



EPiC Series in Computing

Volume 63, 2019, Pages 121–129

Proceedings of 32nd International Conference on  
Computer Applications in Industry and Engineering



# Efficient Capacity-Constrained Multicast in RC-Based P2P Networks

Koushik Maddali<sup>1</sup>, Banafsheh Rekabdar<sup>1</sup>, Swathi Kaluvakuri<sup>1</sup> and Bidyut Gupta<sup>1</sup>

Department of Computer Science  
Southern Illinois University  
Carbondale, IL, USA

[koushik@siu.edu](mailto:koushik@siu.edu), [brekabdar@cs.siu.edu](mailto:brekabdar@cs.siu.edu), [swathi.kaluvakuri@siu.edu](mailto:swathi.kaluvakuri@siu.edu),  
[bidyut@cs.siu.edu](mailto:bidyut@cs.siu.edu)

## Abstract

Application level multicast is independent of router infrastructure unlike router-based IP multicast. The existing DHT-based application level multicast protocols work efficiently as long as there is almost no churn; otherwise, their performances start degrading drastically, because DHT – based architecture cannot handle churn effectively. Besides, most of DHT-based multicast protocols consider single data source and do not consider peer heterogeneity. In this work, we have considered an existing non-DHT based P2P architecture, viz., Residue Class based (RC-based) architecture which has already been shown to perform much better than some well-known DHT-based architectures from the viewpoints of speed of unicast communication and churn handling. We have presented a highly efficient capacity-constrained and any source multicast protocol suitable for the RC-based P2P architecture as mentioned above.

**Keywords – P2P network; Residue class; Capacity constrained, multicast protocol**

## I. INTRODUCTION

Problems associated with the global deployment of multicast-capable routers, lack of support for network management, and also the scalability problem caused by the simultaneous presence of large number of multicast sessions, are some of the main reasons why the deployment of router-based IP multicast has been slow. Consequently, researches have started considering application level multicast as an alternative to IP multicast, because the former one can be deployed fast as it does not depend on router infrastructure [1], [2], [3]. Multicast protocols proposed in [4], [5], [6] focus on designing an

optimized overlay multicast tree per multicast source. They can work well for certain applications, such as software distribution. There exist some interesting multicast protocols designed to work in DHT-based architectures [7], [8] However, none of these above mentioned works considers node heterogeneity. In addition, their performance degrades sharply as frequency of node (peer) movements in and out of the network increases.

In non-DHT-based structured P2P networks, overhead of maintaining the structural properties in the presence of churn has been shown to be much less compared to that in DHT-based structures [9], [10]. Note that frequent joining of new peers and leaving of existing peers is known as churn. However, very few such structures exist in the literature [9], [10]. Authors in [10] have proposed a unique way of designing non-DHT based P2P network architecture; they have designed a very effective multicast protocol suitable for the architecture, which is characterized by (1) any source multicast and (2) host heterogeneity, i.e. peers can have different capacities/degrees.

The work in [9] has used some idea from the work in [10] and modified it appropriately to make it suitable for their proposed non-DHT based architecture. A brief and clear description of how the protocol in [10] works and how our present contribution relates to that work has appeared in Section 3. However, some relevant preliminaries about the architecture [9] have appeared in Section 2, because we think that it will be helpful to have a better understanding of the materials stated in Section 3. The multicast protocol along with its performance have appeared in Section 4.

## II. PRELIMINARIES

In this section, we start with a brief description of the Residue Class (RC)-based hierarchical structured P2P architecture [9]; note that we design our multicast protocols for this architecture. This architecture is a two-level structured one for interest-based peer-to-peer system [9], [11], [15].

### A. Two Level Hierarchy

A resource is defined as a tuple  $\langle R_i, V \rangle$ , where  $R_i$  denotes the type of a resource and  $V$  is the value of the resource. A resource can have many values. For example, let  $R_i$  denote the resource type 'songs' and  $V$  denote a particular singer. Thus  $\langle R_i, V \rangle$  represents songs (some or all) sung by a particular singer  $V$ . In the proposed model for interest-based P2P systems, it is assumed that no two peers with the same resource type  $R_i$  can have the same tuple; that is, two peers with the same resource type  $R_i$  must have tuples  $\langle R_i, V' \rangle$  and  $\langle R_i, V'' \rangle$  such that  $V' \neq V''$ . However, this constraint can easily be relaxed [9]. In addition, let  $S$  be the set of all peers in a peer-to-peer system. Then  $S = \{P^{R_i}\}$ ,  $0 \leq i \leq n-1$ . Here  $P^{R_i}$  denotes the subset consisting of all peers with the same resource type  $R_i$  and no two peers in  $P^{R_i}$  have the same value for  $R_i$  and the number of distinct resource types present in the system is  $n$ . Also for each subset  $P^{R_i}$ ,  $P_i$  is the first peer among the peers in  $P^{R_i}$  to join the system.

The two level overlay architecture is shown in Fig. 1 and at each level structured networks of peers exist. At level-1, we have a ring network consisting of the peers  $P_i$  ( $0 \leq i \leq n-1$ ). The number of peers on the ring is  $n$  which is also the number of distinct resource types. This ring network is used for efficient data lookup and so is named as transit ring network. At level-2, there are  $n$  numbers of completely connected networks (groups) of peers. Each such group, say  $G_i$  is formed by the peers of the subset  $P^{R_i}$ , ( $0 \leq i \leq n-1$ ), such that all peers ( $\in P^{R_i}$ ) are directly connected (logically) to each other, resulting in the network diameter of 1. Each  $G_i$  is connected to the transit ring network via its group-

head  $P_i$ . Each peer on the transit ring network maintains a global resource table (GRT) that consists of  $n$  number of tuples. GRT contains one tuple per group and each tuple is of the form  $\langle \text{Resource Type, Resource Code, Group Head Logical Address} \rangle$ , where Group Head Logical Address refers to the logical address assigned to a node by the proposed overlay P2P architecture. Also, Resource Code is the same as the group-head logical address. Any communication between a peer  $p_i \in G_i$  and  $p_j \in G_j$  takes place only via the respective group-heads  $P_i$  and  $P_j$ .

In this context, it may be noted that any structured P2P system, the mathematical model used to build the architecture defines neighborhood relations among peers. The mathematical model is intimately related to the efficiency of different data lookup schemes used in a given structured P2P system. We now state a brief sketch of the mathematical model used in this approach to realize the architecture [9].

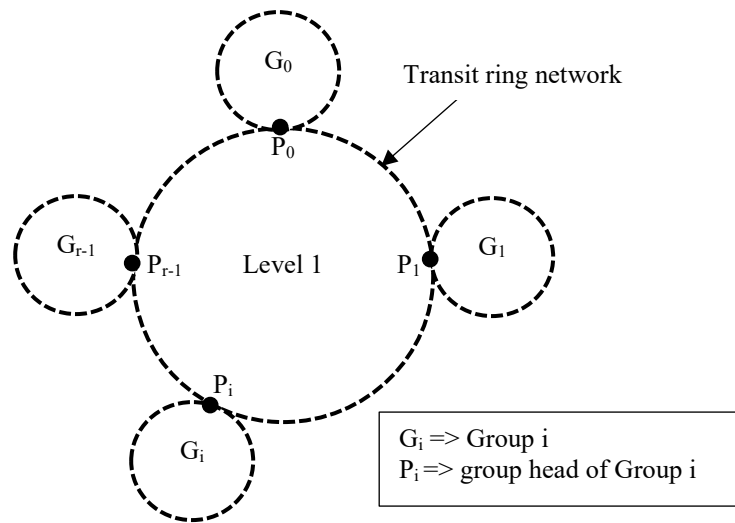


Fig. 1 A two-level structured architecture with distinct resource types

### B. Relevant Properties of Modular Arithmetic

Consider the set  $S_n$  of nonnegative integers less than  $n$ , given as  $S_n = \{0, 1, 2, \dots, (n-1)\}$ . This is referred to as the set of residues, or *residue classes (mod  $n$ )*. That is, each integer in  $S_n$  represents a residue class (RC). These residue classes can be labelled as  $[0], [1], [2], \dots, [n-1]$ , where  $[r] = \{a: a \text{ is an integer, } a \equiv r \pmod{n}\}$ .

For example, for  $n = 3$ , the classes are:

$$\begin{aligned} [0] &= \{\dots, -6, -3, 0, 3, 6, \dots\} \\ [1] &= \{\dots, -5, -2, 1, 4, 7, \dots\} \\ [2] &= \{\dots, -4, -1, 2, 5, 8, \dots\} \end{aligned}$$

Thus, any class  $r \pmod{n}$  of  $S_n$  can be written as follows:

$$[r] = \{\dots, (r-2n), (r-n), r, (r+n), (r+2n), \dots, (r+(j-1)n), (r+jn), (r+(j+1)n), \dots\}$$

In the P2P architecture, the numbers belonging to different classes are used as the logical addresses of the peers; therefore, for the sake of simplicity only the positive integer values are used. Before we present the mechanism of logical address assignments, we state the following relevant property of residue class [9].

**Lemma 1.** *Any two numbers of any class  $r$  of  $S_n$  are mutually congruent.*

### C. Assignments of Overlay Addresses

Assume that in an interest-based P2P system there are  $n$  distinct resource types. Note that  $n$  can be set to an extremely large value a priori to accommodate large number of distinct resource types. Consider the set of all peers in the system given as  $S = \{P^{R_i}\}$ ,  $0 \leq i \leq n-1$ . Also as mentioned earlier, for each subset  $P^{R_i}$  (i.e. group  $G_i$ ) peer  $P_i$  is the first peer with resource type  $R_i$  to join the system. In the overlay architecture, the positive numbers belonging to different classes are used to define the following:

- (a) Logical addresses of peers in a subnet  $P^{R_i}$  (i.e. group  $G_i$ ). Use of these addresses justifies the fact that all peers in  $G_i$  are directly connected to each other (logically) forming an overlay network of diameter 1. In graph theoretic term, each  $G_i$  is a complete graph.
- (b) Identifying peers that are neighbors to each other on the transit ring network.
- (c) Identifying each distinct resource type with unique code.

The assignment of logical addresses to the peers at the two levels and the resources happen as follows:

- 1) At level-1, each group-head  $P_r$  of group  $G_r$  is assigned with the minimum nonnegative number ( $r$ ) of residue class  $r \pmod n$  of the residue system  $S_n$ .
- 2) At level-2, all peers having the same resource type  $R_r$  will form the group  $G_r$  (i.e. the subset  $P^{R_r}$ ) with the group-head  $P_r$  connected to the transit ring network. Each new peer joining group  $G_r$  is given the group membership address  $(r + j.n)$ , for  $j = 0, 1, 2, \dots$
- 3) Resource type  $R_r$  possessed by peers in  $G_r$  is assigned the code  $r$  which is also the logical address of the group-head  $P_r$  of group  $G_r$ .
- 4) Each time a new group-head joins, a corresponding tuple  $\langle$ Resource Type, Resource Code, Group Head Logical Address $\rangle$  is entered in the global resource table (GRT).

Below we have summarized the main characteristics of the architecture in the form of Lemmas and observations [9].

**Remark 1.** GRT remains sorted with respect to the logical addresses of the group-heads.

**Definition 1.** *Two peers  $P_i$  and  $P_j$  on the ring network are logically linked together if  $(i + 1) \pmod n = j$ .*

**Remark 2.** *The last group-head  $P_{n-1}$  and the first group-head  $P_0$  are neighbors based on Definition 1. It justifies that the transit network is a ring.*

**Definition 2.** *Two peers of a group  $G_r$  are logically linked together if their assigned logical addresses are mutually congruent.*

**Lemma 2.** *Diameter of the transit ring network is  $n/2$ .*

**Lemma 3.** *Each group  $G_r$  forms a complete graph.*

### III. MOTIVATION AND OUR CONTRIBUTION

In [10], a non-DHT-based P2P architecture has been proposed in which a capacity-constrained multicast protocol has been designed. The architecture is an unrestricted ring of  $N$  nodes (peers) in a sense that a node can be anywhere on the ring unlike in DHT-based ring, such as Chord [12]. In the later, location of a node is unique, which makes the process of joining an existing structure by a new node very complex. According to the work in [10] each node  $n_i$  has its successor on the ring and also each node randomly selects  $c_i$  number of other nodes on the ring as its immediate logical neighbors;  $c_i$  is the degree/capacity of node  $n_i$ . We have highlighted below some significant contributions of the work.

First, any node can be a source and propagation of a multicast packet from a source node follows a tree structure created implicitly even though there is no explicit multicast tree creation unlike in the classical multicast protocols that use either the source-based tree approach [13] or the shared tree approach [14]. Therefore, it contributes to the communication speed of the multicast packets. Besides, the complicated process of maintaining an explicit tree can altogether be avoided.

The second one is that the value of the depth (number of levels) of the implicit tree can be controlled and it can be a system parameter. Depending on the value of the depth the protocol can combine tree traversal (in its phase 1) and sequential traversal along the ring (phase 2); effectively it transforms the multicast problem to a broadcast one within the scope of the overlay network.

The third one is that it has considered host heterogeneity; therefore, has taken into consideration the degree/capacity of a node while forwarding multicast packets to other nodes.

Authors in [10] have used the following definitions of hop complexity and communication complexity, which we shall use in our contribution. The hop complexity is the number of overlay hops needed by a multicast packet to arrive at any node and the communication complexity is the expected number of copies of a multicast packet that are transmitted in order to ensure that it arrives at all nodes. Note that the number of copies increases by one each time an intermediate node forwards the copy to its neighbor. Let  $c$  be the average node capacity. Then, the hop complexity has been shown to be  $O(\log_c N)$  that is asymptotically optimal; the communication complexity is  $[N + O(N/\log_c N)]$  that is close to  $N$ . On the other hand, if only sequential traversal is followed to send a multicast packet to any receiver, these are  $N/2$  and  $N$  respectively.

We now illustrate briefly the working of the approach. Let a multicast source peer, say  $P_s$  with degree, say 2, has randomly selected two nodes  $P_i$  and  $P_j$  on the ring as its immediate neighbors. Let the respective degrees of  $P_i$  and  $P_j$  be 2 and 3.

First,  $P_s$  delivers the packet to  $P_i$  and  $P_j$ . Node  $P_i$  in turn delivers the packet to its two randomly selected neighbors, say  $P_k$  and  $P_l$ , which do the same to their respective randomly chosen neighbors. Node  $P_j$  does the same. When a node receives a duplicate packet, it sends a control packet asking the sender not to forward any more. Therefore, implicitly packet propagation follows a tree architecture. The work also suggests the following way to transmit the packet. In the above example, assume that  $P_i$  and  $P_j$  do not deliver the packet to their randomly selected immediate neighbors; instead, they forward the packet to their respective successors on the ring and a receiving node drops a packet if it is a duplicate packet, otherwise it forwards the packet to its successor. Therefore, it is actually a sequential traversal on the ring after the initial delivery of the packet by  $P_s$  to  $P_i$  and  $P_j$ . Note that peers  $P_s$ ,  $P_i$ , and

$P_j$  effectively have formed a tree of depth one with  $P_s$  as its root and the other two as the leaves. Observe that any multicast group member never misses a packet and any packet arrives at every node. Therefore, it is clear that multicasting is being done via broadcasting.

#### Our Contribution

We have considered designing a highly efficient capacity-constrained overlay multicast protocol; it is designed specifically for an existing RC - based hierarchical and structured P2P architecture [9]. Such an architecture has been the choice for its structural advantages. The protocol is not restricted to a single data source and it incorporates peer heterogeneity as well. To determine the performance we shall use hop complexity and communication complexity as the parameters. As pointed out earlier that the architecture in [10] is a ring that consists of  $N$  peers in the overlay network. Our architecture is a 2-level one. Number of nodes (group-heads)  $n$  on the level-1 ring is just the number of distinct resource types and in any group (cluster) at level 2 there can be any number of nodes. Note that the number of distinct resources  $n$  is much smaller than the total number of nodes  $N$  on the ring in [10]. Therefore, for ring traversal at level 1 in our architecture, communication complexity is just  $n$  instead of  $N$  and the hop complexity is just  $n/2$  instead of  $N/2$ . It has inspired us to use some idea from [10], especially transforming the multicast problem to a broadcast one and appropriately augmenting it with ours to design a highly efficient any source capacity-constrained multicast protocol suitable for the RC-based architecture with much less hop and communication complexities compared to the work in [10]. In this paper, we shall consider inter-group multicast communication as it is much more complex than the intra-group one.

## IV. OVERLAY MULTICAST PROTOCOL

We shall use the following notations in the proposed protocol. Assume that the  $k^{\text{th}}$  peer in group  $G_i$  (with group-head  $P_i$ ) wants to multicast packets. This source peer is denoted as  $p_{i,k}^s$  and the group-head  $P_i$  will then be denoted as  $P_i^s$  and be called as the source group-head, even though it is not the actual source of multicast. The degree/capacity of  $P_i^s$  is denoted as  $c_i^s$ . In a similar way, the  $m^{\text{th}}$  receiver peer (a multicast group member) in group  $G_j$  is denoted as  $p_{j,m}^r$  and the corresponding group-head as  $P_j^r$ . This group-head will be called a receiver group-head, because even though it may or may not be a multicast group member, it must deliver multicast packets received from other group-heads to peers like  $p_{j,m}^r$ . We assume that the receiver peer  $p_{j,m}^r$  sends a join request to its group-head in order to receive multicast packets sent for this multicast group.

Note that any group-head can itself be a multicast group member or a multicast source as well; besides, a receiver may exist in the same group in which a source peer exists. Finally, in our architecture, each group-head has a global resource table (*GRT*) that has every group-head's logical as well as IP addresses. This piece of information is essential to our proposed protocol. In this protocol,  $n_r$  denotes the number of receiver group-heads and *mcast\_msg* denotes a multicast message. We will consider the following two possible cases: (1)  $c_i^s \geq n_r$  and (2)  $c_i^s < n_r$ . The protocol appears in Figs. 2a and 2b.

*Protocol\_OM-cast*

Case 1:  $c_i^s \geq n_r$

1. source Peer  $p_{i,k}^s$  unicasts *mcast\_msg* to group-head  $P_i^s$
2.  $P_i^s$  unicasts *mcast\_msg* to each of the  $n_r$  receiver group-heads  
*//  $P_i^s$  gets the IP addresses from its GRT; one logical hop to each group-head*
3. **if** a receiver group-head  $P_j^r$  is also a multicast group member, **then**  
it keeps a copy of *mcast-msg* to itself  
 $P_j^r$  unicasts *mcast\_msg* to each receiver  $p_{j,m}^r$  in its group  
**else**  
 $P_j^r$  unicasts *mcast\_msg* to each receiver  $p_{j,m}^r$  in its group  
  
*// one logical hop from group-head to each receiver since a group is a complete graph*  
*total number of copies of mcast\_msg is ( $n_r$  + number of receivers in groups)*

Fig. 2a Multicast protocol: Capacity of the group head ( $c_i^s$ )  $\geq$  number of receiver group-heads ( $n_r$ )

Case 2:  $c_i^s < n_r$

1.  $P_i^s$  unicasts *mcast\_msg* to each of the  $c_i^s$  group-heads  
*//  $P_i^s$  gets the IP addresses of the  $c_i^s$  receiver group-heads from its GRT;*  
*one logical hop to each group-head.*
2.  $P_i^s$  forwards *mcast\_msg* to its successor on the level-1 ring
3. Each receiver group-head  $P_j^r$  unicasts *mcast\_msg* to each receiver  $p_{j,m}^r$  in its group  
*// one logical hop from group-head to each receiver since a group is a complete graph*
4. **if**  $P_j^r$  is a multicast group member, **then**  
keeps a copy of *mcast-msg* to itself  
forwards *mcast\_msg* to its successor on the level-1 ring  
**else**  
forwards *mcast\_msg* to its successor on the level-1 ring
5. **if** a successor group-head receives a duplicate message, **then**  
it discards the message and stops forwarding  
**else**  
**if** a successor group-head is both a receiver group-head and a multicast group member, **then**  
keeps a copy of *mcast-msg* to itself;  
it unicasts *mcast\_msg* to receivers in its group;  
forwards the *mcast-msg* to its successor on the ring;  
**else**  
**if** a successor group-head is a receiver group-head, **then**  
it unicasts *mcast\_msg* to receivers in its group  
forwards the *mcast-msg* to its successor on the ring  
**else**  
it forwards the *mcast-msg* to its successor on the ring

Fig. 2b Multicast protocol: Capacity of the group head ( $c_i^s$ )  $<$  number of receiver group-heads ( $n_r$ )

### Performance

We observe that when  $c^s_i \geq n_r$ , a packet takes a maximum of 3 overlay hops to reach a multicast group member from a source. However, in the case of  $c^s_i < n_r$ , the hop complexity is  $O(n)$ . It is interesting to note that the capacity of the source group-head has apparently no significant effect on the hop complexity, which is not the case in [10] which offers a hop complexity of  $O(\log_c N)$ . We note that in general  $n \ll N$ , therefore, our proposed architecture offers improved performance than the one in [10].

In [10] the communication complexity is  $[N + O(N/\log_c N)]$  that is close to  $N$ . In our proposed work, maximum number of copies from a source to a multicast group-member is 3 when  $c^s_i \geq n_r$ . When  $c^s_i < n_r$ , it is  $(n + (c^s_i + 1) + 1)$  and it is close to  $n$  since in general,  $c^s_i \ll n$ . Therefore, we conclude that our architecture offers improved performance than in [10] in the case of communication complexity as well.

## V. CONCLUSION

In this paper, we have considered an existing RC - based P2P architecture for its structural advantages over DHT-based architectures. We have used some of its unique structural characteristics to design an efficient overlay multicast communication protocol. The multicast protocol is not restricted to a single data source and it incorporates peer heterogeneity. We have used the idea of transforming the multicast problem to a broadcast one [10] and appropriately augmented it to design the proposed protocol with less hop and communication complexities compared to the work in [10]. Future work is directed at building architecture for P2P federation using the idea of this work.

## REFERENCES

- [1] Y.H. Chu, S. Rao, S. Seshan, and H. Zhang, "A Case for End System Multicast," IEEE J. Selected Areas in Comm, vol. 20, no. 8, Oct. 2002.
- [2] J. Jannotti, D. Gifford, K. Johnson, M. Kaashoek, and J. O'Toole, "Overcast: Reliable Multicasting with an Overlay Network," Proc. Symp. Operating Systems Design and Implementation (OSDI'00), Oct. 2000.
- [3] C.K.S. Banerjee and B. Bhattacharjee, "Scalable Application Layer Multicast," Proc. ACM SIGCOMM'02, Aug. 2002.
- [4] J. A. Dejan Kosti, A. Rodriguez, and A. Vahdat, "Bullet: High Bandwidth Data Dissemination Using an Overlay Mesh," Proc. Symp. Operating Systems Principles (SOSP,03), Oct. 2003.
- [5] S. Banerjee, C. Kommareddy, B.B.K. Kar, and S. Khuller, "Construction of an Efficient Overlay Multicast Infrastructure for Real-Time Applications," Proc. INFOCOM'03, Mar. 2003.
- [6] A. Riabov and L. Z. Zhen Liu, "Overlay Multicast Trees of Minimal Delay," Proc. Int'l. Conf. Distributed Computing Systems (ICDCS)'04, Mar. 2004.
- [7] R. Zhang and Y. C. Hu, "Borg: A Hybrid Protocol for Scalable Application-Level Multicast in Peer-to-Peer Networks," Proc. Int'l. Workshop Network and Operating System Support for Digital Audio and Video (NOSSDAV'03), 2003.
- [8] S. Ratnasamy, M. Handley, R. Karp, and S. Shenker, "Application Level Multicast Using Content-Addressable Networks," Proc. Int'l. Workshop Networked Group Comm (NGC'01), 2001.
- [9] Swathi Kaluvakuri, Koushik Maddali, Indranil Roy, Banafsheh Rekabdar, Ziping Liu', Bidyut Gupta, "Design of RC-based Low Diameter Hierarchical Structured P2P Network Architecture," to appear in Proc. Proc. IEEE 17<sup>th</sup> Int. Conf. Industrial Informatics (IEEE INDIN), July 2019, Helsinki, Finland



- [10] Shiping Chen, Baile Shi, Shigang Chen, and Ye Xia, "ACOM: Any-Source Capacity-Constrained Overlay Multicast in Non-DHT P2P networks," *IEEE Tr. Parallel and Distributed Systems*, vol. 18, no. 9, pp. 1188-1201, Sep. 2007.
- [11] A. Rowstron and P. Druschel, "Pastry: scalable, distributed object location and routing for large scale peer-to-peer systems," *Proc. IFIP/ACM Intl. Conf. Distributed Systems Platforms (Middleware)*, pp. 329-350, 2001.
- [12] I. Stocia, R. Morris, D. Liben-Nowell, D. R. Karger, M. Kaashoek, F. Dabek, and H. Balakrishnan, "Chord: a scalable peer-to-peer lookup protocol for internet applications," *IEEE/ACM Tran. Networking*, vol. 11, No. 1, pp. 17-32, Feb. 2003.
- [13] Stephen E. Deering and David R. Cheriton, "Multicast Routing in Datagram Internetworks and ExtendedLANs", *ACM Transactions on Computer Systems (TOCS)*, Vol. 8, No. 2, pages. 85-110, May 1990.
- [14] Tony A. Ballardie, "Core Based Tree Multicast Routing Architecture", *Internet Engineering Task Force(IETF), RFC 2201*, September 1997.
- [15] M. Yang and Y. Yang, "An Efficient Hybrid Peer-to-Peer System for Distributed Data Sharing", *IEEE Trans. Computers*, vol. 59, no. 9, pp. 1158-1171, Sep. 2010.