Is Deviant Behaviour the Norm on P2P File-Sharing Networks?

Daniel Hughes[1], Stephen Gibson[2], James Walkerdine[1], Geoff Coulson[1]

[1]*Computing Department,* [2]*Psychology Department, Lancaster University, Lancaster, UK*

{ hughesdr, walkerdi, geoff } @comp.lancs.ac.uk | s.j.gibson1@lancaster.ac.uk

Abstract

Major international law-enforcement initiatives are currently underway to fight the distribution of illegal pornography over the Internet. It is commonly believed that peer-to-peer (P2P) file-sharing networks are particularly implicated in this area—largely because of the anonymity that they afford their users. This anonymity is thought to encourage individuals to engage in trafficking illegal pornography because it weakens social pressures that would otherwise inhibit such behaviour. In this paper we present empirical evidence suggesting that this conventional wisdom may be flawed. Extrapolating from our analysis, we argue that the vast majority of the P2P-mediated illegal pornography is produced by a small subset of the P2P community who have little or no interaction with the wider law-abiding community.

1. Introduction

Since the release of Napster [1] in 1999, peer-to-peer (P2P) file-sharing communities have been growing in both number and scale. Today, several deployed P2P file-sharing networks (e.g. [2], [3], [4]) boast users numbering in the millions. Due to both scalability concerns and legal issues, today's P2P networks have moved away from the semi-centralized approach typified by Napster towards more scalable and anonymous P2P architectures [5]. These fully decentralized networks, because they exist in the absence of any central authority, provide a new and interesting context for the expression of human social behaviour.

However, the activities of the individuals who compose P2P communities are sometimes at odds with what is considered acceptable by authorities in the 'real world'. The most obvious example of this is the phenomenon of copyright infringement, and the resulting copyright enforcement activities by organizations such as the Recording Industry Association of America (RIAA) [6]. This issue has generated lively debate in the P2P research community, and has given rise to a significant body of work devoted, on the one hand, to policing user behavior [7] and, on the other hand, to creating radically anonymous networks in which the enforcement of any form of control is impossible [8]. Copyright infringement has also exercised the legal world in the shape of recent challenges to the legality of P2P file-sharing systems [9]. More broadly, fundamental questions have been raised about the nature of copyright law and its enforcement, which, despite vigorous debate, have yet to be resolved.

A more clear-cut example of online activity which society finds unacceptable is the use of P2P networks to distribute illegal pornography. Major law-enforcement efforts are currently underway in both Europe and North America [10] that target the distributors of such material; and these have resulted in a number of high-profile prosecutions. In addition, there have been large-scale public awareness campaigns [11] regarding the dangers that the Internet poses to children. Furthermore, a recent attempt by the California legislature to outlaw P2P file-sharing listed amongst its justifications the sharing of illegal pornographic material [9].

To gauge the form and extent of P2P-based sharing of illegal pornographic material, we conducted an extensive analysis of pornography-related resource-discovery traffic in the Gnutella P2P network [12]. The results of this analysis are given in this paper. We chose Gnutella because it is a good example of a large-scale, decentralized, anonymous, P2P file-sharing system; and it also has a well-studied user-base and an open protocol specification. Specifically, Gnutella was chosen over the Fastrack [2] and eDonkey [3] networks as the latter feature username and password authentication and therefore cannot be considered anonymous. However, while our experiments only address Gnutella, there is no reason to suppose that Gnutella users download any more or less pornography than users of other anonymous P2P networks; therefore our results may be considered indicative of what one might expect elsewhere.

From the results of our analysis, we indeed find that a small yet significant proportion of Gnutella activity relates to illegal pornography: for example, 1.6% of searches and 2.4% of responses are for this type of material. But does this imply that

such activity is widespread in the file-sharing population? On the contrary, we also find that this activity is carried out by a small, yet particularly active sub-community of users, and that searching for and distributing illegal pornography is *not* a behavioral norm.

The remainder of this paper is structured as follows: Section 2 provides necessary background on the operation of the Gnutella network. Section 3 provides background on the known socio-psychological effects of anonymity on online behaviour. Section 4 then describes our experimental set-up and the results of our experimental analysis. Following this, section 5 discusses the implications of our findings and section 6 suggests directions for future work.

2. The Gnutella Network

Gnutella is an open protocol designed to support the discovery and transfer of files among its users. Gnutella and similar decentralized file-sharing systems are considered to be more anonymous than earlier semi-centralized systems such as Napster [1] as these earlier systems made use of 3rd party indexing servers to store information about each peer and the files it was making available to the network. In entirely decentralized networks like Gnutella, no such entity has knowledge of the peers or files available on the network.

In technical terms, the Gnutella protocol builds an unstructured, decentralized application-level network [5]. As with any decentralized P2P network, participating peers are required to forward network maintenance and file discovery messages, and to share files on the network. The protocol itself is very simple and uses just five message types as shown in Table 1.

Table 1 – The Gnutella protocol messages

Message	Description
PING	Flooded onto the network to discover new peers. A peer that is willing to accept a new incoming connection will respond with a PONG.
PONG	The response to a PING. Contains connection information and data regarding the number and size of files this peer is sharing.
QUERY	A search message with a plain-text payload. If a peer receiving a QUERY has matching data, it generates a QUERYHIT.
QUERYHIT	A response to a QUERY. Contains the information needed to acquire the requested data.
PUSH	A mechanism to support downloads from firewalled peers. If both peers are behind a firewall, no file transfer is possible.

Having connected to the Gnutella network using PING and PONG, a peer's subsequent activities fall into two distinct phases: *i*) discovering resources, and *ii*) transferring resources (files).

To discover resources, a requesting peer forwards a QUERY message to its neighbors; each neighbor then forwards this message to some of *its* neighbors, and so on; thus 'flooding' the message onto the network. If a peer is able to satisfy an incoming QUERY (i.e. it is sharing a file which matches a search-term contained in the QUERY) it responds by sending a QUERYHIT message back along the same path. QUERYHITs contain the information required to subsequently acquire the requested file: i.e., the network address and port of the responding peer.

Having received one or more appropriate QUERYHITs, the requester selects a suitable peer, opens an HTTP connection to it, and dowloads the target file directly using HTTP GET. Thus file transfer itself takes place outside of Gnutella proper.

3. The Effects of Anonymity on Online Behaviour

A significant amount of research has been devoted to the effects of anonymity and perceived unidentifiability in 'computer mediated communication' (CMC) [13]. Some researchers have argued that anonymity results in a general increase in the likelihood of engaging in 'deviant' or 'disinhibited' online behaviour [14]. For example, Demetriou & Silke [15] found that well over half of individuals visiting a website for non-pornographic material nevertheless attempted to access pornography

when presented with the opportunity to do so. Such findings suggest that when online people may find it harder to insulate themselves from the temptation to engage in behaviours that may ordinarily incur strong social disapproval or sanction.

Other researchers, however, have suggested that the consequences of anonymity in CMC may be better understood in terms of group-specific social norms [16]. According to this view, anonymity will only lead to deviant or illegal behaviour (as defined by general societal norms) if the norms associated with a particular group identity allow for it. Thus, an individual will only be more likely to engage in behaviour that runs *counter* to general social norms when anonymous online if that behaviour *conforms* to group-specific social norms.

These two approaches offer competing predictions for the likely distribution patterns of illegal pornographic material on Gnutella. According to the first approach (which emphasises the generally negative effects of anonymity), no clear pattern should be detectable in the way users search for and serve deviant material. Users who share such material will simply be acting in an individually disinhibited way and any user of Gnutella is therefore a potential provider and user of such material. In contrast, according to the second approach (which draws attention to the importance of group-specific social norms), there should be a pattern detectable in the behaviour of those searching for and serving deviant material. Specifically, such users would be expected to form a distinct 'sub-class' within the wider class of Gnutella users.

For example, the effects of online anonymity on the behaviour of someone with a sexual interest in children would be to facilitate their downloading of images of childhood sexual abuse to the extent that such behaviour is normative for someone with such a sexual preference. Conversely, anonymity would *not* be expected to produce such behaviour in someone who does not identify themselves as having such a sexual preference. The crucial distinction is that whereas in the former approach the simple act of using Gnutella (or the Internet in general) is assumed to leave an individual potentially at risk of engaging in deviant behaviour, in the latter approach the anonymity afforded by online activity merely facilitates behaviour associated with group norms which are more-or-less already inscribed.

If it is true that the anonymity in P2P networks has generally negative affects upon user behaviour, then this would lend support to those who argue for wide-ranging legal restrictions on P2P technology, as was argued in recent legal action in California [9]. If, however, this behaviour is due to the influence of pre-inscribed group norms, it may be that the sharing of illegal sexual material on P2P file sharing networks merely reflects deeper issues in society and that more subtle approaches to discouraging such behaviour are required.

4. The Experiments

Our experimental work was based on intercepting and analysing QUERY and QUERYHIT messages on the Gnutella network. Essentially, analysing QUERY messages tells us what people are searching for, and analysing QUERYHIT messages tells us what people are offering to provide.

As each peer in Gnutella participates in routing all network messages, we can intercept these messages simply by deploying a modified peer onto the network which logs all the QUERY and QUERYHIT messages it routes. Using such a peer (based on the Jtella classes [17]), we monitored Gnutella traffic over a one month period between February 27th and March 27th 2005. We maximized the typicality of our sample base by connecting to the network as an ultra-peer [19], by maintaining a large number of incoming and outgoing connections, and by periodically re-connecting to different areas of the network.

The legality of various types of pornographic materials varies from country to country, and even within countries there are disagreements over the precise letter of the law [20]. Therefore, for the purposes of the present study we limited our definition of illegality to those materials depicting *practices* which are clearly illegal under UK and international law. These practices are rape, incest, bestiality and the sexual abuse of children.

Our experiments addressed two questions which are explored in sections 4.1 and 4.2 respectively:

1. What proportion of Gnutella traffic relates to illegal pornography?
2. Is this the activity of a deviant sub-community?

4.1 What proportion of Gnutella traffic relates to illegal pornography?

To answer this question we examined samples from three Saturdays that fell within our monitoring period (the Saturdays of March 5th, 12th and 19th). We selected Saturdays due to the relatively higher level of traffic observed during weekends. From

these samples, we randomly extracted 10,000 QUERYs and 10,000 QUERYHITs, and then manually classified these as relating to either illegal pornographic material or to 'other' material. The samples used in this classification, along with other raw data, are available online at http://polo.lancs.ac.uk/p2p/deviant.

To ensure accurate classification, the messages were independently classified, according to the criteria given above, by two independent reviewers. Our approach was to classify messages as relating to illegal pornographic material if they could *only* be interpreted as referring to such material. Despite this conservative approach, some level of misclassification was inevitable due to the nature of plain-text searches. For example, while it is possible that a query for '*young girl*' may be intended to retrieve questionable material, it could also refer to legal material (for example a song) and therefore such queries were not selected. Furthermore, due to the illegal nature of the content being shared, it is possible that individuals may use code words in order to avoid detection. This seems particularly likely if those sharing such material do indeed form a distinct sub-community.

The results of the classification are given in Tables 1 and 2. It can be seen that there is a high degree of correlation between the independent reviewers' classifications (p=0.3 for QUERYs; and p=0.8 for QUERYHITs).

Table 1: Traffic relating to illegal material: reviewer 1

	5th March		12th March		19th March	
Queries	1.2%	122	1.6%	156	1.6%	158
QueryHit	2.1%	206	3.0%	295	2.0%	195

Table 2: Traffic relating to illegal material: reviewer 2

	5th March		12th March		19th March	
Queries	1.4%	142	1.8%	184	1.7%	174
QueryHit	2.3%	234	3.0%	297	2.1%	208

The results were as follows:

- An average of 1.6% of QUERY messages were classified as relating to illegal pornography. The minimum value we observed was 1.2% on March 5th, rising to 1.8% on March 12th. The standard deviation between samples was 0.2%.

- An average of 2.4% of QUERYHIT messages relate to illegal pornography. The minimum value observed was 2% on March 19th, rising to 3% on March 12th. The standard deviation was 0.7%.

The disparity between the numbers of QUERY and QUERYHIT messages is primarily due to the fact that QUERYHIT messages refer to multiple files which may have matched a single received QUERY.

4.2 Is this activity the result of a deviant sub-community?

To assess whether or not individuals who share illegal pornography form a sub-community within the wider Gnutella community, we first produced a ranked list of the top 20 pornography-related search terms. From this list, we identified peers who responded with QUERYHITs on our selected dates (the Saturdays within our sampling period). This yielded a list of peers which could reasonably be assumed to be distributing illegal material.

We selected 100 unique hosts at random from the above set and determined whether or not they participated in sharing other material. Figure 2 shows the proportion of illegal material that was being served by these 100 peers over the one month period. As can be seen, the majority of peers that share illegal pornography (57%) *share no other material whatsoever*, while only 17% share less than 50% of illegal materials. Table 4 shows the in-between points.
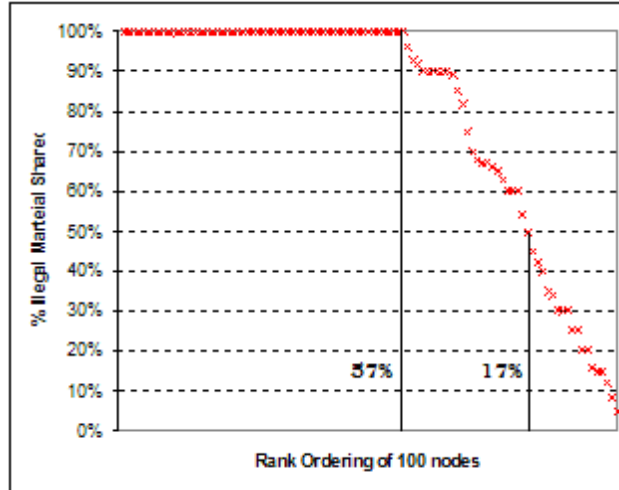
Figure 2: Percentage of illegal material shared

Table 4: Percentage of illegal material shared

Illegal QUERYHITS	Total % of peers sharing
100%	57%
> 90%	66%
> 50%	83%
> 25%	91%

Unfortunately, it is not possible to associate QUERY traffic with specific peers in the same way that it is with QUERYHIT traffic [12] and therefore it is not possible to ascertain whether the peers serving illegal material are the same as those generating QUERY messages searching for this material.

5. Discussion

We found that 1.6% of search traffic and 2.4% of response traffic was related to illegal pornography. While this is a small proportion, it remains significant, particularly given the large size of the Gnutella network. We also found strong evidence that those peers who share illegal pornography form a deviant sub community: 57% of peers that share such material share no other material, while only 17% share less than 50% illegal material.

This second finding can be explained in terms of those socio-psychological theories that emphasise the importance of group-specific social norms. The existence of a sub-community of users sharing illegal pornographic material suggests that the majority of Gnutella users are not involved in the sharing of such material. Furthermore, although the anonymity afforded by Gnutella surely makes sharing such material more attractive to this sub-community, it seems unlikely that anonymity alone is enough to trigger such 'deviant' behaviour. It might therefore be speculated that, in this context, anonymity facilitates, but does not cause, 'deviant' behaviour.

Our findings have significant implications for the debate over the legality and future survival of P2P networks. As mentioned in section 1, there is a growing trend of legal action targeting the online distribution of illegal pornography. This, together with the significant amount of this material available on P2P networks means that P2P file-sharing is likely to be increasingly targeted by such activities.

Despite this, our research has shown that those responsible for the distribution of the illegal material are a small and separate sub-community. Our findings suggest that no action need be taken with regard to P2P file-sharing networks as a whole *if* it possible to effectively target this sub-community without encroaching on the wider file-sharing community [21]. Furthermore, recent research suggests that a significant number of users are migrating to P2P networks that are harder to

police [26]. Legal attacks such as [9] may simply accelerate this process, forcing the deviant sub-community onto technologies where enforcement may be more difficult or even impossible.

6. Future Work

While Gnutella is, in many respects typical of anonymous P2P file sharing networks, it would be beneficial to verify our results across different P2P file sharing networks. For example, do those networks which emphasise anonymity such as Freenet [8], have a greater amount of traffic relating to illegal pornography? Significant benefits may also be gained from extending our study, both over a longer time period and by increasing the depth of analysis. The preliminary study reported in this paper provides only a 'snapshot' of the current situation. Extending it over a longer period, e.g. 12 months, would expose any underlying trends that the present study has missed. There are other interesting phenomena that a longer term study might illuminate. For example, do high-profile prosecutions [10] or public awareness campaigns [11] actually reduce the level of illegal pornography being shared? Such questions may prove important in determining effective law enforcement approaches. It is also possible to extend our study to obtain more information about the make-up of the deviant sub-community. For example, does the volume of pornographic material being shared by a peer relate to its geographical location? Do cultural attitudes and the laws of a peer's host-country play a significant role in shaping online behaviour? Finally, the question of whether users who share other materials also form distinct sub-communities should be addressed. While it appears counter-intuitive, it may be that users sharing legal material, (e.g. jazz fans or punk fans) form communities that are just as isolated within the broader network as the community of user sharing illegal pornography. Anonymous file-sharing networks are an ideal environment in which to evaluate CMC-related questions of this type.

7. References

[01] Napster. http://www.napster.com

[02] Kazaa. http://www.kazaa.com

[03] eDonkey. http://www.edonkey.com

[04] Limewire. http://www.limewire.com

[05] "Dependability Properties of P2P Architectures", Walkerdine, J., Melville, L., Sommerville, I. Proc. 2nd IEEE International Conference on Peer-to-Peer computing (P2P'02), Linkoping, Sweden, September, 2003.

[06] The Recording Industry Association of America (RIAA) http://www.riaa.com.

[07] "An Architecture for Peer-to-Peer Economies", Strulo, B., Smith, A., Farr, J., Proc. 3rd IEEE International Conference on Peer-to-Peer computing (P2P'03). Linköping, Sweden, September, 2003.

[08] "Freenet: a Distributed Anonymous Information Storage and Retrieval System", Clark, I., Sandberg, O., Wiley, B., Hong, T., International Workshop on Design Issues in Anonymity and Unobservability, LNCS 2009, Springer, New York, 2001.

[09] "A Bill to Outlaw the Selling, Advertising, and Distributiion of Peer-to-Peer (P2P) File-sharing Software", A proposed change to the California legal code, US Senate, January 2005, http://info.sen.ca.gov/pub/bill/sen/sb_0051-0100/sb_96_bill_20050114_introduced.html.

[10] "Hundreds targeted over child porn", BBC News, March' 05 http://news.bbc.co.uk/1/hi/world/europe/4354573.stm

[11] The Internet Watch Foundation http://www.iwf.org.uk/

[12] "The Gnutella Protocol Specification v0.4". http://www9.limewire.com/developer/gnutella_protocol 0.4.pdf, 2000.

[13] "Identifiability and Self-Presentation: Computer-Mediated Communication and Intergroup Interaction", Douglas, K. M., McGarty, C., British Journal of Social Psychology, 40, 399-416, 2001.

[14] "Causes and Implications of Disinhibited Behavior on the Internet", Joinson, A. in J. Gackenbach (Ed.), "Psychology and the Internet: Intrapersonal, Interpersonal, and Transpersonal Implications", London, Academic Press, 1998.

[15] "A Criminological Internet "Sting": Experimental Evidence of Illegal and Deviant Visits to a Website Trap", Demetious, C., Silke, A., British Journal of Criminology, 43, 213-222, 2003.

[16] "Social Identity, Normative Content and 'Deindividuation' in Computer-Mediated Groups", Postmes, T., Spears, R., Lea, M., in N. Ellemers, R. Spears, & B. Doosje (Eds.), "Social identity: Context, commitment, content", Oxford: Blackwell, 1999.

[17] "JTella", http://jtella.sourceforge.net/.

[18] "Lancaster's Peer-to-Peer Website" ,http://polo.lancs.ac.uk/p2p.

[19] "The Gnutella Protocol Specification v0.6", http://rfc-gnutella.sourceforge.net/.

[20] "Computer Pornography: A Comparative Study of the US and UK Obscenity Laws and Child Pornography Laws in Relation to the Internet", Akdeniz, Y., International Review of Law, Computers & Technology, 10, 235-263, 1996.

[21] "Internet Content Regulation: UK Government and the Control of Internet Content", Akdeniz, Y., Computer Law & Security Report, 17, 303-317, 2001.

[22] "Social Categorization and Intergroup Behaviour", Tajfel, H., Billig, M.G., Bundy, R.P., Flament, C., European Journal of Social Psychology, 1, 149-178,1971.

[23] "Improving Quality of Service on Gnutella", Hughes, D., Warren, I., Coulson, G., Technical Report COMP-005-2004, Computing Department, Lancaster University, 2004.

[24] "Free Riding on Gnutella Revisited: the Bell Tolls?", Hughes, D., Coulson, G., Walkerdine J., IEEE Distributed Systems Online, June 2005.

[25] "Free Riding on Gnutella", Adar, E., Huberman, B., First Monday, October 2000. http://www.firstmonday.dk/issues/issue5 10.

[26] "Is P2P Dying or Just Hiding?", Karagiannis, T., Broido, A., Brownlee, N., Faloutsos, M., Proc. Globecom 2004, Dallas, U.S., December 2004.

[27] "Looking at the Server-Side of Peer-to-Peer Systems" Qiao Y., Lu D., Bustamante F. E., Dinda P., Proc. 7th Workshop on Languages, Compilers and Run-time Support for Scalable Systems, October 2004.

[28] "The Sybil Attack" Douceur, J., Proc. 1st Internet Workshop on Peer-to-Peer Systems (IPTPS02), Cambridge, MA (USA), March 2002.