Application of Improved Genetic Algorithm in Reliability Optimization of Multi-agent Intrusion Detection

Shaokun Liu[1, a], Lina Yu[2,b] and Yi Fang[3,c]

1Hebei College of Industry and Technology, Department of Computer Technology Shijiazhuang, China

2 Hebei College of Industry and Technology, Department of Computer Technology Shijiazhuang, China

3 Liao Cheng Electric Power Supply Company Limited Customer Center of Dong Chang Liaocheng, China

[a]liushaokun602@126.com, [b] hbyln@126.com, [c] Xiaotingting123456@163.com

Keywords: Improved Genetic Algorithm, Reliability Optimization of Multi-agent Intrusion Detection, Application.

Abstract. Detection agents in a dynamic network environment has difficult to optimize the reliability, and this problem is a typical Nondeterministic Polynomial Completeness puzzle. This paper proposed an improved genetic algorithm to solve the optimization problem, conscientiously introduced the greedy algorithm, and therefore better improved the efficiency of algorithm optimization algorithms in the optimization process. First, construct a mathematical model, multi-agent intrusion detection system between agents as the initial parameters. Then, use the improved genetic algorithm to optimize the model obtained optimal solution to complete optimization objectives.

Introduction

With the development of network technology, the intrusion behavior has become more and more complex and diverse, because of a single intrusion detection agent sometimes can't completely to illegal intrusion behavior timely and effectively, it needs multiple invasion agents to exchange the data and information among each other to finish the testing task. But in the dynamic network environment how to safely and effectively realize the cooperation between intrusion detection agent in testing, has been a focus in agent system.

Model

The proposing of multi-agency process optimization problem. Intrusion detection system has multiple intrusion detection agents, and three factors are needed to be considered, that is, the reliability of the intrusion detection agents, the time of the data transmission and the time of the testing data. Computational complexity in the mathematical model is also need to be considered, using a relatively simple algorithm as far as possible to reduce the additional computation time in consuming in the process and better to improve the optimization efficiency. The agent reliability of participating in the intrusion detection system, the time of the data transmission, the time of testing data are supposed to be known. Now proposed the following three questions:

(1) How to maximize the ultimate reliability of the detection system resource constraints.

(2) How to get the minimized system resource footprint under reliability constraints .

(3) How to change this abstract question into the clear and complete mathematical model.

Analysis of multi-agency process optimization. The detection agent optimization problem can be described as follows: Known the n detection agents may participate in testing activities, in a particular detection time, the corresponding reliability of the of each agent and their average detection time (including the time of the data transmission and the detection data time) are known and $r_i > 0$ and $t_i > 0$ ($i = 1,2, ..., n$)(r_i indicates the reliability of the i-th unit of detection agents; t_i represents the time required by the i-th detection agents unit detects). Because the system time and system resource

consumption is proportional, the amount of the detection time can be measured as a system resource footprint. The occupied time of the system directly response the number of the used system resources .

Basic assumptions of multi-agency process optimization .

(1)Detect network is always smooth without network congestion and transmission interruption and other unforeseen circumstances.

(2)According to the change of the usage of the network and detection agents time to realize data updating, i.e. the average detection time is known.

(3)The detection agent from time to time in normal working condition, such as a detection agent problem occurs, the reliability of the detection agents corresponding value update, stop work time value is set at 0.

Multi-agency process optimization model.

The mathematical model of the problem can be expressed as the following model.

$$\max R(x_1, x_{2,} \cdots, x_n) = \sum_{i=1}^{n} x_i r_i \tag{1}$$

$$s.t. \sum_{i=1}^{n} x_i r_i \le T$$

$$x_i \in \{0,1\} \quad (i = 1,2,\cdots,n)$$

$$\min T(x_1, x_{2,} \cdots, x_n) = \sum_{i=1}^{n} x_i t_i \tag{2}$$

$$s.t. \sum_{i=1}^{n} x_i r_i \ge R$$

$$x_i \in \{0,1\} \quad (i = 1,2,\cdots,n)$$

Equation (1) expressed as the reliability of the final inspection of maximizing system resource constraints and equation (2) is the reliability constraints minimization of system resources occupancy.In the two equations:

x_i is the 0-1 decision variables. $x_i = 1$ means the i-th detection agents to participate in the testing activities. $x_i = 0$ means the i-th detection agents not to participate in the testing activities;

$R(x_1, x_{2,} \cdots, x_n)$ indicates the reliability of the actually detected;

R indicates the reliability of the known defined;

$T(x_1, x_{2,} \cdots, x_n)$ indicates the actual testing time;

T represents a limited time are known;

n represents the number of agents may participate detection;

$i = 1,2,\cdots,n$ represents the serial number that agents may participate in the detection of the corresponding.

This can be more due to consider, if the form of mathematical expression in the process of computing the reliability of overly complex, will increase the computational complexity, causing the excessive amount of additional calculations, therefore, the reliability of mathematical models is established using a relatively simple algorithm,good to improve the efficiency of the optimization, but also can easily compare the reliability of the optimization results.

Such optimization problem is a typical Np complete (Nondeterministic polynomial Completeness) problem, the problem-solving method, either in theory or in practice have a certain significance, such as the allocation of resources in the management and investment decisions, loading problems, etc. for such optimization problems. Such method for solving the optimization problem is a heuristic algorithm (greedy algorithm), can also be used genetic algorithm to solve the problem.

Using the enumeration method for solving optimization problems containing n-detect proxy unit need to search for $2^n$the solving greedy algorithm (Greedy Algorithm), the first detection agent unit

reliability density eighty-two states $\rho_i = r_i / t_i (i = 1,2,\cdots,n)$ (i= 1,2, ..., n) the values in descending order, then in accordance with the order to select the appropriate detection agent unit to join the collaborative detection, (1) in the system resource constraints, until far beyond the limitations of the system resources (2), the reliability constraints, until less than the final detection reliability constraints. using this method can only be approximate optimal solution, but can not guarantee to be able to get the optimal solution. using genetic algorithm, if the problems mall can get the optimal solution or approximate optimum solution, but when demand for the solution of the optimization problem of large-scale, not get better results with a simple genetic algorithm, even in many cases to get results than by the results obtained by the greedy algorithm to be poor. this is mainly the sake of too much of the search space of the problem, and the search direction indicated by a simple genetic algorithm in such a large search space ineffective at this time, if a genetic algorithm to add some expertise of the problem will help guide its search space quickly, to help improve the quality of the genetic algorithm to solve an improved genetic algorithm to solve multi-agent optimization problem, and significantly improve the efficiency of genetic algorithm to solve the problem.

Improved genetic algorithm in multi-agent intrusion detection process optimization and results analysis

The hypothetical detection agents cell reliability is R, detect proxy unit always used time (including the transmission time and the detection time) is T, the time used for the model (l) the formula: detecting does not exceed s (model of the formula (l), for example model (2) the same reason). Its corresponding value were as follows:

$$R = r_i = \begin{Bmatrix} 0.920,0.908,0.898,0.892,0.880,0.880,0.865,0.862,0.860,0.858, \\ 0.855,0.830,0.825,0.822,0.820,0.818,0.815,0.810,0.805,0.801, \\ 0.800,0.800,0.798,0.796,0.795,0.790,0.788,0.782,0.780,0.777, \\ 0.775,0.773,0.772,0.770,0.769,0.766,0.765,0.763,0.760,0.758, \\ 0.756,0.750,0.730,0.720,0.715,0.710,0.708,0.705,0.703,0.701 \end{Bmatrix}$$

$$T = \{t_i\} = \begin{Bmatrix} 0.80,0.82,0.85,0.80,0.82,0.80,0.86,0.80,0.85,0.85, \\ 0.80,0.85,0.80,0.88,0.80,0.82,0.82,0.80,0.80,0.82, \\ 0.80,0.88,0.75,0.72,0.75,0.7\$,0.70,0.72,0.70,0.80, \\ 0.75,0.60,0.60,0.70,0.60,0.75,0.60,0.65,0.60,0.70, \\ 0.70,0.75.0.75.0.70.0.70,0.60,0.74.0.64.0.72.0.71 \end{Bmatrix}$$

The unit optimization problem for the detection agent with improved genetic algorithm to optimize the operating parameters of population size M to take a genetic algorithm, the maximum termination of algebra T, the crossover probability Pc, mutation probability Where were:

$\{M, T, P_c, P_m\} = \{50, 500, 0.6, 0.2\}$

Table 1. A detection agent unit optimization results containing 50

Time limit(S)	Decision variables $x_i = \{i = 1,2,\cdots,50\}$	Results (Reliability/Total time)
8.5	00000001010100000000111000 01010000110000100010000000	9.804 8.440
10	00000001000100000000111001 01010001110010100010000000	11.446 9.960
13.1	00000001010100100000111001 01010001111100110010000010	15.047 13.050
17.6	01000001011111011000110111 01010001110110010110000001	19.614 17.530
23.4	01100001111111011000111111 01010011111111011011001011	25.798 23.370

Conclusion

Whether it is in terms of the quality of solutions, or in solving the speed, and the improved genetic algorithm parameter sensitivity than simple genetic algorithm performance has greatly improved. Its main reason for the improvement is that chromosome over evolutionary process constraints limit the invalid, have become one of the approximate optimal solution to satisfy the constraints after the treatment of the greedy algorithm, which approximate optimal solution as the good performance of the seed genetic larger probability to the next generation. Thus, in the initial stage of the genetic algorithm, it is possible to quickly reach good Solutions nearby, so that after the evolutionary process as a basis for the good solution. , Spreadsheet results show that, when the relatively small scale of the problem, this improved genetic algorithm performance is not, and sometimes may even be worse than simple genetic algorithm to improve. However, the mathematical model is based on multi-agent intrusion detection, the number of agents, improved genetic algorithm showed greater advantage.

References

[1] Whitley, L.D.(ed.) Foundations of genetic algorithms. San Mateo, CA: Morgan Kaufmann, (l993).
[2] Whitley, L.D., and Vose, M.D. (eds.) Foundations of genetic algorithms. San Mateo, CA: Morgan Kaufmann, (1995).
[3] Melanie Mitchell. An introduction to genetic algorithms. Cambridge, MA: The MIT Press, (l996).
[4] David A Coley. An Introduction to Genetic Algorithms for Scientists and Engineers. World Scientific Publishing, Co.Pte. Ltd. (2002).
[5] KRUEGEL C, VALEUR F, VIGNA G, et al. Stateful intrusion detection for high-speed networks[A]. Proceedings of the IEEE Symposium on Security and Privacy. Berkeley, California, USA: IEEE Computer Society Press, (2002), p.285-294.
[6] FISK M, VARGHESE G. An Analysis of Fast String Matching Applied to Content-Based Forwarding And Intrusion Detection.University of California-San Diego, (2002).
[7] Intrusion prevention systems: the next step in the evolution of IDS [EB/OL]. http://www. aecurityfocus.com/infocus/1670.

Published by Atlantis Press, Paris, France.