SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

August 2021, Vol. 64 182301:1–182301:14
https://doi.org/10.1007/s11432-020-3016-x

A novel identity resolution system design based on Dual-Chord algorithm for industrial Internet of Things

Renchao XIE[1,2*], Zhiyuan WANG[1], Fei Richard YU[3],
Tao HUANG[1,2] & Yunjie LIU[1,2]

[1]*School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing* 100876*, China;*
[2]*Department of Future Network, Purple Mountain Laboratories, Nanjing* 211111*, China;*
[3]*Department of Systems and Computer Engineering, Carleton University, Ottawa* ON K1S 5B6*, Canada*

Received 20 May 2020/Accepted 23 July 2020/Published online 1 June 2021

Abstract The industrial Internet of Things (IIoT) has attracted enormous attention from academics and industries, which is a significant component of the future transformation of industrial systems. The analysis, understanding, and application of all kinds of fragmented data in IIoT is one of the critical technologies that need to be studied in the future. The identity resolution technology is the most promising technology to achieve this goal at present. Although many identity resolution systems based on domain name system (DNS) have been proposed, there are technical problems with a single root node causing system performance bottlenecks and slow resolution rates. In this paper, we focus on a widely used distributed structured peer-to-peer (P2P) object name system (ONS) and redesign a novel identity resolution system based on a Dual-Chord algorithm for IIoT. The support vector machine (SVM) algorithm is first introduced in the system designed in this paper to implement the clustering of the P2P network. And then, the Dual-Chord algorithm is proposed, which can get better performance than the classical Chord algorithm. This system resolves the mismatch between logical and physical addresses and improves the resolution rate of the system. Finally, through PeerSim network simulation software, this paper evaluates the simulation performance of the improved identity resolution system based on the Dual-Chord algorithm. Simulation results prove that the system proposed in this paper is a more efficient identity resolution system.

Keywords industrial Internet of Things, identity resolution, distributed Hash table, small world model

Citation Xie R C, Wang Z Y, Yu F R, et al. A novel identity resolution system design based on Dual-Chord algorithm for industrial Internet of Things. Sci China Inf Sci, 2021, 64(8): 182301, https://doi.org/10.1007/s11432-020-3016-x

1 Introduction

With the deep integration of new-generation information technology and the manufacturing industry, the industrial Internet of Things (IIoT) is considered to be an essential and important part of the future industrial system transformation [1]. The core of the IIoT is to build a new industrial manufacturing and service system with all elements, all industrial chains, and all value chains by building comprehensive connectivity of people, machines, material to promote a series of new business models and a new round of changes in the global industrial system [2, 3].

Now we have witnessed the increasing use of intelligent devices and communication applications in industrial production monitoring and their impact on industrial design, manufacturing, production, transportation, and sales activities [4]. With the development of the world economy, more and more countries pay attention to IIoT, and expect to use advanced technology to improve the traditional manufacturing industry [5]. One of the key technologies and application directions of IIoT, is the understanding and application of all kinds of split data involved in intelligent manufacturing. It is necessary to know and

* Corresponding author (email: Renchao_xie@bupt.edu.cn)

grasp the source, flow process, and use of the data when trying to understand it. At present, it can be seen that the only hope to meet this task is the identity resolution technology.

The identifier resolution system is an essential hub for the realization of IIoT. It is responsible for the distribution, registration, management, analysis, and routing of the identity of goods and supports the full life cycle management of equipment, people, and materials in IIoT [6]. At present, the mainstream identification technologies include electric protocol code (EPC) [7], entity code for IoT (Ecode) [8], handle [9], which are proposed by different organizations. These identity resolution system's starting point is to mark an object and provide information query and then develop into a kind of underlying information architecture which is as crucial as domain name system (DNS) on the Internet.

Although many DNS-based identity resolution systems have been proposed, for example, EPC, which has been used in many companies and enterprises in the world, also has technical problems with a single root node causing system performance bottlenecks and slow resolution rates [10]. This paper explores the improvement of the identity resolution scheme, and focuses on the EPC coding system based on the EPCglobal structure. Its object name system (ONS) resolution system is more representative in the field of identity resolution, which has a vast item information exchange network worldwide. Therefore, this paper studies the ONS identity resolution system, and deals with these problems in the DNS based ONS resolution system, such as the root server overload and the bottleneck caused by single point failure. This paper proposes to use the peer-to-peer (P2P) paradigm in the ONS resolution system, especially the structured distributed P2P paradigm, which has a better comprehensive performance. The overall contributions of this paper are listed as follows.

- First, we consider a novel identity resolution system. In this system, we use the routing strategy based on node performance and design the quantitative way of node performance. In the network with better performance of upper nodes, we use the bidirectional Chord routing algorithm based on node distance and introduce the index caching mechanism into the network with the general performance of lower nodes, which significantly improves the query performance and stability of the identity resolution system.
- Next, in order to solve the problem of mismatch between physical address and logical address in the proposed architecture, this paper introduces a small world model and support vector machine (SVM) algorithm, realizes the clustering of feature nodes, and solves the problem of node hierarchy in the architecture.
- Finally, extensive simulation results of the Dual-Chord algorithm and Chord algorithm performance are illustrated and analyzed to demonstrate the performance of the proposed system.

The remaining parts of this paper are organized as follows. Related work and existing developments to identity resolution are presented in Section 2. The design principles and algorithms used and the workflow of the system are designed in Section 3. Following that, we use the PeerSim simulation program to simulate and evaluate the performance of the system designed in this paper in Section 4. Finally, we conclude the article in Section 5.

2 Related work

The identity resolution system is an essential hub for the realization of the IIoT, which is responsible for the distribution, registration, management, resolution, and routing of the identity of items. It supports the full life cycle management of equipment, people, and materials in the industrial Internet. Therefore, among the participants of the industrial Internet, identity resolution is paid more and more attention.

In academic circles, many kinds of identity resolution schemes have been proposed. According to their evolution, they can be divided into two categories. One is the improved path based on DNS. Typical representatives are the ONS of EPCglobal organization [7], the object identifier resolution system (ORS) jointly established by International Telecommunication Union (ITU) and International Organization for Standardization (ISO) [11], the uCode resolution system proposed by Japan Internet of Things Application Organization [12] and the Ecode independently developed by China [8]. The other is an innovation path unrelated to DNS, e.g., identifier resolution technology for specific industries of IIoT, such as handle identifier resolution technology for radiofrequency electronic tags maintained by DONA foundation [9, 13, 14], and ubiquitous identifier (UID) resolution technology proposed by Tokyo University [15, 16]. There are still many ORS identity resolution systems under ITU and ISO international organizations, but the relevant ORS coding standards have been published [17, 18].

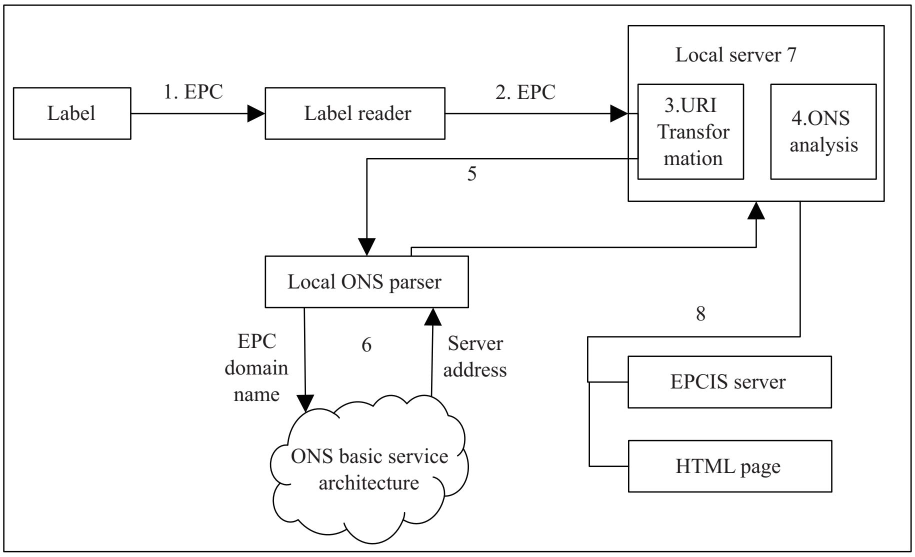

Figure 1 ONS service query flow chart.

Nevertheless, in practical applications of industrial production, the ORS identifier applies only to the electronic communications and medical fields, and Handle only succeeds in product traceability and digital libraries. However, the EPC coding system under the EPCglobal organization is used in many companies and enterprises in the world, so the ONS resolution system in the field of identity resolution is quite representative. There are two kinds of identifier resolution systems of ONS based on the EPC system, i.e., global identity system and local identity system of the company or industry. VeriSign company is responsible for the management of the global identity system. The identity resolution system of the company or industry can use the identity resolution system of the global scope to obtain the object information of the goods and exchange the data flow, which is the most widely used and the most mature operation [10].

At present, ONS mainly uses the existing Internet domain name service (DNS) domain name resolution system to resolve the domain name. First, the electronic label EPC in each item is transformed into the domain name standard format. Then the domain name is resolved to obtain the EPCIS information server address storing the EPC code details. The specific query process is shown in Figure 1. ONS can provide a global resolution and query service like DNS. It resolves the EPC code read by the item tag identifier into one or more resource locators similar to the Internet address. It then further finds out the details of the corresponding item or commodity. Because ONS relies on the domain name resolution system of the Internet to query information, the format of the resource locator of the ONS system is the same as that of the DNS system. The uniform resource locator is used to query the IP address of the EPC information server, i.e., EPCIS, and search the detailed information related to the article. However, it can be seen that the DNS-based ONS resolution system has some problems such as overload of root nodes and bottlenecks caused by single-point failure. Therefore, domestic and foreign research institutions began to use P2P technology in the resolution system, uniquely structured distributed P2P technology. In the Conference of Sensor Networks and Trusted Computing in 2010, Sergei Evdokimov [19] and other experts summarized the implementation principle of five kinds of discovery services (DS) and the advantages and disadvantages of various aspects of performance. They pointed out that the ONS resolution service based on structured distributed P2P technology has a better comprehensive performance. The ONS resolution system based on P2P does not need to convert the EPC code into the corresponding form in DNS, but directly maps the EPC code to the EPCIS information server by Hash operation [20]. At the same time, the P2P-on resolution system cancels the central server, and each server has a relatively balanced load. In general, the P2P-on resolution system dramatically reduces the query delay and improves the query efficiency [21].

3 The design of identity resolution scheme

Structured distributed P2P is the best choice for identity resolution in the IIoT, which has a complete and widely used model system. It does not set up index servers and uses virtual addresses to identify the location of nodes in the network, which overcomes the shortcomings of limited search scope and low scalability in unstructured P2P. It cancels the function of the index server, and the virtual logical address of each node is obtained by Hash processing the node's IP address. Therefore, the virtual logical address of this P2P network is fixed, and the corresponding structured P2P network system's structure is fixed. It does not change with the joining or exiting of nodes. Because node information is stored in the routing table after the Hash operation, a structured distributed P2P network is also called distributed Hash table (DHT) network.

The main advantages of DHT are that it does not need domain name conversion to solve the problem of insufficient address space caused by the use of IP address in a DNS-based identity and resolution system. A point-to-point system overcomes the system bottleneck caused by the overload of the root server, better optimizes the query process, and improves resolution efficiency. However, there are still some problems to be solved when the existing structured P2P system is applied to IIoT scenarios.

(1) The mismatch between the DHT and the underlying network. Because the structure of the P2P network is based on the arrangement of virtual addresses obtained by Hash mapping, there are logically adjacent nodes, but they are far apart. There may be a phenomenon of divergence in query, which affects query efficiency.

(2) Performance, capacity, and interest heterogeneities of nodes lead to low resource utilization. Nowadays, the P2P network is built on the premise of peer-to-peer of all nodes. That is to say, whether the performance of the nodes is a large server or a personal computer, it has to undertake the functions of initiating query requests and responding to retrieval in the P2P network. Nevertheless, this obviously does not match the computing performance and storage capacity of each node in the actual situation.

(3) Dynamic changes of the network. The joining and exiting of nodes will increase query delay and generate additional node information maintenance overhead. There will be interest, location, preference correlation between different nodes.

3.1 Matching physical address with logical address

The problems faced by the structured P2P-ONS parsing system are analyzed at the beginning of Section 3. The critical problem is that the virtual address does not match the physical address. However, from the essence of IIoT, it is to query and analyze the attributes of items in various industries after they are identified and stored. There will be interest, location, preference correlation between different nodes. Therefore, we consider introducing the small world model into the structured P2P-ONS identity resolution system to solve the first problem.

3.1.1 *Small world model*

The small world theory originated from some social network studies. Stanley Milgram [22] first proved the existence of the small world model through the phenomenon of "six-degree separation" in the 1960s.

Steven Strogatz proposed a new network model called WS small world model to analyze "small world phenomenon" through WS small world model. WS small world model has excellent significance to P2P networks. It points out that peer-to-peer nodes can randomly link with long-distance nodes when maintaining short links with neighbor nodes.

The small world model is applied to the P2P network, which makes the nodes in the network correlate according to some characteristics. This feature is the stored resource keywords in the IIoT identity resolution system. The nodes with high correlation store and analyze similar identity information. Therefore, by analyzing the correlation between the characteristics of the nodes, when assigning logical addresses, the nodes with high correlation can be placed in the adjacent positions of the logical network. Furthermore, this P2P network has been called a network with "small world" characteristics. As shown in Figure 2, the partition of node clusters is based on the specified characteristics, and the central node determines the cross-group connection of the whole network with the highest degree of clustering Cv. Different groups of nodes can quickly establish connections through the central node, so the network with small-world characteristics can reduce the length of feature paths and improve retrieval efficiency.

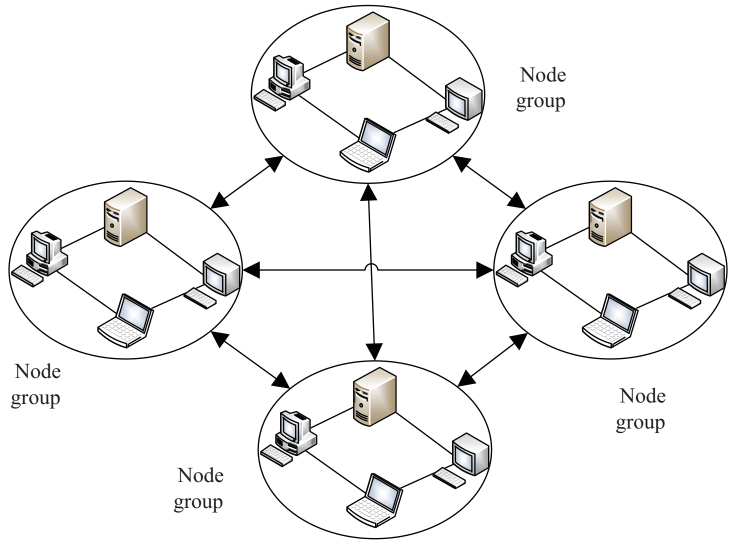

Figure 2 (Color online) Small world model of P2P network.

3.1.2 *Realization of small world model*

In a structured distributed P2P network, it is necessary to divide the network into different groups according to a special network. It is found that when classifying neighbor nodes in the P2P network, the accuracy of the forward fitting support vector machine reaches more than 90%. Only knowing that the personal information of the peer nodes is less than 2%, the selection of the relevant neighbor nodes can be completed efficiently [23].

SVM is a kind of generalized linear classifier that classifies data according to supervised learning. Its decision boundary is the maximum margin hyperplane for solving learning samples [24].

If there is a hyperplane in the feature space of the input data as the decision boundary, the learning objectives are divided into positive and negative classes, and the distance between the points of any sample and the plane is greater than or equal to 1.

Assuming that there are N nodes in a given P2P network, some number of nodes, significantly fewer than the total number of nodes n, are selected at random to serve as a training node. All nodes store certain information, and let the keyword vector of each node contain information tokens be x_i. The learner is given the keyword features present for each training node along with the corresponding classification labels (y). For our neighbor selection task, the label is a binary decision, $y \in \{\pm 1\}$, which defines whether each node can parse the given information, such as the information belonging to the car and whether node t can parse the information labels (u_i). Using the training data, the learner develops a model that uses a small number of features θ in order to predict future connection decisions. Then using these keywords, the nodes in the non-training set of the network are tested, and the nodes in the network are clustered based on interest and divided into different groups.

The algorithm flow of applying support vector machine to the small world model is as follows.

(1) Randomly select T nodes, others nodes as trainers, where $T \ll N$.

(2) Receive information tokens x_t from each $t \in T$.

(3) Determine utility of each training node $y_t = \mathrm{sign}(u_i(t))$.

(4) Let $X = \cup x_t$ and $k = |X|$.

(5) Find features $\theta_1, \ldots, \theta_d \in X$, where $d \ll k$, which best predict $\overline{y_t} = y_t\ \forall t \in T$.

(6) Issue to test set $M \in (N - T)$, predict whether to divide each node into different groups.

In this way, the nodes in the P2P network are clustered according to their interests, thus possessing the characteristics of "small world". In the identity resolution system of IIoT, different node domains can be divided according to the above process according to the identity attributes. If there are more nodes in the domain, smaller subnets can be further divided into the domain. The performance of nodes can be evaluated in each region, and different functions can be allocated according to the performance of nodes. Each P2P network is divided into different levels, and different routing strategies are adopted according to the characteristics of nodes in different levels of networks.

Table 1 Dual Chord routing table description

Table item	Definition
Finger[k]	$(N + 2k)$mod 2^m $(0 \leqslant k < m)$
R_Finger[k]	$(N - 2k + 2m)$ mod $2m$ $(0 \leqslant k < m)$
Successor	Next node in identifier ring
Predecessor	The previous node in the identifier ring

3.2 Improved Chord scheme

In this subsection, we will further analyze the "short-chain effect" and the network hierarchy provided by a high degree of aggregation and investigate the improvement of the Chord algorithm in detail. Given the heterogeneity of network nodes, we propose a hierarchical P2P network [25–27]. Different improved Chord routing algorithms are used in the same level network. Based on the improved Chord algorithm, a two-level identity resolution system is designed to solve the remaining problems at the beginning of this subsection.

3.2.1 *Performance analysis of classical Chord algorithms*

• Average routing hops. Average routing hop count is an important index to measure query efficiency. There are N nodes in the P2P network defined here. The search process of the Chord algorithm is similar to half search. Therefore, the mathematical expectation of the number of routing hops found in a query is $\frac{1}{2}\log_2 N$, that is, the average number of routing hops:

$$\text{Average\_Hop}_{\text{Chord}} = O(\log_2 N). \tag{1}$$

• System stability. The primary measure of system stability is the number of affected nodes in the network when a node joins or exits. When a node joins the network, it is necessary to modify the routing table information of related nodes. Because the routing table of nodes in the Chord network has $\log_2 N$ entries, the number of affected nodes is $O(\log_2 N)$. Similarly, when a node exits the network, it only needs to transfer the resource information stored in the exit node to its successor node and update the routing table of the related node. In this paper's system, the exit operation only affects $O(\log_2 N)$ nodes, so the complexity is $O(\log_2 N)$ level.

3.2.2 *Dual-Chord algorithm*

In the classical Chord algorithm, the query is carried out clockwise on the Chord ring. It is easy to think of a scheme to improve the efficiency of the Chord lookup. After a judgment, the direction of the search on the Chord ring becomes bidirectional. This is the improvement of the Dual-Chord algorithm [28].

Based on the original Chord routing table, Dual Chord adds a counter-clockwise routing table opposite to the original routing table, called R_Finger. In the Dual Chord network, all the routing table entries of the nodes are shown in Table 1.

Dual Chord has clockwise and counter-clockwise directions in routing queries. The nodes in the Finger table are selected clockwise for the next hop, and the next hop node is selected in R_Finger table counter-clockwise. Because Chord's ring structure, whether clockwise or anticlockwise, can always reach the designated nodes on the ring, but most of the time, the two choices need to go through different nodes, which is the key to Dual Chord's performance improvement.

Definition 1. Suppose that the nodes on the Chord ring clockwise constitute a set $N_1, N_2, \ldots, N_k$, where $k = 2^m$ and m are positive integers. The logical distance between nodes N_i and N_j is defined as

$$|\overrightarrow{(N_i, N_j)}| = (j - i) \bmod k. \tag{2}$$

Logical distance can be understood as the number of hops to be experienced from one node to another, which can directly reflect the distance between two nodes.

• Dual Chord routing lookup process. The Dual Chord routing process is basically the same as that of Chord, except that there is one more alternative path provided by the R_Finger table. In Dual Chord routing, N_1 node queries keyword Key, compares the logical distance between the next hop node and Key in clockwise and anticlockwise directions, chooses the direction of a smaller node and Chord routing along that direction. Here, closer logical distance means fewer hops for queries. The specific algorithm process is as follows: node identifier is defined as NodeID, resource object ⟨Key, Value⟩ key-value pair formal storage. Now assume that local node N receives a request for query keyword Key.

(1) The Key is queried at the local node N and returned directly if it is found.

(2) Determine whether the keyword Key is satisfied, N.NodeID $\leqslant$ Key $\leqslant$ N.Successor.NodeID, if satisfied, then clockwise routing, Key stored in the successor of node N, query the next hop node for N.Successor, return the previous step. Otherwise, proceed to the next step.

(3) Judge the direction of the search if you jump counterclockwise to step (4). Otherwise, if the search direction is clockwise, the Finger table of the node is searched to find the node N_1 with the smallest logical distance from the key in Finger, that is

$$|\overrightarrow{N_1.\text{NodeID}, \text{Key}}| = \min(|\text{Finger}[i].\text{NodeID}, \text{Key}),\ i = 0, 1, \ldots, m. \tag{3}$$

Assuming that the node in the Finger table after N_1 is N_2, make the following judgment.

$$|\overrightarrow{N_1.\text{NodeID}, \text{Key}}| < |\overrightarrow{\text{Key}, N_2.\text{NodeID}}|. \tag{4}$$

If formula (4) is valid, clockwise query is carried out. The next hop routing node is P_1 and the return step (1); otherwise, the next hop node is P_2 and the return step (1).

(4) A counterclockwise query is made to find the R_Finger table of the node and the node N_1 with the smallest logical distance from the Key in the R_Finger table.

$$|\overrightarrow{N_1.\text{NodeID}, \text{Key}}| = \min(|\text{R\_Finger}[i].\text{NodeID}, \text{Key}),\ i = 0, 1, \ldots, m. \tag{5}$$

If formula (5) is valid, the next hop routing node is P_1 return step (1); otherwise, the next hop node is P_2, and the next hop node is P_1 return step (1).

• Performance analysis of Dual-Chord algorithms. Assuming that Dual Chord searches in the first half or the second half of a P2P network with N nodes at a time, the order of the routing table is as follows:

$$k \leqslant \log_2 \frac{N}{4}. \tag{6}$$

As can be seen from (6), each forwarding reduces the query area to a quarter of the current area, so the average number of routing hops for Dual Chord routing is

$$\text{Average}_{\text{HopDChord}} = \log_4 N = \frac{1}{2}\log_2 N. \tag{7}$$

Formula (7) shows that compared with the classic Chord algorithm, the average query hops of the Dual-Chord algorithm are reduced by half. This will bring a significant performance improvement in the super large number of node networks of industrial Internet identification and resolution systems. However, the logical distance of the Dual-Chord algorithm is calculated based on the logical address of the Chord ring, ignoring the difference in node performance. In the actual network, nodes with poor performance and high latency may be selected, which will bring unpredictable query delays. Therefore, the Dual Chord routing algorithm is generally applied in the network with good node performance.

• Chord algorithm with additional index caching. The last paragraph describes the Dual-Chord algorithm used in the network with better node performance. For the huge industrial Internet, more nodes are generally or even poorly performing. In order to further improve the performance of the whole analytical system, the regional index mechanism is introduced in this network to reduce the performance requirements of nodes.

The core principle of the index caching mechanism is to increase the number of information providers in the network with low storage space overhead, thus improving the search performance. This strategy is especially useful when used in structured distributed P2P networks with poor performance of nodes. According to some caching strategies, such as least recently used (LRU), resource index items are added to nodes in the network. According to the resource index items, resources can be located directly and query efficiency can be improved. At the same time, it can improve the load balance and relieve the query pressure of peer nodes with frequent requests.

In the Chord network, a part of the nodes can be selected to add a cache table, which is saved, as shown in Table 2.

Table 2 Cache table

Table item	Definition
Key	The resource keyword of the identifier to be queried after Hashing operation
NodeID	Store the node ID to be queried, that is, the node IP address is hashed
Node.IP	Storing the IP of the node to be identified for query

When Chord routing is performed after adding a cache table, the query results are first searched in the local cache table. If the search is successful, the query results are returned directly to the requester. If not, the routing process is carried out and forwarded to other nodes. Finally, if the required resource keywords are successfully found, the query results need to be added to the local cache table in order to prepare for it. After that, we can quickly find and maintain the information stored in the cache table according to the LRU principle.

3.3 Two-level identity resolution system based on improved Chord algorithm

In Subsection 3.2, this paper discusses two improved schemes of the Chord algorithm for different performance networks. It demonstrates the improvement of routing efficiency compared with the classical Chord algorithm in specific scenarios. In this subsection, a two-level P2P-ONS identity resolution system is proposed. The small world model in subsection 3.1 is applied to divide the P2P network into subnets and classify it according to the performance of the nodes [29]. The Dual Chord routing algorithm is used in the upper network with better performance of the nodes, and an additional index buffer is used in the lower network with poorer performance of the nodes. The Dual Chord routing algorithm is used in the upper node network with better node performance. The additional index buffer mechanism is used in the lower network of the lower layer network with poor node performance.

3.3.1 *System model*

By dividing P2P networks into different subnetworks according to their interests, a network with "small world" characteristics is formed. Then, we can evaluate the performance of the nodes in the subnet, and select the nodes whose performance meets the specified conditions from each subnet node as the super-node.

In order to select super-nodes reasonably, the following selection criteria are defined by the information of nodes known by the network as a registration message.

Definition 2. Node_{cal}, node computing power is determined by the performance of node CPU, bandwidth of network and memory size. Node computing power can be represented by

$$\text{Node}_{\text{cal}} = \alpha_1 \times \text{CPU} + \alpha_2 \times \text{Bandwidth} + \alpha_3 \times \text{Meomory}. \tag{8}$$

Among them, α_1, α_2 and α_3 represent the weights of CPU processing capacity, network bandwidth, and memory size under different selection criteria respectively, which accord with

$$\alpha_1 + \alpha_2 + \alpha_3 = 1. \tag{9}$$

Definition 3. Node_{cap} node storage capacity is the size of the node's external memory capacity.

Definition 4. Node_{sta} node stability, the smaller the average failure time of nodes, the better the stability of nodes.

Definition 5. $\text{Node}_{\text{perf}}$ is determined by Node_{cal}, Node_{cap} and Node_{sta}, and quantified by

$$\text{Node}_{\text{perf}} = \beta_1 \times \text{Node}_{\text{cal}} + \beta_2 \times \text{Node}_{\text{cap}} + \beta_3 \times \text{Node}_{\text{sta}}. \tag{10}$$

Among them, $\beta_1, \beta_2, \beta_3$ represent the weights of computing power, storage capacity and stability of nodes under different selection criteria respectively, satisfying the following requirements:

$$\beta_1 + \beta_2 + \beta_3 = 1. \tag{11}$$

Then, according to the size of $\text{Node}_{\text{perf}}$, select the largest $\text{Node}_{\text{perf}}$ in each subnet as a super node. The super node is built into the upper layer network, and the Dual Chord routing algorithm is adopted in the upper layer network. Then, $\text{Node}_{\text{perf}}$ is selected from the subnet to introduce the index caching

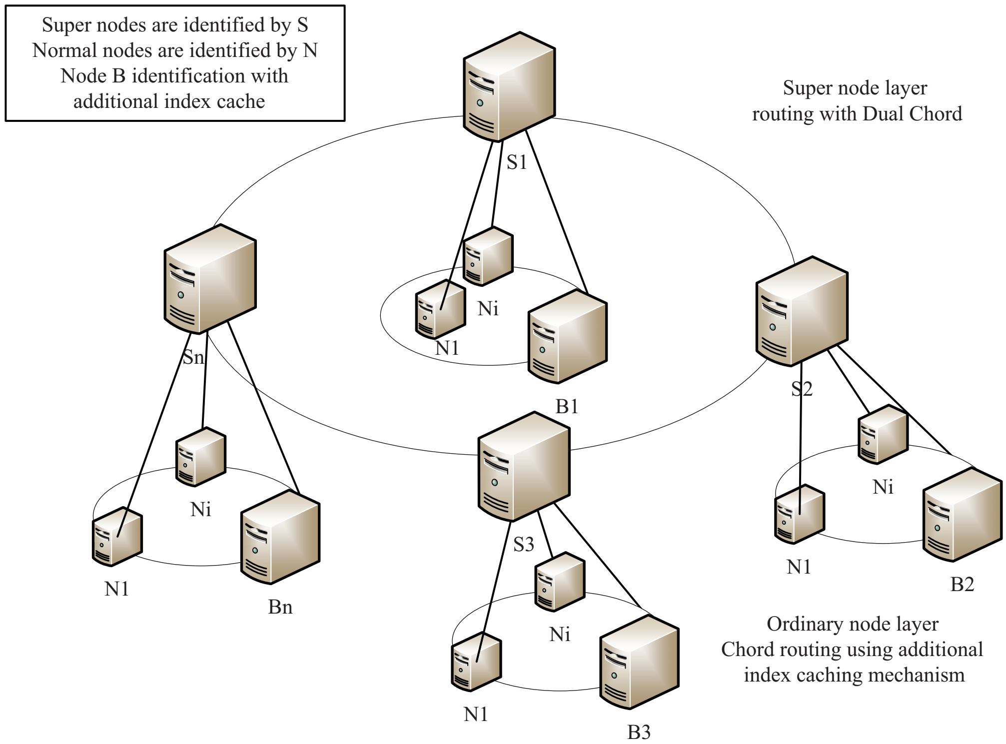

Figure 3 (Color online) Model diagram of two-level identification analysis system based on improved Chord algorithm.

mechanism as the index caching node, and the rest as the index caching node. The nodes of common nodes form the lower network together and adopt the Chord routing algorithm with an additional index caching mechanism. All ordinary nodes need to store the information of the local subnet super-node and index cache node. The index cache node also needs to store the information of the local subnet super-node. The structure diagram of the identity and resolution system is shown in Figure 3.

3.3.2 *System analytical process*

When resolution information in this system, a two-level routing query will be adopted. When the user initiates an identity parsing request.

(1) Find the NodeID information related to the label to be parsed in the local cache, and if it is found, return it directly to the requester. Otherwise, carry out the Chord routing process of the additional index cache mechanism.

(2) Local P2P-ONS nodes generate query requests from the ObjectID and related Node IDs to be searched, send query requests to the index caching nodes in the subnet, and return directly to the requester if there is a resource identity index, and then proceed to step 3 if it fails.

(3) The query request is forwarded to the super-node of the subnet. The super-node performs the Dual Chord routing process. If it finds the identity, it returns result to the requester. Otherwise, it will continue to forward to the next super-node until the search succeeds, or the predetermined forwarding threshold is exceeded.

(4) Finally, if the required resource keywords are found successfully, the query results need to be added to the local cache table in order to search quickly and maintain the information stored in the cache table according to the LRU principle.

The workflow diagram of the whole system is shown in Figure 4.

3.3.3 *Performance analysis*

In this system, assuming that the N nodes P2P network is divided into M sub-networks, the average number of nodes per sub-network is N/M.

• Average routing hops. After the query request is initiated in this system, if the identifier to be parsed

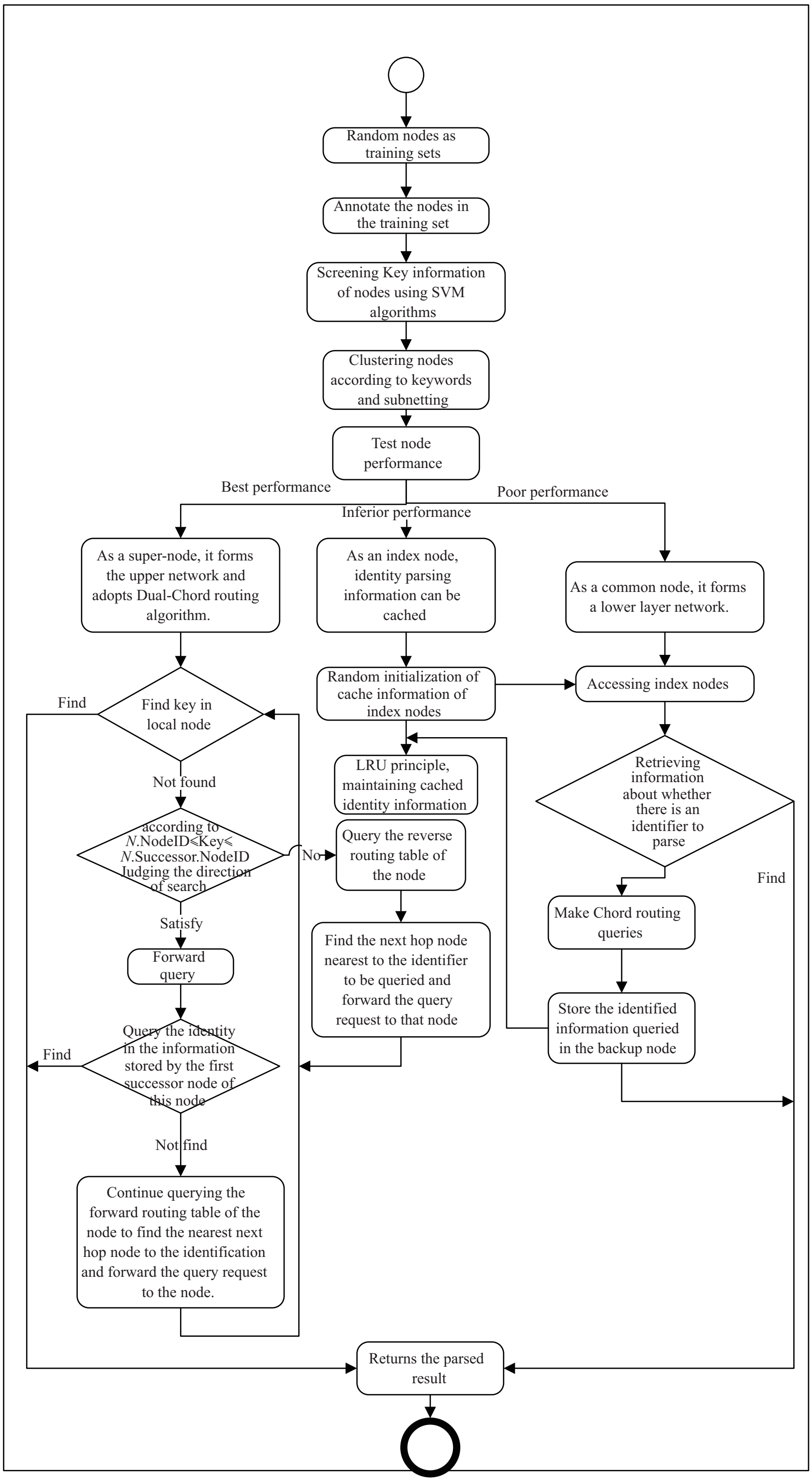

Figure 4 Workflow diagram of two-level identity resolution system.

is in the subnet, the average hops of the search process for the classical Chord routing are

$$\mathrm{Hop1} = O(\log_2 N/M). \tag{12}$$

If it is not in this subnet, it needs to be forwarded to the subnet of the upper network for query, and then parsed query in the corresponding subnet of the lower network, because the upper network uses Dual

Chord routing algorithm for query, and because the upper network is composed of the largest $\text{Node}_{\text{perf}}$ node in each subnet, there are M subnets. In a network composed of subnets, there are M super nodes in the upper network, so when querying in the upper network, the average number of routing hops is

$$\text{Hop2} = O(\log_4 M) = O\left(\frac{1}{2}\log_2 M\right). \tag{13}$$

Assuming the probability that the identifier to be resolved in this subnet is p, which is usually a value close to 1, the average number of routing hops for a query is

$$\text{Average}_{\text{Hop}} = O(\log_2 N) + \left(1 + \frac{p}{2}\right) O(\log_2 M). \tag{14}$$

Compared with the classical Chord routing network, the system reduces the average number of routing hops:

$$\Delta\text{Average}_{\text{Hop}} = \left(1 - \frac{p}{2}\right) O(\log_2 M). \tag{15}$$

Formula (15) shows that the higher the number of subnets partitioned, the greater the probability of finding in local subnets, and the greater the improvement of routing performance of the system. Because of the "small world" characteristics of the system network, the matching problem between logical and physical addresses has been taken into account in the partitioning of subnets, so the parsing identifier can be queried in local subnets. The probability p significantly improved compared with the traditional way.

• System stability. In this system, the network fluctuation caused by the joining and exiting of nodes will also be greatly reduced. When the ordinary node changes dynamically, it only affects the nodes in the subnet, that is, the number of affected nodes is $O(\log_2 N/M)$. When the super node changes dynamically, not only the ordinary nodes of the group are affected, but also other super nodes in the upper network are affected, so that the number of affected nodes is $O(\log_2 N/M) + O(\log_2 M)$. In summary, the average number of affected nodes in the network is

$$\text{Aver}_{\text{NumAffect}} = O(\log_2 N) + \left(\frac{M}{N-1}\right) O(\log_2 M). \tag{16}$$

Eq. (16) shows that the stability of the system is greatly improved when M/N is compared with that of the classical Chord routing network.

Therefore, regardless of the efficiency of routing queries or network stability, the two-level ID resolution system based on Dual Chord and additional index caching mechanism designed in Section 3 is superior to the original Chord system. In the vast network of industrial Internet, the division of subnets can achieve a huge number. It can be further divided into the third-level network in the second-level network. From (16), it can be seen that the design of the system in the more substantial network performance improvement is more obvious, so it is very suitable for IIoT's identity resolution scenario.

4 Simulation

4.1 Simulation environment

Usually, network simulation is a simple and easy way to verify the performance of the system. With the rapid development of network technology, the current simulation technology can deal with various complex realities, such as large-scale topology, multi-layer protocol environment. The two-level improved Chord identity resolution system for the IIoT proposed in this paper is based on the P2P network. Industrial Internet is a large-scale network, and the fluctuation of nodes in the network is generally significant. It is very tough to test it in a real environment. The emergence of PeerSim provides a powerful tool for the simulation of a P2P network [30–32]. The whole code file structure is shown in Figure 5.

The creation of nodes in the Dual Chord network mainly involves the setting of node attributes. Unlike classical Chord, Dual Chord needs to consider the forward and directional routing tables and related attributes because of the existence of reverse routing. The nodes in this paper are generated by random numbers and obey normal distribution. In this paper, the query probability in the node area group is set to 40%. The cache capacity is set to 50. The minimum query delay is set to 0 ms, the maximum query delay is designed to 50 ms. Moreover, the node arrival rate and departure rate are set to 15 per second.

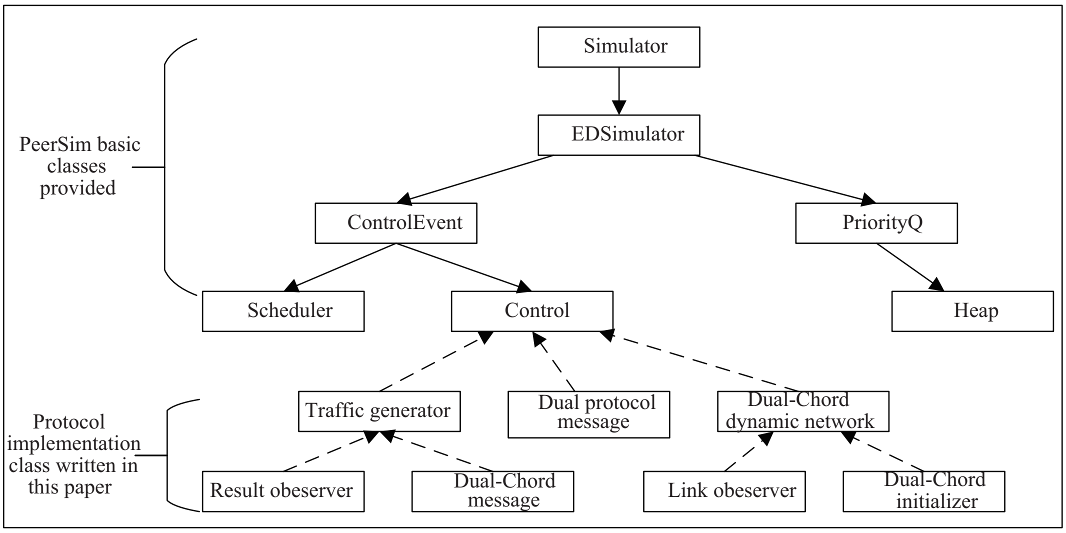

Figure 5 File structure of simulation code.

4.2 Simulation results and analysis

4.2.1 *Average routing hops*

After much coding, this paper implements the Dual-Chord algorithm and classical Chord algorithm and monitors routing hops in result observer class. When PeerSim is used for simulation, the number of network nodes is set to 1000. The distribution of query hops in 1000 queries between the Chord system and Dual Chord system is observed, as shown in Figure 6. After that, the number of network settings is increased from 500 to 10000, and the simulation results are shown in Figure 7.

From Figures 6 and 7, we can see that the simulation results basically satisfy the theoretical analysis. The average routing hops of the classical Chord algorithm is $1/2\log_2 N$, while that of the Dual-Chord algorithm is close to $1/2\log_4 N$. From the curve trend in the graph, we can see that the average routing hops of the Chord algorithm are along with the network nodes. Increasing its growth trend is also tending to be stable, which verifies that the Chord algorithm is used for routing queries in large-scale networks, and the Dual-Chord algorithm further enlarges this advantage.

4.2.2 *Average query delay*

Similar to the process of simulating the average number of routing bars, it is still necessary to increase the number of network nodes from 0 to 5000. However, in the configuration file, the minimum delay and maximum delay of nodes in the transport layer protocol are set to 10 ms and 20 ms, respectively.

The average query latency simulated by the Chord algorithm and Dual-Chord algorithm is shown in Figure 8.

From Figure 8, it is evident that the average query delay of the Dual-Chord algorithm is 10 ms lower than that of the Chord algorithm with the increase of the number of nodes.

4.2.3 *System stability*

In the implementation of protocol code, the stabilizing function for joining and exiting nodes in the network is defined, and the stabilization parameters are defined to indicate the number of stabilizing functions that need to be executed by the node state change system to measure the stability performance of the network. In the simulation, using the dynamic parameters in the configuration file, 0 to 15 nodes are added and removed, respectively. The simulation results are shown in Figure 9.

It can be seen from Figure 9 that although Dual Chord introduces a reverse routing table, which increases the complexity of the algorithm, it is basically the same as the classical Chord algorithm in terms of system stability.

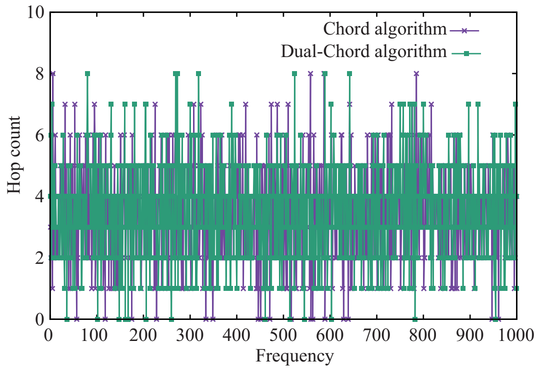

Figure 6 (Color online) Query hop distribution simulation contrast diagram.

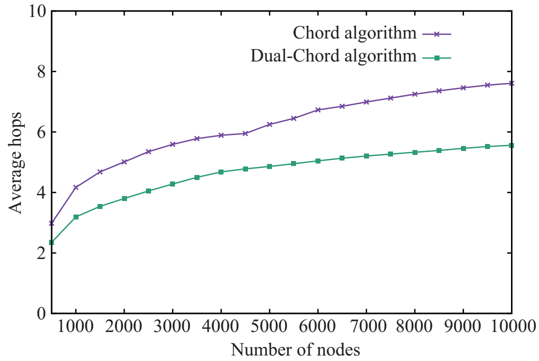

Figure 7 (Color online) Chord and Dual Chord average routing hop count comparison.

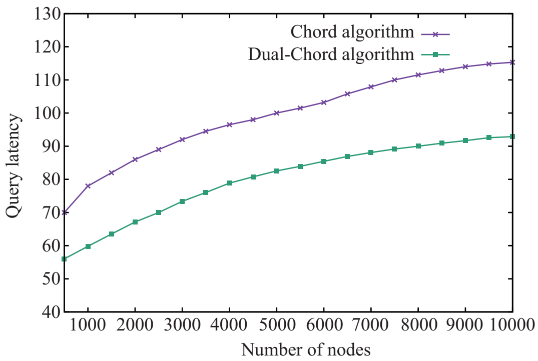

Figure 8 (Color online) Chord and Dual Chord query delay distribution comparison.

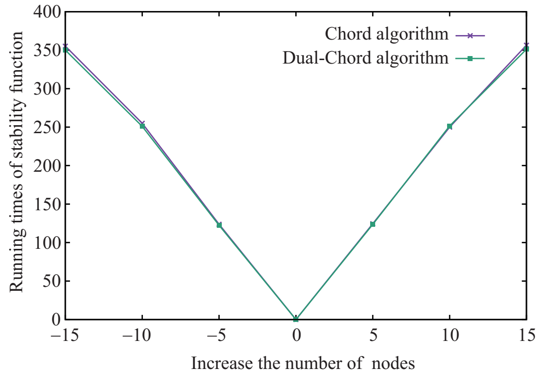

Figure 9 (Color online) Chord and Dual Chord system stability comparison.

5 Summary and prospect

This paper introduces the concept and related research of IIoT and discusses in detail the critical technology of the IIoT identity resolution system. Based on referring to the existing identity resolution system, a more efficient identity resolution system is proposed. To sum up, the main innovations of this paper are: (1) improving the classic Chord routing algorithm in identity query; (2) introducing small-world model and support vector machine algorithm to cluster nodes in view of the heterogeneity of IIoT; (3) on this basis, a novel identity resolution system based on Dual-Chord algorithm for IIoT is proposed.

Because the development of IIoT is still in its infancy, there is still much room for improvement in the technology of identity resolution. Although this paper has made significant improvements based on the existing identity resolution system, there is still much work to be done in the future.

(1) The system in this paper does not consider the attenuation of the node's own performance in the design, and the nodes in the upper network may be aging. Therefore, the next step can consider introducing a periodic maintenance mechanism to detect the performance of network nodes and update the distribution of network nodes regularly to ensure the routing performance of the whole system.

(2) The system in this paper has not been validated by a real network. Although PeerSim's discrete event-based simulation system can complete the implementation protocol, the node size of the simulation network is not large enough. In the future research, a more extensive simulation network needs to be built, which will be more useful for promoting the application of IIoT.

Acknowledgements This work was supported by Natural Science Foundation of Beijing (Grant No. 4212004), National Key Research and Development Program of China (Grant No. 2019YFB1804403), and MIIT of China 2019 (Innovative Identification and Resolution System for Industrial Internet of Things).

References

1 Wang K, Wang Y, Sun Y, et al. Green industrial Internet of Things architecture: an energy-efficient perspective. IEEE Commun Mag, 2016, 54: 48–54

2 Li Z, Kang J, Yu R, et al. Consortium blockchain for secure energy trading in industrial Internet of Things. IEEE Trans Ind Inf, 2018, 14: 3690–3700
3 Xu L D, He W, Li S. Internet of Things in industries: a survey. IEEE Trans Ind Inf, 2014, 10: 2233–2243
4 Hossain M S, Muhammad G. Cloud-assisted industrial Internet of Things (IIoT) — enabled framework for health monitoring. Comput Netw, 2016, 101: 192–202
5 Louis Columbus. IIoT Will Overcome Adoption Barriers, Soar To 200 Billion dollars By 2021. Forbes. 2019. https://www.forbes.com/sites/louiscolumbus/2019/03/13/iiot-will-overcome-adoption-barriers-soar-to-200b-by-2021/#387523453513
6 Qin E, Long Y, Zhang C H, et al. Cloud computing and the Internet of Things: technology innovation in automobile service. In: Proceedings of International Conference on Human Interface and the Management of Information. Berlin: Springer, 2013
7 Traub K, Armenio F, Barthel H, et al. The GS1 EPC Global Architecture Framework Version 1.6. GS1 EPCglobal, Technical Report GS 1, 2014
8 Identification system for Internet of things – Entity code: GB/T31866-2015. (2015-09-11)[2019-08-16]. https://euagenda.eu/upload/publications/identifiers-in-internet-of-things-iot.pdf
9 Sun S, Lannom L, Boesch B. Handle System Overview: RFC 3650. Technical Report, The Internet Society (ISOC)− IETF, 2003. https://tools.ietf.org/html/rfc3650
10 Bo C, Wang D. Design and implementation of efficient distributed name service in EPC network. In: Proceedings of IEEE Conference Anthology, 2013
11 Li Y, Tang K, Gao F. Object Information Processing Method and Apparatus, and ORS Application Gateway: U.S. Patent Application 15/513,662. 2017-10-26
12 Yan Z W, Kong N, Tian Y, et al. A universal object name resolution scheme for IoT. In: Proceedings of 2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing. 2013
13 Sun S, Reilly S, Lannom L. Handle system namespace and service definition: RFC 3651. IETF, (2003-11-11)[2019-08-16]. https://datatracker.ietf.org/doc/rfc3651/
14 Lannom S S S R L, Petrone J. Handle system protocol (ver 2.1) specification: RFC 3652. IETF, (2003-11-11)[2019-08-16]. https://tools.ietf.org/html/rfc3652
15 UID Center. Ubiquitous ID architecture. 2006. https://www.uidcenter.org/wp-content/themes/wp.vicuna/pdf/UID-CO00002-0.00.24_en.pdf
16 UID Center. Ubiquitous code: ucode. 2009. https://www.uidcenter.org/wp-content/themes/wp.vicuna/pdf/UID-00010-01.A0.10_en.pdf
17 ISO/IEC. Information technology — open systems interconnection — object identifier resolution system (ORS): ISO/IEC 29168-1. 2011. https://www.iso.org/obp/ui/#!iso:std:45247:en
18 Ding D, Li M B, Zhu Z. Object naming service supporting heterogeneous object code identification for IoT system. In: Proceedings of 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC), 2018
19 Evdokimov S, Fabian B, Kunz S, et al. Comparison of discovery service architectures for the Internet of Things. In: Proceedings of IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing, 2010. 237–244
20 Schollmeier R. A definition of peer-to-peer networking for the classification of peer-to-peer architectures and applications. In: Proceedings of the 1st International Conference on Peer-to-Peer Computing (P2P'01), Lingkoping, 2001. 101–102
21 Xu H, Wang S P, Wang R C. P2P ONS: a distributed object naming service architecture based on P2P for EPC network. In: Proceedings of Advances in Information Sciences and Service Sciences, 2011. 3: 37–47
22 Milgram S. The small world problem. Psychology Today, 1967, 1: 61
23 Beverly R, Afergan M. Machine learning for efficient neighbor selection in unstructured P2P networks. In: Proceedings of the 2nd USENIX Workshop on Tackling Computer Systems Problems With Machine Learning Techniques, Berkeley, 2007. 1–6
24 Noble W S. What is a support vector machine? Nature Biotech, 2006, 24: 1565–1567
25 Woungang I, Tseng F H, Lin Y H, et al. MR-Chord: improved Chord lookup performance in structured mobile P2P networks. IEEE Syst J, 2015, 9: 743–751
26 Singh M, Kumar C, Nath P. Mobility pattern based Chord (MP-Chord) for enhanced lookup performance in mobile P2P networks. Wirel Personal Commun, 2019, 109: 1971–1985
27 Zhao W, Liu X, Zhang S, et al. Hierarchical P2P based RFID code resolution network: structure, tools and application. In: Proceedings of International Symposium on Computer Network and Multimedia Technology, Wuhan, 2009. 1–7
28 Sun D-P, Wang Y-T. Improvement of Chord routing algorithm. Mech Electr Eng Mag, 2007, 12: 30
29 Amirazodi N, Saghiri A M, Meybodi M. An adaptive algorithm for super-peer selection considering peer's capacity in mobile peer-to-peer networks based on learning automata. Peer-to-Peer Netw Appl, 2018, 11: 74–89
30 Montresor A, Jelasity M. PeerSim: a scalable P2P simulator. In: Proceedings of 2009 IEEE 9th International Conference on Peer-to-Peer Computing, Seattle, 2009. 99–100
31 Kazmi I, Bukhari S F Y. PeerSim: an efficient & scalable testbed for heterogeneous cluster-based P2P network protocols. In: Proceedings of 2011 UkSim 13th International Conference on Computer Modelling and Simulation, Cambridge, 2011. 420–425
32 Dinh T T A, Lees M, Theodoropoulos G, et al. Large scale distributed simulation of P2P networks. In: Proceedings of the 16th Euromicro Conference on Parallel, Distributed and Network-Based Processing (PDP 2008), Toulouse, 2008. 499–507