Analysis of Scale Effects in Peer-to-Peer Networks

Yung-Ming LI, Yong TAN, and Yong-Pin ZHOU

***Abstract* – In this paper, we study both the positive and negative scale effects on the operations of Peer-to-Peer (P2P) file sharing networks, and propose the optimal sizing (optimal number of peers) and grouping (optimal number of directory intermediary) decisions. Using analytical models and simulation, we evaluate various performance metrics to investigate the characteristics of a P2P network. Our results show that increasing network scale has a positive effect on the expected content availability and transmission cost, but a negative effect on the expected provision and search costs. The tradeoffs among all the performance measures can be balanced, and consequently applied to determine the network size that maximizes the overall utility of a content sharing P2P community. We also investigate the impacts of various P2P network parameters on performance measures as well as the optimal scale decisions. Furthermore, we extend the model to examine the grouping decision in the networks with symmetric interconnection structures, and compare the performance between random and location based grouping polices.**

***Index Terms* – peer-to-peer networks, network operations, performance evaluation, distributed file systems**

I. INTRODUCTION

Peer-to-peer (P2P) technologies link social networks into cooperative ventures that share information (audio, video and graphic files), computer resource (computing cycles, hard disk space, and network bandwidth), as well as communication and collaboration (instant messaging). Members of a P2P community exchange information or other resources directly with each other, with very little or no use of a centralized or dedicated server. Many P2P services exist today, such as file sharing service (Gnutella and Freenet), grid computing service (Popular Power and Distributed Net), instant messaging service (AOL, Yahoo! and MSN), and online collaboration service (Groove Networks).

Among various P2P applications, file sharing is probably the most popular. P2P file sharing applications accounted for five of the top 10 downloads from the download.com web site in the last week of June 2002, together constituting 4.5 million downloads [14]. In contrast to the traditional web server based content delivery paradigm, this emerging "bottom-up" mode of information distribution, leveraging the resources on the peer nodes, is considered to be superior. P2P file sharing networks have attracted many users and much press attention, along with the ire from media firms who feel threatened by the illegal exchange of digital music and movie files. Recent empirical evidence [15] also suggests that the increasing use of unauthorized P2P file sharing causes the revenue decline in record industry. However, it also reveals the superior power of utilizing P2P network as content distribution medium.

P2P technologies have many operational characteristics that make them appealing: First, they rely on peer nodes, not the central servers, to deliver content, and therefore are more scalable. Second, on a large P2P network, it is likely for any node to find another node with the desired content that is "close", so transmission delay may be lowered as well. There are, however, drawbacks inherent in the P2P networks, due to the same decentralized structure. First, because each peer node can modify their contents freely, it may be costly to find desired contents. Second, since P2P users obtain contents from each other, the availability of these contents completely depends on the peer nodes being logged on. So, content reliability may be an issue.

In many ways, the size of a P2P network can impact many of these factors. A large network could alleviate the content reliability problem because the same content is likely to exist on multiple peer nodes. A large network would also reduce the transmission delay, as the closest service node will become closer as the network contains more nodes. On the other hand, on a large-scale P2P network, the number of queries may cause congestion at the directory server (if any) as well as network traffic congestion (one query may be forwarded multiple times before a suitable service node is found), due to the limited capacity and network bandwidth. Therefore, determining the "right" network scale is very important for P2P operations.

In this paper, we propose four metrics to evaluate the impact of network scale on the operational performance of P2P networks: content availability, search delay, provision delay, and transmission delay. Using these metrics and balancing all the tradeoffs, we examine the overall scale effect (network externality) and suggest optimal scale decisions, from the P2P network organizer's perspective. In particular, we focus on the impact of local peer parameters, such as P2P participants' computing and bandwidth capacities, local content provision amount, content request pattern and frequency, and sharing propensity.

Y.-M. Li is with the University of Washington Business School, Seattle, WA 98195-3200 USA (e-mail: ymli@ u.ashington.edu).

Y. Tan is with the University of Washington Business School, Seattle, WA 98195-3200 USA (telephone: 206-616-6785, e-mail: ytan@ u.ashington.edu).

Y.-P. Zhou is with the University of Washington Business School, Seattle, WA 98195-3200 USA (e-mail: yongpin@ u.ashington.edu).

Adar and Humerman [1] surveyed the Gnutella network and found 70% of peers on the network are free riders. The prevalence of free riders, who take but never contribute, not only reduces the aggregate content availability but also increases the workload at non-free-riding nodes. In this paper, we will also investigate the impact of dispersion of P2P users, content variety distribution, and content popularity distribution.

Several P2P structures exist and vary by their search algorithms [20]. A centralized P2P architecture, such as Napster, has the scale problem because of the difficulty in scaling the central directory server. Pure decentralized P2P architectures, such as Gnutella v0.4, while easily scalable because search is carried out among peer nodes, have to deal with excessive network traffic due to decentralized broadcast-type search. Newer generations of P2P software, such as KaZaA and Gnutella v0.6, use a combination of centralized and pure decentralized network structure: the peer nodes are grouped and served by supernodes (or, super-peers). Various groups are interconnected via supernodes to forward requests [19].

In this paper, we analyze the promising supernodes P2P networks structure. In particular, because the scale of a P2P plays an important role in determining the network's performance, we investigate two important operational issues of the P2P networks: sizing and grouping decisions. Sizing refers to the determination of the optimal size for a P2P community for any given supernode (i.e. the optimal number of peers connected to the same supernode). Grouping refers to the partition of a fixed number of nodes into multiple P2P communities (i.e. the optimal number of supernodes, given a number of peers). An important factor in the grouping decision is the interconnection structure among groups. In particular, we compare the performance between random and location based grouping decisions, which could be supported by new P2P protocols.

While much of the P2P research has been directed to the technological issues such as search algorithm and topology design, there is little attention been paid to the operational aspects of the P2P networks. In this paper, we use probabilistic distributions and queueing model to formulate the characteristics of typical P2P system dynamics, and present several main performance metrics for evaluating the P2P networks. These metrics allow us to study both the positive and negative scale effects on the operations of P2P file sharing networks, and suggest the optimal scale (sizing and grouping) decisions.

The remainder of the paper is organized as follows. Section II reviews related literatures on P2P networks. Section III gives a formal description of the model, outlines the system parameters, and proposes performance metrics. In Section IV, we analyze the scale and parameter effects on the proposed performance metrics and present the simulation results. Section V examines the impact of various system parameters on the optimal scale decision. Section VI extends the model to analyze grouping decision in various interconnection structures. Section VII concludes our finding and offers directions for future research.

II. Literature Review

There are a number of papers on the technical aspects of P2P networks. These papers focus mainly on developing efficient communication protocols, network topologies, and search algorithms [17][24]. Supernode structure is a promising structure P2P networks developed to improve the search efficiency in pure P2P networks. Sningh *et al* [23] present the incentives for several participants, especially the service providers, to deploy a supernode infrastructure. Yang and Garcia-Molina [26] evaluate the performances and present practical guidelines for the design of an efficient supernode network. Singh *et al* [22] present incentives for deploying the supernodes and propose a topic-based search mechanism to improve the effectiveness of supernodes.

On the topic of networks scale, Asvanund *et al* [2] empirically analyze the network externality in P2P music sharing networks and suggest that larger networks are not always better. Yang and Garcia-Molina [25][27] design various content-sharing P2P search architectures and compare the maximum number of users that can be served on them. Butler [4] investigates the effect of membership size and communication activity on the sustainability of online social structure. The results of this study suggest that networked communication technologies will provide benefits to balance the opposing impacts form membership size. These work provide valuable empirical evidences on the scale effect, but they do not present underlying operational metrics for evaluating the performance of networks and gaining insights about optimal scale decisions.

Regarding the grouping of P2P networks, Asvanund *et al* [3] propose a scheme for club membership management based on content similarity and physical location. Ledlie *et al* [13] develop a hierarchically grouped system that can self-organize to overcome unreliability. Khambatti *et al* [8] use attribute based clustering models to simulate how self-configuring communities are formed. Their simulation results demonstrate that community structures in a random network can be efficiently discovered based on the attribute and link information of peers.

Recently, a few researchers have started to explore the social and economical aspects of P2P free riding phenomenon and incentive mechanism design. For example, Golle *et al* [6] construct a formal game theoretic model to develop and analyze several payment mechanisms to encourage the file exchanges activities. Krishnan *et al* [9][10] propose a plausible model to analyze the existence of free-riding behaviors in P2P file sharing networks. However, the framework, assuming a constant sharing cost in the absence of any query forward interconnection, does not explicitly discuss the impacts of system parameters on network structures. While most of the researches on the P2P networks in technological domains assume that users will follow prescribed protocols without deviation, Shneidman and Parke [21] advocate a P2P model in which users are rational and self-interested. They develop a new operating mechanism that allows the users to behave rationally while still achieving good overall system outcomes. Using economic incentive model, Jackson and Wolinsky [7] examine whether efficient (value maximizing) social networks will form when self-interested individuals can choose to form or sever links.

Additionally, many reputation and trust systems are proposed to provide the incentive for cooperation without involving a pricing scheme [5]. For example, Ranganathan *et al* [16] propose a multi-person prisoner's dilemma model to investigate the behavior of user and develop pricing-based and reputation-based mechanism to improve the system performance. Wang and Vassileva [28] propose a Bayesian network based model to build reputation that's based on the recommendation in P2P network. Kung and Wu [11] present a reputation-based P2P admission system, using the eigenvector approach, to allow only those nodes that have made reasonable service contributions to receive services from others.

However, to the best of our knowledge, little attention has so far been given to the operational aspects of P2P networks. In the paper, we focus on the scale issue, and develop analytical model to examine how the network size and system parameters affect the performance of P2P networks and optimal sizing and grouping decisions.

III. The Model

We consider a content-sharing P2P network in which the participants are categorized as regular peer nodes and supernodes. A supernode and a number of regular peer nodes form a community. Only supernode maintains up-to-date information on all resources available in the community. Every content request (query) is generated at one of the peer nodes, and first processed at the local supernode on a first-come, first-served basis. For every query it processes, the supernode recommends a provision node that has the desired file and is "close" to the request node (determined by the bandwidth between these two nodes). Once this information is passed on to the requesting and provision nodes, download occurs directly between these two nodes. There could be many supernodes but each peer node is connected to only one supernode whenever it logs on. If the query cannot be satisfied from the local supernode, it will be forwarded to other supernodes.

A. Operating Policy

Figure 1 depicts the operations of a supernode-based P2P network. On a snapshot of a network, a peer node A in community G_1 needing a file that it doesn't own will send a content request to the local community center. The supernode of the community G_1, SP_1, searches its directory database and responds with a list of nodes that share the requested content (e.g. nodes B and D), along with the transmission information (approximate delay). It also recommends the node with minimum transmission delay as the provision node (node D). After that, the requesting node A downloads the content directly from provision node D. If the request is not satisfied (no node shares the requested content in the local community G_1), the query will be forwarded to other interconnected supernodes, SP_2 and SP_3, based on various peering policies (such as parallel or sequential forward).

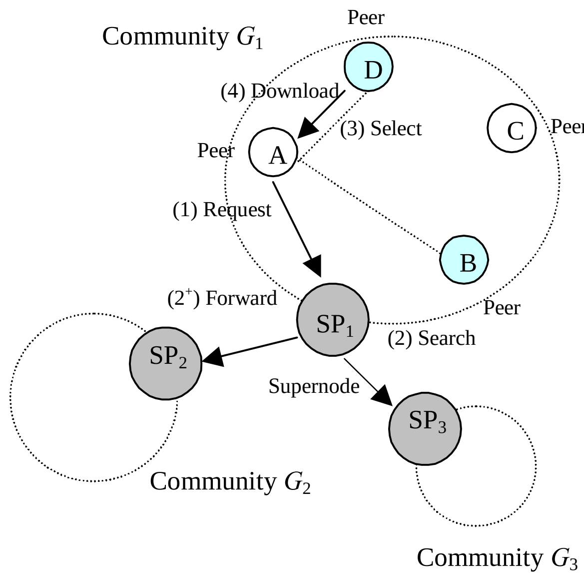


Fig. 1. Operational sequences of hybrid P2P networks (supernode structure).

B. Network Dynamics and Distributions

Since each peer node is a content consumer as well as a content provider, the dynamics of a P2P network is highly dependent on the parameters of local peer nodes. The parameters, listed in Table I, are used to describe the network dynamics such as content provision distribution, content request distribution, and bandwidth (or, transmission delay) distribution.

TABLE I
MODEL PARAMETERS

n	Number of active P2P users (peers)
s	Number of supernodes (groups)
M	Number of file varieties
α	File popularity distribution parameter (Zipf coefficient)
β_i	Average number of files stored in peer node i
$\gamma_{i,m}$	The probability that peer node i share its file m
μ_0	Search service rate of supernode
μ_i	Provision service rate of a regular peer node i
λ_i	Content request rate by peer node i
ρ	Dispersion degree of peers' positions (radius of a circle)

<u>Content Provision Distribution</u>. More popular contents are assumed to be stored and requested with higher probability. It is interesting to investigate how the variation of content popularity affects the operational performance of P2P networks. We assume that there are M same-size files in the

P2P community, denoted by ($F_1, F_2, \cdots, F_M$). These files are ranked in descending order by their popularity. Let θ_m be the probability that a new random request asks for file $F_m, \forall m \in \{1, \cdots, M\}$. We assume that θ_m follows a Zipf-like distribution: $\theta_m = m^{-\alpha} \cdot \theta_0$, $\forall m \in \{1, \cdots, M\}$, where $\theta_0 = 1 / \sum_{m=1}^{M} m^{-\alpha}$ is the normalization factor, and $\alpha \geq 0$ is a parameter for the relative popularity distribution. The greater the value of α, the larger the variations of popularity among the files. When $\alpha = 0$, the distribution is uniform, i.e. each file has the same popularity.

Assume there are n active peer nodes in certain P2P community. We introduce binary random variables, $X_{i,m}$, to indicate the availability of file F_m *stored* on node i, and $\overline{X}_{i,m}$, the availability of file F_m *shared* on node i. Explicitly,

$$X_{i,m}\left(\overline{X}_{i,m}\right) = \begin{cases} 1, & \text{if node } i \text{ has (shares) file } F_m; \\ 0, & \text{otherwise,} \end{cases}$$
$$\forall\, i \in \{1, \cdots, n\}, m \in \{1, \cdots, M\}.$$

Let $P_{i,m} = E(X_{i,m})$ and $\overline{P}_{i,m} = E(\overline{X}_{i,m})$, then they denote the probability that node i has (shares, respectively) file F_m. Obviously, $X_{i,m} \geq \overline{X}_{i,m}$ and $P_{i,m} \geq \overline{P}_{i,m}$. Furthermore, we assume that θ_m, the probability of some node requesting file F_m, and P_m, the probability of some node storing the file, are directly proportional, i.e., $P_{i,m} = \beta_i \cdot \theta_m$. Here, the parameter β_i represents the intensity of content availability and the value of β_i is limited by $\beta_i \cdot \theta_1 \leq 1$. β_i can be interpreted as the average number of files stored in a peer node i. A peer node can decide whether to share its own files (i.e., allow upload). Therefore, the content availability at a peer node becomes $\overline{P}_{i,m} = P_{i,m} \cdot \gamma_{i,m}$, where $\gamma_{i,m}$ is the probability that peer node i decides to share file F_m.

Content Request Distribution. Let λ_i be the request rate for all files by node i, and $\lambda_{i,m}$ be the request rate for file F_m by node i. Because a node will request a particular file at most once, the arrive rates, $\lambda_{i,m}$, should be interpreted as the rates at which these requests are to occur soon. Our queueing analysis corresponds to a "snapshot" of the network. We assume that the content requests follow a Zipf-type distribution similar to the content provision distribution. Hence, the request rate for file F_m can be derived as $\lambda_{i,m} = \theta_m \cdot \lambda_i$.

Search and Provision Processes. The two most time-consuming activities during the entire processes of P2P content distribution are search and download. The former occurs at the supernode, and the latter occurs between peer nodes. We use standard queueing model to evaluate the delay caused by each activity. Both the services times for the search process at the supernode and the provision process at a peer node i are assumed to follow the exponential distribution with service rate μ_0 and μ_i respectively. Because the requests follow a Poisson process, the search at the supernode can be modeled as an M/M/1 queue. Moreover, because the departure process of an M/M/1 queue is also Poisson (see Corollary 5.6.2 in [18]), the requests that are forwarded to the peer nodes for download also follow a Poisson process. Therefore, we can model the provision process at each peer node also as an M/M/1 queue.

Transmission Delay Distribution. P2P technologies utilize the aggregate bandwidth from edge nodes for content transmission to avoid the congestion at dedicated servers. Therefore, the effective bandwidth is scalable with the number of active users. It is difficult to exactly estimate the effective bandwidth and corresponding transmission latency of download activity. However, using network coordinate mapping technologies, such as the Global Network Positioning (GNP) approach [12], the coordinate-based positions of P2P networks can be used to approximately predicate Internet "transmission distance." Considering the properties of dynamic uptime and position of peer nodes (for example, the users are logged on only for a short time period or use a mobile computer), we assume that, at a snapshot, the active peer nodes are uniformly located on a circle with radius ρ. A higher ρ indicates greater dispersion of the users. Furthermore, the transmission delay (distance) between content request node i and provision node j, $T_{i,j}$, is an *i.i.d.* random variable uniformly distributed on $[0, \rho \cdot \pi]$.

Provision Policy. If more than one nodes can provide a file, the node with the minimum transmission delay to the request node will be selected as the provision node. That is, when node i requests file F_m, the community center suggests the optimal provision node j^*, where,

$$j^* = \arg\min_j T_{i,j} \text{ s.t. } \overline{X}_{j,m} = 1.$$

C. Performance Metrics

The performance metrics are established based on the benefits and costs of each activity (request, search, download, and transmission) during the entire process of P2P content distribution. Hence, the system performance metrics include content availability at the requested peer node, the search delay at supernodes, the provision (upload) process delay, and the transmission delay on the networks. Since each peer is both content consumer and provider, all these metrics are associated with the number of active peers, their behaviors (such as request frequency and sharing decision), and the service capacity of supernodes, provision nodes, and network. In our analytical model, we assume all the peer nodes are identical (in probability) and the sharing decision distribution for every file is the same. That is,

$$P_{i,m} = P_m, \overline{P}_{i,m} = \overline{P}_m, \lambda_{i,m} = \lambda_m, \mu_i = \mu, \beta_i = \beta, \gamma_{i,m} = \gamma,$$
$$\forall\, i \in \{1, \cdots, n\}, m \in \{1, \cdots, M\}.$$

Content availability. Content availability (or hit rate) is an important measure of the quality of content provision. Content availability is defined as the probability that an arbitrary request can be satisfied on the P2P networks. Hence, it depends on the content stored and shared on peer nodes. Let $H(n,m)$ be the expected content availability of file F_m in the local community with n nodes. We have,

$$H(n,m) = 1 - \prod_{i=1}^{n}(1-\overline{P}_m)\,.$$

The overall expected content availability can be written as

$$H = \underset{m}{E}\big(H(n,m)\big) = \sum_{m=1}^{M}\theta_m \cdot \Big(\prod_{i=1}^{n}(1-\overline{P}_m)\Big).$$

Search delay. The cost of waiting time occurs at the supernodes. In a hybrid supernode P2P network structure, all content requests not found on the request node will be forwarded to the local supernode. The local supernode responds with the transmission information of nodes in the same community that have the requested content, or, if none exist, forwards the unsatisfied requests to other interconnected supernodes in remote communities. Therefore, the search delay includes the expected total system waiting time served at local and remote supernodes, given certain network topology and request forward protocol. Since content request from each peer node is a Poisson process, the aggregate content (search) request arrival at the supernode is also a Poisson process whose arrival rate, Λ_0, is the accumulation of all the arrivals, $\Lambda_0 = n\cdot\sum_{m=1}^{M}(1-P_m)\cdot\theta_m\cdot\lambda$. Using a queueing model, we can express the search delay of an isolated community as, $S = (\mu_0 - \Lambda_0)^{-1}$, where μ_0 is service rate of the supernode.

Provision delay. The provision delay occurs at the provision node, due to congestion. One provision node may be serving several requests (providing download) simultaneously, so the expected provision delay is estimated from the aggregated content request rate and the process capacity of the provision nodes. Specifically, we define a binary random variable, $Z(i,j,m)$, which is 1 if and only if j is selected as the provision node for file F_m requested from node i. Hence, the probability that node j is selected as the provision node for file F_m is

$$\begin{aligned}&\Pr\big(Z(i,j,m)=1\big)\\ &\quad = \Pr(X_{i,m}=0;\overline{X}_{j,m}=1;T_{i,j}<T_{i,k},\forall k\neq i,j,\text{s.t.}\overline{X}_{k,m}=1).\end{aligned}$$

Assuming that all nodes are identical, we find

$$\begin{aligned}&\Pr\big(Z(i,j,m)=1\big)\\ &\quad = (1-P_m)\cdot\overline{P}_m\cdot\sum_{k=0}^{n-2}\frac{1}{k+1}\cdot\binom{n-2}{k}\cdot\overline{P}_m^{\,k}\cdot(1-\overline{P}_m)^{n-2-k}\\ &\quad = (1-P_m)\cdot\big(1-(1-\overline{P}_m)^{n-1}\big)/(n-1).\end{aligned}$$

Therefore, the aggregate request arrival rate at peer node j (given that j shares F_m) can be obtained as,

$$\begin{aligned}\Lambda_j &= \sum_{i\neq j}\sum_m \lambda_{i,m}\cdot Z(i,j,m)\\ &= \sum_{m=1}^{M}\theta_m\cdot(1-P_m)\cdot\big(1-(1-\overline{P}_m)^{n-1}\big)\cdot\lambda.\end{aligned}$$

Finally, the expected provision delay can be written as, $D = (\mu - \Lambda_j)^{-1}$.

Transmission delay. This delay is estimated by comparing the "network distance" from locations of the provision node to the request node. It depends strongly on the number of active nodes, because the more the active nodes the more likely it is to find a provision node closer to the request node. Let τ be the maximum transmission delay ($\tau = \rho\cdot\pi$) and denote the expected minimum transmission delay among k nodes by $T_{\min(k)}$. Using order statistics, we have,

$$T_{\min(k)} = \int_0^{\tau} t\cdot k\cdot\big(1-F(t)\big)^{k-1}\cdot f(t)\cdot dt = \frac{\tau}{k+1}\,,$$

where $f(t)$ and $F(t)$ are the PDF and CDF for the transmission delay, which has a uniform distribution. Next, $T_{\min}(n,m)$, the expected transmission delay for file F_m in the network with n nodes can be evaluated as

$$\begin{aligned}T_{\min}(n,m) &= E(T_{\min(k)} \mid F_m \text{ is available})\\ &= \frac{1}{H(n-1,m)}\sum_{k=1}^{n-1}T_{\min(k)}\binom{n-1}{k}\overline{P}_m^{\,k}(1-\overline{P}_m)^{n-1-k}\tau.\end{aligned}$$

After some simplifications, we have,

$$T_{\min}(n,m) = \frac{1-(1-\overline{P}_m)^{n-1}\cdot\big(1+(n-1)\cdot\overline{P}_m\big)}{\big(1-(1-\overline{P}_m)^{n-1}\big)\cdot n\cdot\overline{P}_m}\cdot\tau.$$

We further average the above expression over m to obtain the expected transmission delay,

$$T = \underset{m}{E}\big(T_{\min}(n,m)\big) = \sum_{m=1}^{M}\theta_m\cdot T_{\min}(n,m)\,.$$

IV. Performance Analysis

All the performance metrics proposed above are inherently relevant to the scale of the network. We first investigate the scale effects in this section, and then examine the effects from the network parameters in next section. All the proofs can be found in the appendix.

A. Analysis of Scale Effects

Using the assumption described above, we have the following results for the performance of a P2P network when its scale (size) changes. Table II summarizes network scale effects on the performance.

Proposition 1. (Content availability) *The expected content availability concavely increases with the scale of the P2P networks.*

A larger P2P network will improve the content availability but the benefit is marginally diminishing. Besides content availability, larger P2P networks also results in more expected content replicas. For example, the average number of file F_m in the community with n nodes can be described as,

$$R(n,m) = \sum_{k=1}^{n} k \cdot \binom{n}{k} \cdot \overline{P}_m^{\,k} \cdot (1-\overline{P}_m)^{n-k} = n \cdot \overline{P}_m$$

Therefore, the average number of replicas of a file increases with network size.

PROPOSITION 2. (SEARCH DELAY) *The search delay convexly increases with the size of the P2P networks.*

A larger P2P network will incur a higher search delay. The search cost displays diseconomy of scale. Given a limited capacity, the search delay at community center is the bottleneck of a large scale of P2P networks.

PROPOSITION 3. (PROVISION PROCESS DELAY) *The expected provision process delay increases with the size of the P2P networks.*

Larger P2P networks will incur higher provision delay. Moreover, the cost is marginally decreasing when provision capacity μ, or network size *n* is large enough. The aggregated request arrival rate for a certain file F_m at peer node *j*, $\Lambda_j(n.m)$, is written as

$$\Lambda_j(n,m) = (1-P_m) \cdot H(n-1,m) \cdot \lambda_m \,.$$

This indicates that a larger number of active peers in the networks make higher content availability, result in a higher aggregated request arrival rate, and consequently cause higher provision process congestion. However, $\Lambda_j(n,m)$ is bounded by $(1-P_m)\cdot\lambda_m$. Therefore, provision delay is bounded. This suggests that provision delay is not the critical factor that limits scalability.

PROPOSITION 4. (TRANSMISSION DELAY) *The expected transmission delay convexly decreases with the size of the P2P networks.*

Larger P2P networks will reduce the expected transmission delay, but the benefit is marginally decreasing. We can rewrite $T_{\min}(n,m)$ as,

$$T_{\min}(n,m) = \left(\frac{1}{R(n,m)} - \frac{(n-1)\cdot\big(1-H(n-1,m)\big)}{n \cdot H(n-1,m)} \right) \cdot \tau \,.$$

As investigated in Proposition 1 that there will be more content replicas as more peer nodes join in the community. In a traditional client-server network structure, there is no extra benefit if more than one content replica are cached at the same content server. However, in P2P file sharing networks, higher degree of content replicas indicates that better transmission performance may be achieved, by selecting a closer provision node with less transmission time.

TABLE II
EFFECTS OF SCALE ON THE P2P NETWORK PERFORMANCE METRICS

	H	S	D	T
Δ_n	+	+	+	–
Δ_n^2	–	+	*	+

* : – if μ or *n* is large enough.

B. *Analysis of Parameter Effects*

The impacts of system parameters on various performance measures are described in Proposition 5 and summarized in Table III.

PROPOSITION 5.

(CONTENT INTENSITY) *Higher content intensity (higher β) always yields better performance in content availability, transmission, and search. Higher β improves the provision performance only when the community size is sufficiently large.*

(SHARING LEVEL) *Higher sharing ratio (higher γ) improves the performance in content availability and transmission, but increases provision delay.*

(REQUEST FREQUENCY) *Higher content request rate (higher λ) results in larger provision and search delays.*

(UPLOAD CAPACITY) *Higher provision capacity (higher μ) reduces provision delay.*

(POSITION DISPERSION) *Transmission delay decreases with higher degree of location dispersion (higher ρ) among P2P users.*

TABLE III
IMPACTS OF SYSTEM PARAMETERS ON THE P2P NETWORK PERFORMANCE METRICS

	∂H	∂T	∂D	∂S
$\partial\beta$	+	–	+/–	–
$\partial\gamma$	+	–	+	0
$\partial\lambda$	0	0	+	+
$\partial\mu$	0	0	–	0
$\partial\rho$	0	+	0	0
$\partial\alpha$	–/+	–	+/–	–

We find that a higher variation of popularity among the files, or a higher α, results in better transmission and search performance. However, the effect of popularity variation (α) on content availability and provision delay can be negative or positive. Figure 2 shows that content availability decreases with α if α is below a critical value α^*. It can be observed (in Figure 3) that α^* becomes larger when there are more nodes (higher *n*) or smaller number of file varieties *M*. Similarly, provision delay is found to decrease with α only when α is greater than a threshold value.

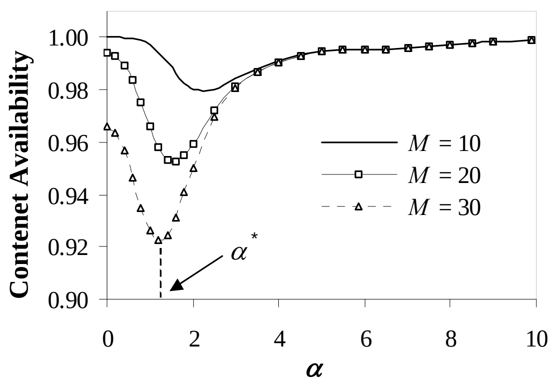


Fig. 2. Effect of α on content availability.

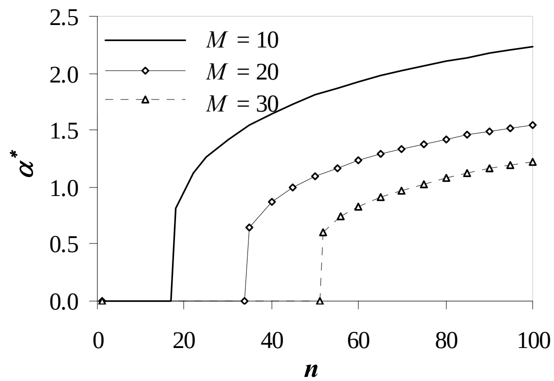


Fig. 3. Effect of n, M on critical value α^*.

C. Simulation Validation

The simulation validation is based on our analytical model, but we also relax the assumptions that content distribution across files, or request behaviors among users, is identical. For each simulation run, we generate 10,000 content requests. The simulations are repeated under different network sizes and various parameters such as the degree of variety of Zipf content distribution (α), the intensity of content availability (β), content request arrival rate (λ), and the number of file varieties (M). Typical parameter values are given as

$$\alpha = 1,\ \beta = 1,\ \gamma = 1,\ \lambda = 1,\ M = 10,\ \mu = 10,\ \tau = 1\ .$$

The simulation results are consistent with our analytical model. Larger networks have higher content availability (hit rate) and lower transmission delay. However, higher hit rate results in more file transfer activities, and hence increases the expected process delay.

Figure 4 shows that the expected provision process waiting time is stable with the network size when the request rate λ is low. When λ is high, serious congestion would occur at some peer nodes and the expected waiting time increases significantly.

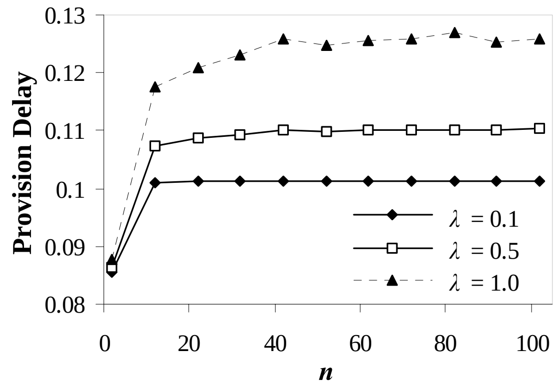
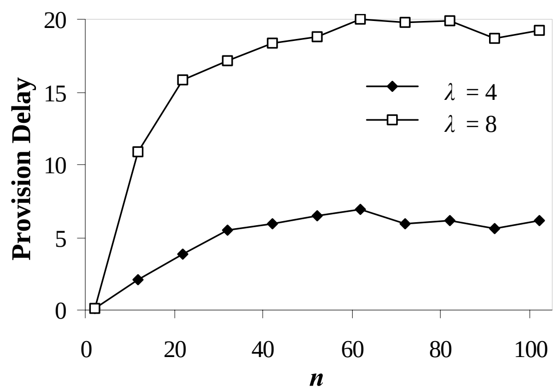


Fig. 4. Effect of λ on provision delay $D(n)$.



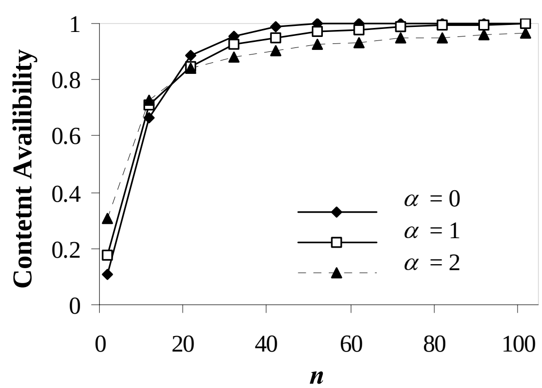


Fig. 5. Effect of α on content availability $H(n)$ and transmission delay $T(n)$.

Figure 5 plots the effect of Zipf distribution coefficient α. Higher α results in lower content availability when the community size is large, however, content availability decreases with α when n is smaller. The results are consistent with the analytical observation, described in Section 4.2, that the critical value α^* increases with community size, and content availability decreases (increases) with α if α is below (above) α^*. Higher α always results in lower transmission delay.

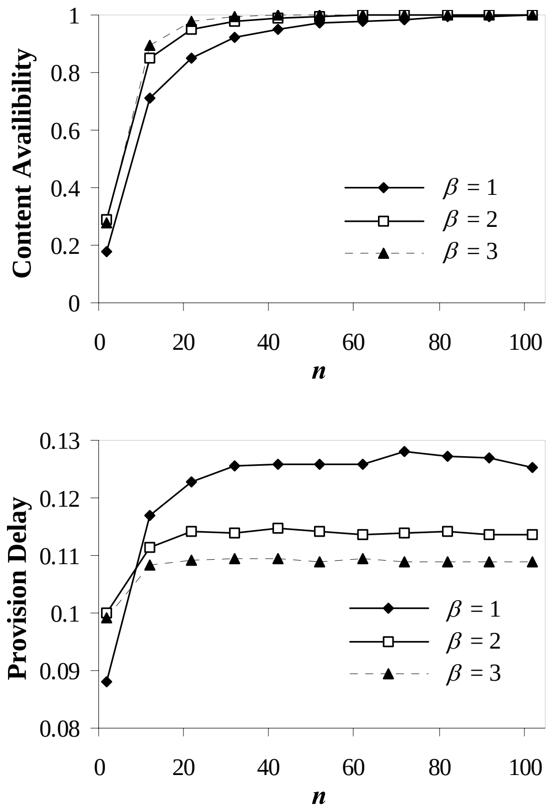
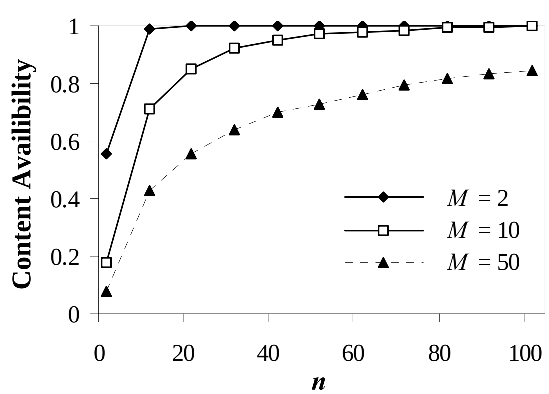


Fig. 6. Effect of β on content availability $H(n)$ and provision delay $D(n)$.

Figure 6 shows that higher β yields higher content availability. This would induce higher request rate at provision nodes. However, at the same time, higher β also indicates lower probability that a node needs content from other nodes, and consequently reduces request rate at provision nodes. The overall impact on provision delay is therefore determined by these two competing factors. When n is large, the impact of β on content availability becomes less significant, so higher β turns out to reduce provision delay. The opposite is true when the community size is small.



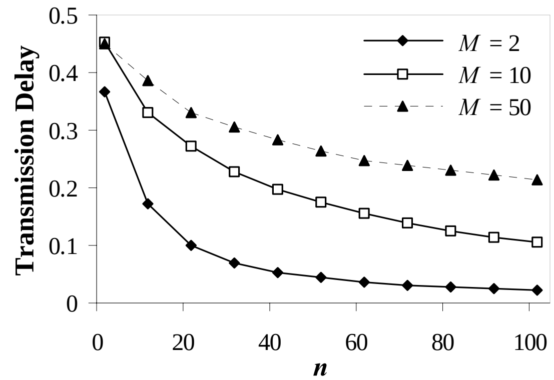


Fig. 7. Effect of M on content availability $H(n)$ and transmission delay $T(n)$.

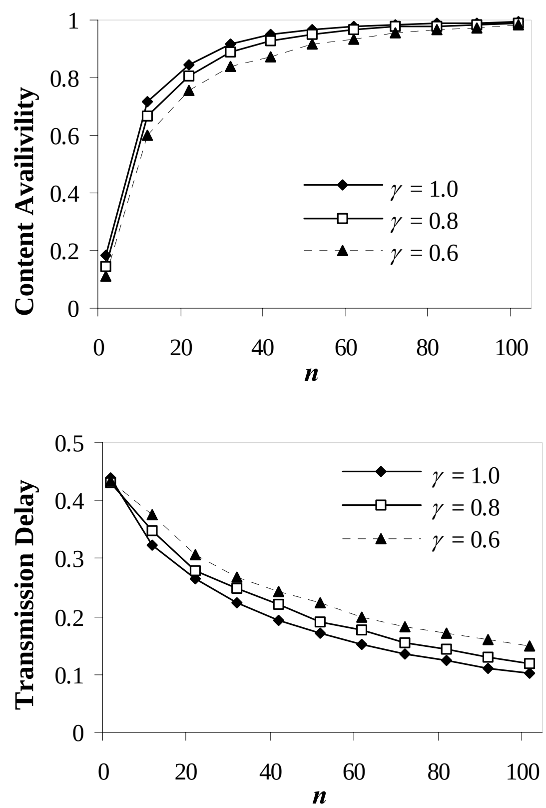


Fig. 8. Effect of γ on content availability $H(n)$ and provision delay $D(n)$.

Figures 7 shows that a smaller M (lower number of files) results in a higher hit ratio and lower expected transmission delay. As is shown in Figures 8, a higher sharing level γ always results in higher content availability and consequently lower expected transmission delay.

V. The Optimal Community Size

In the previous sections, we investigate the impacts of number of active users in a P2P network on various performance metrics. Our analyses show that these performance measures often display opposite behaviors. This suggests possible tradeoffs, which, if balanced, can determine a network size that optimizes the overall utility of a file sharing P2P community. Next, we examine two operational decisions: the optimal sizing (number of peers) for a given supernode in this section, and the optimal grouping on multiple supernodes for a given number of peers, which will be discussed in next section.

From the scale effect analysis, we know that search activity at the supernode is the performance bottleneck. Therefore the maximum size for an isolated P2P community is bounded, due to the fact that search delay increases convexly with number of P2P users. The optimal scale is determined so as to maximize the expected utility of a network, which depends on content availability value function and delay cost functions. Let *V*(.) be the value function of content availability, and *C*(.) the cost function of waiting times. *C*(.) is assumed to be convex, while the value function *V*(.) is concave in content availability. The optimal community size can be obtained as,

$$n^* = \arg\max_n V\big(H(n)\big) - C\big(T(n) + D(n) + S(n)\big)$$

<u>*Numerical results.*</u> In the following, we present some numerical results. The value and cost functions are assume to be linear, that is,

$$U = v \cdot H(n) - c \cdot \big(T(n) + D(n) + S(n)\big).$$

Typical parameter values are $\alpha = 0, \beta = 1, \gamma = 1, \lambda = 1, M = 10$, $\mu = 10,\ \mu_0 = 100,\ \tau = 0.1,\ v/c = 5$. We vary the values of parameters to investigate the impacts of parameters on the community size decision. Table IV summarizes the direction of the changes in optimal community size n^* with respect to various system parameters.

TABLE IV
DIRECTION OF CHANGE FOR OPTIMAL COMMUNITY SIZE n^*

α	β	γ	λ	M	μ_j	μ_0	v	ρ
+/–	+/–	–	–	+/–	+	+	+	+

It is intuitive that the optimal community size n^* increases with the capacities of the supernode and the peer nodes, but decreases with content request frequency λ (Figure 9). Figure 10 shows n^* decreases with the sharing level of P2P users. It is because, when the users are more willing to share their contents, a smaller number of P2P users are required to achieve the equivalent content availability level. Higher content intensity (average number of local cached contents) β also results in smaller n^* when γ is high. As investigated in Section 4.2, if the content sharing level is low ($\gamma < 0.2$), higher β will result in less provision delay, and therefore a larger P2P community can be operated.

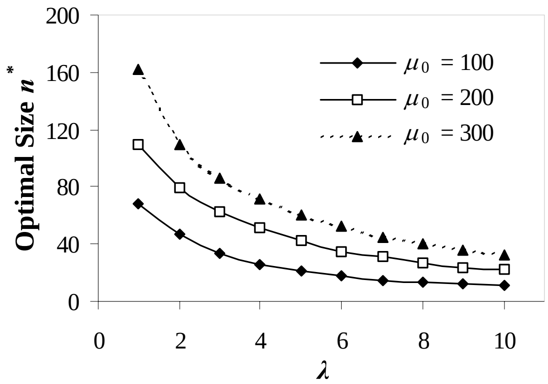


Fig. 9. Effect of λ and μ_0 on optimal size.

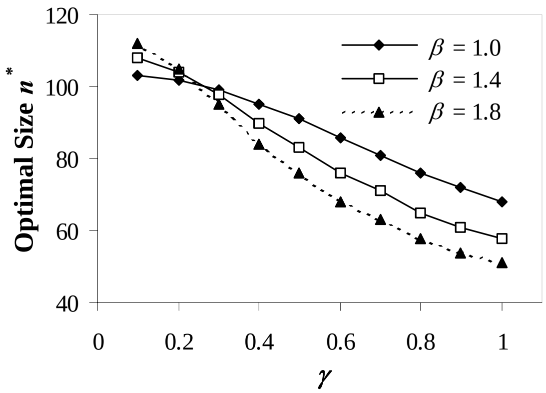


Fig. 10. Effect of γ and β on optimal size.

If the value of content increases relative to the cost of waiting, the community is motivated to achieve a higher content availability by increase its size. If the upper bound of transmission delay τ (or dispersion of user's locations ρ) is high, one trends to operate a larger community to reduce the expected transmission delay (shown in Figure 11). Figure 12 shows that the optimal size increases with the variation degree of content popularity α and the number of content varieties *M*, when α is high (greater than 0.2). The community becomes larger with *M* when α is less than that critical value. This observation is consistent with the previous findings for the property of the critical value α^*.

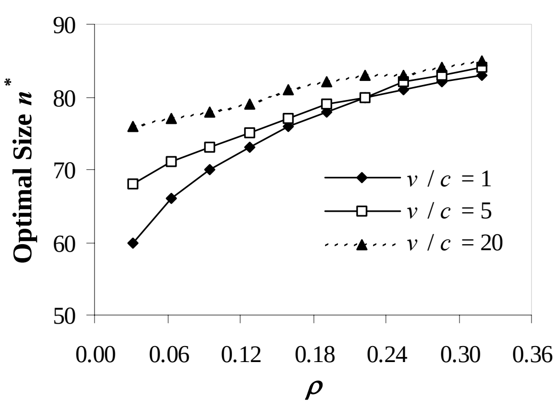


Fig. 11. Optimal size vs. ρ and v/c.

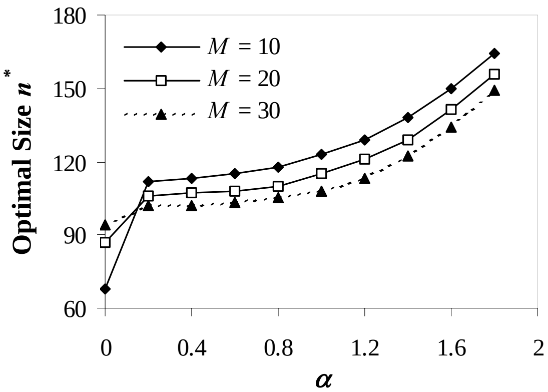


Fig. 12. Effect of α and M on optimal size.

VI. Interconnection and Grouping Decision

Results from the above performance analysis indicate the size of P2P networks is mainly bounded by the search capacity of a supernode, and a larger network would also result in a higher provision delay (upload activity). To improve the search and consequent overall performances, one can invest to expand the capacity of a centrally operated supernode. However, in reality, the search performance is usually enhanced by leveraging multiple decentralized supernodes. Assuming that each user is connected to only one supernode, a larger number of supernodes will segment the entire population of users into groups with smaller sizes. Small group size reduces search delay, but at the same time, lowers content availability and degrades transmission performance. To achieve certain quality level of content availability, unsatisfied content requests are forwarded to interconnected supernodes. Certainly, this will impose extra search load on the interconnected supernodes. It would be interesting and important to examine various grouping approaches so as to identify the best operational performance.

The performances of P2P networks with multiple supernodes are strongly associated with interconnection structures. To get managerial insights, we analyze several specific symmetric structures. Similar to regular peer nodes, we assume that supernodes are uniformly located in the same domain, with identical capacity. Each node is assumed to be connected to one of the supernodes randomly, such that each supernode will serve equal number of nodes. In the next section, we will discuss the situation where nodes are connected to the closest supernode.

A. Interconnection Structure

The performance of a network depends on its interconnection structure. Smaller degree of interconnection renders a larger transmission delay, and also requires higher TTL (time-to-live, or number of hops on supernodes each request is allowed) to achieve an equivalent content availability level. Higher degree of interconnection improves the performance of transmission, but at the cost of higher search load imposed on interconnected supernodes. To illustrate the interconnection performance, we analyze the operations of three specific symmetric network structures, as depicted in Figure 13.

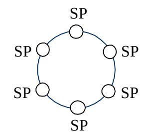
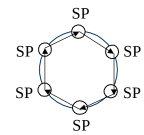
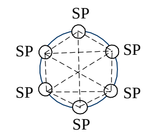
	Isolated	Chained	Complete
			
Out Degree	0	1	No. of Supernodes – 1
TTL	0	No. of Supernodes – 1	1

Fig. 13. Symmetric P2P network structures.

Isolated structure. There is no interconnection among the supernodes. To alleviate search delay, P2P users are partitioned into several isolated groups. The out-degree for each supernode is 0, as is the TTL (time-to live). Isolated structure has best search delay, but the worst hit rate and transmission performance.

Chained structure. All supernodes are connected, but each supernode is directly connected to exactly one supernode. Chained structure achieves full content availability. Content requests are searched in one group and, if not satisfied, forwarded to the next. The transmission performance is the same as that of the isolated structure. The out-degree for each supernode is 1, while the TTL is the total number of supernodes deployed less 1.

Complete structure. All supernodes are fully connected with each other. Content requests are forwarded to *all* interconnected groups at the same time if they are not satisfied by the current group. The complete structure has the best transmission performance, but the highest search load on other supernodes. The out-degree for each supernode is the total number of supernodes less 1, and the TTL is 1.

B. Performance Analysis

Utilizing the analytical metrics presented earlier, we formulate the metrics for evaluating the performance of P2P network with multiple supernodes, and investigate the scale effect of grouping with various interconnection structures. The analytical formulations are summarized in Table 5, where s is the total number of supernodes adopted. It is straightforward to show that the chained and complete structures have the same content availability and provision delay. The transmission delays for chained and isolated structures are also identical.

TABLE V
PERFORMANCES OF P2P NETWORK WITH MULTIPLE SUPERNODES

	$\overline{H}(n,s)$	$\overline{T}(n,s)$	$\overline{D}(n,s)$	$\overline{S}(n,s)$
ISOLATED	$H(n/s)$	$T(n/s)$	$D(n/s)$	$S(n/s)$
CHAINED	$H(n)$	$T(n/s)$	$D(n)$	$\overline{S}_{\text{chained}}(n,s)$
COMPLETE	$H(n)$	$\overline{T}_{\text{complete}}(n,s)$	$D(n)$	$\overline{S}_{\text{complete}}(n,s)$

In Table V, the expected search delay for chained structure is,

$$\overline{S}_{\text{complete}}(n,s) = \begin{cases} \dfrac{1-(1-H(n/s))^s}{H(n/s)\cdot(\mu_0-(2-H(n/s))\cdot(n/s)\cdot\lambda)}, & \text{if } s>1; \\ \dfrac{1}{\mu_0-n\cdot\lambda}, & \text{if } s=1. \end{cases}$$

Here, each group has a probability of $1-H(n/s)$ to forward a request to its adjacent group. This will give each group an overall request rate of $(1+(1-H(n/s)))(n/s)\lambda$. One request has a probability of $1-(1-H(n/s))^s$ to be satisfied. If satisfied, on average, one request is forwarded $1/H(n/s)$ times. Similarly, for complete structure, the expected search delay can be written as,

$$\overline{S}_{\text{complete}}(n,s) = \frac{2-H(n/s)}{\mu_0-(1+(s-1)\cdot(1-H(n/s)))\cdot(n/s)\cdot\lambda},$$

while the expected transmission delay is,

$$\overline{T}_{\text{complete}}(n,s) = (H(n/s)\cdot T(n/s)+(H(n)-H(n/s))\cdot T(n-n/s))/H(n).$$

Numerical results. We numerically evaluate the impacts of number of groups, s, on search and transmission performance with various network structures. The parameter values are set as $\alpha=0$, $\beta=1$ M=10, n=100, λ=1, μ_j=10, μ_0=100, and $\tau=1$. Figure 14 shows that chained structure always outperforms complete structure in search, but complete structure always has better transmission performance than chained structure. Interestingly, the expected search delay is a convex function of number of groups for both chained and complete structures. Therefore, there exists optimal number of groups that would minimize the expected search delay in these two structures. Regarding the expected transmission delay, as expected, small number of groups is preferred for all structures, though for complete structure large number of groups may perform equally well.

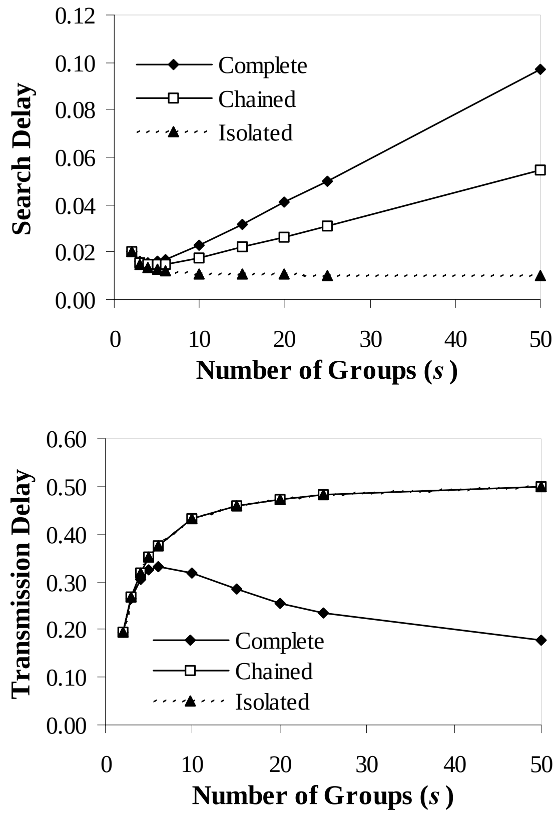


Fig. 14. The impacts of number of groups on expected search and transmission delays.

C. Location Based Grouping (LBG) vs. Random Based Grouping (RBG)

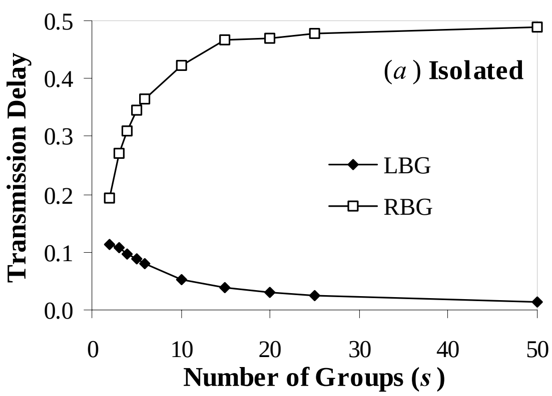
In the previous section, each peer node is assumed to be randomly connected to one of the supernodes. This policy is adopted by most of the current P2P technologies, such as Gneutella v0.4 protocol. Here, we propose a new scheme where each peer node is served by the *closest* supernode. The policy can be realized, for instance, in the case where peer nodes are served by a supernode provided by their local ISPs. The advantage of this scheme is the significant improvement of transmission performance as files are downloaded from nearby peer nodes.

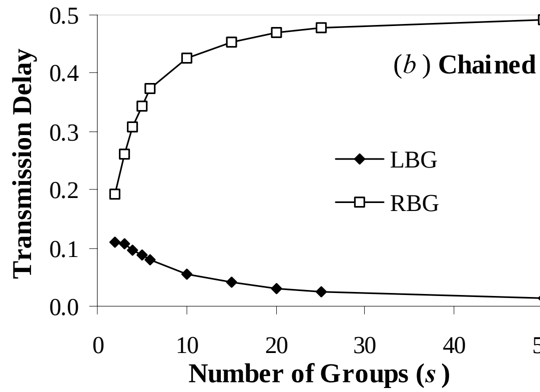
It would be interesting and useful to compare the impacts of number of groups on transmission performances of random based grouping (RBG) and location based grouping (LBG). We assume that there are multiple supernodes (groups), i.e. $s \geq 2$, serving the entire P2P user population. A request node (located at position X) and a provision node (located in position Y) are located in separate groups that are an h-hop distance apart. X and Y are random variables drawn from uniform distributions, $X \sim U(0, 2\rho\pi/s)$ and $Y \sim U(2\rho\pi h/s, 2\rho\pi(h+1)/s)$. If h = 0, both nodes are served by the same supernode. The case where h = 1 indicates that these two nodes are served by different groups with direct interconnection. Let $\hat{F}(t\,|\,h,s)$ denote the CDF for the expected transmission delay between any two nodes in various groups with h-hop distance. $\hat{F}(t \mid h,s)$ is defined as $P\{|X-Y| \le t\}$. Using a convolution of distributions, we have the resulting PDF for transmission delay $\hat{f}(t \mid h,s)$ as,

$$\hat{f}(t \mid h=0, s \ge 2) = \frac{s}{\rho\pi} - \frac{s^2 t}{2\rho^2\pi^2},\ 0 \le t \le 2\rho\pi / s;$$

$$\hat{f}(t \mid h \ge 1, s \ge 2) = \begin{cases} 0, & \text{if } 0 \le t < 2\rho\pi(h-1)/s; \\ \dfrac{(1-h)s}{2\rho\pi} + \dfrac{s^2 t}{4\rho^2\pi^2}, & \text{if } 2\rho\pi(h-1)/s \le t < 2\rho\pi h/s; \\ \dfrac{(1+h)s}{2\rho\pi} - \dfrac{s^2 \cdot t}{4\rho^2\pi^2}, & \text{if } 2\rho\pi h/s \le t < 2\rho\pi(h+1)/s; \\ 0, & \text{if } t \ge 2\rho\pi(h+1)/s. \end{cases}$$

The transmission performance (or the expected minimum transmission delay) of P2P network is determined by number of candidate provision nodes. Given that k nodes share the file requested, the minimum transmission delay (using k^{th} order statistics) of LBG is $2\rho\pi / s(2k+1)$, which is strictly superior to the performance achieved in RBG, $\rho\pi/(k+1)$. When the total number of P2P users is sufficiently large, the number of provision nodes can be approximated as $k \approx n \cdot p / s$, where p is the probability that a peer node user share the requested file. The expected minimum transmission delays become $\rho\pi/(np + s/2)$ for LBG, and $\rho\pi/(np/s+1)$ for RBG respectively. Given a fixed number of P2P users, it can be observed that the transmission performance of RBG worsens as the number of groups increases, while the performance of LBG improves with number of groups.





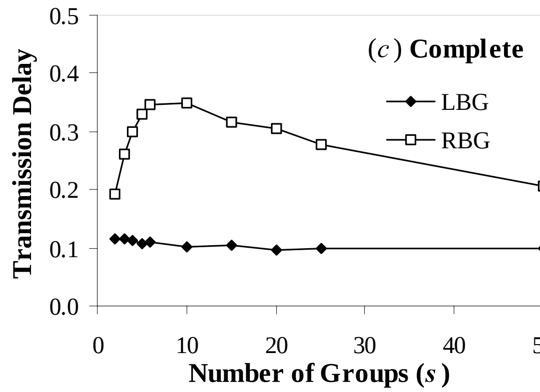


Fig. 15. The impacts of number of groups and LBG on transmission delay.

Next, we numerically investigate the impacts of number of groups on the transmission delay for various interconnection structures, using the same parameter value setting in Section 6.2. Figure 15 shows that, consistent with intuition, the performance of LBG is better than that of RBG, particularly as the number of groups increase. The transmission delays of RBG increase with number of groups for isolated and chained structures. On the contrary, the transmission delay decreases with number of groups when LBG is adopted. This decrease is more significant for isolated and chained structures.

VII. Summary and Conclusions

In this paper, we have developed an analytical model to evaluate the impacts of network scale and system parameters on the performance of P2P networks. Our analytical and simulation results show positive scale effects on content availability and transmission delay, and negative scale effects on provision and search delays. Furthermore, content availability, transmission delay, and provision delay are found to marginally decrease, while search delay is marginally increasing. This suggests that search congestion, rather than provision, is the primary factor that restricts network scalability. Balancing these performance measures, an optimal size is determined to maximize the overall utility of the P2P community. The optimal community size increases with P2P participant capacity, ratio of content value to waiting cost, as well as location dispersion degree of P2P users, but decreases with content request frequency, and content sharing level of

P2P users. Higher degree of variation for content popularity distribution can result in either larger or smaller community size, depending on the number of content varieties.

We further extend the analytical model to examine grouping decision in the networks with symmetric interconnection structures. Comparing the performances between chained and complete structures, we find that chained structure has better search performance, but complete structure provides better transmission performance. There exists optimal grouping size that minimizes search delay for both chained and complete structures. The transmission delay in chained structure is found to increase with number of groups. Finally, we compare the impacts of location based grouping (LBG) and random based grouping (RBG) approaches on transmission delay. LBG has better performance, which even improves with number of groups.

In our model, we assume that nodes (regular peer nodes and supernodes) are symmetric and could be coordinated by the central planner. Investigating the emerging P2P structure under heterogeneous players (peers and super peers) with incentive compatibility is a planned future extension. It will be interesting to study how our results will change if the players of a P2P network are rational to decide the sharing, grouping and interconnection decision. In this paper, the performances are developed from the central operational perspective. It is important to investigate the performance loss due to decentralized decision.

Besides the investigation of factors that affect P2P networks operations, an interesting topic for future research is to study the evolution dynamics of the content distribution among peer nodes. As peer nodes exchange contents, popularity of various files changes. It would be important and interesting to study how these changes occur over time, and what impact this may have on the P2P network performance.

APPENDIX

A. Proof of Proposition 1

For the expected content availability in the network with *n* nodes, $H(n)$, we have,

$$\Delta_n H(n) = H(n) - H(n-1) = \sum_{m=1}^{M} \theta_m \cdot \left(1-\overline{P}_m\right)^{n-1} \cdot \overline{P}_m > 0;$$

$$\begin{aligned}\Delta_n^2 H(n) &= \Delta_n H(n+1) - \Delta_n H(n)\\ &= -\sum_{m=1}^{M} \theta_m \cdot \left(1-\overline{P}_m\right)^{n-1} \cdot \overline{P}_m^{\,2} < 0.\end{aligned}$$

B. Proof of Proposition 2

For the expected search delay in the network with *n* nodes, $S(n)$, we have,

$$\Delta_n S(n) = \frac{\sum_{m=1}^{M} (1-P_m)\cdot\theta_m\cdot\lambda}{\left(\mu_0 - \Lambda_0(n)\right)\left(\mu_0 - \Lambda_0(n-1)\right)} > 0;$$

$$\Delta_n^2 S(n) = \frac{2\cdot\left(\sum_{m=1}^{M} (1-P_m)\cdot\theta_m\cdot\lambda\right)^2}{\left(\mu_0 - \Lambda_0(n+1)\right)\left(\mu_0 - \Lambda_0(n)\right)\left(\mu_0 - \Lambda_0(n-1)\right)} > 0.$$

C. Proof of Proposition 3

For the expected provision process delay, $D(n)$, we have,

$$\Delta_n D(n) = \frac{\sum_{m=1}^{M} (1-P_m)\cdot(1-\overline{P}_m)^{n-2}\cdot\overline{P}_m\cdot\theta_m\cdot\lambda}{\left(\mu - \Lambda(n)\right)\left(\mu - \Lambda(n-1)\right)} > 0;$$

$$\begin{aligned}\Delta_n^2 D(n) = \sum_{m=1}^{M} &\frac{(1-P_m)\cdot\overline{P}_m\cdot\theta_m\cdot\lambda\cdot(1-\overline{P}_m)^{n-2}}{\left(\mu - \Lambda_j(n+1)\right)\left(\mu - \Lambda_j(n)\right)\left(\mu - \Lambda_j(n-1)\right)}\\ &\times\left(-\overline{P}_m\cdot\mu - (1-\overline{P}_m)\cdot\Lambda_j(n-1) + \Lambda_j(n+1)\right).\end{aligned}$$

When μ or n is sufficiently large so that $\mu > \left(\Lambda_j(n+1) - (1-\overline{P}_m)\cdot\Lambda_j(n-1)\right)/\overline{P}_m$, $\Delta_n^2 D(n) < 0$.

D. Proof of Proposition 4

$$\Delta_n T(n,m) = \frac{(1-\overline{P}_m)^{n-2}(n-1)^2\,\overline{P}_m^{\,2} - \left(1-(1-\overline{P}_m)^{n-1}\right)^2}{n(n-1)\overline{P}_m\left(1-(1-\overline{P}_m)^{n-1}\right)\left(1-(1-\overline{P}_m)^{n-2}\right)}\tau.$$

We need to find the sign of the expression, $(1-\overline{P}_m)^{\frac{n}{2}-1}(n-1)\overline{P}_m - 1 + (1-\overline{P}_m)^{n-1}$. It is straightforward to show that this expression is a strictly decreasing function of $\overline{P}_m$, with a maximum value of 0 at $\overline{P}_m = 0$. Since $\overline{P}_m > 0$, $\Delta_n T(n,m) < 0$. Similarly, we can show that $\Delta_n^2 T(n,m) > 0$.

E. Proof of Proposition 5

1) With respect to β

$$\partial H(n,m)/\partial\beta = n\cdot\left(1-\overline{P}_m\right)^{n-1}\cdot\theta_m\cdot\gamma > 0;$$

$$\begin{aligned}&\frac{\partial T_{\min}(n,m)}{\partial\beta}\\ &\quad= -\left(\frac{\theta_m\cdot\gamma}{n\cdot\overline{P}_m^{\,2}} + \frac{n-1}{n}\cdot\left(\frac{(n-1)\cdot\theta_m\cdot\gamma\cdot\left(1-\overline{P}_m\right)^{n-2}}{\left(1-\left(1-\overline{P}_m\right)^{n-1}\right)^2}\right)\right)\cdot\tau < 0;\end{aligned}$$

$$\begin{aligned}&\frac{\partial\Lambda_j(n,m)}{\partial\beta}\\ &\quad= -\theta_m\left(1-\left(1-\overline{P}_m\right)^{n-1}\right) + (1-P_m)(n-1)(1-\overline{P}_m)^{n-2}\theta_m\gamma.\end{aligned}$$

If $-\left(1-\left(1-\overline{P}_m\right)^{n-1}\right) + (1-P_m)\cdot(n-1)\cdot(1-\overline{P}_m)^{n-2}\cdot\gamma < 0$, which is satisfied when n is sufficiently large or γ sufficiently small, we have,

$$\frac{\partial D(n,m)}{\partial\beta} = \frac{\partial\Lambda_j(n,m)/\partial\beta}{(\mu_j - \Lambda_j(n,m))^2} < 0$$

$$\partial\Lambda_0/\partial\beta = -\theta_m^2\cdot n\cdot\lambda < 0$$

$$\frac{\partial S}{\partial \beta} = \frac{\partial \Lambda_0 / \partial \beta}{(\mu_0 - \Lambda_0)^2} < 0 .$$

2) With respect to γ

$$\partial H(n,m) / \partial \gamma = n \cdot \left(1 - \overline{P}_m\right)^{n-1} \cdot P_m > 0 ;$$

$$\frac{\partial T_{\min}(n,m)}{\partial \gamma} = -\left(\frac{P_m}{n\overline{P}_m^{\,2}} + \frac{n-1}{n} \left(\frac{(n-1)P_m \left(1-\overline{P}_m\right)^{n-2}}{\left(1-\left(1-\overline{P}_m\right)^{n-1}\right)^2} \right) \right) \tau < 0 ;$$

$$\frac{\partial \Lambda_j(n,m)}{\partial \gamma} = (1 - P_m) \cdot \lambda_m \cdot (n-1) \cdot \left(1 - \overline{P}_m\right)^{n-2} \cdot P_m > 0 ;$$

$$\frac{\partial D(n,m)}{\partial \gamma} = \frac{\partial \Lambda_j(n,m) / \partial \gamma}{\left(\mu_j - \Lambda_j(n,m)\right)^2} > 0 .$$

3) With respect to λ

$$\partial \Lambda_j(n,m) / \partial \lambda = (1 - P_m) \cdot \left(1 - (1 - \overline{P}_m)^{n-1}\right) > 0 ,$$

Hence

$$\partial D(n,m) / \partial \gamma = \left(\partial \Lambda_j(n,m) / \partial \lambda\right) / \left(\mu_j - \Lambda_j(n,m)\right)^2 > 0 .$$

It is straightforward to show that

$$\partial \Lambda_0 / \partial \lambda = \sum_{m=1}^{M} \theta_m \cdot (1 - P_m) \cdot n > 0 \text{, AND } \partial S / \partial \lambda > 0 .$$

4) With respect to μ

$$\partial D(n,m) / \partial \mu_j = -\left(\mu_j - \Lambda_j(n,m)\right)^{-2} < 0 .$$

REFERENCES

[1] Adar, E. and B. A. Huberman, “Free-riding on Gnutella,” *First Monday 5(10).*

[2] Asvanund, A., K. Clay, R. Krishnan, and M. D. Smith, “An Empirical Analysis of Network Externalities in P2P Music-Sharing Networks,” *23rd Annual International Conference On Information Systems (ICIS’02)*, Barcelona, Spain, December 15-19, 2002.

[3] Asvanund, A., S. Bagla, M. H. Kapadia, R. Krishnan, M. D. Smith, and R. Telang, “Intelligent Club Management in Peer-to-Peer Networks,” *Workshop on Economics of Peer-to-Peer Systems*, June 5-6, Berkeley, CA, 2003.

[4] Butler, B. S. “Membership Size, Communication Activity, and Sustainability: A resource- Based Model of Online Social Structure,” *Information Systems Research*, Vol. 12, No. 4, December 2001.

[5] Dellarocas, C. and P. Resnick, “Online Reputation Mechanisms: A Roadmap for Future Research,” *Workshop on Economics of Peer-to-Peer Systems*, June 2003.

[6] Golle, P., K. Leyton-Brown, I. Mironov, and M. Lillibridge, “Incentives for sharing in peer-to-peer networks,” *Proceedings of ACM Conference on Electronic Commerce*, 2001.

[7] Jackson, M. O. and A. Wolinsky, “A strategic model of social and economic networks,” *Journal of Economic Theory*, 71:44-74, 1996.

[8] Khambatti, M., P. Dasgupta, and K. Ryu, “Peer-to-Peer Communities: Formation and Discovery,” *Fourteenth IASTED International Conference on Parallel and Distributed Computing and Systems*, Cambridge, 2002.

[9] Krishnan, R., M. D. Smith, Z. Tang, and R. Telang, “The Virtual Commons: Why Free riding Can Be Tolerated in File Sharing Networks,” *23rd Annual International Conference On Information Systems (ICIS’02)*, Dec 15-19, 2002.

[10] Krishnan, R., M. D. Smith, Z. Tang, and R. Telang, “The Impact of Free-Riding on Peer-to-Peer Networks,” *Proceedings of the 37th Hawaii International Conference on System Sciences*, 2004

[11] Kung, H. T. and C. H. Wu, “Differentiated Admission for Peer-to-Peer Systems: Incentivizing Peers to Contribute Their Resources,” *Workshop on Economics of Peer-to-Peer Systems*, June 5-6, Berkeley, CA, 2003.

[12] Ng, T. S. E. and H. Zhang, “Predicting Internet Network Distance with Coordinates-Based Approaches,” *Proc. of the 21th IEEE INFORM INFOCOM*, 2002, pp170-179.

[13] Ledlie, J., J. Taylor, L. Serban, and M. Seltzer, “Self-organization in Peer-to-Peer Systems,” *10th EW SIGOPS*, September 2002.

[14] Lee, J., “An End-User Perspective on File-Sharing Systems,” *Communications of the ACM*, February 2003, Vol. 46, No.2.

[15] Liebowitz, Stan. “Will MP3 downloads annihilate the Record Industry? The Evidence so Far,” *Advances in the study of Entrepreneurship, Innovation, and Economics Growth*, JAI Press, New York, 2003.

[16] Ranganathan, K., M. Ripeanu, A. Sarin, and I. Foster, “To Share or Not to Share: An Analysis of Incentives to Contribute in Collaborative File Sharing Environments,” *Workshop on Economics of Peer-to-Peer Systems*, June 5-6, Berkeley, CA, 2003

[17] Ratnasamy, S., P. Francis, M. Handley, R. Karp, and S. Shenker, “A scalable content-addressable network,” *Proceeding of ACM SIGCOMM*, August 2001.

[18] Ross, S. M., *Stochastic Processes*, Second Edition, John Wiley & Sons, Inc., 1996.

[19] Schoder, D. and K. Fischbach, “Peer-to-Peer Prospects,” *Communications of the ACM*, Feb. 2003, Vol. 46, No. 2.

[20] Schollmeier, R., “A Definition of Peer-to-Peer Networking for the Classification of Peer-to-Peer Architectures and Applications,” *Proceedings of the IEEE*, 2001.

[21] Shneidman, J. and D. C. Parke, “Rationality and Self-interest in Peer to Peer Networks,” *Proceedings of the Second International Workshop on Peer-to-Peer Systems (IPTPS’03)*, 2003.

[22] Singh, M. P., B. Yu, and M. Venkatraman, “Community-Based Service Location,” *Communications of the ACM*, Vol. 44 No.4, April 2001.

[23] Sningh, S., S. Ramabhadran, F. Baboescu, and A. C. Snoeren, “The Case for Service Provider Deployment of Supernodes in Peer-to-Peer Networks,” *Workshop on Economics of Peer-to-Peer Systems*, June 5-6, Berkeley, CA, 2003.

[24] Stoica, I., R. Morris, D. Karger, M. F. Kaashoek, and H. Balakrishnan, “Chord: A Scalable Peer-to-Peer Lookup Service for Internet Applications,” *Proceeding of ACM SIGCOMM*, August 2001.

[25] Yang, B. and H. Garcia-Molina, “Improving Efficiency of P2P Search,” *In the Proceedings of the 28th conference on Distributed Computing Systems*, July 2002.

[26] Yang, B. and H. Garcia-Molina, “Designing a Supernode Network,” *In Proceedings of the 19th International Conference on Data Engineering (ICDE)*, Bangalore, India, 2003.

[27] Yang, B. and H. Garcia-Molina, “Improving Search in Peer-to-Peer Systems,” In *Proceedings of the 22nd International Conference on Distributed Computing Systems (ICDCS)*, Vienna, Austria, July 2002.

[28] Wang Y. and J. Vassileva, “Trust and Reputation Model in Peer-to-Peer Networks,” *Third International Conference on Peer-to-Peer Computing (P2P'03)*, 2003.