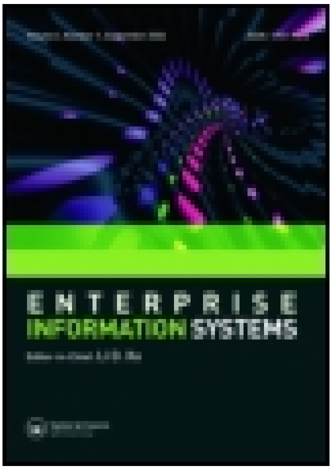

Enterprise Information Systems
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/teis20

# Study on the evolutionary optimisation of the topology of network control systems

Zude Zhou [a] , Benyuan Chen [a] , Hong Wang [b] & Zhun Fan [c]
[a] School of Information Engineering, Wuhan University of Technology , Wuhan, 430070, China
[b] Department of Management , North Carolina A&T State University , Greensboro, NC, 27411, USA
[c] Department of Management Engineering , Technical University of Denmark , Lyngby, 2800, Denmark
Published online: 02 Mar 2010.

*Enterprise Information Systems*
Vol. 4, No. 3, August 2010, 247–264



# Study on the evolutionary optimisation of the topology of network control systems

Zude Zhou[a], Benyuan Chen[a]*, Hong Wang[b] and Zhun Fan[c]

[a]*School of Information Engineering, Wuhan University of Technology, Wuhan 430070, China;* [b]*Department of Management, North Carolina A&T State University, Greensboro, NC 27411, USA;* [c]*Department of Management Engineering, Technical University of Denmark, Lyngby 2800, Denmark*

(*Received 20 August 2009; final version received 3 December 2009*)

Computer networks have been very popular in enterprise applications. However, optimisation of network designs that allows networks to be used more efficiently in industrial environment and enterprise applications remains an interesting research topic. This article mainly discusses the topology optimisation theory and methods of the network control system based on switched Ethernet in an industrial context. Factors that affect the real-time performance of the industrial control network are presented in detail, and optimisation criteria with their internal relations are analysed. After the definition of performance parameters, the normalised indices for the evaluation of the topology optimisation are proposed. The topology optimisation problem is formulated as a multi-objective optimisation problem and the evolutionary algorithm is applied to solve it. Special communication characteristics of the industrial control network are considered in the optimisation process. In respect to the evolutionary algorithm design, an improved arena algorithm is proposed for the construction of the non-dominated set of the population. In addition, for the evaluation of individuals, the integrated use of the dominative relation method and the objective function combination method, for reducing the computational cost of the algorithm, are given. Simulation tests show that the performance of the proposed algorithm is preferable and superior compared to other algorithms. The final solution greatly improves the following indices: traffic localisation, traffic balance and utilisation rate balance of switches. In addition, a new performance index with its estimation process is proposed.

**Keywords:** enterprise computing; industrial network; evolutionary algorithm; topology optimisation; network calculus; delay distribution; enterprise applications



## 1. Introduction

The combination of manufacturing and information technology promotes the development of the network manufacturing. In the network manufacturing environment, equipment acts as a resource node of an enterprise or even a global manufacturing network, receiving commands and executing tasks on the manufacturing networks. And switched Ethernet, which is the most widely used option

*Corresponding author. Email: chenby@whut.edu.cn

ISSN 1751-7575 print/ISSN 1751-7583 online

DOI: 10.1080/17517570903527861
http://www.informaworld.com

nowadays, has many features for this function. By using switched Ethernet in the industrial context, we have only one network type at each level of the computer integrated manufacturing architecture which facilitates the information system management inside and outside an enterprise. Another example of network applications is the enterprise-wide information systems such as enterprise resource planning, decision support systems, executive support systems, digital dashboards and production monitoring systems (Hsu and Wallace 2007, Li *et al*. 2008, 2009). These systems need to collect various types of data and information from all over the organisations where computer networks play very important roles. Some real-time systems such as on-line analytical processing systems also need computer networks to collect and process data on-line (Xu *et al.* 2005, 2008, Luo *et al.* 2007). In addition, the industrial applications now need to exchange more complex information, such as images or videos, which requires more bandwidth. In contrast with the field bus, switched Ethernet provides a large amount of bandwidth.

For a network system, the network delay can be a major factor that affects the system's performance, especially for a real-time control system. Reducing the delay's influence to the network is the key issue for research on the network control system. In switched Ethernet, the end-to-end delay of the data frame depends on the network topology and the queue delay of frames in the switches. Once the topology is selected, the transmission and propagation delay of the frames and the switching delay of the switches are fixed. If a data frame does not queue in the switch, it will have a minimum fixed delay. Therefore, topology optimisation is the first step to ensure the real-time communication and the reliability of the performance of the network. Although the switched Ethernet itself has many features for real-time communications, such as the micro-segmentation of network traffic and the full-duplex link, a less optimal design of the network topology can still generate bottlenecks and slow down the network traffic.

Research on network topology optimisation, focused on different networks such as the communication network, control network, etc., is based on different optimisation theories such as graph theory and various search algorithms. This article concentrates on the control network, and the optimisation method is based on the evolutionary algorithm. Zhang *et al.* point out that the cable distance, traffic distribution, traffic balance and network delay are the main factors to be considered (Zhang *et al*. 2001). They also indicate that the topology optimisation considering all four factors is a non-linear programming problem that is very complex and hard to describe. During an optimisation process based on genetic algorithms, the four factors are assigned different weights and the total cost is the weighted sum of the four. The four weights can be selected according to demands. However, the four factors are not absolutely independent and they belong to different categories. Thus, it is neither accurate nor appropriate to describe the optimisation objective by merely using the weighted sum method. Zhang *et al.* select traffic allocation and traffic balance as two main factors, among many other influential factors (Zhang *et al.* 2005). The network partition problem is then formulated as a multi-objective optimisation problem. In the objective function, the authors add a weight coefficient before the traffic balance factor to compromise between the two objectives. This article also uses the weighted sum method to convert the two objectives into a single one, which may result in the complementation of the two objectives during optimisation and prevent both objectives from being optimised simultaneously if the two objectives are contradictory to each other. The study by Hu *et al.* also focuses on control network and adopts a traffic control strategy (Hu *et al.* 2007). What is different from Zhang *et al.* is that, even though two weight coefficients are used in the modelling of the objective function, the weight values are determined by experiments and experiences, which can introduce subjectivity and uncertainty to the method implemented (Zhang *et al.* 2005). The control network modelling of Li *et al.* is based on graph theory and its solution is based on the graph partition strategy (Li *et al.* 2007). The strategy takes the minimisation of the inter-subnet traffic as the optimisation objective and the traffic balance together with the port utilisation rate balance of switches as two approximate constraints, which means the traffic and node number of each subnet approaches the average as closely as possible. The method controls the port utilisation rate balance of each switch well, but it does not consider the optimisation of the redundant topology. Han *et al.* also describe the network partition as the graph partition and use the genetic algorithm as the optimisation tool, aiming at the minimisation of the subgraphs' relations (Han *et al.* 2006). But the article uses a single objective modelling without considering other factors that can affect the real-time performance of the control network. Al-Bassam *et al.* take the distance and load of links as criteria and design the network topology based on the genetic algorithm (Al-Bassam *et al.* 2006). The objective function is designed as the product of factors. The article focuses on ordinary communication network, so it refers little to the real-time performance and reliability of the network. The spectral partition algorithm proposed by Krommenaker *et al.* is based on graph theory, and the main criterion of its partition is the equipartition of the vertices and minimisation of the sum of edge weights (Krommenaker *et al.* 2001). But in the illustrative example, the article assumes the communication traffic between any two pieces of equipment is equal, which may restrict its real-world application to some extent. Krommenaker *et al.* also carry out the modelling based on graph theory and then solve it based on the genetic algorithm, aiming at minimising the inter-subnet communication traffic (Krommenaker *et al.* 2002a, b). What is the same as Han *et al.* is that they do not consider the factors comprehensively (Han *et al.* 2006). The network topology is also equivalent to the graph by Chou *et al.* (Chou *et al.* 2001). They mainly analyse the operator design, which affects the performance of the genetic algorithm and provides the simulation results. But their model does not aim at a specific network. Sem and Malhotra's study the multi-objective network design (Sem and Malhotra 2008). The two optimisation indices are traffic loss and reliability of links. They take the sum of losses as the core index and the reliability as the penalty. Sayoud *et al.* concentrate on the minimisation of the total cost of the network, which is the sum of the node cost, the link cost and the link maintenance cost (Sayoud *et al.* 2001). Yang, Cheng and Wang proposed to use GAs with immigrants and memory schemes to solve the dynamic SP routing problem in MAGETs (Yang *et al.* 2010).

Based on the research stated above, this article will study the topology optimisation of the control network more comprehensively. The rest of this article is organised as follows. In Section 2, the optimisation problem is analysed thoroughly, including the network topology that will be studied, main factors that affect the network performance, optimisation criteria and the relations between them. In Section 3, the optimisation problem is formulated. In Section 4, the optimisation algorithm is introduced and explained in detail. Illustrative examples are presented in Section 5. A new performance index and its estimation process are proposed in Section 6. Section 7 concludes the article.

## 2. Optimisation problem analysis

### 2.1. Network topology

In the optimisation of the control network's topology, the real-time performance, reliability and expansibility of the system should be considered. Star topology and linear topology are the most commonly used methods in network control systems. In this article, the two-level hybrid topology of switched Ethernet as shown in Figure 1 is studied. When multi-level topology is needed, the network can be designed starting from below to above. The network architecture consists of a star topology where a federative switch interconnects all end switches and a linear topology serving as the redundant topology. When the Spanning Tree algorithm detects a failure on the backbone, the linear topology replaces the star topology.

### 2.2. Optimisation criteria

There is no definite or unified rule for criteria adoption in network topology optimisation. Generally, there are traffic indices, delay indices, cost indices, etc. This article aims at the control network so the real-time performance is the index that should mainly be considered. The traffic criterion is implicit in the delay criterion, so there is no need to consider the two criteria separately. When it comes to the internal relationship between the two criteria, the delay composition of switched Ethernet should be taken into account. Let $D_{\text{frame}}$ be the frame transmission delay, $D_{\text{propagation}}$ be the signal propagation delay, $D_{\text{interval}}$ be the inter-frame delay, $D_{\text{transmit}}$ be the transmission delay of the switch and $D_{\min}$ be the minimum end-to-end delay. The maximum delay of switched Ethernet with multiple switches given by Lee and Lee (2002) and Lee *et al*. (2004) is as follows:

$$\begin{aligned} D = D_{\min} &+ (N_{\text{smax}} - 1)(D_{\text{frame}} + D_{\text{propagation}} + D_{\text{transmit}}) \\ &+ (N_{\text{qmax}} - 1)(D_{\text{interval}} + D_{\text{frame}}) \end{aligned}$$

where $N_{\text{smax}}$ is the number of switches on the longest path between the source and destination, and $N_{\text{qmax}}$ is the number of nodes in the whole network. For $N_{\text{qmax}}$, because of the randomness of the communication, which node will launch the communication and when it does this, as well as at which switch the data frames will

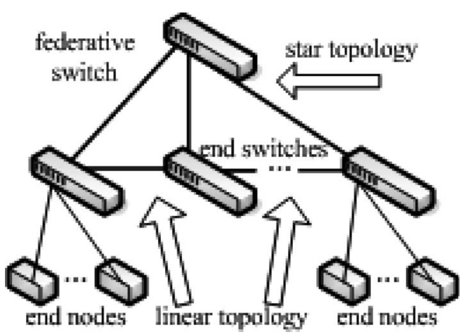


Figure 1. Network topology studied.

aggregate, are uncertain. What can be confirmed is that the maximum queue delay of a frame is the sum of its queue delays in every switch in the worst case.

If the delay criterion is adopted, the expression of average delay should be derived to be used as the objective function for optimisation. The traditional delay analysis model for general networks is usually based on queuing theory, in which the flow from the source to the end is modelled as the Poisson process. However, it cannot accurately describe the data information produced by equipment in the control system. For the delay-non-determinism of the network control system, the accurate expression of average delay cannot be easily achieved. And it may be too pessimistic to use the maximum delay as the index. On the other hand, from the analysis of the delay composition, the main control variable of network topology optimisation is $N_s$, while $N_q$ has a large uncertainty. What the topology optimisation can do to $N_q$ is allocate the devices with larger traffic to others in the same subnet. Therefore, the topology optimisation can control only a part of the delay effectively. Under this consideration, this article uses a traffic criterion which generally includes the following indices: traffic localised index and traffic balanced index. The former tries to achieve the maximisation of the traffic in subnets and to reduce the inter-subnet communication. The latter tries to balance the load of each subnet so as to evenly distribute the total communication burden to subnets. A higher balanced index results in a better communication performance of the network. It is beneficial to the maintenance of network equipment and is convenient for the system extension. Unfortunately, the two indices contradict each other and cannot be optimised simultaneously. Therefore, when the traffic criterion is applied, a comprehensive solution to this multi-objective optimisation problem is needed.

The cost is not the main index this article considers, but it is necessary in the optimisation process. The cost index contains many aspects. This article mainly considers the number of switches used and the port utilisation rate of each switch. Suppose all switches of the system are of a unified type; the port number of each switch is $G$; the number of devices to be allocated is $M$. Excluding the necessary federative switch, the minimum number of second level switches needed ($P$) can be calculated as follows: if $M\%\ g = 0$, then $P = M/g$, else $P = M/g{+}1$. The redundant topology is taken into account, so $g = G-3$. Thus, the minimum number of subnets necessary ($P$) is determined in this way.

## 3. Formulation of the optimisation problem

The communication traffic between any two field devices to be allocated is the main optimisation basis. The communication traffic among devices can be defined into an $M \times M$ matrix $\mathbf{A}$. Each element $a_{ij}$ shows the communication weight between source device $i$ and destination device $j$, where $a_{ii} = 0$, $a_{ij} \neq a_{ji}$. This is in accord with the one-way communication characteristic of field devices. Matrix $\mathbf{A}$ reflects the communication relationship of devices and is the input of the topology optimisation. The sum of its elements is defined as the total communication traffic input ($U$). Suppose the control network is partitioned into $P$ subnets and all the devices are allocated into each subnet. Let an $M \times P$ matrix $\mathbf{H}$ represent which device belongs to which subnet. An element $h_{ij}$ means: if device $i$ belongs to subnet $j$, $h_{ij} = 1$, otherwise $h_{ij} = 0$. According to matrix $\mathbf{A}$ and $\mathbf{H}$, we can get a $P \times P$ matrix $\mathbf{S}$ that shows the communication relationship of subnets. The element $s_{ij}$ represents the communication traffic from subnet $i$ to subnet $j$. Obviously, the sum of the diagonal elements of **S** represents the total traffic inside the subnets, while the sum of all the elements excluding the diagonal of **S** represents the total inter-subnet traffic.

For the hybrid topology in this study, the main difference in the parameter definition between linear topology and star topology is the load of each subnet ($L_k$). In the star topology, a federative switch interconnects all end switches, which can guarantee that the traffic from source subnet to destination subnet will not pass through any other subnets. Star topology can minimise the total network load and achieve the theoretical minimum average delay. But in the linear topology, subnets are not equal anymore. Those subnets that are located relatively in the middle will take on more retransmitting tasks. So it is necessary to define a decision variable $x_{i,j}^k$: if traffic from subnet $i$ to subnet $j$ through subnet $k$, $x_{i,j}^k = 1$, otherwise, $x_{i,j}^k = 0$. The load expressions of subnet $k$ in the two situations have been defined in (Elbaum and Sidi 1995, 1996):

$$L_k = \sum_{i=1}^{P} (s_{ik} + s_{ki}) - s_{kk} \tag{1}$$

$$L_k = \sum_{i=1}^{P} \sum_{j=1}^{P} s_{ij} \cdot x_{i,j}^k \quad 1 \le k \le P \tag{2}$$

To execute the traffic criterion well, and to make the following definition of the fitness function and treatment of experimental data more convenient, the normalised definitions of two key indices are proposed. Let $T_{\text{local}}$ represent the traffic localised index, $T_{\overline{\text{local}}}$ represent the traffic de-localised index, $T_{\text{balance}}$ represent the traffic balanced index, and $T_{\overline{\text{balance}}}$ represent the traffic unbalanced index. Assume that all traffic of devices are generated in a subnet and destined within this subnet. Here, $\sum_{i=1}^{P} s_{ii} = U$. The traffic is completely localised so that $T_{\text{local}} = 1$ and $T_{\overline{\text{local}}} = 0$. Thus, the traffic localised and de-localised indices can be defined as:

$$T_{\text{local}} = \sum_{i=1}^{P} s_{ii} \Big/ U \tag{3}$$

$$T_{\overline{\text{local}}} = 1 - \sum_{i=1}^{P} s_{ii} \Big/ U = \sum_{i=1}^{P} \sum_{\substack{j=1 \\ j \ne i}}^{P} s_{ij} \Big/ U \tag{4}$$

Assume again that the real communication load after the device allocation ($V = \sum_{k=1}^{P} L_k$) is equally distributed among each subnet. This means the traffic is completely balanced. Here, $\forall k$, $L_k = \frac{V}{P}$, $T_{\text{balance}} = 1$, $T_{\overline{\text{balance}}} = 0$. For the general case, the sum of the load deviation to the average is $\sum_{k=1}^{P} \left| L_k - \frac{V}{P} \right|$. Consider the case in which all communication is in one subnet. The communication load of this subnet is $U$, while the loads of other subnets are 0, and the real communication load of the whole control network $V = U$. So the sum of the deviations is:

$$\sum_{k=1}^{P} \left| L_k - \frac{V}{P} \right| = 2 \cdot \frac{P-1}{P} \cdot U \tag{5}$$

Thus, the traffic balanced and unbalanced indices can be defined as:

$$T_{\text{balance}} = 1 - \frac{P \cdot \sum_{k=1}^{P} \left| L_k - \frac{V}{P} \right|}{2(P-1)U} \tag{6}$$

$$T_{\overline{\text{balance}}} = \frac{P \cdot \sum_{k=1}^{P} \left| L_k - \frac{V}{P} \right|}{2(P-1)U} \tag{7}$$

Constrained by the switch itself, the optimisation model should be subject to the following constraints: (1) devices allocated into each subnet ($g_p$) cannot be more than the available ports of the switch ($g$); (2) the communication load of each subnet ($L_k$) cannot exceed the capacity of the subnet ($C_k$).

## 4. Design of the optimisation algorithm

### 4.1. Objective function

According to the traffic criterion, the traffic localised index and the traffic balanced index are not simple linear relations. Therefore, the topology optimisation of the network control system is equivalent to a multi-objective optimisation problem. In this article, the evolution algorithm is applied. The objective function is defined as follows. $f$ is the specific model and $f_1$, $f_2$ are the two parameters to be optimised.

$$\text{Minimise}\, f = \{f_1, f_2\} \tag{8}$$

where $f_1 = T_{\overline{\text{local}}}$, $f_2 = T_{\overline{\text{balance}}}$

Subject to $L_k < C_k$, $1 \leq k \leq P$

$$g_p \leq g$$

### 4.2. Non-dominated set construction

The non-dominated set is constructed by using the arena algorithm (Zheng 2007). And the algorithm is modified and improved in this article. Directly applying the original algorithm to the topology optimisation process often leads to a low efficiency and a waste of storage. So in the implementation step of the algorithm, the 'mask code method' is designed to avoid the repeated shift accessing and rearrangement of individuals, improving the executing efficiency of the algorithm and saving memory space. The modified algorithm is described as follows:

```
Function establish-NDSet (Pop: population)
{Q = Pop; Sign_count2 = 0; mask initialised;
  while (|Q| > 1) do
  {X ∈ Q; X = succeed(0, mask); Sign = .F.; Sign_count1 = 0;
    while ((Sign_count2 > 0) and (W2 ∈ X)) do
    {Q = Q − X; X = succeed(X,mask); Sign-count2 = Sign-count2 − 1;}
      Y = succeed(X+1,mask);
      for (Y ∈ Q)
```

```
{if (Sign_count2 > 0)
  then {Sign_count2 = Sign_count2 − 1;
    if ((X > Y) OR (W2 > Y)) then {Q = Q − Y; Y = succeed(Y,mask);}
    if (Y > X)
      then {Q = Q − X; X = Y; W1 = Y; Sign = .T. ;
      Sign_count1 = the position of Y; Y = succeed(Y+1,mask);}
    else {Y = succeed(Y+1,mask);}}
else if (X > Y)
  then {Q = Q − Y; Y = succeed(Y,mask);}
else if (Y > X)
  then {Q = Q − X; X = Y; W1 = Y; Sign = .T.;
    Sign_count1 = the position of Y; Y = succeed(Y+1,mask);}
else {Y = succeed(Y+1,mask);}
}end of for-loop
if (|Q| >= 1) then NDSet = NDSet ∪ {X};
if Sign then {Sign_count2 = Sign_count1; W2 = W1;}
}end of while-loop
if (|Q| = 1) then NDSet = NDSet ∪ Q;}
```

Next, the meanings of the variables used are explained briefly: $Q$ implies the population; *mask* implies the mask code of the population; *Sign* records whether the conqueror of this round is changed (only the last record is effective); *Sign_count1* records the position in which the change takes place (only the last record is effective); *V1* records the last conqueror after the changes; $X$ implies the temporary conqueror; $Y$ implies the challenger; *Sign_count2* records the position in which the final conqueror change takes place in the previous round; *V2* records the last conqueror of the previous round.

### 4.3. Diversity maintenance

The crowding density principle is used for the diversity maintenance of the evolutionary population. When a new population is generated, individuals with better fitness and with a smaller crowding density will be reserved and will participate in the next evolution (Chaudhry *et al.* 2000, Feng *et al.* 2003, Jiang *et al.* 2009). An individual with a smaller density, on the contrary, has a larger crowding distance. The crowding distance of an individual can be evaluated by the sum of the distance difference between the individual and its adjacent individuals on every sub-objective. For this, the individuals need to be sorted according to their sub-objective values.

However, the evaluation of the crowding density is a relatively time-consuming step in the algorithm, so this principle is mainly used in the non-dominated set. As for the dominated set, when it is used for a certain individual, random individuals will be introduced into the population to restrain the premature convergence.

### 4.4. Individual evaluation

There are some differences in the definitions of the fitness function and objective function presented before. In the multi-objective algorithm, dominative relation strategy and objective function combination strategy can be used for the determination of an individual's fitness. If the former is adopted, the boundary set level of an individual would need to be analysed in each generation. Generally, the construction of the boundary set is the main factor that affects the efficiency of the evolutionary algorithm. The latter strategy generally includes the sum and product combinations. In the weight sum combination, the boundary points may be easily lost, and it cannot ensure that all sub-objectives achieve optimum values simultaneously. In addition, the determination of the weights is carried out with great subjectivity. The product combination is not inclined to lose boundary points, but it cannot always put all of the non-dominated individuals into the matched library. This article differentiates between the non-dominated and dominated individuals and uses different methods to evaluate and locate them. The population is divided into the non-dominated set and dominated set by set construction. In the non-dominated set, the individual $i$ is evaluated by crowding distance calculation, while in the dominated set, the fitness of the individual $i$ is defined as fitness $(i) = \mathrm{fit}_1 \times \mathrm{fit}_2$, where $\mathrm{fit}_1 = T_{\mathrm{local}}$, $\mathrm{fit}_2 = T_{\mathrm{balance}}$. The dominated individual $i$ is evaluated by its fitness. An individual is considered better if its fitness is larger. The definition of sub-fitness is a little different from that of the sub-objective function to, on the one hand, make this problem more direct and, on the other hand, improve the likelihood that the individuals who are excellent in both traffic indices are conserved.

### *4.5. Coding*

The coding scheme determines how to construct the solution of the optimisation problem and affects the design of genetic operators. The integer coding that provides the intuitive mapping from the representative domain to the problem domain is chosen. The solution of the network partition is encoded into an integral vector. If the $i$th element of an individual is $k$, device $i$ is allocated to subnet $k$.

### *4.6. Genetic operators*

For the selection operator, the article adopts the cross generational elitist selection strategy, namely, to select the optimal individuals according to a certain probability in the mixed population of the previous generation and the new population generated by the crossover.

For the crossover operator, the article adopts the uniform crossover strategy, namely, to exchange the genes in each gene locus of the two matched individuals according to the same crossover probability, thereby generating two new individuals. The uniform crossover strategy is more generalised for taking every point as the potential crossover point. The strategy itself is aligned with the characteristic of the problem to be solved, because each device has an equal probability of being allocated to each subnet. It is necessary to check the validity of the offspring individuals immediately after every crossover operation.

Aimed at the network partition, the mutation operator randomly selects some sites of individuals generated by the crossover operation with a certain probability and randomly allocates them to other subnets different from the original. The mutation range is obviously the possible number range of subnets. After the mutation, it is again necessary to check the validity of the offspring individuals immediately.

### *4.7. Redundant topology optimisation*

The linear topology serves as redundancy, so the input of its optimisation is the optimal individual of the star topology.

The indices of the linear topology optimisation mainly depend on two aspects. The first is minimising the sum load of the subnets. In other words, the real communication load of the whole network $V$ gets minimised. Here, $V-U$ is used as the first evaluation index. The second is balancing the load of each subnet. The sum of the load differences is used. Thus, the optimisation problem can be defined as:

$$\text{Minimise } f = \{f_1,\ f_2\} \tag{9}$$

where $f_1 = V - U,\ f_2 = \sum_{j=i+1}^{P} \sum_{i=1}^{P-1} |Li - Lj|$

Subject to $L_k < C_k,\ 1 \leq k \leq P$

The subnet load in the linear topology can be calculated according to Equation (2). The main problem lies in the determination of the decision variable $x_{i,j}^k$ to different network partitions. The chromosome by integer coding is adopted to denote a concrete partition, and the coding scheme shows the linear order of subnets. A decision formulary of $x_{i,j}^k$ can be determined on the basis of the individual gene locus. Then, the coding scheme and elements of $\mathbf{S}$ are brought into the formulary to calculate $L_k$. The method is described as follows:

Function genp-x (gene locus: $y[P]$)
$\{P$ = number of subnets; $L_k$ = load of subnet $k$;
$L_{korig}$ = self-load;
for ($k = 0;k < P;k{+}{+})\{$
  $L_k = L_{korig}\}$
  for ($i = 0;i <= P-2;i{+}{+})\{$
    for ($j = i+1;j <= P-1;\ j{+}{+})\{$
      for ($k = 1;k < j-i;\ k{+}{+})\{$
$L_{[y[i+k]-1]} += s_{[y[i]-1][y[j]-1]} + s_{[y[j]-1][y[i]-1]};\}\}\}$
$\}$

where $P$ is the number of subnets; $y[P]$ is the gene locus; $L_k$ is the total load of subnet $k$; $L_{korig}$ is the self-load of subnet $k$ (the total load excluding the retransmission tasks of subnet $k$).

## 5. Examples

To validate the algorithm proposed under different communication loads, the article considers two examples, and compares the results with other algorithms. We assume the number of devices to be allocated is 40 and each end switch has 16 ports. According to the derivation in 2.2, the least number of subnets $P$ is 4. The communication traffic between any two devices can be described by the integer weights from 1 to 10. For matrix $\mathbf{A}$, the element $a_{ij} = 1$ represents device $i$ sending the smallest Ethernet frame (64 bytes) to device $j$ every 0.01 s. Thus, the real communication traffic is proportional to the weight. In the algorithm, let the crossover probability be 0.9 and the mutation probability be 0.01.

### *5.1. Example 1*

The communication traffic input used here is 1741. We consider a bad topology organisation: device 1–10, 11–20, 21–30 and 31–40 are allocated into four different subnets. In this case, the chromosome is 1111111111222222222233333333334444444444; the total traffic inside subnets is 348 and the inter-subnet traffic is 1393; the real communication load of the whole network is 3134; the loads of subnets are 739, 871, 831 and 693; the traffic de-localised index and the traffic unbalanced index are 0.800115 and 0.103389, respectively. We consider the MMRB algorithm proposed by Li *et al.* (2007). The optimal chromosome given by it is 2214332411311414134332234133224224342114; the total traffic inside subnets is 739 and inter-subnet traffic is 1002; the real communication load of the whole network is 2743; the loads of subnets are 702, 686, 691 and 664; the two indices are 0.575531 and 0.017231. Then, consider the MMKP algorithm proposed by Li *et al.* (2007). The optimal chromosome given by it is 433211423313324231211341241134244212 4432; the total traffic inside subnets is 745 and the inter-subnet traffic is 996; the real communication load of the whole network is 2737; the loads of subnets are 691, 664, 697 and 685; the two indices are 0.572085 and 0.015700.

Now consider the algorithm proposed in this article. We ran the program 10 times and obtained 10 distributions of non-dominated solutions as shown in Figure 2.

According to the distribution of the non-dominated set in Figure 2, we compromise between the two indices and choose the solution needed in the boundary set. But if the port utilisation ratio balance of end switches is taken into account, it is necessary to choose the partition scheme in which the numbers of devices allocated in each subnet are more balanced. Based on this demand, individual 4323221121112 423344111413214323443234423 is the relatively optimal solution. Its two indices are 0.557725 and 0.009190; the total traffic inside subnets is 770 and the inter-subnet traffic is 971; the real communication load of the whole network is 2712; the loads of subnets are 676, 686, 668 and 682. It can be seen that, compared with the random partition, MMRB and MMKP algorithms, the solution of the algorithm proposed here is much better on both traffic localised and traffic balanced indices. In fact, if

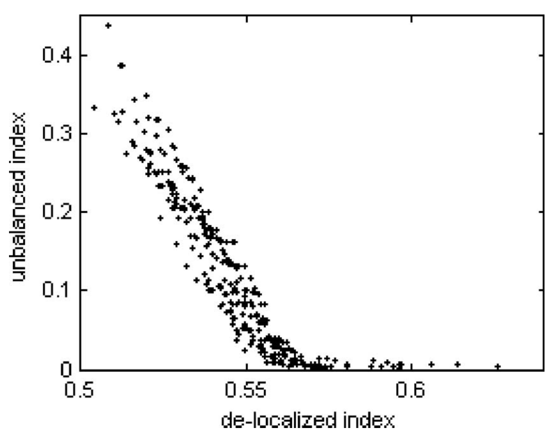


Figure 2. Non-dominated set (Example 1).

not considering the port utilisation ratio balance of end switches, designers can select more optimised solutions on both indices than on the chosen individual. For practical application, based on the boundary set and distribution in Figure 2, choose individuals by making compromises among the two traffic indices and the physical balance of switches according to demands.

### 5.2. *Example 2*

We further carry out the evaluation and comparison and change the device communication matrix input. The communication traffic input this time is 349. Still, the random partition is considered first: device 1–10, 11–20, 21–30 and 31–40 are allocated into four different subnets and the corresponding chromosome is 1111 111111222222222233333333334444444444. The two indices are 0.994269 and 0.152818; the total traffic inside subnets is 2 and the inter-subnets traffic is 347; the real communication load of the whole network is 696; the loads of subnets are 194, 152, 156 and 194. Obviously, most of the traffic needs to be retransmitted among subnets. It means that the federative switch needs to retransmit most of the traffic. We then consider the solution given by the genetic algorithm proposed by Hu *et al.* (2007). The chromosome that corresponds to the solution is 421221 4243312314414424443213132211233211. The two indices are 0.530086 and 0.225406; the total traffic inside subnets is 164 and the inter-subnets traffic is 185; the real communication load of the whole network is 534; the loads of subnets are 189, 137, 90 and 118.

Next, we ran the program 10 times randomly and obtained 10 distributions of non-dominated solutions as shown in Figure 3.

By making compromises between the two indices according to the distribution of non-dominated solutions in Figure 3, choose a relatively optimal solution. Also, if the port utilisation ratio balance of end switches is taken into account, it is again necessary to choose the partition scheme in which the numbers of devices in each subnet are relatively balanced. So individual 23114123442133122444344221 14113342433312 is selected. Its two indices are 0.120344 and 0.108883; the total traffic inside subnets is 307 and the inter-subnet traffic is 42; the real communication

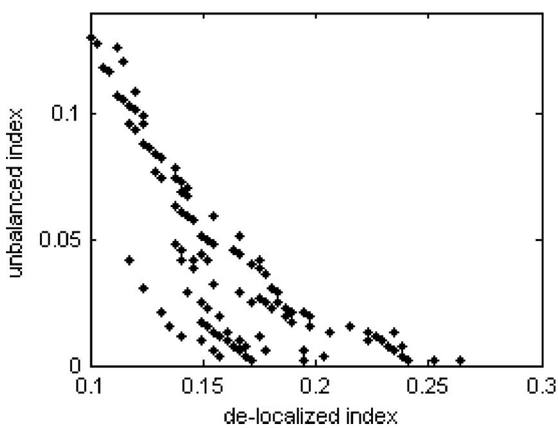


Figure 3. Non-dominated set (Example 2).

load of the whole network is 391; the loads of subnets are 123, 101, 83 and 84. Compared with the algorithm mentioned above, the solution given by this algorithm shows a great improvement in every respect. And if not considering the physical balance of switches, designers can choose better solutions in the boundary set to make both indices even more optimised.

### *5.3. Redundant topology*

For example 1, the individual 4323221121112423344111413214323443234423 is the optimal solution in star topology optimisation. It is used as the input of the linear topology optimisation. When the number of subnets is small, all possible solutions can be enumerated. The distribution of solutions is shown in Figure 4.

From Figure 4, the optimal solution can be easily found. In addition, the sequence of subnets is symmetrical so this optimal point corresponds to two individuals, namely, 2413 and 3142. Here, the real communication load of the whole network is 3302; the two objective values are 1561 and 1232; the loads of subnets are 961, 686, 668 and 987.

For example 2, we use individual 2311412344213312244434422114113342433312 as the input. After the linear topology optimisation, all solutions can be enumerated. The distribution is shown in Figure 5.

From Figure 5, we choose individual 1432 and 2341 as the final solutions. The real communication load of the whole network is 411; the two objective values are 62 and 97; the loads of subnets are 123, 101, 94 and 93.

The two examples showed that the proposed algorithm performs well. In comparison with other algorithms, the final solution has a substantial improvement on traffic localising, traffic balancing and the port utilisation balancing of switches.

## 6. Preliminary study of a new performance index

Besides the previously mentioned optimisation scheme, another simpler method will be discussed in this section. The purpose of compromising the optimisation of the two traffic indices is to decrease the network delay in essence. On one hand, the inter-subnet traffic must be transmitted via the federative switch, so it causes a much larger delay than the traffic inside subnets. A higher traffic localised index means less

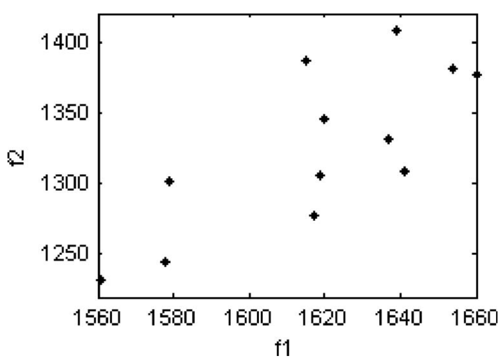


Figure 4. Solution set of linear topology (Example 1).

inter-subnet traffic. On the other hand, if a subnet has relatively less communication burden, the possible queue length of the end switch will be much shorter and the queuing delay will decrease. A higher traffic balanced index means the total communication burden is distributed to subnets more evenly. So the multi-object optimisation can be converted to a single object optimisation if the two traffic indices' collaborative optimisation is merely convert to the delay index' optimisation. Thus, the optimisation process can be simplified. But the expression of the mean delay cannot be achieved easily because the delay is uncertain, and it is improper to describe the data arrival property in the control network by the Poisson or Bernoulli distributions. So, the crux of the solution changes from the multi-object optimisation to the delay estimation of the control network.

In this article, the mean delay of the control network is estimated based on stochastic network calculus so as to evaluate the individuals corresponding to different network partitions. The relationship between the upper bound of the delay distribution and the input aggregate and service curve should be derived first. Then, the expression of the upper bound $P(D(0)>d)$ can be derived based on stochastic comparisons. $D(0)$ denotes the delay of a flow that arrives in the switch at time 0. $P(D(0)>d)$ is the probability that $D(0)$ exceeds $d$. The delay distribution is a complementary distribution. This means it considers all of the flows that are queued in the same buffer, including the flow to be inspected.

To illustrate of the estimation process, the single level star topology as shown in Figure 6 is studied, where the switch is denoted as SW and the end nodes are denoted as $ST_i$, $I = 0,2,\ldots,I$. The first come first served (FCFS) strategy is adopted as the scheduling algorithm of the switch.

Take $ST_0$ as the destination node and $ST_i$ ($i = 1,2,\ldots,I$) as the source nodes. Study the communication delay from $ST_i$ ($i = 1,2,\ldots,I$) to $ST_0$.

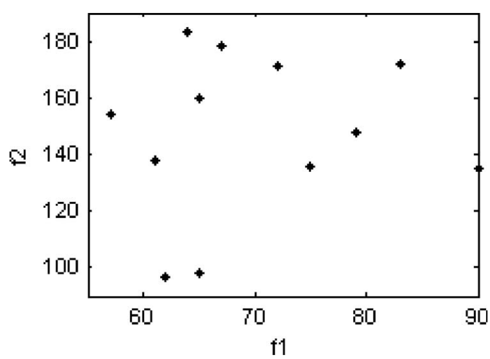


Figure 5. Solution set of linear topology (Example 2).

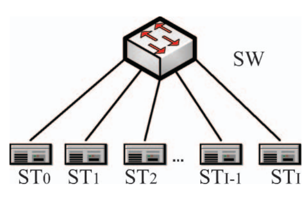


Figure 6. Single level star topology.

Assume that all input flows are restricted by their arrival curves respectively, while SW offers the service curve to the aggregate of input flows. Let $\bar{\alpha}_i$ be the sustainable rate of the data arriving at the $i$th input port, $\bar{\alpha} = \sum_{i=1}^{I} \bar{\alpha}_i$.

Assume $\exists \tau < \infty$, such that $\forall s \geq \tau$, $\beta(s) \geq \alpha(s) \cdot \tau$ is called the upper bound on the duration of a busy period. The existence of $\tau$ is the stability condition of the system. $\forall K \in N$, let $T_K(\tau)$ be the set of partitions of [0, $\tau$] in $K$ intervals. In other words, $T_K(\tau) = \{(s_0, s_1, \ldots, s_K)$: $0 = s_0 \leq s_1 \leq \ldots s_K = \tau$.

Based on the stochastic network calculus (Vojnovic and Le Boudec 2002, 2003), $\forall K \in N$, $\forall \underline{s} \in T_K(\tau)$, $d \in [0, h(\alpha, \beta)]$. The delay distribution bound of the aggregate flow through the switch can be expressed as:

$$P(D(0) > d) \leq \sum_{k=0}^{K-1} \exp\left(-\frac{2 \cdot \left((\beta(s_k + d) - \bar{\alpha} \cdot s_{k+1})^+\right)^2}{\sum_{i=1}^{I} \alpha_i^2(s_{k+1})}\right) \tag{10}$$

To estimate the mean delay, we need to calculate the upper bound of $P(D(0) > d)$ with $d = \Delta d, 2\Delta d, 3\Delta d, \ldots$ ($\mu$s) until $P(D(0) > d) = 0$. Obviously, $P(D(0) > 0) = 1$ and as $d$ increases the probability $P(D(0) > d)$ decreases, which means $P(D(0) > d)$ $P(D(0) > d + \Delta d$. So we consider

$$P(D(0) = d) = P(D(0) > d - \Delta d) - P(D(0) > d + \Delta d)$$

The size of $\Delta d$ reflects the number of sample delay points.

Assume that the other 20 end nodes communicate with $ST_0$, namely, $I = 20$. The port rate of SW is 100 Mbps. To simplify the analysis, we assume that each node sends their data with the same law. The frame size of switched Ethernet is from 64 to 1518 bytes and, when calculating the arrival curves, we need to add seven preamble bytes, one byte for the start-of-frame delimiter and 96 bits of frame intervals. Assume $ST_i$ ($i = 1,2, \ldots, I$) sends the smallest possible Ethernet frame every 10 ms (periodic real-time data). The corresponding arrival curve is $\alpha_i^{cr}(t) = 8400t + 84$. In addition, $ST_i$ sends burst real-time data with a rate that is 1/10 of the periodic data's rate and at most 10 frames successively. Thus, its arrival curve is $\alpha_i^{br}(t) = 840t + 840$.

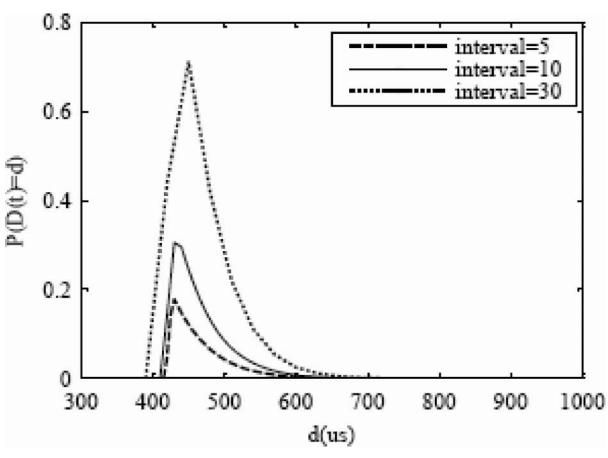


Figure 7. Stochastic delay distribution.

Following the steps introduced above, we can get the delay distribution as shown in Figure 7, where interval denotes the value of $\Delta d$ and the delay unit is in microseconds ($\mu$s).

From Figure 7, the delay distribution range is from 390 $\mu$s to 930 $\mu$s. For different $\Delta d$, there is a slight difference in corresponding peak delay and its probability but the variation range is basically from 430 to 450 $\mu$s. The upper bound of the delay is 930 $\mu$s. According to the sample delay points and relevant probabilities, designers can estimate the mean delay to evaluate each individual acquired in the optimisation.

## 7. Conclusion

As one of the most important tools in enterprise applications, computer networks play more and more critical roles today. Optimisation of network designs that allows networks to be used more efficiently in industrial environment and enterprise applications remains an interesting research topic. This article aims at the network topology of the control system based on switched Ethernet, takes the evolutionary algorithm as the optimisation tool, and studies the hybrid topology optimisation of the control network which consists of the star topology and the linear topology (redundant topology). Starting with the criteria of the topology optimisation, the article introduced the factors that affected the real-time performance of the control network and analysed the delay criterion, traffic criterion and cost criterion together with the internal relations between them. Based on the definition of performance parameters of the control network, two normalised evaluation indices for the topology optimisation were provided, namely, the traffic de-localised index and traffic unbalanced index. Then, the problem was formulated as a multi-objective optimisation problem and the evolutionary algorithm was adopted to solve it. In the optimisation, communication characteristics of the control network, such as the existence of the controller and the one-way communication of field devices, were especially considered. Both the main working topology and redundant topology were optimised. The improved arena algorithm proposed for the construction of the non-dominated set was verified to be more efficient. Because the construction of the boundary set and the calculation of crowding density were identified as the main temporal overheads, for the evaluation of individuals, the integrated use of the dominative relation method and the objective function combination method were adopted to reduce the cost of the algorithm. Specific examples showed that the proposed algorithm performed well, and in comparison with other algorithms, the final solution had improvements in every respect. The proposed optimisation method can be well applied to the control network. In addition, a new performance index was studied as a preliminary step. By this, the crux of the matter changed from the multi-object optimisation to index estimation. The estimation was based on stochastic network calculus and effective steps with illustrations were given. As for future work, we plan to extend the analyses in Section 6 and are also interested in better ways of deriving the expression of the average delay.

## Acknowledgements

The authors are grateful to NSFC (National Natural Science Foundation of China) for the projects under grant No.50675166 and No.50620130441.

## References

Al-Bassam, B., Alheraish, A., and Bakry, S., 2006. A tutorial on using genetic algorithms for the design of network topology. *International Journal of Network Management*, 16 (4), 253–262.

Chaudhry, S., Varano, M., and Xu, L., 2000. Systems research, genetic algorithms and information systems. *Systems Research and Behavioral Science*, 17 (2), 149–162.

Chou, H., Premkumar, G., and Chu, C., 2001. Genetic algorithms for communications network design-an empirical study of the factors that influence performance. *IEEE Transactions on Evolutionary Computation*, 5 (3), 236–249.

Elbaum, R. and Sidi, M., 1995. Topological design of local area networks using genetic algorithms. *In*: *Proceedings of the 14th annual joint conference of the IEEE computer and communications societies*, Boston, USA, Piscataway, NJ: Institute of Electrical and Electronics Engineers, Vol. 1, 64–71.

Elbaum, R. and Sidi, M., 1996. Topological design of local area networks using genetic algorithms. *IEEE/ACM Transactions on Networking*, 4 (5), 766–778.

Feng, S., *et al.*, 2003. An intelligent agent with layered architecture for operating systems resource management. *Expert Systems*, 20 (4), 171–178.

Han, J., *et al.*, 2006. An approach of industrial Ethernet network system design with hybrid niche genetic algorithm. *In*: *Proceedings of the World Congress on Intelligent Control and Automation*, Dalian, China. Piscataway, NJ: Institute of Electrical and Electronics Engineers, Vol. 1, 3699–3703.

Hsu, C. and Wallace, W., 2007. An industrial network flow information integration model for supply chain management and intelligent transportation. *Enterprise Information Systems*, 1 (3), 327–351.

Hu, X., Zhu, D., and Wang, B., 2007. Using improved genetic algorithms to design switched industrial Ethernet topology. *Computer Engineering and Science*, 29 (9), 9–12.

Jiang, Y., *et al.*, 2009. Influencing factors for predicting financial performance based on genetic algorithms. *Systems Research and Behavioural Science*, 26 (6), 661–673.

Krommenaker, N., Rondeau, E., and Divoux, T., 2001. Study of algorithms to define the cabling plan of switched Ethernet for real-time applications. *In*: *Proceedings of the 8th IEEE symposium on emerging technologies and factory automation*, Antibes-Juan les Pins, France. Piscataway, NJ: Institute of Electrical and Electronics Engineers, 223–230.

Krommenaker, N., Divoux, T., and Rondeau, E., 2002a. Using genetic algorithm to design switched ethernet industrial network. *In*: *Proceedings of the 2002 IEEE international symposium on industrial electronics,* Piscataway, NJ: Institute of Electrical and Electronics Engineers, Vol. 1, 152–157.

Krommenaker, N., Rondeau, E., and Divoux, T., 2002b. Genetic algorithms for industrial ethernet network design. *In*: *Proceedings of the 4th IEEE International Workshop on Factory Communication Systems*, Stevenage, UK: Institution of Engineering and Technology, 149–156.

Lee, K. and Lee, S., 2002. Performance evaluation of switched Ethernet for networked control systems. *In*: *Proceedings of the 28th Annual Conference of the IEEE Industrial Electronics Society,* Piscataway, NJ: Institute of Electrical and Electronics Engineers, Vol. 4, 3170–3175.

Lee, S., Lee, K., and Kim, H., 2004. Maximum communication delay of a real-time industrial switched Ethernet with multiple switching hubs. *In*: *Proceeding of the 30th Annual Conference of the IEEE Industrial Electronics Society*, volume 3, 2327–2332.

Li, F., Zhang, Q., and Zhang, W., 2007. Graph partitioning strategy for the topology design of industrial network. *IET Communications*, 1 (6), 1104–1110.

Li, S., Wang, X., and Zhang, D., 2008. Node localisation in wireless sensor network based on self-organising isometric embedding. *Enterprise Information Systems*, 2 (3), 259–273.

Li, S., Wang, X., and Zhang, D., 2009. An adaptive distributed localisation in wireless sensor network. *Enterprise Information Systems*, 3 (2), 117–132.

Luo, J., *et al.*, 2007. Flood decision support system on agent grid: method and implementation. *Enterprise Information Systems*, 1 (1), 49–68.

Sayoud, H., Takahshi, K., and Vaillant, B., 2001. Designing communication networks topologies using steady-state genetic algorithms. *IEEE Communications Letters*, 5 (3), 113–115.

Sem, K. and Malhotra, S., 2008. Multi criteria network design using genetic algorithm. *In*: *Proceedings of the 2008 IET International Conference on Wireless, Mobile and Multimedia Networks*, Mumbai, India. Stevenage, UK: Institution of Engineering and Technology, 56–60.

Vojnovic, M. and Le Boudec, J., 2002. Elements of probabilistic network calculus for packet scale rate guarantee nodes. *In*: *Proceedings of 15th International Symposium on Mathematical Theory of Networks and Systems*, Indiana, USA, 1–13.

Vojnovic, M. and Le Boudec, J., 2003. Bounds for independent regulated inputs multiplexed in a service curve network element. *IEEE Transactions on Communications*, 51 (5), 735–740.

Xu, L., *et al*., 2005. A polychromatic sets approach to the conceptual design of machine tools. *International Journal of Production Research*, 43 (12), 2397–2421.

Xu, L., Liang, N., and Gao, Q., 2008. An integrated approach for agricultural ecosystem management. *IEEE Transactions on Systems, Man and Cybernetics Part C*, 38 (4), 590–599.

Yang, S., Cheng, H., and Wang, F., 2010. Genetic algorithms with immigrants and memory schemes for dynamic shortest path routing problems in mobile ad hoc networks. *IEEE Transactions on Systems, Man and Cybernetics Part C*, 40 (1), 52–63.

Zhang, S., *et al*., 2001. A new algorithm of solving large-scale network topological design based on genetic algorithm. *Journal of China Institute of Communication*, 22 (6), 27–33.

Zhang, Q., *et al*., 2007. Using genetic algorithms to partition switched industrial Ethernet. *Control and Instruments in Chemical Industry*, 32 (3), 31–35.

Zheng, J.H., 2007. *Multi-objective evolutionary algorithm and its application*. Beijing: Science Press.