Computer Networks 100 (2016) 124–140

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

Popularity-based scalable peer-to-peer topology growth☆

Gurhan Gunduz[a,1,*], Murat Yuksel[b]

[a] Computer Engineering Department, Pamukkale University, Turkey
[b] Computer Science and Engineering, University of Nevada - Reno, Reno, NV 89557, USA

ARTICLE INFO

Article history:
Received 22 May 2015
Revised 3 February 2016
Accepted 14 February 2016
Available online 23 February 2016

Keywords:
Peer-to-Peer networks
Scalable
Popularity-based
Overlay networks

ABSTRACT

Peer-to-peer (P2P) networks have gained importance and spread significantly during the last decades resulting in raised research interest in this area. The basic premise of P2P design is higher scalability and many existing large-scale applications, such as Twitter and Skype, use a form of P2P design. Although *structured* P2P designs enable one to guarantee finding of every item (rare or popular), they do not scale beyond a point and support from servers are needed. This breaks the decentralized design of the P2P system and results in a hybrid scheme. *Unstructured* P2P networks, on the other hand, can scale to much larger nodes but yet cannot give time guarantee for finding a rare item.

Topological characteristics of unstructured P2P networks have impact on the efficiency of a search for items. Earlier studies have shown that popularity of the nodes has significant impact in the self-organization of the overlay topology. It is well known that unstructured P2P designs are very good when searching for a popular item. But, when a new trend is emerging, it is a rare item until it becomes highly popular. During that transitional period, unstructured designs suffer from inefficiency in finding *to-be-popular* (TBP) items. Such TBP items could be stuck at nodes with small degrees and thus become hard to find. We hypothesize that if the overlay topology is established and grown with a more proactive consideration of the items' popularity, the delay in finding TBP items could be reduced significantly. Further, such topology growth will reduce the search time for popular items. Thus, the overall search performance of the P2P system will significantly improve as well, since most of the searches are for popular items. In this paper, we investigate incorporating item popularity into the overlay topology in a scalable way.

1. Introduction

Peer-to-peer (P2P) networks have gained importance and spread significantly during the last decades resulting in raised research interest in this area. The basic premise of P2P design is higher scalability and many existing large-scale applications, such as Twitter and Skype, use a form of P2P design. In P2P networks, a peer can search and download resources from other peers while contributing own resources to others in the same network. The approach of peering distributes the load and functions across the participant peers/nodes, and hence, attains a much better scalability in comparison to centralized designs. The peers typically use an overlay topology to search and find each other to share or exchange their resources. Although overlaying has excellent advantages in providing fast deployment of protocols and flexibility in function placement, it can cause significant holdbacks in terms of performance if not carefully handled. Particularly in unstructured P2P networks [1], the characteristics of the overlay topology

☆ This work was supported in part by United States NSF awards AST-1444077 and CNS-1321069, and Republic of Turkey's TUBITAK.

* Corresponding author. Tel.: +905055827396; fax: +1 775 7841877.
E-mail addresses: ggunduz@pau.edu.tr (G. Gunduz), yuksem@cse.unr.edu (M. Yuksel).

[1] Dr. Gunduz was a visiting researcher at the University of Nevada - Reno during most of this work.

http://dx.doi.org/10.1016/j.comnet.2016.02.017

have profound impact on the efficiency of flooding-based P2P search and applications.

Several algorithms have been designed for constructing the topology of unstructured P2P networks [2]. These algorithms typically devise a distributed mechanism by defining how a new node joins or an existing node leaves the network. With the assumption that every node is a peer and the same, most of the recent work on this topic looked at using node degree for organizing the topology. In fact, every node or peer in these networks may be different than others in terms of computation power, bandwidth, capacity and the resources that it is sharing. One implicit way of determining which node has more resources is to measure how popular the nodes are. This imbalance of resources naturally gives rise to a skew in the popularity of the nodes, as some nodes become more popular due to their larger resources and/or the popularity of the items it is maintaining. Intuitively, if a node has more and especially popular items, then more peers will be interested in interacting with that node. This intuition also can be used as the foundation for relating popularity of the nodes to their degrees, since peers will be interested in establishing and maintaining a link to a resourceful node only. Further, some nodes may become more popular if they have items that everybody is looking for. Other peers will want to create a connection with these popular nodes. Therefore, the term "popular" in this context means having relatively higher number of links compared to other nodes in the system.

Another key issue the existing P2P topology construction algorithms need to address is the efficient and effective treatment of trends and memes in the recent social networks [3,4]. Handling such trends and memes in a scalable way alongside with the growing usage of device-based networking and sharing applications (such as Twitter) is a challenge [5]. It is well known that unstructured P2P designs are very good when searching for a popular item. But, when a new trend is emerging, it is a rare item until it becomes highly popular. During that transitional period unstructured designs suffer from inefficiency in finding to-be-popular (TBP) items. Such TBP items could be stuck at nodes with small degrees and thus become hard to reach. We hypothesize that if the overlay topology is established and grown with a more proactive consideration of the items' popularity, the delay in finding TBP items could be reduced significantly. Further, such topology growth will reduce the search time for popular items. Thus, the overall search performance of the P2P system will significantly improve as well, since most of the searches are for popular items.

The primary focus of this research is to find scalable ways of *embedding the item popularity* into the overlay topology so that flooding-based search on overlay topologies is efficient. We introduce a popularity parameter for each node, which is based on the popularity of the items that the node has. This information is used when a new node wants to join the network. Topologies are generated based on this factor along with degree. In order to give our method a practical perspective, we limit the number of neighbors of a node to a predefined value indicated by hard cutoff [6]. The ad-hoc nature of peers is exploited by rewiring the network upon failure or departure of a peer. We also investigate the effect of these popularities and rewiring of ad-hoc nodes on the P2P search efficiency.

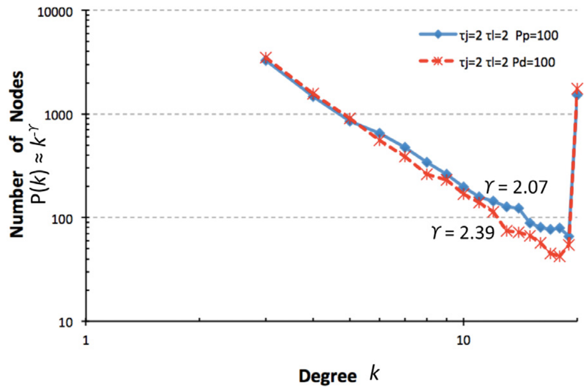

Fig. 1. Comparison of degree distribution of popularity ($P_p = 100$) and degree ($P_d = 100$) based P2P overlay topologies with a degree cutoff $k_c = 20$: Popularity-based overlays have fatter tail, indicated by the smaller power exponent, $\gamma = 2.07$.

Our test results show the pattern in Fig. 1 for a high-level comparison of degree-based topology and our popularity-based topology. Detailed analyses of the results show us the following: The number of minimum and close-to-minimum degree nodes is higher in the degree-based topology. This is not desirable since having only few neighbors will affect the search performance of the topology. As the node degree increases, our popularity-based topology has more nodes than the degree-based topology. Having high degree nodes results in more connected topology, which eventually increases the search performance. The only advantage of degree-based topology is that the number of maximum degree nodes is slightly higher than our popularity-based topology. Even though this feature helps degree-based topology for improved search performance, it is not enough to perform better than our popularity-based topology

1.1. Contributions

Popularity of the nodes has significant impact in the self-organization of the overlay topology. Power-law graphs have super-hubs which can be considered as the "center" of the topology. The most profound property of these super-hubs is their large degree. Yet the research shows mixed results [7] on power the existence of power-law behavior in P2P networks. It was reported [8,9] that power-law behavior emerges in unstructured P2P networks if nodes that are close to the edge of the topology are excluded. Several studies [10–12] showed small-world topology for P2P networks. Small world topologies do not have super-hubs. But, the "center" is the node(s) where betweenness centrality is the maximum. Even for small-world networks, betweenness centrality is higher at nodes equally farther away from the leaf nodes, which usually causes those nodes to be at the cross-section of many end-to-end paths and hence have a higher degree [13].