Scale-Free Networks to Search in Unstructured Peer-To-Peer Networks

[1]Yadma srinivas reddy, [2] Mr.D. Raman,
1 (M.Tech, Cse, Vardhaman college of engineering)
2 (Associate professor , CSE, Vardhaman college if engineering)

Abstract: *Unstructured peer-to-peer file sharing networks are very popular in the market. They introduce network traffic. The resultant networks are not perform search efficiently because existing overlay topology formation algorithms are creates unstructured P2P networks without performance guarantees. In this paper, We compares structured and unstructured overlays, Showing through analytical and simulation results how an unstructured solution relying on a scale-free topology is an effective option to deploy for offering services based on equivalent servants. Based on this result, The EQUivalent servAnt locaTOR (EQUATOR) architecture, which overcomes the issues related to the deployment of a scale-free topology for service location in a real network, mainly due to the static nature of the ideal scale-free construction algorithm and the lack of a global knowledge of the participating peers. Simulation results confirmed the effectiveness of EQUATOR, showing how it offers good lookup performance in combination with low message overhead and high resiliency to node churn and failures. Some possible future works are introduced and are related to some complementary issues ranging from the proximity-aware selection of servants to the introduction of proper incentives to encourage nodes to join the EQUATOR overlay and offer their resources.*

Keywords: *peer to peer systems, Random Networks, Scale-Free Networks, Communication Networks, Socio-Technical Networks, Diffusion.*

I.   Introduction

PEER-TO-PEER networks have been widely Spred out in the Internet, and they provide different services such as file sharing, information retrieval, media streaming. P2P applications are popular because they primarily provide low entry barriers and self-scaling. Object search is an important task in P2P applications. Gnutella is a popular P2P search protocol in the mass market. Because Gnutella networks are unstructured, Peers participating in networks connect to one another randomly, peers search objects in the networks through message flooding. To flood a message, an inquiry peer broadcasts the message to its neighbours. The broadcast message is associated with a positive integer time-to-live (TTL) value. Upon receiving a message, the peer decreases the TTL value associated with the message by 1 and then relays 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. Conceptually, if j has the requested objects and is ready to supply them, then j either directly sends i the objects or returns the objects to the overlay path where the query message traverses from i to j. The search performance in Gnutella is like unstructured P2P networks. Existing orthogonal techniques in the literature for improving search performance in unstructured P2P networks include indexing, replications, super peer architectures and overlay topologies, among others. In this paper, we primarily study the square-root topology technique for unstructured P2P networks, aiming to enhance search efficiency and effectiveness.

The P2P network minimize the overlay path length between any two peers to reduce 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 that existing P2P file sharing networks exhibit the power-law file sharing pattern. our proposal has the following unique features:

In a constant probability, the search hop count between any two nodes is $O(\ln c_1 N)$,where $1 < c_1 < 2$ is a small constant, and N is the number of active peers participating in the network.

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

Whereas some solutions require centralized servers to help organize the system, our proposal needs no centralized servers to participate in. our solution is mathematically provable and provides performance guarantees. Moreover, we suggest a search protocol to take advantage of the peer similarity exhibited by our proposed overlay network. Noticeably, one extra finding in our performance analysis reveals that semantic P2P

networks, which not only attempt to cluster semantic closest peers but also suggest to interconnect peers selected uniformly at random, cannot work well by efficiently exploiting the similarity of peers for searches. Particularly, if any peer i on the path has to find another peer j, which is more similar to the destination peer d than i, to receive and forward the query toward d. Aside from having a rigorous performance analysis, our theoretical analysis is validated in simulations. we compare our proposal with two representative distributed algorithms. With our similarity aware search protocol, we conclude that the overlay networks that exploit the similarity of participating peers can considerably reduce the query traffic than the search protocol based on blind flooding.

II.   Our Proposal: Scale-Free Networks:

Network Models
          ER- and scale-free networks as well as their main characteristics and the processes by which they are created. ER-networks, because of their diminishing importance for modelling real world networks, are only briefly introduced to show their important differences from scale-free networks. The basics of network analysis are discussed first to enable a better understanding.

2.1 Basics of Network Analyses
          In an undirected network graph, the degree of a node is defined as the number of edges it possesses. If all the nodes of a network that share the same degree are counted and the results are sorted by increasing degree a function like that is most likely to occur. There are no nodes with degree zero, because these nodes would not be connected to the network. The network exhibits a majority of nodes with degree one to three and rather fewer nodes with a degree greater than seven. The degree distribution of networks does not have to be continuous. If one node of the network were randomly chosen, the probability of obtaining a node with only one to three edges is much higher than that of obtaining a node with a degree higher than seven. Therefore it is possible to define a probability distribution function that returns the probability that node v has j edges within network N. The concept of the degree distribution of a network has important consequences for the properties of a network and will be frequently used in the paragraphs that follow. It will be seen that deviation from the normal distribution will lead to new results in terms of diffusion within networks. Clustering within networks is another important factor when analyzing networks. It is interesting to know how well nodes are interconnected within a specified area of the network. Using the ratio between existing and possible relations, a clustering coefficient may be computed.
          If a node has z nearest neighbours, a maximum of edges is possible between them. Watts and Strogatz defined the clustering coefficient for node v as the ratio of the number I of existing nodes to the possible number of edges between the direct neighbours of node v As for the degree distribution, the clustering coefficient plays an important role when analyzing networks in terms of important properties like diffusion, which
The path length between two nodes of a network is defined as the number of edges between them. The minimal path length is the shortest path between two nodes. The average path length is the average of all the minimal path lengths between all pairs of nodes in a network.

2.2 ER-Networks
          Since the seminal paper of Erdo and Renyi in 1959, Random network theory dominated scientific thinking [Bara03]. Real world networks had been thought to be too complex to understand and therefore held to be random. In the absence of other well-understood network models, random
Networks were widely used when modelling networks. The process of creating an ER-network depends on probability p. For a network with n nodes each possible pair of distinct nodes are connected with an edge with probability p.
          An ER-network has the property that the majority of nodes have a degree that is close to the average degree of the overall network and that there is not much deviation from the average below and above it. It has been shown that the distribution of links follows a Poisson distribution. Knowing the degree distribution, the average path length, and the clustering coefficient of ER-networks, it is feasible to analyze their different behaviour compared to scale-free networks. ER-networks are still used in some types of models. The following chapters will show that ER-networks may not be applic Able for every purpose, because they lack Some of the properties of other classes of Networks.

2.3 Scale-Free Networks
          When Albert et al. started to map the internet in 1999, they did not know that they were about to influence network research in a sustained way. Because of the Diverse interests of every internet user and The gigantic number of web pages, the linkages between web pages were thought to be randomly linked as a random network. The results of their study have disagreed With this expectation in a surprising way. Only a few pages

have the majority of links, whereas most pages are only very sparsely connected. More than 80% of all Pages visited have 4 links or less, only 0.01% of the pages are linked to more than 1000 other pages. If the nodes with one, two, three, etc. connections are counted, and the numbers are plotted into a chart as in the middle of the distribution of edges becomes visible. There are many nodes with only a few links and a few nodes with a large number of links. If the distribution of edges is plotted in a logarithmic chart, as in right, the power-law nature of the distribution appears in a straight curve with a slope of G from equation. The difference from ER-networks now becomes evident. The probability distribution function of the degree j of scale-free networks is described with j > 0 and g > 0, with g called the scale-free exponent. The term scale-free is a mathematical expression which stands for power-law distributions as in equation. The power-law distributions belong to the Class of leptokurtic or fat-tailed distributions. They deviate from the Poisson distribution in two ways: they have higher peaks and fat tails.

III.   Experimental Results for Equator

This section presents some simulation results on the EQUATOR architecture. We first validate our overlay construction algorithm, which we show to result in a scale-free topology. We also show how EQUATOR is comparable to the ideal Barabasi-Albert network in terms of lookup performance. We then elaborate on the system parameters, also focusing on the lookup and advertisement overhead at nodes. Finally, we investigate the behaviour of our solution in different scenarios triggered by different kinds of peers.

3.1 Structured overlays

We first investigate the possibility to deploy a structured overlay based on a general DHT, as it has been proposed in [19] for the P2PSIP architecture. Since in our scenario all peers provide the same functionality, the number of copies predominates over the number of distinct

Services and therefore the ability of DHTs to locate a specific resource are of little help. Therefore, [19] proposes to use the DHT in a more clever way: queries are performed by randomly selecting a target key and then moving in the overlay to reach this target. Since it does not cause further complexity and possibly improves the system performance, we introduce an additional feature to this querying mechanism: during the lookup process, any node encountered along the path is checked for availability and can be selected as a servant for the querying user. Notice that this operating mode makes the approach independent of the adopted DHT. In fact, only the overlay topology is of interest in our context. In other words, we adopt the topology of a generic DHT, with a fixed number of neighbours for each node, but we use a different routing mechanism. This solution will be however referred to as DHT in the rest of the paper. The idea of using a DHT for our scenario of equivalent servants is especially interesting in case a DHT has to be implemented anyway for some other services. For example, P2PSIP already uses a structured overlay to index all possible targets of a multimedia communication, i.e., all the user agents registered in the SIP domain. Using the same DHT to locate, if necessary, a relay node to support the communication may be a considerable advantage for that application, which needs to maintain only one overlay structure that can be used for both functions.

3.2 Unstructured overlays

An efficient unstructured overlay is characterized by high lookup performance and small amount of traffic required to maintain the overlay. Both parameters are influenced by the topology and the operating principles of the overlay. This section elaborates on these aspects in the context of services based on equivalent servants, proposing to adopt a scale-free topology and motivating this choice. An interesting lookup solution that avoids the deleterious traffic overhead generated by flooding-based queries is the adoption of a service lookup based on random walks encompassing a bounded number of nodes. Within this technique, the service request is forwarded, at each node, to a peer randomly selected among its neighbours. If the encountered node is available or knows an available servant, the procedure terminates. The knowledge of nodes can be improved through proper advertisement messages containing the node itself and other participating peers, thus implementing a so called epidemic dissemination algorithm.

The effectiveness of random walks depends on the overlay topology adopted in the system. Among other possibilities, a scale-free topology may offer interesting features. In a scale free network, the node degree distribution follows a power-law $P(n) = cn^r$, where $P(n)$ is the probability that a node has n connections and c is a normalization factor. Hence, only few nodes have a high degree, i.e., are aware of the existence of a large number of participating peers. The idea is that directing random walks toward hubs means looking for the service where there is a great knowledge of servants. This ensures high lookup performance with respect to an overlay based on a balanced Degree distribution where service requests are randomly distributed among peers.

This result derives from a well-known property of queuing systems, which says that a unique M/G/k/k queuing system servicing an arrival process with rate performs better than k separated M/G/1/1 systems each one servicing an arrival process with rate k. In essence, concentrating the traffic on some nodes that have a deep

knowledge of the network provides better performance than accurately distributing the requests among all nodes, as random solutions try to do. Effectiveness of random walks in scale free networks due to the greater knowledge of resources available at the hubs. In order to achieve high lookup performance, hubs should have a deep knowledge about the other participating peers: the greater the number of peers known by a given node, the higher the probability for a user to find an available servant in a short time.

Since the epidemic dissemination is based on flooding, the overlay topology has a deep impact not only on peers known by each node, but also on the resulting network efficiency. In fact, the greater the average path length between nodes, the higher the depth of the flooding that is needed for an adequate spread of the information, which may cause an unsustainable load on the network. The scale-free topology also ensures a good efficiency of epidemic dissemination algorithms as exhibits a small average path length. In essence, a large number of advertisement messages reach the hubs even with a small dissemination depth (namely, the number of hops encompassed by advertisement messages before elapsing) and a small out-degree. Another interesting feature of scale-free networks is that they can scale to an arbitrarily large network size without Modifying the degree distribution of nodes, which continues to follow the same law? This ensures that new hubs are automatically created when the network size grows, therefore maintaining the above described properties.

In essence, scale-free networks potentially combine the advantages of centralized indexing and Totally distributed solutions. One of the most popular mechanisms to build a scale-free network was proposed by Barabasi and Albert and for this reason is referred to as Barabasi-Albert model. Let m denote the out-degree of a node and d denote its in-degree. The Barabasi-Albert model requires a set of m0 nodes to be already in the system at the beginning of the process. Then, each entering node connects to m existing nodes, chosen proportionally to their popularity. This process is known as Preferential attachment. This network formation algorithm results in a scale free network characterized by a node degree distribution $P(n) = cn-3$ and an average path length which behaves The Barabasi-Albert model is used as a reference in the rest of the paper. Although in general $P(n) = P(m + d)$, in this case we are interested in the in degree of a node as it represents its popularity, i.e., it counts the number of nodes that send their advertisements to it. Thereby, without losing in significance, we consider $P(n) = P(d)$ i.e., the distribution of the in-degree of nodes in the following. The Barabási-Albert model is an ideal network formation algorithm that requires a global knowledge of the existing nodes. Clearly, this is not feasible in a real network. Hence, while this section shows the effectiveness of a scale-free solution will present an overlay construction algorithm based on a limited network knowledge which approximates the Barabasi-Albert model.

3.3 Proactive Replication

The square-root topology is complementary to the square-root replication. In situations where 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, but the combination of the two techniques can provide significant performance benefits. We conducted an experiment where we proactively replicated content according to the square-root replication. Each peer was assigned capacity equal to twice the content they were already storing, and this extra capacity was used to store proactively replicated copies. We then connected peers in the square-root, high-skew power-law, and low-skew power-law topologies, and measured the performance of random walk searches. Again, *G*= 10. The results are shown in Figure 2. 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.

Proactive 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, in our experiment, the square root topology with proactive replication required 68 percent fewer messages than the square root topology without replication.

3.4 Other search walk techniques

Next, we examined the performance of other walk-based techniques on different topologies. We compared three other techniques based on random walks:
• Biased high degree: messages are preferentially forwarded to neighbors that have the highest degree.
• Most results: messages are forwarded preferentially to neighbors that have returned the most results for the past 10 queries.
• Fewest result hops: messages are forwarded preferentially to neighbors that returned results for the past 10 queries who have traveled the fewest average hops .

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 authors studied: fewest result hops requires the least bandwidth, while most results has the best chance of finding the requested number of matching documents.

Each case 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 square-root 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. Overall, the square-root topology provides the best performance, even with the extremely efficient biased high degree/neighbor indexing combination. Moreover, 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. Again, 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% of the system peers on average while still finding sufficient results. Finally, the results so far assume state-keep ing, where peers keep state about where the search has been. Then, peers can avoid forwarding searches to neighbours that the search has already visited. We also ran experiments for no state keeping. The results demonstrate that the square-root topology is better than power-law topologies, whether or not statekeeping is used.

3.5 Other Topologies

We also tested the square-root topology in comparison to several other network structures. First, 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. Moreover, the cost of maintaining the square-root topology is low, as we discuss in Section 4, requiring easily obtainable local information. Thus, 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. The superpeers 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.   Conclusions

We compares structured and unstructured overlays, demonstrating through analytical and simulation results how an unstructured solution relying on a scale-free topology is an effective option to deploy for offering services based on equivalent servants. On the basis of this result, The EQUivalent servAnt locaTOR (EQUATOR) architecture, which overcomes the issues related to the deployment of a scale-free topology for service location in a real network, mainly due to the static nature of the ideal scale-free construction algorithm and the lack of a global knowledge of the participating peers. Simulation results confirmed the effectiveness of EQUATOR, showing how it offers good lookup performance in conjunction with low message overhead and high resiliency to node churn and failures. Some possible future works are introduced and are related to some complementary issues ranging from the proximity-aware selection of servants to the introduction of proper incentives to encourage nodes to join the EQUATOR overlay and offer their resources.
• Peer-to-peer computing takes advantage of existing desktop computing power and networking connectivity, allowing economical clients to leverage their collective power to benefit the entire enterprise
• Large number of participating servants and user's single entity directly handles all possible servants and consequently offers the best performance

References

[1]. Q.Lv ,P.Cao, E.Cohen, K.Li, and S. Shenker, "Search and replication in unstructured peer-to-peer networks," *SIGMETRICS Perform. Eval. Rev.*, vol. 30, no. 1, pp. 258–259, 2002.

[2]. R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Rev. Mod. Phys.*, vol. 74, pp. 47–97, Jan. 2002.

[3]. L. A. Adamic, R. M. Lukose, A. R. Puniyani, and B. A. Huberman, "Search in power-law networks," *Phys. Rev. E*, vol. 64, no. 4, pp. Sep 2001.

[4]. Guido Marchetto, Luigi Ciminiera, Marco Papa Manzillo, Fulvio Risso, Livio Torrero, "Locating Equivalent Servants over P2P Networks"

[5]. Oliver Hein, Michael Schwind, Wolfgang Konig, "Scale-Free Networks" The Impact of Fat Tailed Degree Distribution on Diffusion and Communication Processes.

[6]. Barabasi, Albert-Laszlo; Albert, Reka: Emergence of Scaling in Random Networks. In: Science 286 (1999), pp. 509–512.

[7]. Barabasi, Albert-Laszlo; Bonabeau, Eric: Scale-Free Networks. Scientific American 2003.

[8]. Ebel, H.; Mielsch, L.-I.; Bornholdt., S.: Scale-free topology of e-mail networks. In: Phys. Rev. E 66 (2002), pp. 035103.