schemes proposed throughout studies for efficient encryption of sensitive data.

Keywords **- Data security, Encryption, Genetic Algorithm, Crossover, Mutation, Genetic Inspired Cryptography.**

I. INTRODUCTION

For the past decades information sharing and transfer has exponentially increased. Data has become the most crucial resource for any organization. Safe and secure transfer and storage of data is must in all domains ranging from two people's conversation to a government information base, or cloud data storage. In the information technology (IT) world, data that is transmitted through communication media can be stolen, corrupted or lost. Therefore, there has to be means of protection of information that are transiting from a point A to a point B, to avoid any tampering phenomena [1]. Especially in the Internet era, the need for a security model that will provide data integrity, confidentiality and authenticity, is increased. To achieve that goal, the most used technique is cryptography also referred to as data encryption. It is a technique of encoding and decoding messages so that they cannot be interpreted by anybody except the sender and the intended recipient. There are several encryption methods which can be classified into symmetric and asymmetric cryptography methods. Traditional cryptographic symmetric algorithms are suitable for encrypting data to store it in a secure way. Asymmetric algorithms are more secured but they are preferred for encrypting the keys rather than the data because of their less speed. With the constant evolution of technology, the impact of security threats also kept increasing. Single traditional cryptography schemes are not reliable any more suitable to guarantee high levels of security. This is why researchers for the past few years have been studying evolutionary computing methods that can mitigate the vulnerability of data. Among such modern techniques, Genetic algorithms (GA) are some of the most recommended. This technique derives from artificial intelligence and can be used to solve difficult problems [2].

1.1 Cryptography

The use of encryption/decryption is as old as the art of communication. As information has always been related to power, the need to protect important information for wrongful access and use has always been existent. In previous time (at war time) cipher was called wrongly as a code, which was employed to prevent the enemy from getting the secret information that transmitted over communication systems [3]. Cryptography is the art of protecting information by transforming it into an unreadable format called ciphertext [4]. The purpose of cryptography is to protect transmitted information from being read and understood by anyone except the intended recipient. What this means is that, when unauthorized individuals happen to get access to encrypted message, they should not have any mean of deciphering it. In modern days, cryptography algorithms can be classified into three categories: Symmetric-key algorithms, Asymmetric-key algorithms and Hashing. Key cryptography schemes help to achieve Confidentiality while hashing ensures Integrity of data [5]. Decryption is the opposite process of encryption, performed to restore the original plain data. It should only be done by intended users.

1.1.1 Asymmetric encryption

Also known as public key cryptography, asymmetric encryption uses a pair of keys: Private Key and Public Key. One is used for encrypting and the other one for decrypting. In most cases, the public key is used by sender for encryption and the private is used for decryption. Asymmetric algorithms are more secure as they always maintain one key secret and the other shared. But they run slower than symmetric algorithms, which makes them more suitable for authentication or key exchange. The most well-known asymmetric-key algorithms, especially for cloud are: RSA (Rivest-Shamir-Adleman), Diffie-Helman Key Exchange, and Elliptic Curve Cryptography [5].

1.1.2 Symmetric Encryption

Also known as secret key cryptography, this type of encryption model involves the use of a single secret key, shared between sender and receiver. This single key will be used for both encryption and decryption [6]. The advantage of symmetric algorithms is they are too computing power-consuming, and work with high speed in encryption. It is the reason why they are preferred for data encryption, and most importantly, for processing large streams of data. Symmetric-key algorithms can be furthered divided into two categories: Block cipher and Stream cipher. In block cipher, input is taken as a block of plaintext of fixed size. Key of fixed size is applied on block of plain text and then the output block of the same size as the block of plaintext is obtained. In case of stream cipher, the plaintext is encrypted one bit at a time. Some popular Symmetric-key algorithms includes: Data Encryption

Standard (DES), Triple-DES, Advanced Encryption Standard (AES), Blowfish algorithm, International Data Encryption Algorithm (IDEA).

1.2 Genetic Algorithm

Genetic algorithms belong to the family of evolutionary algorithms along with genetic programming, evolution strategies and evolutionary programming. They are based on the notions of natural selection and natural genetics [7]. Genetic algorithms have proven their powerful and reliable optimization technique in a wide range of applications such as bioinformatics, cryptography, computational science, engineering, economics, game strategies, chess problems, manufacturing, mathematics, physics, and other fields [8][3].

Genetic Algorithms try to follow nature to a great extent by mimicking its randomness. Taking inspiration from the biological evolution, GA tries to find from an initial population the best solution which is the one that matches with the objectives. GA is nowadays a popular method for avoiding local optima in improving search [9].

1.2.1 Genetic Algorithm operations

Genetic Algorithms contain operations which are bio-inspired such as selection, crossover and mutation. They are used to optimize search problems by generating high-quality solutions [10]. A Genetic Algorithm consists of five basic operations: Initial Population generation, Selection, Fitness function, Crossover and Mutation [11] [12].

1) Initial Population Generation

This is the starting step where a set of individuals is randomly created [12]. In the context of encryption, these individuals can be a set of random numbers for further generation of a strong key. In application, a Pseudorandom Number Generator (PNRG) could be used for generating a sequence of numbers. Some PRNG files are Blum Blum Shub, Wichmann-Hill, Complementary-multiply-with-carry, Inversive congruential generator, ISAAC (cipher), Lagged Fibonacci generator, Multiply-with-carry, Naor-Reingold Pseudorandom Function, Park–Miller random number generator [13].

2) Fitness function

Fitness is a very important function of Genetic Algorithm because when it is good, the fitness function helps explore more accurately the search space. On other hand, bad fitness functions are confined to local optimum solution. We can classify fitness functions as Constant fitness function and Mutable fitness function. The fitness function is an objective function which determines whether or not a chromosome is suitable or close enough to an anticipated value [12].

3) Selection

This operator picks from the population specified individuals to be the parents, which will be used for reproduction. The selection is based on fitness value, and higher the value, more is the chance to be selected [14] [3]. Some selection methods are Roulette-wheel Selection, Tournament Selection, and Truncation Selection.

4) Crossover

This operator has the same significance as that of crossover in natural genetic process. It assists in linking two parents from the selected parents to generate new chromosome(s). The crossover can be used as encryption technique by being applied directly on plaintext. It could also be applied on the set of numbers that will help produce key [12]. We have three forms of crossover including single point crossover, two-point crossover and uniform crossover.

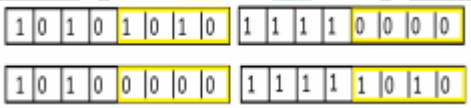

Fig.1 Single Point Crossover

Fig.2 Two Point Crossover

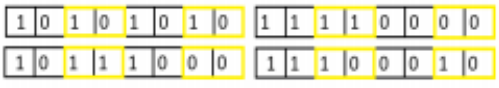

Fig.3 Multipoint crossover

5) Mutation

The last operation of genetic algorithm is mutation. It is similar to biological mutation and is used to maintain diversity from one generation of population to the next. It alters one or more gene value from a chromosome from its original sate. For example, flipping of bits can be performed, where '0' mutates to '1' and vice versa. In encryption problems, mutation is employed for inducing disorder into the vector [11].

Figure 4 describes the flowchart of genetic algorithms with respect to their main operations.

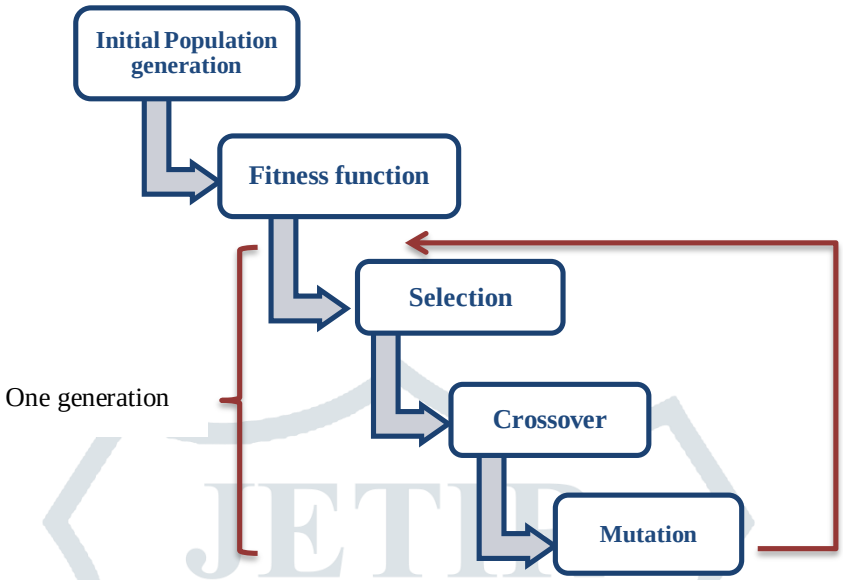

Fig.4 Flowchart of Genetic Algorithm

1.2.2 Structure of basic genetic algorithm [8]

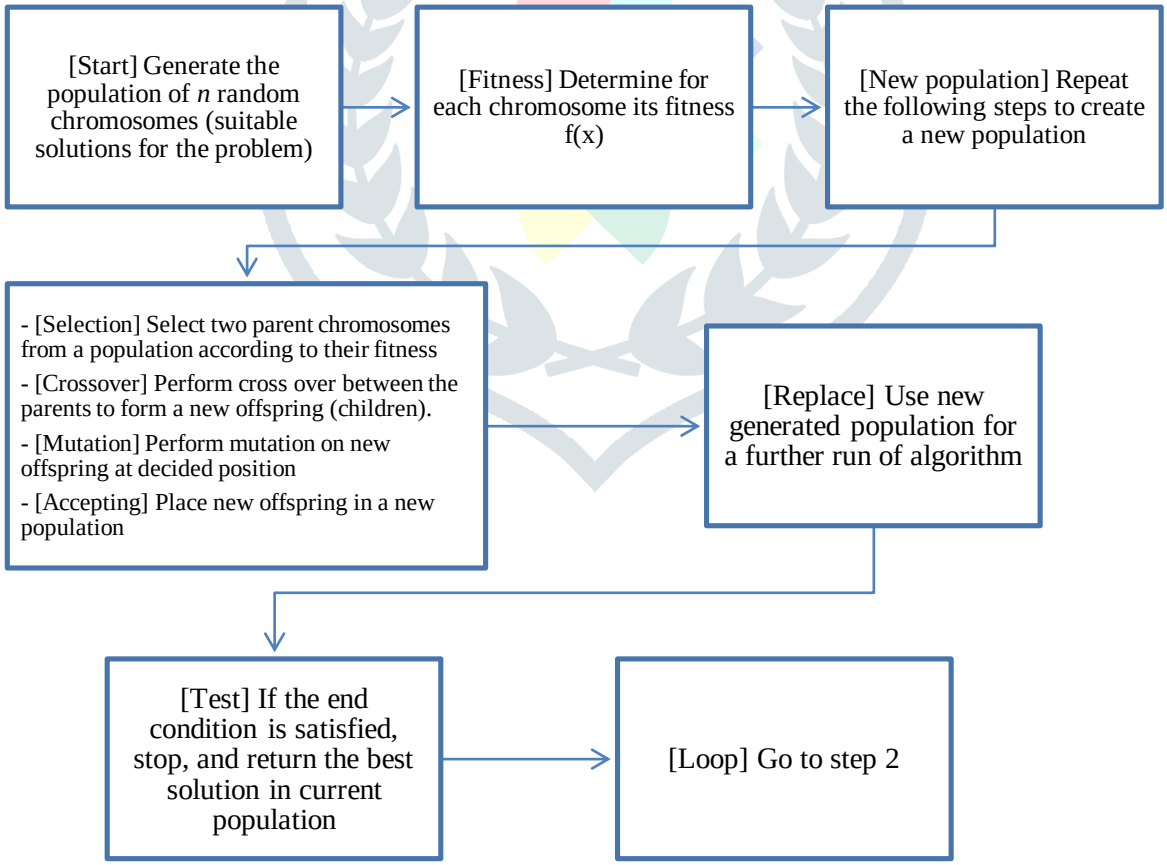

Fig.5 Basic structure of Genetic Algorithm

1.2.3 Genetic Algorithm Inspired Cryptography (GIC)

Genetic Algorithms also find application in cryptography in the concept of genetic optimization [9]. The random value generation required at each session of genetic operations, helps generate a secret key that is used in a symmetrical block cipher scheme. The principal operations involved in GIC are substitution and permutation of bits of plaintext based on the randomly

generated key. In the context of cryptography, the genetic operations are used to increase the security of the key by generating and using random numbers.

Studies have showed that when GIC is combined with another strong cryptography system, the generated algorithm is more complex and therefore more immune to attack [9].

II. Related Work

Numerous studies have been conducted to explore the contribution of genetic algorithm to the information security field. In this section, we analyze some related works about genetic inspired cryptography.

Abdallah et al. (2019) [1] pointed out in their work that encryption is a need in today's world therefore strong cryptography techniques are required to face the constantly evolving threats. The symmetric key encryption method used in this work is based on genetic algorithm which has proven to be a very high performance cryptography algorithm. Two genetic operations are involved in the process, namely crossover and mutation. The two of them require the generation of random numbers which are used for generation of a one-time symmetric key. A permutation factor is also randomly generated and applied to successive blocks of text to make the algorithm more unpredictable for the intruder.

Nazeer et al (2018) [14] proposed a cryptography method that involves Genetic Algorithms to overcome the drawbacks induced by the implementation of traditional symmetric and asymmetric cryptosystems as well as modern hash function based systems. Genetic algorithm is used as a mean to improve strength of the initial key. Key Generation, Data diffusion and data encryption are the steps required for encryption. First, a key is generated through random number generator and by applying genetic operations. Next, data diffusion is performed by using again genetic operators. Finally to encrypt the data, logical operators are applied between the diffused data. When compared to other algorithms such as DES and AES, the algorithm proved to be more secure, because it requires more time to break.

Amalarethinam et al. (2018) [2] aimed their research towards the proposition of a new algorithm for data security named KGGA Algorithm (Key Generation using Genetic Algorithm). This new algorithm uses the concepts of Genetic Algorithm as a tool for generation of a strong and optimal key that will be used for symmetric encryption of data. The optimal key is the one that satisfies the specified fitness function. Once it is found, the key is encrypted using Asymmetric Addition Chaining Cryptographic Algorithm (ACCA) which is an enhanced version of RSA algorithm. The encrypted key generated by ACCA algorithm is used as key for any one of the symmetric key algorithms like AES, DES, Blowfish etc. This technic finds its relevance in the fact that key plays a major role in encryption process especially in cloud environment.

Srikanth et al. (2017) in [10] dealt with the integrity and confidentiality of data that is transmitted through various media. The paper presents the concepts of Genetic Algorithms and proposed a cryptography scheme using ASCII conversion, binary conversion, pseudorandom number generation, crossover method and mutation. Different keys are generated according to which the type of crossover technique is performed. The keys are pseudorandom numbers generated using multiplicative congruential generator method.

Harba et al. (2017) in [3] presented a new encryption/decryption method based on genetic algorithm and ASCII to achieve two levels of security. They highlighted the fact that most genetic algorithm based models suffer from some problems such as time delay on the network communication channels and lack of robustness. In their presented model, the first key is generated by converting the text message into ASCII values and the second key will be randomly produced from genetic algorithm. After that the values of the first key and the second key are added together to obtain the ciphertext. This will be merged with the second key to obtain the secret message that will be sent to the receiver. The randomness of the two encryption keys ensures the secure transmission of the message.

Alkharji et al. (2017) [12] described and examined the effectiveness of a method to use Genetic Algorithm to generate keys for fully homomorphic encryption scheme. The advantage of a fully homomorphic scheme is the ability to support computations on ciphertext directly without exposing plain data. The genetic operations produce a key vector from which random prime numbers are picked to produce a 4096-bits RSA key. Results of this work have shown that a Genetic Algorithm based key provides more randomness thus more security than other known methods. Since the keys generated are strong, non-repeating and of high quality, it makes the scheme more difficult to break.

Jhingran et al. (2015) [6] have surveyed various research works in the field of genetic inspired cryptography (GIC) for text and image encryption. They highlighted the merits and demerits of each work and pointed out the fact that integration of genetic operations with other cryptography techniques has proven to increase efficiency of the security schemes. The benefits of several works are speed, increased security due to randomness and complexity of key. However it has been found that throughput can be decreased and implementation can be difficult.

Hassan et al. (2014) [9] introduced a hybrid encryption scheme which combines symmetrical system using Genetic Optimization and asymmetrical system using RSA. Genetic Algorithm is used to generate a strong key that has good resistance to cryptanalysis. In encryption process, plaintext is broken into blocks, before genetic operations such as mutation and crossover are used to encrypt them. The random parameters involved in these operations are used to produce a very complex key. Then asymmetric encryption helps encrypt the key. Such an approach has proven to provide high data security and high feasibility for practical implementation.

Naik et al. (2014) [11] worked on exploiting the randomness involved in crossover and mutation processes for enhancing the symmetric approach of genetic inspired cryptography. This is done by generating an asymmetric key pair for encryption and decryption of messages. The key length, hence the strength of the algorithm are dictated by the number of crossover points and mutation points, together with permutation factor and random byte to be used in the generation of a private key. The proposed algorithm employs 4 crossover points, 3 mutation points, one permutation factor and one random byte to generate a 36-bits key. The random byte is used for generation of a private key. The permutation factor agreed upon sender and receiver is used for permuting the key bits to increase security. The randomness together with permutation and generation of private key makes the algorithm difficult to break. However encrypting data in one block can be a liability because getting access to the key compromises the whole information. Another point to consider is that, if an attacker gets access to the key, he could perform brute force attack to crack it.

Sindhuja K et al. (2014) [15] pointed out the fact that genetic algorithms are reliable and powerful optimization technique that can be used in a wide variety of applications among which cryptography. They proposed a symmetric key cryptosystem based on

Genetic Algorithm. The proposed algorithm is based on three key operations: right shift, matrix addition, modulo operation and genetic operations. Here plaintext and key (user input) are converted into matrices that are added to create an additive matrix. Then a linear substitution function is applied on the additive matrix to create an intermediate cipher. Two point crossover and mutation are then applied on the intermediate cipher to encrypt the data. This method is simple and easy to implement.

III. RESEARCH GAPS

On the basis of our studies, following gaps have been identified to be further studies topics:

- Data is often encrypted at once, which makes it easy for an attacker to decrypt it if he gets access to the key
- The previous point also makes the solution unable to be applied in a parallelized environment to enhance the performance
- The private decryption key generated could be decrypted using a super calculator if it intercepted by an attacker during transfer, which is possible is key is not encrypted.

IV. PROPOSED SOLUTION

To face the insufficiencies identified, following points describe a proposed hybrid data encryption solution based on enhanced genetic operations and asymmetric encryption. Figure 6 describes the proposed security framework.

4.1 GA operations on split data

When whole data or file is encrypted at once, it makes it more vulnerable because in case of successful attack by a hacker, the data would be entirely available. Instead, we propose the process of data splitting before encryption. Data will be divided into smaller parts that will be treated independently. Each part will require separate encryption processes. To decrypt a whole data, all parts will have to be decrypted and merged. This will make it difficult for an attacker to get access to it.

Another benefit of this method is that working with smaller blocks of data makes the number of GA operations increase. So for encryption of a file, more GA operations will be required, which also increases the number of random numbers needed. As randomness increases confidentiality also increases [16]. This system will be also be useful for future work where parallel treatment or dynamic and distinct storage would be experienced and analysed.

Finally, another improvement can be found in adding a compression solution could be involved to leverage the burden of massive data transmission. In a limited resources environment, lossless compression technique will reduce size of data being stored or transported.

4.2 Key Generation

For key generation, we propose to exploit the randomness involved in GA operations to generate a key. The process will require random numbers corresponding to crossover and mutation positions. To enhance the security, a permutation factor will also be used to permute the bits of key after each encryption run. Another random value will be generated as operation order, to determine whether crossover will be performed before or after mutation. This will increase the randomness of the encryption pattern.

4.3 Encryption

Data encryption will be performed by reading and encrypting two blocks of data at a time. Encryption will be performed by applying GA operations (crossover, mutation) on the blocks according to the key pattern. After one encryption run, permutation will be performed on the key bits for the next run till the end of the file. Since this model follows a symmetric scheme, the performance will be good.

For enhanced security, the key will be encrypted using an asymmetric scheme such as RSA or Blowfish algorithm. This will allow keeping the performance high while guaranteeing a strong security.

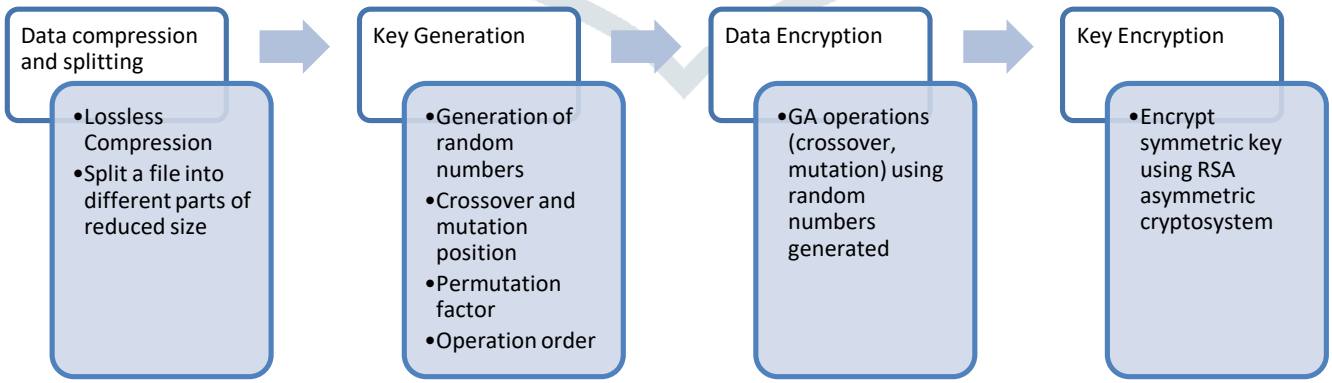

Fig.6 Flowchart of proposed security framework for data encryption

V. CONCLUSION

In this paper, we studied the applications of Genetic Algorithms in the field of cryptography. We pointed various benefits of implementing such an evolutionary computing method for increasing security level of data. We also identified some research gaps which can guide future works.

VI. FUTURE WORK

For future work, we will propose a cryptography algorithm for encryption and decryption of data, using the randomness involved in genetic operations to generate a strong key that can be used for encryption of blocks of data. The security will also be enhanced by pairing this method with an asymmetric scheme such as RSA.

REFERENCES

[1] Abdallah, A. and Ibrahim, M. 2019. Text Encryption Using Genetic Algorithm. International Journal of Computer Science and Network, 8(1).
[2] Amalarethinam, G. and Leena, H. M. 2018. A new key generation technique using GA for enhancing data security in cloud environment. International Journal of Cloud Computing, 7(1).
[3] Harba, H. S., Abb, T. and Harba, E. S. 2017. Randomly Encryption-Decryption Using Genetic Algorithm and ASCII code. Al-Kut Univ. College Journal.
[4] Veetil, A. T. 2015. An Encryption Technique Using Genetic Operators. International Journal Of Scientific & Technology Research, 4(7).
[5] Trivedi, J. and Shah, A. 2017. Survey on Efficiency of Encryption Algorithms for Cloud Data Security. IJARIIE, 3(6).
[6] Jhingran, R., Thada, R. and Dhaka, S. 2015. A study on Cryptography using Genetic Algorithm. International Journal of Computer Applications, 118(20).
[7] Sivanandan, S. N. and Deepa, S. N. 2008. Introduction to Genetic Algorithm. Springer Verlag Berlin Heidelberg.
[8] Kaur, S. and Bhardwaj, V. 2014. Study of Genetic Algorithms. International Journal of Science and Research (IJSR), 3(8).
[9] Hassan, A., Shalash, A. and Saudy, N. 2014. Modifications on RSA cryptosystem using Genetic Optimization. International Journal of Recent Research and Applied Studies, 19(2).
[10] Srikanth, P., Mehta, A., Yadav, N., Singh, S. and Singhal, S. 2017. Encryption and Decryption Using Genetic Algorithm Operations and Pseudorandom Number. International Journal of Computer Science and Network, 6(3).
[11] Naik, P. and Naik, G. 2014. Asymmetric Key Encryption using Genetic Algorithm. International Journal of Latest Trends in Engineering and Technology, 3(3).
[12] Alkharji, M., Al Hammoshi, M., Hu, C., and Liu, H. 2017. Genetic Algorithm based key Generation for Fully Homomorphic Encryption. In Information Institute Conferences, Las Vegas, NV, April 18-20.
[13] Das, S. T., Jash, S., Patra, D. and Paul, P. 2014. A Cryptography Algorithm Using the Operations of Genetic Algorithm & Pseudo Random Sequence Generating Functions. International Journal of Advances in Computer Science and Technology, 3(5).
[14] Nazeer, M. I., Mallah, G., Shaikh, N., Bhatra, R., Memon, R. and Mangrio, M. 2018. Implication of Genetic Algorithm in Cryptography to Enhance Security. International Journal of Advanced Computer Science and Applications, 9(6).
[15] Sindhuja, K. and Pramela, D. S. 2014. A symmetric key encryption Using Genetic Algorithm. International Journal of Computer Science and Information Technologies, 5(1).
[16] Mall, S. and Saroj, S. K. 2018. A new security framework for cloud data. 8th International Conference on Advances in Computing and Communication. Procedia Computer Science, 143:765-775.