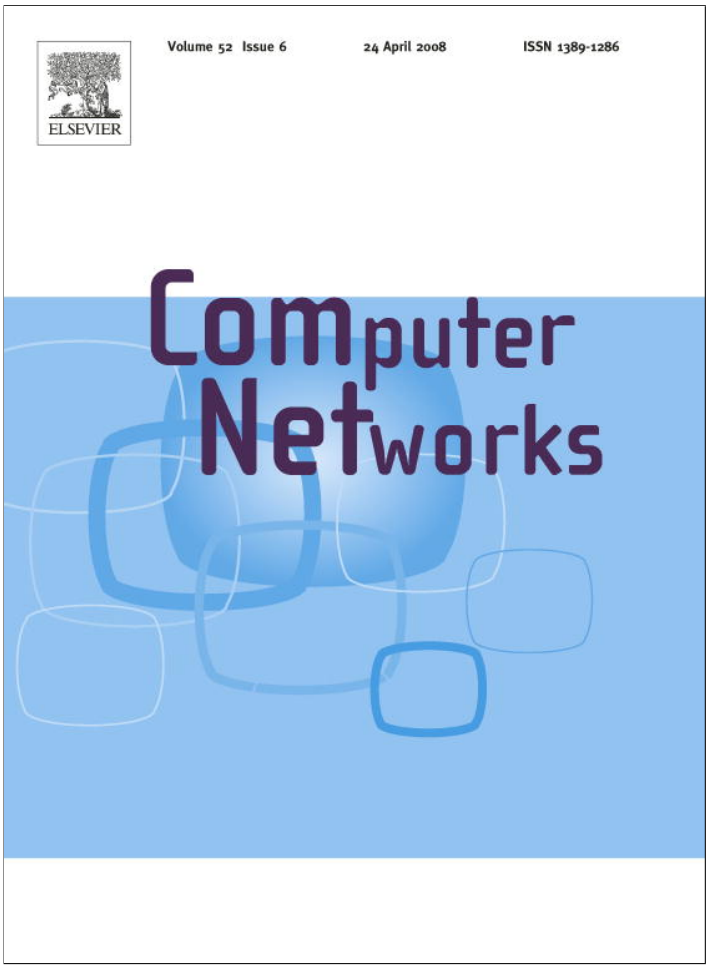

Available online at www.sciencedirect.com
ScienceDirect
Computer Networks 52 (2008) 1272–1280
Computer Networks
www.elsevier.com/locate/comnet

The P2P war: Someone is monitoring your activities

Anirban Banerjee *, Michalis Faloutsos, Laxmi Bhuyan

Department of Computer Science and Engineering, University of California, Riverside, CA 92507, United States

Received 2 January 2008; accepted 9 January 2008
Available online 1 February 2008

Responsible Editor: E. Ekici

Abstract

In an effort to legally prosecute P2P users, the RIAA and MPAA have reportedly started to create decoy users: they participate in P2P networks in order to identify illegal sharing of content. This has reportedly scared some users who are afraid of being caught and prosecuted. The question we would like to answer is how prevalent is this phenomenon: how likely is it that a user will run into such a "fake user" and thus run the risk of a lawsuit? The first challenge is identifying these "fake users". We collect this information from a number of free open-source software projects which are trying to identify such addresses by forming the, so-called, blocklists. The second challenge is to quantify the probability of a user contacting such a fake user by conducting a large scale experiment in order to obtain reliable statistics. Using PlanetLab, we conduct active measurements, spanning a period of 90 days, from January to March 2006, spread over three continents. Analyzing over 100 GB of TCP header data, we quantify the probability of a P2P user contacting fake users. We observe that 100% of our peers run into entities in these lists. In fact, 12–17% of all distinct IPs contacted by any node were listed on blocklists. Interestingly, a little caution can have significant effect: the top five most prevalent blocklisted IP ranges contribute to nearly 94% of all blocklisted IPs we ran into. Avoiding these can reduce the probability of a user being tracked to about 1%. In addition, we examine the identity of these blocklisted IPs. The majority of blocklisted IPs belong to the commercial and government domains and are nearly 2.5 times more than IPs belonging to educational, spyware or adware entities. Interestingly, less than 0.5% of all unique IPs contacted, belong explicitly to media companies. However, this may not be reassuring for P2P users, since the other blocklist users (government or commercial) could be collaborating with media companies.

Keywords: Peer-to-peer; Blocklist; RIAA; Monitoring

1. Introduction

Content providers, such as the RIAA and MPAA, have escalated their fight against illegal P2P sharing [3,14–16,22,23] with the use of fear: there have been a number of lawsuits against individual P2P users [4–7]. To make this more effective, these

* Corresponding author. Tel.: +1 9512310557.
E-mail addresses: anirban@cs.ucr.edu (A. Banerjee), michalis@cs.ucr.edu (M. Faloutsos), bhuyan@cs.ucr.edu (L. Bhuyan).
URLs: http://www.cs.ucr.edu/~anirban (A. Banerjee), http://www.cs.ucr.edu/~michalis (M. Faloutsos), http://www.cs.ucr.edu/~bhuyan (L. Bhuyan).

1389-1286/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.comnet.2008.01.011

organizations and their collaborators have started "trawling" in P2P networks: creating "fake users", which participate in the network, and thus, identify users who contribute towards illegal content sharing. However, the extent of this deployment tactic has not been quantified up to now, and this forms the focus of our work.

In response to this approach, the P2P community has spawned several projects which attempt to identify such "fake users", and enable P2P users to avoid them. In more detail, there is a community based effort to maintain lists of suspect IPs, which are called *blocklists*. Blocklists are published by organizations which provide anti-RIAA software or by groups which focus on security [10]. Additionally, a number of free, open-source, software projects have enabled P2P users to avoid these blocklisted IPs automatically. Such software is easy to download and is compatible with most popular P2P clients using various networks as BitTorrent, eDonkey–eMule, Gnutella [1,2,9,10,32,18,28]. Note that it is not our intention here to examine how accurate and comprehensive these lists are, though this would be interesting and challenging future work. Our claim is that, the information that we use in our work is what is readily available to P2P users. We present Fig. 1a and b which denote the significant numbers of P2P users who download and employ these blocklists to avoid contact themselves with fake users.

The question we attempt to answer is, how prevalent is the phenomenon of fake users. Simply put, how likely is it that a user without running any additional software will run into such a "fake user"? The answer to this question can help us: (a) understand the effort that content providers are putting in trawling P2P networks and (b) justify the effort of the P2P community to isolate "fake users". To the best of our knowledge, this phenomenon has not been quantified before.

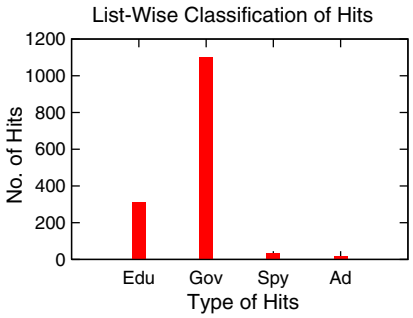

Fig. 2. Classification of blocklist hits according to their type. We observe that hits on the commercial and government blocklist is significantly larger than hits on the other blocklists.

We conduct an extensive measurement study, employing PlanetLab [13] for a period of 90 days. We analyze more than 100 GB of TCP header data, monitoring clients connected to the Gnutella network and use the most popular blocklists on the Internet [2,10,32]. We create and deploy P2P clients which insert about 100 popular song queries from well-known music charts [38,31,30] into the P2P network. Here onwards, we refer to IPs of fake users listed on these blocklists as blocklisted IPs and users exchanging data with them as *being tracked*. A blocklisted IP is said to be *hit* every time a user interacts with it. Our results can be summarized as follows:

1. *Pessimistic result*: A user without any knowledge of blocklists, will almost certainly be tracked by blocklisted IPs. We found that all our clients exchanged files with blocklisted IPs. In fact, of all distinct IPs contacted by any client, 12–17% were found to be listed on blocklists.

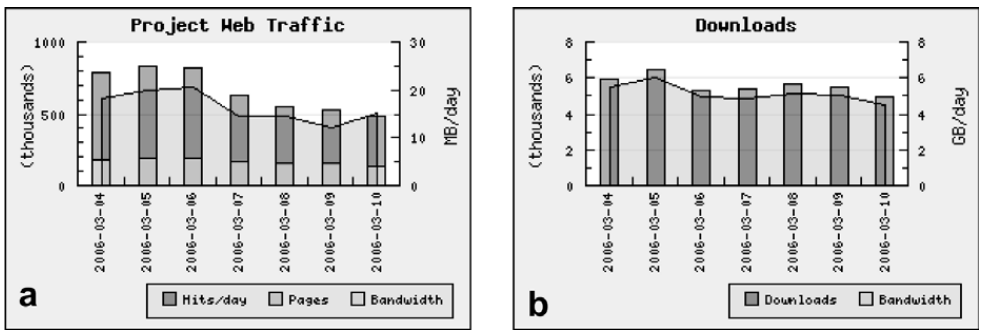

Fig. 1. Activity statistics for Peerguardian, compiled 10th March 2006: (a) The total webtraffic directed towards the Peerguardian webpage at Sourceforge. (b) The amount of downloads for the software, ranging from 4.5 to 6.5 GB (approx.) per day, from the same site [2].

2. *Optimistic results*: We find that *a little information goes a long way*: Avoiding just the top 5 blocklisted IPs reduces the chance of being tracked to about 1%. This is a consequence of a skewed preference distribution: we find that the top 5 blocklist ranges encountered during our experiments contribute to nearly 94% of all blocklist encounters.
3. *Most blocklisted IPs belong to government or corporate organizations*: We quantify the percentage of hits, to blocklisted IPs of each type, i.e. government and corporate, educational, spyware proliferators and Internet advertisement firms. We find that the number of hits which belong to government and corporate lists are nearly 2.5 times more than educational, spyware and adware lists. This is displayed in Fig. 2.
4. *Very few blocklisted IPs belong directly to content providers*: We find that 0.5% of all blocklisted IP hits could actually be traced back to media companies, such as Time Warner Inc.
5. *Geographical bias*: We find that there is a geographical bias associated with how users encounter entities listed on blocklists. Users located on the two opposite coasts, east and west, of mainland US, in Europe and Asia, hit blocklisted entities according to different patterns.
6. *Equal opportunity trawling*: We find that Ultrapeers[1] and leaf nodes have equal probability of associating with a blocklisted IP, with less than 5% variation in the average number of distinct blocklisted IP hits. This comes in contrast to the popular belief that UPs are tracked more aggressively by blocklisted entities [11,12], than leaf users.

The rest of the paper is organized as follows: Section 2 details the relevant literature applicable to our research, followed by Section 3, which discusses the experimental setup. This is followed up by Section 4, which investigates the probability of a user being tracked. Section 5 deals with unearthing geographical bias followed by Section 6 which addresses the Ultra/Super peer versus leaf node debate. This is followed by Section 7.

[1] Ultra-peers are high bandwidth nodes that act as local centers, facilitate low bandwidth *leaf* nodes, and enable the scalability of Gnutella-like networks.

2. Relevant literature

P2P networks are a prevalent application in the Internet. There exists a plethora of P2P networks, such as FastTrack, Gnutella [15], BitTorrent, eMule/eDonkey along with extremely an long list of clients, written in all possible languages for nearly all operating systems [14,16]. P2P networks have recently been touted as the future of content distribution technologies [17]. However, the fact remains that, these overlay networks, still do act as significant enablers in the dissemination of copyrighted material. Organizations such as the RIAA and MPAA have been extremely vociferous in their support for anti-P2P policies, since it is these organizations that lose out on revenue due to the exchange of copyrighted songs and movies [6,8,20,21].

Recently, a slew of reports in the electronic and printed media have led to members of P2P communities pondering over the ramifications of such illegal resource sharing [19]. To reduce the threat of a possible lawsuit, users have resorted to downloading and deploying anti-detection software. This software blocks computers owned by these organizations from communicating with P2P users [9,2]. This kind of software no longer allows entities monitoring P2P users to log the IPs of users. There is a large number of such free software, easily available, from popular websites, for many different P2P clients, networks and operating systems.

Previous work on modelling and analysis of P2P systems [24–27], has focused on developing a viewpoint based on performance metrics of such overlay systems. Our work differs greatly from these earlier efforts. Our goal is to quantify the probability of a P2P user of being tracked by entities listed on the most popular blocklists. To the best of our knowledge, our research is the first which specifically targets an in-depth study of whether such a threat is a reality for a generic P2P user. Moreover, our work is quantifies *who do we talk to* while connected on these P2P networks, when sharing copyright-material. Additionally, we intend to verify reports suggesting that some so-called organizations enlisted by the RIAA *target UPs in preference to leaf nodes* [11,12], in order to break the backbone of the entire overlay structure.

3. Who is watching?

In this section, we discuss the experimental setup we employ and quantify the most prevalent on blocklisted entities in P2P networks.

We find that:

1. The majority of the most active blocklisted entities encountered are hosted by organizations which want to remain anonymous.
2. Content providers such as the RIAA do not participate in large scale eavesdropping into P2P networks using their own IPs.

We initiate our experiments in a manner so as to be able to emulate the typical user and yet be able to measure large scale distributed network wide inter-node interaction characteristics of such P2P networks. We measure statistics based on trace logs compiled from connections initiated using PlanetLab to gather traces in a geographically distributed environment. The duration of measurements spanned more than 90 days, beginning January 2006. We initiate connections using 50 nodes spread not only across the continental US (35 nodes), but also Europe (10 nodes), and Asia (5 nodes) in order to determine any geographical nuances associated with, which entities on blocklists seems to be more active than others, in specific locations. We customized mutella 0.4.5 clients [29], a vanilla console based Gnutella client, and initiate connections to the Gnutella network. Moreover, clients were made to switch interchangeably from UP to leaf modes in order to verify if network wide inter-node behavior of UPs is significantly different from leaf nodes.

Our queries in the P2P network were based on lists of popular songs, from Billboards hot 100 hits [30], top European 50 hits [31] and Asian hits [38]. Each node injected about 100 queries during every run. In the process, we analyzed more than 100 GB of TCP header traces, using custom scripts and filters to extract relevant information which helps us develop a deeper insight into who do we interact with, while sharing resources on P2P networks.

Before we present results obtained from our measurements we must discuss what BOGON IPs [36] mean as they hold special significance to the collected information. BOGON is the name used to describe IP blocks not allocated by IANA and RIRs to ISPs and organizations plus all other IP blocks that are reserved for private or special use by RFCs. As these IP blocks are not allocated or specially reserved, such IP blocks should not be routable and used on the Internet, however, some of these IP blocks do appear on the net primarily used by those individuals and organizations that are often specifically trying to avoid being identified and are often involved in such activities as DoS attacks, email abuse, hacking and other security problems.

Table 1 lists the top fifteen entities that we encounter on the P2P network while exchanging resources, throughout the duration of our active trace collection. Surprisingly, these entities operate from BOGON IP ranges. This observation is made on the basis of the various popular blocklist resources, and suggests that *these sources deliberately wish to conceal their identities while serving files on P2P networks*, by using up IP ranges which cannot be tracked down using an IP-WHOIS lookup to locate the operator employing these anonymous blocks. Only three out of the top fifteen entries in Table 1 do not use unallocated BOGON IP blocks and are listed on PG lists [2], the rest of the BOGON entities, are listed on both Trustyfiles [32] and Bluetack [10] lists. Most of the BOGON IP ranges point to either ARIN or RIPE IP ranges. We must, however, mention that these BOGON IP ranges were found to point back to these generic network address distribution entities at the time of our experiments. It is quite possible that these ranges may have now been allocated to firms or individuals and may no longer remain truly anonymous. We observe that 99.5% of blocklisted IPs contacted either belong to BOGON, commercial entities, educational institutions while only about 0.5% of all blocklisted IPs we came in contact with could actually be traced back to record companies, such as Time Warner Inc. This is an indication of the small presence of record companies themselves, snooping on P2P users in a proactive manner.

Table 1
Listing of top 15 blocklist entities encountered on P2P network

Rank	Top 15 hit ranges	Type
1	72.48.128.0–72.235.255.255	Bogon
2	87.0.0.0–87.31.255.255	Bogon
3	88.0.0.0–88.191.255.255	Bogon
4	72.35.224.0–72.35.239.255	FuzionColo
5	71.138.0.0–71.207.255.255	Bogon
6	70.229.0.0–70.239.255.255	Bogon
7	70.159.0.0–70.167.255.255	Bogon
8	70.118.192.0–70.127.255.255	Bogon
9	216.152.240.0–216.152.255.255	Xeex
10	216.151.128.0–216.151.159.255	Xeex
11	70.130.0.0–70.143.255.255	Bogon
12	87.88.0.0–87.127.255.255	Bogon
13	71.66.0.0–71.79.255.255	Bogon
14	87.160.0.0–87.255.255.255	Bogon
15	70.82.0.0–70.83.255.255	Bogon

FuzionColo listed in Table 1, is understood to propagate self installing malware, and in general as an anti P2P entity [33,34,37]. xeex [35] is more of a mystery. It hosts an inconspicious site which provides absolutely no information as to what the company is really involved in. Going by the discussion groups hosted on the PG website, xeex does turn up frequently in blocklist hits for a large number of users. Other individuals or organizations deliberately employing BOGON IPs to participate in the exchange of resources on P2P networks are certainly attempting to hide, possibly from the RIAA. Another vein of reasoning would suggest that they could be the ones who keep track of what users download.

Tables 2 and 3 display the top five entities on the (a) educational and research institutions list and the (b) government and commercial organizations lists. We observe that FuzionColo and xeex appear prominently in this categorization along with two other commercial organizations which host servers on ed2k and Gnutella networks. We find that hits to entities listed on commercial and government blocklists are much more frequent than hits on any other different kind of blocklists such as Internet ad companies, educational institutions and others. Even though the number of IPs which belong explicitly to content providers such as the RIAA may be small, the fact that IPs listed on commercial and government blocklists are providing content to P2P users is of concern. The scenario in which commercial organizations are hired by content providers to collect user profile data cannot be ruled out. Furthermore, the possibility that these commercial organizations, such as the ones listed in Table 3 are not aware of P2P traffic emanating from their servers does not seem very plausible since some of these blocklisted entities kept tracking our clients nearly every time files were exchanged. It is clear that these commercial IP ranges, which serve files to P2P users, have a very large cache of popular in-demand media and have extremely low downtime. In fact, the number of hits to commercial and government blocklisted entities is nearly 2.5 times greater than hits to any other kind of blocklisted IPs.

Table 2
Listing of top 5 educational entities encountered on P2P networks

Rank	Top 5 educational hit ranges
1	152.2.0.0–152.2.255.255 – University of North Carolina
2	64.247.64.0–64.247.127.255 – Ohio University
3	129.93.0.0–129.93.255.255 – University of Nebraska
4	128.61.0.0–128.62.255.255 – Georgia Tech
5	219.242.0.0–219.243.255.255 – CERNET

Table 3
Listing of top 5 commercial entities encountered on P2P networks

Rank	Top 5 commercial hit ranges
1	72.35.224.0–72.35.239.255 – FuzionColo
2	216.152.240.0–216.152.255.255 – xeex
3	216.151.128.0–216.151.159.255 – xeex
4	38.113.0.0–38.113.255.255 – Perf.SystemsInted2k
5	66.172.60.0–66.172.60.255 – Netsentryed2kserver

4. Probability of being tracked

In this section, we estimate the probability of a typical P2P user being tracked by entities listed on these blocklists. This gives an idea of what percentage of entities encountered while surfing P2P networks are not considered trustworthy. We observe the following during our study:

1. 100% of all our nodes were tracked by entities on blocklists and on average, 12–17% of all distinct IPs contacted by any of our clients were listed on blocklists.
2. Popularity of blocklisted IPs tracking P2P users follows an extremely skewed distribution.

As illustrated in Fig. 3, the percentage of IPs listed on blocklists is quite significant, about 12–17% of all distinct IPs contacted, per node. In fact, this trend was reflected throughout the duration of our measurements, which suggests that the presence of blocklisted entities on P2P networks is not an ephemeral phenomenon. Furthermore, we observe that the frequency of popularity for

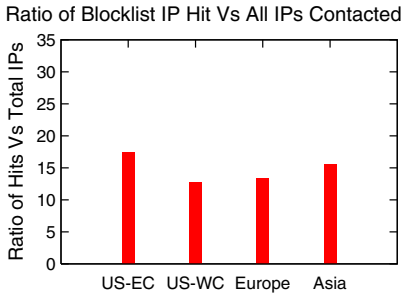

Fig. 3. Percentage of distinct blocklist IPs contacted out of the total number of distinct IPs logged.

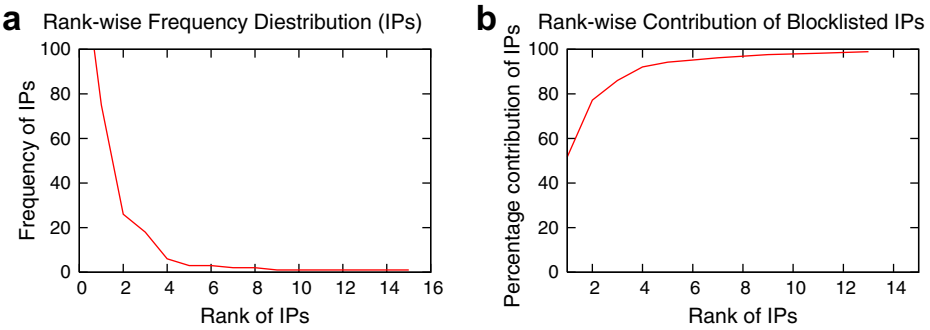

Fig. 4. (a) Frequency of popularity of blocklisted IPs, following a skewed distribution. (b) Percentage contribution by blocklisted IPs. The five most popular blocklisted IPs contribute to nearly 94.2% of all blocklist hits.

blocklisted entities follows a skewed distribution as displayed in Fig. 4a. A small number of entities register a large number of hits while most blocklisted entities are infrequently visible on P2P networks. This fact is of great consequence to users who wish to avoid contact with blocklisted entities and thus reduce their chances of running into anti-P2P entities. Avoiding the five most popular blocklisted IPs leads to a drastic reduction in the number of hits to blocklisted IPs, approximately by 94%. This interesting statistic is displayed in Fig. 4b. In fact, avoiding just these top 5 blocklisted IPs can reduce the chances of a user being tracked significantly, down to nearly 1%. Users can use this fact to tweak their IP filters to increase their chances of safely surfing P2P networks and bypassing the most prevalent blocklisted entities. This is critical considering that, a naive user, without any information of blocklists will almost certainly be tracked by blocklisted entities. Also, the fact that 100% of all nodes regardless of geographical location were tracked by blocklisted IPs, indirectly points to the completeness of the blocklists we collected from the most popular sources.

5. Geographical distribution

In this section, we focus attention towards the issue regarding whether geographical bias exists in our active measurements with respect to entries on blocklists being encountered while our clients connect to the P2P networks from various geographical locations. To achieve this, we needed a geographically diverse set of P2P users. We employed over 50 different nodes on PlanetLab, encompassing the continental US, Europe and Asia. We monitor individually, PlanetLab nodes located in the continental US as nodes situated on the east coast (US-EC) and on the west coast (US-WC), to observe if there is any variation in behavior within mainland US and, surprisingly, we do observe such a difference as discussed below.

In Fig. 5a, we study the effect of geographical location on how blocklisted IPs track P2P users. We observe that the percentage of blocklisted IP hits is highest in US-WC followed by US-EC, Asia and Europe. The percentage of blocklisted IP hits, per node, as a percentage of total hits to IPs contacted by each node, located on the US west coast

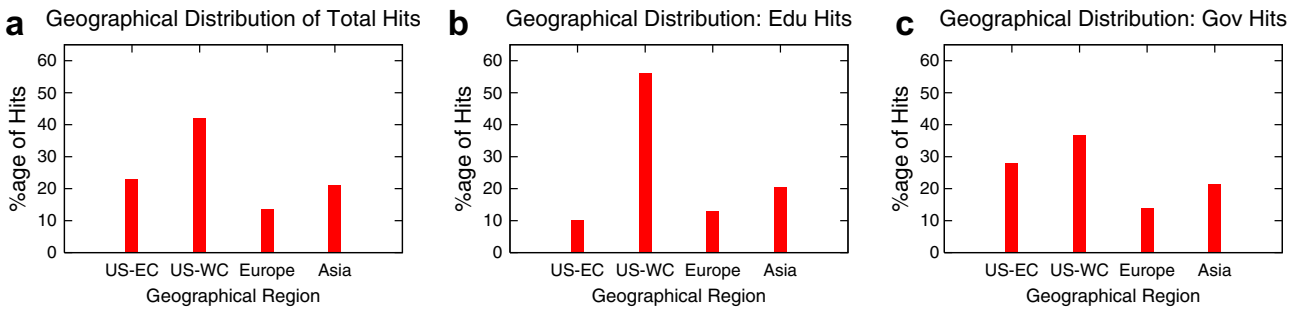

Fig. 5. (a) Distribution of blocklisted IPs contacted in different geographical zones; (b) distribution of blocklisted IPs contacted, on educational lists, in different geographical zones; and (c) distribution of blocklisted IPs contacted, on government and commercial lists, in different geographical zones.

seems to be nearly twice of that for nodes located on US east coast. This suggests that users accessing the P2P network from these two vantage points, within the mainland US, encounter different levels of tracking activity. We believe this observed inequality stems from the difference in user behavior and possible difference in levels of monitoring activities by entities on the blocklists could directly be responsible for such a skewed trend. Fig. 5b depicts the distribution of blocklisted IP hits from the "educational" range, comprising of academic and research institutions. Again, we observe a similar trend, nodes located on US-WC have a up a higher percentage of blocklist hits compared to nodes located on US-EC, Asia and Europe. In fact, the difference in measurements between US-WC and US-EC is more than five times than that of readings gathered from US-EC.

Fig. 5c depicts the distribution of blocklisted IP hits in the government and commercial domain. Once again, we observe that the probabilities for nodes situated on US-WC are higher than nodes on US-EC, Asia and Europe. The period of observation, the UTC time when data was logged, the number of queries in the P2P network, the order in which queries were injected were identical for all nodes. This suggests that, throughout the duration of our experiments, a consistent skewed distribution between US west coast and US east coast can be due to difference in user behavior and the differing degree of local tracking activity in these different geographical settings. Nodes located in Europe consistently registered a lower number of blocklisted IP hits when compared to nodes located in Asia. We always attempt to maintain a balance while logging data using PlanetLab and deploy our code on nearly the same number of nodes in different geographical settings, log data during synchronized time periods using automated scripts bootstrapped via crontab. The only difference while gathering measurements in these settings was that we used different lists of queries which were injected into the P2P network for nodes located in separate continents. For nodes located in Europe, we constructed query lists based on European 50 hits [31], and for nodes in Asia we constructed query lists based on Asian hits [38]. The magnitude of difference observed between nodes in Europe and Asia was found to be more or less consistent across the different types of blocklisted IPs. However, they were significantly different from measurements gathered across the mainland US.

6. Effect of role on the probability of being tracked

This section delves into whether, according to popular perception in P2P communities [11,12], the probability of being tracked by blocklisted entities varies with the role played by a P2P node. The question we answer is: *are UPs are tracked with higher probability by entities on blocklists versus regular leaf nodes*. We find that the role of the node does not seem to have an effect on its probability of being tracked by blocklisted IPs. To examine this, we repeatedly configured nodes to shift from UP to leaf mode and back over a number of cycles in order to obtain connectivity patterns for each mode of operation. The uptime for each mode was identical and experiments were repeated to smoothen out any temporal fluctuations in observed data. We observe in Fig. 6a the mean number of distinct IPs contacted by leaf nodes and UPs in various geographical locations. We find that leaf nodes, located in the US, seem to interact with a larger number of distinct IPs than do UPs. However, this is not the case in either Europe or Asia, where UPs connect to larger number of distinct IPs than leaves. This observation could be due to the false perception, hyped primarily in the US that UPs are being watched with more vigor by entities on the blocklists compared to leaf users. Since significant legal action against users of P2P networks has been directed towards users in the US, it is obvious that peers would refrain from voluntarily switching their P2P client's mode of operation to become a UP. Therefore, we see a much lesser intensity of UP interaction within P2P networks in the US. While in Asia, where the threat of legal action has yet to create a dent in the minds of P2P users, it is evident that users will hardly shy away from switching clients to UP mode or at least deliberately prevent clients from doing so. Hence, we observe much larger numbers of peers communicating with UPs in Asia. We believe that the same vein of thought holds true for the scenario for Europe based nodes, albeit to a lesser extent.

Ultra-peers do not encounter more blocklisted entities than leaf nodes in a consistent manner. In Fig. 6b, we compare the percentage of blocklisted IP hits as recorded by UPs and leaf nodes. The percentage shows how many of the total number of IPs encountered are blocklisted IPs. This metric depicts if there is any correlation between UPs being tracked preferentially over leaf nodes irrespective of geographical location. We find that UPs in

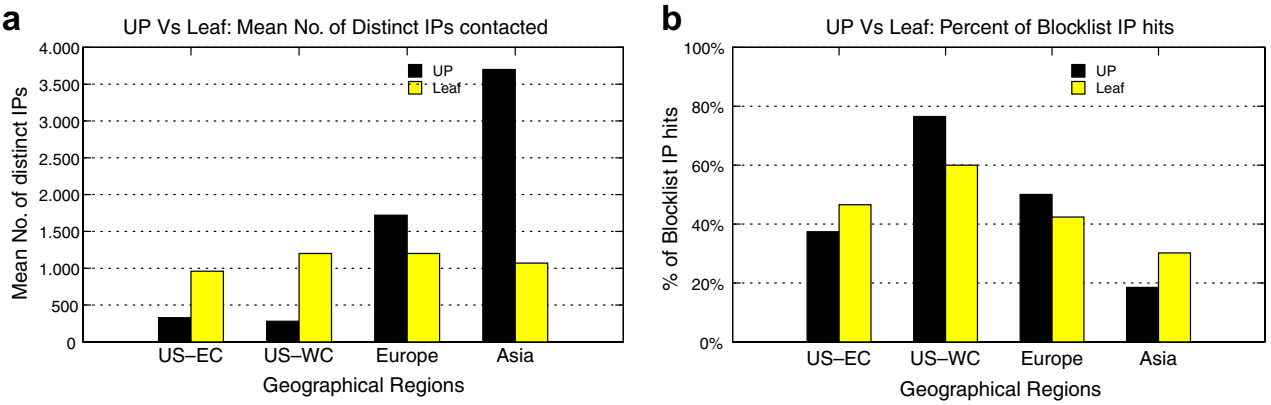

Fig. 6. UP vs. leaf: The black bar signifies UPs while the yellow bar signifies leaf users: (a) comparison of average number of distinct IPs contacted by UPs and leaves and (b) comparison of percentages of blocklisted IP hits as encountered by UPs and leaf users.

US-WC encounter higher numbers of blocklisted IP ranges versus leaf nodes. This trend is consistent with Europe based nodes. However, for US-EC and Asia based nodes we observe that UPs encounter lesser numbers of blocklist IPs compared to leaf nodes. In fact we observe less than 5% variation in the average number of blocklisted IP hits by UPs versus leaf nodes on these P2P networks and thereby do not find any supporting evidence for claims of UPs being preferentially tracked by entities on these lists vis-a-vis leaf nodes. From our experiments we understand that a UP has nearly the same probability of running into blocklisted entities as leaf users and do not find any significant variation in the number of blocklisted entities contacted by either. It must be noted though that our experiments do suggest a difference in P2P user behavior between US-WC and US-EC as has been discussed in previous sections.

7. Conclusion

To the best of our knowledge, this work is the first to quantify the probability that a user will be tracked by blocklisted IPs, and thus, potentially run the risk of a lawsuit. Using PlanetLab, we conduct large-scale active measurements, spanning a period of 90 days, from January to March 2006, spread over three continents, yielding over a 100 GB of TCP packet header data. We find that a naive user is practically guaranteed to be contact blocklisted IPs: we observe that 100% of our peers run into blocklisted users. In fact, 12–17% of all distinct IPs contacted by a peer are blocklisted IPs. Interestingly, a little caution can have significant effect: the top five most prevalent blocklisted IPs contribute to nearly 94% of all blocklisted IPs we ran into. Using this information users can reduce their chances of being tracked to just about 1%. At the same time, we examine various different dimensions of the users such as the geographical location and the role of the node in the network. We find that the geographical location, unlike the role, seems to affect the probability of encountering blocklisted users. Finally we examine, who are the blocklisted IP addresses. Interestingly, we find that 0.5% of all distinct IPs belong explicitly to media companies. The major of the blocklisted users seem to belong to commercial and government organizations and a sizeable portion of the most popular belong to anonymous BOGON ranges.

Our work is the first step in monitoring the new phase of "wars" between the content providers and the P2P community. It will be very interesting to continue to monitor the evolution of this conflict. For example, one could analyze the accuracy and completeness of the blocklists, and the speed with which a new blocklisted entity is flagged.

References

[1] <http://www.sourceforge.net>.
[2] <http://peerguardian.sourceforge.net>.
[3] E.K. Lua, J. Crowcroft, M. Pias, R. Sharma, S. Lim, A survey and comparison of peer-to-peer overlay network schemes, IEEE Commun. Survey and Tutorial, March 2004.
[4] <http://news.dmusic.com/article/7509>.
[5] <http://www.betanews.com/article/MPAASuesUsenetTorrentSearchSites>.
[6] <http://importance.corante.com/archives/005003.html>.
[7] <http://www.mp3newswire.net/stories/napster.html>.
[8] <http://news.com.com/2100-1027-995429.html>.
[9] <http://sourceforge.net/projects/peerprotect>.
[10] <http://bluetack.co.uk/blc.php>.
[11] <http://www.boycott-riaa.com/article/9316>.
[12] <http://slashdot.org/articles/02/05/25/0324248.shtml>.

[13] <http://www.planet-lab.org>.
[14] T. Karagiannis, A.Broido, M. Faloutsos, K.C. Claffy, Transport layer identification of P2P traffic, in: ACM Sigcomm IMC'04, 2004.
[15] E. Markatos, Tracing a large-scale peer-to-peer system: an hour in the life of Gnutella, in: 2nd IEEE/ACM International Symposium on Cluster Computing and the Grid, 2002.
[16] S. Sen, J. Wang, Analyzing peer-to-peer traffic across large networks, in: ACM SIGCOMM IMW, 2002.
[17] Thomas Karagiannis, Pablo Rodriguez, Dina Papagiannaki, Should Internet service providers fear peer-assisted content distribution? in: IMC'05, Berkeley, 2005.
[18] Kurt Tutschku, A measurement-based traffic profile of the edonkey filesharing service, in: PAM'04, Antibes Juan-les-Pins, France, 2004.
[19] <http://www.techspot.com/news/16394-record-labels-launch-legal-action-against-kazaa.html>.
[20] <http://www.mpaa.org/CurrentReleases/2004_12_14_Wwde-P2PActions.pdf>.
[21] Valerie Alter, Building Rome in a Day: What Should We Expect from the RIAA? 56 HASTINGS COMM. & ENT. L.J. 155.
[22] Jane Black, The Keys to Ending Music Piracy, BUS, WK, January 27, 2003. <http://www.businessweek.com/bwdaily/dnflash/jan2003/>.
[23] RIAA Gives Advance Warning to Song-Swappers Before Lawsuits are Filed, 2003. <http://www.antimusic.com/news/03/oct/item77.shtml>.
[24] Thomas Karagiannis, Andre Broido, Nevil Brownlee, K.C. Claffy, Michalis Faloutsos, Is P2P dying or just hiding, IEEE Globecom, 2004.
[25] Krishna P. Gummadi, Richard J. Dunn, Stefan Saroiu, Steven D. Gribble, Henry M. Levy, John Zahorjan, Measurement, modeling, and analysis of a peer-to-peer file-sharing workload, in: Proceedings of the SOSP-19, October 2003.
[26] J. Chu, K. Labonte, B.N. Levine, Availability and locality measurements of peer-to-peer file systems, in: Proceedings of the ITCom'02.
[27] F. Clvenot-Perronnin, P. Nain, Stochastic fluid model for P2P caching evaluation, in: Proceedings of the IEEE WCW, 2005.
[28] <http://azureus.sourceforge.net/plugin_details.php_plugin_safepeer>.
[29] <http://mutella.sourceforge.net/>.
[30] <http://www.billboard.com/bbcom/charts/chart_display.jsp?fThe_Billboard_Hot_100>.
[31] <http://www.mp3hits.com/charts/euro>.
[32] <http://www.trustyfiles.com>.
[33] <http://isc.sans.org/diary.php?date=2005-04-11>.
[34] <http://www.winmxworld.com/tutorials/block_the_RIAA.html>.
[35] <http://xeex.com>.
[36] <http://www.completewhois.com/bogons/index.htm>.
[37] <http://phoenixlabs.org>.
[38] <http://www.mtvasia/Onair>.

Anirban Banerjee is a graduate student with the Department of Computer Science and Engineering at UC Riverside since 2004. His interests encompass content distribution via P2P networks and web security.

Michalis Faloutsos is a faculty member at the Computer Science Department in the University of California, Riverside. He got his bachelor's degree at the National Technical University of Athens and his M.Sc. and Ph.D. at the University of Toronto. His interests include, Internet protocols and measurements, multicasting, cellular and ad-hoc networks. With his two brothers, he co-authored the paper on powerlaws of the Internet topology (SIGCOMM'99), which is in the top 15 most cited papers of 1999. His work has been supported by several NSF and DAPRA grants, including the prestigious NSF CAREER award. He is actively involved in the community as a reviewer and a TPC member in many conferences and journals.

Laxmi Bhuyan has been a professor of computer science and engineering at the University of California, Riverside since January 2001. Prior to that he was a professor of computer science at Texas A&M University (1989–2000) and program director of the Computer System Architecture Program at the US National Science Foundation (1998–2000). He has also worked as a consultant to Intel and HP Labs. His current research interests are in the areas of network processor architecture, Internet routers, and parallel and distributed processing. He has published more than 100 papers in related areas in reputable journals and conference proceedings. His brief biography and recent publications can be found on his Web page at http://www.cs.ucr.edu/~bhuyan/. He is a fellow of the IEEE, a fellow of the ACM, and a fellow of the AAAS.