The value of contributions in discretionary databases

Mike Klaas
University of British Columbia
klaas@cs.ubc.ca

Abstract

Interest in the study of discretionary databases has grown as numerous large-scale public file-sharing systems have emerged as mainstream internet applications. In peer-to-peer systems, Usenet, and other discretionary databases, the study of *incentives* is of particular importance, at it allows us to model user behaviour and thus predict system performance. We examine linear models of consumption utility and determine that they fail to capture important aspects of observed behaviour. Instead, we propose a framework based on *topic classes* that can be used to generate a wide variety of utility models. Based on this framework, we quantify the *contribution value* of an item. We primarily use the Usenet file-sharing system to motivate the discussion, and show how file value can be used to solve the capacity problem in Usenet.

Introduction

Shared databases are an increasingly important means of facilitating dissemination and collection of information for groups in organizations and in the general public. A shared database is *discretionary* if users contribute to the database voluntarily. Social dilemmas arising from discretionary databases have been studied long before the advent of modern peer-to-peer (P2P) systems (Sweeney 1973; Kalman, Fulk, & Monge 2000). An infamous example of a social dilemma is the *free-rider* problem, which occurs when the database suffers from under-supply due to users not perceiving individual utility from contributing, choosing instead to free-ride. The issue has gained recent prominence as analysis of P2P systems has determined that they too suffer from an incentive problem. For instance, a recent study of the Gnutella network found that almost 70% of users are free-riders (Adar & Huberman 2000).

There is a great deal of work that looks at the incentive problem in P2P systems. Usually, this involves building a model of the incentives and analyzing it using game theory, simulation, or some other approach. Our primary objective is to develop a flexible incentive framework for discretionary databases that can be used in the modelling of specific systems. In particular, we argue that file size in file-sharing systems is an inherently non-linear term in agent utility. Another important contribution is the inclusion of the concept of *collective identification* from the discretionary database literature into our framework, which allows it to be included in formal analysis. We apply our model to Usenet and find it has significant explanatory power in that domain.

One of the goals in incentive analysis is to design new systems that do not suffer from problems identified in current systems. In doing so, the concept of *contribution value* is often an integral component. In P2P systems, for instance, it has been proposed that contributors be rewarded in some manner based on the value of their contribution to the system. We present a quantified value of a contribution in discretionary databases which exhibits many desirable properties for use in system design. We demonstrate its application to social dilemmas in P2P applications and in Usenet.

Related Work

There has been a significant amount of recent work in modelling P2P systems. (Ge *et al.* 2003) presents a broad performance model which includes incentives. (Fuqua, Ngan, & Wallach 2003) and (Mfeldman *et al.* 2003) concentrate on motivational issues. The latter gives additional insight into the disincentives that contribute to the free-rider problem, pointing out that sharing files can also hurt the capability of the user to consume files due to asymmetric bandwidth. Several papers have appeared that argue that game theory is the correct tool for incentive modelling in P2P networks (Golle *et al.* 2001; Buragohain, Agrawal, & Suri 2003). We agree with that stance, and apply it to the setting of discretionary databases in general.

Kalman et al. (Kalman, Fulk, & Monge 2000) present a treatise on motivation models in discretionary databases. They theorize that the social dilemmas inherent in such systems can be resolved by ensuring that a user identifies with other users of the system, and with the collective as a whole. This is a strong analogue to the way Usenet operates, and provides insight into why people share files on current P2P systems considering the inherent disincentives.

Problem Statement

As in previous work, we will build a game-theoretic model of the discretionary database, and use the concept of Nash equilibrium to analyze social dilemmas that occur in the database. Let $\mathcal{I} = \{a_1, a_2, \ldots, a_n\}$ be the set of agents that may participate in the system. We model the actions of these agents over a fixed time interval as a Baysian game with actions, types, and rewards as defined below. Agents contribute and consume amounts of *content* divided into individual *items*.

We are primarily targeting databases where the *size* of the content is an important factor in consumption and contribution utility, as in file-sharing systems such as peer-to-peer networks and Usenet. In this setting, items are simply files.

Usenet

What started as a means of distributing textual discussion in the 1970's has become a significant means of sharing files, with over 1 TB/day contributed to the system[1]. The file sharing side of Usenet is an interesting case study because it is a system that emerged without ever being designed for such use. Indeed, all infrastructure that currently exists to support file transfer arose in response to user activity rather than having inspired it.

Usenet file transfer is similar to P2P networks in that it is a distributed discretionary database for file transfer, but differs from them in several important ways:

- One main problem in P2P networks (freeloading) is not as important an issue on Usenet. Downloading does not put a critical load on the system, and often the heaviest users pay proportionately for their use.
- When shared, files are uploaded to the system even if they are not wanted by any consumers. In a peer-to-peer system, there is no intrinsic cost to sharing a file that is never requested. This means that providing an incentive to share desirable content is even more important on Usenet.
- Unlike in P2P networks, the performance of the system degrades as a function of total uploading volume as every contributed file is propagated to thousands of Usenet nodes.
- Usenet exhibits a high level of community spirit, despite being semi-commercial. This creates a greater incentive to contribute to the system due to a greater identification with the system and due to palpable positive or negative feedback from other users.

Before examining the system in greater detail, we will present the general motivational framework for discretionary databases.

Actions

The participating agents have *consuming* actions and *contributing* actions. Since we are modelling a single time slice of a repeated game, the amount agents choose to contribute or consume can be thought of as contribute and consumption *rates*, respectively.

Utility

An agent's utility is composed of several factors divided into two types: utility of consumption actions (ν_i) and utility of contributory actions (μ_i). Total utility is the sum of these two factors $u_i = \mu_i + \nu_i$. We will assume risk-neutral agents throughout (linear relationship between utility and money).

[1]As cited March 19, 2004 on `alt.binaries.news-server-comparison`. Current statistics can be obtained at `http://newsfeed.visi.com/`

Consumption utility Utility of consuming can be broken down as follows:

- *Content Retrieved* The amount of data successfully retrieved from the system. Agents may only be interested in a set amount of content, after which they are made no happier by downloading more.
- *Variety* The greater selection of content to choose from, the happier the agent.
- *Bandwidth* Represents the cost of consuming data as a combined function of bandwidth and the agent's time.
- *Explicit Cost* There may be a explicit cost in terms of "points" or dollars.

One of the principle motivations of this paper is to develop a model for consumption utility in a discretionary databases which better reflects reality. The general setting is as follows: Let $\mathcal{Q}$ be the set of content available for consumption. Then the valuation function $\nu_i(C)$ gives the utility of agent $a_i \in \mathcal{I}$ for consuming some $C \subseteq \mathcal{Q}$.

The simplest model is linear in the quantity consumed ($\nu_i(C) \propto size(C)$). This has the advantage of not having to account for the cost of consuming (since it is assumed to be rolled into the proportionality constant). One problem is that we are assuming that the agent values all subsets C of the same size equally. In actuality an agent will value different contributions to the system at different levels depending on their interests. This approach, augmented with a *benefit matrix* which allows differential valuation, is taken in (Buragohain, Agrawal, & Suri 2003).

We define the matrix $\mathbf{B} = \{b_{ij}\}$ to be an n by n matrix with entry $b_{ij} \geq 0$ indicating a_i's interest in the content a_j provides. The expression for a_i's utility becomes $\nu_i(C) = \sum_{j \neq i} b_{ij} c_j$, where c_j is a_j's contribution in C. We can drop the condition that $j \neq i$ by noting that $\forall i$, $b_{ii} = 0$.

This model is much more useful, but still suffers from problems. The main criticism is that the proportionality of utility and size of content does not match our intuition for the way discretionary databases are used. Firstly, it violates *variety*; agents should be happier when they have a wider choice of content. Secondly, linear models put too much emphasis on size. If, for instance, a file-sharing system tried to optimize social welfare (the sum of agent utilities across the system), then a user downloading a cd-image ($\approx 2^{29}$ bytes) will be given the same weight as over a thousand users downloading a jpeg picture ($\approx 2^{18}$ bytes). This emphasis may not be desired.

We propose the following model that attempts to consider both of these factors. The notion of variety can be thought of analogously to the notion of *substitutability* in auction theory, which occurs when the total utility for obtaining two goods is less than the sum of the utilities of obtaining the goods individually. In a document database, the utility of consuming an item will be less if the agent has access to very similar documents already. The degree to which two items are substitutable will vary enormously in the database. In fact, it is unlikely that two items are completely independent or perfectly correlated. We will assume that there are classes of items whose substitutability is given by a function and which are otherwise additive. Formally, let $\mathcal{Q} = \bigcup_i^k Q_i$, such that the $\{Q_i\}$ form a partition of $\mathcal{Q}$ (ie., they are pairwise disjoint). Then $C \subseteq \mathcal{Q}$ is defined as $C = \bigcup_i^k C_i$ with $C_i \subseteq Q_i$, and agent a_i's utility for consuming C is:

$$\nu_i(C) = -\phi_i(C) + \sum_{j=1}^{k} w_{ij}\theta_i(size(C_j)), \qquad (1)$$

where ϕ_i is a function giving the cost of consuming C, and w_{ij} are a_i's value for the different classes. To model substitutability, we require that the marginal utility gain in a class goes to 0 as the total consumption in the class increases. Although not necessary in the general setting, we will make additional continuity and differentiability assumptions in our analysis.

Assumption 1. *For any $a_i \in \mathcal{I}$, the* ***class utility*** *function $\theta_i : [0, \infty) \times [0, \infty)$ satisfies:*

1. $\theta_i \in \mathcal{C}^2$
2. $\theta_i(0) = 0$
3. $\frac{\partial \theta_i(x)}{\partial x} > 0$ *(θ_i is non-decreasing)*
4. $\frac{\partial^2 \theta_i(x)}{\partial x^2} < 0$ *(marginal utility gain goes to 0)*

Two examples of functions that satisfy these properties are $\theta_i(x) = \sqrt{x}$ and $\theta_i(x) = \log(1 + x)$.

We are not the first to use diminishing returns in valuations. *Downward-sloping* valuations are an important and well-studied class of valuations in combinatorial auction theory. In addition, information-theoretic arguments for diminishing-return valuations for general resources have been presented in the network bandwidth allocation literature (Lazar & Semret 1998).

Contribution utility The classic social dilemma in discretionary databases is the free-rider problem and has been observed to occur on several P2P networks. Usenet is a theoretically similar but somewhat more dire situation. Unlike some P2P systems, sharing is not default behaviour of the application, and requires tangible cost to perform. Also, whereas most P2P systems are non-commercial (which might inspire some level of altruistic "donations"), access to binary Usenet is dominated by commercial servers. Worst still, on these commercial servers, users must pay to consume and are given no credit for contributing!

This makes Usenet an intriguing case to analyze, as it does not appear to suffer from a contribution problem despite the disadvantages aforementioned. We believe the principal reason lies in Fulk's insight that users may *identify* with the collective and consequently derive some personal satisfaction in maximizing the *common good*. This effect manifests on Usenet at a much greater level than on P2P systems for several reasons. First, Usenet was originally a discussion medium, and files are intermixed with discussion in a newsgroup. This, combined with the very division of the system into newsgroups, contributes to foster a sense of community within a newsgroup which in turn contributes to the identification of a user. Additionally, users can and do provide explicit positive and negative feedback to contributors within the group.

Here is a summary of factors that influence the utility of contributing to a discretionary database:

- *Selfish/Altruistic Motive* The personal satisfaction of agent for contributing to the system, or the agent's satisfaction gained for contributing as much personal content as possible to the system[2].
- *Bandwidth* Cost to contribute, in time and bandwidth.
- *Explicit Reward* In a micro-economic system, contributions may be explicitly rewarded.
- *Reputation* The utility gained or lost by the positive or negative feedback of others.

Selfishness and altruism are equivalent from the perspective of our model, and we will combine them with the intrinsic cost into a factor $\phi_i(c_i)$. $\phi_i(c_i)$ is usually positive, meaning that the cost to contribute outweighs the selfishness/altruism factor. We will assume that $\phi_i(0) = 0$.

Thus *feedback* is a crucial concept to capture when modelling Usenet. We we model this by considering feedback to be the total consumption utility derived by all other players from the content provided by this agent. We will assume feedback is given with some fixed probability, which gives (with the constant omitted):

$$\mathbb{E}(\mu_i) = -\phi_i(c_i) + \sum_{j \neq i}^{n} \tilde{\nu}_{ji} \qquad (2)$$

where $\tilde{\nu}_{ji}$ is a_i's contribution to a_j's utility.

Application to Usenet

We will now apply this framework to develop an incentive model for Usenet file sharing. Before analyzing the game, we have to specify parameters left undefined in the previous section.

The cost to contribute will be assumed to be linear and governed by a factor γ_i unique to each agent. That is, $\phi_i(x) = \gamma_i x$. We will also only consider $\phi_i(x) > 0$. We will also assume that all agents share a common class utility function $\theta_i = \theta_j = \theta$. A crucial decision is to decide how to partition content into classes to apply equation 1. We will assume that each contributor contributes a set of partially-substitutable content, and that there is no substitutability among contributors. This simply involves assigning each agent's contribute to its own partition. This assumption also allows us to return to using $\mathbf{B}$ to represent an agent's desire for different content classes.

By rolling the cost function into the sum, we can re-write equation 1 as follows to obtain a_i's utility for consuming:

$$\tilde{\nu}_{ij} = b_{ij}\theta(c_j) - \gamma_i c_j \qquad (3)$$

$$\nu_i = \sum_{j}^{n} \tilde{\nu}_{ij} \qquad (4)$$

Note that we obtain the desirable property that $\tilde{\nu}_{ij}$ can decrease (and even attain negative values). Once the marginal

[2] Selfishness can be manifest in several forms. One user we interviewed confided that her principal motivation for sharing content on KaZaA was to "show off" her personal music collection.

utility gain is less than γ_i, the additional content present lowers the performance of the system and makes it less likely for a_i to find content that interests him.

Agent a_i's utility for contributing c_i is given by

$$\mu_i = -\gamma_i c_i + \sum_{j \neq i}^{n} \tilde{\nu}_{ji} \tag{5}$$

Nash Equilibrium

A *Nash Equilibrium* of a game is a set of agent strategies in which any single agent cannot gain by unilateral deviation (that is, considering every other agent's strategy as fixed, every agent is playing their *best response* strategy). While computing Nash equilibria is normally a difficult problem, we are aided by the structure of the setting. Due to the feedback factor, any time an agent contributes to another agent's utility (whether positively or negatively), the contribution is manifested in their own utility in the agent's μ_i term. Also, the gain in ν_i from other agents' actions is not affected by a_i's action. Thus to find a Nash equilibrium we simply maximize μ_i.

By combining equations 3 and 5, differentiating, and setting to zero we obtain:

$$\frac{\partial \theta}{\partial c_i} = \frac{\sum_k^n \gamma_k}{\sum_{j \neq i}^n b_{ij}} \tag{6}$$

Whether or not this has a closed-form analytic solution depends on the exact form of θ. If $\theta(x) = \sqrt{x}$, then

$$c_i = \left(\frac{\sum_{j \neq i}^n b_{ij}}{2 \sum_k^n \gamma_k} \right)^2 \tag{7}$$

Rogue Agents

Not all agents participating in usenet are reputation-motivated. This can be a characteristic of the agent, or a characteristic of the group, as the level of community spirit varies considerably in usenet groups. Agents that are not motivated by reputation do not have the feedback factor in their contribution utility μ_i, which means that their decision to contribute depends only on $\phi_i(c_i)$. They will behave to maximize $-\phi_i(c_i)$. If this occurs when $c_i = 0$, the agent will not contribute.

A mix of reputation-motivated and rogue agents seems to be a good model for a Usenet group. The appropriate mix depends heavily on the group. For example `alt.binaries.sounds.mp3` is one of the most community-oriented binary newsgroups on the network, to the point of having an enforced charter and voluntary posting limits. It is likely the group can be closely-modelled by exclusively reputation-motivated agents.

Contribution Valuation

In the previous sections we have developed a general model for utility in discretionary databases and applied it to Usenet. In this section we will discuss how these ideas could be used as components when designing new systems.

It is immediately apparent that our utility model as it stands is not ideal for incorporation into a monetary payment scheme (for instance). This is because it involves several quantities and concepts that are difficult to determine. The interest weights w_{ij} are impossible to measure directly, and the partition of $\mathcal{Q}$ into non-substitutable classes $\{Q_i\}$ seems difficult to get a handle on, considering the partition could be in theory be completely different for each agent.

Value of an Item

The expected value to the system of an item contributed has many potential uses. It can be used to reward the contributor (whether monetarily, or through improved service), to determine contributor statistics, or to tune the performance of the system by guiding resource allocation.

There are three key values relating to a item f that will be useful. They are the item's size f_{size}, the contribution count f_{up}, and the consumption count f_{down}. The precise nature of these values will depend on the system in question. In the case of a temporal database like Usenet, f_{down} and f_{up} are the raw counts in some rolling history, while in a P2P system the might represent integrals of downloads and sharing over time, respectively. We count all contributions of the same item together since we are certain that they are perfectly substitutable.

Intuitively, an item's value should increase in f_{down} and decrease in f_{up}. Also, the item must be in a single topic class for any given agent, so the utility from the size can be no greater than $\theta(f_{size})$. This gives us a good guess at the value function:

$$v_{first}(f) = \frac{f_{down}}{f_{up}} \cdot \theta(f_{size}) \tag{8}$$

Our intuition turns out to be a good approximation, as the following proposition shows.

Proposition 2. *Let $f \in \mathcal{Q}$ be an individual contribution to the system. Then the expected value $v(f)$ when f is considered part of a topic class is bounded by $v_{first}(f)$, and is within a λ-factor of $v_{first}(f)$, where λ is the expected number of items in a topic class.*

Proof. The expected value $v(f) = \frac{f_{down}}{f_{up}} \frac{\theta(\lambda f_{size})}{\lambda}$. Hence,

$$\begin{aligned} \frac{v(f)}{v_{first}(f)} &= \frac{f_{down}}{f_{up}} \frac{\theta(\lambda f_{size})}{\lambda} \cdot \frac{f_{up}}{f_{down}} \frac{1}{\theta(f_{size})} \\ &= \frac{\theta(\lambda f_{size})}{\theta(f_{size})} \cdot \frac{1}{\lambda} \\ &\geq \frac{1}{\lambda} \end{aligned}$$

since $\theta(\lambda f_{size}) > \theta(f_{size})$ by assumption 1. □

For many choices of θ the situation is more rosy. In the case where $\theta(x) = \sqrt{x}$, $v(f)$ is within a $\sqrt{\lambda}$-factor of $v_{first}(f)$.

Use in P2P incentivizing

The use of the integral of content shared over time as a measure of contribution has been proposed by several authors in the literature (Golle *et al.* 2001; Buragohain, Agrawal, & Suri 2003). Golle notes that this provides an incentive for agents to share unpopular files and to do so during times of low demand, and suggest scaling factors to remedy these defects. The use of our measure of contribution factor provides a specific implementation of these factors. Let the reward for sharing an file be:

$$\int v_{first}(f,t)dt = \theta(f_{size})\frac{\int f_{down}(t)dt}{\int f_{up}(t)dt} \quad (9)$$

where $f_{up}(t)$ is the number of users from which f is available, and $f_{down}(t)$ is the number of users downloading f at time t. $f_{up}(t) \geq 1$, so the reward is well-defined.

There are many advantages to this reward function. Users gain more by sharing popular files, but only as long as the supply isn't too great. Rare files will be downloaded less frequently, but the supply will also be smaller, thus providing no disincentive to share rare files. Finally, the number of accesses will be low during off-peak times, thus there is no incentive to only connect during those periods.

Competition for system resources

Earlier, we showed how positive and negative feedback from other users can influence the contribution rate for reputation-motivated agents. This is a form of collective regulation of individual action. Unfortunately, the way Usenet is structured, this cannot solve the resource allocation problem which we define below. We will show how this problem might be alleviated using the concept contribution valuation as defined in the previous section.

Let us return to the user-benefit matrix $\mathbf{B}$. Some insight can be gained by examining its structure. First, in a general file-sharing system, most of the entries will be 0 because of the wide variety of content available. Second, $\mathbf{B}$ is likely to appear symmetric, in the sense that if $b_{ij} = 0$, there is a high probability that $b_{ji} = 0$, and likewise for non-zero entries. This is due to the idea that if there is interest between two agents in one direction, it is likely that their content is topically similar, so there will be a mutual interest expressed. For the same reason, $\mathbf{B}$ will appear to be transitive ($b_{ij} > 0 \wedge b_{jk} > 0 \rightarrow b_{ik} > 0$ with high probability).

These structural properties mean that we can permute the indices of agents to obtain a block-diagonal form (with the caveat that $b_{ii} = 0$):

$$\mathbf{B} = \begin{pmatrix} \blacksquare & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \blacksquare & \cdots & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \blacksquare \end{pmatrix}$$

where boldface **0**'s denote block of zeros. This corresponds to a partitioning of the agents into groups of mutual topical interest. This effect is most apparent on Usenet, where the partitions correspond to newsgroups (or sets thereof).

This is a problem in Usenet because agents are only motivated by feedback from other agents in the same group. However, the system resources are shared by all groups. Since there is little to no feedback or collective identification between an agent and a group to which he doesn't belong, there is no incentive for groups to act non-selfishly in acquiring system resources.

Modelling the competition

We will model the bandwidth competition as a Baysian game, and assume that groups of agents can through feedback achieve a desired collective action. Thus we consider *entire groups* as individual players in this game.

Let κ be the amount of the a system performance resource, such as the total bandwidth per time slice that can be propagated without loss. A player's utility is proportional to the expected amount downloaded[3]. Let $c_i \in [0, k_i]$ be the amount uploaded per time slice for player i, and denote by s_{-i} the strategy of all other players . If $\sum_i c_i \leq \kappa$, then an agents utility is $u_i \propto c_i$. Otherwise, the contribution exceeds the system's capacity. Let $f_i(c_1, c_2, \ldots, c_n, \kappa)$ be a function that calculates the expected fraction of player i's content that isn't dropped. Then the expected utility for player i is:

$$\mathbb{E}(u_i) \propto \begin{cases} c_i & \text{if } \sum_j c_j \leq \kappa \\ c_i \cdot f_i(c_1, c_2, \ldots, c_n, \kappa) & \text{otherwise} \end{cases} \quad (10)$$

In Usenet, the most common way to choose f_i is to drop articles with uniform probability, which means that

$$\forall i,\ f_i = f(c_1, c_2, \ldots, c_n, \kappa) = \frac{\kappa}{\sum_j c_j}$$

Figure 1 shows what single player's utility for different values of contributions from all players.

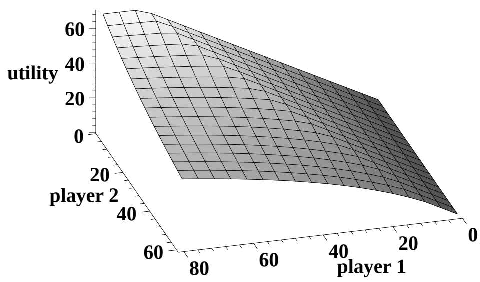

Figure 1: Player 1's utility surface in the 2-player bandwidth competition game for $\kappa = 70$

Proposition 3. *In the n-player bandwidth competition game, $S = \{k_1, k_2, \ldots, k_n\}$ is the unique Bayes-Nash equilibrium.*

Proof. Consider player i keeping c_{-i} fixed. For $c_i \leq k_i$, we have

$$\mathbb{E}(u_i) \propto \frac{c_i \cdot \kappa}{c_i + \sum_{j \neq i}(c_j)} \propto \frac{c_i}{c_i + \sum_{j \neq i}(c_j)}$$

[3] This holds in our model as long as the number of classes is proportional to c_i

But the partial derivative wrt c_i is

$$\frac{\partial \mathbb{E}(u_i)}{\partial c_i} \propto \frac{\sum_{j \neq i} c_j}{\left(c_i + \sum_{j \neq i} c_j\right)^2} > 0$$

so $\mathbb{E}(u_i)$ increases monotonically in c_i and has a global maximum at $c_i = k_i$ (since $c_i \in [0, k_i]$). Since this is a dominant strategy equilibrium, it is unique. □

This is a problem for servers as users in a group have incentive to post amounts of content beyond what they value to gain a larger slice of the global resource pie. This increases load unnecessarily and causes other performance problems.

Incentivizing reasonable behaviour

In this section we will outline a scheme to allocate Usenet resources in such a way that doesn't create an unbounded incentive to upload content. Consider an auction where c_i is the amount of good divisible κ that player i desires. We will use a Progressive Second Price (PSP) auction as mechanism which are proven to be efficient and incentive compatible (Kounavis 1997; Lazar & Semret 1998). Instead of eliciting a numeric bid value, we will take the marginal collective valuation of the content uploaded as the bid value in some unit that we will call "points". If group i posts c_i of content in files F_i, then its bid is:

$$s_i = \langle c_i, b_i \rangle = \left\langle \sum_{f \in F_i} f_{size}, \frac{1}{c_i} \sum_{f \in F_i} \frac{\theta(f_{size}) f_{down}}{f_{up}} \right\rangle$$

Groups with the highest collective valuation will be allocated resources from the top down:

$$a_i(s) = \min \left\{ c_i, \left[\kappa - \sum_{k | b_k > b_i} a_k(s) \right]^+ \right\}$$

Each group will have to pay a cost in points that is equal to the "social opportunity cost" of the other groups:

$$p_i(s) = \sum_{j \neq i}^{n} \left(b_j \frac{a_j(s_{-i}) - a_j(s)}{\sum_{k \neq i}^{n} a_k(s_{-i}) - a_k(s)} \right)$$

These points will be used to determine how much retention the group's content receives, which is the length of time the files are kept on the server before being expired. Groups with a higher marginal valuation will be willing to pay a higher cost in retention.

Note that this scheme is vulnerable to manipulation, as a user could post a large quantity of useless data in a competing group, thus lowering their marginal valuation and making their own group's content more likely to have higher valuation. This can be overcome by requiring a certain number of accesses before a file is counted in a group's bid.

Conclusion

We have proposed a motivational framework for discretionary databases that has greater expressive power than linear models and better corresponds to observed practice. We have used it to quantify the utility of collective identification and feedback and applied the model to Usenet and determine that it has good explanatory power in that setting. Additionally, we determined a quantification for the value of a contribution in a discretionary database, and showed how it can be used to solve social dilemmas in peer-to-peer networks and Usenet file transfer.

References

Adar, E., and Huberman, B. 2000. Free riding on gnutella. *First Monday* 5(10).

Buragohain, C.; Agrawal, D.; and Suri, S. 2003. A game theoretic framework for incentives in P2P systems. In *Proc. of the Third IEEE International Conference on Peer-to-Peer Computing*.

Fuqua, A. C.; Ngan, T.-W. J.; and Wallach, D. S. 2003. Economic behavior of peer-to-peer storage networks. In *Workshop on Economics of Peer-to-Peer Systems*.

Ge, Z.; Figueiredo, D. R.; Jaiswal, S.; Kurose, J.; and Towsley, D. 2003. Modeling peer-to-peer file sharing systems. In *Proc. of Infocom*.

Golle, P.; Leyton-Brown, K.; Mironov, I.; and Lillibridge, M. 2001. Incentives for sharing in peer-to-peer networks. In *Proc. of The Second workshop on Electronic Commerce (WELCOM'01)*.

Kalman, M. E.; Fulk, J.; and Monge, P. 2000. A motivational model for resolving social dilemmas in discretionary databases.

Kounavis, M. E. 1997. Auctions for resource sharing.

Lazar, A. A., and Semret, N. 1998. Design, analysis and simulation of the progressive second price auction for network bandwidth sharing. *ISDG* 8th.

Mfeldman, M. F.; Lai, K.; Chuang, J.; and Stoica, I. 2003. Quantifying disincentives in peer-to-peer networks.

Sweeney, J. 1973. An experimental investigation of the free-rider problem. *Social Science Research* 2.