

The Effect of Peer Selection with Hopcount or Delay Constraint on Peer-to-Peer Networking*

S. Tang, H. Wang, and P. Van Mieghem

Delft University of Technology, P.O. Box 5031, 2600 GA Delft, The Netherlands
{S.Tang,H.Wang}@ewi.tudelft.nl, P.F.A.VanMieghem@tudelft.nl

Abstract. We revisit the peer selection problem of finding the most nearby peer from an initiating node. The metrics to assess the closeness between peers are hopcount and delay, respectively. Based on a dense graph model with i.i.d regular link weight, we calculate the probability density function to reach a peer with minimum hopcount and asymptotically analyze the probability to reach a peer with the smallest delay within a group of peers. Both results suggest that a small peer group size is enough to offer an acceptable content distribution service. We also demonstrate the applicability of our model via Internet measurements.

1 Introduction

The idea of peer-to-peer (P2P) networking creates a reciprocal environment where, by sharing storage, bandwidth and computational capacity with each other, mutual benefit between end-users is possible. With the distribution of storage and retrieving functionality to peers in the P2P network, the process of selecting a best peer (in cost, bandwidth, delay, etc.) among a group of peers to start content retrieval becomes a vital procedure.

Our model considers the hopcount and delay as the major criteria for peer selection. The problem is confined as follows: given a underlying network of size N , over which m peers with the desired content are randomly scattered, what is the distribution of the hopcount and delay respectively to the most nearby peer from a requesting node? A *requesting peer* refers to the peer who initiates the downloading request. By solving the above problem, we expect to answer the fundamental question of how many peers are needed to store the replicas of a particular file so that the most nearby peer can always be reached within j hopcount or t delay.

Modeling of the peer selection problem is presented in section 2. We complement our model by verifying its applicability from a series of substantive experiments in Section 3. In Section 4, we conclude the paper.

* This work has been partially supported by the European Union in CONTENT NoE (FP6-IST-038423) and NWO Robunet (643.000.503).

2 Problem Description and Modeling

2.1 Modeling Assumptions

We model the number of hops and the latency to the nearest peer among a set of m peers based on three assumptions: (a) a dense graph model¹ for the underlying network, (b) regular link weight around zero, and (c) i.i.d. link weight distribution on each link.

The shortest path (SP) from a source to a destination is computed as the path that minimizes the link weights² along that path. In [4, Chapter 16.1], it is shown that a regular link weight distribution - regular means a linear function around zero - will dominate the formation of the shortest path tree (SPT), which is the union of all shortest paths from an arbitrary node to all the other destinations. A uniform recursive tree (URT) is asymptotically the SPT in a dense graph with regular i.i.d. link weights (e.g. exponential link weights) distribution [6]. A URT of size N is a random tree that starts from the root A , and where at each stage a new node is attached uniformly to one of the existing nodes until the total number of nodes reaches N .

2.2 Hopcount Distribution to the Nearest Peer

A) Theoretical analysis

The number of hops from a requesting peer to its most nearby peer, denoted by $h_N(m)$, is the minimum number of hops among the set of shortest paths from the requesting node to the m peers in the network of size N . Let $H_N(m)$ be the hopcount starting from one, excluding the event $h_N(m) = 0$ in the URT. Since $\Pr[h_N(m) = 0] = \frac{m}{N}$, we have

$$\Pr[H_N(m) = j] = \Pr[h_N(m) = j | h_N(m) \neq 0] = \frac{1}{1 - \frac{m}{N}} \Pr[h_N(m) = j] \quad (1)$$

with $j = 1, 2, \dots, N$ and $\Pr[h_N(m) = j]$ recursively solved in [4, p. 427]. However, the recursive computation involves a considerable amount of memory and CPU-time which limits its use to relatively small sizes of $N \leq 100$.

Fig. 1 illustrates $\Pr[H_N(m) = j]$ versus the fraction of peers $\frac{m}{N}$ for different hops j with network size varying from $N = 20$ up to $N = 60$. The interesting observation from Fig. 1 is that, for separate hops j , the distribution $\Pr[H_N(m) = j]$ rapidly tends to a distinct curve for most of the small networks ($N \leq 100$) and that the increase in the network size N only plays a small role. Further, the crosspoint of curve $j = 1$ and $j \geq 2$ (the bold line) around $\frac{m}{N} = 15\%$ indicates that in small networks (i.e. $N = 20$), the peer fraction should always be larger than 15% to ensure $\Pr[H_N(m) = 1] > \Pr[H_N(m) \geq 2]$.

¹ The dense graph is a heterogenous graph with the average degree $E[D] \geq p_c N \approx O(\log N)$ and a small standard deviation $\sqrt{\text{Var}[D]} \ll E[D]$, where $p_c \sim \frac{\log N}{N}$ is the disconnectivity threshold of the link density [4, Chapter 15.6.3].

² The link weight w_{ij} assigned to a link (i, j) between node i and node j in a network, is a real positive number that reflects certain properties of the link, i.e. distance, delay, loss, or bandwidth.

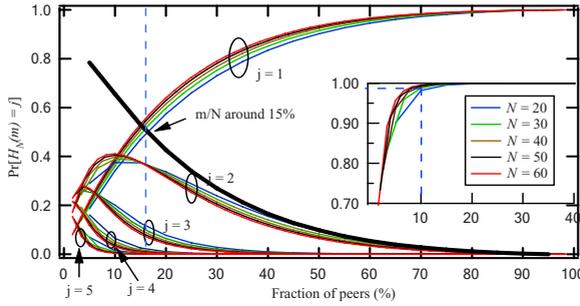


Fig. 1. $\Pr[H_N(m) = j]$ versus the fraction of peers $\frac{m}{N}$ from network size $N = 20$ to 60. The bold line is the pdf of $H_N(m) \geq 2$ for $N = 20$. The inserted figure plots the $\Pr[H_N(m) \leq 4]$ as a function of peer fraction $\frac{m}{N}$ for network sizes $N = 20$ to 60.

To avoid the recursive calculation in (1), we compute $\Pr[H_N(m) = j]$ approximately by assuming the independence of the hopcount from the requesting node to the m peers when m is small³. The approximation is expressed as

$$\Pr[H_N(m) \leq j] \approx 1 - (\Pr[H_N > j])^m \tag{2}$$

where $\Pr[H_N > j]$ is the probability that at least one of the peers is j hop away (or not all peers are further than j hop away). As explained in footnote³, we confine the estimation of (2) with large N and small m , whereas the exact result is applicable for $N \leq 100$ with all m . We discuss the usage of (2) and its asymptotic result for very large network in more detail in [7].

B) Application of $\Pr[H_N(m) = j]$

We apply (1) to estimate the peer group size for a certain content delivery service. For instance, if the operator of a content distribution network (CDN) with 40 routers has uniformly scattered 4 servers (peer fraction around 10%) into the network, he can already claim that approximately in 98% of the cases, any user request will reach a server (the term of server and peer are interchangeable in this case) within 4 hops ($j \leq 4$) as seen in the inserted figure of Fig. 1. Placing more servers in the network will not improve the performance significantly.

2.3 Weight of the Shortest Path to the First Encountered Peer

A) The asymptotic analysis

In [4, p. 349], the shortest path problem between two arbitrary nodes in the dense graph with regular link weight distribution (e.g. exponential link weights)

³ The path overlap from the root to the m peers in the URT causes correlation of the hopcount between peers. When m is small compare to the network size N , the path overlap is expected to be small, and so is the correlation of the hopcount. The larger the m , the more dependent of the hopcount from the root to the m peers becomes.

has been rephrased as a Markov discovery process. It evolves as a function of time from the source and stops at the time when the destination node is found. The transition rate in this continuous-time Markov chain from state n with n already discovered nodes, to the next state $n + 1$ is $\lambda_{n;n+1} = n(N - n)$. The inter-attachment time τ_n between the inclusion of the n -th and $(n + 1)$ -th node in the SPT for $n = 1, 2, \dots, N - 1$ has exponential distribution with parameter $n(N - n)$.

The exact probability generating function (pgf) of the weight $W_{N;m}$ of the shortest path from an arbitrary node to the first encountered peer among m peers can be formulated as $\varphi_{W_{N;m}}(z) = \sum_{k=1}^{N-m} E[e^{-zv_k}] \Pr[Y_m(k)]$, where $\Pr[Y_m(k)]$ represents the probability that the k -th attached node is the first encountered peer among the m peers in the URT, $v_k = \sum_{n=1}^k \tau_n$ denotes the weight of the path to the k -th attached node. And the corresponding generating function of v_k is $E[e^{-zv_k}] = \prod_{n=1}^k \frac{n(N-n)}{z+n(N-n)}$. The formation of the URT with m attached peers indicates $\binom{N-1}{m}$ ways to distribute the m peers over the $N - 1$ position (other than the source node). The remaining $m - 1$ peers should always appear in the position that are larger than k -th position. Hence, there are $\binom{N-1-k}{m-1}$ ways to distribute the $m - 1$ peers over the $N - 1 - k$ position. This analysis leads us to express $\Pr[Y_m(k)]$ as

$$\Pr[Y_m(k)] = \frac{\binom{N-1-k}{m-1}}{\binom{N-1}{m}}$$

The asymptotic probability of $\varphi_{W_{N;m}}(z)$ with proper scaling is thus derived⁴ as

$$\lim_{N \rightarrow \infty} \Pr[NW_{N;m} - \ln \frac{N}{m} \leq y] = e^{-my} m^{m+1} e^{me^{-y}} \int_{me^{-y}}^{\infty} \frac{e^{-u}}{u^{m+1}} du \quad (3)$$

which converges to the Fermi-Dirac distribution function as shown in [5, Section 3]

$$\lim_{N \rightarrow \infty} \Pr[NW_{N;m} - \ln \frac{N}{m} \leq y] = \frac{1}{1 + e^{-y}} \quad (4)$$

for large m as shown in Fig. 2. It illustrates that a relatively small peer group $m \approx 5$ is sufficient to offer a good service quality because increasing the number of peers can only improve the performance marginally, i.e. logarithmically in m .

B) Application

We use the Fermi-Dirac distribution (4) to estimate the minimum number of peers m needed to satisfy the requirement of $\Pr[W_{N;m} \leq y] \geq \eta$, which means that in η of the cases, the delay to the nearest peer is no larger than y . Rewriting (4), yields $\Pr[W_{N;m} \leq y] \approx \frac{\frac{m}{N} e^{Ny}}{1 + \frac{m}{N} e^{Ny}} \geq \eta$, from which we find

$$\frac{m}{N} \geq \frac{\eta}{(1 - \eta)} e^{-yN} \quad (5)$$

⁴ The joint probability of the pair $(H_N(m), W_{N;m})$ as calculated in [2] is shown to be asymptotically independent. Hence, for large N , both hopcount via (1) and delay via (3) is sufficient to compute $\Pr[h_N(m) = j, W_{N;m} \leq y]$.

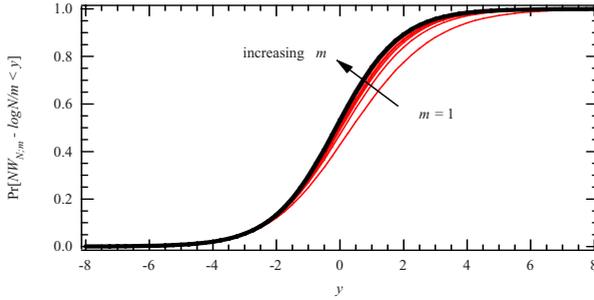


Fig. 2. The convergence of the pdf with scaled random variable $NW_{N,m} - \ln \frac{N}{m}$ towards the Fermi-Dirac distribution (in bold) for increasing $m = 1, 2, 3, 4, 5, 10, 15, 20$.

Consider an online P2P music sharing system, which takes advantage of the VoIP service for example. The standard [3] suggests a mouth-to-ear delay < 150 ms. Let $N = 80$, $y = 150$, and $\eta = 99.99\%$, with (5), we find that in this network, the declaration that in 99.99% cases the delay to the nearest peer is smaller than 150ms can always be achieved if there are 5 peers sharing the same music ($\frac{m}{N}$ around 6%).

3 Discussion on Modeling Assumptions

In this section, we discuss the major two assumptions made for the URT model: 1) a dense graph to mimic the underlying network and 2) i.i.d. regular link weight. The Internet might be denser than what has been measured, taking into account all sorts of significant sampling bias, such as insufficient sources for traceroutes suggested in [1]. We will also give indications on the link weight distribution and the applicability of the URT model in P2P network in this section.

3.1 Link Weight Structure of Networks

We use the data from the National Road Database provided by the Dutch transport research center to give an indication on the link weight distribution in a transportation network. The link weight of the Dutch road is evaluated as the physical distance between two roadsections. In Fig. 3, we fit the link weight distribution $F_w(x)$ of the Dutch road network with a linear function. A regular (linear) link weight distribution is found within a small range $[0, \epsilon]$, where $\epsilon \sim 0.03$, which gives evidence to the assumption of regular link weight structure around zero. Given that the link weight structure in the Internet are tunable, we claim that the assignment of regular link weights in Section 2.1 is reasonable.

3.2 Applicability of the URT Model

We have carried out a series of experiments by using the *traceroute* data provided by iPlane⁵ to give further indication on how well the URT model matches the real

⁵ <http://iplane.cs.washington.edu/>

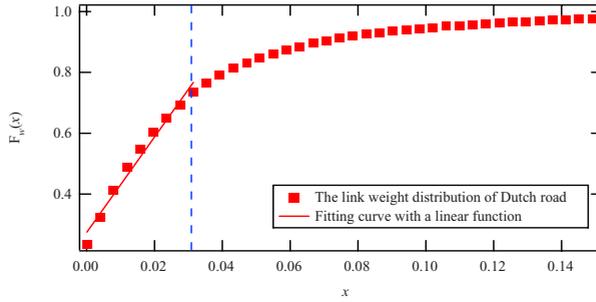


Fig. 3. $F_x(x) = \Pr[w \leq x]$ of the link weight of the Dutch transportation network with the x axis normalized between $[0, 1]$ ($x \in [a, b] \xrightarrow{\text{normalize}} x \in [\frac{a}{b}, 1]$). The correlation coefficient $\rho = 0.99$ suggests a high fitting quality.

network. iPlane performs periodic measurements by using PlanetLab nodes to probe a list of targets with global coverage. We use the iPlane measurement data executed on 8th June 2007. We extract the stable traces from 52 Planetlab nodes that are located in different Planetlab sites. Assuming the *traceroutes* represent the shortest paths, we construct a SPT rooted at each PlanetLab node (as a source) to m peers (destinations), resulting in 52 SPTs in total. By using a map with all aliases resolved in iPlane, we obtain the router-level SPTs. The m peers are randomly chosen over the world, and the hopcount (H_{SPT}) and degree (D_{SPT}) distribution are obtained by making a histogram of the 52 SPTs (each with m destinations).

A) Experimental results on node degree distribution

Three sets of experiments with $m = 10, 25$ and 50 are conducted to examine the degree distribution of the sub-SPT because the number of peers in a P2P network is not expected to be large⁶. We observed from the experiments that an exponential node degree distribution is, if not better, at least comparable to the power law degree distribution that has been reported in most of the published papers.

In Fig. 4, we fitted $\Pr[D_{SPT} = k]$ for $m = 10$ with a linear function on both log-lin and log-log scales. The linear correlation coefficients used to reflect the fitting quality are ρ_α on the log-lin scale and ρ_β on the log-log scale respectively. The quality of the fitting on the log-log scale ($\rho_\beta = 0.99$) is only slightly higher than that of the log-lin scale ($\rho_\alpha = 0.98$), which questions the power law degree distribution of a small subgraph of the Internet topology. A similar phenomenon is also observed for $\Pr[D_{SPT} = k]$ for $m = 25$ and $m = 50$. Due to space limitation, we only provide the correlation coefficients for $m = 25$ and $m = 50$ in Table 1. Again, the quality of the fitting seems to be comparable on both scales.

A discrepancy with the first three experiments occurs if we increase the peer size to 500 [7]. For larger subgraphs, a clear power law, rather than an exponential

⁶ Measurements on PPLive, a popular IPTV application [8] reveal that the number of active peers that a peer can download from is always smaller than 50.

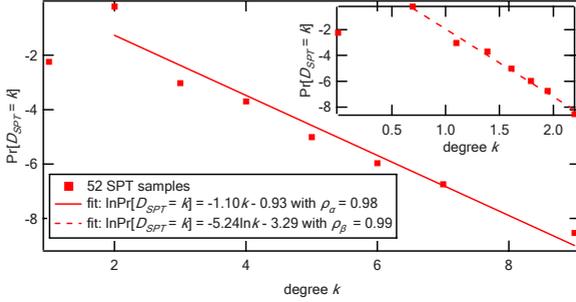


Fig. 4. The histogram of degree D_{SPT} for 52 SPTs with $m = 10$ peers based on PlanetLab traceroute data on log-lin and log-log scale in the inset. ρ_α represents the linear correlation coefficient on log-lin scale and ρ_β is the one on log-log scale.

Table 1. Correlation coefficient for both log-lin (ρ_α) and log-log (ρ_β) scale of $m = 10, 25$ and 50

	ρ_α	ρ_β
$m = 10$	0.98	0.99
$m = 25$	0.95	0.99
$m = 50$	0.95	0.99

distribution dominates the node degree. More discussion on the node degree distribution of both experimental and simulation results can be found in [7]. We conclude that the node degree of a subgraph with small m cannot be affirmatively claimed to obey a power law distribution. At least, it is disputable whether the exponential distribution can be equally good as the power law.

B) Hopcount distribution in the Internet

The probability density function of the hopcount from the root to an arbitrary chosen node in the URT with large N can be approximated as the following according to [4, p. 356].

$$\Pr[H_N = k] \approx \Pr[h_N = k] \sim \frac{(\log N)k}{Nk!} \tag{6}$$

where H_N indicates the event that $k \geq 1$.

We plotted the pdf of the hopcount with $m = 50$ (50 traceroute samples for each tree) in Fig. 5 (a), in which we see a reasonably good fitting with (6). An even better fitting quality is found in Fig. 5 (b) if we increases the number of traceroutes samples by randomly selecting $m = 8000$ destinations for each tree, because more traceroutes gives higher accuracy. We conclude that the hopcount distribution of the Internet can be modeled reasonably well by the pdf of hopcount (6) in the URT.

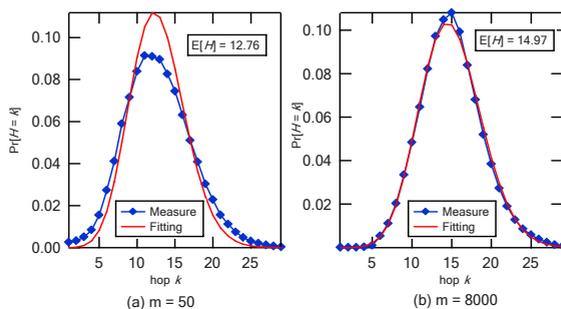


Fig. 5. The histogram of hopcount derived from 52 SPTs for $m = 50$ (a) and $m = 8000$ (b) are fitted by the pdf (6) of the hopcount in the URT. The measured data for (a) and (b) are fitted with $\log(N) = 12.76$ and $\log(N) = 14.97$ respectively.

4 Conclusion

We obtain the hopcount and delay distribution to the most nearby peer on the URT by assigning regular i.i.d. link weights (e.g. exponential link weights) on a dense graph. Both results suggest that a small peer group is sufficient to offer an acceptable quality of service. Via a series experiments, we show the applicability of the URT model, based on which the pdfs for hopcount and delay have been derived. To summarize, with a small group of peers ($m \leq 50$), the URT seems to be a reasonably good model for a P2P network.

Acknowledgments

We would like to thank Neil Spring, who is so kind to provide us the iPlane data.

References

1. Clauset, A., Moore, C.: Accuracy and scaling phenomena in Internet mapping. *Physical Review Letters* 94, 18701 (2005)
2. Hooghiemstra, G., Van Mieghem, P.: The weight and hopcount of the shortest path in the complete graph with exponential weights. *CPC* (submitted)
3. ITU-T Rec. G.114, One-way transmission time (May 2003)
4. Van Mieghem, P.: *Performance Analysis of Communications Networks and Systems*. Cambridge University Press, Cambridge (2006)
5. Van Mieghem, P., Tang, S.: Weight of the Shortest Path to the First Encountered Peer in a Peer Group of Size m . *Probability in the Engineering and Informational Sciences (PEIS)* 22, 37–52 (2008)
6. Van der Hofstad, R., Hooghiemstra, G., Van Mieghem, P.: First Passage Percolation on the Random Graph. *Probability in the Engineering and Informational Sciences (PEIS)* 15, 225–237 (2001)
7. Tang, S., Wang, H., Van Mieghem, P.: Peer selection with hopcount and delay constraint, Delft University of Technology, report20080222 (2008)
8. Hei, X., Liang, C., Liang, J., Liu, Y., Ross, K.W.: Insights into PPLive: A measurement study of a large-scale P2P IPTV system. In: *Workshop on Internet Protocol TV (IPTV) services over World Wide Web*, Edinburgh, Scotland, May 23 (2006)