

Research on P2P Network Resource Search Model

Jing Leng^{1, a} and Tonghong Li^{2, b *}

^{1,2} Department of Information Technology, Hubei University of Police, Wuhan 430034, China;

^aemail:daleng0127@sina.com;

* corresponding author

Keywords: P2P networks; Search delay model; Simulation

Abstract. In this paper, the search delay model of P2P networks is proposed. By analyzing the search process of structured P2P network and using Markov chain, the relationship between the mean node session time and the search delay of the P2P network is obtained. This model shows the inherent law between P2P Churn and network performance, and it is verified by simulation.

Introduction

The main line of the development of the resource search technology of P2P networks is intertwined with the index information placement policy for P2P network resources and the process of research on the organization mechanism of P2P networks. At present, the research of resource search technology of P2P networks is mainly concentrated in two directions: The first one is to continue to study how to improve the quality of resource search by improving the organization mechanism, resource and the distribution mechanism of resource index of the P2P network. The second one is to try to find out the optimal search features of various P2P network models in order to improve the search performance. Various P2P search technologies currently used have some defects: The unstructured P2P network commonly uses controlled flooding mode, which consumes too much bandwidth and is unreliable. The search mechanism used in structured P2P network is difficult to support fuzzy search and semantic search. The structured organization mechanism used to improve the search efficiency brings high maintenance costs and is difficult to cope with the high volatility of P2P networks and other problems, such as churn.

In this paper, the principle of P2P network resource search mechanism, the mainstream P2P network model and its search mechanism and algorithm are studied, the search process in Chord protocol is analyzed[1], and based on the Markov model, a P2P network resource search model is proposed. The model reveals the impact of Churn on the search performance of P2P networks, gives the relationship between the mean session time of nodes and the search delay in P2P networks, and carries out the verification by simulations.

Overview of Resource Searching in P2P Networks

The P2P network, as an overlay network, is a content addressable application layer network[1]. The main goal of resource search and positioning is to find the “content pointer” corresponding to the “content” (For example, the IP address of the node in which the resource is located). In P2P networks, each addressable resource has an identifier ResourceID and a corresponding pointer P containing its address. The system saves <ResourceID, P>. When users need to search for resources, they should search for P according to the resource identifier ResourceID, so as to locate resources.

Classification of Resource Search Mechanism in P2P Networks

The organization mechanism of P2P network and the distribution mechanism of resource and resource index in the P2P network often restrict the basic methods and basic algorithms of the resource search mechanism. The current mainstream P2P system adopts resource search mechanism, which can be roughly divided into four categories: centralized lookup, flooding, selective

forwarding, and deterministic search based on mapping relationship between resource index information and its location.

Deterministic Search Based on the Mapping Relationship between the Resource Index Information and Its Location

The deterministic search based on the mapping relationship between the resource index information and its location is the basic way for the searching resources of structured P2P networks. Among the various schemes that construct the mapping relationship between the resource index information and its location, those whose impact is larger are all based on the distributed hash table(DHT). The basic idea of DHT is that when establishing the resource index table, the identification of resources is replaced by the hash value of resources. Each index entry is expressed as $\langle \text{key_id}, \text{URL}, \dots \rangle$ form. Among them, key_id is called a resource key value, which is a hash value of a resource name (or other feature information). url is the positioning information of the node that actually owns the resource(Such as IP address and TCP port or UDP port). In this way, after a resource index hash table is established, if you want to find a resource, you can find the address of the node that actually owns the resource in this table by giving the key value of the resource. At the same time, a hash value is allocated for each node as the identifier of the node, and a mapping relationship is established between the node identifier and the resource key value, which enables nodes to store and maintain the resource index information of a key value in a certain range according to this mapping relationship. In this way, the resource index table of the whole system is divided into several parts, and each node in the system is responsible for storing and maintaining a small part[3].

The P2P network resource search model based on DHT typically has the Chord model.

Chord Network Resource Searching Model

The Chord model makes its node identifiers form a ring topology. Every node and resource can obtain a m -bit identifier through the consistent hashing table. In the search process, the search message stops at the nearest predecessor node from the search key value, and the predecessor node returns its successor node as the search result. In a Chord network with a base value of b , each node maintains a routing table composed of $(b-1)\log_b(n)$ pointers. The node periodically detects the state of its routing table pointer pointing to the node. When the node which is pointed by the routing table pointer is a failure node, the node selects a new node through a search process to replace the failure node and maintain the validity of the nodes in the routing table. Each node maintains a list of successor nodes composed of several nodes at the same time to maintain the Chord's ring topology[4].

The Chord model practices the idea of a distributed hash table in a simple way. When a resource search is carried out, the number of times that the related messages are forwarded in the overlay network is $O(\log_2 N)$, which can be used to build a large-scale P2P system.

Search Delay Model

The Concept of P2P Churn

There are continuous nodes entering and leaving the P2P network, which we call P2P Churn, and it is one of the essential features of the P2P network. P2P Churn causes nodes in routing tables and the successor node list to fail, and serious churn will lead to P2P search timeout and even network partition[2]. At present, P2P Churn has become one of the main research directions in academia.

When Churn increases, the nodes in the network will enter and leave the P2P network at a higher frequency. We use the node session time as the indicator to measure Churn, and when the Churn increases, the node session time is shortened. According to the search process of the Chord protocol, the author establishes a "search delay model", and reveals the relationship between the node session time and the search delay through the model.

Analysis of the Search Process of P2P Network under Churn

Firstly, we analyze the impact of Churn on the search process in the network. According to the above analysis, when Churn increases, the mean session time of nodes is reduced[2]. When a P2P node P_0 initiates a search and the search path contains a failure node F , the search process is shown in Fig.1.

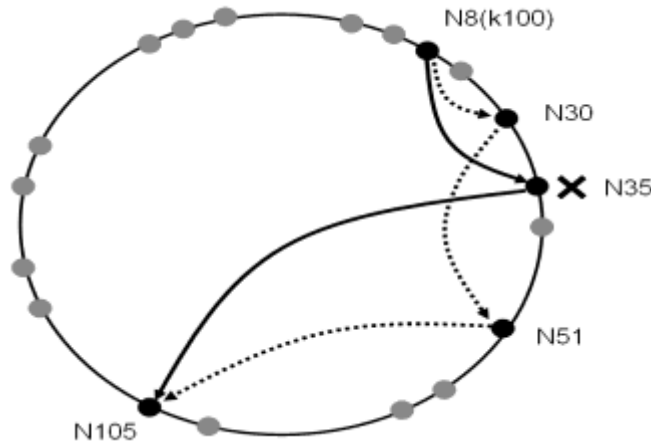


Figure 1. The search process in Chord

Node N8 initiates search, looking for key value $K=54$. Node N8, based on its own routing table, finds the P2P node N42 with the maximum node identifier but not more than K , and forwards the search message to the node.

If the P2P node N42 is online, node N8 will successfully send the search message to N42, and it takes time T_{relay} . N42 will continue to forward the search message.

If the P2P node N42 is not online, and after the time $T_{timeout}$, the P2P node N8 that initiates the search does not receive the response, a new neighbor node, N38, will be re-selected to initiate a search.

The above process should be repeated until the search message is forwarded to the destination node, or the timer expires, then the search process is completed.

Construction of Markov Chain

In the above search process, the search is initiated by the P2P node P_0 , and the P2P nodes that the search messages pass through are in turn $P_1, P_2, P_3, \dots, P_{hop}$, as shown in Fig. 2(1). Among them, hop is the average number of hops for searching messages in the P2P network, $hop = O(\log N)$.

We use P_{avail} to represent the probability of a P2P node online. The P2P node P_0 initiates the search, and the probability of the node P_1 online is P_{avail} , that is, the probability that the search message is successfully forwarded to the node P_1 is P_{avail} , the required time is T_{relay} . The probability that the search message is unsuccessfully sent to the node P_1 is $1 - P_{avail}$, and the time needed to detect the node P_1 failure is $T_{timeout}$. When search message forwarding fails, the node P_0 needs to reassign the next hop node P_1' according to the routing table, constructs a new search path, and continues to forward the search message through the above process, as shown in Fig. 2(2).

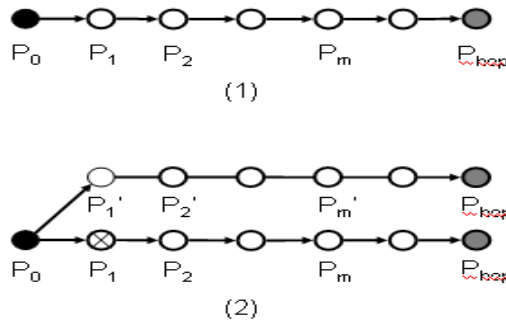


Figure 2. The path to search

Each node in the search path, after receiving the search message, will forward the message to the next hop node in the same way as the node P_0 until the search message is forwarded to the destination node or search timeout. Through the above process, the Markov chain[6] is constructed, and the following state transfer matrix is obtained.

$$P = \begin{pmatrix} 1 - P_{avail} & P_{avail} & 0 & 0 & 0 \\ 0 & 1 - P_{avail} & P_{avail} & 0 & 0 \\ 0 & 0 & 1 - P_{avail} & P_{avail} & 0 \\ 0 & 0 & 0 & 1 - P_{avail} & P_{avail} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (1)$$

In Eq. 1, the above matrix has a hop row and a hop column.

In the process of completing a search, the search message needs to experience m failure nodes. A_m represents the number of distribution of m failure nodes on the search path. T_m indicates the time required to complete this search when the search path has m failure nodes. The search message starts from the node P_0 , and after l state transitions ($l = hop + m$), the probability $P\{X_l = hop \mid X_0 = 0\}$ of its arrival at the node P_{hop} is expressed as:

$$P\{X_{hop+m} = hop \mid X_0 = 0\} = A_m (1 - P_{avail})^m P_{avail}^{hop} \quad (2)$$

The time T_m required to forward a search message is:

$$T_m = hop * T_{relay} + T_{timeout} * m \quad (3)$$

The general expression of A_m is:

$$A_m = \sum_{i=1}^{hop-m+1} \overbrace{\dots \sum_{r=1}^{m-1} \sum_{n=1}^r} \dots \sum_{r=1}^m \sum_{n=1}^r n \quad (4)$$

A_m consists of m accumulation.

The Probability P_{avail} of the Node Online

It is assumed that the node session time in the P2P network obeys the x distribution, the node P_i is any node in the P2P network, and the session time is θ_i , therefore, θ_i also obeys the x distribution[7].

In P2P networks, the probability of any node leaving the P2P network is exponentially distributed. For the pointer pointing to the node P_i , the node is online in time T_{stab} , and the probability of the node online in $T_{stab} + t$ can be expressed as:

$$P_{avail} = P\{X > t + T_{stab} | X > T_{stab}\} = \begin{cases} e^{-t/\theta_i}, & 0 < t < T_{stab} \\ 1, & t = 0 \end{cases} \quad (5)$$

Among them, T_{stab} is the synchronization cycle of the node, and θ_i obeys the x distribution(x distribution is any one of the exponential distribution, the Pareto distribution, the uniform distribution, etc.)

It is known from the Eq.5 that the probability of the node P_i online is related to the node P_i session time and the synchronization cycle of the routing table. The P2P network is made up of a large number of P2P nodes. Therefore, the mean session time and its distribution rule of the P2P network depict the session time of the P2P node macroscopically, and make it determine the probability of the nodes online in the network[8].

Search Delay Model

According to the 5.3 Markov matrix, the mathematical expectation T_{lookup} of the search delay in P2P network can be expressed as:

$$T_{lookup} = \sum_{n=0}^m P\{X_{hop+n} = hop | X_0 = 0\} * T_n \quad (6)$$

When they are put into the Eq. 2 and Eq. 3, we can get the following:

$$T_{lookup} = \sum_{n=0}^m A_n * (1 - P_{avail})^n * P_{avail}^{hop} * (hop * T_{relay} + T_{timeout} * n) \quad (7)$$

We sort out the above formula:

$$T_{lookup} = \sum_{n=0}^m A_n * (1 - P_{avail})^n * P_{avail}^{hop} * hop * T_{relay} + \sum_{n=0}^m A_n * (1 - P_{avail})^n * P_{avail}^{hop} * n * T_{timeout} \quad (8)$$

Eq. 5 and Eq.8 together form a search delay model[9].

Verification of Search Delay Model

We plug the parameter values in the p2psim simulation environment into the search delay model consisting of the Eq. 5 and Eq.8, and the results are shown in Fig.3.

As we can see from the figure, the search delay model fits well with the p2psim simulation results. The search delay model shows the relationship between the P2P Churn and the search delay. When the network tends to be stable, the search delay is shorter, for example, when the mean session time of the network node is 4800 seconds, the search delay is 350~370 milliseconds. When the dynamic performance of the P2P network is improved, the search delay is significantly increased. In the simulation, when the mean session time of the node is 1200 seconds, the search delay is 720~800 milliseconds.

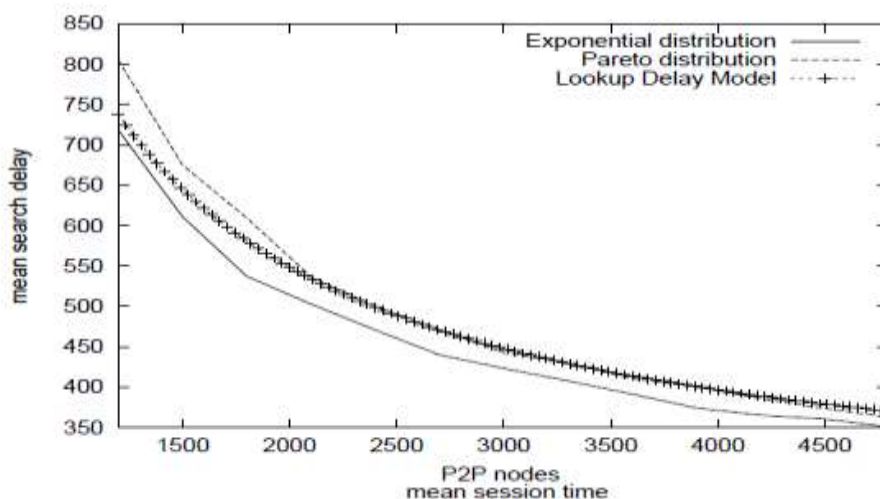


Figure 3. Search delay model

It can be verified by experiments that shortening the synchronization cycle of node routing table can improve the search performance of the network under the condition that the network bandwidth is not constrained[10].

Conclusion

In this paper, the “search delay model” is obtained by analyzing the search process of the Chord protocol and using the Markov random process. Then the correctness and effectiveness of the search delay model are verified by simulation. The search delay model accurately shows the inherent law between P2P Churn and network performance, and has a strong application value. In the future, we will further study the impact of various optimization strategies on the query delay and reflect these effects in the model, for example, we should consider the effect of optimization strategies such as parallel query and neighbor node selection on the search delay query, so that the actual situation of the network can be better reflected.

Acknowledgements

This work is partially supported by the Educational Commission of Hubei Province of China(B2016257),the Public Security Department Of Hubei Province(hbst2014yyxc04),the Hubei University Of Police (2015ZD006).

References

- [1] J.R.Wang: Research on Query Search Mechanism and Trust model in P2P Network[D], Tianjin University, 2008.
- [2] Y.X.Zhang, D.Yang and H.K.Zhang: Research on Churn Problem in P2P Networks [J], Journal of Software, Vol.20,No.5,5,2009. 1362–1376.
- [3] Zhongmei Yao, Dmitri Loguinov. Analysis of Link Lifetimes and Neighbor Selection in Switching DHTs. Parallel and distributed systems. VOL.22.NO.11, NOV. 2011. 1834-1841.
- [4] Ulusoy, O. Research issues in Peer-to-Peer data management. Computer and information sciences. 2007.iscis 2007. 1-8.
- [5] Anceaume E, Ludinard R, Sericola B, Analytic Study of the Impact of Churn in Cluster-Based Structured P2P Overlays. Communications (ICC), 2010 IEEE International Conference. 2010.1-5.
- [6] Riikka Susitaival,Samuli Aalto,Jorma Virtamo. Analyzing the Dynamics and Resource Usage of P2P File Sharing by a Spatio-temporal Model. Computational Science-ICCS 2006. Volume 3994.2006.420-427.
- [7] Wu D, Tian Y, Ng KW. Analytical study on improving DHT lookup performance under churn. In: Proc. of the 6th IEEE Int’l Conf. on Peer-to-Peer Computing (P2P 2006). Piscataway: IEEE Press, 2006. 249–258.
- [8] Martinez-Yelmo I, Bikfalvi A,Cuevas R,Guerrero C & Jaime Garcia(2009)H-P2PSIP: Interconnection of P2P SIP domains for global multimedia services based on a hierarchical DHT overlay network. Computer Networks 53(2009):556-568.
- [9] Zoels S, Despotovic Z & Kellerer W(2008) On hierarchical DHT systems-An analytical approach for optimal designs. Computer Communications 31:576-590.
- [10] Xie C,Chen G,Vandenberg A & Pan Y (2008) Analysis of hybrid P2P overlay network topology. Computer Communications 31(2):190-200.