# A Novel, Distributed, and Reliable Search Policy for Unstructured Peer-to-Peer Networks

Dr. Shanmugam Kannan[1], Goski Sathish[2]

[1]*Associate Professor,* [2]*Assistant Professor, Department of IT*
[1,2]*Malla Reddy Engineering College (A), Hyderabad, Telangana, India.*

## *Abstract*

*Finding a document or resource in an unstructured peer-to-peer network can be an exceedingly difficult problem. In this paper we propose a query routing approach that accounts for arbitrary overlay topologies, nodes with heterogeneous processing capacity, e.g., reflecting their degree of altruism, and heterogenous class-based likelihoods of query resolution at nodes which may reflect query loads and the manner in which files/resources are distributed across the network. The approach is shown to be stabilize the query load subject to a grade of service constraint, i.e., a guarantee that queries' routes meet pre-specified class-based bounds on their associated a priori probability of query resolution. An explicit characterization of the capacity region for such systems is given and numerically compared to that associated with random walk-based searches. Simulation results further show the performance benefits, in terms of mean delay, of the proposed approach. Additional aspects associated with reducing complexity, estimating parameters, and adaptation to class-based query resolution probabilities and traffic loads are studied.*

***Keywords:*** *P2P network, Query routing and Capacity region*

## 1. Introduction

Peer-to-Peer (P2P) systems continue to find increasing and diverse uses as a distributed, scalable and robust framework to deliver services, e.g., file sharing, video streaming, expert/advice sharing, sensor networks, databases, etc. One of the basic functions of such systems is that of efficiently resolving queries or discovering files/resources. This is the problem addressed in this paper. There is a considerable body of work exploring the design of efficient search/routing mechanisms in structured and unstructured P2P networks, see e.g.,[1]–[10]. In structured networks, peers/files/resources are organized to form overlays with specific topologies and properties. Search mechanisms that perform name resolution based on distributed hash table (DHT) coordinate systems can be devised to achieve good forwarding-delay properties, see e.g.,[2]. In such systems, the query traffic may depend on how keys are assigned. So, load balancing requires proactive/reactive assignments of keys to peers and data/service objects, e.g., [11], and possibly exploiting network hierarchies [10]. Fundamentally, in such networks the difficulty of search/discovery is shifted to that of maintaining the structural invariants required to achieve efficient query resolution particularly in dynamic settings with peer/content churn or when reactive load balancing is required. Unstructured networks, by contrast, are easier to setup and maintain, but their mostly *ad hoc* overlay topologies make realizing efficient searches challenging. In a purely unstructured P2P network, a node only knows its overlay neighbors. With such limited information, search techniques for unstructured networks have mostly been based on limited-scope flooding, simulated random walks, and their variants [3]–[5]. Much research in this area has focused on evaluating these search techniques based on the contact time, i.e., number of hops required to find the target, using the spectral theory of Markov chains on (random) graphs, see e.g., [4]–[6].

 Copyright © 2019 Authors

Unfortunately, in heterogenous settings where service capacity or resolution likelihoods vary across peers, such search techniques perform poorly under high query loads. The inefficiencies of purely unstructured networks can be partially addressed by hybrid P2P systems, e.g., Fast Track and Gnutella2. Such systems use a simple two-level hierarchy where some peers serve as 'super-peers.' These are high degree nodes which are well connected to other super-peers and to a set of subordinate nodes in a hub-and-spoke manner [12]. Though such systems have advantages in terms of scalability, proposed search techniques are still based on variants of flooding and random walks. The work of [7] proposes an approach where peers cache the outcomes of past queries as informed by reverse-path forwarding. The idea is to learn, from past experience, the best way to forward certain classes of queries, i.e., to intelligently "bias" their forwarding decisions by correlating classes of queries with neighbors who can best resolve them. This approach involves considerable overhead, is not load sensitive, and has not yet given guarantees on performance. Although, as will be clear in the sequel, our results are not exclusive to hybrid P2P networks, these will serve as the focus of the paper. We assume that each super-peer contributes a possibly heterogenous amount of processing resource for resolving queries for the network—incentives for doing so are outside of the scope of this paper, see e.g., [8], [9]. Super-peers serve their subordinates by resolving queries or forwarding them to other super-peers. Super-peers can resolve queries by checking the files/resources they have, as well as those of their subordinate community. In our approach we also introduce a notion of query classes. These might, for example, represent types of content, such as music, films, animations, documents, or some other classification of files/resources relevant to the application at hand. The idea is that such a grouping of queries into classes can be used as a low overhead approach to make useful inferences on how to relay queries. Given a hybrid P2P topology and query classification, we propose a novel query resolution mechanism which stabilizes the system for all query loads within a 'capacity region', i.e., the set of loads for which stability is feasible. Essentially, our policy is a biased random walk where forwarding decision for each query is based on instantaneous query loads at super-peers. To balance the load across heterogeneous super-peers, the policy aims at reducing the differential backlog at neighboring super-peers, while taking into account the class and history information to improve the query's resolvability. Our policy draws upon standard backpressure routing algorithm, which is used to achieve stability in packet switching networks, e.g., see [13], [14].

In previously studied backpressure-based systems, the goal is to deliver packets to the corresponding destinations. By contrast, our aim is to provide a grade of service in resolving queries with no fixed destinations. The random nature of the location of query resolution in the network leads us to deal with expected queue backlog instead of current queue backlog. Further, in P2P systems, the probability of resolution of a query at a given node depends on the query's history, i.e., the path that led it to the current node. These characteristics of P2P systems are not captured in previous works on backpressure by Tassiulas and Ephremides [13] and the subsequent enhancements, see e.g., [14]–[21]. To summarize, our approach differs from standard work on backpressure in that we incorporate the following different issues that arise in P2P search: (a) we model the uncertainty in the locations where a query may be resolved depending upon where the file/object of interest are placed, (b) we guarantee a grade of service to each query under such uncertainties, (c) we incorporate the information about a query's resolvability available through the knowledge of its history

## 2. Existing System

- In a purely unstructured P2P network, a node only knows its overlay neighbors. With such limited information, search techniques for unstructured networks have

 Copyright © 2019 Authors

mostly been based on limited-scope flooding, simulated random walks, and their variants.

- Much research in this area has focused on evaluating these search techniques based on the contact time, i.e., number of hops required to find the target, using the spectral theory of Markov chains on (random) graphs, see e.g., Unfortunately in heterogenous settings where service capacity or resolution likelihoods vary across peers, such search techniques perform poorly under high query loads.
- The inefficiencies of purely unstructured networks can be partially addressed by hybrid P2P systems, e.g., FastTrack and Gnutella2.

**Disadvantages**

- In structured networks the difficulty of search/discovery is shifted to that of maintaining the structural invariants required to achieve efficiency.
- In query resolution particularly in dynamic settings with peer/content churn or when reactive load balancing is required.
- Standard backpressure-based routing our policies suffer from a major drawback: each node needs to share the state of its potentially large number of non-empty queues with its neighbors.
- Complexity problem will be also raised.

## 3. Proposed System

Given a hybrid P2P topology and query classification, we propose a novel query resolution mechanism which stabilizes the system for all query loads within a 'capacity region', i.e., the set of loads for which stability is feasible. Essentially, our policy is a biased random walk where forwarding decision for each query is based on instantaneous query loads at super-peers. To balance the load across heterogeneous super-peers, the policy aims at reducing the differential backlog at neighboring super-peers, while taking into account the class and history information to improve the query's resolvability. Our policy draws upon standard backpressure routing algorithm, which is used to achieve stability in packet switching networks. Further, proposed a query forwarding mechanism for unstructured (hybrid) P2P networks with the following properties. It dynamically accounts for heterogeneity in super-peer's 'service rate,' reflecting their altruism, and query loads across the network. To the best of our knowledge, this is the first work to rigorously account for such heterogeneity in devising a search mechanism for P2P networks. It is based on classifying queries into classes. This classification serves as a type of name aggregation, which enables nodes to infer the likelihoods of resolving class queries, which, in turn, are used in learning how to forward queries. Our approach is fully distributed in that it involves information sharing only amongst neighbors, and achieves stability subject to a Grade of Service (GoS) constraint on query resolution. The GoS constraint corresponds to guaranteeing that each query class follows a route over which it has a reasonable 'chance' of being resolved. We provide and evaluate several interesting variations on our stable mechanism that help significantly improve the delay performance, and further reduce the complexity making it amenable to implementation.

Resource discovery is a challenging issue in unstructured peer-to-peer networks. Blind search methods including flooding and random walks, are the two typical algorithms used in such systems. Blind flooding is not scalable because of its high communication cost. On the other hand, the performance of random walks approaches largely depends on the random choice of walks. Some informed mechanisms use additional information, usually obtained from exciting queries, for routing. Such approaches can reduce the traffic overhead, but they limit the query coverage. We propose a simple index scheme to enhance search in unstructured P2P network. The index scheme uses a data structure

 Copyright © 2019 Authors

"Bloom Filters" to index files shared at each node, arid then gossip to one another to change their Bloom filters. In effect, each node indexes a random set of files in the network, thereby lowing every query to have a constant to be successfully resolved within search space. We propose two schemes which can be used to improve the search performance in unstructured peer-to-peer networks. The first one is a simple caching mechanism based on resource descriptions. Peers that offer resources send periodic advertisement messages. These messages are stored into a cache and are used for routing requests. The second scheme is a dynamic Time-To-Live (TTL) enabling messages to break their horizon. Additional aspects associated with reducing complexity, estimating parameters, and adaptation to class-based query resolution probabilities and traffic loads are studied.

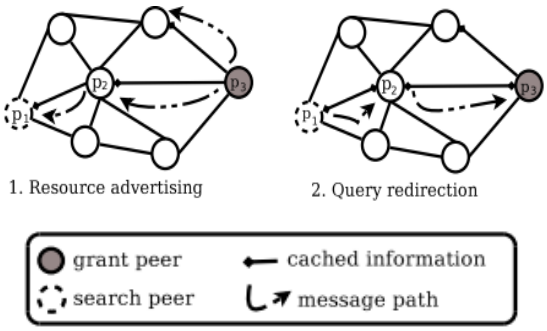
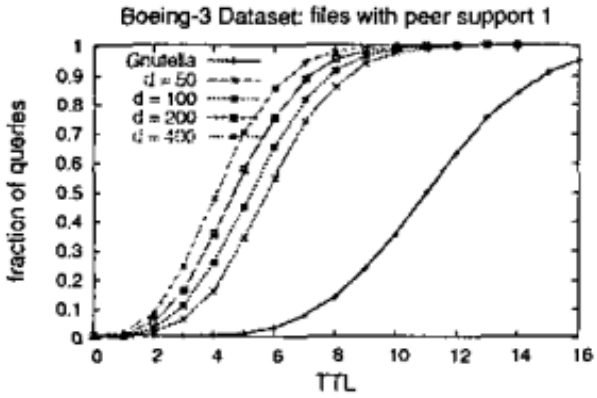



Figure 1. The basic operations of routing approach

**Advantages**

1. A simple but efficient cache system based on resource description.
2. A cache-based routing scheme aiming at forwarding messages towards resource location.
3. A dynamic TTL which enables messages to break their horizon when necessary while still maintaining a low overhead.
4. A set of schemes which can be used to improve the performance of existing search protocols based on caching and TTL.

Performance of Surrogate Queries to study how many queries should he reissued in the "Surrogate Queries" algorithm. The setting is similar to the previous one. Note that security is treated as a separate issue in the paper. Because peers are quite autonomous, all peer-to-peer networks are vulnerable to various security attacks, e.g., a peer responds with a 'poisoned' file, or blocks queries and fakes search results.




## 4. Conclusion

 Copyright © 2019 Authors

To summarize, we provided a novel, distributed, and reliable search policy for unstructured peer-to-peer networks with super-peers. Our backpressure-based policy can provide capacity gains of as large as 68% over traditional random walk techniques. We also provided modifications to the algorithm that make it amenable to implementation.

## References

[1] Wikipedia, "Peer-to-peer," 2011 [Online]. Available: http://en. wikipedia.org/wiki/Peer-to-peer

[2] I. Stoica *et al.*, "Chord: A scalable peer-to-peer lookup protocol for internet applications," *IEEE/ACM Trans. Netw.*, vol. 11, no. 1, pp. 17–32, Feb. 2003.

[3] X. Li and J. Wu, "Searching techniques in peer-to-peer networks," in *Handbook of Theoretical and Algorithmic Aspects of Ad Hoc, Sensor, Peer-to-Peer Networks*. Boca Raton, FL, USA: CRC Press, 2004.

[4] C. Gkantsidis, M. Mihail, and A. Saberi, "Random walks in peer-topeer networks," in *Proc. IEEE INFOCOM*, 2004, pp. 120–130.

[5] C. Gkantsidis, M. Mihail, and A. Saberi, "Hybrid search schemes for unstructured peer to peer networks," in *Proc. IEEE INFOCOM*, 2005, pp. 1526–1537.

[6] S. Ioannidis and P. Marbach, "On the design of hybrid peer-to-peer systems," in *Proc. ACM SIGMETRICS*, Annapolis, MD, USA, Jun. 2008, pp. 157–168.

[7] P. Patankar, G. Nam, G. Kesidis, T. Konstantopoulos, and C. Das, "Peer-to-peer unstructured anycasting using correlated swarms," in *Proc. ITC*, Paris, France, Sep. 2009, pp. 1–8.

[8] R. Gupta and A. Somani, "An incentive driven lookup protocol for chord-based peer-to-peer (P2P) networks," in *Proc. Int. Conf. High Perform. Comput.*, Bangalore, India, Dec. 2004, pp. 8–18.

[9] D. Menasche, L. Massoulie, and D. Towsley, "Reciprocity and barter in peer-to-peer systems," in *Proc. IEEE INFOCOM*, 2010, pp. 1–9.

[10] B. Mitra, A. K. Dubey, S. Ghose, and N. Ganguly, "How do superpeer networks emerge?," in *Proc. IEEE INFOCOM*, 2010, pp. 1–9.

[11] D. Karger and M. Ruhl, "Simple efficient load balancing algorithms for peer-to-peer systems," in *Proc. 16th ACMSPAA*, 2004, pp. 36–43.

[12] B. Yang and H. Garcia-Molina, "Designing a super-peer network," in *Proc. IEEE ICDE*, 2003, pp. 49–60.

[13] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks," *IEEE Trans. Autom. Control*, vol. 37, no. 12, pp. 1936–1948, Dec. 1992.

[14] M. J. Neely, E. Modiano, and C. E. Rohrs, "Dynamic power allocation and routing for time varying wireless networks," in *Proc. IEEE INFOCOM*, 2003, pp. 745–755.

[15] L. Ying, S. Shakkottai, A. Reddy, and S. Liu, "On combining shortest-path and back-pressure routing over multihop wireless networks," *IEEE/ACM Trans. Netw.*, vol. 19, no. 3, pp. 841–854, Jun. 2011.

[16] M. Alresaini, M. Sathiamoorthy, B. Krishnamachari, and M. Neely, "Backpressure with adaptive redundancy (BWAR)," in *Proc. IEEE INFOCOM*, Mar. 2012, pp. 2300–2308.

[17] L. Bui, R. Srikant, and A. Stolyar, "A novel architecture for reduction of delay and queueing structure complexity in the back-pressure algorithm," *IEEE/ACM Trans. Netw.*, vol. 19, no. 6, pp. 1597–1609, Dec. 2011.

[18] B. Ji, C. Joo, and N. Shroff, "Delay-based back-pressure scheduling in multihop wireless networks," *IEEE/ACM Trans. Netw.*, vol. 21, no. 5, pp. 1539–1552, Oct. 2013.

[19] Y. Cui, E. Yeh, and R. Liu, "Enhancing the delay performance of dynamic backpressure algorithms," *IEEE/ACM Trans. Netw.*, 2015, to be published.

 Copyright © 2019 Authors

[20] L. Georgiadis, M. J. Neely, and L. Tassiulas, "Resource allocation and cross-layer control in wireless networks," *Found. Trends Netw.*, vol. 1, no. 1, pp. 1–144, 2006.

[21] Y. Cui, V. Lau, R. Wang, H. Huang, and S. Zhang, "A survey on delayaware resource control for wireless systems—Large deviation theory, stochastic Lyapunov drift, distributed stochastic learning," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1677–1701, Mar. 2012.

[22] S. Asmussen, *Applied Probability and Queues*. NewYork, NY, USA: Springer, 1987.

[23] H. J. Kushner and G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*. New York, NY, USA: Springer, 2003.

 Copyright © 2019 Authors