Research on resource discovery and data traffic control based on peer to peer networks

Yang Yansong, Zhang Ning, Li Jun and Zhu Chenxu

College of Information, Beijing Union University, Beijing, China

_____

ABSTRACT

With the widespread popularity of the Internet, a rapid increase in network bandwidth and the computing power attracted more attention. P2P (Peer-to-Peer Network) network and its application has gradually become a research hot spot. In P2P network, resource discovery and network traffic technology play an important role for a high efficiency network. In this paper we analyze the resource discovery, proposed a novel method to find network resource. With the new method, the load balancing between the network nodes can be guaranteed. In addition, an investigation, through the network traffic generated by P2P application, was examined. In order to improve the operating efficiency of the P2P network, we presented a new routing method combined with the shortest path to predict traffic congestion state and local information. Simulation results show that our proposals are feasible.

**Key words**: Peer to Peer networks; Resource discovery; Traffic control; Load balancing; Routing
_____

INTRODUCTION

The rapid development of the Internet, is triggering a new network technology innovation. With the widespread popularity of the Internet, a rapid increase in network bandwidth and the computing power attracted more attention. P2P (Peer-to-Peer Network) network and its application has gradually become a research hot spot. P2P network is to establish a logical overlay network on top of a layer of the Internet [1]. It features very significant, Including non-centralized network node (to avoid network bottlenecks), scalability (network nodes are added, the overall expansion of resources and services the ability to synchronize), robustness (partial node failure has little effect on other parts), lower cost computing and storage capacity higher, peer network nodes, the network load balancing, transmission of information makes the private information dispersed eavesdropping and reduce the possibility of leaks, freedom of user behavior and a wealth of information content.Powerful service node has direct access to the backbone network [2]. A large number of services are being gradually transferred to the nodes near the center of the network. Not only has the backbone network transmission capacity, but also to the intelligence network with storage, computing and other functions change [3-5]. As the center of the optical backbone network architecture, that provides user access control, quality of service, negotiation mechanism and also to provide better SLA support [6]. The mass intelligence nodes, which are at the network edge provide ubiquitous, ever-present perception. Real-time processing of large amounts of information will be collected to be passed elastic computing data center for information processing [7]. Transmission network carries more than just information, share content, access services and other technical aspects of the function in the future, but also carries a commercial activity on the network. According to the 2012 Global Network traffic monitoring report, P2P, WWW have occupied 75 percent of network traffic [8]. Internet users has more than just consumers of content, but also through P2P, micro blogging and other ways to become a content provider. Users contribute content on the Internet, and make the spread of content. Internet into a more open, more interactive, user-supplied content and participate in the joint construction stage [9]. Each peer P2P networks (peer) must take on the role of two functions, they are both the service provider and a service user. Ownership and control of resources distributed to each node in the network. Behavioral characteristics of P2P networks including peer connection mode, Popularity, perturbation [10]. T.Karagiannis et al found two

_____

characteristics of P2P network transport layer connection: One is about 2/3 of the P2P applications using both TCP and UDP protocols. Second, in the P2P file-sharing networks, it is usually using only one TCP connection for file transfer between peers. F, Comtantinou et al studies indicate that P2P network having a greater diameter compared with the logical network topology formed by other network. Some P2P applications when used to initiate a large number of connections increase suddenly popular, which is a characteristic behavior of the P2P network [11]. Popularity refers to the number of the network to establish a connection with a host at the time within t. In this paper, based on the P2P networks we conducted analysis deeply [12]. In P2P networks, we use a hybrid transmission routing topology model to achieve routing and locate resources. In this model, a node will be divided into two categories, the agent nodes and resource nodes. Resource node, that will provide all of the shared resource in system. It also needs to ensure that the load balancing is for each resource node. So that resources are evenly distributed [13]. In terms of network traffic, network traffic generated by P2P applications has become a major part of the Internet traffic [14]. The rapid growth of P2P network traffic increased network bandwidth consumption. It is near-symmetrical network traffic patterns caused by network congestion [15]. A more serious problem is that P2P applications initiated by network attacks, it is likely to cause failure of the Internet. Therefore, the identification and analysis of P2P network traffic has become a common concern hot topic of scientific research personnel, network management, network service providers. For quick relief network data traffic, it is very necessary to design an efficient routing algorithm. Through deep study, we proposed a new routing algorithm, which is a combination of the shortest path and predicted traffic information for local congestion state, thereby it improves the operating efficiency of the network.

NETWORK RESOURCE DISCOVERY
The current P2P network search, the main goal is to improve search quality from a client or reduce the search from the network management point of consumption, this search method, often focusing on the improvement on the search algorithm, while ignoring the transmission delay between nodes and downloading from the physical location of the delay resulting performance of the search, Two node ID may be close to the actual physical location far, one jump in a P2P layer may be multi-hop to the IP layer, this increases the query forwarding process between nodes in the request delay and download delay. The computer network is shown in figure 1.
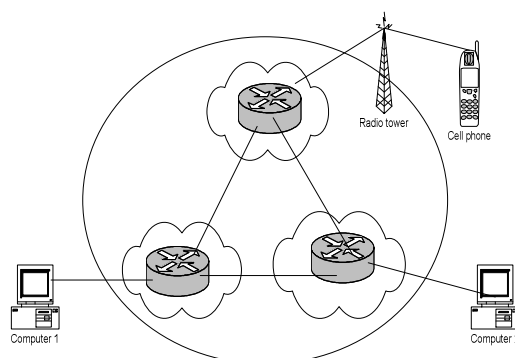
Fig. 1: Communication structure of computer network

P2P resources search algorithm based on node position feedback, which main idea is to divide network node into group. Group or Clustering is the original node disorder divided by a common attribute. Conducting resource search, a node which set up the find can initiate a request submitted to the nodes of the cluster or the domain of interest. If the lookup fails, the search was extended to a wider area, or flooding Search. In recent years, P2P technology is widely used, it is one of the most popular applications in unstructured P2P file environment (unstructured P2P) under shared. The main reason is that the unstructured P2P network topology search mechanism is simple and easy to implement. The main objective of P2P file-sharing system in continuous improvement and evolution, improvement is: find a better search mechanism, making the system both good search results, there are high search efficiency. In fact, as a resource to find performance of P2P is an important research P2P application. Agent-based distributed hierarchical routing mechanism star topology layer node index table positioning strategy and ring topology layer, constitutes a double-layer structure routing policy. In the index table positioning strategy star topology layer, proxy node played a crucial role for the network to quickly locate the information resource query target, and provides a shortcut. In a distributed hierarchical routing mechanism ring topology layer, the layer-dependent resource node is routing table and routing information query sequence to determine the resource location. More than two routing strategies are complement, each node carry out their duties, resulting in a P2P network system layers and form a stable query routing strategy. For each route choice mechanism, it is through the work of the state agency to determine the nodes, namely the agency work state and the agent failed state. Resource node information update is primarily responsible for the network login and exit of each node. When a new node

_____

joins the network, the node sends a request to the proxy node login information, then the corresponding node agent login information recorded at the same time to add the node to the shared resource information resource index table to prepare for resource retrieval. When a node leaves the network resources, the same request to the proxy node sends a quit message, the node to be deleted then the information recording and corresponding shared resource information and resources from the node information table in the index table. For the above processing strategy , it applies only to the case agent nodes work properly; However, when the agent node fails, it will not be completed properly add and delete node information, which will cause the node information table records the actual status of each node in the network inconsistencies. Another major portion of the agent node updates the index table is updated resource index information. For this part of the update process, the original old algorithm early in the design, try to use to update the strategy. The basic principle is in this case, each node is in the event of a change (adding a node, or decrease a node), then this node will go to send a message to the node agent, told the agent node information table changed, added a new member, or reduce a member. In this way, an obvious drawback is that, you need to spend some network resources. In addition, the agent node increases the processing time for information. What is worse is that if there are multiple nodes simultaneously take action to send updates, which will cause network congestion channel. This will seriously affect the normal travel of the network, reducing the efficiency of the network.
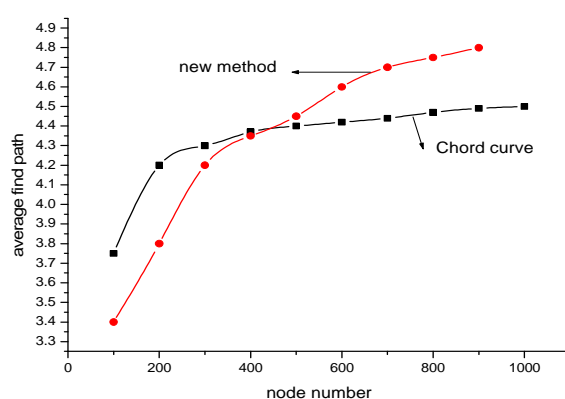
Fig. 2: The relation between node number and average find path

The new improved algorithm, adopted by the need is to update the way. It can be the agent node update task to carry out a processing time. In the new network, there is a new node is added, the agent node in the network is resource retrieval first, proxy node does not temporarily update its information. When a node is access to the relevant information from the new resources, then it will take an action. All resources are updated to the node resource index table. Another situation is when the network exit nodes, then to conduct inquiries, it first check out the failed node, Specifically, when the discovery of a failed node, it sends a connection request fails, then the node relevant information will be automatically deleted. Action time of agent node also is after the inquiry to be carried out. This new routing strategy proposed in this paper, is good for the full utilization of network resources. It can save network resources and can be a good solution to the proxy process node failures that may occur in the case of inconsistency in resource information, thus it avoids the impact on the resource update. In addition, due to the condition of each node access is controlled, and therefore, it will not cause network congestion. In order to assess the actual effect of the new algorithm, the performance of the new algorithm is conducted for a series of simulation experiments, in order to verify that the new algorithm is to improve the performance of the network system. We evaluated by comparing various indicators, such as: the average amount of information transmission, routing, time, to verify the validity of resources routing policy. After analysis we concluded that "the average search path" in this article to build the routing policy is better results. Meanwhile, in order to reflect the relative strengths of the other routing policy, this paper will also check it with Chord network routing algorithm for locating experimental for comparison. For the network, the data transmission cost is relatively large, and therefore the bandwidth used on larger. It is mainly because the node mapping use disposable Chord hash into a unified logical space, this eliminates the contact node physical location, and it does not make full use of the advantages of data locality. Every time a user downloads the file, who may be distant from the actual physical location of the node to download. The two-tier structure of the new algorithm routing mechanism is to reduce the length of the path and query information transmission cost routing process consumed. Positioned directly limited by the index table or query routing table, it can get the position of the node resources, so that the average search path and data transmission bandwidth have declined. We conducted a simulation experiment, the number of nodes in the network from small to large increases from 0, have been increased to 1000, and then view the results of the average search path. Here is important to stress that we are used to represent the average number of hops to search the effect of the average search path. In general network, the average search hops should be between 3 to 6, which is normal. If the result is outside the scope of this value, it

indicates that there may be a prerequisite for one or more errors. It must be a prerequisite for re-correction until it can get the result of the normal range so far. Our experimental results are shown in Figure 2. The results can be seen from the figure, the start of the first stage, the curve Chord algorithm is higher than the results obtained we propose a new method. The intersection curve of two algorithms at the position of about 410, after about 410, the curve is higher than the results obtained by the Chord algorithm. That is, in the latter part of the entire range of the node, the value of the proposed new algorithm is significantly higher than the Chord algorithm.

In our new algorithm, each routing table we have designed a plurality of layers and each layer can be stored in another node in the m position. According to the new routing algorithms and strategies, each resource node routing table, have provided important support information. And the obvious fact is that, m value is bound to affect the size of the network nodes to find routing efficiency. Our results also confirmed this point. As shown in Figure 3. It can be clearly seen from Figure 3, the mean value of the resource search path U decreases with increasing the number of nodes m of the network gradually, and this result is consistent with our expectations. From the change in the curve in Figure 3, we also found that with the increasing number of network nodes, network resources it takes to find the behavior has also been reduced, the cable is a slow downward trend, and finally became a relatively flat. It also shows the fact that although it takes a certain behavior to find valuable network resources, but with the increasing number of network nodes, network resources to find the amount spent on behavior change is not large. In other words, an increase in the number of network nodes, there is no significant impact on network resources it takes to find the behavior.
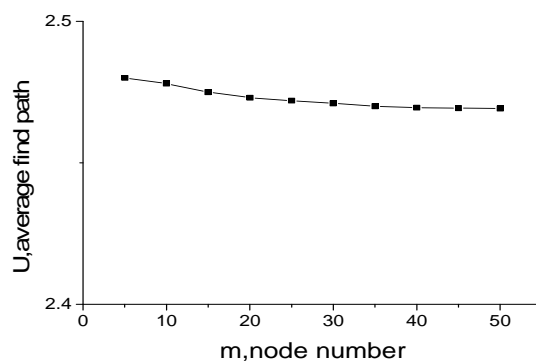

Fig. 3: The relation of the average find path and node number

In P2P network topology, it uses a hybrid routing transmission model to achieve the route and location of resources. In this model, the nodes are divided into two categories, resource nodes and agent nodes. Resource node, the system will provide all the shared resources, but also need to ensure that each resource is load balancing node, so that resources are evenly distributed. Proxy node, as a structural model of another type of node, compared to resource node, is very different in terms of functionality. On the one hand, this type of node stores the same data, but the data are no longer the resource data, but the position information of each resource node of the system. Each resource node on the resource index information is stored; such node also bear part of the resource locator function, in order to help the resource nodes faster and more stable for storage and retrieval of resources. In order to improve the resource retrieval speed and reduce the consumption of the search process, reducing energy resource update, here we also proposed hierarchical routing service strategy and its updating policies. Through the transmission route in the system structure of the model to achieve P2P layer simulation, while Chord comparison with analog data, it can be drawn to improve in terms of the speed of the resource location, and decreased consumption in terms of updating. However, due to the special nature of the agent node which provides additional data storage, making the system applied in the case of transmission structure model, adds the certain storage consumption.

NETWORK TRAFFIC CONTROL
With the proliferation of network size and the amount of information, congestion becomes a common reality of the dynamic characteristics of the performance of the network in the form. Many real networks, such as Internet and network public traffic occur congestion frequently, leading to a transmission time with a significant increase and overall network performance occur sharp decline. Therefore, the network characteristics were studied by many researchers. Not only a large number of data streams generated by network nodes, especially concurrent data streams, is the main reason leading to network congestion generated, and some characteristics of the network itself has also produced and spread of congestion, and such as the capacity of the network nodes, the node packet processing speed, the bandwidth of the communication link of the network topology and the like will play a role. Although it is possible to achieve the effect of improving congestion by upgrading router hardware, but in reality it can not be unlimited hardware upgrades. In addition to economic considerations, the rapid growth of the network size also

_____

makes hardware upgrades can bring improvement effect is very limited, so we need to consider other ways to improve the performance of the network, such as designing more efficient routing strategy. P2P also has important applications for streaming network. Centralized P2P network structure is the central server as the core information resources on a central server, the client is sent directly to the server, access to information resources. Such as Server / Client mode, database systems, Web services, e-commerce services. In a centralized P2P streaming service system, the system records through a central index of current directory server nodes in the P2P network status information and maintain the network directory. The central server is responsible for searching and locating resources, data transmission and interaction are done directly between the nodes. Centralized P2P network system to search and locate resources, the transmission and interact shown in Figure 4.
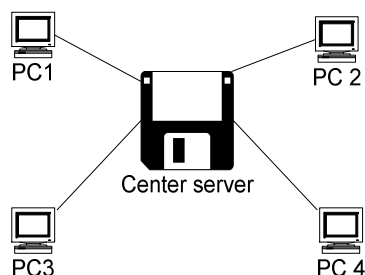

Fig. 4: The center network system

To unstructured P2P networks, there is no central server, which is based on random graphs way to connection between the structure of production nodes in distributed network. The system has a strong ability of anti-jitter tolerant. As shown in Figure 5. Generally, it uses flooding (Flooding) way to search target in the whole network resources, support vague, with regular expressions and more complex keyword queries to TTL (Time to Live) values to control the search depth.
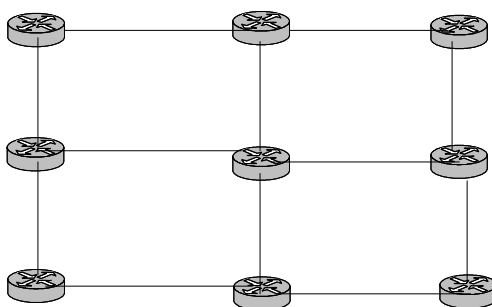
**Fig. 5: The peer to peer network structure**

The shortest path routing algorithm is a kind of the most widely used routing strategies. By strategy, in which the packet is always along the path passed closest to the destination node. But the algorithm never considers whether or not the network is in a state of congestion. Even in routing by the number of node, performance fell sharply, due to the accumulation of a large number of data packets that fail to pass it a packet, the shortest path routing policy did not provide a mechanism to choose new routing path for the packet. Because the most prone to congestion nodes located in node is opposite bigger, so the shortest path routing algorithm will be more congested node congestion, leading to decline in network performance further. For fast relief in the network data flow, design an efficient routing algorithm is very necessary. A good routing algorithm should have some ability to perceive congestion and regulatory, enables it to early detection of network congestion occurs, and timely adjust the routing policy in order to control at early date. Network dynamic process description is as follows: suppose network starting without load, for each time step K a packet network, and randomly selected source node and destination node, the ability of each node packets processed by node n to the importance of the assignment, each node per unit time can be at most n, a packet to its next routing nodes, packets generated after, is added to the end of the queue, the node queue that contains both the packet and other nodes to arrive. Assuming that node queue length is unlimited, each node as the packet is passed, the first to search within the adjacent nodes, If there is a target node adjacent nodes, it is passed directly to the destination node. If not, then the calculation of all of the neighbor nodes is done in the node weights, according to the minimum weight so that pass it on to the next node. After the packet to reach the target node will disappear in the network. Here has the sense of local congestion ability of routing policy. Assumes that the packets is from the source node S to pass the destination node t, first of all, the algorithm needs to compute nodes s weights of all the neighboring nodes, in order to make the choice of which one sent to the adjacent node. Figure 6 shows the P2P

network structure.
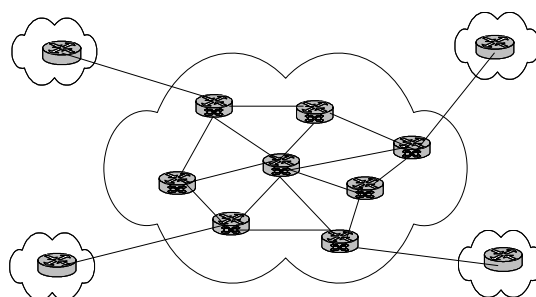

Fig. 6: P2P network structure

In real communication network, the important router nodes usually has a better performance, the ability of processing data packets are more, so can be by Ni, the value to reflect the situation. Ni, the value is different according to the importance of nodes in the network. Because of the node degree is an important level of node in the network. Traffic identification algorithm based on port is the most common recognition algorithm, principle is the main use of the data flow of packet head information. According to the different protocols it uses different ports for identification. At present, most of the network data capture and protocol analysis software has the function, such as Ethereal, Wireshark, Sniffer, and so on. According to the specific application protocol, port number, it is corresponding to identify specific data stream. Such as a TCP data flow using port 80, according to the relevant provisions of the management of the port, it will mark it as a Web data traffic. Communication network performance can be achieved by the entire network for packet processing and transmission capacity measure, however, network communication ability is not the same as the communication capability of a single node. Here, the capability of communication for a single node is n, and the communication capability of the entire network with a critical discharge of packets Kc to measure, namely the critical discharge packet Kc, is continuous phase transition from free state to congestion state. Peer-to-peer (P2P) network traffic identification and analysis has become a common concern issues for scientific research personnel, network management and network service providers. In terms of scientific research, from two dimensions of time and space for P2P network traffic measurement and analysis, not only can know all kinds of the heat of the P2P applications, can also be carried out on the user's network behavior research and statistical analysis, to ensure the safety of the network, to design a robust and efficient network to provide the scientific basis. In terms of network management, accurate identification, extraction and describe the characteristics of P2P network traffic, building the corresponding mathematical model of network traffic can not only provide effective scheduling strategy for network management, at the same time it can also ensure that the P2P application quality of service (QoS) of the business, provide a basis for effective control of P2P network flow. In terms of Internet service providers, P2P network flow measurement and identification can provide the necessary reference basis and bandwidth for network construction planning. There is a free state concept in the network traffic. Free state, the concept so-called network refers to such a situation, while at the same time, the number of packets by production, as well as the number of packets reaches the destination node, we call it the free state of the network. Relative to the network of free state, the other is the congestion state of the network. So-called network congestion state, it is to point to in the packet, the number of packets reaches the destination node is only a fraction of the number of packets, at this point, the redundant data packets will pile up in the system, until finally the whole system paralysis. For freedom in the network configuration and network congestion state, we can describe quantitatively. When J = 0, this network is in a free state, if J > 0, suggests that the network into the congestion state. We can also use the total number of packets S (t) in the network and network packets K generated to describe the network status. Among them, there is a critical state Kc, its meaning is, when it is in the network number Kc of packets generated , the network packets number generated will be greater than the number of packets to the finish, now the network will be in the position of the transition point, we call it a critical state. Apparently, according to our definition, the meaning of the K value is very obvious, when K > Kc, network must be in a state of congestion, at this moment, it is much bigger than the number of packets in the network to reach the number of packets in the end. When K < Kc, the network is in free state, that is to say, at this moment, the network packets is almost equal to the number of data packets to the finish. Particularly we want to point out is, the larger the value of J, the network will show that the bigger the network congestion state, when J = 1, it shows that packets didn't come out at this point in the network, the network is completely in the congestion state. Figure 7 shows the relation of total packets s(t) and time. Figure 8 shows The relation of s(t) and t, K=1200.
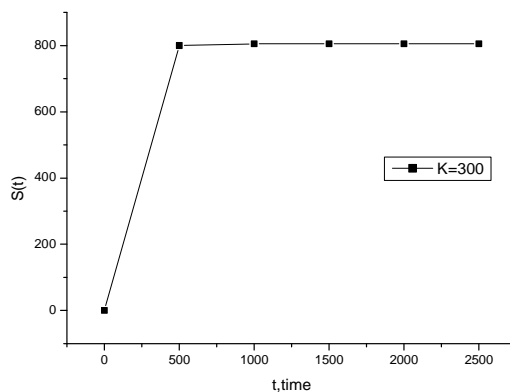
Fig. 7: The relation of total packets s(t) and time
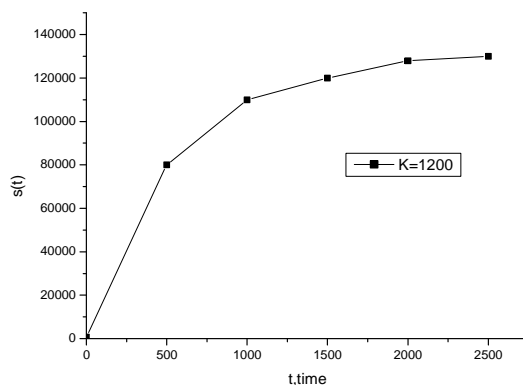
Fig. 8: The relation of s(t) and t, K=1200
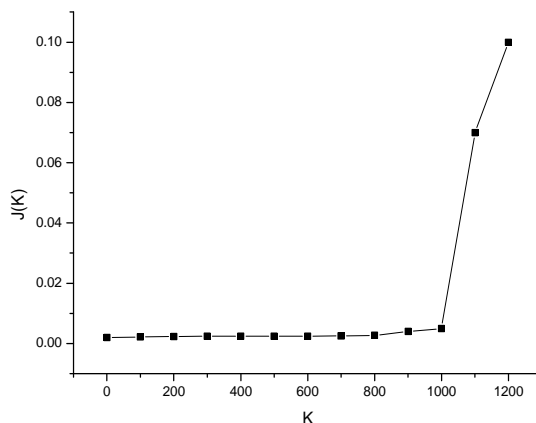
Fig. 9: The J(K)in new routing method

From figure 9 we can find, when the K value is small, the curve of the network is horizontal, it shows that the network at this point is in the normal operation state, and it is no network congestion. However, when the K value is higher, the network will begin to enter a degree of congestion state, for example, when K = 1000, it have taken place in a certain degree of network congestion. K = 1200, then the network congestion state is very obvious. Figure 10 shows Old routing method with the shortest path routing.
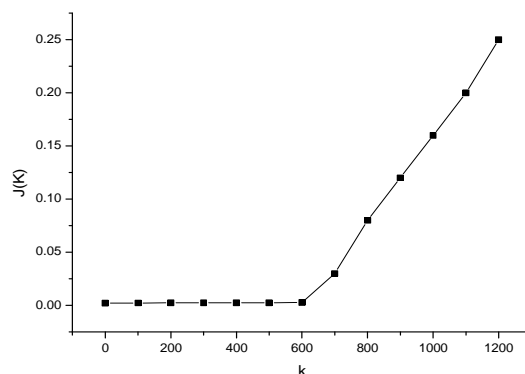
Fig. 10: Old routing method with the shortest path routing

CONCLUSION

In recent years, P2P technology has been rapid development. In P2P network, how to make full use of network resources, how to reduce unnecessary costs, network resources and network routing node optimization search, the P2P network efficient operation, these are very important problems for network. In the process of network resource requirements, network node need to send a resource request to the server inside the network, and then by the server to resources within the search, and it finally will be sent to the resources outside the network nodes. The data transmission strategy depends on the server, and thus caused the collapse due to server failure brought about by the network. This paper proposed a new P2P network routing lookup method, the method ensures the load balance between nodes, at the same time improve the resource searching performance. In P2P network routing uses a mixed topology transmission model, to implement the routing and location of resources. In this model, the nodes are divided into two categories, resources node and agent node, thus it will make the optimization of the network efficiency.

REFERENCES

[1] Zhong M., Shen K., Seiferas J., *IEEE Transactions on Computers*, **2008**, 57 (5), 619-633.
[2] Ming X., Shuigeng Z., Jihong G., *Journal of Network and Computer Applications*, **2010**, 33 (2), 115-127.
[3] Ramachandran K., Sikdar B., *IEEE Transactions on Parallel and Distributed Systems*,**2010**, 21 (3), 361-378.
[4] Yigitel M. A., Incel O. D., Ersoy C., *Computer Communications*, **2011**, vol.34, 1991-2001.
[5] Hu Q., Tang Z., *Journal of Multimedia*, **2011**, vol. 6. no.2, 122-128.
[6] Wang G., Gui X. L., *Journal of Multimedia*, **2012**, vol.7, no.4, 279-288.
[7] Tang Y., Xiang Y., et al. *Journal of Computers*, **2012**, 7(1),11-18.
[8] Liu J., Wang X., Yang K. Zhao C., *Journal of Computers*, **2012**,vol 7, No 7, 1671-1680.
[9] Li W., Xia X., Gu D., et al. *Journal of Computers*, **2011**, vol 6, No 2 , 216-223.
[10] Zhu S., Li G., Han J., *Journal of Computers*, **2012**, vol.7, 1555-1563.
[11] Gamer T., *Computer Networks*, **2012**, 56 (1), 169-185.
[12] Ciccarelli G., Cigno R., *Computer Network*, **2011**, 55 (15), 3517-3532.
[13] Tian C., Yang B., *Future Generation Computer Systems*, **2011**, 27 (8), 1135-1141.
[14] Li D., Guo C., Wu H., *IEEE/ACM Transactions on Networking*, **2011**, 19(1), 102-114.
[15] Shiravi A., Shiravi H., Tavallaee M., *Computers and Security*, **2012**, 31(3), 357-374.