



GPS Based Topology Matching Algorithm For P2P Systems

B LalithaDepartment of CSE
JNTUACEA, Anantapur, India**Dr. Ch D V Subba Rao**Department of CSE
S V University, Tirupati, India

Abstract: *In a p2p environment network mismatch is a generally occurring scenario which will hinder the performance of nodes in that environment. This paper proposes a topology matching algorithm that uses the GPS system which aims to alleviate the mismatch in P2P systems without reducing the search scope and convergent speed.*

Keywords: *Peer-to-Peer, Topology Mismatch, GPS, PeerSIM*

1. INTRODUCTION:

A peer-to-peer (abbreviated to P2P) computer network is one in which each computer in the network can act both as a client as well as a server for the other computers in the network, allowing shared access to various resources such as files, peripherals, and sensors without the need for a central server. Each network type requires all computers in the network to use the same or a compatible program to connect to each other and access files and other resources found on the other computer. P2P networks can be used for sharing content such as audio, video, data, or anything in digital format. Each computer or device in the network is referred to as 'node'. A node is a connection point, either a redistribution point or a communication endpoint. Peer-to-Peer (P2P) computing has emerged as a popular model aiming at further utilizing Internet information and resources. However, the mechanism of peers randomly choosing logical neighbors without any knowledge about underlying physical topology can cause a serious topology mismatch between the P2P overlay network and the physical underlying network. The topology mismatch problem brings great stress in the Internet infrastructure. It greatly limits the performance gain from various search or routing techniques.

Traditional topology optimization techniques identify physically closer nodes to connect as overlay neighbors, but could significantly shrink the search scope. Recent efforts have been made to address the mismatch problem without sacrificing the search scope, but they either need time synchronization among peers or have a low convergent speed. This paper proposes a GPS based topology matching algorithm which aims to alleviate the mismatch without reducing the search scope and convergent speed.

The overlays used to implement an unstructured peer to peer network are setup in a random manner which causes a significant mismatch between the overlay network and the underlying physical network. This mismatch often results in iterative transmission of queries which causes congestion in the network. Currently there are plenty of methods available to solve the topology mismatch problem.

This paper analyses several traditional optimal solutions available that are proven to solve the mismatch problem and comes out with a new method to solve the mismatch problem.

2. Related Work:

A Two-Hop Solution of Solving Topology Mismatch [2] algorithm is used to solve mismatch in decentralized and unstructured network such as Guntella, KaZaA etc., THANCS to dynamically connect closer peers and disconnect distant ones. This algorithm has 2 main components. Piggybacking neighbour distance on queries and neighbour comparison and selection. Each peer probs distance with its immediate logical neighbour based on network delay and use that to connect or disconnect other peers.

A near optimal algorithm attacking the topology mismatch problem in unstructured peer-peer network [3]. This is used in Gthutella to improve message flooding in an overlay. This uses an optimal algorithm to solve topology mismatch based on the metropolis hasting method. This algorithm optimizes an existing overlay network by minimizing broadcast delay and maximizing broadcast scope.

Location awareness in unstructured peer-peer systems [4] is used in decentralized p2p networks. These uses a location aware topology mismatch (LTM) algorithm. In LTM each peer issues a detector in a small region so that the peers

receiving the detector can record relative delay information based on the delay information, a receiver can detect and but most of the inefficient and redundant logical links and add closer nodes as its direct neighbours.

Large scaling Unstructured p2p with heterogeneity aware topology and routing [5] improves topology of the overlay network using heterogeneity awareness. This is widely used in KaZaa application. An overlay is implemented with multi-tier heterogeneity aware topologies. This paper develops an analytical model to construct multitier network connection aware topologies and an probabilistic selective routing algorithm that enhances the routing performance.

Reducing network overhead with common junction methodology [6] is used in Decentralized unstructured P2P for decentralized traffic. Here Common Junction methodology (CJM) is used to solve the topology mismatch. CJM reduces network overhead by optimizing the overlay traffic at underlay level. CJM finds common junction between available paths and traffic is only routed through the common junction and not through the conventional identified paths.

Building a Scalable Bi-partite P2P overlay network [7] in Guntella and KaZaA reduces search scope and redundant traffic. Scalable Bi-partite overlay (SBO) scheme to optimize the overlay topology by identifying and replacing the mismatched connections. SBO use an efficient strategy for distributing optimization tasks to peers with different colours (white and red). The colour of the peer is assigned at the bootstrapping stage, and a peer will only be connected with peer of different colour. A white peer probes the cost information of its red neighbours. The cost information of peers within 2 hops, each peer will built a MST which is the base for the white peer to replace or cut mismatched connections.

Semantic based and location aware unstructured P2P (SLUP) [8] is used for Lengthy queries and increased bandwidth in decentralized unstructured P2P, to shorter the length of the query and to reduce the bandwidth consumed. This uses semantic based and location aware unstructured P2P model. All the nodes are clustered into domains according to the physical distance then the nodes in the domain are organized into groups based on semantic similarities of shared resources. When a query occurred it will first be routed in the nodes with close physical distance. Moreover, the query will be forwarded to the groups who have related content with query in the domain of the resources is not found in the domain, the query is then forwarded to the other domains.

2.1 Probabilistic Models for Peer-to-Peer networks:

There are several probabilistic models to describe the working of P2P networks. The models are used to explain the dynamics of this P2P streaming network. Here we make use of marginal probability models, which presents a good platform for performing deep analysis of the p2p system behavior.

2.1.1 The Marginal Probability Model

The marginal probability model, was proposed by Zhou et al. [1]. It assumes that the states of individual buffer locations are independent from each other. Hence the shared allotment of a particular buffer state can be written in terms of the product of the marginal probabilities of all buffer locations, reducing the dimension of the state space from $2N$ to N .

Assuming identical and independent peer state distribution for all peers, let p_i denote the steady-state marginal probability that buffer location i has a chunk in the current time slot.

$$p_i = \sum_{x \in \{X: x(i)=1\}} p_s(x)$$

The marginal probabilities $\{p_i, i \in \{1, 2, \dots, N\}\}$ are defined by the following equations [1]:

$$\begin{aligned} p_1 &= 0 \\ p_2 &= \frac{1}{M} \\ p_{i+1} &= p_i + p_i(1 - p_i)s_i(\delta), \forall i \in \{2, 3, \dots, N-1\} \end{aligned}$$

where $s_i(\delta)$ denotes the probability that one piece will be downloaded to the buffer position i in a given time slot, given a deterministic downloading policy δ . In later sections, the term $s_i(\delta)$ will take on specific expressions as we focus on particular classes of downloading policies.

3. Solution:

This paper proposes a P2P topology matching algorithm based on GPS and Location based services as shown in Fig 1. All the peers who wish to be part of a P2P network should register with a central server by providing its location information. The central server uses the location information of the peers and maps their position on the google map. The three major steps are 1) Peer Registration, 2) Peer Connection Establishment and 3) Peer Disconnection.

3.1. Process by which peers register

The peers who want to form a p2p network provide their location information in terms of co-ordinates to a central server.

- The central server stores all the location information in a database.
- Using the peer information in the database the server maps the position of the registered peers on the google map.
- The registered peers can view the position of the other registered peers from the google map.
- The peers can then choose a relatively close peer by analyzing the position of other peers on the map and send a connection request to it.

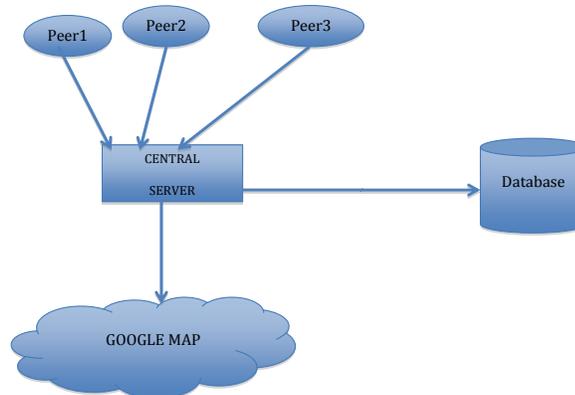


Fig 1: High Level Architecture Diagram

3.2. Establishing Connection

The peer sends a connection request message to the peers it wants to establish a connection with. Several connection attempts may be necessary because nodes may not be willing to accept new connections or may have left the network. If the receiving peer sends back a “connection-accepted” message a connection is established and the peers are connected in a network. Once the connection is established the nodes can then send “resource-query” message with a specified resource name. This message is flooded across the P2P network. If a receiving peer has a resource that matches the resource criteria specified in the “resource-query” message, it responds with a “resource-available” message which is back propagated to the sender. The peers can then exchange resources such as files using an appropriate transfer protocol. The receiving peer can accept or deny the connection. If a connection request is accepted the receiving peer sends a “connection-accepted” message else it sends a “connection-denied” message.

3.2.1. Authentication

A digital certificate based authentication mechanism is used by the central server to make sure only trustworthy peers are allowed in the P2P network. For this use of digital certificates is a mechanism and in this proposed methodology it is used.

Digital Certificates are the Most Common Method of Online Authentication. The most common method for authenticating the parties of an e-commerce transaction is via the exchange of digital certificates. A digital certificate is an electronic document that validates the identity of its owner. It is not a casual validation, nor is the digital certificate self-assigned. Rather, an independent entity, known as a certificate authority conducts the validation process to confirm the identity of an applicant, and then issues the digital certificate. As the final step in this process, the certificate authority endorses the certificate with its digital signature, sealing the contents contained therein. This digital signature is arguably the most important part of a digital certificate. It is not only the official "stamp" of the verification methods employed in creating this electronic document, it is also a unique representation of the certificate authority itself: the digital signature is a distinctive mark that cannot be replicated by another entity. When affixed to a digital certificate, the certificate authority's signature affirms that the registration and issuance process requirements have been satisfactorily met by the applicant, and that the identity of the certificate holder is valid as represented. Moreover, the user identity and credentialing information proscribed in the digital certificate cannot be manipulated once sealed; if attempted, then it is easily detectable.

It can be easily concluded, therefore, how much value this digital signature adds to an issued digital certificate, giving it the greatest assurance possible for both identity and document authenticity. As a result, it is the principle component of certificate authentication methods conducted by other parties. The Web is abounding with organizations and businesses; clients cannot possibly identify, much less confirm, which of these are legitimate and verifiable entities. But clients can identify a trusted third-party certification authority; particularly when SSL security technology is integrated into most Web browsers and servers, streamlining authentication security processes. For this reason, client browsers authenticate a presented digital certificate - and the identifiable information contained therein - by verifying the digital signature is that of a trusted certificate authority.

3.3. Peer Disconnection

A timeout mechanism is used to determine if a peer is still alive in the network. All the messages which are exchanged in the p2p network have a specified timeout value. If the receiving peer doesn't reply within the specified time out value the peer is disconnected from the network. There is also another provision where a peer wishing to leave a network sends a “leaving” message to all its neighbors indicating that it is disconnecting from the network.

Once the peer which wants to leave the network is identified, it is disconnected from the p2p network and a new neighbor can be chosen from the central server or the resources can be exchanged using the existing network.

4. Implementation:

The proposed work is implemented using PeerSim simulator and Newcast Protocol

4.1. PeerSim Simulator

Peer-to-peer (P2P) systems can be extremely large scale (millions of nodes). Nodes in the network join and leave continuously. PeerSim can cope with these properties and can reach extreme scalability and to support dynamism. In addition, the simulator structure is based on components and makes it easy to quickly prototype a protocol, combining different pluggable building blocks, that are in fact Java objects.

PeerSim 1.0 supports two simulation models: the cycle-based model and a more traditional event-based model. Here the cycle based model is used. This model is a simplified one, which makes it possible to achieve extreme scalability and performance, at the cost of some loss of realism. Several simple protocols can tolerate this loss without problems. The simplifying assumptions of the cycle based model are the lack of transport layer simulation and the lack of concurrency. In other words, nodes communicate with each other directly, and the nodes are given the control periodically, in some sequential order, when they can perform arbitrary actions, such as call methods of other objects and perform some computations.

Initialization needs to be performed, that sets up the initial states of each protocol. The initialization phase is carried out by Control objects that are scheduled to run only at the beginning of each experiment. In the configuration file, the initialization components are easily recognizable by the init prefix.

After initialization, the cycle driven engine calls all components (protocols and controls) once in each cycle, until a given number of cycles, or until a component decides to end the simulation. Each object in PeerSim (controls and protocols) is assigned a Scheduler object which defines when they are executed exactly. By default, all objects are executed in each cycle. However, it is possible to configure a protocol or control to run only in certain cycles, and it is also possible to control the order of the running of the components within each cycle.

In gossip-based aggregation, each node periodically selects a neighbor to communicate with (using an underlying overlay network) and in each such communication step they mutually update their approximations of the value to be calculated, based on their previous approximations.

A fixed P2P random topology composed of 50000 nodes is created. The values to be aggregated (averaged) at each node are initialized using a linear distribution on the interval (0, 100). Here to a numeric index called protocol ID, by the PeerSim engine. This index does not appear in the configuration file, but it is necessary to access protocols during a simulation.

4.2. Newcast Protocol

Newcast is an epidemic content distribution and topology management protocol. Every peer in the system has a partial view of the set of nodes which is in fact a fixed size set of node descriptors. Each descriptor is a tuple consisting of a peer address and a time-stamp recording the time when the descriptor was created.

Each node updates its state by choosing a random neighbor and exchanging views. After the exchange both peers merge the two views and keep the freshest entries only. In this manner, old information (descriptors) is removed from the system. This process allows the protocol to repair the overlay topology removing dead links with minimum effort which is very useful in highly dynamic system where nodes join and leave continuously.

Now the aggregation protocol runs on top of Newcast and a few other extensions are added. For example, there is a Control object that changes the network size, shrinking it by cutting out 500 nodes each time it is called, that is, once in each cycle from cycle 5 until cycle 10. The global parameters are the same as in the previous

5. CASE STUDY:

A number of peers send connection requests to a server running on a particular port. The server accepts the connection requests from various peers and issues a wait notification to them. The server then pairs the peers which are waiting on the same port. Now the connected peers can proceed to exchange messages.

5.1. GPS Module

The peers who want to form a p2p network provide their location information in terms of co-ordinates to a central server. The central server stores all the location information in a database using the peer information in the database registered peers on the google map. These registered peers can view the position of the other registered peers from the google map. These peers can then choose a relatively close peer by analyzing the position of other peers on the map and send a connection request to it.

6. Results:

The below graphical representation shows the connection establishment time take by the proposed model and the time frame when the peer comes out of the connection. The peer withdrawal doesn't affect any of the peer-peer transmission.

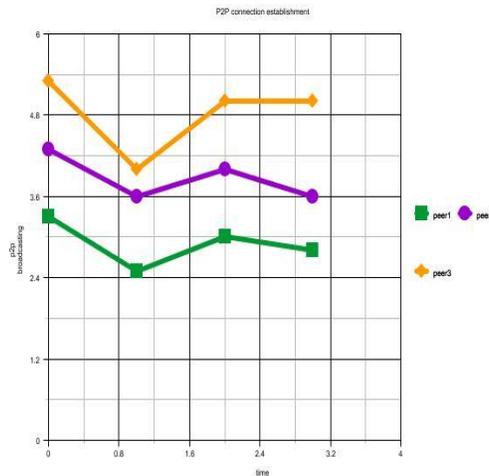


Fig 2: Time vs Peers

7. Testing:

Testing is performed with test cases derived regarding the 6 scenarios such as unavailability of resources, choosing distance peer as neighbor, server crash, request for available resources, TTL becoming zero before destination is reached, choosing a peer with same area code as neighbor.

8. Conclusion:

The quality of overlays significantly affects search efficiency of P2P systems. At least 60 percent of the queries and their responses are delivered to the data requestor along mismatched paths, increasing both traffic cost and search response time. Our GPS based Topology matching algorithm addresses these problems effectively and helps alleviate the topology mismatch problem. P2P networks using our algorithm has shown improved search scope, optimized query routing, optimized query path, minimized network congestion and optimized overlay connections.

9. Future Work:

Further this algorithm can be extended to solve the topology mismatch in Mobile ad-hoc networks. Hopefully it will have the same success which it has had in the personal computer based peer to peer networks.

References:

- [1] Yipeng Zhou, Dah-Ming Chiu, and John C. S. Lui, "A Simple Model for Analyzing P2P Streaming Protocols," ICNP 2007, pp. 226–235.
- [2] Yonhao Liu "A Two-Hop Solution to Solving Topology Mismatch," IEEE Transactions on Parallel And Distributed Systems Vol. 19, No. 11, November 2008.
- [3] Hung-Chang Hsiao, Hao Liao, and Po-Shen Yeh, "A Near-Optimal Algorithm Attacking the Topology Mismatch Problem in Unstructured Peer-to-Peer Networks," IEEE Transactions on Parallel and Distributed Systems, Vol.21, No.7, July 2010.
- [4] Yunhao Liu, Xaiomei Liu, Ni, L.M., Xiaodong Zhang, "Locatio-Aware Topology Matching in P2P Systems" Vol.4 INFOCOM 2004. Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies.
- [5] Srivatsa, M., Gedik, B., Ling Liu "Large Scalilg Unstructured Peer-to-Peer Networks with Heterogeneity-Aware Topology and Routing," Vol.17, No 11 IEEE Transactions on Parallel and Distributed Systems.
- [6] Bhushan, Shashi Dave, M., Patel, R.B. "CJM: A Technique to Reduce Network Traffic in P2P Systems," ACE 2010 International Conference on Computing & Processing.
- [7] Yunhao Liu, Lionel M. Ni "Building a Scalable Bipartite P2P Overlay Network," Vol. 18, No. 9, IEEE Transactions on Parrallel and Distributed Systems.
- [8] Xin Sun, Kan Li, Yushu Liu, Yong Tian "SLUP: A Semantic-Based and Location-Aware Unstructured P2P Network," IEEE International Conference on High Performance Computing and Communications.