Global Journals LaTeX JournalKaleidoscope™
Artificial Intelligence formulated this projection for compatibility purposes from the original article published at Global Journals. However, this technology is currently in beta. *Therefore, kindly ignore odd layouts, missed formulae, text, tables, or figures.*

A Square Root Topologys to Find Unstructured Peer-To-Peer Networks

Yadma Srinivas Reddy[1]

[1] Vardhaman College of Engineering

Received: 7 December 2012 Accepted: 5 January 2013 Published: 15 January 2013

Abstract

Unstructured peer-to-peer file sharing networks are very popular in the market. Which they introduce Large network traffic. The resultant networks may not perform search efficiently and effectively because used overlay topology formation algorithms are creating unstructured P2P networks are not performs guarantees. In this paper, we choosen the square-root topology, and show that this topology. Which improves routing performance compared to power-law networks. In the square-root topology shows that this topology is optimal for random walk searches. A power-law topology for other types of search techniques besides random walks. Then we interoduced a decentralized algorithm for forming a square-root topology, its effectiveness in constructing efficient networks using both simulations and experiments with our prototype. Results show that the square-root topology can provide a good and the best performance and improvement over power-law topologies and other topology types.

Index terms— peer to peer connections, unstructured overlay networks, search, random walks.

1 Introduction

EER-TO-PEER networks have been widely used in the internet, and they provide more services like file sharing, information gathering, media streaming. P2P applications are more popular because they firstly gives low entry barriers and self-scaling. P2P applications are dominating 20 percent of Internet traffic. Object search is the big task in P2P applications. Gnutella is a popular P2P search protocol in the market. Gnutella networks are unstructured, The peers participating in networks connect to one another randomly, peers search objects in the networks By using message flooding. To flood a message, peer broadcasts the message to its neighbouring peer. The broadcast message is associated with a positive integer time-to-live value. By receiving a message, the peer decreases the TTL value with the message by 1 and then stops the message with the updated TTL value to its neighbours, except the one sending the message to j, if the TTL value remains positive. Aside from forwarding the message to the neighbours, j searches its local store to see if it can provide the objects requested by peer i. if j has the requested objects and is ready to send them, then j directly sends i the objects or returns the objects to the overlay path where the query message moves from i to j. The Gnutella search performance is like unstructured P2P networks. used orthogonal techniques is for improving search performance in unstructured P2P networks. replications, super peer architectures and overlay topologies, among others. we use the square-root topology technique for unstructured P2P networks, to enhance search efficiency and effectiveness. The P2P network to minimize the overlay path length between any two peers to dicrease the query response time. The probability of peer j being the neighbor of peer i increasing if j shares more common interests with i. we first observe the existing P2P file sharing networks shows the power-law file sharing process. this proposal has the following unique features.

In a constant probability, the search hop count between any two nodes is Oðlnc1 NÞ,where 1< c1<2 is a small constant, and N is the number of active peers participating in the network.

In a constant probability of approximately 100 percent, peers on the search path from the querying peer to the destination peer progressively and effectively exploit their similarity.

1 INTRODUCTION

Whereas some solutions require centralized servers to help to set the system, our proposal needs no centralized servers to participate . our solution is mathematically provable and gives performance guarantees. we using a search protocol to take advantage of the peer similarity exhibited by overlay network. one extra finding in our performance analysis shows that P2P networks. the search path length is OðN c2 Þ (where $0 < c2 < 1$) if any peer i on the path has to search another peer j, which is to be same as to the destination peer d than i, to receive and forward the query toward d. from having a best performance analysis, our theoretical analysis is existed in simulations. we compare our proposal with two representative distributed algorithms. With our similarityaware search protocol, that the overlay networks that shows the similarity of participating peers can considerably dicrease the query traffic and the search protocol based on blind flooding. Our proposal: the square-root topology

A peer-to-peer network with N peers. Each peer k in the network has degree dk. The total degree in the network is D, where D =_Nk=1 dk.

Equivalently, the total number of connections in the network is D/2. We used the square-root topology as a topology where the degree of each peer is proportional to the square root of the popularity of the peer's content. if we define gk as the proportion of searches submitted to the system that are satisfied by content at peer k, then a square-root topology has dk ? ? gk for all k. consider a user submits a search s that is existed by the by content at a particular peer k. Until the search is processed by the network, we do not know which peer k is. How many hops will the search message take before it arrives at k. The expected length of the random walk depends on the degree of k: Lemma 1. If the network is connected and nonbipartite, then the expected number of hops for search s to reach peer k is D/dk.

Where the probability of transition from state i to state j depends only on i and j, and not on any other history about the process. The states of the Markov chain are the peers in the system, and 1 ? i, j ? N. Associated with a Markov chain is a transition matrix T that shows the probability that a transition occurs from a state i to another state j. this transition probability is the probability that a search message that is at peer i is next sends to peer j. With random walks, the transition probability from peer i to peer j is 1/di if i and j are neighbors, and zero. it depends only on the node degrees, and not on the structure. the expected length of a walk does not depend on which peers are connected to which other peers. This property exicutes from the fact that the Markov chain converges to the same stationary distribution of which vertices are connected.

Thismodel shows peers forward search messages to a randomly chosen peers, even if that search message has just come from that neighbor or has already visited this neighbor. This process simplifies the Markov chain analysis. Already used process for random walks have noted that avoiding previously visited peers can improve the efficiency of walks, and we state this possibility in simulation results in the next section. Using the transition matrix, we can calculate the probability that a search message is at a given peer at a given point in time. First, we define an N element vector V0, called the initial distribution vector, the kth entry in V represents the probability that a random walk search starts at peer k. The entries of V sum to 1. Given T and V0, we can calculate V1, where the kth entry represents the probability of the search being at peer k after one hop, as V1 = TV0. In general, the vector Vm, representing the probabilities that a search is at a given peer after m hops, is recursively defined as Vm = TVm?1. Under the conditions of the lemma, Vm converges to distribution vector Vs, representing the probability that a random walk search visits a given peer at a particular point in time. It shown that the kth entry of Vs is dk/D. In the steady state, the probability that a search message is at a given peer k is dk/D.

The search routing as a series of experiments, by choosing a random peer k from the population of N peers with probability dk/D. The successful experiment occurs when a search chooses a peer with matching content. The expected number of experiments before the search message successfully reaches a particular peer k is a geometric random variable with expected value 1 dk/D = D dk. This is the result comes by Lemma 1.

If a given search requires D/dk hops to reach peer k, we assume that a search will be satisfied by a single peer. We define Gk to be the probability that peer k is the goal peer, gk ? 0 and _N k=1 gk = 1. The gk vary from peer to peer. The proportion of searches seeking peer k is gk, The expected number of hops that will be taken by peers seeking peer k is D/dk, The expected number of hops taken by searches is:H = _N k=1 gk ? D dk**(1)**

It turns out that H is minimized when the degree of a peer is proportional to the square root of the popularity of the documents at that peer. This is the square-root topology.

Theorem 1:H is minimized when dk = D ? gk _N i=1 ? gi**(2)**

Proof:

We use the method of Lagrange multipliers to minimize equation (1). Recall the constraint that all degrees dk sum to D, the constraint for our optimization problem isf = (_N k=1 dk) ? D = 0.

We must find a Lagrange multiplier ? that satisfies ?H = ??f (where ? is the gradient operator).

First, treating the gk values as constants,?H =_N k=1?D?gk? d ?2 K? u?k (**3**)

Where u?k is a unit vector. Next,??f = ? _N k=1 u?k = _N k=1 ?u?k**(4)**

Because ?H = ??f, we can set each term in the summation of equation (3) equal to the corresponding term of the summation of equation (4 If we change the dummy variable of the summation in equation (**??**) from k to i, and substitute back into equation(1), we get equation (2). Theorem 1 shows that the square-root topology is the optimal topology over a large network, does not impact performance substituting equation (2) into equation (1) eliminates D. any value of D that ensures the network is connected is sufficient. Result shows more of which

peers are connected to which other peers, because of the properties of the stationary distribution of Markov chains. Peer degrees must be integer values, Therefore, the optimal peer degrees must be calculated by rounding the value calculated in equation (2).

2 III.

Experimental Results on the Square-Root Topology

In analysis of the square-root topology is based on an idealized model of searches and content. peer-topeer systems are less idealized, searches may match content at multiple peers. In this we present simulation results to get the performance of a square-root topology. We use simulation because we wish to exbit the performance of large networks and it is difficult to deploy that many live peers for research on the Internet. Our first metric is to count the total number of messages sent under each search method. Searches terminate when enough results are found, where enough is defined as a user specified goal number of results G.

3 The results show:

? Random walks perform best on the square-root topology, requiring up to 45 percent fewer messages than in a power-law topology. The square-root topology also results in up to 50 percent less search latency than power-law networks, even when multiple random walks are started in parallel. ? The square-root topology is the best topology when replication is used, and the combination of squareroot topology and replication provides higher efficiency than technique alone. ? Other search techniques based on random walks, such as biased high-degree, biased towards results or fewest result hop neighbors, and random walks with state keeping performed best on the squareroot topology, decreasing the number of messages sent by as much as 52 percent compared to a power-law topology.

? The square-root topology shown better than other topology structures, including a constant degree network, and a topology with peer degrees directly proportional to peer popularity. In super-peer networks the square-root was the best topology for connecting the super-peers. we first sets our experimental setup, and then present our results.

4 a) Experimental Setup

Our results were exhibited by using a discreteevent peer-to-peer simulator. In this simulator models individual peers, documents and queries, also the topology of the peer-to-peer overlay. Searches are send to individual peers, and then walk around the network according to the routing algorithm. Square-root topology is based on the popularity of documents stored at different peers, it is important to accurately model the number of queries that match each document, and the peers at which each document is stored. It is difficult to gather complete query, document and location data for tens of thousands of real peers. Therefore, we used the content model described in, which is based on a trace of real queries and documents, and more accurately describes real systems than simple uniform or Zipfian distributions. We downloaded text web pages from 1,000 real web sites, and evaluated keyword queries against the web pages. Then we generated 20,000 synthetic queries matching 631,320 synthetic documents, stored at 20,000 peers, such that the statistical properties of our synthetic content model matched those of the real trace. The resulting content model allowed us to simulate a network of 20,000 peers. In this simulation, we submitted random queries chosen from the set of 20,000 to produce a total of 100,000 query submissions. we describe the details of this method of generating synthetic documents and queries, and provide experimental evidence that the content model, though synthetic, results in highly accurate simulation results. The synthetic model retains an accurate distribution of the popularity of peer content, which is critical for the construction of the square-root topology. We conducted an experiment to examine the performance of random walk searches in different topologies. This queries matched documents stored at different peers, and had a goal G = 10 results. We compared three different topologies.

? A square-root topology, generated by assigning a degree to each peer based on equation (2), and then creating links between randomly chosen pairs of peers based on the assigned degrees. ? A low-skew power-law topology, generated using the PLOD algorithm. In this network, ? = 0.58. ? A high-skew power-law topology, generated using the PLOD algorithm, with ? = 0.74.

Random walks in the square-root topology require 8,940 messages per search, 26 percent less. than random walks in the low-skew power-law topology and 45 percent less than random walks in the high-skew power-law topology. In the power-law topologies, searches tend toward high degree peers, even if the walk is truly random and not explicitly directed to high degree peers. These high degree peers also have the most popular content, Result is that searches have a low probability of going to the peer with matching content, and the number of hops and thus messages increases. If the power-law distribution is more skewed, then the probability that searches will congregate at the wrong peers is higher and the total number of messages are necessary to get to the right peers increases. Even though random walks perform best in the square-root topology, a large number of messages need to be sent. the result is a significant improvement over traditional Gnutella style search, flooding in a high-skew power-law network, with a TTL of five in order to find at least ten results on average, requires 17,700 messages per search. The above results are for simple, unoptimized random walks. Adding optimizations such as proactive replication or neighbor indexing reduces the cost of a random walk search, and results for these techniques show

that the square-root topology is still best. Another issue with random walks is that the search latency is high, as queries may have to walk many hops before finding content. To deal with this, Lv et al propose creating multiple, parallel random walks for each search. Since the network processes these walks in parallel, the result is reduced search latency. We ran experiments where we created 2,10, 15, 20, 30, and 100 parallel random walks for each search, and measured search latency.

Table **??** : Parallel random walks: search latency (ticks)

These results are shown in Table **??**.The squareroot topology provided the lowest search latency, regardless of the number of parallel walks that were generated. The improvement for the square-root topology was consistently 27 percent compared to the low-skew power-law topology, and 50 percent compared to the high-skew power-law topology. Even when searches are walking in parallel, the square root topology helps those search walks quickly arrive at the peers with the right content.

5 c) Proactive Replication

The square-root topology is complementary to the square-root replication. It is feasible to proactively replicate content, the square-root replication specifies that the number of copies made of content should be proportional to the square root of the popularity of the content. The square-root topology can be used whether or not proactive replication is used, the combination of the two techniques can provide significant performance benefits. We conducted an experiment where Replicated content according to the square-root replication. Then we connected peers in the square-root, high-skew power-law, and low-skew power-law topologies, and states the performance of random walk searches. Again, G= 10. As expected, proactive replication provided better performance than no replication. Proactive replication performs best with the square-root topology, requiring only 2,830 messages per search, 42 percent less than in the low-skew power-law network and 56 percent less than in the high-skew power-law network. replication makes more copies of the documents that a search will match, while the square-root topology makes it easier for the search to get to the peers where the documents are stored. The combination of the two techniques provides more efficiency than either technique alone. For example, the square root topology with proactive replication required 68 percent fewer messages than the square root topology without replication.

6 d) Other search walk techniques

We examined the performance of other walkbased techniques on different topologies. We compared three other techniques based on random walks: In each case, ties are broken randomly. For the biased high degree technique, we examined both neighbour indexing and no neighbour indexing. Although describes several ways to route searches in addition to most results and fewest result hops, these two techniques represent the best that the result hops requires the least bandwidth, while most results has the best chance of finding the requested number of matching documents. The square-root topology is best. The most improvement is seen with the biased high degree technique, where the improvement on going from the high-skew power-law topology to the squareroot topology is 52 percent. Large improvements are achieved with the fewest result hops technique and most results. The smallest improvement observed was for the biased high degree technique with neighbor indexing. square-root topology offers a 16 percent decrease in messages compared to the lowskew power-law topology. The square-root topology provides the best performance, even with the extremely efficient biased high degree/neighbor indexing combination. The square-root topology can be used even when neighbor indexing is not feasible. The combination of square-root topology, square-root replication and biased high degree walking with neighbor indexing provides even better performance. Our results indicate that this approach is extremely efficient, requiring only 248 messages per search on average. The square root topology is better than the power law topology when square-root replication and neighbor indexing are used. Using all three techniques together results in a searching mechanism that contacts less than 2 percent of the systems peers on average while still finding sufficient results. The results so far assume statekeeping, where peers keep state about where the search has been. peers can avoid forwarding searches to neighbours that the search has already visited. The results demonstrate that the square-root topology is better than power-law top logies, whether or not statekeeping is used.

7 e) Other Topologies

We also tested the square-root topology in comparison to several other network structures. We compared against two simple structures.

? Constant-degree topology: every peer has the same number of neighbors. In our simulations, each peer had five neighbors. ? Proportional topology: every peer had a degree proportional to their popularity gk.

Our results show that the square-root topology is best, requiring 10 percent fewer messages than the constant degree network, and 7 percent fewer messages than the proportional topology. Although the improvement is smaller than when comparing the square-root topology to power-law topologies, these results again demonstrate that the square-root topology is best. The cost of maintaining the square-root topology is low, as we discuss in Section 4, requiring easily obtainable local information. It clearly makes sense to use the square-root topology instead of constant degree or proportional topologies. A widely used topology in many systems is the super-peer topology. In this topology, a fraction of the peers serve as super-peers, aggregating content information from several leaf pears. Then, searches only need to be sent to super-peers. They are connected using a normal

unstructured topology. We ran simulations using a standard superpeer topology, in which searches are flooded to super peers. We compared this standard topology to a super peer topology that used the square-root topology and random walks between super peers. The results indicate a significant improvement using our techniques the square root super peer network required 54 percent fewer messages than a standard super-peer network.

IV.

8 Conclusions

We have presented the square-root topology, and shown that implementing a protocol that causes the network to converge to the square root topology, rather than a power-law topology, can provide significant performance improveents for peer-to-peer searches. In the square-root topology, the degree of each peer is proportional to the square root of the popularity of the content at the peer. Our analysis shows that the squareroot topology is optimal in the number of hops required for simple random walk searches. We also present simulation results which demonstrate that the squareroot topology is better than power-law topologies for other peer-to-peer search techniques. we presented an algorithm for constructing the square-root topology using purely local information. Each peer estimates its ideal degree by tracking how many queries match its content, and then adds or drops connections to achieve its estimated ideal degree. Results from simulations and our prototype show that this locally adaptive algorithm quickly converges to a globally efficient square-root topology. Our results show that the combination of an optimized topology and efficient search mechanisms provides high performance in unstructured peer-to-peer networks. [1]

Figure 1:

[1] EA Square Root Topologys to Find Unstructured Peer-To-Peer Networks

1

Parameter	Value
Number of peers	20,000
Documents	631,320
Queries submitted	100,000
Goal number of results	10
Average links per peer	4
Minimum links per peer	1

Figure 2: Table 1 :

[Ratna et al.] , S Ratna , P Francis , M Handley , R Karp , S Shenker , : A Scalable , Network .

[Kalnis et al.] *An adaptive peer-to-peer network for distributed caching of OLAP results*, P Kalnis , W Ng , B Ooi , D Papadias , Tan .

[Chawathe et al.] Y Chawathe , S Ratnasamy , L Breslau , N Lanham , Shenker . *S: Making Gnutella-like P2P systems scalable*,

[Yang and Garcia-Molina] *Designing a super-peer network*, B Yang , H Garcia-Molina .

[Stoica et al.] *H: Chord: A scalable peer-to-peer lookup service for internet applications*, I Stoica , R Morris , D Karger , M F Kaashoek , Balakrishnan .

[Loo et al.] *I: Enhancing P2P file-sharing with an Internet-scale query processor*, B Loo , J Hellerstein , R Huebsch , S Shenker , Stoica .

[Loo et al.] *J: Enhancing P2P file-sharing with an Internet scale query processor*, B Loo , R Huebsch , I Stoica , Hellerstein .

[Rowstron and Druschel] *P: Pastry: Scalable, decentralized object location and routing for largescale peer-to-peer systems*, A Rowstron , Druschel .