Multi-objective Topology Optimization Based on Mapping Matrix and NSGA-II for Switched Industrial Internet of Things

Jielin LI*, Ming CHEN†
* School of Mechanical Engineering, Tongji University, Shanghai, 201804, China
Email: neew@163.com
†Sino-German College of Applied Sciences, Tongji University, Shanghai 201804, China
Email: 1210309@tongji.edu.cn

Abstract—With the deep integration of Information and Communication Technology (ICT), the network performance of existing switched Ethernet cannot meet the growing control and management demands of the industrial production. To solve the time-delay and sub-network transmission balance problem, a multi-objective Non-dominated Sorting Genetic Algorithm (NSGA-II) based on mapping matrix is proposed, in which a new topology model is established to reflect the flexible Ethernet topology format. Meanwhile an adjacency mapping matrix is introduced as genetic code to embody the actual topology structure and several special properties of the mapping matrix caused by the industrial topology characteristics are proved. On the basis of these properties, an asexual crossover and double swap mutation operation is proposed to achieve the population evolution. Besides, in the genetic procedure of the proposed algorithm a self-adaptive measure and an elite reservation strategy are applied to accelerate the convergence speed. The validity of proposed algorithm is demonstrated by benchmark test and the case verification shows that the comprehensive performance of the switched Ethernet is improved.

Index Terms—Internet of Things (IoT); NSGA-II; Switched Industrial Ethernet; Network Topology; Mapping Matrix

I. INTRODUCTION

With the proposal of the fourth industrial revolution [1], the deep integration between ICT and industrial production area has been paid much attention to. The network connection layer is the communication bridge between the management layer and the production infrastructure, so in order to achieve the data acquisition in IoT and the information handling in CPS (Cyber Physical System) [2], [3], a stable and reliable network topology structure is essential throughout the entire production process. Especially in the mechanical manufacturing, there exist particular requirements for the network performance. For instance, to monitor the status information in the normal running and trigger emergence stop when accident happens, the analysis of the real-time information uploaded from the manufacturing field depend much on the reliable transmission performance of the network in the production system. Besides the consideration of security monitoring, the motion control system and the virtual simulation system also rely much on the real-time signals received from the production devices. Additionally, the organization mode of distributed and parallel production demands much more frequent interaction between different workshops as well as factories, and the manufacturing mode of customization and individuation demands much more intimate collaborative cooperation between different departments in enterprises, which can generate enormous amount of data flow in the industrial production network system.

The switched Ethernet has been applied widely in the industrial control field thanks to its distinguishing features of the acceptable fault tolerance, high communication speed and broadcast storm restrict [4]-[6]. But in the design procedure, the designers are usually not concerned with the topology structure of the switched Ethernet, and in the following deployment and extension procedure, from the subconscious the engineers are apt to add the new network nodes to the existing nodes with a high connectivity under the influence of the Matthew effect. This kind of network construction mode neglects the consideration of the network performance, which results in high time delay, transmission speed reduction, and information flow blockage, and then reduces the stability and reliability of horizontal device interconnection and vertical networking integration. Therefore, the multiple objectives optimization of switched industrial Ethernet topology structure should be given higher priority, which is the research focus of this paper.

Although many remarkable achievements have been made, the time delay and sub-network transmission balance problem is still a crucial issue because the current work mainly depends on the hierarchical topology model with Fieldbus technology. To solve this problem, we investigate the topology structure composed of industrial switches and connection devices and the integral structure is taken into the optimization model. The contributions of this paper are summarized as follows:

1) We present a more flexible network topology model compared with the hierarchical model, which can better reflect the actual network structure in the industrial IoT. According to this topology model, the mapping rules of the network topology structure are proposed and the corresponding theorems and corollaries are proved.

2) Based on the adjacency mapping matrix and the communication properties under the industrial Ethernet

*Corresponding author.
protocol PROFINET, the multi-objective optimization model is presented. In addition, the fault tolerance strategies on the basis of optimization are presented.

(3) We propose a multi-objective optimization algorithm based on the characteristics of mapping matrix and NSGA-II. In this algorithm, an asexual crossover and double swap mutation operation is proposed to achieve the population evolution, which guarantees the offspring can meet the constraint requirements. The benchmark test shows that the proposed algorithm can obtain superior optimization solutions, and the case verification shows that the performance of the network is evidently improved after the topology optimization compared with randomly generated network.

The rest of this article is given below. In section 2, the research status of the network especially the switched industrial network is introduced. In section 3, the incidence relation between the mapping matrix and the actual network topology is proposed, and the special characteristics of the mapping matrix are presented and proved. Section 4 establishes the optimization model of the topology problem and section 5 proposes the multi-objective optimization algorithm based on NSGA-II and the mapping matrix. The efficiency of the proposed algorithm is tested by simulation experiment in section 6 and section 7 gives the final conclusions.

II. RELATED WORK

The classic algorithms such as integer programming, branch-and-bound algorithm can achieve the global optimal solutions, but as a kind of NP-hard problem, the network topology optimization problem needs exponential time to be solved by using exact algorithms. Garroppo et al. [7] assessed the mixed integer programming models under different green network topology scenarios and the solutions obtained can satisfy the QoS requirements. Farvaresh and Sepehri [8] applied the classic branch and bound algorithm to achieve global optimal solutions in Bi-level discrete network design problem and a lower bound for the upper-level objective was developed. Humpola et al. [9] proposed a mixed-integer nonlinear formulation to solve the topology optimization problem in gas network design by using valid inequalities based on the directed graph theory, which can speed up the computation processes on average by 35%. To reduce the computation cost and extend the authenticity of mathematical model, approximate algorithms such as the metaheuristic methods are applied in the network design problem which may not get the global results but can satisfy the engineering requirements. Evolutionary algorithms such as GA (Genetic Algorithm) [10], Differential Evolution [11] and swarm-intelligence-based algorithms such as PSO (Particle Swarm Optimization) [12], Ant Colony Optimization [13], Cuckoo Search [14] have been widely utilized in network design and optimization problems, but unfortunately these methods cannot be applied directly in the industrial network topology optimization problem because of the special optimization characteristics of the industrial Ethernet topology architecture.

In order to improve the network performance of the switched industrial Ethernet, researchers have made many attempts. Georges et al. [15] presented a GA based method to minimize end-to-end delays by designing a switched Ethernet architecture and distributing the industrial devices on the network switches. The objective function was defined by a deterministic network calculus theory and enables to ascertain the bounded delays. However their studies may be more reasonable if an encoding representation method with different quantities of devices connected to the federative switches had been utilized to build up the relationship between the chromosome and the network topology. With the proliferation of hierarchical network connection format in the production field, Zhang and Zhang [16] proposed a hierarchical distribution model of the industrial Ethernet application and described the multi-objective optimization problem in the network partition: the internetwork communication reduction and the network traffic balance. On the communication characteristics of industrial control network, the GA was proposed to search optimal solution of the optimization problem.

After the industrial Ethernet topology model was proposed by Zhang and Zhang, other researchers also proposed some algorithms based on the master-slave form of the Ethernet network to solve the multi-objective optimization problem. Carro-Calvo et al. [17] presented a novel GA and switch-device encoding method to solve the industrial Ethernet network partition problem. The good performance of the approach was proved by the comparison with many genetic operators. But the topology model was based on the two-level industrial Ethernet, so it was not so flexible and structure was relatively fixed with a top-level switch and slave switches. Kim et al. [18] proposed a Nash GA for solving a hierarchical spanning tree network problem. By finding an optimal configuration of backbone network, the proposed algorithm based on Nash game could be employed in designing the backbone topology in a hierarchical link routing domain. Zhou et al. [19] discussed the methods of the network control system based on switched Ethernet in an industrial context. Through the analysis of the optimization criteria and the definition of performance parameters, an arena algorithm was proposed to solve the multi-objective optimization problem of switched Ethernet.

Other researchers [20], [21], also have made some attempts to improve the performance of the industrial network from different perspectives, but till now the transmission mechanism has changed a lot with the development of the communication protocols, so the time delay model of the switches should be updated. Besides, the topology structure is much more flexible and the connection model is proved to represent the trend of flat organization rather than the former hierarchical structure. Furthermore, for the convenience of the calculation and processing, the strict mathematical model should also be established because the network topology of current industrial switched Ethernet is large scale with massive devices rather than small scale local distribution.

III. NETWORK TOPOLOGY MODEL AND CORRESPONDING MAPPING MATRIX

According to the Ethernet features, the network nodes can be divided into two kinds: the branch nodes, i.e., the switches and
the leaf nodes. For instance, Fig. 1 shows an actual topology condition of the industrial Ethernet. The switches constitute the backbone (surrounded by dotted line) of the network and play the role of store and forward function of the information transmission. The PLC (Programmable Logic Controller), I/O device, industrial robot, driver, CNC, PC, IPC (Industrial Personal Computer), RFID sensor, HMI (Human Machine Interface), and industrial camera constitute the leaf nodes of the network. Through corresponding interfaces of hardware and software these devices can achieve the information transmission between the machines as well as men and the machines. Compared with the previous strict hierarchical master-slave connection format of the Ethernet network, the current relationship between the devices in the network are more likely the service provider and the consumer, and there exist no central switches in the decentralized net structure and more likely the service provider and the consumer, and there exist no central switches in the decentralized net structure and every switch can be directly connected with any workstation. The transmission between the machines as well as men and the software these devices can achieve the information transmission. The PLC (Programmable Logic Controller), I/O devices, the role of store and forward function of the information backbone (surrounded by dotted line) of the network and play a significant rate which makes new demands on network transmission performance of the switched industrial Ethernet.

Fig. 1 An actual network topology of switched industrial Ethernet

These nodes and the edges between them constitute the undirected graph \( V(G, E) \), in which \( G \) means the node set and \( E \) means the edge set. For the convenience of the following mathematical processing of the topology optimization, the transmission load and the undirected graph \( V(G, E) \) should be abstractly described.

A. Mapping rule of network topology

For the industrial Ethernet network, the majority of the information transmission is periodic and the process data size can be estimated in unit time. The transmission load between leaf nodes and the connection condition in \( V(G, E) \) can be elaborated by the mapping matrix.

1) Communication mapping matrix of leaf nodes

Supposing there are \( n \) leaf nodes in the industrial network, the communication mapping matrix \( A \) can be defined as follows:

\[
A = \begin{bmatrix}
0 & \cdots & a_{1n} & \cdots & a_{kn} \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
\vdots & \ddots & 0 & \cdots & a_{ln} \\
0 & \cdots & a_{1l} & \cdots & 0 \\
a_{n1} & \cdots & a_{nl} & \cdots & 0
\end{bmatrix}
\]

in which \( a_{ij} \) means the directed communication quantity (it can be regarded as a data size with the unit Kbps) of the process data from the originating node \( i \) to the target node \( j \). For the pure upstream/downstream nodes such as data acquisition devices, the corresponding elements in the mapping matrix can be defined as unidirectional, i.e., \( a_{ij} \neq 0, a_{ji} = 0 \). For the bidirectional communication nodes such as PLC, smart camera and RFID (Radio Frequency Identification) reader, the position elements can be defined as \( a_{ij} \neq a_{ji} \) because generally the communication quantity between node pair differs along different orientations. The diagonal elements are defined as 0, which means the node has no communication demand with itself.

2) Adjacency mapping matrix of \( V \)

When there is an edge between a node pair, i.e., the nodes are connected with each other, the position element is defined as \( X_{kl} = 1 \), otherwise \( X_{kl} = 0 \). Supposing there are \( N \) switches as the branch nodes in the network, then there may be edges established by the physical medium and communication protocols between the \( N \) switches themselves, as well as the \( N \) switches and the \( n \) leaf nodes. Different from the routers used in Internet, the industrial switches are usually utilized as a network node connected directly with the control device rather just a transition node. And owning to the control requirement in the industrial field \( n \) leaf nodes cannot be connected with each other directly. Because the graph \( V(G, E) \) is undirected, the connection edge from node \( k \) to node \( l \) is also the edge from node \( l \) to node \( k \), i.e., \( X_{kl} = X_{lk} \). The graph \( V \) has no loop inside, the leaf node can only be connected with some branch nodes and the connection degree of the node itself is defined as 0, i.e., \( X_{kk} = 0 \). Therefore, the element in the mapping matrix can be defined as follows:

\[
X_{kl} = \begin{cases} 
1, & k \sim l \& k \neq l \& (1 \leq k \leq N \mid 1 \leq l \leq N) \\
0, & \text{otherwise}
\end{cases}
\]

in which \( k \sim l \) means node \( k \) is connected with node \( l \) directly.

Then the entire mapping adjacency matrix can be obtained:

\[
X = \begin{bmatrix}
X_{1N} & \cdots & \cdots & X_{1(N+n)} \\
\ddots & \ddots & \ddots & \vdots \\
\vdots & \ddots & 0 & \cdots \\
X_{(N+n)1} & \cdots & \cdots & X_{(N+n)(N+n)}
\end{bmatrix}
\]
Obviously $X$ is the mapping matrix of tree structure $V(G,E)$ in graph theory. For this particular tree structure, $G$ contains $N$ internal vertices and $n$ leaves, and $E$ contains $N+n-1$ edges respectively. Due to the connection particularity of the industrial Ethernet, the above mapping matrix has some special characteristics compared with general adjacency matrix.

**B. Theorem and corollary proof**

**Theorem:** If $X^d_{ki} = 1$, then the position element $A, X^{cd}_{ki} = 0$; $B, X^{d+1}_{ki} = 0$; $C, X^{d+2}_{ki} > 1$, in which $X^d$ means $d$-order power of the mapping matrix $X$, $k$ and $l$ are the corresponding element positions and $k \neq l$.

**Proof:** For conclusion A, proof by contradiction. If $X^d_{kl} = 1$, supposing the position element $X^{cd}_{kl} \neq 0$, then $X^{cd}_{kl} > 1$ can be concluded under Theorem 3 in the Appendix. $X^d_{kl} = \sum_{i=1}^{\text{N+n}} X^{d-1}_{kl} X_{il}$, the distance from the node $i$ which is connected directly with node $l$, to the node $k$ is $d-1$ or $d+1$. For the node with distance $d+1, X^{d+1}_{kl} = 1$, then according to the above assumption, $X^{d-1}_{kl} > 1$, which leads to contradiction $X^{d-1}_{kl} \neq 0$. Therefore the hypothesis is not correct and the conclusion A is proved.

For conclusion B, $X^{d+1}_{kl} = \sum_{i=1}^{\text{N+n}} X^{d-1}_{kl} X_{il}$, the node $i$ which is directly connected with the node $l$, must be on the extension line of path $k-l$ or on the path $k-l$ path with the distance $d$. For the node $i$ on the extension line, $X^{d+2}_{kl} = 1$, then $X^{d+2}_{kl} = 0$ can be got according to the above Theorem A. As a result $X^{d+1}_{kl} = \sum_{i=1}^{\text{N+n}} X^{d-1}_{kl} X_{il}$ can be simplified to $X^{d+1}_{kl} = X^{d}_{kl}$, in which $i$ is the only node directly connected with the node $l$ on the path $k-l$ i.e., the node $i$ is the only node directly connected with the node $l$ on the path $k-l$. Therefore the hypothesis is not correct and the conclusion B is proved.

For conclusion C, $X^{d+2}_{kl} = \sum_{i=1}^{\text{N+n}} X^{d+1}_{kl} X_{il}$. For a certain node $i$ on the extension line of path $k-l$, $X^{d+1}_{kl} = 1$, so $X^{d+2}_{kl} > 1$ can be drawn. But according to Theorem 3 in the Appendix, $X^{d+2}_{kl} = 1$, therefore the position element $X^{d+2}_{kl} > 1$, proved.

**Corollary 1:** If $X^d_{kl} = 0$, then the position element $A, X^{d+odd}_{kl} = 0$; $B, X^{d+odd+1}_{kl} > 1$, in which $k \neq l$ and $odd$ is an odd number.

**Proof:** Using mathematical induction. When $odd = 1$, the conclusion is correct according to aforementioned Theorem. Supposing there exists an odd number $odd$ which makes $X^{d+odd}_{kl} = 0, X^{d+odd+1}_{kl} > 1$ when $X^d_{kl} = 1$. Now consider the condition of $odd + 2$ and $odd + 3$.

$X^{d+odd+2}_{kl} = \sum_{i=1}^{\text{N+n}} X^{d+odd+1}_{kl} X_{il}$. $i$ is still the node directly connected with the node $l$ on the extension line of path $k-l$ or on the path $k-l$. For the node $i$ on the extension line, $X^{d+odd+1}_{kl} = 1$, the conclusion $X^{d+odd+1}_{kl} = 0$ can be drawn on the basis of the assumption. And there must exist only a node $i$ on the path $k-l$, $X^d_{kl} = 1$. Then the equation $X^{d+odd+2}_{kl} = \sum_{i=1}^{\text{N+n}} X^{d+odd+1}_{kl} X_{il}$ can be simplified to $X^{d+odd+2}_{kl} = X^{d+odd+1}_{kl}$. Homoplastically, $X^{d+odd+1}_{kl} = X^{d+odd}_{kl}$, in which $i$ is the only node directly connected with node $i$ on the path $k-i$. In the same way, $X^{d+odd+2}_{kl} = X^{d+odd+1}_{kl} X_{ik}$, in which $q$ is the node directly connected with node $k$, i.e., $X_{ik} = 1$. According to the above assumption, $X^{d+odd+1}_{kl} = 0, so X^{d+odd+2}_{kl} = 0$.

**Corollary 2:** The diagonal position elements in the odd-order power of $X$ are 0; the diagonal position elements in the even-order (above 3) power of $X$ are more than 1, in particular, the two-order power condition is aforementioned in Theorem 1 in the Appendix.

**Proof:** $X^{odd}_{kk} = \sum_{i=1}^{\text{N+n}} X^{odd-1}_{kl} X_{ik}$, in which $i$ is the node directly connected with node $k$. $X^{odd-1}_{kl} = 0$ can be drawn according to the Corollary 1, so $X^{odd}_{kk} = 0$.

**IV. MULTI-OBJECTIVE OPTIMIZATION MODEL FORMULATION**

The mathematical model of multi-objective optimization can be summarized as follows:

Minimize $F(X) = \left[f_1(X), f_2(X), ..., f_u(X)\right]$  \hspace{1cm} (4)

Subject to $X \in \left\{ h(X) = 0, g(X) \leq 0 \right\}$  \hspace{1cm} (5)

where $F(X)$ means the fitness function set and $u$ is the number of the objectives. $X$ means the solution space. $h(X)$ means the strong constraint set and $g(X)$ means the weak constraint set.

**A. Optimization criteria**

In the Ethernet the switches take the responsibility of information flow transmission based on the MAC and IP identification between different sub-networks. In the industry area there are two main types of switch: the cut-through switch and the store-forward switch. For the cut-through switch the data frame is directly sent to the next interconnected node without any error information verification, and as a result the transmission speed is fast but there may be some mistakes in the data packet. Contrarily, the store-forward switch stores the data frame when it receives the send request and then checks if there exists any error inside, so correspondingly the transmission speed is slowed down but the information correctness is guaranteed which is of prime importance for the industrial applications. In consideration of the security and safety in the industrial production, store-forward switch is chosen as the connection node of the industrial Ethernet.

The switched Ethernet network plays the role of the information transmission bridge between the real physical world and the virtual cyber world. For the real-time monitoring, timely management and conformable logistic scheduling, to reduce time delay of the information transmission is the top priority of the network topology optimization. The information
transmission time through the network is determined mainly by three parts: physical medium transmission time, switch mechanism and transport protocol. The transmission time of the physical medium such as optical fiber and shielded twisted pair is approaching the speed of light, so the medium transmission time can be ignored. With regard to the transport protocol, PROFINET is taken into consideration [22]-[25]. PROFINET is an open standard based on industrial Ethernet proposed by the international organization PROFIBUS&PROFINET. As a kind of real time Ethernet, PROFINET comprises TCP/IP and IT standard, and it can seamlessly integrate current different field buses. The predominated performance make PROFINET the mostly applied protocol in the industrial control field.

For the information transmission through the switches based on PROFINET, when the RT (Real Time) data gets stuck by the NRT (Not Real Time) data at the ith switch, then the entire time delay of RT data packet at the Nth switch on the transmission path in the network is as follows:

\[
\begin{align*}
    t_{RT-NRT}(n) & = \sum_{n=1}^{N} t_{i}^{RT-NRT}(n) \\
    & = i \cdot t_{i}^{RT} + t_{queue-RT(i)} + (N - i) \cdot t_{send-NRT}
\end{align*}
\]

in which \(t_{i}^{RT} = t_{s&f-RT} + t_{trans-RT}\) is the store&forward and transmission time of RT data packet through a switch. \(t_{queue-RT(i)}\) means the queuing delay of RT data packet at the ith switch. \(t_{send-NRT} = t_{s&f-NRT} + t_{trans-NRT}\) means the store&forward and transmission time of NRT data packet through a switch, which is proportional to the size of the data frame. \(t_{i}^{RT-NRT}(n)\) means the time delay of the RT data packet at the nth switch because of the NRT data packet blocking.

Now the worst condition is taken into consideration, i.e., the data blocking appears at the first switch node, then:

\[
\begin{align*}
    t_{RT-NRT}^{N} & = t_{RT}^{1} + t_{queue-RT(1)} + (N - 1) \cdot t_{send-NRT} \\
    & = N \cdot t_{send-NRT} + (t_{s&f-RT} + t_{trans-RT} + t_{queue-RT(1)} - t_{s&f-NRT} - t_{trans-NRT}) \\
    & = N \cdot t_{send-NRT} + (t_{queue-RT(1)} - t_{queue-NRT})
\end{align*}
\]

in which \(t_{queue-RT(1)}\) means the queuing time of RT data packet at the first switch node and \(t_{queue-RT}\) means the queuing time of RT data packet at the following switch nodes. According to the experiment test [26], generally \(0 < (t_{queue-RT(1)} - t_{queue-RT}) \ll t_{trans-NRT}\), so the following approximate equation of the RT data time delay can be drawn:

\[
    t_{RT-NRT}^{N} \approx N \cdot t_{send-NRT}
\]

That is to say, the transmission time delay of the real time data packet in the Ethernet network is proportional to the packet size of the information and the quantity of the switches between the target nodes. Corresponding measures should be carried out to restrict the maximum transmission time delay in the whole network. Furthermore, the production devices of IoT in industrial site should gain access to the industrial control and management network on demand, so considering the scalability of the network, the information load of every switch should be balanced.

As a consequence, the optimization target of network topology can be summarized as follows:

- To insure the real time on-site control capability, the leaf nodes which interact frequently with each other should be under the same switch and the whole transmission load of large data packets between switches should be reduced.
- Balance the information transmission load of each sub-network to insure the scalability and reconfiguration of the network.

B. Mathematical formulation of the optimization problem

Because of the special connection requirements of the industrial Ethernet topology, the longest path must lie between the leaf nodes. According to Theorem 3 in the Appendix, the maximum path length (diameter) \(D\) is as follows:

\[
    D = \{\max(d) \mid \exists X^k_{il} = 1, \ N + 1 \leq k < l \leq N + n\} \quad (9)
\]

The information transmission load in the network after \(m\) switches is as follows:

\[
    \omega_m = (m - 1) \cdot \sum_{i=1}^{n} B^{m+1}_{ij} \cdot a_{ij} \quad (10)
\]

in which \(1 \leq m \leq D - 1, i,j = 1,2,...,n\). According to Corollary 1, if \(X^{(m+1)}_{(N+i)(N+j)} = 1\), then \(B^{(m+1)}_{ij} = 1\), and if else, \(B^{(m+1)}_{ij} = 0\).

Then the fitness function of transmission load between sub-networks in the whole network is as follows:

\[
    f_1 = \sum_{m=1}^{D-1} \sum_{i=1,j=1}^{n} (m - 1) \cdot B^{m+1}_{ij} \cdot a_{ij} \quad (11)
\]

With regard to the transmission load of each switch, if both leaf nodes are linked with the same switch, then the path length is 2; if only one leaf node exist under the switch, then the path length is 1. Hence the transmission load of each switch can be defined as follows:

\[
    \omega(s) = \sum_{i=1,j=1}^{n} (X_{s(N+i)} \cdot a_{ij}) \quad (12)
\]

in which \(s = \{1,2,...,N\}\). Then the fitness function of the transmission load difference between switches can be defined as follows:

\[
    f_2 = \sum_{s=2}^{N} |\omega(s) - \omega(s - 1)| \quad (13)
\]

Distinct from the loosely connected Internet, there are relatively strict requirements in the industrial Ethernet network. Strong constraint: There is no isolated leaf node in the network. In the branch node mapping area of \(X\) the quantity of 1 is \(N - 1\), and in the leaf node mapping area the quantity of 1 is \(n\).

Weak constraint:
• Every branch node should be connected with at least one leaf node.
• The maximum connection degree of branch node is limited, here assumed as 8, i.e., every switch has 8 ports.
• The interconnection of the leaf nodes. According to Corollary 1 and Corollary 2, easy to know the necessary and sufficient conditions of interconnection is as follows: 
\[ \forall (X_{ki}^D - 1 + X_{kl}^D) \geq 1, N + 1 \leq k < l \leq N + n. \]

C. Fault tolerance strategy

For industrial environment, fault tolerance of Ethernet is very important and the network survivability depends on amount of available resources such as switches or physical connections (copper or fiber cable). In order to prevent network paralysis caused by the failure of network devices, besides improving the stability and reliability of the devices, two kinds of strategies are taken into consideration: the device redundancy and the path redundancy.

• The device redundancy mainly refers to the hot standby of the hardware and software. Once some failures of the main devices appear, the redundancy system can run just in time to ensure the normal operation of the network. This kind of parallel redundancy strategy can support status switching without time delay, but the device costs are usually so high that the device redundancy strategy is deployed at the positions which have direct and important influence on the performance of the network system. Here the node with maximum connection degree obtained according to the Theorem 1 in the Appendix can be chosen as the deployment position;

• The path redundancy strategy, represented by the ring redundancy should be applied in the topology model. Because usually it is unable to predict which network device will have failure, the maximum distance, i.e., the transmission path with the longest length can be chosen to set up ring redundancy link. For any certain network structure established according to the Pareto frontier, the maximum path length \( D \) can be obtained based on equation (9), and then both end switches on the maximum path are connected with each other to form a ring redundancy structure. Once some node failure occurs in industry site, the network can rapidly achieve reconfiguration according to the redundancy mechanism of transmission protocol, so the network can still keep running with performance loss until the failure is removed.

V. MAPPING MATRIX BASED MULTI-OBJECTIVE NSGA-II ALGORITHM DESIGN

NSGA-II which was proposed by Deb K etc. in 2002 [27], has solved the bottleneck problem in the application of NSGA-I and been applied in many multi-objective optimization problems [28]-[30]. By computation simulation, NSGA-II was proved that it can approach the final Pareto solutions with higher distribution and better convergence compared with other multi-objective evolutionary algorithms such as SPEA (Strength Pareto Evolutionary Algorithm), PAES (Pareto Archived Evolutionary Strategy). NSGA-II utilizes the traditional float coding method to achieve the multi-objective solutions, which is usually used to solve the problems with continuous search space and cannot solve the highly discrete optimization problem of network topology, so the standard NSGA-II should be revised. The proposed algorithm utilizes the non-dominating sorting and crowding distance assignment processes of NSGA-II, but compared with NSGA-II, our proposed algorithm mainly focuses on the evolutionary processes of the population such as the initialization, the crossover operation, the mutation operation of the branch nodes and leaf nodes. These evolutionary mechanisms are presented aiming at solving this specific optimization problem of network topology and can ensure the smooth implementation of the population evolution. Based on several special characteristics of the mapping matrix proved through rigorous mathematical derivation and the evolutionary processes, the Pareto frontier is achieved finally. The complete flow chart of proposed algorithm is as shown in Fig. 2, and it will be described with details in the following part of this section. In Fig. 2, the parameter \( GEN \) means the total iteration times and the parameter \( Gen \) means the current iteration time. Once the current iteration time reaches the predetermined termination condition \( GEN \), then the whole program ends and returns the final results.

![Flow chart of proposed algorithm](image)

Fig. 2 Flow chart of proposed algorithm

A. Original population initialization

The population initialization is to initialize the original mapping matrix of the Ethernet network. In accordance with the structure features of scale-free network under the influence of the Matthew effect, the following constructing algorithm is applied to build the original network up.

For the branch nodes:

**Step 1:** Two switch nodes are chosen randomly to be the original network topology nodes;

**Step 2:** Every time a new node is added to the existing branch node whose connection degree is still in the limitation range, i.e., less than 8. The selection of the existing nodes is based on the selection probability rather than randomly. Supposing the quantity of current nodes is \( Num \), then the selection probability of node \( i \) is \( X^2_{ii} / \sum_{k=1}^{Num} X^2_{kk} \);

**Step 3:** If the sum of nodes has not reached \( N \), go back to step 2.

For the leaf nodes: The leaf nodes are connected to the
existing branch nodes as attachments. Every time a new leaf node is added to the branch node whose connection degree is less than 8. Supposing the quantity of current nodes is \( N + n_{\text{num}} \), then the selection probability of branch node \( j \) is \( X_{jj}^2 / \sum_{k=1}^{N+n_{\text{num}}} X_{kk}^2 \). Repeat the above process until the sum of nodes reaches \( N + n \).

Fig. 3 An example of the original network initialization

For example, Fig. 3 shows a concrete initialization process of a new network. When a new branch node is added to the existing switch backbone network, it is highly likely to be connected to the node \( i \) or \( i' \) because these two nodes have the same maximum connection degree 3. Then after several similar operations the original network \( a \) is constructed as the network \( b \) shown in Fig. 3. When a leaf node is added to the current network, it is most likely to be connected to the node \( j \) and \( j' \), because these two branch nodes have the maximum connection degree 4.

After the network construction is finished by the connection of the nodes, the restriction verification should be carried out. If the topology structure does not satisfy the actual connection demands, rebuild the network up until the qualified chromosome complex with \( P \) individuals is generated.

**B. Non-dominated sorting and crowding distance assignment**

The standard non-dominated sorting and crowding distance assignment introduced in NSGA-II is utilized to sort the two-objective solution set. Firstly solution set is sorted by the non-dominated degree of the two-objective fitness functions. If some solution vectors have the same non-dominated degree, then calculate the crowding distance and sort the solution vectors again. The sorting procedure is introduced in detail in the reference [27].

**C. Tournament selection**

As the selection target is to achieve the chromosomes with minor non-dominated degrees, the traditional roulette method is not applicable. Here the tournament selection method is applied to perform the selection operation. Firstly the chromosomes with lower degrees are selected as the winners in the tournament. If the chromosomes have the same non-dominated degree, then the chromosomes with larger crowding distance are selected as the winners. Each time a chromosome with good genetic traits is selected as the parent generation until the mating pool is filled.

**D. Crossover and mutation**

1) **Asexual crossover operator**

The optimized solution is a 0-1 adjacency matrix which has obvious discrete characteristic. Due to the particularity of the multi-dimensional space solution, the offspring generated by traditional binary crossover between the chromosomes usually cannot satisfy the constraint requirements. Therefore, combining the special characteristics of Ethernet topology structure, an asexual crossover operation is proposed to get the next generations: For \( i, j \leq N, X_{ij}^a \) and \( X_{ij}^b \), as well as \( X_{il}^a \) and \( X_{lj}^b \) interchange directly with each other, in which \( l = 1, ..., N, i \neq j, i \neq l, j \neq i ; \) for \( N < i, j \leq N + n \), \( X_{il}^a \) and \( X_{ij}^b \), as well as \( X_{il}^b \) and \( X_{lj}^a \) interchange directly with each other, in which \( l = 1, ..., N, i \neq j \). That is to say, two different nodes (branch nodes or leaf nodes) are selected randomly and then the position crossover operation is implemented. This kind of crossover operation is implemented inside uni-chromosome rather than the conventional crossover operation between two parent chromosomes, so it is called single-parent asexual crossover.

2) **Double swap mutation operator**

If the traditional binary bit-flip mutation operation between two chromosomes is applied as mutation strategy in the genetic process, the offspring generation is unavailable because the offspring individuals cannot satisfy the constraint conditions. Therefore the double swap mutation operation is proposed to implement the mutation process. The operation procedure is as follows in detail: for the connected directly branch nodes \( i \) and \( j \), i.e., \( i, j \leq N \), let \( X_{ij} = 0 \), and then select a node \( k \) which is not connected with \( i \), let \( X_{ik} = 1 \). In other words, two coupling chain appear along with the disconnection of nodes \( i \) and \( j \), and then within the constraint ranges, the node \( i \) of one chain is connected again with a randomly selected node \( k \) of the other chain. For a leaf node \( i \), it must be connected with a branch node \( j \) according to Theorem 2 in the Appendix. The node \( i \) is disconnected and then connected with a new branch node within the constraint range.

Fig. 4 shows a concrete example of the asexual crossover operation and double mutation operation. For the crossover operator, the randomly selected branch nodes 1 and 2 transpose the position with each other, and the randomly selected leaf nodes 3 and 4 transpose the position with each other. After the crossover operation, the randomly selected edge between the branch nodes 5 and 6 is disconnected and then the node 6 with a coupling chain is connected to the randomly selected branch node 7. Similarly, the randomly selected edge between the leaf node 8 and the branch node 9 is disconnected and then the leaf node 8 is connected with randomly selected branch node 10.
After the above crossover and mutation operation, the network is transformed into the newly generated topology model shown in Fig. 4.

Fig. 4  An example of the crossover and mutation operation

For the parent chromosomes with non-dominated degree $\text{deg}$, the above asexual crossover and double swap mutation strategy should be implemented for $\text{deg} \times (\text{floor}(\log_{10}(\text{GEN} - \text{Gen})) + 1)$ times, so that the chromosomes can make adaptive adjustment corresponding to the non-dominated degree in the genetic procedure, and the convergence speed is accelerated with the population diversity guaranteed.

E. Elite reservation

In the crossover and mutation operation, the entire parent individuals must carry out the asexual crossover and double swap mutation operation. Some measures of elite reservation strategy should be taken to prevent the next generation individuals with good traits from being replaced in the genetic process of the whole population. The elite chromosome selection strategy is as follows: The offspring generation $S$ generated in the mating pool is combined with the current chromosomes $P$ and then all of them should be sorted by the non-dominated degree and the crowding distance. Then $P$ chromosomes with lower non-dominated degree are selected and reserved as the elite individuals. If the non-dominated degree is the same, then the chromosome with larger crowding distance is selected.

F. Complexity analysis

The complexity analysis of proposed algorithm is mainly divided into two parts: the time complexity and space complexity. For the time complexity, the most time consuming part is the fitness function calculation process in the main loop of the algorithm because this process contains the matrix operation and the non-dominated sorting. To calculate the optimization objectives, the time complexity is $T_1 = O(SNn(N + n)^4)$, in which $S$ is the offspring quantity of the mating pool, $N$ and $n$ are the quantities of branch nodes and leaf nodes respectively. The complexity of the non-dominated sorting is $O(u(P + S)^2)$, and the complexity of the crowding distance assignment is $O(u(P + S)\log(P + S))$, so the whole time complexity of sorting part is $T_2 = O(u(P + S)^2)$, in which $u$ is the optimization objective quantity and $P$ is the population quantity. In the worst-case scenario, $T_1$ consumes more computation time than $T_2$, so the whole time complexity of proposed algorithm is $O(SNn(N + n)^4)$.

For the space complexity, $O(S(N + n)^3)$ computation storage is required for the matrix operation and the sorting storage requirement is $O(S^2(N + n)^2)$. Therefore for a network with small-scale nodes, the whole space complexity of proposed algorithm is $O(S^2(N + n)^2)$, and for a network with large-scale nodes, the space complexity is $O(S(N + n)^3)$.

VI. SIMULATION RESULTS AND DISCUSSION

A. Benchmark test

In view of the complexity of the network topology problem, there are no theoretical optimal solutions for a similar concrete example jet so that it is hard to find out the exact optimization gap between the results achieved by proposed method and the optimal solutions. But to verify the efficiency of proposed algorithm, a benchmark test is carried out by using the industrial network optimization case presented by Zhang and Zhang [16], though our presented mathematical model can more accurately reflect the real network optimization problem of the switched industrial Ethernet. Supposing there exit 4 switches, 40 nodes of the process data, and the communication matrix between leaf nodes in [16] is applied. The generation quantity $P$ is assumed as 100, iteration times as 1000, tournament scale as 2, the mating pool capacity as 100, crossover and mutation rate as 0.9 respectively. Fig. 5 displays the achieved Pareto results by proposed algorithm.

Carro-Calvo et al. [17] updated the solutions got by Zhang and Zhang, and according to the Pareto non-dominated solution definition the final solution sets are as follows: $S_1$: $f_1 = 1022$, $f_2 = 21$; $S_2$: $f_1 = 1046$, $f_2 = 7$. The Pareto frontier solutions got by the proposed algorithm in Fig. 5 is as follows in Table I.

<table>
<thead>
<tr>
<th>TABLE I</th>
<th>OPTIMIZED RESULTS OBTAINED BY PROPOSED ALGORITHM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Obtained Solutions</td>
</tr>
<tr>
<td></td>
<td>$f_1$</td>
</tr>
<tr>
<td></td>
<td>$f_1$</td>
</tr>
<tr>
<td></td>
<td>$f_1$</td>
</tr>
</tbody>
</table>

Obviously the solutions obtained by the proposed algorithm can dominate the results got by Zhang and Zhang [16] and Carro-Calvo et al. [17] weakly. Furthermore from Fig. 5 we can see that the proposed algorithm can obtain relatively complete Pareto frontier compared with existing results, and the solutions
not only locate in a local area but has a much wider distribution. The distribution range of objective function \( f_1 \) is from 889 to 1037 and \( f_2 \) is from 2 to 641, which brings more topology choices for the engineering network establishment directly. This benchmark test verifies the usability of proposed algorithm.

**B. Case verification**

Supposing there are 30 switches and 100 process data transmission leaf nodes. The communication matrix is a random generated matrix with the transmission packet size between 0 and 4096 Bytes. The generation quantity \( P \) is assumed as 200, iteration times as 1000, tournament scale as 2, the mating pool capacity as 200, crossover and mutation rate as 0.9 respectively.

For more clarity, Fig. 7 displays the detail information of the obtained optimal Pareto frontier of the two-objective fitness functions. Because of the highly discrete characteristic of the delay parameter in the switched Ethernet network, the Pareto frontier demonstrates a degree of clustering behavior. The fitness function \( f_1 \) can make sure that the data packets of NRT are delivered mainly in sub-networks and the transmission load between sub-networks are restricted to a certain value. The fitness function \( f_2 \) is used to balance the workload between sub-networks under different switches so that the switches in the Ethernet topology structure are neither overloaded nor relatively idle. The optimal results with a degree of clustering behavior around the Pareto frontier help make the performance selection of the switched Ethernet topology possible at the industry engineering site. If the time delay reduction is the primary need in the production field such as in the real time control system, then the optimal solutions around point 1 can be selected and the network should be constructed in accordance with corresponding mapping matrix. If the network performance of the information flow balance and good extendibility are emphasized, then the optimal solutions around point 3 can be selected. However, as a kind of heuristic algorithm, the optimal results achieved from the proposed algorithm cannot be determined whether they are the theoretical optimal solutions or not.

**VII. CONCLUSION**

The revised NSGA-II based on mapping matrix can significantly ameliorate the network performance of the switched industrial Ethernet through the special self-adaptive crossover and mutation operations. The transmission load of
the data frames between sub-networks is reduced and the real-time of the network is guaranteed. At the same time, the information transmission load difference between sub-networks is decreased and as a result the information flows under the switches are balanced. Based on the optimal Pareto frontier, the structure designers and network engineers can select the optimal network topology structures in accordance with the actual industrial production demands. However this engineering practical problem is not a standard test problem so there are no theoretical Pareto optimal solutions yet. Besides, the information flow in the network is usually dynamic and unpredictable, so a more accurate assumption of the flow should be proposed to reflect the real network transmission condition, and the topology optimization problem with more than two objectives should be the research emphasis of next stage.

APPENDIX

**Theorem 1:** The sum of the kth row vector or the kth diagonal element of square X is the connection degree of node $k$.

**Proof:** For the square $X, X_{kk} = \sum_{i=1}^{N+n} X_{ik}X_{ik}, X_{ii} = X_{ii}$, so $X_{kk} = \sum_{i=1}^{N+n} X_{ik}X_{ik}$. According to the definition of mapping matrix, if node $k$ is connected with node $i$, the position element $X_{ki} = 1$, otherwise $X_{ki} = 0$. Thus $X_{kk} = \sum_{i=1}^{N+n} X_{ik}X_{ik}$. The latter part of the equation is the sum of the kth row vector, and the diagonal element $X_{kk} = 0$, proven.

**Theorem 2:** In the branch node mapping area of $X$ (surrounded by solid line) the quantity of $I$ is $N - 1$; in the leaf node mapping area of $X$ (surrounded by dotted line) the quantity of $I$ is $n$ and in every column vector only one I exists.

**Proof:** Because $V(G, E)$ is a undirected simple graph, $N$ branch nodes must be connected with each other by $N - 1$ edges; for the $n$ leaf nodes, each node should be connected with a certain branch node.

**Theorem 3:** The necessary and sufficient condition of the position element in d-order power of $X$, i.e., $X_{d,ki} = 1$ is that the distance between node $k$ and node $l$ is $d$.

**Proof:** When $d = 1$, obviously the conclusion is correct. Supposing $d = i$, the conclusion is also correct, now consider the condition of $d = i + 1$. According to $X_{d+1, ki} = \sum_{i=1}^{N+n} X_{d+1,ki}X_{i,l}$, the connection path $k - l$ with the distance $i + 1$ can be considered that the node $k$ is connected with the node $i$ by $d$ edges firstly and then the node $i$ is connected with node $l$.

Necessary condition: If there is a path between node $k$ and node $l$, there must exist only a such node which makes $X_{d, ki} = 1, X_{d, li} = 1$ according to the definition of a simple graph, so $X_{d+1, ki} = 1$.

Sufficient condition: If $X_{d+1, ki} = 1$, because $X_{d+1,ki}$ are integers, there must only a such node which makes $X_{d, ki} = 1, X_{d, li} = 1$, which means the path distance between node $k$ and node $l$ is $i + 1$.

REFERENCES


Jielin Li was born in Shandong Province, China, in 1989. He received the B.S. degree in mechanical design and theory from School of Mechanical Engineering, Tongji University, Shanghai, China, in 2012. Since 2012, he has been pursuing the Ph.D. degree of mechanical manufacturing and automation in School of Mechanical Engineering, Tongji University, Shanghai, China. His research interests include modeling and optimization for smart production system in discrete manufacturing, especially the network topology structure optimization of IoT.

Ming Chen was born in Shanghai, China, in 1964. He received the B.S. degree in engineering machinery, M.S. degree in mechanical design and theory, and Ph.D. degree in mechanical manufacturing and automation from School of Mechanical Engineering, Tongji University, Shanghai, China, in 1987, 1990 and 2006 respectively. Currently he is a professor with Sino-German College of Applied Sciences and the chief of “Industry 4.0”-Smart Plant Lab. His research interests include PLM technology with industrial big data, MES technology with sensor networks, and digital development technology of smart products.