A DISTRIBUTED AND MEASUREMENT-BASED FRAMEWORK AGAINST FREE RIDING IN PEER-TO-PEER NETWORKS

Murat Karakaya, İbrahim Körpeoğlu and Özgür Ulusoy
Computer Engineering
Bilkent University Ankara, Turkey
{muratk,korpe,oulusoy}@cs.bilkent.edu.tr

Abstract. Peer-to-peer networks have attracted a significant amount of interest as a popular and successful alternative to traditional client-server networks for resource sharing and content distribution. However, the existence of high degrees of free riding may be an important threat against P2P networks. In this report, we propose a distributed method with the aim of reducing the degree of free riding in P2P networks. We primarily focus on locating free riders and taking actions against them. We propose a model in which each peer monitors its neighbors, makes decisions and takes appropriate actions. We specify three different free riding types and their symptoms observable from neighboring peers. We employ simple formulas to determine if a peer exhibits any kind of free riding. The counter actions to be applied to free riders are defined. We combine the mechanisms proposed to detect free riders and take appropriate actions in an Event-Condition-Action rule and a state diagram. Furthermore, we describe possible attacks to the proposed mechanisms and show how the system can handle them. By reducing the amount of free riding in a P2P network, we expect to increase quality of service, availability of content and services, robustness of the system, network load balancing, and scalability of the network.

1 Introduction

Peer-to-peer (P2P) networks have attracted a significant amount of interest both in the Internet community and in the academic world as a popular and successful alternative to traditional client-server networks for resource sharing and content distribution. P2P networks are implemented as overlay networks on top of the existing Internet infrastructure. There have been many system proposals and applications in the main functional areas of P2P paradigm such as data placement, file lookup, replication, etc. Most of these efforts aim to provide efficient, effective and fast exchange of files between peers. Although there are different architectural designs and applications for P2P file sharing, in nearly all P2P systems files are stored at peers, searched through the P2P network mechanisms, and exchanged directly between peers using the underlying network mechanisms. In an ideal case, a file that is downloaded by a peer is automatically opened for sharing with other peers. However, peers can, and frequently do, turn off this property and stop sharing a downloaded file to economize on their own resources such as bandwidth. Therefore, the primary property of P2P systems, the implicit or explicit functional cooperation and resource contribution of peers, may fail and lead to a situation called *free riding*.

As a P2P concept, *free riding* means exploiting P2P network resources (through

searching, downloading objects or using services) without contributing to the P2P network at desirable levels. Researchers have observed the existence of high degrees of free riding in P2P networks and they suggest that free riding may be an important threat against the existence and efficient operation of P2P networks. Adar and Huberman argue that "free riding leads to degradation of the system performance and adds vulnerability to the system. If this trend continues copyright issues might become moot compared to the possible collapse of such systems" [3]. Free riding is a serious problem for commercial P2P systems that would like to charge the users for accessing the resources and make profit in this way. If free riding can not be prevented, such systems can not count on the altruistic contribution of all peers, as happens in today's free systems [5], to have a successful operation and use of the P2P network. In those systems, to increase the system value to users by improving the number and variety of the available files, some incentive schemes would be required.

There may be various reasons and motivations behind free riding. For example, peers with a Network Address Translation (NAT) address may act as a free rider. Bandwidth limitation would be another cause. Another reason would be the peers' concern of sharing "bad" or "illegal" data on their own computers. Some peers may concern about security if they share something.

Free riding may cause several negative side effects on P2P networks. In a free riding environment, a small number of peers will serve for all other peers. Therefore, many download requests will be directed towards these peers which may lead to scalability problems [10]. Renewal of content or presenting interesting content may decrease in time, thus the number of shared files may become limited or may grow very slowly. Fault-tolerance property of P2P networks may be adversely affected due to the fact that a very small portion of the peers provides most of the content[1]. This also leads to a client-server like paradigm [11, 13] and decreases P2P network advantages. Quality of search process may degrade due to increasing number of free riders in the search horizon. As the peers age in the network, they begin not to find interesting files and may leave the system for good with all the files that shared earlier [10, 6]. Moreover, the large number of free riders and their queries will generate a great amount of P2P network traffic, which may lead to degradation of P2P services. Furthermore, underlying available network capacity and resources will be decreased by free riders, which will cause extra delay and congestion to non-P2P traffic.

In this report, we propose two mechanisms to cope with free riding. The first mechanism primarily focuses on locating and detecting free riders, whereas the second one deals with taking actions against them. We propose a model in which each peer monitors its neighbors, makes decisions, and take actions accordingly. We make several distinctions between free riders and contributors to enforce free riders to comply with cooperation. As the first step of our work, we propose several design criteria which should be met by any P2P system aiming to prevent free riding. Then, we specify three free riding types and their symptoms observable from neighboring peers. We present simple methods and formulas to determine if a peer exhibits any kind of free riding activity. The counter actions which will be applied against free riders are also specified. We then integrate the hints that suggest a peer to be free rider and the counter-actions that can be applied to such a peer using a finite state diagram that shows the possible states and the transitions between them. We also represent the transitions using an Event-Condition-Action (ECA) rule that enables automatic execution of counter-

[1] 1% of the peers provides 37% of the content [3].

actions upon updates and depending on the current conditions. We identify three possible counter-actions that can be applied against a peer that is exhibiting free-riding behavior.

The organization of the report is as follows. Sections 2 and 3 are devoted to the related work and background, respectively. In Section 4, following the discussion of the design criteria and performance metrics, the mechanisms for locating free riders and taking actions against them are described. In the last part of this section, integration of the proposed mechanisms in the form of a finite state machine (FSM) and an ECA rule is presented. Section 5 discusses the possible attacks against the proposed mechanisms by free riders. The conclusion is presented in Section 6.

2 Related Work

User traffic on Gnutella network is extensively analyzed in [3] and it is observed that 70% of peers do not share any file at all. Furthermore, 63% of the peers who share some files do not respond to any queries. That is, they are sharing some files but nobody is interested in them and therefore no queries are generated searching for these files. Another interesting observation is that 25% of the peers provide 99% of the whole content in the network. Adar and Huberman propose some ideas to prevent free riding [3]. Two of these ideas are automatic replication of data in the network as in FreeNet [14] and automatically sharing downloaded files as in Kazaa [16]. However, the authors also point out the drawbacks and their concerns about the possibility of practical use of these ideas. They propose to implement a market-based architecture that allows a peer to exchange computer resources. However, they do not propose any specific model or method for applying their ideas.

In a more recent work, Saroui et al. confirm that there is a large amount of free riding in Gnutella network as well as in Napster [13]. Another interesting observation is that 7% of the peers together provide more files than all of the other remaining peers. The authors suggest that the system should not treat its peers equally, on the contrary it should provide the right incentives and rewards for peers to provide and exchange data.

In [10], Ramaswamy and Liu concentrate on how to prevent free riding. They propose to calculate a utility function for each peer in order to estimate its usefulness to all community. According to the result of the function, P2P network will permit a peer to search and download a file or just reject its request. The function is based on two parameters; the total size of the files downloaded and uploaded by the peer. The difference of two values determines the utility of the user to the system. If the user requests a file to download with a size less than the utility value, then it is permitted to download. Otherwise, it is refused. There are two ways to increase the utility value, either by uploading new files or by waiting for some time for a bonus utility value. When a peer downloads a file, its utility is decreased by the amount of the size of the downloaded file.

With the proposed method, free riders can not download files from the system if the utility value is lower than the size of the requested file. However, there can be some ways to walk around the utility values. For example, a user can share some small files with fake names resembling popular file names. Other users can download these files and the peer gets utility values for them. Moreover, the proposed method depends on accurate information about peers which is provided by the peers themselves. A P2P network depending on such a protocol can be misreported and cheated by

rewriting some malicious client programs. For example, KazaA P2P client program [16] implements a method like the one proposed in [10] in order to prioritize users request at the source peer. However, a modified version, Kazaa Lite [15], has been released and it maliciously declares its user to be a "Supreme Being" which is the peer with the highest participation level. Therefore, we think that this method can not fully prevent free riding. Any method proposed to hinder free riding should be designed in such a way that it should not solely depend on user-submitted information, or it should create the right incentives for the peers to report accurate information [15]. This is because, free riders may possibly not tell the truth about themselves, if they are not given any incentive to do so[2].

In a recent work [17], Vishnumurthy et al. suggest using a single scalar value, called Karma, to evaluate a peer's utility to a system like in [10]. Each peer has an account consisting of Karma. When a peer uploads a file to a requesting peer, it gets some amount of Karma from the requesting node. On the other hand, if it downloads, it gives some amount of its Karma to source peer. The account of a peer is replicated by a group of peers, called the bank-set, in order to ensure Karma against loose and tampering. The transfer of Karma between peers is executed through bank-set of each peer. The main difference from the work in [10] is that the utility value of a peer is not stored at the peer itself but at some other peers.

However, to make the scheme work, a group of peers must be known to store Karma value. Whenever a peer's Karma changes, a predefined number of these peers should be reachable. Therefore, the identification of the peers should be known and be permanent. However, unstructured P2P networks do not support permanent and reliable identification mechanisms. Thus, the prototype of the proposed scheme was implemented on top of Pastry, which is a Distributed Hash Table (DHT), in other words a structured P2P network.

In our work, we do not propose to use any scoring value for a peer's utility to the system. Therefore, we don't have to bother with storing, retrieving, and saving a utility value. At each peer, we just store the information about the neighbors' messages which are routed by the peer itself. Furthermore, we do not require the explicit cooperation of any group of peers to make the system work. Each peer executes the same kind of mechanisms alone and does not depend on any other peer's cooperation. Our proposal can be implemented on both types of P2P networks, that is structured and unstructured networks.

3 Background

In this section, we describe some basic P2P systems concepts and protocols upon which our approach and schemes are built. We focus on unstructured P2P networks like Gnutella, because of their popularity and well-known protocols [4]. Unstructured P2P networks have the distinct properties that can be summarized as [2]:

- no central coordination
- no central database

[2]For example, about 30% of the Napster users do not report their bandwidths or misreport with less values [13]. Another example may be given from Seti@Home project [8]. Some peers modify their client programs so that they appear to the system as if they were doing more work than what they are actually doing. In this way, they abuse the scoring system and their names are displayed at the top of computation units contributed list.

- no peer has a global view of the system
- global behavior emerges from local interactions
- all existing data and services should be accessible
- peers are autonomous
- peers and connections are unreliable

These features enabled unstructured P2P networks to be very successful, but also brought some problems. Among the problems of such networks is the so-called reputation problem. In an unstructured P2P network such as Gnutella, peers interact with unknown peers and have no information about their reputations. In other words, they do not know to what extent they can trust the other peers and the data provided by them. As a result, the detection of free rider peers and actions against them can not be easily implemented. In this work, we propose a distributed and local solution which does not require any persisting information about peers.

3.1 Probable Causes of Free Riding

The motivations and reasons behind free riding may be different. For example, peers with a Network Address Translation (NAT) address may act as a free rider. Multiple computers share the same domain of IPs through NAT. Therefore, if both peers are using NAT-based IP, they cannot download files from each other. As stated in [3], 16% of observed peers are using NAT-based IP and this corresponds to about 2% of the transactions. These peers cannot upload files and therefore they become free riders even they share files.

Bandwidth limitation would be another cause. In the first sight, one may think that peers have scarce bandwidth and so they do not want to share it by uploading files. However, the analysis of network traffic [3] shows that there is no strong correlation between bandwidth and free riding. That is, a peer with a large amount of bandwidth has the same tendency to be a free rider as a peer with poor connection.

Another reason would be the peers' concern of sharing "bad" or "illegal" data on their own computers even they use them. Because, they do not want to be responsible for the data item on any occasion of surveillance. In addition, peers may download content, use it (watch, listen, read, etc.) and delete it immediately. For some types of files such as videos, they do not need to refer to the same file more than once. Moreover, peers may worry about their resources and do not want to share them. If they can use the system for free, then they may not want to contribute anything to the system.

Some peers concern about security if they share something. Active cooperation with P2P network might frighten peers. One important incentive would be the fact that most of the existing P2P systems do not care about the active cooperation of the peers. Protocols are designed as if each peer were volunteer to cooperate and each peer contributes to the system equally. Therefore, in most P2P networks, all peers enjoy the equal and same services even though some of them do not obey the expectations. Peers neither benefit from sharing files to community nor face any loss from not serving.

3.2 Participation Levels

In general, free riding occurs as low or no participation of peers to P2P network. Participation levels to P2P network can be classified as follows:

- Zero content contribution: Peers do not share any content.
- Uninteresting content contribution: Peers share some content but nobody searches for them.
- Only replication: Peers share only downloaded content, and do not create new content.
- Network traffic regulation: Peers may act as super peers which route queries and provide integrity of the whole network. They do not necessarily contribute content.
- Large number of replication: Some peers reserve large capacity for sharing files, while they download less or no content.
- Original content contribution: Peers upload original and new content to network. They can produce content by themselves or by just transferring new content. This process can be legal or illegal.
- System design: Firms or individual programmers develop and distribute P2P protocols and client programs. Even they seem to be unrelated to participation level, they determine how the peers interact with the whole system. If the system designers do not obey the existing system protocols or do not care about cooperation and contribution of peers while developing P2P protocols and client programs, there could not be any participation at all and free riding may exist in all forms.

In our work, we aim to increase the participation level of any peer to reduce the amount of free riding effects on the network. The participation types given above can be observed as a single or combination of several.

4 Mechanisms Against Free Riding

We think that in order to reduce the amount of free riding and increase cooperation among peers, availability of a set of mechanisms is required. Many existing P2P systems have implicitly assumed the altruistically contribution of peers, however the reality is different. We believe that in current P2P systems there are not enough incentives for peers to participate and contribute. Therefore, we try to create an environment in which peers will be monitored about their contributions to a P2P network and enforced to act in more cooperation and to contribute to continue to use the services and resources of the P2P network.

We are not attempting to eliminate all possible kinds of free riding from a P2P system completely[3]. We are aiming to improve the situation and reduce the ill-effects of the problem by increasing the participation levels of peers, and hence their contributions, as much as possible.

[3]For example, the scope of our work does not aim promoting or enforcing new content contribution by peers. We have two reasons for this. First, in reality it may be impractical to judge easily if a file is new to the system or not. Second, it could not be possible for every peer to create and contribute original content.

4.1 Design Criteria

While developing some mechanisms to prevent or diminish free riding, one should consider several issues some of which are listed below:

- **Simplicity:** The actions observed and reactions to them should be simple to implement and manage.
- **Decentralism:** Making decisions and taking actions should be executed in a decentralized way.
- **Cooperation:** Besides decentralism, cooperation should be intensified with coordination among peers.
- **Low overhead:** Methods should not cause much overhead. Non-free riders should not devote much resources to prevent free riding.
- **Abuse-proof:** Peers may try to walk around mechanisms by misreporting their status or implementing their own client programs. Mechanisms must not depend on information provided by peers solely. Instead, mechanisms should depend on P2P paradigm to collect information about peers.
- **Fairness:** Peers with low bandwidth connections may not contribute even they are willing to do so. The peers with NAT-based IPs also behave like free riders. Furthermore, there is an asymmetry between upload and download bandwidths for most of the peers which results in better download speeds and quality compared to uploads. For these reasons, mechanisms and policies applied against free riders should be fair and smart enough to distinguish the peers which are not real free riders.

4.2 Performance Metrics

As stated before, the aim of this work is to prevent the amount of free riding that can occur in a P2P system to some degree. In this way, we are targeting to reduce free riding occasions and their effects to the system performance. With that we can have better performing P2P networks, where participation and contribution levels are high.

The performance improvement of a P2P system has to be measured by using some metrics. We would like to use the following metrics to evaluate the performance of a P2P system that applies our schemes to prevent and reduce free riding.

- Quality of Service: If possible, Quality of Service (QoS) for non-free riders should be increased, while diminishing them for free riders. QoS parameters may be specified as, search time, hit quality and quantity, and download time.
- Availability: We expect that by increasing the cooperation between peers for eliminating free riding and by forcing free riders to contribute, availability of content and services in P2P network can be increased. For example, a scheme that could replicate popular items on free riders would increase hit ratio for those items, even though the original content providers leave the system.

Descriptor	Description	Content
Ping	Used to actively discover hosts on the network. A servent receiving a Ping descriptor is expected to respond with one or more Pong descriptors.	Nothing
Pong	The response to a Ping. Includes the address of a connected Gnutella servent and information regarding the amount of data it is making available to the network.	IP and port of responding host, number and size of files shared
Query	The primary mechanism for searching the distributed network. A servent receiving a `Query` descriptor will respond with a `QueryHit` if a match is found against its local data set.	Minimum speed requirement of the responding host;search string
QueryHit	The response to a `Query`. This descriptor provides the recipient with enough information to acquire the data matching the corresponding `Query`.	IP and port, speed of responding host; number of matching files and their indexed result set
Push	A mechanism that allows a firewalled servent to contribute file-based data to the network.	Responding host id; file index; IP and port of requesting peer

Table 1: Gnutella Protocol Descriptors

- Load Sharing and Scalability: Content provider peers can be bottleneck to due excessive search and download operations they are involved in. Via increased cooperation in P2P system, the load on peers can be also shared by peers that would otherwise be free riders. This will help the system to be a more scalable so that larger number of search queries and download operations can be executed on the system successfully.

- Robustness: Mechanisms against free riding can make a P2P network more robust against disconnections and legal attacks, which will increase network population in terms of available content and also in terms of number of nodes that are reachable. This will expand the search horizon and will increase the hit ratio in search operations.

4.3 Locating and taking actions against free riders

We propose a system in which every peer passively monitors the other peers. In the proposed system, peers can be classified into two different roles (see Figure 1). In the first type of role, a peer functions as a monitoring peer, P_M, which monitors and records the number of messages coming from and going to neighboring peers. The messages are implemented with descriptors in Gnutella Protocol [4] (See Table 1), the protocol upon which our solution is based. At the same time, each peer is a controlled peer, P_C, which means that its messages are monitored and recorded by its neighboring peers.

By examining the messages from a neighbor, and compiling the information recorded about the neighbor and its related messages, a monitoring peer may suspect a neighboring peer to be a free rider. Then it can take counter actions against this suspected peer (i.e. controlled peer).

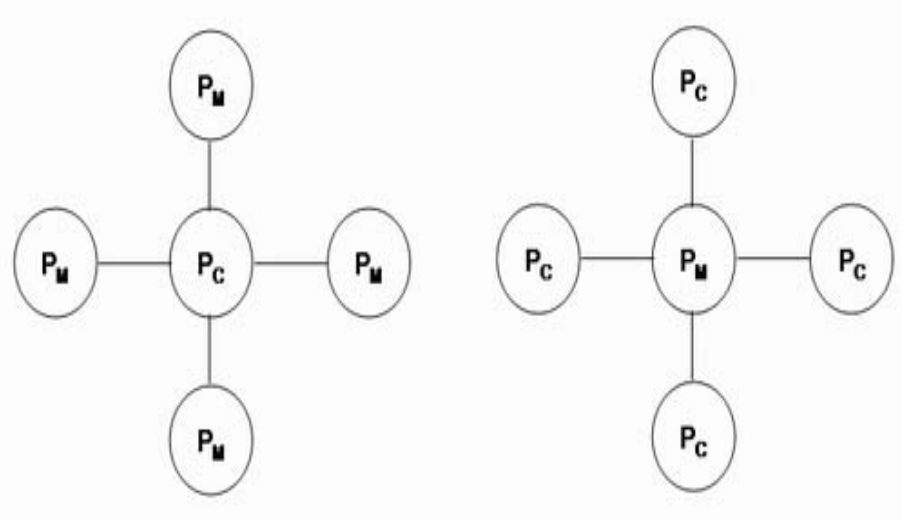

Figure 1: Peers are in two roles: monitoring and controlled.

Symbol	Description
Q_P	Number of `Query` descriptors submitted by peer P.
RQ_P	Number of `Query` descriptors routed by peer P.
TQ_P	Number of `Query` descriptors routed towards peer P.
QH_P	Number of `QueryHit` descriptors submitted by peer P.
RQH_P	Number of `QueryHit` descriptors routed by peer P.
SQH_P	Number of `QueryHit` descriptors satisfying queries of peer P.
N_P	Number of `Notify` descriptors submitted for peer P.

Table 2: Observed Descriptors

4.3.1 Locating Free Riders

In order to determine if a P_C is a free rider or not, we may exploit several clues that may be derived from the behaviors of the neighbors. For each clue, we need to maintain some information about each neighbor and its behaviors. Due to the small-world phenomena, average number of neighbors is expected to be about 3-4 [7]. Therefore, this process does not impose high overhead on peers.

The information that is maintained about neighbors of a peer consists of some statistical counters which are presented in Table 2. These counters are updated when messages are received from the neighbors and when messages are sent towards the neighbors. The clues about the neighboring peers (if they are free riders or not, and the types of free riding they exhibit if they are free riders) are derived from the values of these counters.

One issue to consider is whether there exists enough time to collect information and make decisions. It is known that in a P2P network, peers can join and leave the system at any time. We can find some related work in the literature about the network topology dynamics and peer characteristics of P2P applications. In [11], it is stated that about 40% of the peers leave the Gunetella network in less than 4 hours, while only 25% of the peers are alive for more than 24 hours. In another work [13], the median session duration of both Napster and Gnutella clients is about 60 minutes. In a similar work [6], 90% of average session lengths of Kazaa clients is found to be about 30 minutes. In summary, it can be assumed that peers stay connected long enough to collect information about them and take necessary actions.

Another issue is whether a monitoring peer can snoop and monitor enough number of messages that are coming from or going towards the neighboring peers. In [9], it is reported that the average number of queries per second for three peers located at different geographic locations is about 50. Also, about 30 query responses per second are recorded. Query response ratio is ranging around 10%-12%. This shows that a peer will have enough number of messages forwarded over itself to judge if a neighboring peer is a free rider or not.

4.4 Free Riding Types

In the following discussion, we mention several possible free riding types and identify some clues that can be used to detect them.

- **A peer does not share anything at all or shares uninteresting files.** It may be observed that a neighboring peer does not return any `QueryHit` messages to the queries that it receives. There may be two reasons for that: either the peer does not have any files matching the queries, or the peer does not share any files at all. To decide if a peer is a zero-content (or an uninteresting content) contributor, whenever the monitoring peer initiates a search or routes a search on behalf of another peers by sending a `Query` message to its neighbors, the monitoring peer also increases the value of the respective TQ counters[4] for its neighbors. The monitoring peer also observes and counts the `QueryHit` messages received from the neighboring peers. If the monitoring peer receives a `QueryHit` message that has the IP address of one of its neighbors in it, the monitoring peer increases the value of the QH counter maintained for that peer in the log table. Receiving a `QueryHit` message originating from a neighboring peer indicates that the neighboring peer is sharing an interesting file that is requested.

  The monitoring peer then compares the values of TQ and QH counters maintained for a neighboring peer, to decide if that peer is a free rider that is not sharing any files (a *non-contributor*). More specifically, for this decision to be made, the monitoring peer may compare the QH/TQ ratio against a threshold value and decide that the neighbor is a free rider of type non-contributor if the ratio is smaller than the threshold. Several different approaches for setting up a threshold value may be proposed[5]. Below, we formulize the condition that is required to judge if a neighboring peer is a free rider or not.

  Furthermore, to remove the warm-up period and to obtain valid statistical information we propose to use a threshold value for the number of forwarded `Query` messages to the observed peer, τ_{TQ}. A monitoring peer start deciding about a neighboring peer after this threshold.

  if ($\mathrm{TQ}_P > \tau_{TQ}$) and ($\frac{QH_P}{TQ_P} < \tau_{non\_contributor}$) **then**
  peer P is considered as a `non-contributor`
  endif

[4]Different counter types used for locating free riders, including TQ, are described in Table 2.

[5]We may set up a constant value for unsatisfied query number, ($TQ_P - QH_P$), e.g. 100. Or we may use a time based threshold, e.g. 10 minutes. If there is no `QueryHit` message from the peer in that period of time, we may treat this peer as non-contributor.

- **A peer consumes more resource than that it shares.** A monitoring peer counts the `QueryHit` responses (QH) originated from its neighbors and successful `QueryHit` messages (SQH) destined to and received by its neighbors. The comparison of these two numbers reveals if any of the neighboring peers consumes more than it shares. More specifically, a threshold value, $\tau_{consumer}$, can be compared against the ratio of these two numbers. If the ratio QH/SQH is smaller than the threshold, a decision that the neighboring peer is a free-rider of type *consumer* can be made.

  if ($\mathrm{TQ}_P > \tau_{TQ}$) and ($\frac{QH_P}{SQH_P} < \tau_{consumer}$) **then**
  peer P is considered as a `consumer`
  endif

- **A peer drops others' queries.**

  A monitoring peer counts `Query` and `QueryHit` messages forwarded by each of its neighbors. If these two values are very low for a neighboring peer, it can be assumed that the neighboring peer does not have enough connections or it drops `Query` and/or `QueryHit` messages. $\tau_{dropper}$ is used as a threshold value. We call this type of free rider as a *dropper*.

  if ($\mathrm{TQ}_P > \tau_{TQ}$) and ($\frac{RQ_P+RQH_P}{TQ_P} < \tau_{dropper}$) **then**
  peer P is considered as a `dropper`
  endif

4.4.1 Actions against Free Riders

If a peer identifies another peer as a free rider, it can take some counter-actions against it. We specify three level of actions. Level 1 action is the least restrictive for the free rider. Level 3 action is the most restrictive for the free rider.

- **Level 1 Action: Decrementing TTL value:** Normally, when a peer receives a `Query` message from a neighboring peer, it first executes the search on local files for a match, then the `Query` is forwarded to the other neighboring peers. Before the `Query` message is forwarded, its TLL value is decremented by one. To act against a suspected free rider, the monitoring peer can play with the TTL value for `Query` messages that are received from the suspected peer, i.e. it can decrement the TTL value by more than one before forwarding. In this way, the search horizon of the free riding peer is narrowed down. This also reduces the overhead that `Query` messages are imposing on the network. This counter-action is applied to a peer that exhibits only one type of free riding, i.e. it is either a non-contributor, or a dropper, or a consumer.

- **Level 2 Action: Ignoring requests:** A free rider can be punished by the monitoring peer by ignoring the searches (i.e. the `Query` messages) originating from that free riding peer. The `Query` messages originating from the free rider peer can be partially or totally ignored. Ignoring a `Query` message means not searching the local files for a match and not forwarding the `Query` any more. In other words, the `Query` message is simply dropped. We can do this action parametric, so that the probability of ignoring (dropping) the `Query` messages can be adaptive and

tunable. However, a monitoring peer should be careful about the origin of the `Query` messages while dropping them. It has to drop only the messages that are originated from a free riding peer[6]. This counter-action is applied to a peer that is exactly exhibiting two types of free-riding (for example a peer that is both a consumer and a dropper).

We expect that ignoring the requests of free riders (fully or partially) does not only punish the free riders, but also improves the overall system performance. If not controlled, Query messages may become a significant fraction of overall network traffic. For example, as it is pointed out in [12], an 18 bytes of search string in a Query message may cause 90 megabytes of data to be forwarded by the P2P network peers. As another example, [1] states that total number of messages including the responses triggered by a single Query message can be as large as (assuming 4 connections per peer):

$$2 * \sum_{i=0}^{TTL} C * (C-1)^i = 26240 \quad (1)$$

In addition to these estimations, the network traffic measurement at the University of Wisconsin (http://wwwstats.net.wisc.edu) shows that P2P traffic (in and out) constitutes 9.3% of the total campus traffic whereas http-based traffic is about 52% in March 2004. Thus, P2P network traffic constitutes a considerable bandwidth usage in the Internet.We believe that decreasing the number of queries submitted by free riders may help improving the performance and scalability of P2P networks and the underlying internet.

- **Level 3 Action: Disconnecting from network:** If a peer is sure that a neighboring peer is a free rider that is exhibiting all types of free riding, the peer may drop the connection with that peer. In that way, the peer saves its resources which can later be allocated to another peer. The difference between ignoring (all or partial) search request and disconnection is that, in the preceding method, if any change in behavior of the peer is observed, the punishment can be cancelled. However, when disconnection is executed, the disconnected peer should reconnect to the system through a new peer.

4.5 Putting all together

In the previous sections we have discussed the clues to detect free riding types and possible counter actions that can be taken against free riders. We now would like to integrate them together using an ECA rule and a FSM. As described in section 4.4, a free-riding peer can be a non-contributor, or a dropper, or a consumer, or a combination of these. A neighboring peer can be also a good behaving peer, in which case it is not a free-rider and it will not show any of the mentioned free riding types.

[6]This issue then becomes how to decide if received Query is originated at a free rider neighbor or at some other peer. Usually, P2P protocols try to hide the identity of the originator of a Query message [4]. So, we can not check the identify of the originator by looking to the content of the `Query` message. However, we may exploit the TTL value of the request. If it is 7 (which is the value that originator of the message sets), it means that the request is originated from the neighboring peer. If it is less than 7, it is only forwarded by the neighboring peer.

As the first step towards a formal description, we will use three boolean variables to denote if a neighboring peer is non-contributor, or a dropper, or a consumer or not. We call these three boolean variables as N (for non-contributor), D (for dropper), and C (for consumer). If the counter values maintained at the log table of the monitoring peer indicate that the neighboring peer is a non-contributor, then N has the value 1, otherwise it has the value 0. If counter values at the log table indicate that the peer is a dropper, then D has the value 1, otherwise D has the value 0. If the counter values indicate that the peer is a consumer, then C has the value 1, otherwise it has the value 0.

When the counters maintained for the neighboring peer P at the log table of the monitoring peer change, the values of these variables (N, D, and C) may also change. For example, if (QH_P/TQ_P) ratio was first smaller than the respective threshold (i.e. N = 1), and later becomes greater than that threshold, peer P becomes no longer a non-contributor and the value of N changes from 1 to 0.

At any moment in time, depending on the counter values maintained in the log table of a monitoring peer (and hence depending on the values of the above mentioned three boolean variables) we may have one of the following eight conditions shown in Table 3 holding for a neighboring peer P.

N	D	C	Condition
0	0	0	C0
0	0	1	C1
0	1	0	C2
0	1	1	C3
1	0	0	C4
1	0	1	C5
1	1	0	C6
1	1	1	C7

Table 3: Conditions

If, for example, condition C0 holds at a given time, that means there is no free-riding at all at that time. If one of the conditions C1, C2, or C4 holds at a given time, that means only one type of free riding is exhibited by the neighboring peer P. This means, peer P is either a non-contributor, or a dropper, or a consumer. In other words, either N, or D, or C has the value of 1, and the other two variables have the value of 0. If one of the conditions C3, C5, or C6 holds at a given moment, that means the peer is exhibiting exactly two types of free riding. In other words, both N and C, or both N and D, or both D and C are 1. If condition C7 holds at a given moment, that means the peer is showing all types of free riding, i.e. the peer is a consumer, a dropper, and at the same time a non-contributor.

A monitoring peer may apply the appropriate counter-action policy against a neighboring peer depending on the values of N, D, and C (i.e. depending on the current condition defined by these three variables). Table 4 shows what action is to be taken against the neighboring peer under which condition. If, for example, condition C0 holds, there is no free-riding and therefore no counter-action is applied against the peer. If one of the conditions C1, C2, or C4 holds, then the level 1 counter-action is applied. If one of the conditions C3, C5, or C6 holds, then the level 2 counter actions is applied. If condition C7 holds, the level 3 counter action is applied and the peer is disconnected.

We will represent each row at the above table with a state in the monitoring peer. When there are updates on the log table counters, the state of the monitoring peer may

Conditions	Action Level	Action Description
C0	Level 0	No counter-action
C1, or C2, or C4	Level 1	Reduce TTL by more than 1
C3, or C5, or C6	Level 2	Ignore Requests Partially
C7	Level 3	Disconnect

Table 4: Conditions and Counter-Actions

change, since the condition may change. In each state (i.e. each row of table above) we will apply a different counter-action to a peer. We have four states: S0, S1, S2, and S3 (as shown in Table 5). When the monitoring peer is in state S0 for a neighboring peer P, only the condition C0 may hold and no counter action is applied. When the monitoring is in state S1, one of the conditions C1, C2, or C4 may hold, and the level 1 counter action is applied.

State	Conditions	Action Level
S0	C0	Level 0
S1	C1, or C2, or C4	Level 1
S2	C3, or C5, or C6	Level 2
S3	C7	Level 3

Table 5: States, Conditions, and Counter-Actions

If we state briefly, the level of counter action to be applied depends on the current state. At state S1 the level 1 counter action is applied, at state 2 the level 2 counter action is applied, and at state 3 the level 3 counter action is applied. We may have a transition between two states when the condition (values of N, D, and C) changes upon updates on the log table. The Figure 2 shows the whole state diagram that shows all possible transitions. If, for example, the monitoring peer is in state S1 for a neighboring peer and an update occurs on the log table that causes a new condition to appear, the monitoring peer can make a transition to either the state S0, or the state S2, or the state S3 depending on the new condition. If the new condition is C5, the new state becomes S2; if the new condition is C7, then the new state becomes S3; if the new condition is C0, then the new state becomes S0.

On the above state diagram, if the monitoring peer makes a transition towards right, it means it is increasing the level of counter-action that is to be applied to the neighboring peer. If the monitoring peer makes a transition towards left, it means that it is decreasing the level of counter action that is to be applied to the neighboring peer.

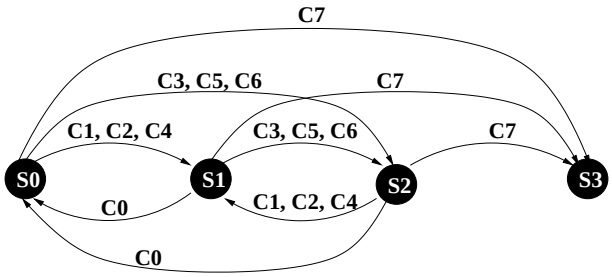

Figure 2: State diagram

We can express the transitions that a monitoring peer has to make upon an update on the log table and a change on the condition using an ECA rule (Figures 3, 4, 5, 6). Upon a transition, the monitoring peer moves to a new state where it applies a different counter action policy towards the neighboring peer.

---

define rule *StateTransition*
on update log_table($QH_P, TQ_P, ...$)
if (*State* == *S0*) **then**
    Execute ConditionalAction-0
elseif (*State* == *S1*) **then**
    Execute ConditionalAction-1
elseif (*State* == *S2*) **then**
    Execute ConditionalAction-2
endif
endrule

Figure 3: ECA rule that governs the state transitions

5 Possible Attacks and Counter Measures

As pointed out in section 4.1, an important design issue for mechanisms against free riding is having abuse-proof property. Aiming to investigate this issue, in this section we discuss a list of possible counter attacks against free riding prevention mechanisms. We also discuss how we can defend against those kind of attacks.

- **Fake `QueryHit` Messages:** A free rider can cheat its neighbors by replying to some queries with `QueryHit` messages fraudulently as if it has the requested file. But when the requesting peer asks for the file, it may just refuse the connection. In this way it may seem to the network as it is serving well, since neighboring peers may not be aware of unsuccessful download and cheating (download path may be different than the `Query` and `QueryHit` paths). In the log tables of the neighbors, the malicious provider may seem to be a non-free rider because of its `QueryHit` replies.

  Given the descriptors in Table 1, it may not be possible for a neighboring peer to observe and perceive this kind of fake messages. Because, download occurs between two peers outside the P2P network and there is no feedback mechanism or reputation concept in unstructured P2P networks. To prevent this kind of

ConditionalAction-0:

if ($Condition == C7$) **then**
    State = S3;
elseif ($Condition == C1$
    $\vee$ $Condition == C2$
    $\vee$ $Condition == C4$) **then**
    State = S1;
elseif ($Condition == C3$
    $\vee$ $Condition == C5$
    $\vee$ $Condition == C6$) **then**
    State = S2;
else
    Do not change state
endif

Figure 4: Contitional Action 0

fake `QueryHit` messages, we propose to use a new descriptor (see Table 6). The descriptor is used to report a malicious peer to its neighbor to inform that the peer is believed to be a cheater. When a querying peer is refused by a responding malicious peer during the attempt of the download, the querying peer may send a `Notify` descriptor through the P2P network to reach the neighbor of the malicious peer. To avoid an increase in the network traffic, the querying peer does not broadcast the descriptor message. Instead, it forwards the descriptor to only one neighbor which has delivered the `QueryHit` message, containing the IP of denying peer. Any intermediate peer on the way to the last peer, forwards the `Notify` message to only one neighboring peer based on the `Query` Descriptor Id. The last peer is the neighbor of the suspected peer. It checks and logs the `Notify` message and takes necessary actions against the suspected malicious peer. The `Query` descriptors are stored in the network for some time to route `QueryHit` descriptor in the same backward path. Therefore, we do not enforce to store new information on the peers. However, the time for deleting records from routing tables should be extended such that an unsuccessful download attempt can be executed.

There could be some side effects of the proposed `Notify` descriptor. As stated before, small number of peers provide large amounts of files in the system and they are posed to heavy download traffic. Furthermore, some peers have very limited upload bandwidth. These peers may refuse more connections when they reach the maximum number of connections. If a download from such a peer can not be initiated, submitting a `Notify` descriptor for this peer would be unfair and incorrect. To hinder these kinds of false notifications, we propose to use a ratio of the `Notify` messages to `QueryHit` messages for a given peer. If it exceeds a predefined parameter, e.g. 80%, then proper actions can be taken against the peer.

- **Fake Files**

ConditionalAction-1:

if ($Condition == C0$) **then**
    State = S0;
elseif ($Condition == C3$
    $\vee$ $Condition == C5$
    $\vee$ $Condition == C6$) **then**
    State = S2;
elseif ($Condition == C7$) **then**
    State = S3;
else
    Do not change state
endif

Figure 5: Conditional Action 1

Descriptor	Description	Content
`Notify`	Used to report a suspected peer that refused to upload the file it provided in `QueryHit` descriptor in respond to a given `Query` descriptor.	`Query` Descriptor Id; Suspected peer IP;File Index

Table 6: New Protocol Descriptor

Free riders could share dummy files with popular names in order to cheat querying peers. These files can be very small in size to incur upload overhead. In that way, free rider peers can conceal themselves. We believe that this situation can be prevented by using the `Notify` descriptor as proposed above.

6 Conclusion

In this work, we have proposed a distributed and measurement based method to reduce the degree of free riding in unstructured P2P networks. We have first specified possible free riding types and counter actions that can be taken against free riders. Then, we have proposed mechanisms which can detect free riders and employ counter actions against them. Furthermore, we have combined these mechanisms into a formal framework by using an ECA rule and a finite state machine showing what kind of counter action is to be applied under which condition. The mechanisms proposed for reducing the amount of free riding meet the essential requirements of P2P paradigm, such as distributed computing, anonymous connections, unreliable connections, and so on. We have also proposed a new descriptor (a `Notify` message) to be used against the counter attacks by malicious peers to the proposed mechanisms.

By reducing the amount of free riding in a P2P network, we expect also to increase the quality of service that peers can get from the network, the availability of content and services, the robustness of the system, the balance of the load on network peers and elements, and the scalability of the network.

ConditionalAction-2:

if ($Condition == C0$) **then**
    State = S0;
elseif ($Condition == C1$
    $\vee$ $Condition == C2$
    $\vee$ $Condition == C4$) **then**
    State = S1;
elseif ($Condition == C7$) **then**
    State = S3;
else
    Do not change state
endif

Figure 6: Conditional Action 2

We are currently developing a simulation program to implement and evaluate the proposed system and different mechanisms proposed to handle free riding problem in P2P networks.

References

[1] Karl Aberer and Manfred Hauswirth. An overview of peer-to-peer information systems. *WDAS*, 2002.

[2] Karl Aberer and Manfred Hauswirth. Peer-to-peer information systems: Concepts and models, state-of-the-art, and future systems. *18th International Conference on Data Engineering (ICDE)*, 2002.

[3] Eytan Adar and Bernardo A. Huberman. Free riding on gnutella. *http://www.firstmonday.dk/issues/issue5_10/adar/*, 2000.

[4] Clip2. The gnutella protocol specification v0.4 (document revision 1.2). *http://www9.limewire.com/developer/gnutella protocol0.4.pdf*, Jun. 2001.

[5] Philippe Golle, Kevin Leyton-Brown, and Ilya Mironov. Incentives for sharing in peer-to-peer networks. *Proceedings of the Electronic Commerce'01*, 2001.

[6] Krishna P. Gummadi, Richard J. Dunn, Stefan Saroiu, Steven D. Gribble, Henry M. Levy, and John Zahorjan. *Measurement, Modeling, and Analysis of a Peer-to-Peer File-Sharing Workload*. In the Proceedings of the 19th ACM Symposium on Operating Systems Principles (SOSP-19), October 2003.

[7] M. Jovanovic, F.S. Annexstein, and K.A. Berman. Scalability issues in large peer-to-peer networks - a case study of gnutella. *Technical Report, University of Cincinnati*, 2001.

[8] Leander Kahney. Cheaters bow to peer pressure. *http://www9.wired.com/news/tecnology/0,1282,41838,00.html*, 2001.

[9] Evangelos P. Markatos. *Tracing a large-scale Peer to Peer System: an hour in the life of Gnutella*, pages 65–74. In the Proceedings of the second IEEE International Symposium on Cluster Computing and the Grid, May 2002.

[10] Lakshmish Ramaswamy and Ling Liu. Free riding: A new challenge to peer-to-peer file sharing systems. *36th Annual Hawaii International Conference on System Sciences (HICSS'03) - Track7,Big Island, Hawaii*, January 2003.

[11] Matei Ripeanu, Ian Foster, and Adriana Iamnitchi. Mapping the gnutella network: Properties of large-scale peer-to-peer systems and implications for system design. *IEEE Internet Computing Journal special issue on peer-to-peer networking*, 6, 2002.

[12] Jordan Ritter. Why gnutella can't scale. no, really. *http://www.darkridge.com/ jpr5/doc/gnutella.html*, February 2001.

[13] Stefan Saroiu, P. Krishna Gummadi, and Steven D. Gribble. *A Measurement Study of Peer-to-Peer File Sharing Systems.* In the Proceedings of the Multimedia Computing and Networking 2002 (MMCN'02), January 2002.

[14] FreeNet Web Site. http://www.freenet.com/, 2004.

[15] Kazaa Lite Web Site. http://www.k-lite.tk/, 2004.

[16] Kazaa Web Site. htpp://www.kazaa.com, 2004.

[17] Vivek Vishnumurthy, Sangeeth Chandrakumar, and Emin Gun Sirer. Karma: A secure economic framework for p2p resource sharing. *In Proceedings of the Workshop on the Economics of Peer-to-Peer Systems*, June 2003.