An incentive mechanism for message relaying in peer-to-peer discovery

Cuihong Li
Tepper School of Business
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213
cuihong@andrew.cmu.edu

Bin Yu
Robotics Institute
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213
byu@cs.cmu.edu

Katia Sycara
Robotics Institute
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213
katia@cs.cmu.edu

ABSTRACT

Peer discovery via distributed message relaying is an important function of a P2P system in which the resources/services provided by peers and service/resource providers may change frequently. The usual searching protocols create huge burden on communications or cause long response time. Appropriate incentives are required to avoid free-riding and achieve efficient cooperations in message relaying. We present an incentive mechanism that aims at solving both the efficiency and incentive problems of message relaying for peer discovery. In this mechanism rewards are passed from upstream nodes to downstream nodes. A peer is rewarded if a service provider is found via a relaying path composed of this peer. The mechanism allows peers to specifically trade off communication efficiency and reliability, and maintains anonymity and information locality. Some analytical insights are given to the subgame perfect Nash equilibria (SPNE) and best response strategies of this game. An approximation approach is provided to calculate a symmetric SPNE. Experiments show that this incentive mechanism brings a system utility higher than in breadth-first search, and close to the optimal utility in a centralized system. In the incentive mechanism the distribution of relaying efforts over hops is dependent on the convexity of the cost function. With the cost function more convex, peers in earlier hops tend to spend more efforts on relaying.

1. INTRODUCTION

In a peer-to-peer(P2P) system information is highly distributed and stored by individual peers. Peer discovery, the function to find peers that provide certain information or services, is important in a P2P system for peers to exploit the distributed resources owned by other peers. However, to ensure the scalability and robustness of the system, as well as to avoid some legal issues, a centralized database of content of each peer usually does not exist in a P2P system (an exception is Napster). Instead a distributed catalog of content is favored in which each peer only maintains a list of resources/services of their own, and may also of their acquaintances and neighbors. In this distributive model peer discovery is realized via message relaying between peers so that a message for resource searching is propagated in the system with a "word-of-mouth" effect.

Generally there are two ways to perform distributive peer discovery: *breadth-first search*(used by Gnutella) and *depth-first search*(used by Freenet) [17][15]. In *breadth-first search* each search query message is assigned a maximum time-to-live(TTL), which limits the search depth in a number of hops. The requesting peer sends a query message to all its neighbors, and all nodes who receive the message relay it to all their neighbors, who forward the message to the next depth, and so on. The process ends when TTL is reached. In *depth-first search* each node relays the query to a single neighbor at one time, and waits for the response from the neighbor before forwarding the message to another neighbor or forwarding the result back to the requestor or the its upstream node. The search depth is also limited. With BFS the messages are flooded in the system. Therefore the consumption of bandwidth is enormous, although results can be found very quickly. With DFS searches can be terminated once a result is found, and therefore use less bandwidth. But the response time could be very long and is exponential in the depth limit. It is also more difficult to implement than the breadth-first search model.

Although modifications of the searching protocols have been suggested to reduce the bandwidth assumption or improve the response time [17], a problem that has not been considered in the current protocol design is the incentive problem. The protocols assume that peers will follow the design and contribute their bandwidth and resources to relaying messages as required. However, a P2P network is a highly decentralized system and each peer may present a different self-interested entity. A peer may manipulate the local information to take advantage of other peers' resources [8, 13]. For example, a peer may simply drop a message that is sent from other peers for relaying, for the purpose of saving communication bandwidth and energy. Therefore a message relaying P2P system is vulnerable to the *free riding* problem. Since a sound P2P system relies on the contribution of resources from each individual peers, free riding can cause severe degradation of the system performance or even paralyze the system [4, 10, 13]. It is important to design an incentive mechanism that motivates each peer to behave rationally in a system efficient way.

In this paper we present an incentive mechanism of message relaying for peer discovery that overcomes the flooding problem of BFS search while reserving the quick response property and good reliability. We have to first note that although both peer discovery and distributed routing are related to message relaying, they are different problems [5]. In distributed routing the destination of the message is known, but in peer discovery there is no guidance about who the message should be sent to. The unknown destination implies more uncertainty and less control in the message relaying process for peer discovery, and prevents the application of the pricing incentive mechanisms for distributed routing that requires the prior knowledge of routing paths [9].

The mechanism we propose is as follows: The source peer sends the query to some neighbors and promises some payment to each receiver if the resource provider is found via a transmission route composed of the receiver. Depending on the offer each receiver decides the number of neighbors she relays the message to and also the promised payment to its immediate downstream peers. Each of the new receivers again makes the same decisions, until the maximum number of hops is reached. This mechanism does not price the relaying activities, but instead prices the relaying result, which influences the relaying activities. This mechanism is motivated by the following requirements:

- *Communication efficiency:* Transmitting messages not only consumes the bandwidth resource, which could cause delay of communications, but also costs energy, which is especially a concern in wireless networks [6]. A significant part of inefficiency in message propagation is caused by the *overlapping* or *saturation* issue. By *overlapping* we mean that a peer may forward the message to some peers that have received the message from other peers, and these actions only waste the communication resources. Since the number of times that the message is transmitted in the system increases exponentially with respect to each peer's transmission effort (the number of neighbors to forward the message to), the probability of overlapping will soon get close to 1 if each peer makes significant relaying efforts, in other words, the system becomes *saturated* very quickly. To reduce the communication inefficiency caused by overlapping, a peer should explicitly consider the overlapping probability, and be able to adjust the transmission effort with the progress of propagation, or the saturation status of the network.

- *Reliability:* Communication efficiency and reliability are two conflicting goals. The intensity of message relaying is positively correlated with the reliability of peer discovery. On an extreme a service provider will be found with probability 1 if the searching is exhaustive by flooding the messages in the system deeply enough. A communication cost saving could in turn cost the reliability. Therefore an efficient message relaying scheme should tradeoff the communication cost and reliability. A peer should decide the optimal relaying effort by considering both the cost and the expected payoff of finding a service provider. The expected payoff of finding a provider not only depends on the value of the service to the requestor, but also on the reliability of finding a provider, which increases with the coverage of the peers that are exposed to the query.

- *Anonymity and information locality:* Although pricing the scarce resource and charging for the usage of the resource via a micro-payment system is a common approach to provide incentive compatibility [7, 18],[1], such a mechanism for a peer to directly purchase services from one another is not feasible in the message relaying P2P discovery system. In a message relaying system, with a micro-payment mechanism the requester would "buy" relaying behaviors of other peers. But such a mechanism requires that the source peer can identify all the intermediate peers and their transmission efforts, which is not feasible in a decentralized P2P system with anonymity. On the other hand, it is neither easy for the requestor or the mechanism designer to decide the right price to charge for each relaying action as the local environment, such as the number of neighbors, of a peer is not known by the mechanism designer or by a third party. Revelation of these kinds of local information is called in [14] *non-private value revelation*. One way to avoid this revelation problem is to ask a peer to price "items" that only require its own local information. In our mechanism the immediate downstream nodes and their responses, and the input incentive are all local information of a peer.

Some people may find this mechanism similar to the multi-level marketing(MLM) model [3, 2]. The incentive mechanism proposed in this paper keeps the advantage of distributive propagation in MLM, but avoids the pyramid effect. A peer is informed of the stage of the propagation by the current hop number that is carried in the message. A peer will estimate the current system state and foresee its behavior on the future propagation based on its position in the "family tree". But the pyramid effect exists in MLM because people are unaware of their positions in the network and the market status, but instead usually misled by the getting-rich stories of a few big distributors that exaggerate the potential opportunities of making money.

The rest of the paper is organized as follows. In Section 2 we present the model of the message relaying mechanism in peer-to-peer systems. Simulation results are provided in Section 3. Section 4 concludes.

2. THE RELAYING MODEL

The propagation of a message in the system is a sequential process. A peer that has received the message is called a *knower*, otherwise it is an *ignorant*. The number of knowers increases while the number of ignorants decreases along with the propagation. The requestor is initially a knower. The *hop* that a peer is located in is defined as its distance from the requestor on the relaying path when it receives the message at the earliest time. Hop 0 has one single peer, the requestor. We assume that peers in an earlier hop conduct relaying earlier than peers in a later hop. Therefore

[1]For example a well known micro-payment incentive mechanism in a file sharing system is to ask a peer to pay certain price for each unit of resources it downloads from other peers [7]

if a peer receives more than once the same query message from different peers, its hop number is defined as the smallest hop number of these senders plus one. Figure 1 shows an example of the construction of hops. In this figure each arrow represents a transmission of the message between two peers. The arrows in solid lines contribute to the identification of the hop for the receivers and they construct the "family tree" of peers. The arrows in dashed lines represent the situations where the message is sent to a knower, and that does not change the hop of the receiver. A peer will relay the message only once, and only respond to its earliest sender (the immediate upstream node on the family tree), i.e., the sender with the smallest hop number. This avoids repeated relaying and reduces repeated queries[2], and is also used in Gnutella [1]. This is also in the self-interest of a peer since the incentive that is passed along a relaying path decreases by hops. An earlier sender gives a higher reward on average.

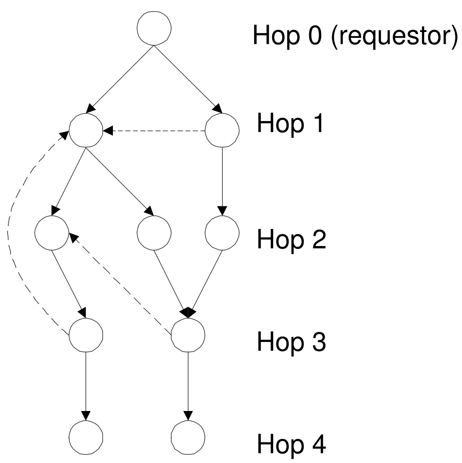

Figure 1: Construction of hops

We make the following assumptions in this initial study:

1. Single provider: There is only one service provider in the network.
2. Homogeneity: Each node, except the requestor, has the same *ex ante* probability to be the service provider. This may be violated in a social network in which a peer learns from the past experience on the expertise and functionality of acquaintances and neighbors. But it is true in a simple network such as a sensor network.

Let $c(k)$ denote the cost of relaying a message to k neighbors. The cost $c(k)$ is an increasing convex function of k. [3] Denote by N the average number of peers in the network. Without confusion we also let N denote the collection of peers. The value of the resource/service to the source peer is v_0. We say peer j is peer i's *downstream* node if j is located on a transmission path which includes i at the upstream, particularly j is i's *immediate downstream* node or a *child* if j is reached by i.

[2]Repeated queries cannot be completely eliminated because a peer cannot identify a knower or an ignorant.

[3]There can be different interpretations for the convexity of the cost function: (1) the resource to spend on a transmission increases with the distance between the sender and the receiver and the closer neighbors are covered before the further neighbors; (2) the cost also includes the cost on discovering neighbors, whose margin increases; (3) the cost on bandwidth is measured by the average time delay, which is a convex function of the arrival rate of transmission tasks [11].

2.1 Equilibrium analysis

Let D_i be the degree of peer i and the transmission effort k_i of peer i is defined on $\mathcal{K}_i = \{0, 1, \ldots, D_i\}$. Let $\mathcal{V} = [0, \bar{v}]$ be the space of incentives and $\mathcal{H} = \{0, 1, \ldots, H\}$ the set of possible hop numbers where H is the maximum hop number. The strategy S_i of a peer i is defined as $(k_i, u_i) = S_i(h_i, v_i) : \mathcal{H} \times \mathcal{V} \to \mathcal{K} \times \mathcal{V}$, where $h_i \in \mathcal{H}$ is the hop number of i, $v_i \in \mathcal{V}$ is the input incentive, $k_i \in \mathcal{K}$ is the transmission effort, $u_i \in \mathcal{V}$ is the output incentive. If i is the requestor, $h_i = 0$. Let SP_i be the strategy space of peer i, and $S = \{S_i\}_{i=1}^N$, $S_i \in SP_i$ be a strategy profile of peers. Denote the expected utility of peer i by U_i, and the expected number of downstream nodes by L_i. The probability that the provider is found via peer i's transmission is equal to L_i/N. Then the expected utility of i is equal to

$$U_i(S) = (v_i - u_i)L_i(S)/N - c(k_i). \tag{1}$$

Note that L_i not only depends on peer i's transmission effort k_i, but also on the other peers'. This is not only because peers compete in searching for the service provider, but also because the total number of downstream nodes of a peer depends on the transmission efforts of its downstream nodes, which again are impacted by the input incentives that are passed on from their upstream nodes.

The strategy space $SP = \prod_{i=1}^N SP_i$ and utility functions $\{U_i : SP \to \mathbb{R}\}_{i=1}^N$ define a message relaying game G among the peers. A *subgame* G_h is the propagation process starting from the hop h to the last hop H, given peers' strategies $\{S_i(l, \cdot)\}_{i \in N}$ at each earlier hop l, $l = 0, 1, \ldots, h-1$. A *subgame perfect Nash equilibrium (SPNE)* is a profile of peers' strategies which constitutes a Nash equilibrium for each subgame G_h, $h = 0, 1, \ldots, H$. In the rest of this section we provide some analytical features of SPNE.

We first study a SPNE of the transmission strategies $\{k_i^*(h_i, v_i|u_i)\}_{i \in N}$ assuming the incentive transfer strategies $\{u_i(h_i, v_i)\}_{i \in N}$ are given. Proposition 2.1 summarizes some features of $k_i^*(h_i, v_i|u_i)$.

PROPOSITION 2.1. [4] *Given the incentive relaying strategy $u_i(h_i, v_i)$ for each peer i, a SPNE of the transmission strategies $\{k_i^*(h_i, v_i|u_i)\}_{i \in N}$ exists. $k_i^*(h_i, v_i|u_i)$ decreases with the increase of h_i or u_i, or with the decrease of v_i.*

Unfortunately we cannot prove the existence of a joint SPNE of both the transmission strategy and the incentive transfer strategy, although the existence is guaranteed if the incentive transfer strategy can be a mixed strategy [12]. Proposition 2.2 provides a feature of a best response strategy. Proposition 2.3 shows the existence and uniqueness of a symmetric SPNE when all peers are symmetric.

PROPOSITION 2.2. *With a best response strategy the increase of the input incentive leads to the increase of the expected number of downstream nodes.*

[4]The proofs of this proposition and other statements are all presented in Appendix.

PROPOSITION 2.3. *If all peers have the same degree, a symmetric SPNE exists and is unique.*

Note that although based on Proposition 2.1, the transmission effort decreases by hops when the input incentive and output incentive are the same, this is not necessary in a SPNE when the output incentive is also a decision variable. This is because k_i and u_i are somehow substitutable. To expand the coverage of descendants, a peer i can increase the number of immediate downstream nodes by increasing k_i, which incurs more transmission cost. Or it can encourage the transmission efforts of its downstream nodes by increasing u_i, which sacrifices the future payoff. Which approach to use depends on their relative expensiveness, and can be impacted by the form of the cost functions and the position of the peer in the propagation (the hop number).

Although the transmission efforts may not decay by hops in the equilibrium, Proposition 2.2 shows that when the input incentive decreases, either the immediate or the future transmission efforts will decrease. Since the input incentive decreases by hops along a relaying path, it means that with this incentive mechanism the flooding problem and pyramid effect are automatically avoided with peers' individual self-interested behaviors.

2.2 Approximation of symmetric SPNE

Based on Equation 1 we can see that a SPNE of the message relaying game, if exists, is not directly computable because of the complexity of L_i, $i \in N$, as a function of all peers' strategies. In this section we present an approximation of L_i, which allows the calculation of an approximated symmetric SPNE. In this solution peers are assumed to have the same degree, and therefore the strategies of peers in the same hop are identical. In the following presentations we only differentiate the strategies by hops with a subscript h.

Suppose there are n_h knowers in the system and m_h knowers in hop h before the peers in hop h forwarding the message, $h = 0, 1, \ldots, H + 1$.[5] Initially the only knower is the requestor before the propagation and hence $n_0 = m_0 = 1$. There is the connection $m_h = n_h - n_{h-1}$ for $h = 1, \ldots, H+1$. If each of the peers in hop h relays the message to k_h neighbors, the expected number of ignorants $\hat{k}_h$ reached by each sender in hop h can be calculated as follows:

$$\hat{k}_h = \begin{cases} 0 & \text{if } k_h = 0 \text{ or } m_h = 0 \\ \frac{N-n_h}{m_h}[1-(1-\frac{k_h}{N-1})^{m_h}] & \text{else.} \end{cases} \quad (2)$$

The process of deriving $\hat{k}_h$ is described in Appendix.

Taking the expectation $\hat{k}_h$ as certainty, we can estimate the expected number of nodes in each hop and the number of knowers as follows:

$$m_{h+1} \doteq m_h \hat{k}_h = (N - n_{h-1} - m_h)[1 - (1 - \frac{k_h}{N-1})^{m_h}], \quad (3)$$

$$n_{h+1} \doteq n_h + m_h \hat{k}_h. \quad (4)$$

Therefore we can estimate the expected number of descendants of a peer in hop h:

$$L_h(n_h, m_h) = \sum_{l=h+1}^{H} m_l/m_h,$$

and the expected utility U_h of a peer in hop h can be approximated by:

$$\begin{aligned} &U_h(k_h, u_h|v_h, n_h, m_h, \{k_l\}_{l=h+1}^{H}) \\ &= (v_h - u_h)L_h/N - c(k_h) \end{aligned} \quad (5)$$

Given U_h for each hop h, we can calculate an approximate SPNE $\{(k_h^*, u_h^*)\}_{h=0}^{H}$ by a backward induction.

3. SIMULATION

In this section we provide experimental results comparing the system performances of the distributed incentive mechanism, the breadth-first searching mechanism and the centralized mechanism. The strategy in the distributed incentive mechanism is based on the approximate symmetric SPNE calculated following the approach in Section 2.2. The approximate optimal policy in the centralized mechanism is calculated by searching all the possible transmission effort profiles of the hops (in a symmetric network) to maximize the total utility based on the approximation of the coverage functions (3) and (4). We are interested in (1) the total utility of the system, (2) the total coverage, i.e., the number of peers exposed to the message, at the end of the propagation, and (3) the distribution of transmission efforts and incentives over the hops in the distributed incentive mechanism, with the change of the query's value (v_0) and different types of cost functions.

We let $N = 50$, v_0 change from 10 to 30, $H = 2$ (three hops) and $D = 6$. We examine two cost functions: a linear function $c_{linear}(k) = 0.1k$, and a strictly convex function $c_{convex}(k) = 0.015k^2$.

- System utility:

The system utilities based on different mechanisms are illustrated in Figure 2. Denote by U_{dist}, U_{BFS} and U_{cen} the system utilities in the distributed incentive mechanism,

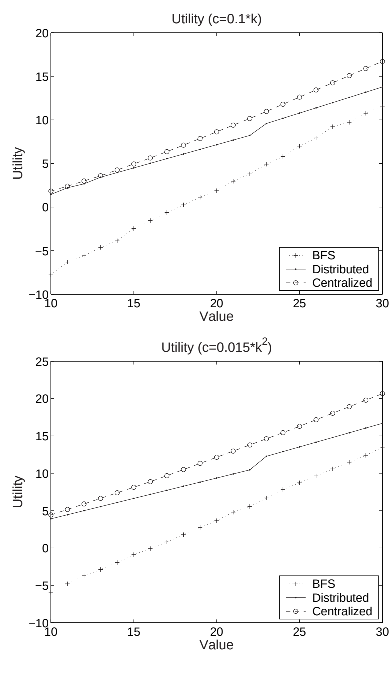

Figure 2: Comparing the system utility

[5]By n_{H+1} and m_{H+1} we count the peers that are reached by the last hop.

in the breadth-first search and in the centralized system. Figure 2 shows that $U_{BFS} < U_{dist} < U_{cen}$, with U_{dist} generally higher than 80% of U_{cen}. Also $U_{dist} - U_{BFS}$ decreases with the increase of the query's value because the exhaustive breadth-first search is closer to a system optimal policy if the value of the query is higher. When the query value is small, the breadth-first search could bring a negative utility.

- Coverage:

The total number of peers that receive the query message during the propagation process based on each mechanism is recorded in Figure 3. Breadth-first search always covers

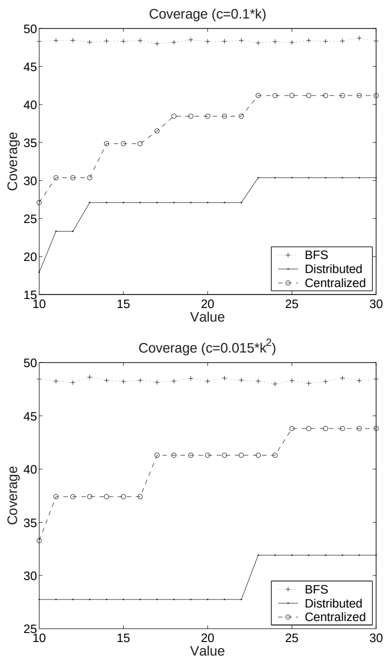

Figure 3: Comparing the coverage

more peers than the other two mechanisms, which implies a higher reliability (but its utility is lower as shown in Figure 2). Because each peer transmits the message to all its neighbors in breadth-first search, the coverage is independent of the query value. In the distributed and centralized mechanisms the coverage increases with the query value. But the distributed mechanism cannot achieve the coordination among peers as required in an optimal propagation, and results in the coverage lower than the coverage in the centralized system.

- Distribution of transmission efforts and incentives:

The transmission efforts and incentive transfer strategies[6] in an approximate symmetric SPNE are illustrated in Figure 4 and 5 respectively.

Compare the upper (with a linear cost function) and lower (with a convex cost function) plots in Figure 4, we can find that with a linear cost function transmission efforts are more concentrated in downstream hops, while with a convex cost function earlier hops spend more efforts in relaying than later hops. Correspondingly Figure 5 shows that with a

[6]We only show the output incentive from the first two hops because in the last hop the output incentive is always zero.

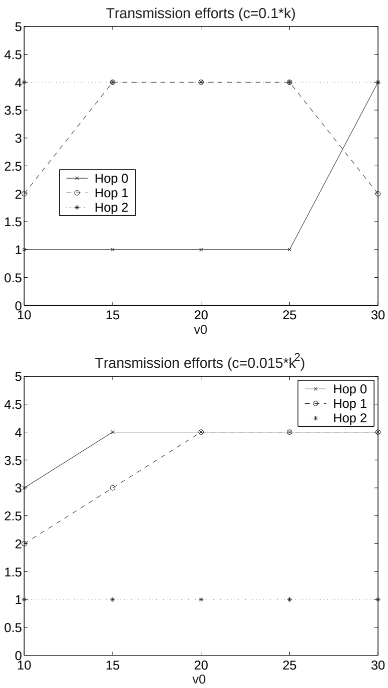

Figure 4: Distribution of transmission efforts

linear cost function peers pass on most of the input incentives to downstream nodes, while with a convex cost function peers keep most of the input incentives. These observations reveal close connections between the transmission distribution and the convexity of the cost function. When the cost function becomes more convex, the cost for inducing downstream propagations increases faster than the cost to directly increase the immediate downstream nodes. Actually when the cost is linear, a last hop peer will search all its neighbors as long as the average payoff is no less than the unit relaying cost.

4. CONCLUSION

In this paper we present an incentive mechanism for message relaying in peer-to-peer discovery. In this problem the common micro-payment protocol based on the relaying actions is not feasible for an anonymous message relaying process. By pricing the searching result but not the searching behavior our mechanism provides appropriate incentives for distributed message relaying, and achieves tradeoff between the searching cost and reliability.

In our near future work we are interested in evaluating the incentive mechanism and understanding in greater depth how are the strategies and performances impacted by the system parameters (such as the cost function, TTL, the system scale, and the query value, etc.) with more possible analytical and simulation work. In this paper we have made the assumptions that there is one single service provider, peers are homogeneous, simple and do not discriminate their neighbors. When there are more than one potential service provider, special attention has to be given to the propagation stopping rule and other related issues. For example, should the propagation be stopped once a service provide is

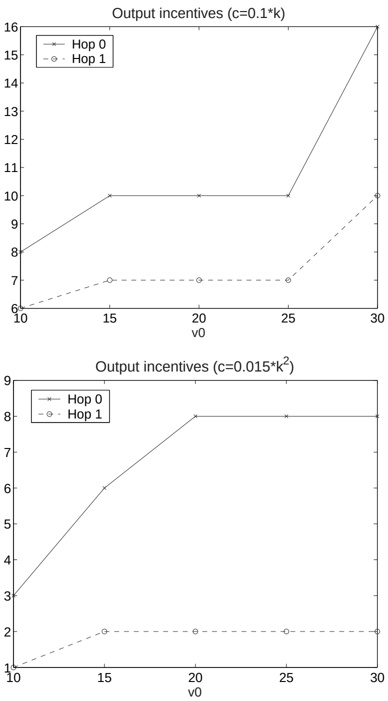

Figure 5: Distribution of incentive transfers

found, and how? If more than one provider is found, how are contributors be rewarded? In our further future work we will also study the adaptive strategies of learning peers in such an incentive mechanism in more complex situations such as where peers are heterogeneous and are able to accumulate knowledge from past experience.

5. REFERENCES

[1] The Gnutella protocol v0.4.

[2] Multi-level marketing(a.k.a. network marketing & referral marketing). http://skepdic.com/mlm.html.

[3] What's wrong with multi-level marketing? www.vandruff.com/mlm.html.

[4] Eytan Adar and Bernardo Huberman. Free riding on gnutella. *First Monday*, 5(10), 2000.

[5] Sam Chandan and Christiaan Hogendorn. the bucket brigade: Pricing and network externalities in peer-to-peer communications networks. In *Telecommunications Policy Research Conference*, 2001.

[6] Anthony Ephremides. Ad hoc networks: not an ad hoc field anymore. *Wireless communications and mobile computing*, 2:441–448, 2002.

[7] Philippe Golle, Kevin Leyton-Brown, Ilya Mironov, and Mark Lillibridge. Incentives for sharing in peer-to-peer networks. In *Proceedings of the Second International Workshop on Electronic Commerce*, 2001.

[8] Chaki Ng, David Parkes, and Margo Seltzer. Strategyproof computing: Systems infrastructures for self-interested parties. In *The First Workshop on Economics of Peer-to-Peer Systems*, Berkeley, CA, June 2003.

[9] C. H. Papadimitriou. Algorithms, games, and the internet. In *Annual ACM Symposium on Theory of Computing (STOC)*, 2001.

[10] Lakshmish Ramaswamy and Ling Liu. Free riding: A new challenge for peer-to-peer file sharing systems. In *Proceedings of the 36th Hawaii International Conference on Systems Science*, 2003.

[11] Sheldon M. Ross. *Introduction to Probability Models.* Academic Press, 2000.

[12] Drew Rudenberg and Jean Tirole. *Game Theory.* The MIT Press, 2000.

[13] Jeff Shneidman and David Parkes. Rationality and self-interest in peer-to-peer networks. In *The Second International Workshop on Peer-to-Peer Systems (IPTPS'03)*, Berkeley, CA, February 2003.

[14] Jeff Shneidman, David Parkes, and Margo Seltzer. Overcoming rational manipulation in distributed mechanism implementations. Technical Report TR-12-03, Harvard University, 2003.

[15] Dejan S.Milogicic, Vana Kalogeraki, Rajan Lukose, and etc. Peer-to-peer computing. Technical Report 2002-57RI, Hewlett-Packard Company, 2002.

[16] Donald M. Topkis. *Supermodularity and Complementarity.* Princeton University Press, 1998.

[17] Beverly Yang and Hector Garcia-Molina. Improving search in peer-to-peer systems. In *Proceedings of the 22nd International Conference on Distributed Computing Systems (ICDCS)*, Vienna, Austria, 2002.

[18] Beverly Yang and Hector Garcia-Molina. PPay: Micropayments for peer-to-peer systems. In *Proceedings of the 10th ACM Conference on Computer and Communications Security (CCS)*, Washington D.C., October 2003.

Appendix

Let $L_i(k)$ denote the expected number of descendants of i when its transmission effort is k, given the incentive transfer strategy u_i, the input variables h_i and v_i, and other peers' strategies. Before proving Proposition 2.1, we need to show the following lemma first,

LEMMA 5.1. *$L_i(k+1) - L_i(k)$ decreases with the increase of h_i.*

Proof of Lemma 5.1: Let the increase of L_i be denoted by $\Delta_k L_i(h)$ when $h_i = h$ and k_i is increased from k to $k+1$. $\Delta_k L_i(h)$ is positive. Denote by $\Delta_k \hat{L}_i(h-1)$ the increase of the number of downstream nodes in hops from h to $H-1$. $\Delta_k L_i(h)$ and $\Delta_k \hat{L}_i(h-1)$ involve the same number of hops, but the hops in $\Delta_k \hat{L}_i(h-1)$ are one hop earlier than those in $\Delta_k L_i(h)$. $\Delta_k L_i(h) \le \Delta_k \hat{L}_i(h-1)$ because the probability of sending a message to a knower is smaller at earlier time of the propagation process (corresponding to a smaller hop number). Then $\Delta_k L_i(h-1) \ge \Delta_k \hat{L}_i(h-1) \ge \Delta_k L_i(h)$. $\square$

Proof of Proposition 2.1: We can prove by backward inductions starting from the last hop. When $h_i = H$, $U_i = (v_i - u_i)L_i(k_i)/N - c(k_i)$. Denote by $p_i(k)$ the marginal probability of i reaching an ignorant by relaying the message to one more peer when the transmission effort is k, and $\Delta L_i(k_i) = L_i(k_i+1) - L_i(k_i)$. $p_i(k)$ decreases with k because more peers will become knowers with k bigger and therefore it is more difficult to reach an ignorant in the future. Therefore $\Delta L_i(k) = p_i(k)$ decreases with k, and $L(k)$

is an increasing concave function of k. Since $c(k_i)$ is a convex function of k_i, the optimal transmission effort k_i^* that maximizes U_i exists and is unique, and k_i^* increases with v_i. For a peer i in the last hop, k_i^* is independent of u_i because there is no propagation by its immediate downstream nodes.

Now we assume for a peer i in hop $h_i \geq h+1$, $k_i^*(h_i, v_i)$ exists and is unique, and $k_i^*(l, v_i) \leq k_i^*(l+1, v_i)$ for $l \geq h+1$ and $l \leq H-1$. In the following we prove that the conclusion also holds for a peer i with $h_i = h$ given k_j^* with $h_j \geq h+1$, $j \in N$. The proof is organized in the following three steps.

Step 1: $\Delta L_i(k_i) \geq 0$ and decreases with k_i.

We can analyze based on one sample path of propagation in which peer i sends the message to k_i neighbors. Now let k_i be increased by 1, and denote the new child of i in hop h_{i+1}, if there is any, by j. Denote by L_i and $\hat{L}_i$ the numbers of downstream nodes of i when the transmission efforts of i are k_i and k_i+1 respectively.

Case 1: If j is not a downstream node of i in the sample path, then the new set of downstream nodes of i includes the old set in the sample path.

Case 2: If j is an un-immediate downstream node of i in the sample path, then the hop number of j is reduced. Since the transmission effort decreases with the hop number, j relays the message to more neighbors and the number of downstream nodes of j is increased based on the induction assumption.

Based on the assumption of induction, $L_i(k_i)$ increases with k_i given the transmission equilibrium $k_j^*(h_j, v_j)$ for any peer j with $h_j \geq h+1$.

Since the probability of reaching an ignorant decreases with the number of knowers, $\Delta L_i(k_i)$ decreases with the increase of k_i.

Step 2: k_i^* exists, and k_i^* decreases with u_i and increases with v_i.

Since $\Delta L_i(k_i)$ decreases with the transmission strategies of other peers in the same hop, U_i is a submodular function of the transmission efforts of all peers in the hop h_i, and an equilibrium of their transmission efforts exists [16].

Since $k_i = k_i^*$ maximizes $U_i(k_i) = (v_i - u_i)L_i(k_i) - c(k_i) = 0$ and $\Delta L_i(k_i)$ decreases with k_i, k_i^* will decrease (increase) if u_i (v_i) increases.

Step 3: $k_i^*(h_i, v_i) \geq k_i^*(h_i+1, v_i)$.

Based on Lemma 5.1, if h_i decreases, $\Delta L_i(k_i)$ will increase. This leads to the increase of k_i^* since $L_i(k_i)$ decreases with k_i. □

Proof of Proposition 2.2:

Let $v' - v = \delta > 0$. Denote by (u', k') ((u, k)) the best response strategy, U' (U) the expected utility and L' (L) the expected number of downstream nodes at the best response strategy when $v_i = v'$ ($v_i = v$). Then

$$U' = (v' - u')L'/N - c(k'),$$

$$U = (v - u)L/N - c(k).$$

But we know $U' \geq (v' - u)L/N - c(k) = U + \delta L/N$ and $U \geq (v - u')L'/N - c(k') = U' - \delta L'/N$. Then we have $\delta L/N \leq U' - U \leq \delta L'/N$. Therefore $L \leq L'$ since $U' - U \geq 0$. □

Proof of Proposition 2.3: If there exists a SPNE, the SPNE is symmetric because it is a symmetric game. Given a SPNE of the transmission strategy $k_i^*(u_i)$ based on an incentive transfer strategy u_i, the original game can be transferred to a game with only the incentive transfer decisions, and the utility function is

$$U_i(u_i|S_{-i}) = (v_i - u_i)\frac{L_i((k_i^*(u_i), u_i)|S_{-i})}{N}. \qquad (6)$$

A symmetric SPNE $\{u_i^*\}_{i\in N}$ can be found with backward inductions with respect to the hops. Then $\{(k_i^*(u_i^*), u_i^*)\}_{i\in N}$ is a symmetric SPNE of the original game. □

Derivation of Equation 2:

Suppose the nodes in hop h forward the message sequentially. Denote by $\hat{k}_h^j$ the expected number of new nodes reached by the j-th mover. Let $l_h = k_h(N - n_h)/N$ and $p_h = 1 - k_h/N$. l_h can be interpreted as the number of immediate downstream nodes of a peer assuming receivers of different peers in the same hop will not overlap. p_h can be interpreted as the probability

$$\hat{k}_h^1 = k_h \frac{N - n_h}{N} = l_h$$

$$\hat{k}_h^2 = k_h \frac{N - n_h - \hat{k}_h^1}{N} = l_h(1 - \frac{k_h}{N}) = l_h p_h$$

$$\hat{k}_h^3 = k_h \frac{N - n_h - \hat{k}_h^1 - \hat{k}_h^2}{N} = \frac{k}{N-1}(N - n_h - l_h - l_h p_h) = l_h p_h^2$$

$$\vdots$$

$$\hat{k}_h^j = l_h p_h^{j-1}$$

Then

$$\begin{aligned}\hat{k}_h &= \frac{1}{m_h}\sum_{j=1}^{m_h}\hat{k}_h^j = \frac{l_h}{m_h}\sum_{j=1}^{m_h} p_h^{j-1} = \frac{l_h(1 - p_h^{m_h})}{(1 - p_h)m_h} \\ &= \frac{N - n_h}{m_h}[1 - (1 - \frac{k_h}{N})^{m_h}]\end{aligned}$$