

## PEER TO PEER FILE SHARING USING NETWORK CODING

Ajay Choudhary<sup>1</sup>, Nilesh Akhade<sup>2</sup>, Aditya Narke<sup>3</sup>, Ajit Deshmane<sup>4</sup>

Department of Computer Engineering, University of Pune  
Imperial College of Engineering and Research  
Pune, India

<sup>1</sup>koolacac@gmail.com, <sup>2</sup>akhadenilesh93@gmail.com,  
<sup>3</sup>adityanarke62@gmail.com, <sup>4</sup>ajit.deshmane555@gmail.com

**Abstract:** *Structured peer-to-peer overlay networks provide a solution for the construction of large scale, decentralized applications, including distributed storage, group communication, and content distribution. These overlays are highly flexible; they can route messages correctly even when a large numbers of the nodes crash. Network coding is beneficial technique to improve peer to peer file sharing networks. Our idea is to split the file to be distributed in fixed length blocks, later each block would be encoded and routed independently in the peer to peer network overlay. The overlay networks create a virtual topology on top of the physical topology. Because overlay networks define neighbor nodes by content stored, they can change search from a standard graph-traversal problem into a localized iterative process. Network coding can improve throughput, robustness, complexity, and security. P2P File sharing network are secure as compared to centralized file sharing as there no need to upload the files on the third party server. In network coding each peer is computational node so it can perform encoding and decoding operation rather than simply storing and forwarding. Networks such as Gnutella organize nodes independently of the underlying physical topology; "neighbors" might exist within the same sub network or across the Internet, they are classified as unstructured P2P networks. Our scheme focuses on constructing peer to peer network overlay and improves routing with the help of intermediate nodes. The routing scheme is dynamic and decided by central server. The server constructs graph, finds out best path for routing. Each peer follows server guidelines for routing and participates in peer to peer network.*

**Keywords:** Peer to Peer Network Overlay, File sharing, Network Coding.

## 1. INTRODUCTION

In today's network, messages are generally transferred by routing through intermediate nodes between the source and the destination, i.e., by having intermediate nodes store and forward messages. In fact, routing is not the only operation that can be performed at a node. Recently, network coding has emerged as a promising enhancement of routing to improve network throughput and provide high reliability [2]. Network coding refers to a scheme where a node is allowed to generate output messages by encoding (i.e., computing certain functions of) its received messages. Thus, network coding allows information to mix, in contrast to the traditional routing approach where each node simply forwards received messages.

## 2. LITERATURE SURVEY

There are peer to peer systems already available each one with different approach and different services. P2P systems can be classified into two different classes:

1. structured P2P systems
2. Unstructured P2P systems.

In structured P2P systems, connections among peers in the network are fixed, and peers maintain information about the resources (e.g., shared content) that their neighbor peers possess. Hence, the data queries can be efficiently directed to the neighbor peers that have the desired data, even if the data is extremely rare.

In unstructured P2P systems, connections among peers in the network are formed arbitrarily in flat or hierarchical manners. In order to find as many peers that have the desired content as possible, peers in unstructured P2P systems query data based on several techniques such as flooding and expanding-ring. Three different designs of unstructured P2P systems exist:

- a) Centralized unstructured P2P systems
- b) Decentralized (or pure) unstructured P2P systems
- c) Hybrid unstructured P2P systems

In a centralized unstructured P2P system, a central entity is used for indexing and bootstrapping the entire system. A Bit Torrent network is an example of a centralized unstructured P2P network. Napster, the network that pioneered the idea of P2P file sharing, is another example of a centralized design. In Napster, a server (or server farm) is used to provide a central directory. A peer in the network informs the directory server of its IP address and the names of the contents that it makes available for sharing. Thus, the directory server knows which objects each peer in the network have, and then, creates a centralized and dynamic database that maps content name into a list of IPs.

The decentralized (or pure) unstructured P2P network is an overlay network. An overlay network is a logical network. An edge in this network exists between any pair of peers that maintain a TCP connection. Gnutella is an example of a decentralized unstructured P2P network. In order to join the Gnutella network, a user initially connects to one of several known bootstrapping peers. The bootstrapped peers then respond with the information about

one or more existing peers in the overlay network. This information includes the IP address and port of each peer. The peers in Gnutella are aware only of their neighbor peers [1].

A hybrid unstructured P2P network allows the existence of infrastructure nodes, often referred to as super-peers (or super-nodes or overlay nodes). This creates a hierarchical overlay network that addresses the scaling problems on pure unstructured P2P networks such as Gnutella. A peer in such network can typically change roles over time. KaZaA , is an example of a hybrid unstructured P2P network.

### 3. PROPOSED SYSTEM

To overcome the drawbacks of current system as written above we will go through proposed system of peer to peer file sharing using network coding. In recent literature, network coding has emerged as a promising information theoretic approach to improve the performance of both peer-to-peer (P2P) and wireless networks [4]. It has been widely accepted and acknowledged that network coding can theoretically improve network throughput of multicast sessions in directed acyclic graphs, achieving their cut-set capacity bounds. Recent studies have also supported the claim that network coding is beneficial for large-scale P2P content distribution, as it solves the problem of locating the last missing blocks to complete the download. The proposed system utilizes link capacity by means of encoding message, to obtain max-flow through link.

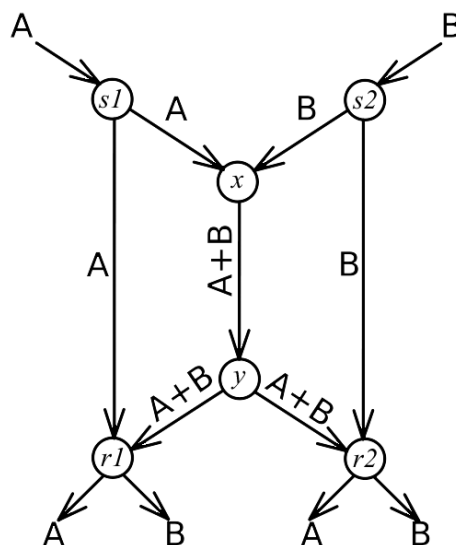


Fig.1: Maximum utilization of link  $x - y$ .

Fig.1 shows data pieces A and B, are to be multi-casted. The node x encodes the messages from s1 and s2 using simple XOR operation, ie,  $f(A,B)=A \oplus B$ . Without network coding x would send A and B separately, that will require one additional time slot. Receiver nodes r1 and r2 performs decoding operation,  $B=f(A,B) \oplus A$ , and  $A=f(A,B) \oplus B$  respectively. Hence

they can recover remaining piece from encoded message. The proposed system is explained as following:

1. Architecture defines two modules, one at server side, and another at peer side. Unlike other P2P distribution networks [2], the server is not source of files in this system.
2. Peers are source of files, which can share files with any other peer.
3. Background process running at peer module participates in P2P sharing activity, which will help others, by means of encoding messages. The message is chunk of file content.
4. Peer will contact server for file availability through control link.
5. The network overlay is a directed graph, in which nodes are peers in the network, and edge is bandwidth between them. Server will construct network overlay dynamically on each peer JOIN/LEAVE.
6. Server has database of files, each file is identified by its checksum. File checksum is the result of a hash function over the file content. This will avoid duplicate entry of same file into the database.
7. The server decides routing of message. It will detect sub graphs which resembles with butterfly network [2] into network overlay graph, and inform intermediate nodes to participate. Hence server assists the peers to download the file.

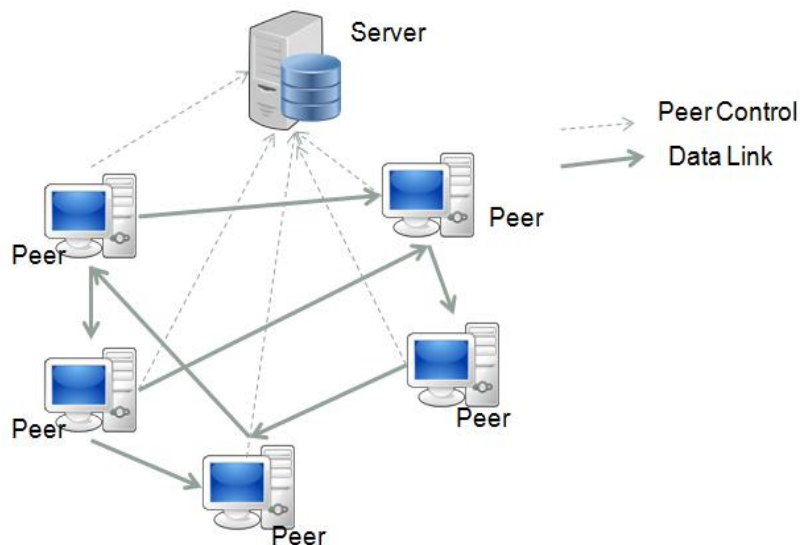


Fig.2: Peer To Peer File Sharing Network

#### 4. ADVANTAGES OF PROPOSED SYSTEM

**(a) Scalability:** Files are distributed through a peer-to-peer network. With the increase of the network size, the total available bandwidth also increases.

**(b) Efficiency:** The linear network coding scheme is deterministic and easy to implement. There is no requirement for peers to collaborate to construct the linear coding scheme on demand.

**(c) Reliability:** The redundant links can greatly improve the reliability of the system with little overhead.

**(d) Topology awareness:** The clustering scheme can greatly reduce link stress and improve throughput. The network overlay construction considers link speed as parameter for topology.

**(e) Heterogeneity support:** In case that links have different link capacities, our scheme can arrange the overlay topology to maximize the utilization of each peer's link capacity.

## 5. CONCLUSION

The network coding based file sharing proven to be efficient as per theory. Effect of network coding can be realized in large networks. The BitTorrent, another P2P file sharing networks also splits the file into pieces but it has few drawbacks. It is decentralized peer to peer network, hence it is difficult to administer, also it causes internet speed and computer performance degradation. The proposed network alleviates these drawbacks, by proposing novel protocol. In this project we are implementing structured, efficient and secure P2P file sharing network, which uses theory of network coding to enhance current P2P networks. We conclude that network coding enhances Peer to Peer networks with efficient routing mechanism.

## ACKNOWLEDGEMENT

We would like to thank our Guide Mrs. Darshika Lothe Madam for her support, co-operation and encouragement extended during publication of paper.

## REFERENCES

- [1]. "Secure Routing for Structured Peer-to-Peer Overlay Networks (2002)", by Miguel Castro , Peter Druschel , Ayalvadi Ganesh , Antony Rowstron , Dan S. Wallach
- [2]. "Applying Network Coding to Peer-to-Peer File Sharing", by Min Yang and Yuanyuan Yang.
- [3]. "Overlay Networks A Scalable Alternative for P2P", by Diego Doval and Donal O'Mahony
- [4]. "Network coding in live peer to peer streaming" by Mea Wang and Baochun Li.