Pervasive and Mobile Computing 40 (2017) 339–358

Contents lists available at ScienceDirect

Pervasive and Mobile Computing

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

Evolution of superpeer topologies — An analytical perspective

Joydeep Chandra [a,*], Bivas Mitra [b], Niloy Ganguly [b]

[a] Indian Institute of Technology Patna, India
[b] Indian Institute of Technology Kharagpur, India

ARTICLE INFO

Article history:
Available online 13 July 2017

Keywords:
Superpeer networks
Degree distribution
Bootstrapping protocols
Webcache
Rate equation
Preferential attachment

ABSTRACT

In superpeer based networks, resourceful peers (having high bandwidth and computational resources) are discovered through the process of bootstrapping, whereby they get upgraded to superpeers. However, bootstrapping is influenced by several factors like limitation on the maximum number of connections a peer can have due to bandwidth constraints, limitation on the availability of information of existing peers due to cache size constraints and also by the attachment policy of the newly arriving peers to the resourceful peers. In this paper, we derive closed form equations that model the effect of these factors on superpeer related topological properties of the networks. Based on the model, we show that existing bootstrapping protocols can lead to a situation where only a small fraction of the resourceful peers gets converted to superpeers, i.e., a large fraction of them remain underutilized; we later validate this statement using real Gnutella snapshots. We observe that as a node attachment policy, newly arriving peers must use a combination of random and preferential attachment strategy so as to ensure proper utilization of the resourceful peers. We also show that the cache parameters must also be suitably tuned so as to increase the fraction of superpeers in the network. Finally, we show that in real Gnutella networks the degree distribution generated using our models suitably fits the corresponding empirical values.

1. Introduction

The popularity of peer-to-peer (p2p) network based applications is growing significantly in recent years. Application of p2p based technologies are being proposed in several emerging network based systems like social networking systems [1,2], vehicular networking [3] and multimedia communication networks [4]. Most of these networks are self-organizing and emergent in nature, i.e., there is a constant evolution of these networks that cannot be centrally controlled. In most of the cases, the networks follow the principle of preferential attachment like natural systems, where the increase in degree of a node is proportional to its current degree. This phenomenon can be observed in several natural systems; like in citation networks the probability that a paper receives a new citation increases with the number of citations it had already received. However, unlike these networks, in technology-mediated networks like p2p systems, unbridled preferential attachment is not possible due to certain network specific constraints that exist for these networks. For example, due to bandwidth limitations, nodes in many of these networks impose a cutoff on the maximum number of connectivities that a node can have [5,6]. Further, in many systems like Gnutella [7], the information about many of the high degree nodes are preferentially stored in a webcache, so that newly arriving peers can obtain information about these peers and connect to them to start the bootstrapping process. However, these caches are finite sized and hence information about only a subset of these nodes

* Corresponding author.
E-mail addresses: joydeep@iitp.ac.in (J. Chandra), bivas@cse.iitkgp.ernet.in (B. Mitra), niloy@cse.iitkgp.ernet.in (N. Ganguly).

http://dx.doi.org/10.1016/j.pmcj.2017.07.005
1574-1192/

are maintained in the caches. This constraint affects the evolution of connectivities as the newly arriving nodes connect preferentially to those nodes whose information are present in the cache. The existing mathematical models of network evolution do not consider these network specific constraints and hence are not suitable to model the evolutionary properties of these application networks. Hence there is a need to rework the existing mathematical models to gain insights about these networks which is the goal of this paper.

Our Contributions: In this paper, we consider three bootstrapping parameters – (a) the degree cutoff of the nodes, (b) the finite sized caches and (c) the connectivity principle, to determine a relationship between these parameters, and the degree distribution and the characteristics of the most high degree nodes of these networks, that are called superpeers. The term superpeers has been conventionally used to denote those peers that have high resources and connect to a large set of other peers providing them different services [8,9]. They function as a set of equals among other superpeers and as servers for the large set of other client peers. Thus resourcefulness as well as large number of connections are both important factors that distinguish a superpeer from other peers. In our case, since we assume that peers have similar bandwidth resources, hence we identify superpeers as those peers that have reached their cutoff degree, i.e. maximum connectivity. Thus these peers provide a maximum share of their bandwidth to other peers in the network. A high fraction of superpeers thus implies that the performance of the network in terms of bandwidth sharing is high. Under such conditions, the average download rate received by the peers is also high. In several emerging applications like vehicular networks and multimedia networks, these superpeers also provides additional services like service discovery and indexing [10,11]. In practical cases, the extent of preferentiality varies across different networks, i.e., the connectivities between the nodes is not purely preferential but has a random component, ϵ attached to it, that we also consider in our model. A part of the work stated in this paper, that includes the model for the topological characteristics for high and low ϵ values and the formalism for finite sized webcache for a single cutoff degree has been published in [12]. In this paper we provide a detailed discussion of these models and also extend the same to include multiple cutoffs and network performance evaluation measures for these bootstrapping parameters. Through these models we show that in naturally evolving networks with several resourceful peers, the fraction of these peers that get converted to superpeers is inherently determined by these bootstrapping parameters like ϵ, the cache size and the degree of preferentiality in attachment. It is observed that the fraction of superpeers in the network can be increased by suitably tuning these bootstrapping parameters.

The outline of the paper is as follows. We present a brief survey of the related works in Section 2. In Section 3, we state and model the bootstrapping protocol followed by peer servants. In Section 4 we propose a formal framework for analyzing the topological properties of the networks starting with a simple assumption that all peers join the network with fixed cutoff degree. We subsequently develop the analytical frameworks with more realistic assumptions in a step by step manner; in Section 5 we extend the theory for the case when webcaches are present and in Section 6 we develop models when nodes have different bandwidths and hence varying cutoff degrees. In each of these sections, we validate the corresponding model with simulations and reveal the insights that can be drawn from the same. In light of the framework developed, an empirical analysis of various topological structures found in real networks is performed in Section 7. Finally, we draw conclusions in Section 8.

2. Related works

A significant amount of work have been done proposing new ways to construct optimal peer-to-peer topology to maintain desired quality of services [13–22]. Several works analyze the importance and challenges in generating optimal topologies [9,23]. Authors in [23] show that peers connecting selfishly to other peers to optimize their individual benefit can lead to a non-optimal unstable network that changes continuously. One of the most cited papers in the topic, [17] has proposed a bootstrapping algorithm which builds a p2p network with small diameter keeping the average degree constant. However, the design heavily depends on a central server that is needed to coordinate the connections between peers. In [22], the authors propose techniques to generate balanced optimal-diameter p2p topologies. Authors in [24] proposed a distributed bootstrapping protocol with random walk based joining and relinking process and have shown that topological structure of p2p networks depends heavily on the node heterogeneity and capacity distribution of the joining nodes; however, no real data analysis was done for the validation study. In [25], the authors proposed a scalable peer-to-peer based bootstrapping protocol for Web Real Time Communication Systems (WebRTC) that uses distributed hash tables. In addition to these protocols, other 'optimal' bootstrapping algorithms [13–15,17–19] have also been proposed; however, these algorithms are difficult to analytically model and analyze. Hence research needs to be directed towards modeling the simple bootstrapping process as node attachment rules and present results to the network engineers about the exact impact of the simple parameters. The present paper attempts to achieve this objective.

There is a great amount of interest among physicists in understanding the growth of complex networks due to the attachment of joining nodes [26–30]. It has been found that degree distribution of most of the large scale networks like Internet, Web network etc., follow power law, hence most of the theories developed to calculate the topology of the emerging network is directed towards explaining the emergence of scale free networks. Barabasi and Albert [28] coined the term "preferential attachment" and analytically showed its relation with the emergence of scale free networks. A large number of works have followed this initial work where researchers have suggested extensions or modifications of the model that makes it a more realistic representation of processes taking place in real-world networks, like the Bianconi and Barabasi [30] fitness model and the models in [26,31] that take into account the aging of nodes so that a link is connected not only preferentially