The Impact of Active Measurement Techniques on Unstructured Peer-to-Peer Networks (Postprint)

Authors: Sha Ying

Date: 2017-03-10T00:00:00+00:00

Abstract

Active measurement technology has become an important fundamental means for studying the characteristic properties and user behavior of peer-to-peer networks. However, the impact of active measurement technology on peer-to-peer networks remains unknown, and there is currently a lack of detailed analysis and formal research in this regard. To address these issues, this paper first introduces the current development and status quo of peer-to-peer network measurement, analyzes the main challenging problems currently faced, and emphatically presents the preliminary progress of our research on the impact of active measurement technology on unstructured peer-to-peer networks. We first propose an active measurement network model, focusing on investigating the impact of active measurement technology on the degree distribution of peer-to-peer networks. The research findings indicate that the increase in active measurement network scale will lead to a phase transition phenomenon in degree distribution. Theoretical analysis and simulation experiments demonstrate that a small-scale high-"degree"active measurement network outperforms a large-scale low-"degree" active measurement network.

Full Text

Abstract

Active measurement has become a fundamental technique for studying the characteristic properties and user behaviors of peer-to-peer networks. However, the impact of active measurement on peer-to-peer networks remains unknown, and detailed analysis and formal research on this topic are still lacking. To address this issue, this paper first introduces the current development and status of peer-to-peer network measurement, analyzes the main challenging problems faced, and highlights the preliminary progress of our research on the impact of active measurement on unstructured peer-to-peer networks. We propose an active measurement network model and focus on studying the influence of active measurement on the degree distribution of peer-to-peer networks. Our research reveals that as the scale of the active measurement network increases, a phase transition phenomenon occurs in the degree distribution. Both theoretical analysis and simulation experiments demonstrate that a small-scale high-degree active measurement network is superior to a large-scale low-degree one.

Keywords: peer-to-peer network; active measurement; degree distribution

1 Introduction

Accurately describing the dynamic behavior of peer-to-peer (P2P) network topologies and analyzing their network performance provides valuable reference for both network operators and P2P protocol developers. A critical foundational task to achieve these objectives is precise network measurement.

Currently, there are two primary approaches for peer-to-peer network measurement: active and passive. Since passive measurement only applies to monitoring P2P traffic flowing through observation points, it cannot cover the entire peer-to-peer network. Moreover, this method is limited to traffic performance measurement and cannot obtain information about network search, routing, or application-layer transmission. Consequently, active measurement plays an irreplaceable role. Generally speaking, active measurement employs network crawlers that actively join peer-to-peer networks to acquire relevant network characteristics and node attributes, such as peer IP addresses, port numbers, and all metadata accessible through P2P protocols. Using active measurement, we can obtain three types of P2P behavioral information: (1) micro-behavioral characteristics, such as P2P network topology, latency, content availability, and upload/download ratios; (2) topological dynamic behavior features, such as changes in stable core overlay connectivity (the set of nodes in the top-level topology with online time exceeding a specific threshold), distributions of peer session times and download durations, and correlations between session times; and (3) churn behavior, i.e., the dynamic changes caused by users joining or leaving the P2P system.

In summary, active measurement has become a fundamental and irreplaceable method for studying P2P network properties such as topology and degree distribution, as well as user behavior characteristics. However, due to its inherent nature, active measurement inevitably impacts the original P2P network's topology, traffic, node strategy selection, and user behavior decisions. Currently, detailed analysis and formal research on these impacts are lacking. Therefore, quantitatively evaluating the accuracy of active measurement and subsequently designing calibration algorithms to reconstruct more accurate and comprehensive information about the measured P2P network are topics of significant research importance.

Our research focuses on the impact of active measurement on peer-to-peer networks. The specific approach is as follows: first, model active measurement nodes and the active measurement network; conduct theoretical analysis of their various impacts on P2P networks by studying the hybrid network formed by superimposing the active measurement network onto the P2P network to obtain formal results; propose active measurement calibration algorithms to acquire more authentic and accurate information about the measured P2P network; and finally, verify the effectiveness of the models and algorithms through simulation experiments and real P2P network environments. The ultimate goal is to establish unified evaluation criteria for active measurement technology through research on its impact on P2P networks and corresponding calibration techniques, making P2P-related research based on active measurement more authentic and reliable, and further guiding P2P network design and optimization. We have obtained preliminary results regarding P2P degree distribution. This research is supported by the National Natural Science Foundation of China (General Program, Grant No. 61070184) and was initiated in 2011.

2 Current Status of Peer-to-Peer Network Active Measurement Technology

This section first introduces current research hotspots in peer-to-peer network measurement, then discusses the remaining challenging issues in this field.

2.1 Research Status and Development Trends

Current research on peer-to-peer network measurement primarily focuses on the following aspects:

(1) **Measurement Data Acquisition Methods and Theoretical Modeling**
There is substantial work in this area, mainly concerning the design of measurement methods (such as network crawlers), analysis of measurement data, simulation of theoretical models, and measurement validation, as seen in references [17], [21], [30], [31], [32], etc. Saroiu et al. [17] pioneered measurement studies on the two most popular P2P systems at the time, Napster and Gnutella, using active network crawlers. Liang et al. [18, 19] conducted in-depth exploration and analysis of KaZaA system behavior, designing specialized crawlers and performing the first measurement and analysis of file pollution in KaZaA [20]. However, these studies lack theoretical analysis of measurement accuracy and related issues. They only mention in data analysis that crawler design affects results: due to the dynamic nature of P2P networks, shorter crawling times yield more accurate network snapshots, but excessively short times limit crawling scope. Stutzbach and Rejaie [28] studied snapshot accuracy issues in Gnutella network crawlers, but their approach primarily relied on actual measurements of how different crawling frequencies, speeds, and durations affect P2P topology, lacking theoretical analysis and mathematical models. Isda

et al. [23] noted that active measurement may send unexpected packets not anticipated in proper protocols, potentially causing issues at scale, such as triggering intrusion detection system alarms. They identified that measurement behavior could become anomalous and impact the system, but did not address how to resolve these system impacts.

Thus, although extensive measurements have been conducted on P2P networks, there is insufficient general attention to measurement process effectiveness and result accuracy, which affects the validity and authenticity of P2P network analysis based on measurement data.

(2) **Sampling Issues in Large-Scale Peer-to-Peer Networks**
Due to the large scale and dynamic nature of P2P systems, directly obtaining global behavioral data is unrealistic, making sampling an effective method for understanding these systems. The key challenge in sampling is how to obtain representative data and how to derive global properties from limited samples.

Stutzbach and Rejaie [1] focused on obtaining representative nodes for node attributes (such as degree, link bandwidth, and number of shared files), noting that P2P system sampling introduces two types of errors: (i) temporal (due to different node lifecycles) and (ii) topological properties (differences in node "degree"). Sampling tends to favor high-degree nodes. According to existing research [2,5], many nodes have very short lifespans while others persist for long periods. As sampling time increases, the sample set accumulates numerous short-lived nodes.

Saroiu et al. [4] employed breadth-first or depth-first approaches for node sampling, both of which bias toward short-lifecycle nodes. Another method is random walk, where a simple scheme involves walking randomly for length r, using the terminal node as a sample point, then walking another r, and so on. Although this method works well for certain graph types, its efficiency is low. According to graph theory, log [6][7] is recommended.

If peer-to-peer networks are mapped as graphs, some problems can leverage graph theory research results. Bollobás [8] and Jerrum [9] defined classes of graphs sharing certain properties (such as degree distribution) and then used specific random algorithms to generate all such graphs. Cooper et al. [10] used this approach to prove that their overlay reconstruction algorithm could produce well-behaved graphs. Krishnamurthy [11] and Stumpf [13] et al. generated representative subgraphs from a large graph through sampling that preserved original graph properties.

Stutzbach and Rejaie [24] focused on the problem of obtaining unbiased representative nodes from an unknown, large-scale, dynamically changing graph. Leskovec et al. [25] studied graph evolution, concentrating on properties such as densification (networks becoming denser) and diameter shrinkage (diameter decreasing as networks grow), and used new graph generators to explain these properties. Their observations spanned years.

Sampling techniques affect the selection of active measurement nodes. However, quantitative theoretical analysis of active measurement errors in P2P networks is still lacking. If active measurement technology distorts the P2P network, even unbiased sampling techniques will yield inaccurate information. Therefore, it is necessary to first address the impact of active measurement on P2P networks before tackling measurement error and calibration issues.

(3) **Statistical Analysis of Search Coverage**
Terpstra et al. [26][27] proposed a probabilistic exhaustive search system called BubbleStorm based on random multigraphs. BubbleStorm divides the entire system into multiple query "bubbles" and data "bubbles"; if the two intersect, relevant content can be found. This system exploits the characteristic that bandwidth differs between nodes. Chawathe [14], Gkantsidis [15], and Lv [16] et al. used random walks as the foundation for unstructured P2P network search, but search only needs to locate specific data during the walk process without concern for whether certain nodes are more likely to be selected, which differs from unbiased sampling.

(4) **Establishing Standards for Validating Measurement-Based Network Research**
Allman [29] proposed reactive measurement, viewing measurement as a process rather than an event, where previous measurement results determine whether and what subsequent measurements are needed. This demonstrates that measurement is a dynamic process and also indicates the need for objective evaluation criteria. Krishnamurthy et al. in 2008 [22] proposed the need to establish validation standards for measurement-based networking research and presented preliminary ideas for such standards. They raised four questions: (i) measurement quality, (ii) appropriateness of statistical analysis, (iii) scientific value of modeling methods, and (iv) completeness of validation. P2P network measurement faces the same issues, and there is an urgent need to establish unified evaluation criteria. To establish such criteria, comprehensive research on the impact of active measurement on P2P networks is first required.

Therefore, research on the impact of active measurement technology on peer-to-peer networks is a critical foundational issue concerning measurement accuracy and effectiveness. It requires establishing mathematical models of active measurement based on thorough analysis of its essential characteristics, theoretically analyzing its impact on P2P networks to obtain formal results, and proposing calibration algorithms to obtain more accurate and comprehensive key characteristics and information about the measured P2P network.

2.2 Current Research Challenges

Although numerous researchers have conducted extensive measurements and analyses of various P2P systems, obtaining many beneficial results, current active measurement technology still faces the following problems:

(1) **Unknown Impact of Active Measurement on Peer-to-Peer Networks**
As P2P systems grow increasingly large, so do active measurement systems. However, the impact of active measurement technology on P2P networks remains unknown.

P2P active measurement typically involves customizing P2P system clients to form network crawlers that actively join the P2P network to acquire relevant network characteristics and node attributes. To obtain this information, network crawlers generally adhere to basic P2P protocols but require certain modifications for measurement efficiency and accuracy. When network crawlers reach a certain scale, they inevitably affect P2P network characteristics and performance metrics, influencing decisions of relevant nodes and servers in the network, which further affects user decisions. The strategies of original P2P nodes are also impacted by the insertion of active measurement nodes. Although connections initiated by each measurement node are clear, the resulting changes in strategies of original P2P nodes due to these added measurement nodes are unknown. Without research on the impact of active measurement on P2P networks, there is no way to judge the effectiveness of the measurement process or the accuracy of measurement data. Currently, no relevant formal analysis or theoretical research has been found.

(2) **Existing Sampling Techniques Cannot Guarantee Correct Information**
As previously mentioned, sampling is an effective method for understanding these systems and deriving global information from local samples. As P2P system scales increase, so does sampling scale. Sampling techniques affect the selection of active measurement nodes. However, quantitative theoretical analysis of active measurement errors in P2P networks is still lacking. If active measurement technology distorts the P2P network, even unbiased sampling techniques will yield incorrect information. Therefore, it is necessary to first understand the impact of active measurement on P2P networks before addressing measurement error and calibration issues, making the use of reasonable sampling techniques meaningful.

(3) **Lack of Objective Evaluation Criteria for Active Measurement Technology**
As P2P measurement work deepens, issues concerning measurement effectiveness validation standards are gradually gaining attention. Despite extensive measurement efforts, there is still a lack of objective evaluation criteria for measurement process effectiveness and result correctness. Without objective evaluation criteria, there is no way to assess the quality of measurement methods, 不利于 measurement data sharing, no way to evaluate measurement data quality, and consequently no way to employ reasonable technical means during data processing to avoid or overcome inherent measurement data defects, ultimately preventing accurate and complete validation of theoretical models through measurement.

In summary, the main problems with current active measurement are: the impact of active measurement technology on P2P networks is unknown; existing active measurement technology cannot obtain global network snapshots; the accuracy of active measurement lacks quantitative theoretical analysis; and as an important research method for P2P networks, active measurement still lacks objective evaluation criteria.

3 Our Work

This section first introduces our research approach, then presents our preliminary results on the impact of active measurement technology on P2P degree distribution.

3.1 Basic Research Approach

We first need to study the behavioral models of active measurement nodes and the active measurement network composed of these nodes. The observed measurement results are essentially the superposition of the measured P2P network and the active measurement network. Based on the first two research components, reasonable decomposition of measurement results can yield more authentic and accurate network characteristics and node attributes of the measured P2P network, thereby establishing active measurement calibration algorithms.

Specific research includes the following aspects: - **Active measurement node research**: including attribute models and behavioral models of active measurement nodes, with emphasis on comparison with normal P2P nodes; - **Active measurement network research**: including active measurement network models, their essential characteristics, and interaction mechanisms with the measured P2P network; - **Impact of active measurement on P2P networks**: analyzing which P2P network properties are affected by active measurement and how, establishing superposition models of active measurement networks and P2P networks and their formal results; - **Active measurement calibration algorithms and validation**: reasonably decomposing measurement results to obtain authentic and accurate network characteristics of the measured P2P network, and verifying model and algorithm effectiveness in simulation and real network environments.

3.2 Impact of Active Measurement Technology on Peer-to-Peer Network Degree Distribution

Degree distribution is an important characteristic for describing peer-to-peer networks. Therefore, we first study the impact of active measurement technology on the degree distribution of unstructured P2P networks as the starting point for this research. We define the degree k_i of node i as the number of other nodes connected to it. The average of all node degrees k_i in the network is called the network' s average degree, denoted as $\langle k \rangle$. The degree distribution of nodes in the network can be described by the distribution function $p(k)$. $p(k)$ represents the proportion of nodes with degree k among all nodes, i.e., the probability that a randomly selected node has degree k.

Numerous studies show that the degree distributions of many real networks can be described by a power-law form $p(k) \sim k^{-\gamma}$. If a function $f(x)$ has the property that for any given constant a, there exists a constant b such that $f(ax) = bf(x)$, then the function is called "scale-free." Power functions are the only functions with this property. This indicates that networks with power-law degree distributions have no obvious characteristic scale, and such networks are called scale-free networks. In a large-scale scale-free network, the vast majority of nodes have relatively low degree, but there exist a small number of nodes with relatively high degree. Therefore, scale-free networks are non-uniform, and those few high-degree nodes are called the network' s "hub nodes."

The overall behavior of active measurement on P2P networks can be viewed as the superposition of the measured P2P network and the measurement network: $G_m = G_0 + G_a$, where G_m is the measured network (the hybrid network formed by superimposing the active measurement network onto the measured P2P network), G_0 is the measured P2P network, and G_a is the active measurement network.

3.2.1 Measured P2P Network G_0 Generation Model Many studies show that unstructured P2P networks exhibit scale-free characteristics, and their degree distribution can be described by $p(k) \sim k^{-\gamma}$. Therefore, we can use the Barabási-Albert model (BA model) [33] to generate the G_0 model. First is the growth characteristic of real networks: networks are open systems where network size is not fixed but continuously increasing. Second, real networks exhibit preferential attachment: new nodes tend to connect to "large" nodes with higher connectivity. Based on these growth and preferential attachment characteristics, the BA scale-free network model construction algorithm is as follows [33]:

(1) **Network growth**: Start with a network of m_0 nodes. At each step, introduce a new node and connect it to m existing nodes.

(2) **Preferential attachment**: The probability Π_i that a new node connects to an existing node i satisfies the following relationship with node i' s degree k_i:

$$\Pi_i = \frac{k_i}{\sum_j k_j}$$

After t steps, this algorithm produces a network with $N = t + m_0$ nodes and mt edges. When N is sufficiently large, the network' s degree distribution can be described by a power-law function with exponent -3, i.e., the probability of nodes with connectivity k is proportional to k^{-3}, and the network' s average degree is $2m$.

When $t \to \infty$, we obtain a time-independent power-law degree distribution:

$$p(k) = 2m^2 k^{-3}$$

3.2.2 Mixed Network G_m Generation Model First, we analyze the characteristics of the active measurement network G_a regarding its impact on P2P network degree distribution: - To improve the accuracy of obtained network snapshots, active measurement duration cannot be too long; - Because measurement time cannot be too long, the lifecycle of active measurement nodes is also short, but within the measurement period, newly added nodes should primarily be active measurement nodes; - Due to the dynamic and large-scale nature of P2P networks, active measurement nodes constitute only a small portion of the entire P2P network; - The degree of active measurement nodes is generally higher than that of normal P2P nodes, as only high-degree measurement nodes can obtain as much information as possible, but within the measurement period, new connections should mainly be between measurement nodes and existing nodes; - Active measurement nodes can typically only obtain local information about the current P2P network and initiate connections to some of these nodes.

Considering these characteristics of the active measurement network G_a, the proposed mixed network G_m generation model primarily simulates the active measurement process. Suppose the measured P2P network has N nodes and M connections, and the local P2P network scale obtained through active measurement is γ. The mixed network G_m generation model can be established by repeating the following steps:

(cid:132) **Step 1**: Establish the local P2P network. Randomly select or use some sampling algorithm to choose γ nodes from the entire P2P network. These γ nodes constitute a local P2P network for the active measurement node, denoted as Δ;

(cid:132) **Step 2**: Add nodes. With probability p, add a new node with m_1 connections to the local P2P network Δ. The probability that any existing node i in the local network connects to this new node at time t is proportional to node i' s average degree, i.e.:

$$\Pi_i = \frac{k_i}{\sum_{j \in \Delta} k_j}$$

(cid:132) **Step 3**: Add connections. With probability q, add m_1 new connections among existing nodes in the local network. One end of each new connection is randomly selected, while the other end is chosen with probability Π_i;

(cid:132) **Step 4**: Remove connections. With probability $1 - p - q$, remove m_2 connections in the local network, which are randomly selected within the local network.

Here, $0 \leq p < 1$, $N \gg m_1, m_2$, and $p + q \leq 1$.

Step 1 can be implemented using various methods to map from the global P2P network to the local network. Step 2 corresponds to adding a new measurement node; since newly added nodes during active measurement are primarily measurement nodes, this model mainly considers measurement nodes. Step 3 corresponds to adding connections between measurement nodes and normal P2P nodes, including both measurement-to-normal and normal-to-normal connections. Due to node performance limitations, the system can only add a limited number of connections per unit time, so we set the parameters for both operations to m_1. The final step simulates the dynamic nature of P2P networks, where some connections between measurement nodes and normal nodes are disconnected per unit time.

Of course, these P2P network models and superposition models of active measurement networks and P2P networks are relatively simple and will gradually adopt more complex models to approach real systems as research progresses.

3.2.3 Mixed Network G_m Degree Distribution This section uses Barabási-Albert continuous mean-field theory to analyze differences in degree distribution between the mixed network and the measured network. From the previous section, at each time step there are m_1 connections added. Therefore, we can consider two extreme cases:

Case 1: When $\gamma = N$, i.e., the local P2P network scale equals the entire P2P network scale.

In this case, the local P2P network is the entire P2P network, and its degree distribution is:

$$p_m(k) \sim ak^{-b}$$

where k_0 is the average degree of the measured P2P network.

This shows that in this case, the mixed network remains a scale-free network. That is, if the active measurement scale is small and the active measurement network can obtain information about the entire P2P network, active measurement only affects coefficient a but not the power exponent b.

Case 2: When $m_1\gamma = 1$, the degree distribution becomes:

$$p_m(k) \sim ae^{-bk}$$

In this case, the mixed network follows an exponential distribution.

From these two extreme examples, we can see that if $m_1\gamma \to 1$, then $p_m(k)$ approaches an exponential distribution; if $m_1\gamma \to N$, then $p_m(k)$ transitions from exponential to scale-free distribution. If the local P2P network scale is sufficiently large and active measurement nodes mainly connect to existing normal P2P nodes ($m_1\gamma \to N$), then $p_m(k)$ approaches a scale-free distribution, meaning the active measurement network has minimal impact on the measured P2P network. However, if the local P2P network scale is small, the mixed network transitions to an exponential distribution, indicating significant impact from the active measurement network on the measured P2P network. Therefore, active measurement networks should obtain information from as many nodes in the measured P2P network as possible.

3.2.4 Simulation Experiments We primarily used MATLAB to simulate the generation models of the measured network G_0 and the mixed network G_m. We use the following notation: G_0 is denoted as $G(N_0, m_0)$, where N_0 is the total number of nodes in G_0 and m_0 is the number of connections added per time step; $p(k)$ and $p_m(k)$ are denoted as P_0 and P_m respectively. G_m is denoted as $G_m(N_m, m_1, m_2, p, q, G_0)$, where N_m is the total number of nodes in G_m, m_1 is the number of connections added per time step, m_2 is the number of connections removed per time step, p is the probability of adding nodes, q is the probability of adding connections, and G_0 is the initial measured P2P network. We simulated G_0 scales from 2,000 to 20,000 nodes. Since results were similar, we only use $G_0(5000, 5)$ as an example below.

Comparison of P_0 and P_m

The degree distributions of $G_0(5000, 5)$ and $G_m(5010, 10, 1, 0.6, 0.3, G_0)$ are shown in [Figure 1: see original paper], where the X-axis represents node degree and the Y-axis represents degree distribution (log-log scale). From the figure, we can see that the degree distribution approximates a straight line, indicating that both networks follow scale-free distributions without obvious differences, showing that small-scale active measurement has minimal impact on P2P networks.

Impact of N_m on P_m

As N_m increases, coefficients a and power exponent b drop sharply, causing a phase transition phenomenon ([Figure 2: see original paper]). As shown in [Figure 3: see original paper], when N_m exceeds 5,070, the degree distribution has severely deviated from the scale-free distribution. We observed that even a very small-scale active measurement network (around 1%) can severely impact the degree distribution (). Therefore, when designing active measurement networks, small-scale high-degree measurement nodes should be used rather than large-scale low-degree ones.

Values of N_0 and N_m at phase transition

[Figure 2: see original paper] Impact of N_m on coefficients a and power exponent b, X-axis shows N_m from 5,000 to 5,200, Y-axis shows a and b respectively.

[Figure 3: see original paper] P_m: $G_0(5000, 5)$; $G_m(5130, 10, 1, 0.6, 0.3, G_0)$

4 Conclusion and Future Work

This paper first introduced the development and current status of peer-to-peer network measurement, identifying that the main problem is the unknown impact of active measurement technology on P2P networks, which affects the judgment and correction of measurement results and hinders the establishment of unified measurement evaluation criteria. We then presented our work on the impact of active measurement on P2P networks: we proposed a basic research approach and obtained preliminary results on how active measurement affects P2P degree distribution. Future work will further investigate the impact of active measurement on other P2P network characteristics, propose active measurement calibration algorithms to obtain correct and authentic measurement results, and establish unified measurement evaluation criteria.

References

[1] S. Daniel, R. Reza, Sampling Techniques for Large, Dynamic Graphs, in Global Internet Symposium, http://mirage.cs.uoregon.edu/pub/sampling-gi06.pdf

[2] M. Izal, G. Urvoy-Keller, E. W. Biersack, et al. Dissecting BitTorrent: Five Months in a Torrent' s Lifetime. In PAM, 2004.

[3] J. Pouwelse, P. Garbacki, D. Epema, and H. Sips. The Bittorrent P2P File-sharing System: Measurements and Analysis. In International Workshop on Peer-to-Peer Systems (IPTPS), 2005.

[4] S. Saroiu, P. K. Gummadi, and S. D. Gribble. Measuring and Analyzing the Characteristics of Napster and Gnutella Hosts. Multimedia Systems Journal, 9(2), 2003.

[5] D. Stutzbach, R. Rejaie, and S. Sen. Characterizing Unstructured Overlay Topologies in Modern P2P File-Sharing Systems. In Internet Measurement Conference, 2005.

[6] C. Gkantsidis, M. Mihail, and A. Saberi. Random Walks in Peer-to-Peer Networks. In INFOCOM,

[7] L. Lovász. Random walks on graphs: A survey. Combinatorics: Paul Erdős is Eighty, 2, 1993.

[8] B. Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. European Journal of Combinatorics, 1, 1980.

[9] M. Jerrum and A. Sinclair. Fast uniform generation of regular graphs. Theoretical Computer Science, 73, 1990.

[10] C. Cooper, M. Dyer, and C. Greenhill. Sampling regular graphs and a peer-to-peer network. In Symposium on Discrete Algorithms, 2005

[11] V. Krishnamurthy, M. Faloutsos, M. Chrobak, L. Lao, J.-H. Cui, and A. G. Percus. Reducing Large Internet Topologies for Faster Simulations. In IFIP Networking, 2005.

[12] V. Krishnamurthy, J. Sun, M. Faloutsos, and S. Tauro. Sampling Internet Topologies: How Small Can We Go? In International Conference on Internet Computing, 2003.

[13] M. P. H. Stumpf, C. Wiuf, and R. M. May. Subnets of scale-free networks are not scale-free: Sampling properties of networks. Proceedings of the National Academy of Sciences, 102(12), 2005.

[14] Y. Chawathe, S. Ratnasamy, and L. Breslau. Making Gnutella-like P2P Systems Scalable. In SIGCOMM, 2003.

[15] C. Gkantsidis, M. Mihail, and A. Saberi. Random Walks in Peer-to-Peer Networks. In INFOCOM,
[16] Q. Lv, P. Cao, E. Cohen, K. Li, and S. Shenker. Search and Replication in Unstructured Peer-to-Peer Networks. In International Conference on Supercomputing, 2002.
[17] S. Saroiu, PK. Gummadi, SD. Gribble, A measurement study of peer-to-peer file sharing systems. In: Proc. of the Multimedia Computing and Networking (MMCN) 2002. January 2002.
[18] J. Liang, R. Kumar, KW. Ross, The KaZaA overlay: A measurement study. In: Proc. of the 19th IEEE Annual Comp Communications Workshop. 2004. http://cis.poly.edu/~ross/papers/KazaaOverlay.pdf
[19] J. Liang, R. Kumar, KW. Ross. Understanding KaZaA. 2004. http://cis.poly.edu/~ross/papers/UnderstandingKaZaA.pdf
[20] J. Liang, R. Kumar, KW. Ross. Pollution in file sharing systems. In: Proc. of the IEEE Infocom 2005. 2005. http://cis.poly.edu/~ross/papers/pollution.pdf
[21] Xiaoming Wang; Zhongmei Yao; D. Loguinov. Residual-Based Measurement of Peer and Link Lifetimes in Gnutella Networks; INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE 6-12 May 2007 Page(s):391 - 399.
[22] B. Krishnamurthy, W. Willinger What are our standards for validation of measurement-based networking research?, HotMetrics 2008 June 6, 2008 Annapolis, MD First Workshop on Hot Topics in Measurement & Modeling of Computer Systems
[23] T. Isda, M. Piatek, A. Krishnamurthy, T. E. Anderson: Leveraging BitTorrent for End Host Measurements, Passive and Active Measurement, 8th International Conference, PAM 2007.
[24] D. Stutzbach, R. Rejaie, On Unbiased Sampling for Unstructured Peer-to-Peer Networks, Internet Measurement Conference(IMC), Proceedings of 6th ACM SIGCOMM conference on Internet measurement. 2006.
[25] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations. In KDD, 2005.
[26] W. W. Terpstra, J. Kangasharju, C. Leng, A. P. Buchmann, BubbleStorm: Resilient, Probabilistic, and Exhaustive Peer-to-peer search, Proc. ACM SIGCOMM 2007.
[27] W. W. Terpstra, C. Leng, and A. P. Buchmann. BubbleStorm: Analysis of Probabilistic Exhaustive Search in a Heterogeneous Peer-to-Peer System. Technical Report TUD-CS-2007-2, Technische Universität Darmstadt, Germany, 2007.
[28] D. Stutzbach, R. Rejaie, Evaluating the accuracy of captured snapshots by peer-to-peer crawlers, Passive and Active Measurement Workshop, Extended Abstract, PAM 2005.
[29] M. Allman, V. Paxson, A Reactive Measurement Framework, Passive and Active Network Measurement (2008), pp. 92-101.
[30] Mo Zhou, Yafei Dai, and Xiaoming Li, "A Measurement Study of the Structured Overlay Network in P2P File-Sharing Systems," International Journal of Advances in Multimedia, vol. 2007, Article ID 27958, 9 pages, 2007

[31] Zhao Yang, Hou Xiaoxiao, Yang Mao, Dai Yafei. Measurement Study and Application of Social Network in the Maze P2P File-Sharing System. In proceeding of International Workshop on Peer-to-Peer Information Management (P2PIM), May 2006, Hong Kong.
[32] Bin Cheng, Xiuzheng Liu, Zheng Zhang, Hai Jin, Lex Stein, and Xiaofei Liao, Evaluation and Optimization of a Peer-to-Peer Video-on-Demand System, Journal of Systems Architecture, Vol.54, No.7, Elsevier Science, July 2008, pp.651-663.
[33] D. J. Watts, S.H. Strogatz. Collective Dynamics of 'Small-world' Networks. Nature, 393(6684):440-442, 1998.

Author Biography:

Sha Ying: Assistant Researcher, Information Security Research Center, Institute of Computing Technology, Chinese Academy of Sciences, shaying@ict.ac.cn

Note: Figure translations are in progress. See original paper for figures.