

**A Modified Depensation Model of
Peer to Peer Networks:
Systemic Catastrophes and other Potential
Weaknesses**

Andrew H. Chen (achen@u.washington.edu)
Andrew M. Schroeder (andypants@hotmail.com)

December 17, 2001
(Revised June 13, 2002)
AMATH 383
Term Paper

Professor K. K. Tung

Abstract

Peer to peer networks have generated significant attention in the recent past, especially file trading networks such as Gnutella and Morpheus. These networks allow users to easily access and exchange a wide array of digital information, potentially including pirated music and other digital media. Also, these networks are decentralized, so there is no central entity that can be held responsible for illegal actions on the network, leading some to believe that they can never be shut down.

However, in this paper we argue that these networks are potentially more vulnerable than previously believed by utilizing a modified version of the “depensation model,” commonly found in ecological models of fish and birds. If the model is accurate, one interesting implication is that P2P networks may be susceptible to “catastrophes,” situations in which the P2P network’s user base has negative natural growth. This implies that it is not necessary to take action against all the users on the network to shut it down, but rather just a subset.

In particular, our analysis of the model leads to four potential strategies, which can be used in conjunction:

1. Randomly selecting and litigating against users engaging in piracy
2. Creating fake users that carry incorrectly named or damaged files¹
3. Broadcasting fake queries in order to degrade network performance
4. Selectively targeting litigation against the small percentage of users that carry the majority of the files

Just as catastrophes can be caused in animal populations by overharvesting, it may be easier for the recording industry to solve the P2P problem than previously thought, by purposely pushing the network over this tipping point.

Problem Description

The promise of the Internet lies in its ability to eventually make information available to everyone, anytime, anywhere. Eventually, any type of media, whether audio, video, literature, or images, can be accessed, shared, and distributed to the whole world in a blink of an eye. In many respects this promise has been fulfilled in the United States, but for many companies, this comes with strings attached.

The ability to distribute information so effectively poses problems for companies that rely on intellectual property, such as creators of software, movies, music, books, etc. As the internet supercedes traditional methods of distribution, there is an increasing danger that profits will fall as people find the companies' products for free. For example, the music industry estimates that it loses \$4.1 billionⁱⁱ in revenue from music piracy.

This problem has gained tremendous attention in the last year, as the Recording Industry Association of America (RIAA) successfully challenged, and eventually shutdown Napster, the most popular music trading service at that time. Since then, Napster has restarted their services, but the ability to trade pirated music was severely handicapped, which has resulted in a 90% decline in their user baseⁱⁱⁱ.

However, these users have not simply stopped trading files; they have just moved onto more advanced file-trading platforms based on peer to peer technology, such as Gnutella and Morpheus. Many have stated that these P2P networks cannot be stopped by the RIAA, but this issue has not been thoroughly studied with a mathematical model.

To do so, a model must be created which not only describes P2P network also the consequences of RIAA interaction with its users. To this end, this paper describes a conceptual model of P2P networks, based on the depensation model found in studies of population dynamics. This may be an accurate model because there are many similarities between P2P networks and certain animal populations such as bird flocks or fish schools. For these animals, there are

advantages to grouping together, such as for protection and confusing predators, and this grouping creates a “critical mass” which must be satisfied for the animal populations to have positive growth. Similarly, there is a critical mass for users in P2P networks, because if there are too few connected to the network, files will be difficult to find. Thus, if the number of users is below this minimum number, the user base will not grow, and the service will eventually collapse. Another similarity is that there is an overall limit to the size of the animal population, imposed by the resources of the environment, such as water, space, etc. Similarly, in P2P networks, there is a maximum number of users limited by bandwidth^{iv}.

By modifying the depensation model found in animal populations, it may be possible to accurately describe P2P networks, and most importantly, the results of “harvesting” the population. This paper aims to address the following questions:

1. How must the depensation model be modified in order to account for conceptual differences between P2P networks and animal populations?
2. What are the conditions necessary to cause catastrophes in P2P networks?
3. What does the model imply about other ways to limit or stop P2P networks?
4. What is the most effective method to stop P2P networks?

Simplifications

To obtain an appropriate model for P2P networks, it is necessary to simplify the problem by making a few logical assumptions.

Infrastructure Assumptions

In particular, because P2P networks can be quite large and are built on a complex and diverse infrastructure, it is not viable to create a model which captures the all of the interaction of agents across the entire Internet. Thus, we make the following assumptions:

1. The P2P network is balanced and homogeneous (equal edges, bandwidth, etc.).
2. Messages sent across the network have common characteristics (size, TTL, etc.), and are sent at similar rates.

User Assumptions

Additionally, it is necessary to make a set of assumptions on the behavior of the users on the network, based on intuitive reasoning. Thus, we assume:

1. The growth of new users depends on the number of users on the network.
2. Not everyone both shares as well as downloads files; there are a significant number of “freeloaders” that only download without sharing.^v
3. Users will stop using the service if their searches for files come up negative.
4. As the network performance starts to degrade, the usage of the service will begin to decline.

However, beyond these behavioral assumptions, it is necessary to make assumptions about the potential actions of the RIAA or other organizations aiming to stop P2P file-trading. This paper will not discuss the specifics of how such an organization would actually go about removing users, other than suggesting what has been done before, litigation. Thus, we assume

1. There is a way to stop users from using the service.
2. If there is an effort to stop individual users, then growth of new users will decline.

The last assumption is based on the idea that whatever method that is used to stop users on the P2P network, it will intimidate or otherwise discourage new users from joining. Even if the effort is more benign, such as introducing a service that competes with the functions of P2P networks, it would most likely cause adoption rates to be lowered.

With these simplifications, it is now possible to create a model to describe P2P networks and the results of removing users from the service.

Mathematical Model

Unmodified Depensation Model

Peer to peer networks, as discussed earlier, have much in common with animal population. Because of this, the depensation model will be modified to describe P2P networks.

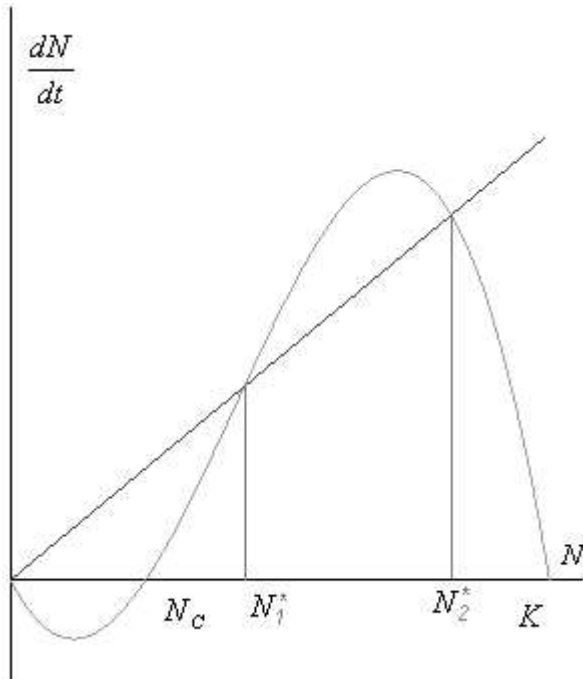
However, to begin with, an unmodified version of the model will be discussed:

$$\frac{dN}{dt} = -rN\left(\frac{N}{N_c} - 1\right)\left(1 - \frac{N}{K}\right) - qEN^{\text{vi}}$$

Table 1:

Symbol	Definition
r	Growth rate
N	Population
N_c	Critical mass for population
K	Carrying capacity for environment
q	Catchability coefficient
E	Harvesting effort

Figure 1:



In the graph shown in Figure 1, the two non-zero equilibrium points are marked. N_1 is unstable, and N_2 is stable.

Modified Depensation Model

From assumption 6, however, there is a modification that needs to be made to this model. The growth rate needs to decrease when effort is increased, as thus, we divide r by $\alpha E + 1$, where α is a parameter indicating how removal effort affects user growth rate. A high number would indicate that removal effort is highly effective in discouraging new users from entering the network, and a low one would indicate a low discouragement.

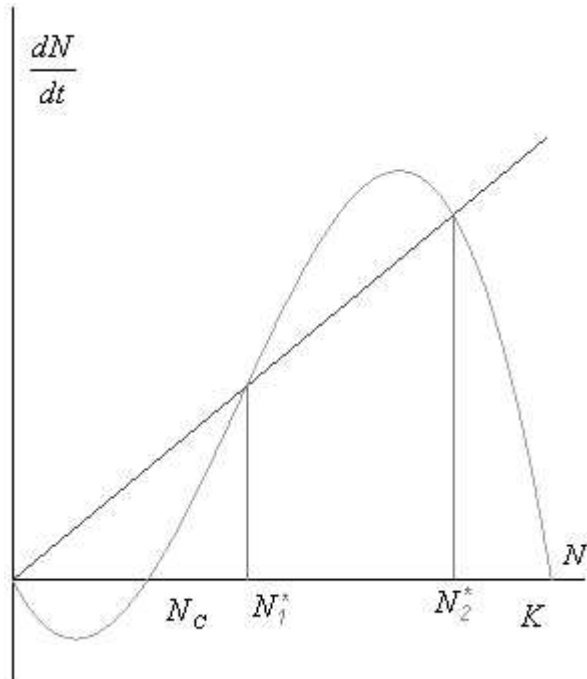
The new equation, followed by descriptions specifically for P2P networks follow:

$$\frac{dN}{dt} = \left(\frac{-r}{\alpha E + 1}\right)N\left(\frac{N}{N_c} - 1\right)\left(1 - \frac{N}{K}\right) - qEN$$

Table 2:

Symbol	Definition
r	Growth rate
α	Relation of removal effort to new growth rate
N	Number of users
N_c	Critical mass of users
K	Maximum number of users supported by the network (carrying capacity)
q	Removal coefficient
E	Effort to remove users from network

Figure 2:



As shown in Figure 2, modifying the model does not change the overall shape, and it looks the same as the unmodified model. One should see the differences in the two models only when the values for effort are factored in, which changes the height of the curve.

Many of these constants must simply be observed empirically, for obvious reasons. It is difficult to say what would happen to the growth rate if users began to be removed from the network, for example. However, some variables can lend themselves to a fair amount of speculation. For example although it would vary for different methods of user removal, the removal coefficient q is probably fairly small, because the most potent method, litigation, is slow-acting and generally targets small numbers of users. At the very best, it may be possible to achieve a high removal coefficient by pursuing larger groups of users through their ISPs, schools, or companies. It is through this type of intuitive reasoning that information can be gathered on these variables without resorting to extensive research.

Critical Mass and Carrying Capacity

Two important numbers, carrying capacity and critical mass, can be decomposed further. Both depend on the idea of the probability of a search hit, which will be referred to as p . This variable p depends on several factors, such as the number of files each user shares, the diversity of these files, and the tastes of users on this network. Although these parameters seem difficult to quantify, especially the last one, one may be able to make a reasonable estimate by sitting on the P2P network and watching searches and responses go by, to see how often users find the files they are looking for.

Before discussing these different costs, here are the variables that will be used:

Table 3:

Symbol	Definition
R	Maximum number of reachable users within T hops, assuming D edges per node
D	Number of edges per node
T	Time-to-live, or number of hops message will traverse
p	Probability of finding a specific file on the network
m	Probability of finding a specific file on the specific node
S_q	Size of query packet
S_r	Size of response packet
Q	Average number of queries/responses per second
K	Average number of responses per query
F	Average size of files transferred

First, one can note that the critical mass is tied closely with the parameter p , because as the number of users increases, the probability of a successful search increases as well. A thorough treatment of this relation is difficult, because the critical mass is simply the number of users necessary to push p over some percentage, completely determined by the psychology of the users. This adds in a lot of complexity because it means the critical mass may shift depending on external factors, such as the availability of substitutes trading files over this particular network.

Users may be able to move to other networks or distribution methods, which would affect the critical mass.

Carrying capacity relates to more concrete factors, fortunately. Because of the assumption that the P2P network is fully balanced, the number of users within T hops of a specific node, assuming each edge has D edges, is given by:

$$R = D \sum_{i=1}^T (D-1)^{(i-1)}$$

From this, the cost of querying and passing on queries is given below, where Q is the number of queries from that node per second, m is the probability the specific node has the file, S_q is the size of the query packet, and R is the number of nodes the query is sent out to:

$$C_q = QS_q R + QS_q R(D-1) + DQS_q = QS_q RD(R+1)$$

The first term in the equation quantifies the fact that everyone within T hops of the node will be issuing queries to the node. The next term details how much capacity is used to pass the response. The final term captures the capacity needed in the node's own searches.

The cost of responding and passing on responses to queries is given below, where p is the probability of users finding the files they want, S is the size of the response packet, and R and Q are the same as before:

$$C_r = RQS_r k \frac{T}{R}$$

$$C_r = QS_r kT$$

where T/R is the probability that of a response being passed upstream through an individual node, from a downstream neighbor.

Finally, after the response, the node connects directly with the other node, and there is a cost for the download and upload. It is given below, where C_d and C_u are the costs of downloading and uploading files, respectively. F is the average size of files transferred across the

network, and m (where $m \gg p$) is the probability that the node has the file the other nodes are looking for.

$$C_{du} = C_d + C_u = pRQF + mQF$$

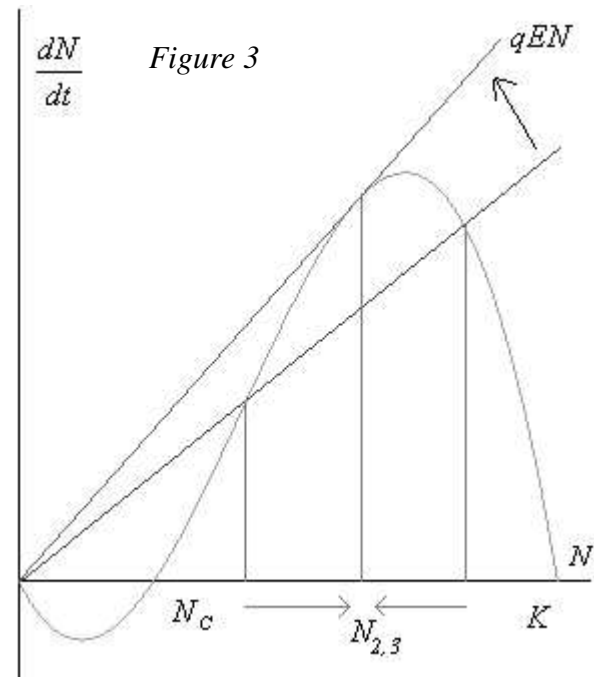
Combining each one of these costs, we can then find the total bandwidth, which is simply each of the costs added together. This is given by:

$$C_T = C_q + C_r + C_{du} = QS_qRD(R+1) + QS_rkT + pRQF + mQF$$

The carrying capacity of the network can now be defined. Because the users on the network are limited by the bound of the bandwidth capacity of each node, when C_T is equal to the bandwidth capacity, the carrying capacity is reached. The carrying capacity may increase or decrease depending on several factors, including how often the users are searching, or how quickly they are able to find the files they want. It is also obvious that decreasing the carrying capacity decreases the total number of trades that can be happening across the network, which limits the potential for piracy.

Catastrophes

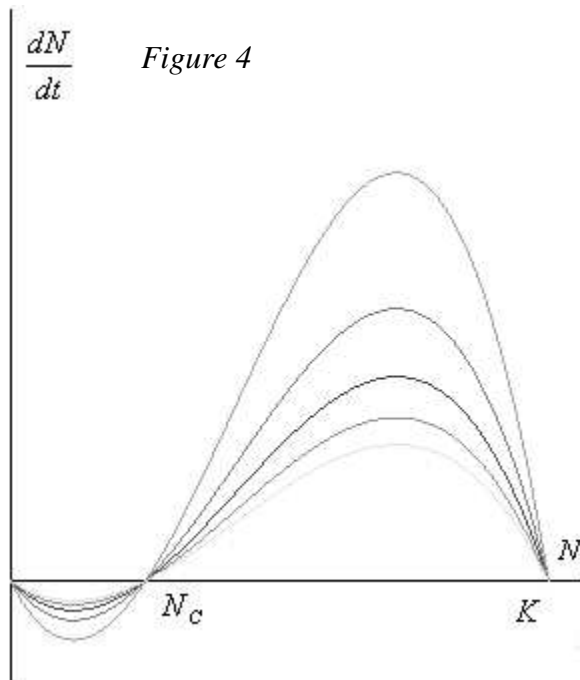
Also, it is important to note that with both models, there is the potential for “catastrophes,” the situation in which harvesting leads to a collapse in the population or user base. For this to happen, the effort must be increased so that the two equilibrium points on the curve (indicated by the intersection of the curve and the effort line) merge into one. This is shown in Figure 3, which shows the unmodified



depensation model.

As the harvesting effort is increased, the two equilibrium points eventually converge to one point, at which there is a catastrophe. At this point, the natural growth rate of the population is negative, and the system collapses. By solving for the equilibrium points, it is then possible to find the conditions necessary to cause this collapse.

Because of the modification to the model, as the effort is increased, the curve itself flattens out. This is reflected in the Figure 4.



Thus, as the effort increases, and the curve flattens, it is easier to achieve a catastrophe than in the unmodified animal population model.

Solution of the Mathematical Problem

Next we must find the equilibrium points. This is done by equating the differential equation to zero and solving for N:

$$\frac{dN}{dt} = \left(\frac{-r}{\alpha E + 1}\right)N\left(1 - \frac{N}{N_c}\right)\left(1 - \frac{N}{K}\right) - qEN$$

$$0 = \left(\frac{-r}{\alpha E + 1}\right)N^*\left(1 - \frac{N^*}{N_c}\right)\left(1 - \frac{N^*}{K}\right) - qEN^*$$

$$N_1^* = 0$$

$$0 = \left(\frac{-r}{\alpha E + 1}\right)\left(1 - \frac{N^*}{N_c}\right)\left(1 - \frac{N^*}{K}\right) - qE$$

$$N^{*2} - (N_c + K)N^* + \frac{(N_c K)(qE(\alpha E + 1) + r)}{r} = 0$$

$$N_{2,3}^* = \frac{(N_c + K) \pm \sqrt{(N_c + K)^2 - 4 \frac{(N_c K)(qE(\alpha E + 1) + r)}{r}}}{2}$$

The collapse occurs when $N_2 = N_3$, and we try to solve for E to find the appropriate amount of effort to cause the catastrophe.

$$(N_c + K)^2 = 4 \frac{(N_c K)(qE(\alpha E + 1) + r)}{r}$$

$$\alpha E^2 + E + \frac{r}{q} \left(1 - \frac{(N_c + K)^2}{4(N_c K)}\right) = 0$$

$$E = \frac{-1 \pm \sqrt{1^2 - 4\alpha \frac{r}{q} \left(1 - \frac{(N_c + K)^2}{4(N_c K)}\right)}}{2\alpha}$$

$$E = \frac{-1 \pm \sqrt{1 + 4\alpha \left(\frac{r}{q}\right) \left(\frac{N_c^2 + K^2}{4(N_c K)} - \frac{1}{2}\right)}}{2\alpha}$$

$$E = \frac{\sqrt{1 + 4\alpha \left(\frac{r}{q}\right) \left(\frac{N_c^2 + K^2}{4(N_c K)} - \frac{1}{2}\right)} - 1}{2\alpha}$$

One of the solutions to E is going to be negative, and because effort cannot be negative, it is discarded.

Results and Conclusions

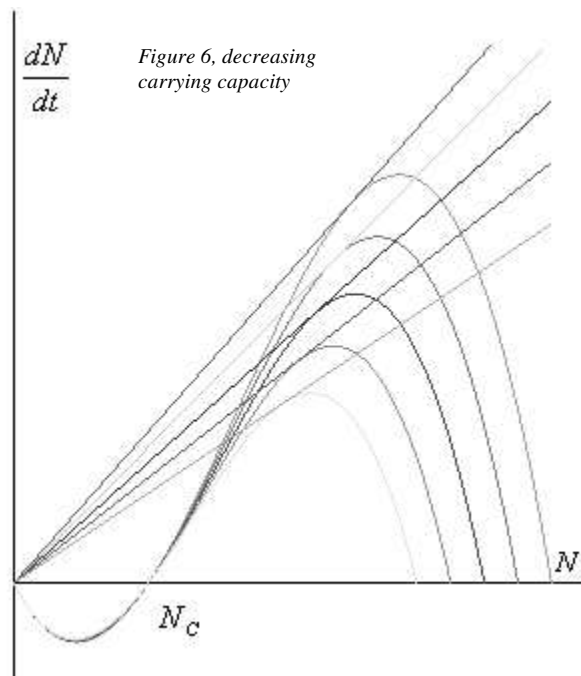
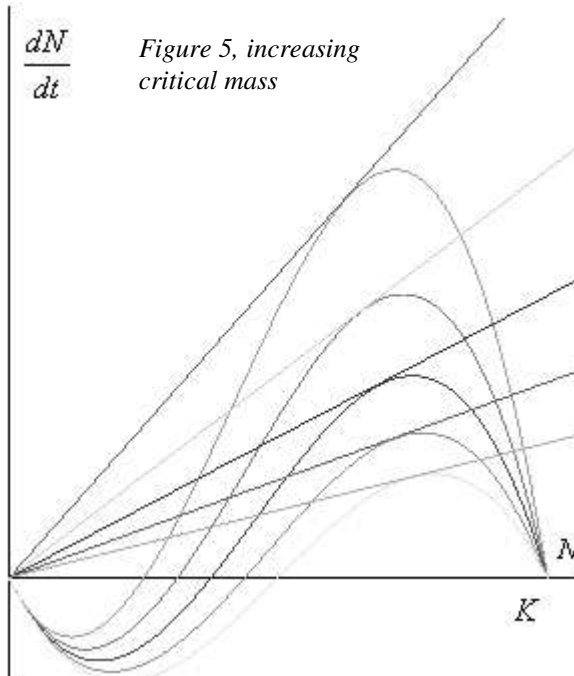
From the solution, which describes the effort needed to achieve a catastrophe, there are several important conclusions that can be drawn. To minimize E, the denominator should be as large as possible. The model indicates that by far, the most effective method of causing a catastrophe would be to raise α . A high α would reflect a method of user removal that would be effective in scaring or otherwise discouraging new users from joining the network. It might work to randomly sue users, and even if the lawsuits were not successful, it might create enough trouble to keep people from wanting to use the network.

Another potential weakness comes from the critical mass and carrying capacity. The larger the critical mass is, and conversely, the smaller the carrying capacity is, the easier the system can be made to collapse. The critical mass is strongly dependent on p , the probability of

users' searches returning hits, and it may be possible to lower this factor by selectively threatening users which carry a large number of files, or by posing as legitimate users while carrying false files. This would frustrate users of the network, causing the critical mass to be smaller. If this were to happen, a smaller effort would be needed to achieve a catastrophe, as shown in Figure 5.

Also, the carrying capacity may be interfered with, also by making p smaller (misses produce further queries), or by broadcasting fake requests across the network. This would consume bandwidth and lower the carrying capacity so that the network would support a lower number of workers. This also results in a smaller amount of effort needed to cause a catastrophe, as shown in figure 6.

Thus, through a combination of removal methods, targeting random users as well as selectively pursuing large file-carriers, P2P networks may be pushed to the tipping point. Empirically this was observed in the Xerox study, in which “nearly 50% of all responses are returned by the top 1% of sharing” and obviously shutting down this top 1% would do great damage to the network. This approach



could be combined with adding stress to the network by posing as legitimate users while flooding the other users with requests, which would consume bandwidth, which may frustrate users, and degrade performance. Because the nature of P2P is very democratic, imposters may be difficult to stop, as they infiltrate and break the system from within.

Improvement

However, whatever the implications of the model may be, it is still somewhat lacking because the model is only conceptual. It is not clear if the assumptions, such as balanced networks, are actually correct. For example, last year the Gnutella network actually collapsed^{vii}, but it happened because the connection speeds were not homogenous across the network, and low bandwidth clients walled off the communication between high bandwidth clients. The model presented in this paper would not account for such phenomenon, because of the homogenous network assumption.

One of the most damaging critiques of this model also relates to the balanced network assumption. Although the Internet is widely interconnected, it is not clear that examining a subset of the network, as this paper does, is a fair analysis of the entire network. If one subset was wiped out, the consequences for the rest of the network may be very damaging, or not at all, depending on the level of interconnectedness. For example, if the entire Internet was shutdown, but a large group that was “clumped” together away from other users continued, they might still be able to trade files. Thus, students in dorms may never be stopped from trading music with each other, although it may be the goal of the companies to keep the phenomenon from exploding to a global level.

Also, there are clear differences between the growth of users and the growth of animals. If one were able to push a fish population below the critical mass in a single day, it would suffer a catastrophe and never recover. However, if one were able to shut down a P2P network for a day,

people might still keep the programs on their computers and try again later, and thus the user base could recover and jump over the critical mass.

Thus, although the effort needed to cause a catastrophe has been solved in this paper, it is also not clear whether or not reaching this point is ever feasible. It may be that potential punishment is not serious enough, or cannot happen widely enough, to deter new users. Or perhaps the carrying capacity of a well-designed P2P network is huge, and amount of flooding can overwhelm the network.

To find the answers to these questions, one must attempt to measure each of the values included in the model empirically, would require extensive research. Thus, many of these issues are potential areas of improvement and research in the future.

References

ⁱ Levine, Dan. *Not the Real Slim Shady*

http://www.salon.com/tech/feature/2002/06/10/eminem_mp3/?x

ⁱⁱ Dickinson, Todd. 2000. *Dickinson on Internet Piracy of Music, Software*

<http://usinfo.state.gov/topical/global/ecom/00072004.htm>

ⁱⁱⁱ Costello, Sam. 2001. *Napster Usage Nosedives 90 Percent*

<http://www.pcworld.com/news/article/0,aid,51903,00.asp>

^{iv} Ritter, Jordon. 2000. *Why Gnutella Can't Scale. No Really.*

<http://www.darkridge.com/~jpr5/doc/gnutella.html>

^v Adar, Eytan and Huberman, Bernardo. 2000. *Free Riding on Gnutella.*

http://www.firstmonday.dk/issues/issue5_10/adar/

Although it is not an innate characteristic of P2P networks to have freeloaders, Xerox research has empirically observed the number to be as high as 70%:

http://www.xerox.co.uk/news/03_10_00b.htm

^{vi} Tung, K. K., 2001. Lecture 3, AMATH 383

<http://amath.washington.edu/courses/383-autumn-2001/lecture3.pdf>

^{vii} Truelove, Kelly. 2001. *Gnutella: Alive, Well, and Changing Fast*

<http://www.openp2p.com/pub/a/p2p/2001/01/25/truelove0101.html>