International Journal of Applied Science and Engineering
2020. 17, 1: 45-58

Performance Improvement of P2P using Location Awareness and Interest Clustering

Shu-Ching Wang[a*], Wei-Ling Lin[a], Hsin-Hung Pan[b] and Po-Chien Lin[a]

[a]*Department of Information Management, Chaoyang University of Technology, Taiwan, R.O.C.*
[b]*Department of M-Commerce and Multimedia Applications, Asia University, Taiwan, R.O.C.*

Abstract: Because Peer-to-Peer (P2P) has the characteristics of providing scalable network services, related applications of P2P are one of the hot topics of network services today. However, too many P2P applications may lead to inefficient use of network resources or extremely poor application execution performance. At the same time, it may also occupy a large amount of bandwidth, causing an increase in the load of Internet Service Providers (ISPs). In order to avoid excessive network traffic burden, ISPs often use a flow control mechanism to prevent P2P from generating excessive traffic, which leads to the degradation of P2P software service quality. Since P4P (Proactive network Provider Participation for P2P) is one of the technologies used for P2P network optimization, it provides communication between ISPs and P2P service providers in a cooperative manner to reduce P2P traffic and ISP operating costs. Therefore, in order to reduce the operating costs of ISPs and the bandwidth required for P2P application services, a three-layer architecture combining location awareness with interest clustering and P4P framework is proposed in this study, which is called ICP4P (interest clusters based P4P). Using the ICP4P proposed by this research will reduce Internet traffic and increase the robustness of the network. In addition, this study provides a candidate peer election mechanism to improve the service quality of P2P. When super peer fails, leaves or overloads, the backup super peer or cooperative peers will be enabled to maintain the stability of whole network.

Keywords: Peer-to-peer network; P4P; quality of service; internet service provider; interest cluster; location awareness.

1. Introduction

Network management has steadily evolved over recent years. Along with the growing need for advanced features in network management solutions, several distribution models were investigated, varying from centralized to fully distributed models [1]. Because Peer-to-peer (P2P) has the characteristics of providing scalable network services, P2P related service applications are one of the hot topics of today's network services [2]. Most of the early P2P related applications were file sharing [2]. However, in recent years, many applications developed on the P2P platform have been used for commercial purposes, including providing MP3 or movie exchange services [3]. A P2P overlay network is a virtual or logical network of overlay peers connected by virtual or logical links, formed on the top of another physical network that is called underlay. Whereas every virtual or logical link is like a path consisting of one or many physical links of underlay [2].

*Corresponding author; e-mail: scwang@cyut.edu.tw
Received 7 February 2020
doi:10.6703/IJASE.202003_17(1).045

Accepted 18 February 2020

In a P2P network, the "peers" are computer systems which are connected to each other via the Internet [4]. Files can be shared directly between systems on the network without the need of a central server. In other words, each computer on a P2P network becomes a file server as well as a client.

P2P network topologies can be divided into centralized and decentralized architectures [5]. A centralized P2P architecture in which shared items of all peers are indexed in a central server. While the queries in this centralized architecture are very efficient and versatile, the main issues are security and scalability. A decentralized P2P architecture can be divided into structured and unstructured structures. For decentralized structured P2P networks, although they do not have a central server, they have a large number of structures. Structured means that the topology of the network is tightly controlled, files are not randomly placed on peers, but placed in specific locations, which makes subsequent queries easier to satisfy. A decentralized unstructured P2P network has no central directory server and no precise control over network topology or file placement. A decentralized unstructured P2P network is formed by peers that join the network following some loose rules, so queries in the system are typically forwarded to random neighbors.

In a structured decentralized architecture, DHT (Distributed Hash Table) is used to map keywords to a certain peer, and then peers for data storage are quickly searched according to various routing strategies. The biggest problem of structured decentralized architecture is the frequent joining and leaving of peers, which will bring a huge burden to the maintenance of DHT [6]. In an unstructured decentralized architecture, data is queried through flooding through TTL (Time To Live), so the system does not create a bottleneck [5].

In the P2P architecture, the data is randomly selected by peer for transmission, so in the process of transmitting data, it often passes through redundant paths, resulting in excessive traffic [3]. Since the data resources are distributed on separate peers in the P2P network [7], how to efficiently access data, reduce the operating cost of ISP (Internet Service Provider) and the bandwidth required for P2P application services will be explored in this study.

In P2P systems, the design of information search methods will affect the efficiency of search. Therefore, Guo *et al*. have proposed to classify users according to their interests, so that peers with the same interests can be located in adjacent areas [8]. Classifying peers by interest allows users to quickly find the information they need, so the time for information search will be reduced and the efficiency of search can be improved [8].

P2P is cost effective, but there are still concerns about scalability and QoS. Therefore, an architecture called P4P (Proactive network Provider Participation for P2P) was proposed by Xie *et al*. [9] that enables effective collaboration between applications and network providers. In other words, P4P enables efficient resource allocation and better performance.

Therefore, a P4P architecture based on interest clusters (interest clusters based P4P, ICP4P) will be constructed in this study to improve the service performance of P2P networks. Through the ICP4P proposed by this research, the Internet traffic can be reduced, and the robustness of the network can be increased. Due to the characteristics of P4P, the suitable data peers for data access can be utilized. In addition, the bandwidth required for P2P application services can be effectively reduced.

The rest of this paper is arranged as follows: Section 2 illustrates the related work of our research, including the concept of P4P, interest cluster and related research results of P2P. The proposed ICP4P is given in Section 3. The operations of ICP4P are shown in Section 4. The experiment design and results analysis of ICP4P are explained in Section 5. Finally, the conclusion is provided in Section 6.

2. Related Work

In this section, the concept of P4P, interest cluster, and related research results of P2P will be explained.

2.1 The Concept of P4P

P4P (Proactive network Provider Participation for P2P) is one of the technologies to optimize P2P networks, which was proposed by Xie *et al*. in 2008 [9]. P4P is a simple and lightweight architecture that enables more efficient collaborative flow control between applications and ISP.

The P4P framework consists of a data-plane component and a control-plane component [10,11]. In the data plane, P4P allows routers on the data plane to give fine-grained feedback to P2P and enable more efficient usage of network resources. In the control plane, P4P introduces iTrackers to provide portals for P2P to communicate with network providers. The introduction of iTrackers allows P4P to divide traffic control responsibilities between P2P and providers, and also makes P4P incrementally deployable and extensible. The potential entities in the P4P framework include iTrackers owned by individual network providers, appTrackers in P2P systems, and P2P peers.

Figure 1 is a schematic diagram of the P4P frame control plane [9]. Among them, iTrackers are owned by personal network providers, appTrackers and P2P clients (or peers) in P2P systems. Not all entities can interact in a given setting. For example, no tracking system has any appTrackers. P4P does not specify the exact flow of information, but only provides a generic messaging framework that controls the encoding of XML in order to achieve scalability. The key component of the P4P framework is iTrackers. iTrackers provides three interfaces that other people can query, including: “info”, “policy”, and “capability”.

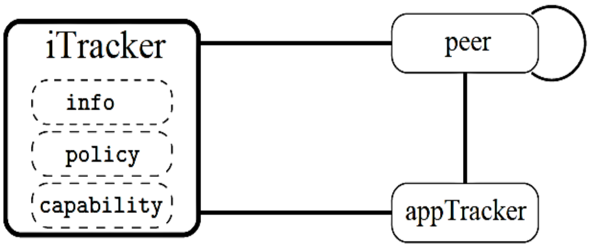

Figure 1. iTracker interfaces and information flow.

The “info” interface allows others (usually peers within the provider network) to obtain the network topology and state. The “policy” interface allows others (such as peers or appTrackers) to get the network's policies and guidelines. Policies specify how network providers use their networks at a high level, usually regardless of P2P applications; and guidelines are specific recommendations for P2P to use network resources. User similarity search in a clustered P2P network of interest is a task that effectively finds a specific target similar to a given user based on common interests [12]. The “capability” interface allows others (such as peers or content providers (via appTrackers)) to request the capabilities of the network provider.

2.2 Interest Cluster

User similarity search in a P2P network is a task that effectively finds a specific target similar to a given user based on a common interest [12]. Since network cost is a major consideration in P2P networks, the algorithm should have network efficiency while maintaining good search quality.

In the decentralized unstructured P2P architecture, when a search is performed, it is broadcast in a flooding manner. In order to reduce the restrictions on flooding, TTL (Time To Live) is used to limit the traffic of the broadcast [13]. When the TTL is 0, the file search is stopped immediately. Since the path distance of each peer to other peers is different, the peer with the file may not be broadcasted because the connection path of the peer is too far.

Since in the P2P system, the data search method will affect the efficiency of the search. The users are therefore grouped according to their interests so that peers with the same interest can be in the same or adjacent clusters. These clusters obtained by grouping by interest are called Interests Cluster. Since the Interests Cluster is grouped by the user's interests, it is possible to quickly find the information needed, reduce the time of data search and improve the efficiency of the search. To improve the scalability of P2P networks in order to respond quickly to customer service requests flexibility, a hierarchical hybrid P2P architecture based on Interests Clustering was proposed by Tu *et al*. [14].

2.3 Related Research Results of P2P

In the past, some scholars used interest clusters to construct peer networks [6, 15, 16]. The related research will be explained in this section.

2.3.1 Peer cluster

In the research of Hunag *et al*., the construction method of Peer Cluster is proposed [15]. In Peer Cluster, the Hypercube architecture is used to solve message duplication and utilize interest clusters to increase search performance. However, in a P2P network constructed by peer based by Hunag *et al*., due to the Hypercube architecture, if the addresses are added in the order in which the peers are added without considering the peers in adjacent locations, the problem of mismatching will also occur. That is, the distance travelled by the file inquiry will be far greater than the shortest physical distance between two points, resulting in poor performance [17].

2.3.2 A Locality awarded Hybrid P2P system with interest Grouping (LaHiG)

LaHiG was proposed by Hu, which is mainly used to solve the problem of mismatching in Peer Cluster [6]. LaHiG is a two-layer architecture. The first layer of LaHiG is a Hypercube built for peers with similar locations in the same interest. The super peers in the first layer are responsible for archival query of the same area with the same interest. The composition of the second layer is divided into two types. If it is a super peer in the same area with different interests, it will be composed as an overlay network; if it is a super peer in the same area with different interests, it will be connected as a Hypercube. The super peers in the second layer are responsible for the archive query of different regions with the same interest, the same regions with different interests, and different regions with different interests.

In the first layer of LaHiG, the peers of the same area with the same interest are connected in a Hypercube manner, and the peer with the best capability value is selected as the super peer, which is responsible for the file query of the same area and also serves as the peer of second layer. The peers in the second layer are connected in a Hypercube manner, and are responsible for querying files in the same area with different interests and different areas with different interests.

2.3.3 Three tiers peer-to-peer network of Location Awareness and Interest Group (TLAIG)

TLAIG is a framework proposed by Lin in 2011 [16] to improve the performance of LaHiG [6]. TLAIG selects the best super peer from the peers in each interest cluster as the main super peer of the third layer. When a large number of peers join at the same time, there are main super peers of the third layer and general super peers that can query the cross-interest and same-interest files respectively.

TLAIG is three-layer architecture. The composition of the first layer of TLAIG is to connect the peers of the same area with the same interest in the way of Hypercube. The super peers of the first layer are responsible for the query of the same area with the same interest. The composition of the second layer can be divided into two types. For super peers in the same area with different interests, they are composed in the form of an overlay network; if they are super peers in different areas, they are connected in the form of a hypercube. The work of the second layer of super peers is responsible for the file query of peers with the same interest in different regions and peers with different interests in the same region. The peers in the third layer are clustered by various interest clusters in the second layer. The peer with the best capability value is regarded as the main super peer. The main super peer is responsible for the archive query of different areas of interest.

In the first layer of TLAIG, the peers in the nearby area with the same interest are connected in a Hypercube manner, and the peer with the best capability value is selected as the super peer, which is responsible for performing file query in the same area, and also serves as the peer of second layer. The peers in the second layer will be connected in two ways. If the peers are in the different area with different interests, they will be connected in the Hypercube mode. If the peers are in the same area with different interests, they are connected in the manner of an overlay network, and the peers in the second layer are responsible for performing archive query on the same areas with different interests. Each interest cluster in the second layer will then select the peer with the best capability value as the peer in the third layer, and the peer in the third layer is responsible for performing file queries of different areas with different interests.

3. The Proposed ICP4P

In this study, a P4P architecture based on interest clusters (interest clusters based P4P, ICP4P) was proposed. ICP4P combines the P4P framework with the interest clusters to enable ISPs and P2P application service providers to communicate in a collaborative manner. ICP4P can reduce the traffic on the Internet and increase the robustness of the network. It can also leverage the P4P features to provide the right data peers for data access, reducing the operating costs of ISPs and the bandwidth required for P2P application services.

In the ICP4P designed by this study, peers are divided into three roles, namely Peer, Super Peer and Backup Super Peer. The various roles of peers are described below:

(1) Peer: A peer managed by a Super Peer in the same interest cluster.
(2) Super Peer: Use the capability function to find the Peer with the best capability value in the cluster as the Super Peer of the cluster, which is responsible for forwarding the file search request of the Peer in the cluster. When selecting a Super Peer, you need to consider the various capability factors of the Peers in the cluster, such as the processing power of the CPU, memory size, network bandwidth, and transmission rate.
(3) Backup Super Peer: A Peer with the next best capability value in the same cluster is regarded as a Backup Super Peer. When a Super Peer fails or leaves the cluster, the Backup Super Peer is responsible for replacing the Super Peer.

3.1 Three Phases of ICP4P

In the process of forming and operating ICP4P, there are three phases: Inactive Phase, Active Phase, and Overload Phase. In the Inactive Phase, the information and interest cluster provided by iTracker are used to construct a hierarchical clustering structure. In the Active Phase, the cluster has been established in the network topology, and it has started to operate normally. During the Overload Phase, the Backup Super Peer is selected as the new Super Peer if the Super Peer fails or leaves the cluster. However, if there is no Backup Super Peer in the cluster, that is, the original Super Peer is the last peer in the cluster, so the entire cluster will be unclustered.

3.1.1 Inactive Phase

In the Inactive Phase, the Peers in the first layer (Cluster layer) will be clustered according to the “interest” and “area” using the information recorded by iTracker. And, the Peer with the best capability value from each area of each interest cluster is selected as a Super Peer, it is responsible for doing archival queries of the same interest. The Peers of the second layer (Hypercube layer) are composed of the first layer of Super Peers. They are connected in the manner of Hypercube. They are responsible for querying the files of the same area with different interests. The third layer is a Peer across the Hypercube management layer. The Peers in the third layer are responsible for connecting with iTracker and forwarding archive queries of different areas with different interests. When the cluster structure is completed, it enters the Active Phase and operates normally.

In short, iTracker is used to collect the location of each Peer and its storage files, which will be used to group Peers into clusters of interest in the Inactive Phase. These Peers then build an ICP4P architecture based on the interests and regions of each Peer. In other words, at this time, the information and interest cluster provided by iTracker will be used to establish a hierarchical clustering architecture.

3.1.2 Active Phase

In the Active Phase, the establishment of the interest cluster has been completed in the network topology and is beginning to function normally. After the Super Peer is selected, the Peer with the second best capability will be treated as a Backup Super Peer. Since the Peer may join or leave the system frequently, when the new Peer joins the system, it will judge whether the capability value of the new Peer is greater than the capability value of the Backup Super Peer. If it matches, the new Peer will be regarded as a Backup Super Peer.

And, the tasks that need to be performed in the Active Phase include the operation of adding new Peers, the operation of leaving Peers in the cluster, the search query of Peers, and the query of files at all levels. In other words, in the Active Phase, the cluster has been established in the network topology, and it has started to operate normally.

3.1.3 Overload Phase

In the Overload Phase, the load of Super Peer in the interest cluster has exceeded the pre-set threshold. In the Overload Phase, the Backup Super Peer selection mechanism will be used to solve overload situation of Super Peer in this study to avoid the stability of the overall network due to overload of the Super Peer.

In this phase, when a Super Peer in the cluster fails or leaves the cluster, the system will select a Backup Super Peer as the new Super Peer and take over the work of the original Super Peer. If there is no Backup Super Peer in the cluster, it means that the original Super Peer is the last Peer in the cluster, so the entire cluster is disbanded.

3.2 The Clustering of ICP4P

With the rapid development of P2P technology, P2P technology is applied to commercial applications such as file transmission, instant communication, and multimedia transmission. However, these applications allow P2P traffic to occupy most of the Internet traffic, while P2P traffic continues to grow rapidly, causing Internet service provider bandwidth to be occupied and increasing the cost of the ISP.

Therefore, a three-layer clustering architecture combined with interest clusters and P4P framework ICP4P was established by this study. ICP4P uses the information provided by the P4P framework to group Peers according to their interest cluster and location of the region. Then, the capability function is used to calculate the capability values of each Peer to select a suitable Super Peer. The Super Peer is responsible for communicating with each Peer in the same interest cluster and Super Peer in other different regions. In each interest cluster, Super Peers in the same region are connected in Hypercube and are responsible for inquiring about files of different interests. In each Hypercube layer, a best-performing Peer is selected as the Peer manager in Cross-Hypercube management layer, responsible for performing cross-interest, cross-regional file queries.

First, the Peers on current P2P networks are used to build the interest clusters by ICP4P. After the interest cluster is built, according to the capability values, the Cluster Manager of each region in different interest clusters will be selected. Then, the Cluster Managers of these clusters will be linked in a Hypercube. Then, the Peer with the best capability value from the Hypercube structure is elected as the Peer manager in the Hypercube management layer. Among them, iTracker will provide information to the Peer manager across the Hypercube management layer to perform inter-Hypercube queries.

The selection of the managers of each level needs to consider the capability factor *V* of the Peer, including the processing power of the CPU (*c*), the size of the memory (*m*), and the transmission rate (*t*), ... and so on. Each capability factor has different importance levels depending on the application, so different weight values are given respectively, *i* is the identifier of Peer *i*, *j* is *j* different weight values, and the sum of the weight values is 1. Figure 2 is an example of ICP4P clustering and Eq. 1 is the capability function.

$$V_i = w_1 * (c_i) + w_2 * (m_i) + w_3 * (t_i); \sum_{j=1}^{3} w_j = 1 \quad (1)$$

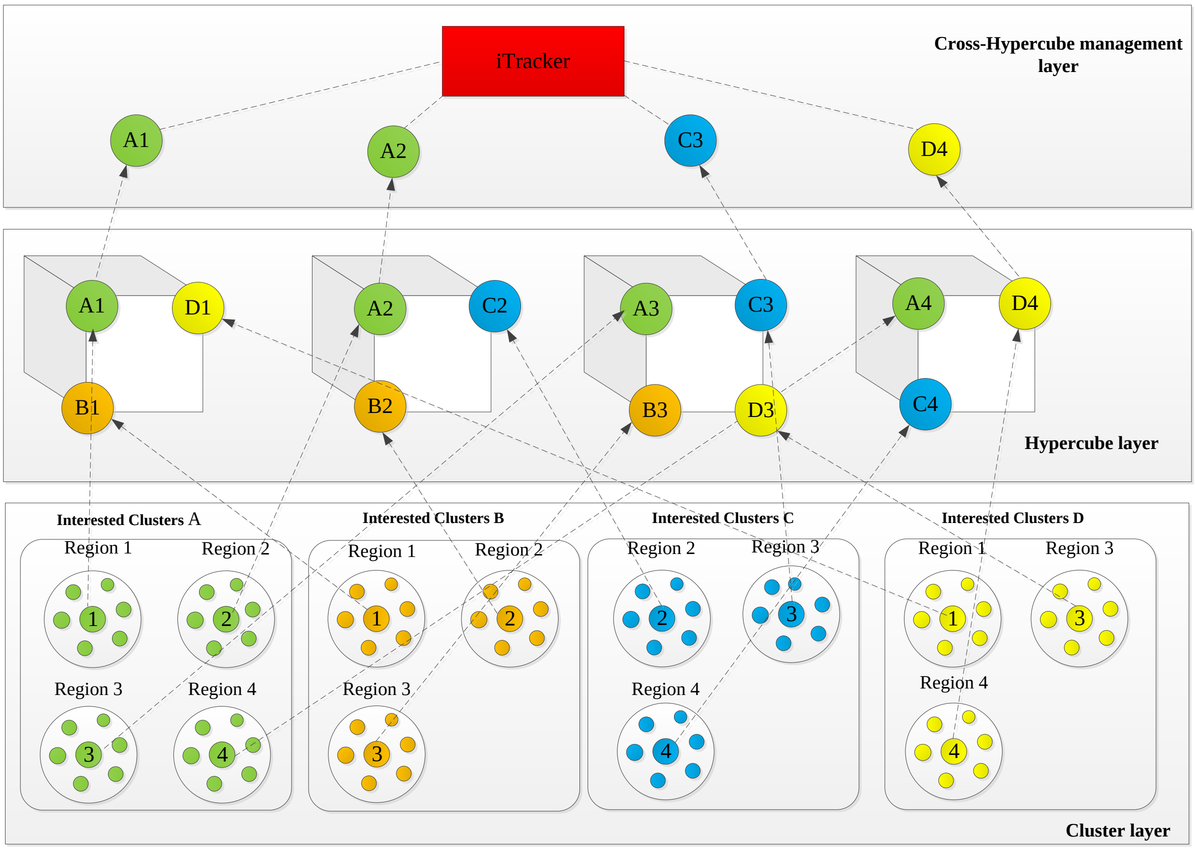

Figure 2. An example of ICP4P clustering.

4. The Operating of ICP4P

In this section, the operation of adding new Peers, the operation of leaving Peers in the cluster, the method of Peers search, and the query of Super Peer files at all layers will be explained separately in ICP4P.

4.1 The Operation of Adding New Peers

If I_α is an interest cluster, *A*, *B*, *C*, and *D* represent different Peers, respectively. When a new Peer *A* wants to join ICP4P, and I_α is the interest cluster that Peer *A* wants to join, the steps to join are as follows:

[Step 1]: First, Peer *A* sends a join request to the local iTracker, and iTracker sends the information of Peer *A* to appTracker for registration.

[Step 2]: Use the information provided by iTracker in [Step 1] to assign Peer *A* to the interest cluster I_α to be added. First, Peer *A* sends a query request to iTracker. iTracker assigns Peer *A* to super Peer *C* with the same interest cluster I_α through the routing table.

[Step 3]: After Peer *A* finds the interest cluster I_α, it uses the routing table to find the Super Peer *D* closest to Peer *A* through the found Super Peer *C*, and then Peer *A* is assigned to Super Peer *D* by Super Peer *C*.

[Step 4]: Super Peer *D* allocates an address of storage Peer *A* and updates the routing table, and sends a "Join Confirm" message to Peer *A*, and updates the routing table of Peer *A*.

4.2 The Operation of Peers Leave the Cluster

If *A*, *B*, and *C* represents different Peers, when Peer *A* is damaged or leaves the cluster, the processing steps are as follows:

[Step 1]: When Peer *A* leaves the cluster, it will find the nearest neighbour Peer *B* from Peer *A* in the routing table of Peer *A*.

[Step 2]: Peer *A* sends a leave request to Peer *B*, and passes the address information recorded by Peer *A* to Peer *B*. After receiving the message from Peer *A*, Peer *B* starts to update its routing table and notifies the neighbours of Peer *A* to change the IP of Peer *A* to the IP of Peer *B*. Finally, Peer *B* sends a "Leave Confirmation" message to Peer *A*.

[Step 3]: If the Peer *A* is a Super Peer, the Backup Super Peer *C* is regarded as a Super Peer in this area, and a new Backup Super Peer *B* is selected again. And all Peers in the cluster will also update the information of Peer *C*, and Peer *C* will update the routing table through information exchange.

4.3 The Search of Peer in ICP4P

If I_α is an interest cluster, *A*, *B*, *C*, *D*, *E*, *F*, and *G* represent different Peers, respectively, and *S*, *I*, and *J* represent Super Peers, respectively. When Peer *A* wants to query the I_α and perform data Peer search, there are three steps:

[Step 1]: When Peer *A* issues an inquiry to I_α, Peer *A* first determines whether it is at I_α.

[Step 2]: If Peer *A* is in I_α, Peer *A* sends a request for the required file to Super Peer *S* of the region.

i. If the I_α where the Super Peer *S* is located has the file desired by Peer *A*, and the file exists in Peer *B*. Then *S* will pass the query requested by Peer *A* to Peer *B* with the file, and Peer *B* will respond to Peer *A* and start transmitting the file.

ii. If the I_α where the Super Peer *S* is located does not have the file desired by the Peer *A*, the Super Peer *S* transmits the message to the Super Peer *I* of the same interest in the Cluster layer. Super Peer *I* finds a Peer *C* that owns the file that Peer *A* wants, and Peer *C* responds to Peer *A* and begins transmitting the file. If Super Peer *I* does not have the information required by Peer *A*, Super Peer *I* will pass the message to the Super Peer of the same region of interest, and the Super Peer will find a Peer *D* with the file that Peer *A* wants, then Peer *D* will respond to Peer *A*. Then, the file is transferred.

[Step 3]: If Peer *A* is not in I_α, Peer *A* will send an inquiry to the Super Peer *S* in this region that it needs the file. The Super Peer *S* will ask the Peer *S'* in the Hypercube layer, and *S'* will ask the Peer to the Cluster layer.

i. If the Peer has the interest file desired by Peer *A*, Peer *E* responds to Peer *A* and starts transmitting.

ii. If the desired file of the Peer *A* cannot be found, the Super Peer *S* passes the query to the Peer *F* in the Cross-Hypercube management layer, and the Peer *F* queries the other management Peers through the iTracker. After the Peer *G* is found, this is the Peer *A*'s desired interest file, Peer *G* will respond to Peer *A* and start transmitting files.

4.4 The Query of Super Peer in Different Layer

In this subsection, the file query mechanism for each layer in ICP4P will be described, including the Cluster layer, the Hypercube layer, and the Cross-Hypercube management layer.

4.4.1 The File Query of Cluster Layer

The file query at the Cluster layer can be divided into two categories, namely, file queries in the same interest cluster in the same region and file queries in the same interest cluster in different regions.

The steps for file query within the same interest cluster in the same region:

[Step 1]: The file query of Peer *A* is sent to the Super Peer *S*.
[Step 2]: The Super Peer *S* finds Peer *B* with the file in the same region.
[Step 3]: Peer *B* passes the file to Peer *A*.

The steps for file query within the same interest cluster in different region:

[Step 1]: Peer *A* issues a file query to the Super Peer *S*.
[Step 2]: The Super Peer *S* cannot find a Peer with an archive in the region and forward the file query to Super Peer *I*.
[Step 3]: Super Peer *I* finds Peer *C* with the file.
[Step 4]: Peer *C* passes the file to Peer *A*.

4.4.2 The File Query of Hypercube Layer

In every region of each interest cluster, the Peer with the best capability value is selected as the Super Peer, and all selected Super Peers in the same interest cluster are connected in the way of Hypercube. When a Peer in the cluster issues a file query of other interests, the local Super Peer searches for Peer in the region firstly, thereby reducing traffic on the ISP.

[Step 1]: Peer *A* issues a file query to Super Peer *S*.
[Step 2]: Super Peer *S* forwards the file query to Super Peer *I*.
[Step 3]: Super Peer *I* finds Peer *B* with the file.
[Step 4]: Peer *B* passes the file to Peer *A*.

4.4.3 The File Query of Cross-Hypercube Management Layer

The best performing Peers connected by the Cross-Hypercube management layer will be selected as the management Peer. The selected Peer is responsible for searching for files in different regions of interest. Also, the selected Peer is linked to the iTracker and other regions with the best features.

[Step 1]: Peer *A* issues a file query to Super Peer *S*. At this time, there is no file inquired by Peer *A* in the cluster, and the requirement is moved to the Hypercube layer.
[Step 2]: If the Hypercube layer does not have the file queried by Peer *A*, the Super Peer *S* forwards the file of Peer *A* to the Super Peer *J* in the Cross-Hypercube management layer.
[Step 3]: Super Peer *J* finds the Peer *D* that owns the file queried by the Peer *A*.
[Step 4]: Peer *D* passes the file to Peer *A*.

5. Experiment

The simulation experiments conducted in this study will be carried out under the following experimental assumptions.

(1) In a heterogeneous P2P network environment, the hardware devices of each Peer may be dissimilar.
(2) Each Peer needs to provide information about this Peer to the iTracker during the Inactive Phase (e.g. CPU processing speed, memory size, ..., etc.).
(3) No damage or failure occurs when the file is transferred.
(4) Each Peer is assumed to be assigned only to one interest cluster.

In this section, the simulation results of this study will be explained. The experimental simulation consists of two parts:

(1) When searching for files, compare the average search time required for file search by “not searching for files with interest clusters” and “using interest clusters”.
(2) The data transmission using the Super Peer architecture, the delay caused by point-to-point.

5.1 The Comparison of the Average Search Time Required for File Search

In the experiment, there are 4 interest clusters, 8 regions, and 1024 Peers. According to Eq. (2), i is the total number of Peers and i=1024, j is the number of factors and j=3. After normalizing the CPU capacity, memory capacity, and transmission rate of the Peer. According to the capability value of Eq. (2), the interest clusters and Super Peers can be elected.

$$V_i = w_1 * (c_i) + w_2 * (m_i) + w_3 * (t_i);\ 1 \leq i \leq 1024;\ \sum_{j=1}^{3} w_j = 1 \qquad (2)$$

In this experiment, the speed of file query is emphasized, so a relatively large weight is given to the processing speed of the CPU, and the weight of the each factor is set to 0.6, 0.2, and 0.2, respectively. As for the interest cluster queries, the 80/20 distribution will be used, where 80% of the queries occur under the interest clusters to which the Peer belongs, and 20% of the Peers have the file of the query.

In the case of file search in this experiment, it will be divided into (1) not searching for files in the interest cluster, and (2) searching for files in the form of interest cluster. Figure 3 shows the comparison results of the experiment. The X-axis is the order of file search, and the Y-axis is the time of file search. The time unit is second.

Figure 3. The time of file searching.

The experimental method is to distribute the files evenly to each Peer, the number of Peers is 1024, and each Peer is assigned 1 file to calculate the total search time of the file. It can be seen from the results shown in Figure 3 that the file is searched without using the interest cluster, because the file is a random search, so the search time required is longer, and the average time is about 0.194066667 seconds. If the file search is performed in the interest cluster, because the files are clustered by interest, the traffic can be reduced and the search efficiency can be improved. The average search time is about 0.1083666677 seconds. From the comparison of the average search time of the experiments, the search method showing the interest cluster is significantly better than the search without the interest cluster.

5.2 The Data Transmission Delay Caused by Point-to-point

This study simulates the data transmission of interest clusters. The experiment uses a total of 16 Peers, which respectively represent 2 different interest clusters. Each interest cluster has 8 Peers, which respectively represent super Peers in different regions. Among the eight Peers, a Super Peer will be selected as a Peer of the Hypercube layer to perform file query to measure the point-to-point delay result of the interest clusters. In the experiment, 10 data packets are transmitted from the interest cluster *A* to the interest cluster *B*. The link path between each Peer is 10Mps bandwidth and 20ms transmission delay time. The file query of each source Peer will pass through the Super Peer. After processing, it is sent to the destination Peer to reduce the query traffic.

Figure 4 shows the result of point-to-point delay using the Super Peer architecture for data transmission. The X-axis is the transmission time, which is about 1 second, and the Y-axis is the point-to-point delay time, which is about 0.8 to 0.9 seconds.

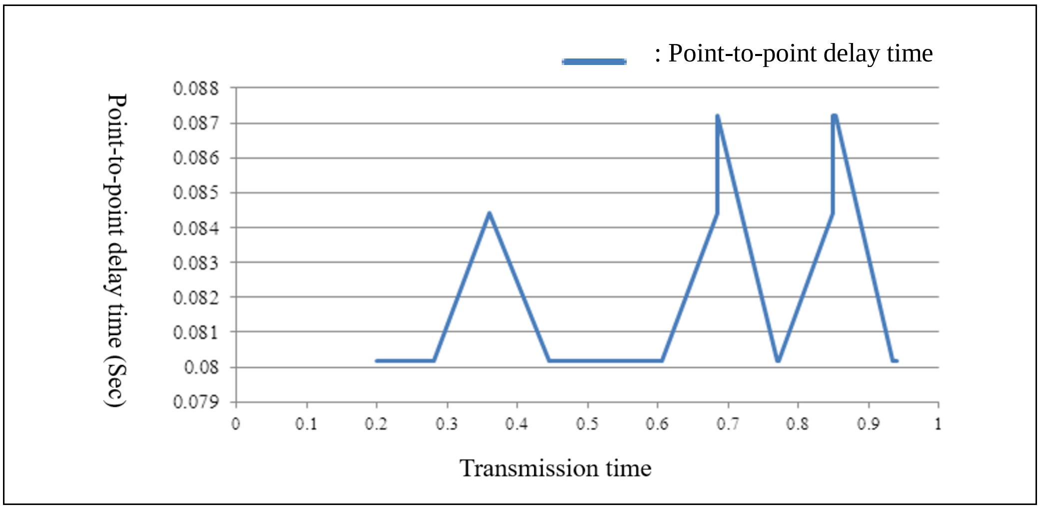

Figure 4. Point-to-point delay time.

6. Conclusion

Since too many P2P applications may result in inefficient use of network resources or inefficient application execution, the ICP4P architecture is built in conjunction with the P4P framework and interest cluster in this study. Through the ICP4P proposed by this research, the traffic generated by P2P application software can be reduced, and the interest cluster can be used to reduce the time of information search and improve the efficiency of search. The ICP4P proposed by the study is a P4P framework that combines interest clusters and enables ISPs and P2P service providers to communicate with each other. ICP4P can not only reduce the traffic of ISPs across regions, but also improve the service quality of P2P.

Due to the Peers in the system usually have multiple interests, the design of the addressing will be considered in future research, making the ICP4P more in line with the actual application. In addition, heterogeneous Peers, such as the size of the storage capacity, CPU capabilities, and bandwidth capabilities, will affect the overall architecture design. Therefore, when selecting a Super Peer, the stability of the Super Peer must also be taken into account, so that changes in the cluster will not have much impact on the system.

Peer Cluster proposed by Huang *et al*. uses the architecture of Hypercube to solve information duplication and utilize interest clusters to increase search performance [15]. Hu proposed a hybrid P2P interest clustering system (LaHiG) with position awareness to solve the mismatching problem of Peer Cluster [6]. Lin proposed a three-layer Peer network (TLAIG) with location awareness and interest clustering to improve the performance of LaHiG [16]. The ICP4P architecture proposed in this research is a P4P framework combined with clusters of interest, which enables ISPs and P2P service providers to have a communication platform. ICP4P can not only reduce the cross-region traffic of ISP, but also improve service quality of P2P.

Acknowledgment

This work was supported in part by the Ministry of Science and Technology MOST 107-2221 -E-324-005-MY3.

References

[1] Nobre, J. C., Melchiors, C., Marquezan, C. C., Tarouco, L. M. R., and Granville, L. Z. 2018. A survey on the use of P2P technology for network management. *Journal of Network and Systems Management*, 26, 1: 189-221.

[2] Ijaz, H., Welzl, M., and Jamil, B. 2019. A survey and comparison on overlay-underlay mapping techniques in peer-to-peer overlay networks. *International Journal of Communication Systems*, 32, 3: e3872.

[3] Neumann, C., Chandrashekar, J., and Le Merrer, E. 2019. *U.S. Patent Application*, No. 16/025: 175.

[4] Steinmetz, R., and Wehrle, K. 2005. What Is This “Peer-to-Peer” About?. *Peer-to-peer systems and applications*. Springer, Berlin, Heidelberg: 9-16.

[5] Yang, J., Zhong, Y., and Zhang, S. 2003. An efficient interest-group based search mechanism in unstructured peer-to-peer networks. In *Proceeding of IEEE 2003 International Conference on Computer Networks and Mobile Computing*: 247-252.

[6] Hu, J. J. 2007. Hybrid P2P interest clustering system with location awareness. Master's Thesis, Department of Information Engineering, National Central University, Taoyuan County, Taiwan

[7] Sahhaf, S., Tavernier, W., Colle, D., and Pickavet, M. 2016. Resilient availability and bandwidth-aware multipath provisioning for media transfer over the internet. In *Proceeding of IEEE 2016 8th International Workshop on Resilient Networks Design and Modeling*: 134-141.

[8] Guo, Y., Liu, L., Wu, Y., and Hardy, J. 2018. Interest-aware content discovery in peer-to-peer social networks. *ACM Transactions on Internet Technology*,18, 3: 1-21.

[9] Xie, H., Yang, Y. R., Krishnamurthy, A., Liu, Y. G., and Silberschatz, A. 2008. P4P: Provider portal for applications. *ACM SIGCOMM Computer Communication Review*, 38, 4: 351-362.

[10] Li, L., Wang, F., Yu, W., Mao, X., Yang, M., and Li, Z. 2018. An improved design of P4P based on distributed tracker. *International Journal of Information and Communication Technology*, 13, 3: 340-350.

[11] Ying, Z., Sining, W., Zhiyong, L., and Qiuhong, S. 2017. Research on P4PCDN automatic construction strategy based on network measurement. *Procedia Computer Science*, 107: 490-497.

[12] Gupta, R., Singha, N., and Singh, Y. N. 2016. Reputation based probabilistic resource allocation for avoiding free riding and formation of common interest groups in unstructured P2P networks. *Peer-to-Peer Networking and Applications*, 9, 6: 1101-1113.

[13] Hauge, M., and Landmark, L. 2018. Limiting the flooding of simplified multicast forwarding to a defined scope. In *Proceeding of IEEE 2018 International Conference on Military Communications and Information Systems*: 1-8.

[14] Tu, Z., Jiang, W., and Jia, J. 2017. Hierarchical hybrid DVE-P2P networking based on interests clustering. In *Proceeding of IEEE 2017 International Conference on Virtual Reality and Visualization*: 378-381.

[15] Huang, X. M., Chang, C. Y., and Chen, M. S. 2006. PeerCluster: a cluster-based peer-to-peer system. *IEEE Transactions on Parallel and Distributed Systems*, *17,* 10: 1110-1123.

[16] Lin, Z. J. 2011, Three-tier peer network with location perception and interest grouping. Master's Thesis, Department of Information Engineering, Truth University, New Taipei City, Taiwan.

[17] Tussyadiah, I. P., and Zach, F. 2017. Identifying salient attributes of peer-to-peer accommodation experience. *Journal of Travel & Tourism Marketing*, *34,* 5: 636-652.