Performance Analysis of Restricted Path Flooding Scheme in Distributed P2P Overlay Networks

Hyuncheol Kim[1], Younghwa Kim[2], Kwangjoon Kim[2], Doo-Won Seo[3]

[1] Dept. of Computer Science, Namseoul Univ., Cheonan, Korea
[2] ASON Tech Team, Electronics and Telecommunications Research Institute, Daejeon, Korea
[3] Dept. of Computer Network, Daeduk College, Deajon, Korea

Summary
The Internet can be seen as a cooperative network composed of millions of hosts spread around the world and it is also a distributed shared resource. In the P2P technology, as opposed to the existing client/server concept, devices are actively connected with one another in order to share the resources and every participant is both a server and a client at the same time. In P2P overlay network, peers are able to directly share and exchange information without the help of a server. This results in a prompt and secure sharing of network resources and data handling. However, flooding algorithm that is used in distributed P2P network generated query message excessively. Our objective in this paper is to proposes a presents a restricted path flooding algorithm that can decrease query message's occurrence to solve P2P network's problems. It includes concepts as well as systematic procedures of the proposed scheme for fast path flooding in distributed P2P overlay networks.
Key words:
P2P, Peer to Peer, Flooding

1. Introduction

The Internet can be seen as a cooperative network composed of millions of hosts spread around the world and it is also a distributed shared resource. The current network status requires, more so than ever, a scheme in which bandwidths are more efficiently managed and network capacities are fully utilized as well as application programs that can send packets more safely to farther distances [1].

Traditional client/server models with a vertical hierarchical structure generate problems such as vulnerability against malicious attacks, inability to be expanded, and bottleneck issues in local networks. Therefore, various researches were conducted on the P2P (Peer-to-peer network) model, which is a horizontal network model in which all hosts that form a network can act as a client and the server at the same time.

Under such change in the computing paradigm, future Internet-based applications include the following three requirements.

First is to provide expandability. Flexible measures are required for requests from users regarding a larger bandwidth, saving and processing functions. Second is the security and credibility issue. Counter against malicious attacking of centralized structure-based systems, anonymity, and protection from hacking are required. Third issue is flexibility and QoS (Quality of Service). Future network to come must be equipped with prompt and simple countermove ability for new services.

The existing client/server-based model falls short of satisfying such requirements. For example, it cannot react against malicious attacks or resource bottleneck effects in focused structures and there are complexity and high-cost issues when attempting to mitigate the original structure [2].

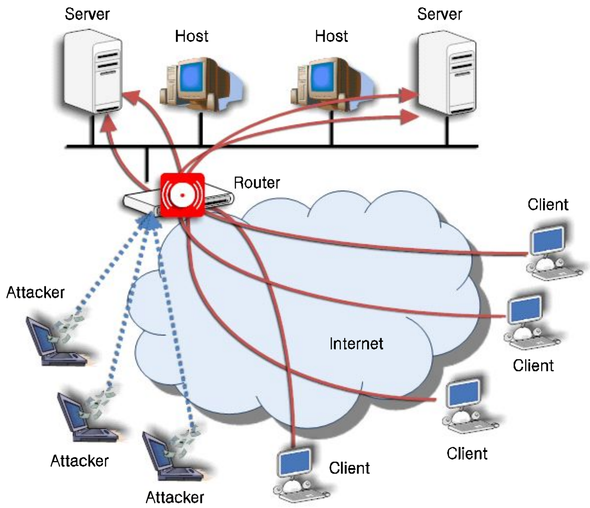

Fig. 1 Traffic Concentration Problem in Client-Server Networks

Manuscript received December 5, 2007
Manuscript revised December 20, 2007

In the P2P technology, as opposed to the existing client/server concept, devices are actively connected with one another as shown in Fig. 2 in order to share the resources and every participant is both a server and a client at the same time. As depicted in Fig. 2, when the peers that exist in a physical network are registered in the P2P service, a virtual network among the registered peers, namely a P2P overlay network, is formed.

In the P2P overlay network, peers are able to directly share and exchange information among themselves without the help of the server. Through this, a safer and a faster network resources sharing and data handling are made possible.

The P2P overlay network's strength is in that it considers its sub-networks merely as a part of the platform. This means that it is not relevant what the sub-networks are made up of as long as there are peripherals that can support the applications. Therefore, it is useful in that it allows a timely service development and distribution.

However, the P2P overlay network has a drawback in that it induces an inefficient operation of network bandwidth because it consumes such a vast amount of network bandwidth. In order to overcome these difficulties, this paper proposes a new flooding algorithm that enhances the overall performance of the network by reducing the excessive query message traffic that occurs in the distributed P2P network, which is a type of P2P overlay network.

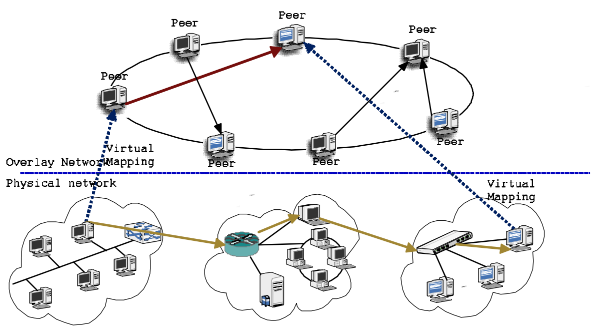

Fig. 2 P2P Overlay Network Architecture

The remainder of this paper is organized as follows. Section 2 examines some previous studies that try to enhance the flooding technique in the existing distributed P2P network and summarizes the problems that it entails. Section 3 describes the characteristics of each P2P overlay network type and it primarily focuses on the flooding method in the distributed P2P network along with a possible solution. Section 4 compares the existing flooding method with the proposed flooding method through simulation in order to analyze and contrast their quantified performances. Section 5 concludes this paper by summarizing some key points made throughout and assessing the representation of analyzed results.

2. Related Works

There are two general strategies used to search for an object: blind search and in-formed search. The semantics of the used information range from simple forwarding hints to exact object locations. The placement of this information can be classified into centralized approaches (e.g., [3]), distributed approaches (e.g., [4], [5], [6]), and hybrid architectures (e.g., [7]).

The semantics of the stored indices in informed approaches can be used for another categorization. Indices might relate to exact object locations (e.g., [8]), probability of discovery through a link (e.g., [6]), number of objects through a link (e.g., [4]), or other metrics (e.g., [9]). Finally, search schemes can be categorized according to the query forwarding method into flooding-based and non flooding-based (e.g., [10]).

As known, the original Gnutella algorithm is flooding scheme, which generates too much traffic in the network. Modified-BFS [5]: Peers randomly choose only a ratio of their neighbors to forward the query. This reduces the average message production, but still contacts a large number of peers.

Iterative Deepening: Two similar approaches that use consecutive BFS searches at increasing depths are described in [10], [11].

These algorithms achieve best results when the search termination condition relates to a user-defined number of hits and it is possible that searching at small depths will satisfy the query.

In a different case, they produce even bigger loads than the standard flooding scheme. In Random Walks [10], the requesting node sends out k query messages to an equal number of randomly chosen neighbors. Each of these messages follows its own path, having intermediate nodes forward it to a randomly chosen neighbor at each step.

The algorithm achieves a message reduction and some load balancing compared to the flooding scheme. However, success rates vary greatly depending on network topology.

In Gnutella2 [12], it is a self-organized system, which consists of super-peer (or hub) and client-peer (or leaf). The leaf only connects to some of hub, while the hub connects too many other hubs and many leaves as well. Hub provides routing information for leaves. Neighboring hubs exchange local repository tables to filter out unnecessary traffic.

In GUESS [7], a search is conducted by iteratively contacting different super-peers and having them ask all their leaves, until a number of objects are found. The number of leaves per super-peer must be kept high, even after node arrivals/departures. This is the most important condition in order to reduce message forwarding and increase the number of discovered objects.

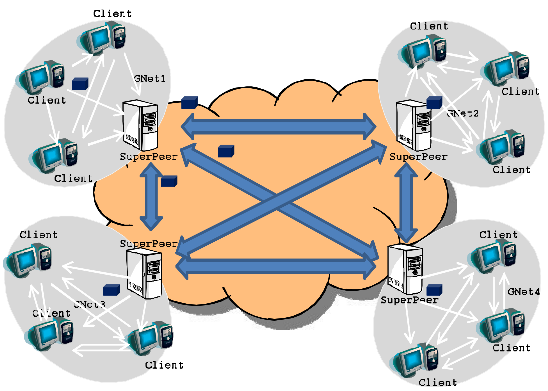

Fig. 3 Problem of Distributed P2P Network (Flooding)

3. Restricted Path Flooding

With the existing method, information data or ping, pong messages in the network generate too much unnecessary data. Also, the network may become overloaded due to the repeating duplicate data because a generated query message uses the flooding method. This will result in a critical confusion and it may interfere with the sending of the actual data.

As of the many distributed P2P network problems, this paper focused on how to resolve the problem that occurs from using the flooding method when generating and delivering query messages. Therefore, we would like to propose the Restricted Path Flooding algorithm, an upgrade from the existing method that has reduced the network traffic occurring in the flooding method.

3.1 Method

In the existing flooding algorithm, when a host wants to send a query message or relay a received query message to other peers, it chose the method that sends the message to all its adjacent peers.

Such method is useful in that it is able to promptly distribute specific information in a vast network. However, the drawback of the flooding method is that it induces large traffics within the network.

Not only this overload the entire network but it also creates bottleneck issues. So the distributed P2P network uses the TTL value to control the area to which a query message is delivered to.

Nonetheless, finding the appropriate TTL value is not a simple task. If the TTL value is too large, it will result in an unnecessary overloading and when it's too small, the appropriate target node will not be found because there are too little a number of nodes to send to.

Another problem with the flooding method is that there is a high probability that duplicates of a query message could be sent in a P2P network with high connectivity. Performing a flooding on duplicate messages will cause the number of nodes that receive the duplicate messages to continuously increase.

Also, since messages will only be sent to similar nodes, it will bring about a difficulty in detecting the target node. To prevent this, the P2P that uses the flooding-style search method is designed so that it discards any duplicate message and does not execute the flooding again.

Nevertheless, since duplicate messages will occur regardless of flooding, the feasibility of this technique is very low. In order to mitigate that problem, this research paper proposes the restricted path flooding algorithm.

Fig. 3 shows these problems of the flooding algorithm and Fig. 4 represents the proposed Restricted Path Flooding algorithm.

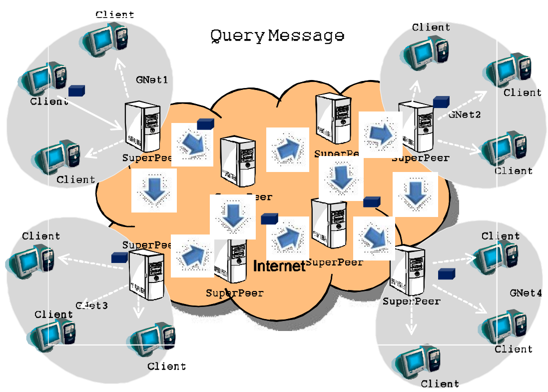

Fig. 4 Restricted Path Flooding

In the restricted path flooding technique, the P2P network is, first, designed to have an overlay structure as shown in Fig. 4 and the query-generating peers are made to send query messages to only the Superpeer.

The Superpeer does a flooding on the query message that includes the list of the adjacent peers' address. Upon receiving the query message, the nodes check if its adjacent nodes exist in the list of addresses included in the query message and it will not send the query message the nodes that exist in the address list. This method is utilized so that a duplicate query message is not delivered to nodes in the distributed P2P network.

When a query message contains a list with over a certain number of addresses, a large bandwidth could be consumed in order to deliver one query message. To prevent this from happening, when the query message tops an arbitrary threshold, the query is message is flooded only to a certain number of nodes randomly selected based on probability p out of all the adjacent nodes. Probability value p is determined by the number of nodes adjacent to the node that will be receiving the query message. This is because the likelihood of retrieving a response will increase proportionally to the number of adjacent nodes.

3.2 Algorithm

Fig. 5 is a flow diagram that represents the restricted path flooding algorithm.

Let's assume that a temporary node generated a query message in order to request for data resources.

The node that sends a query message does a flooding on the query message that it generated to its adjacent nodes where it first configures a TTL value suitable for the network environment, generates a query message that includes the query details and the address of the adjacent nodes, and does a flooding on it.

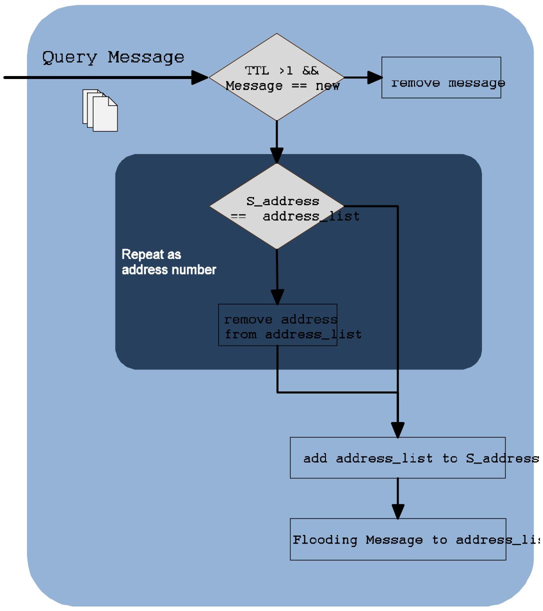

Fig. 5 Restricted Path Flooding Algorithm

When a node receives a query message, it verifies if it has the corresponding data resources. When it locates such resource, it sends back a response otherwise it checks if the TTL value in the query message is greater than 1.

If it is greater than 1, it checks if the query message had been received before and it floods the query message only when it had never been received.

Here, the S_Address stored in the message (the address of the peers to which this message was sent) and the address of the adjacent peers are compared and when there is a redundant address, that address is removed from the list of addresses that it has to send to.

After having compared all the addresses in the list and optimizing the list of addresses to which its query message will be sent, it adds its address list to the previous S_Address list and then sends the message. Therefore it minimizes the chances of sending a duplicate query message to the same node.

4. Performance Analysis

Simulations of various types of topology were conducted in order to evaluate the restricted path flooding algorithm proposed in this research.

4.1 Network Topology

In order to evaluate the performance of the proposed algorithm, we generated three different types of network topology and compared the performance for each.

The most important element when considering network topology is the number of adjacent nodes. This number refers to the average number of nodes adjacent to each node.

Each of the simulated network topologies was structured to have an average of 4 adjacent nodes. As shown in Fig. 6, the first network topology is 2-dimensional mesh topology. In this mesh-type topology, which is composed of 10*10 nodes, all nodes within the 2-dimensional mesh topology excluding the 4 nodes in the corners have 4 adjacent nodes. A typical adjacent node of this topology has about 3.96 nodes.

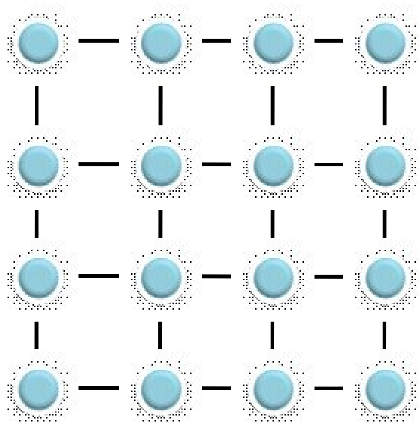

Fig. 6 2-dimensional Mesh Topology

The second network topology is a topology created by the gt-itm topology generator as shown in Fig. 7. This network topology has the form of a 2-phase hierarchical random graph type.

The first phase forms a connective graph with the 10 generated nodes. In this connective graph, nodes are connected with one another with probability p and each node is composed of 10 vertices randomly scattered. The second phase uses each of the generated nodes, composed of vertices, as the root of the new graph in order to generate a new random graph that consists of 10 nodes.

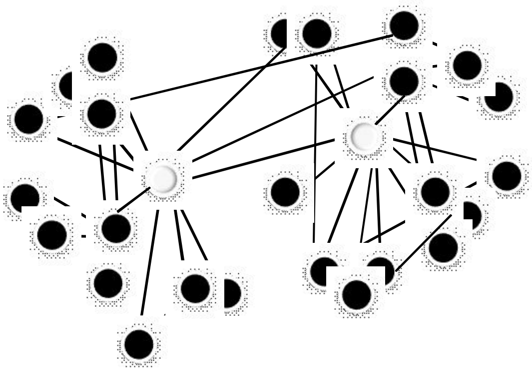

Fig. 7 2-phase Hierarchical Random Graph Topology

The latest results show that the current dispersion P2P network is in the form of power-law type-node dispersion with a reducing number of adjacent nodes due to the power function.

The power-refraction dispersion index of -2.2 is used to represent the node dispersion of the Internet and the power-refraction dispersion index of -2.3 represents the node dispersion type of an unstructured P2P network.

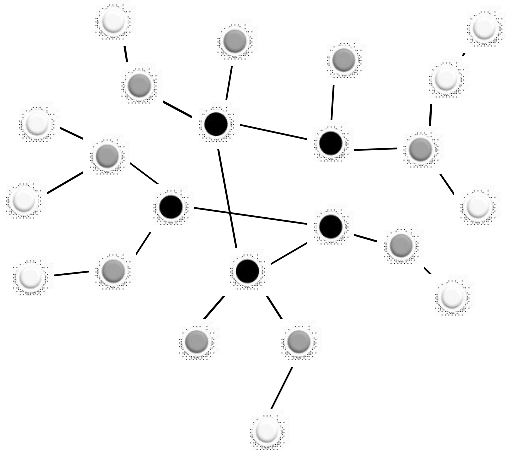

Fig. 8 Power-law Dispersion Topology

As shown in Fig. 8, a network topology generator known as Inet was used and the network topology obtained by this has an average number of adjacent nodes of approximately 4.

First, this simulation compared the amount of traffic and the number of query messages generated by the flooding algorithm and the restricted path flooding algorithm in each network topology defined by the topology up to 50 unit times and compared the amount of traffic, the number of query messages, and the time elapsed until the arrival at the arbitrary target node.

Examining the results obtained through the simulation, it can be seen that the proposed restricted path flooding algorithm generates a fewer number of query messages and less traffic than the standard flooding algorithm.

The difference in the number of query messages generated by flooding and the restricted path flooding is smaller for a lower TTL value and an increase in the TTL value could result in a maximum 58% reduction in the number of generated query messages.

For the power-law dispersion, the reason why there is no change in the number of generated query messages and the amount of traffic after a certain amount of time is because all the nodes have been reached.

Fig. 9 and Fig. 10 compare the number of query messages created, the amount of generated traffic, and the time it takes for the flooding algorithm and the restricted path flooding algorithm to arrive at the tentative target node in each network topology defined in the topology.

When a delivery of a query message to the target node is successfully made, it can be seen that the difference in the number of query messages generated among flooding and restricted path flooding is not significant.

Also, the power-law dispersion in which one node is connected to many nodes showed only 14% of decrease.

Therefore, the longer the time it takes for a query message to get to the target node, the higher efficiency of the restricted path flooding algorithm than the flooding algorithm.

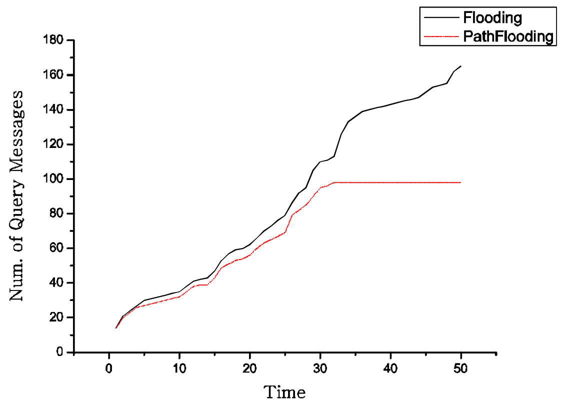

Fig. 9 Number of Query Message in Two Dimensional Mesh Topology

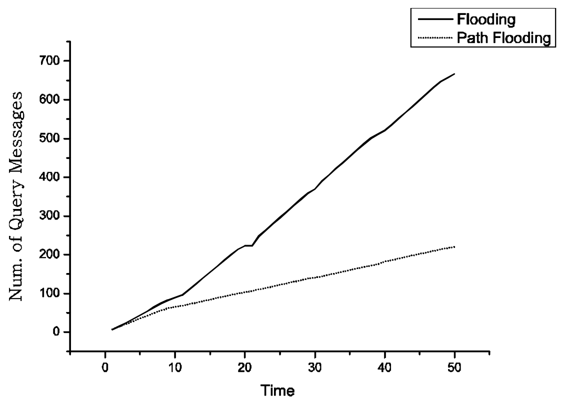

Fig. 10 Number of Query Message in Random Topology

4. Conclusion

The Internet started off as a perfectly symmetrical network, or a P2P network of users who cooperate with each other. As the Internet became vaster, accommodating an enormous number of users who swarmed to the online world, it created issues such as the bottleneck problem between the local network and the backbone caused by the network overloading between a relatively smaller number of servers and many clients.

Such phenomenon resulted in the emergence of the P2P network, thereby resolving the network difficulties.

This paper takes sharing way of incomplete data in order to improve performance in decentralized P2P network. Such attempt could bring about a P2P network with an improved performance.

References

[1] Nelson Minar, Marc Hedlund, \A Network of Peers: Peer-to-Peer Models Through the History of the Internet\, Peer to Peer Harnessing the Benefits of a Disruptive Technology, oreilly, 1999.

[2] H.J. Park, K.R. Park, "P2P Technology trend and Application to Home Network\, ETTRENDS, Num. 21, Vol. 5, Oct. 2006.

[3] Napster, http://www.napster.com.

[4] A. Crespo, H. Garcia-Molina, "Routing Indices for Peer-to-Peer Systems\, ICDCS, Jul. 2002.

[5] V. Kalogeraki, D. Gunopulos, and D. Zeinalipour-Yazti, "A Local Search Mechanism for Peer-to-Peer Networks\, CIKM, 2002.

[6] D. Tsoumakos and N. Roussopoulos, "Adaptive Probabilistic Search for Peer-to-Peer Networks\, 3rd IEEE Intl Conference on P2P Computing, 2003.

[7] S. Daswani and A. Fisk, "Gnutella UDP Ex-tension for Scalable Searches (GUESS) v0.1\, https://www.limewire.org

[8] K. Sripanidkulchai, B. Maggs, and H. Zhang, "Efficient Content Location Using Interest Based Locality in Peer-to-Peer Systems\, INFOCOM, 2003.

[9] Y. Chawathe, S. Ratnasamy, L. Breslau, N. Lan-ham, and S. Shenker, "Making Gnutella-like P2P systems Scalable\, ACM SIGCOMM'03, Aug. 2003.

[10] C. Lv, P. Cao, E. Cohen, K. Li, and S. Shenker, "Search and Replication in Un-structured Peer-to-Peer Networks\, ICS, 2002.

[11] B. Yang and H. Garcia-Molina, "Improving Search in Peer-to-Peer Networks\, ICDCS, 2002..

[12] M. Stokes, "Gnutella2 Specifications Part One\, http://www.gnutella2.com/gnutella2 search.htm..