Università degli Studi di Genova
Dottorato in Informatica
(XX Ciclo)

Progress Report

Design of a Reputation-based Incentive System for P2P Applications

Candidate: Matteo Dell'Amico
dellamico@disi.unige.it
Dipartimento di Informatica e Scienze dell'Informazione
Università di Genova

Advisor: Giovanni Chiola
chiola@disi.unige.it
Dipartimento di Informatica e Scienze dell'Informazione
Università di Genova

November 2006

Abstract

Free riding, the behaviour of exploiting resources contributed by other nodes while not sharing anything valuable, is a plague in many P2P networks. Its main cause is that sharing resources has a cost which is not, or not sufficiently, counterbalanced by benefits.

Incentive systems have been developed in order to reward fair users and punish free riders; the most successful existent implementations obey the principle of *direct reciprocation*: nodes give to other peers a service having a quality which is proportional to the service quality received from them.

The *indirect reciprocation* approach, which is based on evaluating peers based on their reputation (a judgement of past interactions with *any* node), is more powerful. In fact, in the common case of large networks, peer pairs have a low probability of having interacted before, and nodes can be judged more accurately when having access to more data.

The goal of this thesis is to develop an indirect reciprocation system designed to be secure and efficient enough to be implemented in large-scale networks, where nodes are free to create new identities at any time.

The main concerns will be finding good metrics for reputation and algorithms to calculate them efficiently, rewarding peers that give accurate information about their history, providing significant experiments to evaluate the obtained benefits, implementing the designed system in an existing or new P2P network.

1 Introduction

Trust and Resource Allocation A crucial problem in many network applications is determining how to assign *trust*. Users, in fact, are given access to resources (e.g., information, bandwidth, storage space, and/or CPU processing time) when they can be *trusted* to fairly use the application. This problem is even more important in P2P applications, where nodes are both consumers and service providers, and the overall performance of the system is thus dependent on the quality of service given by peers.

In traditional client-server applications, the server administrator usually manages a database describing who the trusted users are, and trust is granted to a user only after the server has authenticated her/him. In large-scale P2P applications, the scenario is completely different: a centralised administration of such a "trust database" would unfortunately deny many of the well-known qualities of decentralised applications, creating a single point of failure and a possible performance bottleneck.

In P2P applications, nodes often have a high degree of freedom regarding the amount of resources they can make available to other peers in the network. Usually, the node pays a small cost in sharing those resources, while peers using those resources will receive the benefit associated with the use of that resource.

In applications developed without regard to trust evaluation, giving each peer the same level of access to resources, there is indeed an incentive to *free riding*: a "selfish" behaviour in which nodes consume resources given by other peers, while not giving their available resources back to the peer community in return. Indeed, measurements done on the Gnutella [2] file-sharing network in 2000 [3] showed that nearly 70% of the nodes did not share any files.

An idea for solving this problem is introducing *reciprocity*. This principle is applicable whenever interaction between actors is repeated. According to this principle, entities have a better attitude towards peers that cooperated with them in the past. Cooperative behaviour is thus rewarded with a better outcome [4].

The reciprocity approach has been successfully implemented in P2P file-sharing networks, both in BitTorrent [10] and in EDonkey2000 [1]. These applications have been so successful that they are evaluated to accont for approximately half of the total Internet traffic [7, 8].

Webs of Trust The reciprocity approach is useful under the assumption that two interacting nodes have high probability of meeting again in the future. In a P2P network, this is true whenever networks are small, nodes usually interact with the same peers, and/or mutual interactions have a sufficiently large duration (e.g., two peers sharing complementary pieces of a big file). Unfortunately, there are many relevant cases in which these assumptions do not hold, and the incentives to cooperation are thus lost.

The notion of *reputation* is introduced to solve this problem. Nodes make available to the whole network their *recommendations* of other peers, earned due to past encounters or trust relationships between node administrators. Using data from recommendation restores incentives to cooperation [21].

In the following, we will call *web of trust* (WoT) the graph formed having peers as nodes and recommendations as edges. Webs of trust are complex networks that can be seen as related both to technological and social networks. Such networks have been extensively studied in recent years [20], and have recurrent characteristics such as the scale-free property (power-law distribution of node degree), and the small-world property (high clustering and low diameter).

Based on a concept of "trust transitivity" (nodes recommended by a trusted node deserve some trust themselves), reputation values are dependant on paths in the web of trust. In obtaining reputation values from webs of trust, some issues are particularly relevant in P2P networks:

Lack of global knowledge It can be impossible for a node of the network to have up-to-date information about the whole web of trust, since they can both be huge (some P2P applications have millions of users) and highly volatile. Possible approaches could be using distributed data structures such as DHTs, or evaluating reputation using only local knowledge for the interested peers.

Cheap identities A malicious user can create a big number of identities, in order to subvert the behaviour of the system (*Sybil attack*), or to erase previous history of misbehaviour (*whitewashing*). The opportunity of whitewashing forces us to take in consideration only good recommendations

[14], while Sybil attacks can be avoided by using metrics that are resilient to that kind of attack, such as maximal flow [9].

Goal The goal of the thesis is to design a system that uses reputation evaluation effectively in a P2P network, in order to significantly encourage cooperation, so that it can be used in a wide application domain. Simulation experiments will be conducted in order to evaluate the effectiveness of the system. This system is planned to be integrated in an existing P2P infrastructure.

The resulting system will be completely decentralised, thus avoiding single points of failure and performance bottlenecks. It will be applicable in situations where nodes have the possibility of arbitrarily creating new identifiers for them, thus making their past, possibly malicious, history inaccessible. Common characteristics of complex networks are taken into consideration, in order to build systems that exploits them. Link analysis techniques [6] will be used in order to calculate metrics that are both significant and resilient to attacks.

2 Neighbourhood Maps

A first contribution of the thesis is the introduction of *neighbourhood maps*, a means for efficient approximated evaluation of reputation metrics in a web of trust.

Reputation values can be seen as dependent from the paths in the web of trust that start from the evaluating node A and reach the node B that is being evaluated. Each node builds its neighbourhood maps by iteratively contacting its neighbours, collecting reputation values for a small ($O\left(\sqrt{n}\right)$) number of highly-rated nodes.

When A wants to rank B, data from the neighbourhood maps of A and B is combined, looking for nodes that appear in both maps. For instance, if the reputation metric to be evaluated is the length of the path representing minimal distance, if a node x is found at distance 2 from A and distance 3 from B, the distance from A to B will be estimated as 5.

Some properties are commonly found in social networks. The "small world" property guarantees that the network is characterised by a low diameter. The "scale-free" property guarantees that some nodes (*hubs*) are highly relevant for the network and are thus likely to be part of many neighbourhood maps. The reasonable assumption that these characteristics will be present in the web of trust creates confidence in the belief that neighbourhood maps are likely to have non-empty intersections even when they are small.

In [12], experimentation on the web of trust of OpenPGP [16] showed that graph distance could be evaluated with exponential convergence, and a relative error less than 1% when the size k of the neighbourhood maps is greater than $2\sqrt{n}$. The web ranking algorithm PageRank [22] could be approximated with a Kendall similarity[1] of 0.8 when $k > 3\sqrt{n}$.

[1] Given two different ranking methods, the Kendall similarity is the probability that, given two random nodes, the two methods put the same node at the top.

In future work, neighbourhood maps will be extended to store the subgraph of the neighbourhood, in order to evaluate different metrics – such as maximal flow – that have desirable behaviour and properties [13], and are difficult to estimate using the present methodology.

3 Layouts in Webs of Trust

3.1 From the Layout to the Network

Models that attempt to explain the emergence of clustering in small-world networks often do it by immersing networks in lattices [24, 18], where nodes have a high probability to be linked to nodes at a short distance, and low probability to be linked to far away nodes. We can justify the lattices as an expression of a layout in an underlying topology, where closeness represents affinity between nodes. Depending on the particular network, two nodes can be seen as affine if they are for instance geographically close, or if they represent people with similarities in behaviour.

The scale-free property can be modeled as a result of a "rich get richer" mechanics: when the networks evolves over time, nodes that have many incoming links are likely to get connected to new nodes that enter the network. In the Barabási-Albert model [5], new nodes connect to an old node j with a probability that is proportional to its degree k_j.

A simple variation on these two models is proposed in our work [11], by extending the Barabási-Albert model and incorporating the idea of layout in it, allowing us to model networks that are both scale-free and small-world. In the resulting model, nodes are randomly placed in a simple topology such as a ring, and a new node i gets connected to an old node j with a probability that is proportional to $k_j^\alpha / d_{ij}^\sigma$, where d_{ij} is the Euclidean distance that separates node i and node j. Using different parameters α and σ gives us a family of models, where both small-world and scale-free characteristics are found when $\alpha = \sigma \geq 1$.

3.2 From the Network to the Layout

Results seen in Section 3.1 suggest that an underlying layout may be "hidden" in networks where high clustering is present. Discovering this layout can be important for our purposes, since we are interested in finding short paths that connect nodes in the WoT. If distance in the layout is related to shortest-path distance in the WoT, we can use the coordinates in the layout as hints for doing efficient "routing" between nodes using the approach seen in [18]. An existent effort in this direction uses a Metropolis-Hastings algorithm in order to obtain a good layout [23].

It is possible to use this idea in combination with the neighbourhood maps seen in Section 2: maps from nodes "half-way" in the path between nodes can be requested, in order to give nodes a clearer view of the networks, enhancing the precision of such techniques given a fixed size for the maps.

The layout should put connected nodes at short distance, while separating as much as possible the others. Algorithms for drawing graphs have the same goals, even if we have the additional constraint of needing an efficient and decentralised algorithm for large-scale systems. Moreover, our layouts are not constrained to 2 or 3 dimensions, like the ones needed for visualisation.

Many well-known algorithms for graph visualisation are based on a "spring model" [17]: repulsive forces pull nodes apart, and attractive forces bind linked nodes together. These algorithms are unfortunately difficult to implement on P2P networks, since each iteration of the algorithm is $O\left(n^2\right)$, and require each node to exchange information about its position with each other node in the network.

Another, more efficient, algorithm, is based on spectral properties of graphs [19]. The algorithm minimises the distance between connected nodes, with the constraint of positions having a finite variance to avoid the whole system from collapsing in a single point. The iterative algorithm resulting from it, at each step:

1. puts each node at the centre of the position of all its neighbours;
2. normalises variance in order to avoid nodes collapsing in a single point;
3. avoids that nodes get placed on a straight line by imposing that the co-variance of the nodes positions gets to zero.

Step 1 is efficiently implementable in a P2P network; steps 2 and 3 can be solved by calculating mean, variance and co-variance using gossiping algorithms over the WoT.

Work is underway in order to evaluate how to efficiently implement the algorithm, for example by skipping steps 2 and 3 in most iterations. Preliminary results show that routing achieves efficiency that is roughly comparable to [23] when dimensions are few, and converge towards the optimum when dimensionality grows.

4 Experimentation

The effectiveness of the developed reputation evaluation means will be evaluated through experimentation both on real and synthetic networks. The synthetic networks will be generated using models such as the ones described in Section 3.1, as well as using well-known and established models. Real networks will be used too, using data such as the OpenPGP web of trust [16, 15], and traces of interactions in on-line transactions systems such as eBay [25].

A first evaluation will be towards the precision of the approximate metrics that are developed using work from Sections 2 and 3.2. This will help understanding the trade-offs involving precision of evaluation versus computational efficiency.

A second step will be based on evaluating the effectiveness of using the reputation systems, using evolutionary dynamics along the style of [4, 13, 21], perhaps differentiating the goals of the peer population (e.g., some peers may just want to receive service from the network, other nodes may be motivated to maximise the diffusion of their material, others may want to disrupt the system). This experimentation will help us in understanding how much the developed system in reality encourages cooperation, with respect to simple strategies and other reputation evaluation systems.

A final goal is the integration of this reputation system in an already existent system (such as, for instance, BitTorrent), in order to evaluate the usefulness of this system in the large.

5 Planned Thesis Structure

The thesis will consist of 3 parts.

1. Reputation evaluation: state-of-the-art and novel methods for evaluating reputation, such as the ones discussed in Sections 2 and 3, will be presented.

2. Mechanisms for Cooperation: methods seen in the first part will be evaluated in an experimental scenery, such as the one discussed in Section 4.

3. Prototype implementation: using results from the previous part, the design for a prototype of a reputation system will be shown.

References

[1] Edonkey2000. http://www.endonkey2000.com.

[2] Gnutella - A Protocol for a Revolution. http://rfc-gnutella.sourceforge.net.

[3] Adar, E., and Huberman, B. A. Free riding on gnutella. *First Monday 5*, 10 (2 Oct. 2000).

[4] Axelrod, R. *The Evolution of Cooperation.* Basic Books, New York, 1984.

[5] Barabási, A.-L., and Albert, R. Emergence of scaling in random networks. *Science 286*, 5439 (Oct. 1999), 509–512.

[6] Borodin, Roberts, Rosenthal, and Tsaparas. Link analysis ranking: Algorithms, theory, and experiments. *ACMTIT: ACM Transactions on Internet Technology 5* (2005).

[7] CacheLogic. The True Picture of Peer-to-Peer Filesharing. http://www.cachelogic.com/research/p2p2004.php, July 2004.

[8] CacheLogic. Peer-to-Peer in 2005. http://www.cachelogic.com/research/p2p2005.php, 2005.

[9] Cheng, A., and Friedman, E. Sybilproof Reputation Mechanisms. In *Third Workshop on Economics of Peer-to-Peer Systems, Philadelphia, PA.* (22 Aug. 2005).

[10] Cohen, B. Incentives build robustness in bittorrent. In *Proceedings of the Workshop on Economics of Peer-to-Peer Systems* (Berkeley, CA, USA, 2003).

[11] Dell'Amico, M. Highly Clustered Networks with Preferential Attachment to Close Nodes. In *Proceedings of ECCS 2006* (Nov. 2006).

[12] Dell'Amico, M. Neighbourhood Maps: Decentralised Ranking in Small-World P2P Networks. In *3rd International Workshop on Hot Topics in Peer-to-Peer Systems (Hot-P2P)* (Apr. 2006).

[13] Feldman, Lai, Stoica, and Chuang. Robust incentive techniques for peer-to-peer networks. In *CECOMM: ACM Conference on Electronic Commerce* (2004).

[14] Friedman, E. J., and Resnick, P. The Social Cost of Cheap Pseudonyms. *Journal of Economics & Management Strategy* (Aug. 17 2001).

[15] Henk P. Penning. PGP pathfinder and key statistics. http://www.cs.uu.nl/people/henkp/henkp/pgp/pathfinder/.

[16] Jörgen Cederlöf. Wotsap. http://www.lysator.liu.se/ jc/wotsap/index.html.

[17] Kamada, T., and Kawai, S. An algorithm for drawing general undirected graphs. *Information Processing Letters 31*, 1 (12 Apr. 1989), 7–15.

[18] Kleinberg, J. M. The small-world phenomenon: an algorithm perspective. In *STOC* (2000), pp. 163–170.

[19] Koren. On spectral graph drawing. In *COCOON: Annual International Conference on Computing and Combinatorics* (2003).

[20] Newman, M. E. J. The structure and function of complex networks. *SIAM Review 45*, 2 (2003), 167–256.

[21] Nowak, M. A., and Sigmund, K. Evolution of indirect reciprocity by image scoring. *Nature 393*, 6685 (1998), 573–577.

[22] Page, and Lawrence. PageRank: Bringing order to the web. Stanford Digital Libraries Working Paper 1997-0072, Stanford University, 1997.

[23] Sandberg, O. *Searching in a Small World.* PhD thesis, Chalmers University of Technology and Göteborg University, 2005.

[24] Watts, D. J., and Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature 393*, 6684 (June 1998), 397–498.

[25] Zhou, R., and Hwang, K. Trust overlay networks for global reputation aggregation in P2P grid computing. In *Proc. IEEE International Parallel & Distributed Processing Symposium (20th IPDPS'06)* (Rhodes Island, Greece, Apr. 2006), IEEE Computer Society. USC, LA.