International Journal of Cyber-Security and Digital Forensics (IJCSDF) 1(1): 59-66
The Society of Digital Information and Wireless Communications, 2012 (ISSN: 2305-0012)

Genetic Algorithm Approach for Risk Reduction of Information Security

Alireza Tamjidyamcholo and Rawaa Dawoud Al-Dabbagh
Artificial Intelligence Department
Faculty of Computer Science and IT
University of Malaya
itm.tamjid@gmail.com
rawaa_aldabbagh@siswa.um.edu.my

ABSTRACT

Nowadays, information systems constitute a crucial part of organizations; by losing security, these organizations will lose plenty of competitive advantages as well. The core point of information security (InfoSecu) is risk management. There are a great deal of research works and standards in security risk management (ISRM) including NIST 800-30 and ISO/IEC 27005. However, only few works of research focus on InfoSecu risk reduction, while the standards explain general principles and guidelines. They do not provide any implementation details regarding ISRM; as such reducing the InfoSecu risks in uncertain environments is painstaking. Thus, this paper applied a genetic algorithm (GA) for InfoSecu risk reduction in uncertainty. Finally, the effectiveness of the applied method was verified through an example.

KEYWORDS

Risk Reduction, Information Security (InfoSecu), Genetic Algorithm (GA).

1 INTRODUCTION

Organizations are increasingly relying on information systems (ISs) to improve business operations, facilitate management decision making, and deploy business strategies. In the current business environment, dependence has increased and a variety of transactions involving the trading of goods and services are being accomplished electronically [17]. Increasing organizational dependence on ISs has led to a corresponding increase in the impact of information security (InfoSecu) abuses. Therefore, InfoSecu is a critical issue that has attracted much attention from both IS researchers and practitioners.

IS practitioners use controls and various countermeasures (such as identifying which IS assets are vulnerable to threats) to prevent security breaches and safeguard their assets from various threat patterns. However, such implementation does not always fully protect against threats due to inherent control weaknesses [18]. Thus, risk assessment and reduction are the important steps to be taken towards InfoSecu risk management (ISRM).

Currently, most researchers are focusing on risk assessment but tend to disregard the risk reduction aspect. As a result of risk assessment alone, IS risk only gets rated but not minimized or reduced since risk reduction is quite complex and full of uncertainty [6]. The issue of uncertainty existing in the risk reduction process is one of the primary factors that influence ISRM effectiveness.

Therefore, it is crucial to address the uncertainty issue in the InfoSecu risk reduction process. To do so, we propose an InfoSecu risk reduction model based on a Genetic Algorithm (GA). According to the preliminary results, our proposed model can effectively reduce the risk derived from uncertain environments.

The rest of this paper is organized as follows: Section 2 reviews the related work, after which the basic concepts of risk assessment are explained. Next is an explanation of the Genetic Algorithm, and Section 4 discusses the proposed model. Section 5 demonstrates the validation of our proposed model. Finally, conclusions are drawn.

2 RELATED WORK

There are several factors that can influence InfoSecu, ranging from human factors to managerial and technical aspects. A deficiency in any of these areas can result in various types of losses such as economical and damage to the business reputation [1]. As an example, according to ISNA, approximately 3 million smart cards have been hacked in Iran [2]. So far, many publications on ISRM risk management and standards are reported in [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16]. However, no research has been done on minimizing or reducing the risk level. Similarly, existing ISRM software solutions such as GStool, Callio, Counter Measures, Cramm, ISAMM, Modulo Risk Manager and RA2 only concentrate on evaluating and managing risk. Moreover, no risk reduction function exists in the aforementioned ISRM software. Thus, our proposed InfoSecu risk reduction model is based on a Genetic Algorithm (GA) to reduce the level of risk.

3 RISK ASSESSMENTS PROCESS

Assessing the relative risk for each vulnerability is accomplished via a process called risk assessment. Risk assessment assigns a risk rating or score to each specific vulnerability. Rating enables one to gauge the relative risk associated with each vulnerable information asset. The risk elements include assets, threats, vulnerabilities and uncertainty. Assets broadly include the people, environment, technology and infrastructure of a system. Threats are things that can go wrong or that can 'attack' the system. Vulnerabilities make a system more prone to be attacked by a threat or allow for the possibility of an attack to more likely have some success or impact. Vulnerabilities are an asset's properties that may be exploited by a threat and include weaknesses. It is not possible to know everything about all vulnerabilities. Therefore, a factor that accounts for uncertainty must always be added to the risk assessment process, which consists of an estimate made by the manger using good judgment and experience. In fact, risks are assessed by examining the likelihood of threats and vulnerabilities and by considering the potential impact of an unwanted security incident and adding uncertainty. The shaded part in Figure 1 outlines the risk assessment steps.

4 GENETIC ALGORITHM

CGA algorithms are search algorithms based on the mechanics of natural selection and neutral genetics. They combine survival of fittest among string structures with a structure yet randomized information exchange to form a search algorithm with some of the innovative flair of human search. In

every generation; a new set of artificial creatures (string) is created using bits and pieces of the fittest of the old; an occasional new part is tried for good measure [17]. They efficiently exploit historical information to speculate on a new search points with expected improved performance. Genetic algorithms have been developed by Johan Holland and his colleagues at the University of Michigan. The goals of their research have been twofold:

1 - To abstract and rigorously explain the adaptive processes of natural system

2- To design artificial systems software that retains the important discoveries in both natural and artificial systems science.The GA has many differences from more normal optimization and search procedures in:

1- GAs work with a coding of the parameter set, not parameter themselves. The GAs require the natural parameter set of the optimization problem to be coded as a finite-length string over some finite alphabet [17].

2- GAs search from a population of points not single point.

3- GAs use payoff (objective function) information, not derivatives or other auxiliary knowledge.

4- GAs use probabilistic transition rules not deterministic rules .A canonical genetic algorithm is composed of three operators: Reproduction, Crossover, and Mutation.

The right side of Figure 1 is a flow diagram of a typical genetic algorithm process.

4.1 Representation

The CGA contains only one main data structure: population of individuals. Each individual affectionately known as a critter represents an element with the domain of the solution space of the optimization problem .The individuals in CGA are simply finite length strings of bits. Each string of 1s' and 0s' is called chromosomes for fixed length n [17]. The chromosome of a given critter is the only source for all the information about the corresponding solution. Since the variable values are represented as binary, there must be a way of converting continuous values into binary values and vice versa. The difference between the actual function value and the quantization measure is known as the quantization error. The mathematical formulae for the binary encoding and decoding of the nth variable P_n are given as follows:

For encoding:

$$P_{norm} = \frac{P_n - P_{min}}{P_{max} - P_{min}}$$

$$chromosome[m] = round\{P_{norm} - 2^{-m} - \sum_{p=1}^{m-1} chromosome[p] \times 2^{-p}\}$$

For decoding:

$$P_{quant} = \sum_{m=1}^{n} chromosome[m] \times 2^{-m} + 2^{-(m+1)}$$

$$Q_n = P_{quant}(P_{max} - P_{min})$$

In each case,

P_{norm} is a normalized variable within the range $0 \leq P_{norm} \leq 1$.

P_{min} is the minimum variable's value.

P_{max} is the maximum variable's value.

$chromosome[m]$ is the binary version of P_n.

$round\{.\}$ rounding the variable's value to the nearest integer value.

P_{quant} is the quantized version of P_{norm}.

Q_n is the quantized version of P_n.

Typically, this means that a string of $1s$ and $0s$ are used to present the decision variables, the collection of which represents a potential solution to the problem.

4.2 Setting GA Parameters

The next decision to make in implementing a genetic algorithm is how to set the values for the various parameters, such as population size, crossover rate, and mutation rate. These parameters typically interact with one another nonlinearly, so they cannot be optimized one at a time. There is a great deal of discussion of parameter settings and approaches to parameter adaptation in the evolutionary computation literature. There are no conclusive results on what is best but most often they use settings similar to those as: Population size 20-30, crossover rate 0.75-0.95, and mutation rate 0.001-0.005. He found that a very small population size was better, especially in light of other studies that have argued for large population size (e.g., Goldberg 1989), but this may be due to the on-line performance measure ; since each individual ever evaluated contributes to the on-line performance, there is a large cost for evaluating a large population [17].

4.3 Fitness Evaluation

Associated with each individual is fitness value. This value is a numerical quantification of how good of solution to optimization problem the individual is .Individual with chromosomal strings representing better solution has higher fitness values, while lower fitness values are attributed to those whose bit string represents inferior solution.

The fitness function can be one of two types: maximization or minimization. Along with the fitness function, all of the constraints on decision variables that collectively dictate whether a solution is a feasible one should be demonstrated. All infeasible solutions are eliminated, and fitness functions are computed for the feasible ones. The solutions are rank-ordered based on their fitness values; those with better fitness values are given more probability in the random selection process.

4.4 Selection Operation

Emphasize a probabilistic survival rule mixed with a fitness dependent chance to have (difference) parameters for producing more or less offspring. Their exist various kinds of different tools of selection operators. Deterministic sampling selection is used. It uses the fitness value of the previous generation (generation 0), and gives a straight forward of choosing offspring for the next generation. The next steps are crossover (recombination) and mutation operations.

4.5 Crossover Operation

The recombination operator of CGA is a variation and exploration operators that work by swapping portions between two individuals [17]. The crossover operator works entirely on the bit representation, completely ignoring the genetic code and the epigenetic apparatus.

4.6 Mutation Operation

This operation is carried out by using the function flip (the biased coin toss) to determine whether or not to change the gene value according to the mutation operation. Of course the flip function

will only come up heads (true) P_m(the probability of mutation is set at 1 / population size) as a result of the call to the pseudorandom number generator random (generates random numbers of interval [0, 1]) within flip itself [17].
The new population denoted in the new chromosome is assigned to the old strings (old chromosomes) as the next generation. In our CGA, we apply genetic operators to an entire population at each generation. Then the process continues until the maximum number of generation is reached, or the optimum solution is found.

5. RISK REDUCTION BASED GENETIC ALGORITHM

The risk identification process starts with an assessment, in which step an organization's assets should be classified and categorized accordingly.
Then, the assets should be prioritized according to their importance. In each step, data is collected from companies through interviewing experts and distributed questionnaires. For classifying and categorizing assets, once the initial inventory is assembled, it must be determined whether the asset categories are meaningful to the organization's risk management program. Such a review may cause managers to further subdivide the categories to create new categories that better meet the needs of the risk management program. Assessing values for information assets is the next step. Once each information asset is identified, categorized, and classified, a relative value must also be assigned to it. In this stage, an expert should assign values to assets. Consequently, assets should be listed in order of importance. Regarding threat identification, any organization is typically faced with a wide variety of threats, some of which comprise acts of human error or failure, compromises to intellectual property, deliberate acts of espionage or trespassing, etc. As part of vulnerability assessment, after identifying the organization's information assets and documenting various threat assessment criteria, every information asset for each threat will start to be reviewed. This review leads to the creation of a list of vulnerabilities that remain potential risks to the organization. In the risk assessment process, the relative risk for each vulnerability is evaluated. Here, a risk rating or score is assigned to each specific vulnerability.
The result of risk assessment is risk rate. If the risk rate for a specific asset is to an allowable extent, there is no need to continue the process with the genetic algorithm. If the result rate is not satisfactory, the rate needs to be deducted to an acceptable extent by applying GA to reach an adequate amount of risk. The GA needs a risk assessment variable to begin the process. Firstly, variables were assigned to GA. Second, the GA was run by the arranged elements. In the third step, the GA result was compared with the adequate risk degree, and if the result was equal to, or less than, the acceptable level, the process was finished. However, if the result did not match or was over the scale of admissible volume, the number of generations or other elements in the GA should be changed until an appropriate point is reached.

6 APPLICATION EVALUATION

The process of risk assessment is extensive and complex. Therefore, for simplification, it was assumed there is

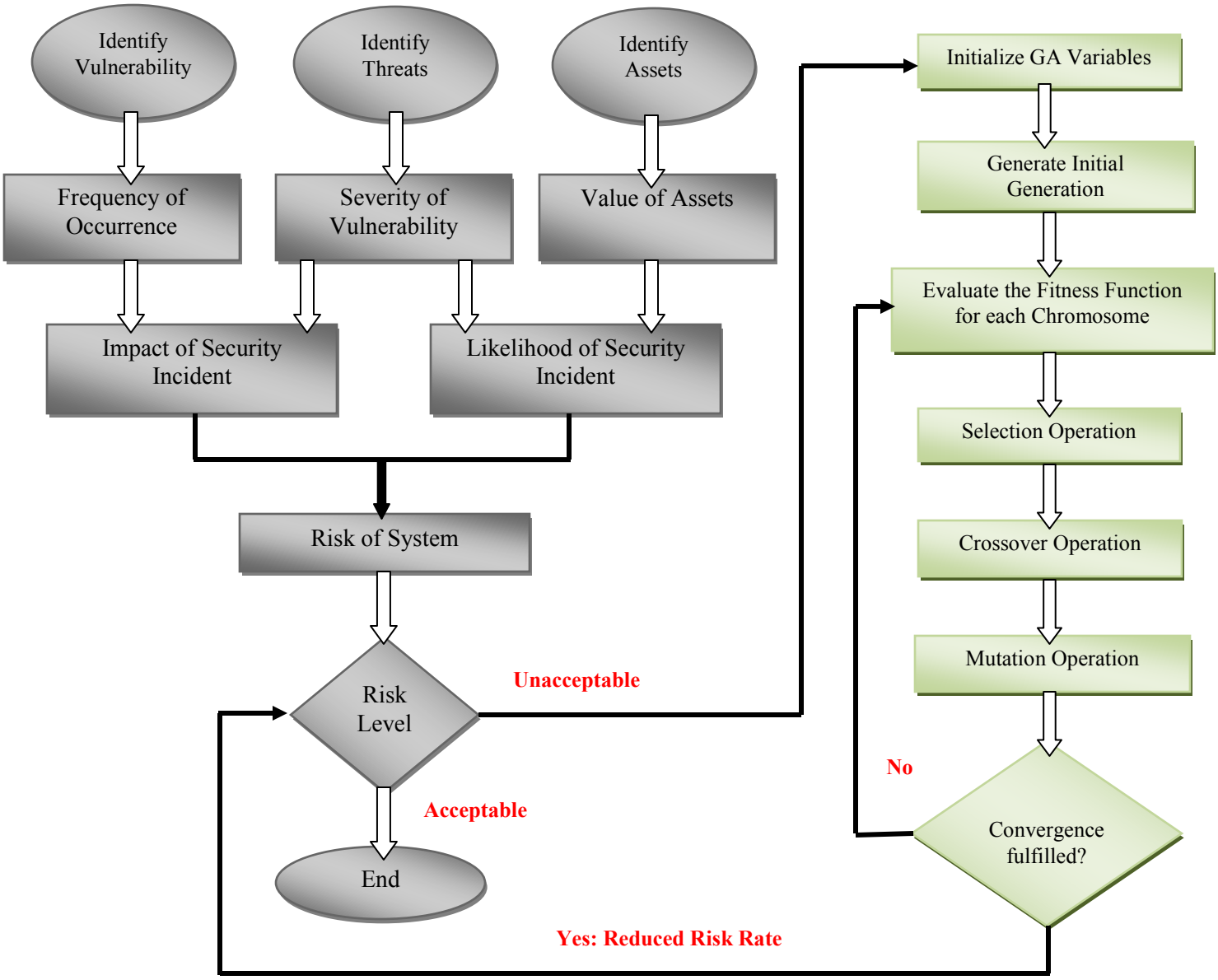

Figure 1: Risk Reduction Process

only one asset with one vulnerability, threat and uncertainty.
The risk assessment formula is,

$$Risk\ Rate = VA \times LV - (VA \times LV) \times MC + (VA \times LV) \times UV$$

Where VA denotes the information asset value (1 to 100). LV shows the likelihood of vulnerability occurrence (0 to 1). MC represents the percentage of risk mitigated by current controls (0% to 100%) and UV refers to the uncertainty of current knowledge of vulnerability is (0% to 100%). It is supposed that VA = 100, LV = 0.5, MC = 0.5 and UV = 0.2. By using GA, we want to decrease rate of risk to 0. Variables of risk assessment are used as fitness function variables.
The fitness function for GA is:

$$Y = Risk\_Function(X)$$

$$Y = X(1) \times X(2) - \big(X(1) \times X(2)\big) \times X(3) + (X(1) \times X(2)) \times X(4)$$

The Genetic Algorithm was simulated in the MATLAB environment. As shown in Figure 2, GA attained a point of zero after 51 iterations.
Other elements, diagnostic information, of the GA are shown below:

```
Fitness Function=@Risk_Reduction
```

$$Number\ of\ variables = 4$$
$$Inequality\ constraints = 0$$
$$Equality\ constraints = 0$$
$$Total\ number\ of\ linear\ constraints = 0$$

Modified option:

```
options.PopulationType          =
'bitstring'
options.PopInitRange = [-1; 1]
options.CrossoverFcn=@crossovert
wopoint
options.MutationFcn={@mutationun
iform []}
options.Display = 'diagnose'
options.PlotFcns={@gaplotbestf@g
aplotbestindiv
@gaplotdistance@gaplotselection
@gaplotstopping}
options.OutputFcns              =
@gatooloutput
```

End of diagnostic information.

7 CONCLUSION

Information security is very complicated and uncertain, and most researchers focus on risk assessment without much concern for minimizing the risk. In this paper, the rate of risk was assessed first, after which GA was applied to reduce the scale of risk. It was shown through an example that GA is effective in reducing the IS risk in organizations.

8 REFERENCES

1. M.E.Whitman and H.J.Mattord, Principal of Information Security, Second, Ed. CRC press, 2009.
2. M.EWhitman and H.J.Mattord, Management of Information Security, Second, Ed. CRC press, 2009.
3. N. Feng and M. Li, “An information systems security risk assessment model under uncertain environment,” Applied Soft Computing, vol. 11, no. 7, pp. 4332 – 4340, 2011,Soft Computing for Information System Security [Online]. Available:http://www.sciencedirect.com/science/article/pii/S1568494610001419
4. S. Kondakci, “A causal model for information security risk assessment,”in Information Assurance and Security (IAS), 2010 Sixth International Conference on, aug. 2010, pp. 143 –148.
5. J. J. Ryan, T. A. Mazzuchi, D. J. Ryan, J. L. de la Cruz, and R. Cooke, “Quantifying information security risks using expert judgment elicitation,” Computers & Operations Research, vol. 39, no. 4, pp. 774 – 784, 2012,Special Issue on Operational Research in Risk Management[Online]. Available: http://www.sciencedirect.com/science/article/pii/S0305054810002893
6. A. G. Gary Stoneburner and A. Feringa. (2002, July) Risk management guide for information technology systems. National Institute of Standars and Technology Technology Administration. [Online]. Available: http://ilna.ir/newsText.aspx?ID=255838
7. Z.-Q. Wei and M.-F. Li, “Information security risk assessment model base on fsa and ahp,” in Machine Learning and Cybernetics (ICMLC), 2010 International Conference on, vol. 5, july 2010, pp. 2252 – 2255.
8. Z. Wang and H. Zeng, “Study on the risk assessment quantitative method of information security,” in Advanced Computer Theory and Engineering (ICACTE), 2010 3rd International Conference on, vol. 6, aug. 2010, pp. V6–529 –V6–533.
9. Z. Xinlan, H. Zhifang, W. Guangfu, and Z. Xin, “Information security risk assessment methodology research: Group decision making and analytic hierarchy process,” in

Software Engineering (WCSE), 2010 Second World Congress on, vol. 2, dec. 2010, pp. 157 –160.

10. W. Shuang, Z. Tong, W. Yuan, and Z. Jian-mei, "Multi-level fuzzy-gray comprehensive evaluation of information security risk," in Management and Service Science (MASS), 2010 International Conference on, aug. 2010, pp. 1 –4.
11. W. Lijian, W. Bin, and P. Yongjun, "Research the information security risk assessment technique based on bayesian network," in Advanced Computer Theory and Engineering (ICACTE), 2010 3rd International Conference on, vol. 3, aug. 2010, pp. V3–600 –V3–604.
12. S. Kraemer, P. Carayon, and J. Clem, "Human and organizational factors in computer and information security: Pathways to vulnerabilities," Computers & Security, vol. 28, no. 7, pp. 509 – 520, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167404809000467
13. A. Asosheh, B. Dehmoubed, and A. Khani, "A new quantitative approach for information security risk assessment," in Computer Science and Information Technology, 2009. ICCSIT 2009. 2nd IEEE International Conference on, aug. 2009, pp. 222 –227.
14. A. Munteanu, D. Fotache, and O. Dospinescu, "Information systems security risk assessment: Harmonization with international accounting standards," in Computational Intelligence for Modelling Control Automation, 2008 International Conference on, dec. 2008, pp. 1111 –1117.
15. X. Wu, Y. Fu, and J. Wang, "Information systems security risk assessment on improved fuzzy ahp," in Computing, Communication, Control, and Management, 2009. CCCM 2009. ISECS International Colloquium on, vol. 4, aug. 2009, pp. 365 – 369.
16. J. H. Holland, Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. University of Michigan Press, Ann Arbor, 1992, iSBN: 0262581116.
17. D. E. Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning. Addison-wesley, 1989, iSBN: 0201157675.

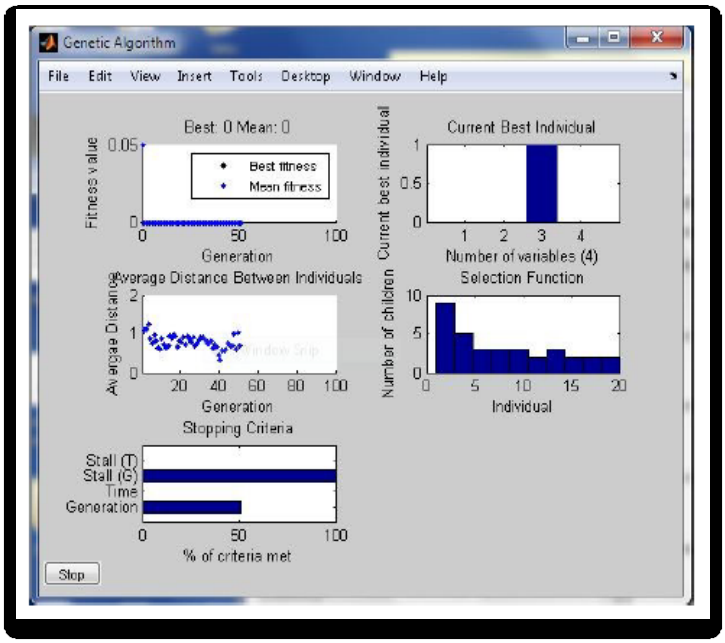

Figure 2: GA operation Results