2015 2nd International Conference on Information Science and Control Engineering

Optimization of a Switched Ethernet Topology based on Genetic Algorithm

Liu Liang, Jiang Han-hong, Rui Wan-zhi, Wang Jie
National Key Laboratory For VIPS Technology
Naval University of Engineering
Wuhan, China
e-mail: myquell@live.com

***Abstract*—In this paper, a multi-objective programming model for optimizing the topology of a practical industrial network control systems is established on the basis of the analysis of IP packets through each link, and then solved by Genetic Algorithm(GA). To ensure the effectiveness of GA, this paper proposes an individual coding scheme and its valid criterion, designs a set of specific genetic operators based on Gene Ring operation, and modifies the default parameters of corresponding switches in the practical system through simulation and calculation. The results show that the model and the algorithm can effectively reduce the network load and optimize the topology without the necessity for changing in the links of the control system by adding cost.**

Keywords-Switched Ethernet, STP, MST, GA, Gene Pool

I. Introduction

Annular structures are generally adopted in practical industrial network control system in order to improve its reliability, which serves as a tree structure in working conditions via Spanning Tree Protocol(STP) in switches. STP is defined in IEEE802.1D, which automatically chooses the path that cost least as the best one(serves as main link) and disconnect other paths(serves as backup links) logically according to switch parameter of bridge ID(BID), path cost(PC) and port ID(PID).

The spanning tree is exclusive and optimal via STP after the wiring of a ring network is determined. However, the optimality, which is based on the default parameter of BID, PC, and PID, doesn't consider reliability of the equipment and working conditions, and may not make the control system run in the optimal state. So we should optimize the network topology further according the practical condition.

For a deterministic network control system, we can first analyze the traffic magnitude and directions of each data flow by capturing IP packets, build a programming model according to the optimization objectives, solve it with suitable algorithms, and then modify the system with the best solution so that we can achieve the Minimum Spanning Tree(MST). MST problem is a sort of NP-complete problems, routinely which could be solved by artificial intelligence algorithms such as GAs, Artificial Neural Network(ANN), Simulated Annealing Algorithm, Ant Colony Algorithm, Evolutionary Computation, etc. A set of GA operators were designed in the paper to solve the programming problem for the network topology, and the result showed the validation.

II. Problem Description

The switch topology of a practical industrial network control system is shown in Figure 1, excluding the sub-controllers. The topology could be considered as a connected graph $G=(V, E)$, $V=\{Va, Vb, \ldots, Vx\}$, is the set of switch nodes, and $E=\{E_1, E_2, \ldots, E_{27}\}$ is the set of switch links, in which the solid line presents the logically active link, while the dotted line presents the backup links generated by STP. The weight of each link is denoted by w_i, i=1, 2, …, 27.

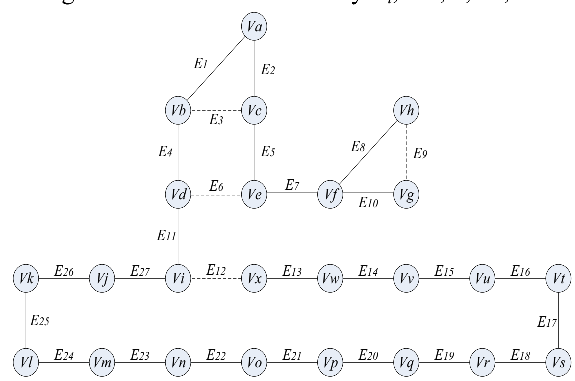

Figure 1. The topology of a practical network control system

The difference among links is insignificant because all the links are 1Gb/s, single-mode optical fiber links. Thus the weight of a link is totally determined by the traffic through it, and each data flow is deterministic, which can be inferred via capturing IP packets at several key nodes. Now we have all the data flow shown in Table 1.

TABLE I. THE MAIN TRAFFIC OF THE SYSTEM IN WORKING STATE

Sending node	Receiving node	Traffic(unit: kb/s)
Va	*Vf*	143
Vb	*Vf*	8291
Vh, *Vg*	*Vf*	4221
Vl	*Vf*	80
Vw, *Vx*	*Vf*	146
Vm~*Vv*	*Vw*	284
Va	*Vb*	211
Vw, *Vx*	*Vb*	8385

Broadcast and multicast traffic is ignored in Table 1 for it is less than 1kb/s. The state of the present network had been inferior by contrasting Figure 1 and Table 1: the average traffic was 14471kb/s, and the maximum was 19902kb/s at link 14. The link load was of great occupancy and imbalance.

978-1-4673-6850-6/15 $31.00 © 2015 IEEE
DOI 10.1109/ICISCE.2015.87

Instead of changing the physical links, we hoped to achieve the following objectives by modifying the backup links:

1. minimize the total traffic(i.e. the average traffic) of the network, because the lower load of network could make the network more stable;

2. minimize the maximum traffic among the links of the network, because the lower maximum load could reduce the collision which had the exponential relationship with network bandwidth occupancy.

let x_i, i=1, 2, …, 27, be 0-1 decision variables, x_i=1 expresses that E_i is a logically active link, while x_i=0 a dead link or an inactive link, and the programming problem could be represented as:

$$\min f(\mathbf{x}) = \sum_{i=1}^{27} w_i x_i + k x_{\max} \quad (1)$$

$$\text{s.t.} \quad \sum_{i=1}^{6} x_i = 4 \quad (2)$$

$$x_1 + x_2 \in [x_3, 2 - x_3] \quad (3)$$

$$\sum_{i=8}^{10} x_i = 2 \quad (4)$$

$$\sum_{i=12}^{27} x_i = 15 \quad (5)$$

$$x_i \in \{0,1\}, i = 1,2,...,27 \quad (6)$$

where k is the weight of the maximum traffic, which is a comprehensive subjective considering. The expressions from (2) to (5) are the criterion to judge the network is acyclic. Taking the status and the optimization objectives of the system into account, this problem is different from the others in that the weight of each link is variable. When a link is changed from an active link to a dead link, the traffic running through it will run to the destination through another link, so the traffic of corresponding links will change, i.e. $w_i = w_i(\mathbf{x})$. Therefore the objectives can't be achieved by Prim or Kruskal algorithms, and some other method will be adopted instead.

III. ALGORITHM DESIGN

The GA is a heuristic random search algorithm of global optimization. It requires little about the properties of the objective function, and can be compatible with classical optimization methods. The shortcoming is lack of convergence theory, and would lead to premature or deception if improper parameters are chose. GA is defined as an eight-tuple: GA=(C, E, P_0, M, Φ, Ψ, Γ, T), where C is the gene coding method, E is the fitness function for evaluating the individual, P_0 the initial population, M the population size, Φ the selection operator, Ψ the crossover operator, Γ the mutation operator, T is the termination condition.

3.1 Gene coding and evaluating the fitness

The gene coding, the basis of GA, is a presentation method to express a solution as chromosomal form, and it has a great impact on the performance of the algorithm. To express each network state as the form of chromosomes visually and to determine the legality of a solution quickly, the gene coding of network link state can be expressed as $C(\mathbf{x})=(x_1 x_2 \ldots x_{27})$ by the decision variables x_i. Thus the equations from (2) to (5) could also be used to judge the legality.

The individual in Figure 1 can be expressed as (0111101 10110111111111111111) by this method. The coding method can also be used for network link state monitoring: let individual $\mathbf{x}$=(011110110111111111111111100), and it expressed a failure state in which the links adjacent to the switch J were disconnected.

The fitness function is used to evaluate the fitness of an individual. Individuals with high fitness in the survival of the fittest selection are easy to produce the next generation, those with low fitness are likely to be eliminated from the population. As a minimum multi-objective optimization problem, it is necessary to be converted to a maximum optimization problem via weighting multiple targets. The fitness function will achieve by transforming the objective function into:

$$E(\mathbf{x}) = \frac{1}{f(\mathbf{x}) - A} = \frac{1}{\sum_{i=1}^{27} w_i x_i + k x_{\max} - A} \quad (7)$$

where A is a conservative estimate of $f(\mathbf{x})$.

3.2 Generating the first population

The initial population is the starting point of the search procedure of GA, which is constituted by a plurality of individuals. The convergence speed may differ from different initial populations. It must ensure that the gene set from all individuals in the population (also known as "gene pool) is equivalent to the gene set of the entire solution space to converge to the optimal solution theoretically, i.e. the search space of the GA must cover the entire solution space. It would be difficult to meet the above conditions if the initial population gene pool is incomplete, because the selection operator and crossover operator will not produce new genes, and the mutation operator will do at a quite lower probability. Though a complete gene pool would be very easy to achieve for a randomly generated initial population in conventional binary gene coding, the minimum unit in this paper is a ring, and each gene ring is of many types, so the initial population should be generated semi-manually and semi-randomly so as to meet the requirement of the diversity and completeness of the population. Population size M is set to a fixed value of 51 in the paper.

According to the schema theorem, the genetic algorithm can only ensure that individuals with low order, short defining length and fitness above average will grow in exponential way in the offspring by selection, crossover and mutation operators, but can not ensure the convergence to the optimal individual, except for that population sample traverses the entire solution space.

3.3 Genetic Operators

Genetic operators have an important impact for the convergence and search speed. The coding rule in this paper is stringent, and improper designs of genetic operator can easily lead to illegal individuals, so it is very important to design the genetic operators.

3.3.1 Crossover Operator

The crossover operator, as a central role in the GA, is an operation which generates new individuals via restructuring part of two parents' gene. New gene rings will not appear via this operator, i.e. the gene pool will stay the same.

Taking the gene coding into account in this paper, you can ensure that the individual are legitimate via the crossover method which exchanges the entire gene ring of two individuals as a unit, while other methods are prone to illegal individuals.

The crossover operator is the main factors affecting the convergence of the GA. The GA can be equivalent to a the limited absorption Markov process in the case of crossover without mutation, and it can be achieved by crossover to converge to the optimal solution under appropriate selection strategy. Crossover probability p_c is a key parameters that impact the behavior and performance of the GA, the greater p_c, the faster the speed of new individuals, but it will also lead to sooner destruction to individual with high fitness, while a smaller p_c will make the search process to slow down. In order to ensure the performance of the GA, adaptive crossover probability in this paper is shown as follows:

$$p_c = \begin{cases} p_{c1} - \dfrac{p_{c1} - p_{c2}(E' - \overline{E})}{E_{\max} - \overline{E}}, E' \geq \overline{E} \\ p_{c1}, \qquad\qquad\qquad\qquad E' < \overline{E} \end{cases} \tag{8}$$

where p_{c1}=0.9, p_{c2}=0.5, $E_{\max}$ is the maximum fitness in the current population, $\overline{E}$ is the average fitness, E' is the larger fitness of the two individuals to crossover.

Procedure of gene ring crossover operator:

(1) all individuals are paired off in population $P_{t\text{-}1}$ except for the best individual of past generations;

(2) traverse all paired individuals and determine crossover operation according to the crossover probability;

(3) randomly select a gene ring to crossover if determined, otherwise go to (4);

(4) population P_{t1} is determined if traverse ends, otherwise go to (2).

3.3.2 Mutation Operator

The mutation operator, generally as an auxiliary operator, modifies genes at some gene locus. The number of gene rings, i.e. the gene pool, will change by this operator.

Relative to the crossover operator capable of global searching, the mutation operator is capable of local random searching, thus the GA could be capable of both global and local searching, which makes the GA to maintain the diversity of population and to prevent the emergence of non-mature convergence. An effective genetic algorithm design should consider the effective co-ordination between mutation operator and crossover operator.

Increasing the mutation probability at the beginning and reducing it at the end can improve search speed. When the mutation probability p_m>0.5, the GA will degenerate to random search with loss of some important mathematical properties and search capabilities of GAs. This paper selected p_m=0.05, and adopted the cautious mutation operator to ensure the effectiveness of the algorithm as follows:

Procedure of cautious mutation operator:

(1) traverse all the gene rings of all individuals except for the best individual of past generations, and determine crossover operation according to the crossover probability for each gene ring;

(2) randomly generate a gene ring replacing the old one if mutation is determined and the gene ring is not the only one in the gene pool, otherwise go to (3);

(3) population P_{t2} is determined if traverse ends, otherwise go to (1).

3.3.3 Selection Operator

The selection operator, which is also called "replication" operator, chooses superior survivals from the whole generation according the rule "survival of the fittest". New individuals will not appear via this operator.

Common selection operators will probably lead to prematurity or deception by way of selecting the same individual of high fitness many times due to randomicity of selection, so a selection operator named Cross Generational Elitist (CGE) Selection without duplicates is adopted in this paper, which mixes the last generation with this generation population in the iterative process, selects the optimum individual in accordance with a certain probability, and ensures the individuals from the new population do not reproduce in order to maintain the diversity of the population genetic pattern. This selection method can effectively improve the behavior of the GA, but is relatively time-consuming, and not suitable for a larger population. An additional individual is used to record the best individual in the past population so as to ensure the convergence of the GA, which does not participate in genetic operations. This is the reason why the population size is an odd number in the paper. Hence the procedure is different and the fitness of all the individuals should be recalculated because the selection operator follows the other genetic operations:

Procedure of CGE selection operator:

(1) generate Roulette Population P_R via mixing P_{t2} with $P_{t\text{-}1}$ excluding the same individuals;

(2) let $\mathbf{x}_P$ be the best individual, the new population P_t= $\{\mathbf{x}_P\}$, P_R =P_R-$\{\mathbf{x}_P\}$;

(3) select a random individual $\mathbf{x}_i$ from the roulette table, P_t= P_t +$\{\mathbf{x}_i\}$, P_R =P_R-$\{\mathbf{x}_i\}$;

(4) population P_t is determined if P_t reaches the population size, otherwise go to (3).

3.4 The termination condition and the flow chart

The termination condition of a GA is generally set to a fixed termination generation T. If the optimization problem is complex, the termination condition could be set to that the average fitness of the population exceeds a certain threshold, or the average fitness of successive generations of growth is below a certain threshold, etc. In this paper, T is set to 50. The flow chart of the algorithm is shown in Figure 2.

IV. SIMULATION RESULT AND SYSTEM VERIFICATION

The Matlab simulation parameters selected population size as 41, generations as 50, and the trend of best individual obtained is shown in Figure 3.

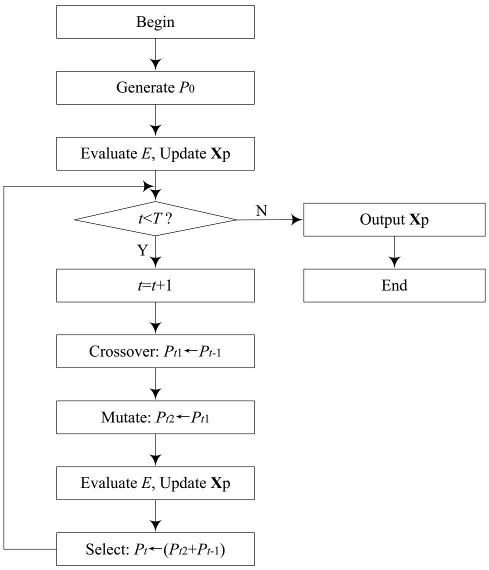

Figure 2. The flow chart of the GA

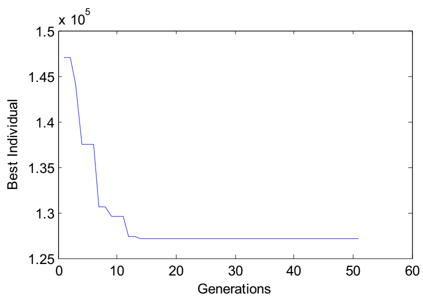

Figure 3. The trend graph of the best individual

In Figure 3, the optimal individual is (101110110111111 111110111111), the average link traffic of the network structure is 4750 kb/s, and the maximum traffic is 17914 kb/s at link 12.The network topology with link traffic is shown in Figure 4.

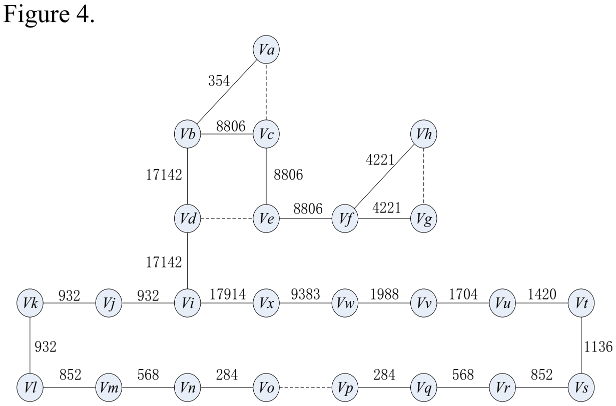

Figure 4. The control system network topology after optimization

If the maximum traffic in the target function is ignored for the optimal individual, the optimal solution obtained through the same algorithm is (10011111011111111110111 111). Its average link traffic is 4718 kb/s, lower than the above optimal solution, but the maximum traffic is 25204 kb/s at link 4, and the probability of collision is so greatly increased that the solution is not suitable as the optimal one.

According to the simulation results, the PCs for six corresponding switches port (*Va*, *Vc*, *Vd*, *Ve*, *Vo* and *Vp*) are modified into 120 (the PC for a 1000Mb/s port is 4, a 100 Mb/s port 19, a 10Mb/s port 100 by default), greater than the other ports so as to match the network control system topology with Figure 4. The system has run in good condition so far after the topology modification, and the network load is consistent with the theoretical value.

V. Conclusion

The GA provides an effective solution and a common framework to network topology optimization problems. However, the focus of GAs is more on theoretical research than on practical engineering applications. In this paper, a multi-objective programming model of an existing deterministic industrial network control system was established, a GA based on gene ring operation was proposed and it was proved effective via Matlab simulation and practical system. The result, while it isn't the topology of minimum average traffic, is more balanced and reliable than that topology, which is in line with the optimal objective of the network system.

References

[1] P. ping s, Vonnahrne E, Jasperneite J. Analysis of switched Ethernet networks with different topologies used in automation systems. Fieldbus Technology, l999, pp.351-358.

[2] Krommenacker N, Divoux T, Rondeau E. Using genetic algorithms to design switched Ethernet industrial networks. Proceedings of the 2002 IEEE International symposium. Vandoeuvreles Nancy, France, 2002, pp.152-157.

[3] Gen M, Cheng R, Oren S S. Network Design Techniques Using Adapted Genetic Algorithm. Advances in Engineering Software, 2001, pp.731-744.

[4] Georges J P, Krommenacker N, Divoux T. Rondeau E. A design process of switched Ethernet architectures according to real-time application constraints. Engineering Applications of Artificial Intelligence. 2006, pp.335-344.

[5] Sayoud H, Takahashi K. Designing Communication Networks Topologies Using Steady-state Genetic Algorithm. IEEE Communication Letters. 2001, pp.113-115.

[6] Dorogovtsev S N, Mendes J F F, Samukhin A N. Structure of growing networks with preferential linking. Physical Review Letters, 2000, pp.4633-4636.

[7] ZHANG Qi-zhi, TANG You-chun, SUN Hua-li, ZHANG Wei-dong. Using Genetic Algorithms to Partition Switched Industrial Ethernet. Control and Instruments In Chemical Industry, 2005, pp.31-34

[8] GU Kai, ZHANG Qi-zhi. Optimization of the ring topology of industrial Ethernet based on genetic algorithm. Control Theory & Applications, 2006, pp.597-600.

[9] ZHANG Qi-zhi, ZHANG Bin, ZHANG Wei-dong. Calculation of maximum delay in switched industrial Ethernet based on network calculus. Control and Decision, 2005, pp.117-120.