2MC-Match: A Topology Matching Technique with 2-means Clustering Algorithm in P2P Systems

Guangyu Shi[1], Youshui Long[1], Jian Chen[1], Hao Gong[1] and Hongli Zhang[2]

[1]*Huawei Technologies Co., Ltd*
Shenzhen, 518129, China
{shiguangyu, longyoushui, jchen, haogong}@huawei.com

[2]*Harbin Institute of Technology*
Harbin, 150001, China
zhl@pact518.hit.edu.cn

Abstract

A peer-to-peer (P2P) system is built upon an overlay network whose logical topology is independent of the underlying physical network. A lot of research work has been presented to address this issue, but most results still have some drawbacks, such as complexity or deployment difficulty. In order to alleviate the mismatching problem, we propose a topology matching technique with 2-means clustering algorithm called 2MC-Match, which uses the 2-means clustering algorithm to classify the Internet peers and build efficient "close" cluster. By performing the measured realistic Internet data (China), we show that 2MC-Match outperforms the well-known GNP in both aspects of accuracy and maintenance cost.

Keywords: P2P, mismatch, 2-means clustering

1. Introduction

The Peer-to-Peer (P2P) technology being widely adopted in today's both academic research and practical service providing, has many potential advantages, including high scalability and cost-effectiveness. In a P2P system, each end node provides services to other participating nodes as well as receives services from them. An attractive feature of P2P is that peers do not need to directly interact with the underlying physical network, providing many new opportunities for user-level development and applications. Nevertheless, the mechanism for a peer to randomly choose logical neighbors, without any knowledge about the physical topology, causes a serious topology mismatch between the P2P overlay networks and the physical networks.

The mismatch between physical topologies and logical overlays is a major factor that delays the lookup response time, which is determined by the product of the routing hops and the logical link latency. Mismatch problem also causes a large volume of redundant traffic in inter-domain between the every ISP, which is also an important reason why ISP may prohibit P2P applications. The network administrator of ISP can't control the traffic of P2P applications causing the operating expenditure of Internet rapidly increased [1]. The work in [2] shows that about 75 percent of the query response paths suffer from the topology mismatch problem in 8000 logical peers on 27000 physical nodes.

The issue of mismatch between physical and logical networks in P2P systems has been the focus of intensive research in recent years. According to the methods of solving mismatch, they can be broadly classified into four categories as follows:

The first representative studies to address the mismatch problem are peer coordinate systems such as GNP [3] and Vivaldi [4], which demonstrated the possibility of calculating synthetic coordinates to predict Internet latencies. GNP relies on a small number (5-20) of "landmark" nodes; other nodes choose coordinates based on RTT measurements to the landmarks. The choice of landmarks significantly affects the accuracy of GNP's RTT predictions. Requiring that certain nodes be designated as landmarks may be a burden on P2P systems. Vivaldi is a simple, light-weight algorithm that assigns synthetic coordinates to hosts such that the distance between the coordinates of two hosts accurately predicts the communication latency between the hosts. But Vivaldi is so complex that hard to deploy in nowaday Internet, and its convergence of coordinate system is a slow process.

The second studies approach is a hierarchical location-based node IDs in P2P systems proposed by [5]. Physical locations and network distances are effec-

The research is partially supported by the National Information Security 242 Program of China under Grant No.2006A18.

tively embedded in the node IDs, inducing congruity between overlay topology and physical topology, and thereby improving routing locality. However, what is the influence on the load balance is a particularly important question concerning embedding location prefixes in node IDs.

The clustering algorithm such as Coral [6] takes a third different approach. In Coral system, the peer discovered the router IP cluster by randomly traceroute then chose the first five routers as the close cluster and joined into the cluster. But this scheme is coarse-grained and has difficulty to distinguish relatively close nodes. Why did Coral choose the first five routers as the criterion of close cluster? In the worst case, a majority of nodes could be clustered into one single bin or the clusters could not be established.

Finally, the mechanisms based on measurement have been emerged in the last few years, such as Bamboo [7]. They select proximity neighbor based on monitoring daily DHT operation continuously. The disadvantage of this scheme is that it is hard to measure and learn adequate neighbor information during the peer's limited life time.

The main purpose of existing mismatch solving technique is reducing the lookup response time in P2P system. However, recent studies [8, 9] have shown that a large volume of inter-domain redundant traffic by mismatch already became the serious problem for ISPs. Due to the complexity of P2P networks, a novel mismatch solving technique can not only consider the reduction of lookup response time but also the influence decrease to backbone networks by inter-domain traffic. How to decrease the traffic on backbone networks is another purpose of a novel technique today.

In order to address the limitations of the above cited work, we propose a topology matching technique with 2-means clustering algorithm called 2MC-Match, which uses the 2-means clustering algorithm to classify the Internet end peers and build efficient "close" cluster. 2MC-Match allows end hosts to perform distance predictions in a timely fashion and is highly scalable. Using the 2MC-Match clustering technique, we can optimize the overlay topology by identifying and replacing the mismatched connections.

The rest of this paper is organized as follows: Section 2 describes 2MC-Match in detail. Simulation methodology and performance evaluation are discussed in section 3. And we conclude the work in Section 4.

2. The 2MC-Match Technique

Optimizing inefficient overlay topologies can fundamentally improve P2P search efficiency and decrease redundant traffic. 2MC-Match technique utilizes several basic characteristics of current Internet paradigm, which can efficiently and accurately classify the peers belong to same ISP into a sort of cluster. Furthermore, 2MC-Match is completely decentralized, scalable, reliable and self-organizing. The peers can cluster together separately without the centralized control and any predefined system parameters. In this section, we present two key aspects that need to be addressed in the design of 2MC-Match in the following.

2.1. The characteristics of Internet paradigm

As a rapid development network, the Internet presents two kinds of basic characteristic. On one hand, with many devices randomly join and leave the network, the topology of Internet is dynamic and variable. On the other hand, the routers and switches split the realistic network to different AS domains. The whole topology of Internet presents a strict hierarchy. In P2P systems, most terminal peers exist at the edge of the Internet and communicate with the other peers across by routers which exist at the core of Internet. So during a long period to observe, the core routers are more stabilization than the terminal peers. Made use of this characteristic, we find that the relationship between peer and routers is more effective than the relationship between overlay networks and the underlying physical networks in topology location aware.

In the mean time, the number of connection degree between the ISP routers is obviously less than the whole number of network degree from the world of Internet, which is a considerable characteristic for us to divide the network routers into "near" routers and "remote" routers from the degree lay. Just like the Fig.1, a mass of inner-routers from different AS domains are connected together by the few "star" edge routers. The degree of these "star" edge routers is far less than the core routers.

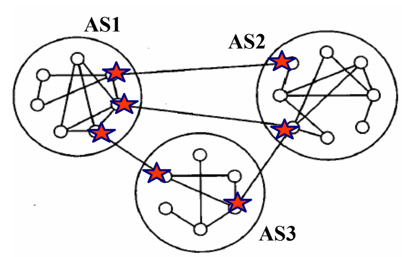


Fig.1 the characteristics of Internet

Utilized above characteristics, 2MC-Match uses the peers' Traceroute result and builds efficient and accurate "close" clusters and "far" clusters in whole Internet topology.

2.2. Traceroute mechanism

According to above mention, in realistic Internet, the stabilized routers are fit for making a mark than dramatic terminal peers. So in 2MC-Match technique, we used the routers' location relationship as the judgment criterion of "near" and "remote" routers. When a peer joining the P2P networks, it randomly picked an Internet IP address and probed it using the traceroute tools. According to the measured data, the peer tracked the return information to a vector data. The data structure of a vector is <IP, Hops, Latency>, where IP is the ip address of every router hop, Hops is the hop number of every middle router in traceroute path and Latency means the link delay between this router and pervious-hop router.

It is worth mentioning that the random choice of probing IP address is related to the result of following clustering. The IP address must be far enough so that the information from the traceroute is adequately efficient to make clustering. So during the traceroute process, the peer chose an IP address with first bit far different from itself, otherwise randomly chose another suitable IP address to probe.

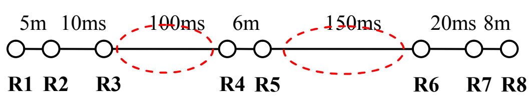


Fig.2 the characteristics of Internet

Fig.2 is an example of path information by R1 traceroute to R8. As is clear from Fig.2, between the R3 and R4, R5 and R6, there are some huge latency leaps than the others. It is possibly means that traceroute message across the different AS domains or different ISP ranges. How to find and utilize these latency leap betweenness to distinguish the "near" and "remote" routers is an important issue for 2MC-Match. 2MC-Match uses the peers' Traceroute result to execute 2-Means Classification, thereafter lets peers to build efficient "close" cluster. 2-means classification algorithm is a special example of k-means classification algorithm [10] with k=2.

2.3. The description of 2MC-Match

2MC-Match uses the peers' Traceroute result to execute 2-Means Classification, thereafter let peers to build efficient "close" cluster. The total approach of 2MC-Match technique is presented in detail as following.

Firstly, when a peer joining the P2P networks, it randomly picked an Internet IP address and probed it using the traceroute tools. The peer tracked the return information to a vector data, with the data structure <IP, Hops, Latency>.

Secondly, the peer used the 2-means classification algorithm to classify the Internet routers based on the Latency attribute in these vectors. The 2-means classification algorithm includes four steps as following.

step1. Peer chose the maximum latency item and minimum latency item in whole vectors as centroids for two initialization sets "first" and "second".

step2. Peer took the latency item in vector to make an absolute distance value with two centroids in turn, and then separately associated the corresponding vector to one vector cluster that has smaller absolute distance value.

step3. Peer calculated the latency mean and variance value of two vector clusters.

step4. If the variance value was larger than the threshold, peer picked two latency mean values as new centroids of "first" and "second" sets, then took a loop from step2 until the two vector clusters achieved the minimum total intra-cluster variance. This produces a separation of the routers into two groups from which the metric to be minimized can be calculated. At last, two "first" and "second" sets confirmed in the end.

```
structure Routerinfo
    ip ipaddress          Δ the ip of routers
    double srcdelay       Δ the delay form this to source
    double hopdelay       Δ the link delay between this router
                             and pervious-hop router

procedure Get2meansCluster(allrouters[1...routenumber])
    firstcenter ← allrouters[min].hopdelay
    secondcenter ← allrouters[max].hopdelay
    oldE ← -1              Δ the previous variance
    E ← 0                  Δ variance
    while (E - oldE) > 0.000001
      firstlist.clear()
      secondlist.clear()
      foreach i ← 0 to allrouters.size
        if ((allrouters[i].hopdelay-firstcenter) <= (secondcenter
         - allrouters[i].hopdelay))
        then firstlist.add(allrouters[i])
        else secondlist.add(allrouters[i])
      end foreach
      firstcenter ← the mean of firstlist.hopdelay
      firstE ← the variance of firstlist.hopdelay
      secondcenter ← the mean of secondlist.hopdelay
      secondE ← the variance of secondlist.hopdelay
      oldE ← E
      E ← firstE + secondE
      ret ← new Vector[2]
    end while
    Ret[0] ← firstlist     Δ the "close" routers clustering
    Ret[1] ← secondlist    Δ the "far" routers clustering
    Retrun Ret[0…1]
```

Fig.3 pseudo code of 2-means classification algorithm

The pseudo code of 2-means classification algorithm in 2MC-Match is shown in Fig.3

Finally, peer chose the router with minimum Hops item in “second” set as a hop threshold. This router and the other routers whose Hops item are larger than the hop threshold all divided into a “remote” router cluster. And then the remaining routers are gathered into another “near” cluster. The peer registers into the P2P overlays with the hop threshold router as Edge Gateway. Through shared vector cluster information such as “near” routers cluster and Edge Gateway, two peers were considered as a close neighbor relationship when their “near” routers cluster both had a same router's IP address at least and then gathered together to form a “close” peer clusters.

In 2MC-Match, each peer clusters has a ClusterId generated by consistent hash when the first 2 peers decided to form the clusters. The ClusterId and its member peer Ids will be PUT into the DHT (or other similar information retrieve means), and the peer can find the other members within the same cluster by DHT GET with its ClusterId remembered during its last online life.

So, in 2MC-Match technique, the judgment criterion for close neighbor peers is considered that whether their “near” routers cluster had same router information, which utilized the information of stabilized routers in P2P systems. The overhead of this technique is the traceroute operation when peer joining the system. On the contrary, the “close” relationship of peers is judged by their Euclidean space in GNP technique. The peer must probe the landmarks to measure the RTT delay as overhead.

2MC-Match allows end nodes to perform distance predictions in a timely fashion and is highly scalable. 2MC-Match builds efficient “close” peer clusters and “far” peer clusters. Using the 2MC-Match clustering technique, we can optimize the overlay topology by identifying and replacing the mismatched connections.

3. Simulations and analysis

In this section, we choose GNP as comparative technique. By performing experiments based on the measured realistic Internet data of China, we get the performance and overhead comparison. Then we discuss some further improvements of 2MC-Match.

3.1. Simulation performance metrics

Both physical topologies and logical overlay topologies which can accurately reflect the topological properties of real networks in each layer are needed in the simulation study. To evaluate the effectiveness of 2MC-Match, we conducted comprehensive simulations. We use the measured realistic Internet data of China, in our simulation.

A well-designed topology matching technique should seek to optimize both efficiency and quality of service. Hence, we mainly focus on three performance metrics: the accuracy of clustering “close” and “far” peers, the average delay in clusters and the average traffic cost.

The accuracy of clustering is one of the important performance metrics with which topology matching algorithms are seriously concerned. The low accuracy of clustering increases the total delay of each cluster and the response time of query in a lookup event. Moreover, the traffic overhead would be heightened when the clustering is not enough accurate. Internet measurement [11] reports that roughly 75%~90% of flows have RTTs less than 200ms. Hence, we define the close peers in realistic networks as those peers whose latency less than 100ms from the source. And the accuracy of clustering is defined as an “AND” operation result between the “close” peers cluster by techniques and the real “close” peers cluster in realistic networks.

The average latency in clusters determines the response time of query in a lookup event. We define the average latency in cluster as average link latency in each “close” peer cluster. The lower average latency obtained, the more efficient the cluster is.

Traffic cost is other parameters that most seriously concern network administrators. Heavy network traffic limits the scalability of P2P networks and is also a reason why a network administrator may prohibit P2P applications. We define the traffic cost as network message used in a topology clustering process of P2P systems, which is mainly a function of consumed network bandwidth and other related expenses.

3.2. Measurement methodology

To evaluate the effectiveness of 2MC-Match, we need an actual topology data which made up of the major ISP routers and most city-level routers in China. We used 9 probing sources to collect the measurement data. The sources are located in Anhui, Gansu, Hainan, Heilongjiang, Hebei, Guangxi, Beijing, Shanghai, Guangzhou, nine main provinces in China mainland. To ensure the effectiveness of probing, we intentionally selected addresses from a large set of IP addresses within China range as target addresses, each /24 subnet we chose 8 IP addresses. The sources triggered 20 processes to send the ICMP traceroute messages to the target nodes and calculated the RTT value by monitoring the response messages. If the sequential 10 hops did not response, the probing process terminated. The difference RTT between previous hop and next hop made an average as link latency.

The measurements from all probing sources were conducted roughly around the same time. The whole measurement lasted three days. As a result, 76116 routers' IP addresses, 128083 links and their latency were obtained. They belong to a diverse set of prefixes originating from ASes across the entire Chinese Internet hierarchy. There were about 33300 edge routers in whole 76116 routers. The measured routers were spread all over China as shown in Fig.4.

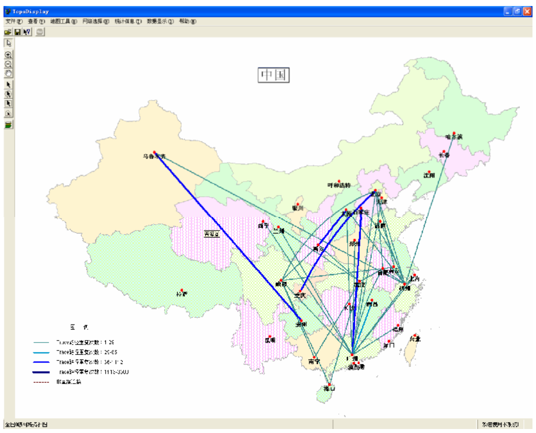


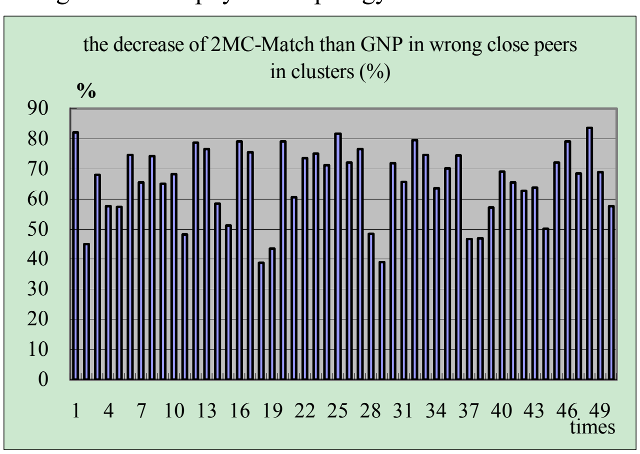
Fig.4 measured realistic Internet topology of China

3.3. Experiment parameters

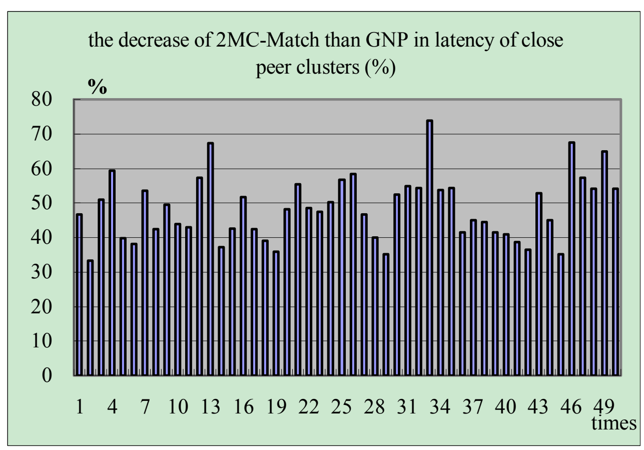
The logical topology represents the overlay topology built on top of the physical topology. We used PlanetSim simulator to generate logical topology, with 10000 peers each round. The peers randomly accessed the edge routes in physical topology. The J-sim simulator provided an interface to connect physical topology and logical topology. On the choice of the landmark number, we considered that in GNP thesis, GNP only set 15 landmarks for the RTT measurement in whole world. So with the China mainland topology, deploying 8 landmarks should be enough to meet the requirement. In our experiments, we set the number of GNP landmarks to 8. And the landmark was chosen randomly in whole nodes of P2P system. For the accuracy each experiment we run 50 rounds, each with different new generated 10000 peers.

3.4. Experiment result

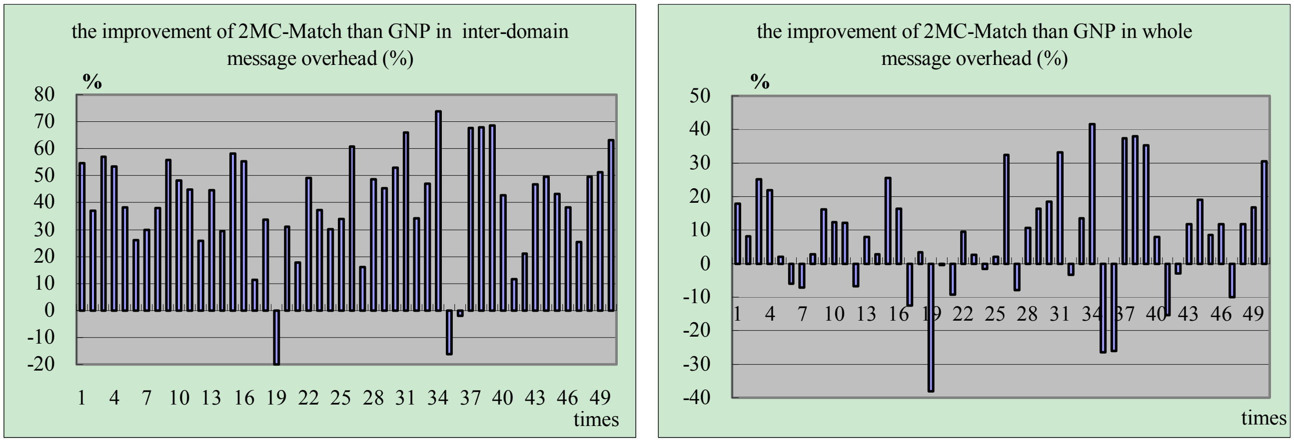
First of all, we need to measure the accuracy of clustering "close" and "far" peers. We choose the "close" peer cluster by topology matching technique to make a comparison with the real "close" peer with latency less than 100ms each other in the realistic network. From Fig.5(a), we find that the number of wrong close peers in close cluster decreases by 65% in 2MC-Match than that of in GNP. It is mainly because in GNP, the landmarks as probing targets are chosen randomly. Every peer calculates own GNP distance and clusters together by probing the landmark, so the number and location of landmarks directly determine the performance of GNP schemes. But in 2MC-Match technique, there is no landmark and the probing target is random. It avoids that the limitation of landmark in peer clustering. Furthermore, through efficiently clustering, 2MC-Match can reduce a large volume of inter-domain redundant traffic than GNP.



(a) the comparison of wrong close peer in clusters



(b) the comparison in latency of close peer clusters



(c) the comparison of inter-domain message overhead

(d) the comparison in whole message overhead

Fig.5 the performance comparison of 2MC-Match and GNP in measured realistic Internet data of China

The accuracy of clustering brings forth the improvement of latency immediately. Because of the "close" peers' number increased and accuracy of "close" peer clusters improved, the average latency in cluster decreases by 48% approximately from the Fig.5(b). The lower average latency means quicker P2P interactions. As is clear from the Fig.5(c) and Fig.5(d), 2MC-Match achieves another favorable performance metric, the maintenance traffic overhead. 2MC-Match consumes less maintenance bandwidth overhead of clusters construction than GNP both in the inter-domain and whole network level. The reason is that in 2MC-Match, each peer just probes only one peer. On the contrary, the peer must measure the delay of every landmark in GNP scheme, so that the number of landmark determines the maintenance overhead of scheme. As one can see in Fig.5(c) and Fig.5(d), 2MC-Match decreases the average maintenance overhead in inter-domain by 40%, in whole by 8%.

4. Conclusion

In this paper we presented a simple and effective peer clustering technique called 2MC-Match that uses the traceroute tools and 2-means clustering to classify the Internet peers. 2MC-Match allows end hosts to perform distance predictions in a timely fashion and is highly scalable. Using the 2MC-Match clustering technique, we can optimize the overlay topology by identifying and replacing the mismatched connections. 2MC-Match technique utilizes several basic characteristics of current Internet paradigm, which can efficiently and accurately classify the peers belong to same ISP into a sort of cluster. Furthermore, 2MC-Match is completely decentralized, scalable, reliable and self-organizing. The peers can cluster together separately without the centralized control and any predefined system parameters. By performing experiments using the measured realistic Internet data of China, we show that 2MC-Match outperforms the existing state of the art system GNP scheme in both aspects of accuracy and maintenance cost.

5. References

[1] Mudhakar Srivatsa, Bugra Gedik, and Ling Liu, "Large Scaling Unstructured Peer-to-Peer Networks with Heterogeneity-Aware Topology and Routing," IEEE Transactions on Parallel and Distributed Systems, 2006, 17(11), pp.1277-1293.

[2] Yunhao Liu, Li Xiao and Lionel M. Ni, "Building a Scalable Bipartite P2P Overlay Network," IEEE Transactions on Parallel and Distributed Systems, 2007, 18(9), pp.1296-1306.

[3] T. S. Eugene Ng and Hui Zhang, "Predicting Internet Network Distance with Coordinates-Based Approaches," Proc. IEEE INFOCOM, 2002.

[4] F. Dabek, R. Cox, F. Kaashoek, and R. Morris, "Vivaldi: A Decentralized Network Coordinate System," Proc. ACM SIGCOMM, 2004.

[5] Shuheng Zhou, Gregory R. Ganger, and Peter Steenkiste, "Location-based Node IDs: Enabling Explicit Locality in DHTs," Technical Report CMU-CS-03-171, 2003.

[6] Michael J. Freedman and David Mazieres, "Sloppy Hashing and Self-organizing Clusters," Proc. IPTPS, 2003.

[7] Sean Rhea, Dennis Geels, Timothy Roscoe, and John Kubiatowicz, "Handling Churn in a DHT," Proc. USENIX, 2004.

[8] Vinay Aggarwal, Anja Feldmann and Christian Scheideler, "Can ISPs and P2P Systems Cooperate for Improved Performance?" ACM SIGCOMM Computer Communications Review, 2007, 37(3), pp.29-40.

[9] G. Shen, Y. Wang, Y. Xiong, B. Zhao, and Z. Zhang, "HPTP: Relieving the Tension between ISPs and P2P," Proc. IPTPS, 2007.

[10] Vance Faber, "Clustering and the Continuous k-Means Algorithm," Los Alamos Science, 1994, 22, pp.138-144.

[11] H. Jiang and C. Dovrolis, "Passive Estimation of TCP Round-Trip Times," ACM Computer Communications Review, 2002, 32(3), pp.75-88.