ISSN : 1226-7244 (Print)
ISSN : 2288-243X (Online)
논문번호 18-03-07

j.inst.Korean.electr.electron.eng.Vol.22,No.3,570~577,September 2018
http://dx.doi.org/10.7471/ikeee.2018.22.3.570
49

# P2P Streaming Media Node Selection Strategy Based on Greedy Algorithm

Yiqi Gui*★, Shuangshuang Ju*, Hwangkyu Choi**★

## Abstract

With the increasing number of network nodes, traditional client/server node selection mechanisms are under tremendous pressure. In order to select efficient cooperative nodes in a highly dynamic P2P network topology, this article uses greedy algorithm to translate the overall optimization into multiple local optimal problems, and to quickly select service nodes. Therefore, the service node with the largest comprehensive capacity is selected to reduce the transmission delay and improve the throughput of the service node. The final simulation results show that the node selection strategy based on greedy algorithm can effectively improve the overall performance of P2P streaming media system.

*Key words: P2P, Streaming, Media node selection, Greedy algorithm, Throughput, Transmission delay*

## Ⅰ. Introduction

With the expansion of network coverage and the popularity of broadband services, online streaming media-based video, audio, remote classroom, video conferencing and other applications, such as pps, pptv, pplive, have achieved rapid development. Streaming media applications have become important services in the Internet [1]. Therefore, the traffic generated by the user puts tremendous pressure on the network. The traditional client/server(C/S) mode will overload the server and cause QoS failure [2].

The peer-to-peer(P2P) streaming media on-demand technology has the advantages of load balancing and scalability. It guarantees high transmission performance and service quality even when the scale of users continues to expand [3]. The P2P network is shown in Fig. 1.

* Dept. of Computer Science, College of Information Engineering, Yangzhou University
** Dept. of Computer Science & Engineering, College of Information Technology, Kangwon National University
★ Corresponding authors
* E-mail : gigi_1368@163.com, Tel : +86-182-4870-9466
** E-mail : hkchoi@kangwon.ac.kr, Tel : +82-033-250-6382
※ Acknowledgment
The research was supported in part by Natural Science Foundation of the Jiangsu Province under grant (No. 0574870053033), Natural Science Foundation of Yangzhou University under grant (No. 0374780015632), and 2017 Korea-China Young Scientist Exchange Program from National Research Foundation of Korea.
This study was supported by 2016 Research Grant from Kangwon National University (No. 520160136).
Manuscript received Aug, 9. 2018; revised Aug, 27. 2018; accepted Aug, 29. 2018

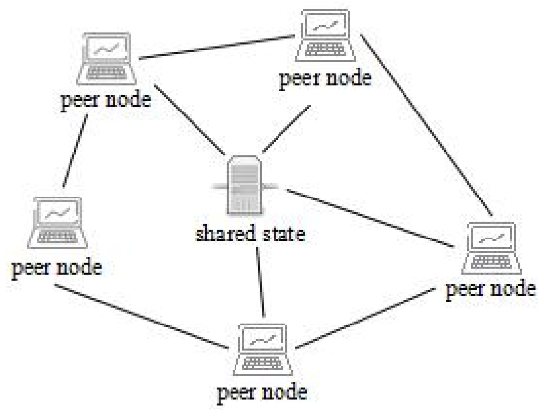



Fig. 1. Peer-to-peer network structure.

As shown in Fig. 1, in P2P networks, multiple computers are connected to each other and are in a peer-to-peer position. Each computer has no master-slave difference. A computer can serve as both a server and a client. The network does not rely on centralized servers. Every computer in the network can obtain tasks and content. Usually these resources and services include: information sharing and exchange, computing resources, storage sharing, network sharing, printer sharing, and so on.

In a P2P system, a node can become a video stream provider and can also retrieve video stream information from its neighbor nodes [4]. If a node needs to obtain a specific "resource", it must first be able to locate which nodes have the required "resources". This part of the function is called the content routing function in the P2P system. When the node determines which nodes have the same "resources", it needs to use some mechanism to acquire resources from multiple nodes in parallel. At the same time, the resource fragments obtained from different nodes should be different, otherwise they lose the significance of parallel transmission.

This paper proposes a node selection strategy based on greedy algorithm. The greedy algorithm is very effective for solving many problems, because the local optimal solution is easy to find, so it is easy to understand and efficient, and many problems can get the overall optimal solution [5]. First we calculate the "attribute distance" of each service node as

$$\sqrt{(Attribute1 - Threshold1)^2 + (Attribute2 - Threshold2)^2 + (Attribute3 - Threshold3)^2}\ .$$

Then we use the greedy algorithm for node selection based on "attribute distance" to improve the overall network throughput and reduce transmission delay.

## II. Related Work

In recent years, a large number of domestic and foreign scholars have conducted extensive research in order to improve the performance of P2P streaming media. The selection of neighbor is one of the challenging areas in P2P computing.

Rao et al. [6] proposed the Multi-Layer Perceptron (MLP) technique for neighbor selection in P2P computing to reduce the communication overhead. Feng et al. [7] proposed an adaptive neighbor node selection algorithm that dynamically describes the node's service capabilities and significantly improves the service performance of the system. However, in terms of service capabilities, only the upstream bandwidth of the node is used as an indicator, and no other attributes of the node are considered. Therefore, the service performance needs to be improved.

Li et al. [8] proposed a node selection strategy based on Intuitionistic Fuzzy Sets(IFS) to overcome the randomness and blindness of traditional strategies, and adopted a method of scoring the reliability of nodes, making full use of powerful nodes and having a good adaptability. The algorithm is suitable for large-scale, low node performance environments, but the algorithm has no obvious advantages when the node performance is good.

Lu et al. [9] has systematically analyzed the existing node-select algorithms. Furthermore, they proposed and implemented a new node selection algorithm based on Dijkstra algorithm, named

NSDA algorithm.

Zhang et al. [10] proposed a method for selecting neighbor nodes by establishing a model of delay and bandwidth. However, the delay between all the nodes in the system needs to be measured. The computational overhead is large and the scalability is poor.

Guan et al. [11] has analyzed the communication delay between the nodes in P2P system in detail, divides the node into different groups according to the different communication delay between the nodes, thus reduces the delay of the search node. This article is optimized on the search node, but it is not applicable to real-time streaming media systems.

Wang et al. [12] studied the location-aware relay discovery and selection problem for large-scale P2P streaming networks. In these large-scale and dynamic overlays, it incurs significant communication and computation cost to discover a sufficiently large relay candidate set and further to select one relay with good performance.

Chen et al. [13] proposed the destination node routing algorithm based on super nodes. The algorithm than the average P2P routing algorithm is faster, which usually only 2~3 hop routing queries could locate the resource nodes.

Budhkar et al. [14] proposed a two-node selection strategy of Tracker layer and peer layer. The node's selection mainly relied on propagation delay, upload capacity and buffer level, and maintained the neighbor node relationship through the server. The maintenance overhead of the algorithm was relatively large.

Liu et al. [15] proposed a Reputation-Aware Super Node Selection Algorithm. This algorithm selects nodes with good machine performance and long online time as super nodes from nodes with high reputation. Although the network reliability was improved by detecting malicious nodes, there is no regular management of the detected nodes, and it is difficult to make predictions about mobility [16].

This paper comprehensively considers the node's uplink and downlink bandwidth, node online duration, node distance and node service capability, and then proposes a P2P streaming media node selection strategy based on greedy algorithm. Therefore, an efficient cooperative node is selected under a highly dynamic P2P network topology.

In this paper, the greedy algorithm is used to transform the overall optimality into multiple local optimal problems, and the service nodes are quickly selected to select the service node with the most comprehensive capabilities. Thereby reducing the transmission delay and improving the throughput of the service node, effectively improving the overall performance of the system.

## III. System Model

This section explains the system model and the overall system flow. First, we define the node model and calculate the corresponding parameter matrix, and then select the node through the greedy algorithm to calculate the overall throughput and transmission delay. This paper only considers the node selection strategy under a single time slot.

First, the request node set is defined as $User=\{user_1,user_2......user_n\}$ . Each node contains three parameters as shown in Table 1.

Table 1. Request node parameters.

| No. | Parameter |
|---|---|
| 1 | Position |
| 2 | Current segment number |
| 3 | Bandwidth of the requesting node |

Then define $m$ service nodes. The service node set is defined as $Peer=\{peer_1, peer_2,.......peer_m\}$ . Each service node contains five attributes as shown in Table 2.

Table 2. Service node parameters

| No. | Parameter |
|---|---|
| 1 | Position |
| 2 | The serial number of the stored segment |
| 3 | Online duration |
| 4 | Current service volume |
| 5 | Bandwidth of the service node |

The request node and service node matrix are as follows: calculate the( $n \times m$ )distance matrix, which is the distance from each request node to each service node. Where $localpeer$ is the location of the request node, and $peer$ is the location of the service node. The distance from the $n$th request node to the $m$th service node is as follows:

$$dist_{n,m} = |localpeer_{(n,1)} - peer_{(m,1)}| + |localpeer_{(n,2)} - peer_{(m,2)}| \quad (1)$$

Calculate the channel capacity matrix, which is still a matrix of ( $n \times m$ ). The channel capacity here refers to the maximum information rate that the channel can transmit. In other words, it refers to the maximum transmission capacity that the channel can achieve. It is determined by the characteristics of the channel [17]. The Shannon formula directly reflects the relationship between channel capacity, signal power, and noise power [18] [19]. According to the Shannon formula, the channel capacity between the $n$th request node and the $m$th service node is calculated as follows:

$$c_{n,m} = B_m * log_2(1 + \sigma_n * p_n) \quad (2)$$

Where $p_n$ is the average of the signal transmission power of the requesting node to other service nodes. Since the service node has a large relationship between bandwidth and throughput, the bandwidth of the service node is selected as $B_m$ . $\sigma_n$ is stand for the $SNR$ of the request node and the $n$th service node. The signal-to-noise ratio is defined as the ratio of the average power of the transmitted signal to the average power of the additive noise. It is defined as follows:

$$\sigma_n = \frac{d_{n,m} * h_0}{B_m * N_0} \quad (3)$$

Where $d_{n,m}$ represents the path loss of the nth request node connecting the $m$th service node and $h_0$ denotes the channel incremental. $h_0$ are different values for different channel conditions and remain unchanged in the same time slot. $N_0$ is the power spectral density and the value is $1*10^{-7}$.

When obtaining an available service node, because the service node has certain restrictions, in order to reduce the transmission time and improve the overall performance, a threshold is set for the channel capacity between the service node and the request node, and when the service node is less than the threshold, the service node is discarded. A threshold is set for the distance between the service node and the request node, and when the threshold is greater than the service node, the service node is discarded.

A threshold is set for the online duration of the service node, and when it is less than the threshold, the service node is discarded. A threshold is set for the number of request nodes that the service node can serve, and when it is greater than the threshold, the service node is discarded. According to the above description, the available matrix is obtained:

$$buffpeer_{n \times m} = \{0,1\} \quad (4)$$

The $buffpeer$ is a matrix that determines if a service node is available. If the service node is available, $buffpeer_n = 1$, and if the service node is not available, $buffpeer_n = 0$ .

When the service node is available, the service node storing the next video segment is found (if the current node of the request node is segment 4, the service node storing the segment 5 is found). If the conditions are met, the service node's comprehensive service capability value is calculated and the comprehensive capability matrix is obtained. It defines as follows:

$$Ability_{n,m} = \sqrt{((dist_{n,m} - 100)/10)^2 + (c_{n,m} - 3)^2 + (time_m - 3)^2} \quad (5)$$

Otherwise, $Ability_{n,m} = 0$.

Where $dist_{n,m}$ represents the distance from the $n$th request node to the $m$th service node; $c_{n,m}$ represents the channel capacity of the $n$th request node to the $m$th service node; $time_m$ represents the online duration of the $m$th service node. At the same time, update the number of request nodes currently served by the service node, determine whether the service limit is reached, and update the available table $buffpeer$ .

Finally, a greedy algorithm is used. According to the comprehensive capability matrix $Ability$, each requesting node selects the service node with the largest comprehensive capacity, obtains the node label, and then calculates the throughput and transmission delay of the node. Specifically expressed as follows:

$$C = \sum_{i=1}^{n} c_n \quad (6)$$

$$TIME = \sum_{i=1}^{n} data_n / c_n \quad (7)$$

Where $C$ represents the total throughput in single time slot in P2P system; $TIME$ represents the total transmission delay in single time slot in P2P system. Based on the above formula, P2P streaming media node selection algorithm GA-PSA based on greedy algorithm is obtained, as shown in the Table 3.

Table 3. Node selection algorithm based on greedy algorithm.

| P2P streaming media node selection algorithm based on greedy algorithm (GA-PSA) |
|---|
| Step1 : Initialize $h, N0, pm$ and other parameters. |
| Step2 : Generate request nodes and service nodes, and generate corresponding parameters for request nodes and service nodes. |
| Step3 : Calculate node distance $dist_{n\times m}$ , channel capacity $c_{n\times m}$ . |
| Step4 : Calculate the available matrix $buffpeer_{n\times m}$ . |
| Step5 : Based on the available matrices, calculate the comprehensive capability matrix $Ability_{n\times m}$ and find the service nodes that meet the requirements. |
| Step6 : Using a greedy algorithm, each requesting node selects the service node with the highest comprehensive ability among the service nodes that meet the requirements. |
| Step7 : Calculate the total transmission delay and throughput. |
| Step8 : The algorithm is completed. |

## Ⅳ. Results and Comparisons

In order to verify the performance of this algorithm, we use MATLAB for simulation. The transmission delay and throughput calculated by this algorithm are compared with the node selection strategy based on intuitionistic fuzzy set, and the performance difference between the two is observed.

The specific simulation conditions are as follows: The bandwidth between the request node and the service node is the average value of the bandwidth of the request node and the service node. The power spectral density between the request node and the service node is $N = 10^7$ and the average power is $pm = 10dbm$. The number of request nodes is $M = 50$ and the number of service nodes is $N = 500$. We understand the sum of the execution times of the code as the time required for the algorithm to calculate the result. Therefore, the algorithm complexity is $O(n^2)$. The experimental process diagram shown in Fig. 2.

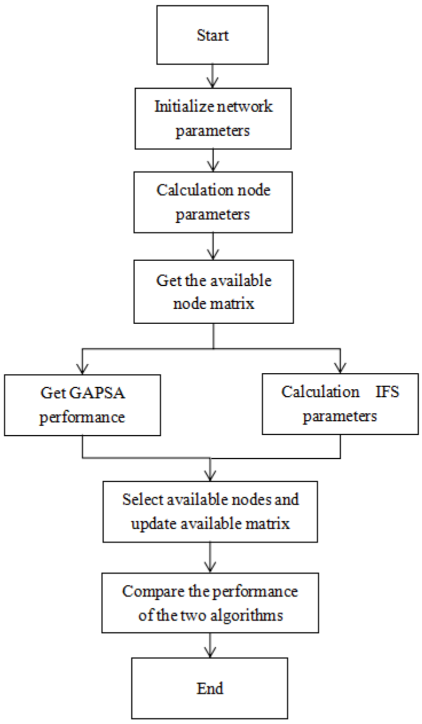



Fig. 2. Experimental process diagram.

The specific simulation results are shown in the following Fig. 3 and Fig. 4.

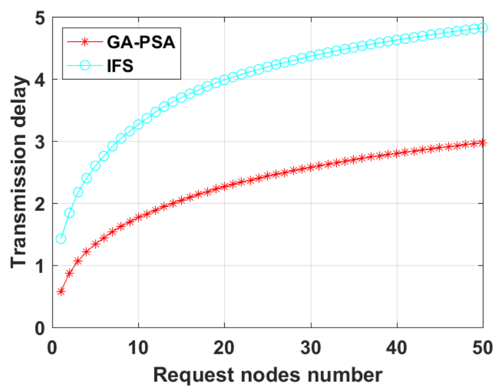



Fig. 3. Relationship between the total transmission delay and the number of request nodes.

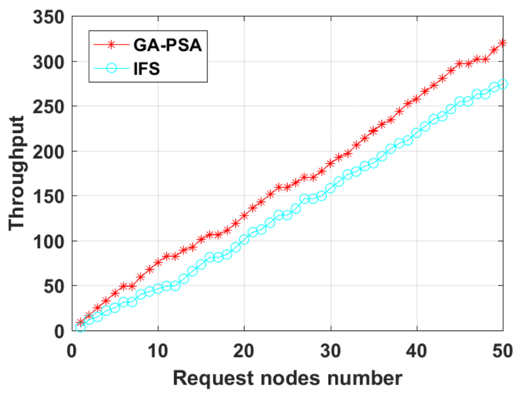



Fig. 4. Relationship between the total throughput and the number of request nodes.

The results show that the P2P streaming media node selection algorithm based on greedy algorithm has higher throughput than the node selection algorithm based on intuitionistic fuzzy set, and has lower transmission delay. This is because the node selection algorithm based on greedy algorithm comprehensively considers the attributes of each dimension of the service node, obtains the comprehensive "attribute distance" of the service node, and selects the node farthest from the threshold "distance". Therefore, it can better reflect the superiority of the selected node.

The node selection strategy based on intuitionistic fuzzy sets obtains the solution set through the scoring function, and then weights the attribute according to its importance to obtain the decision matrix. Because the weighting has a certain deviation, it cannot fully reflect the overall performance of the node. Therefore, the node selection strategy based on greedy algorithm has better performance. In our future work, pure fuzzy logic strategy can be studied due to be a quite effective strategy which is proved by Valdés et al. [20].

## Ⅳ. Conclusion

This paper uses greedy algorithm (GA, Greedy Algorithm) to translate the overall optimization into multiple local optimal problems, and to quickly select service nodes. Therefore, the service node with the largest comprehensive capacity is selected to reduce the transmission delay and improve the throughput of the service node.

In P2P streaming media system, the node selection strategy mostly considers the factors such as node bandwidth, online duration and service quality, but how to integrate these factors to get the optimal node selection solution is still an important issue at present. The next step needs to further explore other node selection indicators in order to make the overall system perform better.

## References

[1] Y. Wei, S. Han, “Streaming Media Super Nodes Selection Algorithm Based on Bandwidth Prediction,” *Computer Applications and Software,* vol.33, no.8, pp.179-182, 2016.

[2] H. Sun, J. Y. Zhang, “P2P Streaming Media Distributing Algorithm Based on Push-Pull Mechanism,” *Software,* vol.34, no.5, pp.43-47, 2013.

[3] R. Fan, G. Huang, “Research on P2P Neighbor Node Selection Algorithm Based on Location-Aware,” *Journal of Guilin University of Electronic Technology,* vol.37, no.4, pp.307-310, 2017.

[4] J. N. Wei, L. Zhuang, “Peer Selection Strategies without Overlay Topology Information in P2P Streaming Media,” *Application Research of Computers,* vol.29, no.4, pp.1536-1539, 2012.

[5] F. Liu, Z. Cai, C. Fu, “Analysis and Study on Common Classical Algorithms,” *Computer Era,* pp. 58-63, 2017.

[6] D. C. Rao, T. Nayak, M. R. Kabat, “Neighbor Selection in Peer-to-Peer Computing using Multi-Layer Perceptron,” *Proc. of Next Generation Computing Technologies (NGCT), IEEE,* pp. 245-250, 2016. DOI:10.1109/NGCT.2015.7375120

[7] Z. Feng, H. Ni, J. Wang, J. You, “Adaptive Neighbor Selection Method for the P2P Media Streaming System,” *J. of Xidian University,* vol.39, no.3, pp.136-143, 2012.

[8] Y. Li, L. N. Wang, “Node Selection Strategy Based on Intuitionistic Fuzzy Sets in P2P Streaming Media System,” *Computer Science,* vol.40, no.6A, 2013.

[9] T. Lu, J. Zhang, L. Zhao, Y. Li, X. Zhang, “NSDA: A Novel Node Selecting Optimal Algorithm Based on Dijkstra,” *Int. J. of Future Generation Communication and Networking,* pp.113-128, 2016. DOI:10.14257/ijfgcn.2016.9.6.11

[10] K. Zhang, K, Li, “LBA: Location and Bandwidth Awareness Overlay for P2P Live Video Streaming,” *Proc. of Networking and Digital Society, IEEE,* pp.327-330, 2010. DOI:10.1109/ICNDS.2010.5479400

[11] W. Lu, L. T. Guan, “Search Model of P2P Bases On Tactics of Ranking in Group,” *J. of Xidian University,* vol.34, no.3, 2007.

[12] C. Zhang, A. Y. Wangla, X. Hei, “Relay Discovery and Selection for Large-Scale P2P Streaming,” *PLOS ONE,* vol.12, no.4, 2017. DOI:10.1371/journal.pone.0175360

[13] Y. Chen, L. Zhang, “Target Node Routing Algorithm in P2P Based on the Quality of Service,” *Intelligent Computer Applications,* vol.5, no.5, 2015.

[14] S. Budhkar, V. Tamarapalli, “Two-Tier Peer Selection Strategy to Minimize Delay in P2P Live Streaming Systems,” *Proc. of 22th National Conference on Communication,* 2016. DOI:10.1109/NCC.2016.7561203

[15] Y. M. Liu, S. B. Yang, W. M. Chen, “The Research of the Reputation-Aware Super Node Selection Algorithm in P2P System,” *J. of the Graduate School of the Chinese Academy of Sciences,* vol.25, no.2, pp.197-203, 2008.

[16] H. J. Cha, H. K. Yang, Y. H. Kim, “A Node Management Scheme for Stable P2P Service in Mobile ad-hoc Networks,” *Peer-to-Peer Networking and Applications,* vol.9, no.3, pp.558-565, 2016. DOI:10.1007/s12083-015-0392-5

[17] X. Yu. “The Calculation of Channel Capacity in Discrete Channels,” *Project Management and Technology,* no.14, 2018.

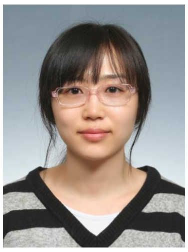
[18] C. J. Liu, Z. Wei, “Research on Application of Shannon Formula in Spread Spectrum Communication,” *Sichuan Journal of Ordnance,* vol.34, no.4, pp.80-83, 2013.

[19] J. H. Yan, J. Ren, “Simulation and Analysis of Spread-Spectrum Communication System Based on System View,” *Electronic Design Engineering,* vol.24, no.11, pp.99-101, 2016.

[20] L. Valdés, S. Montesinos, A. Ariza, M. S. Allende, G. Joya, “Peer Selection in P2P Wireless Mesh Networks: Comparison of Different Strategies,” *Soft Computing,* vol.19, no.9, pp.2447-2455, 2015. DOI:10.1007/s00500-014-1572-6

## BIOGRAPHY

**Yiqi Gui** (Member)



2004 : BS degree in Computer Science, Beihua University in China.
2007 : MS degree in Computer Engineering, Kangwon National University.
2012 : PhD degree in Computer Engineering, Kangwon National University.
2012~ : Research on datamining for multimedia system and network security, Yangzhou University in China.

**Shuangshuang Ju** (Member)

2013 : BS degree in Computer Science, Yangzhou Institute of Technology in China.
2017~ : Currently for MS degree in Yangzhou University in China.
Research on datamining for multimedia system.

**Hwangkyu Choi** (Member)

1984 : BS degree in Electronics Engineering, Kyungbook National University.
1986 : MS degree in Electric and Electrical Engineering, KAIST.
1989 : PhD degree in Electric and Electrical Engineering, KAIST.
1990~ : Professor in Department of Computer Science and Engineering, Kangwon National University.