

Peer-to-Peer: Is Deviant Behavior the Norm on P2P File-Sharing Networks?

Daniel Hughes, *Lancaster University*

James Walkerdine, *Lancaster University*

Geoff Coulson, *Lancaster University*

Stephen Gibson, *York St John College*

Empirical evidence suggests that a small subset of the peer-to-peer community produces the majority of P2P-mediated illegal pornography.

P2P file-sharing networks such as Kazaa (<http://www.kazaa.com>), eDonkey (<http://www.edonkey.com>), and Limewire (<http://www.limewire.com>) boast millions of users. Because of scalability concerns and legal issues, such networks are moving away from the semicentralized approach that Napster (<http://www.napster.com>) typifies toward more scalable and anonymous decentralized P2P architectures.¹ Because they lack any central authority, these networks provide a new, interesting context for the expression of human social behavior.

However, the activities of P2P community members are sometimes at odds with what real-world authorities consider acceptable. One example is the use of P2P networks to distribute illegal pornography. (For another example, see the sidebar.) Both Europe and North America² have major law-enforcement efforts underway that target the distributors of such material, resulting in numerous high-profile prosecutions. Additionally, large-scale public awareness campaigns (<http://www.iwf.org.uk>) warn of the dangers that the Internet poses to children. (The California legislature's recent attempt to outlaw P2P file-sharing listed among its justifications the sharing of illegal pornography.³)

To gauge the form and extent of P2P-based sharing of illegal pornography, we analyzed pornography-related resource-discovery traffic in the Gnutella P2P network.⁴ We found that a small yet significant proportion of Gnutella activity relates to illegal pornography: for example, 1.6 percent of searches and 2.4 percent 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, our results show that a small yet particularly active subcommunity of users searches for and distributes illegal pornography, but it isn't a behavioral norm.

The Gnutella network

Gnutella is an open protocol that supports file discovery and transfer among its users. Gnutella and similar decentralized file-sharing systems are generally considered more anonymous than earlier semicentralized systems such as Napster, which used third-party indexing servers to store information about each peer and the files it made available to the network. In entirely decentralized networks like Gnutella, no such entity has knowledge of the peers or files available on the network.

We chose Gnutella because it's a good example of a large-scale, anonymous P2P file-sharing system. It also has a well-studied user-base and an open protocol specification. Specifically, we chose Gnutella over Fastrack and eDonkey because these networks feature username and password authentication and therefore aren't truly anonymous. However, while our experiments only address Gnutella, there's no reason to suppose that Gnutella users download any more or less pornography than users of other anonymous P2P networks. So, you can consider our results indicative of what you might expect elsewhere.

In technical terms, the Gnutella protocol builds an unstructured, decentralized application-level network.¹ As with any decentralized P2P network, participating peers are expected to forward network maintenance and file discovery messages and share files on the network. This simple protocol uses just five message types (see table 1).

Table 1. The Gnutella protocol messages.

Message	Description
Ping	Floods message onto the network to discover new peers. A peer that's willing to accept a new incoming connection will respond with a Pong.
Pong	The response to a Ping. A Pong 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. A QueryHit contains the information necessary 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: discovering and transferring resources (files).

To discover resources, a requesting peer forwards a Query message to its neighbors. Each neighbor then forwards this message to its neighbors, and so on, thus flooding the message onto the network. If a peer can satisfy an incoming Query (that is, it's sharing a file that matches a search term in the Query), it sends a QueryHit message back along the same path. QueryHits contain the information required to subsequently acquire the requested file—the responding peer's network address and port.

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

Anonymity's effects on online behavior

Researchers have devoted significant effort to studying how anonymity and perceived unidentifiability affect computer-mediated communication.⁵ Some have argued that anonymity generally increases the likelihood of engaging in deviant or disinhibited online behavior.⁶ For example, Christina Demetriou and Andrew Silke found that well over half of the individuals visiting a Web site for nonpornographic material nevertheless attempted to access pornography when given the opportunity to do so.⁷ Such findings suggest that when online, people might find it harder to insulate themselves from the temptation to engage in behaviors that would ordinarily incur strong social disapproval or sanction.

Other researchers, however, have suggested that anonymity's consequences in computer-mediated communication can be better understood in terms of group-specific social norms.⁸ According to this view, anonymity will only lead to deviant or illegal behavior (as defined by general societal norms) if the norms associated with a particular group identity allow for it. So, an individual will only be more likely to engage in behavior that runs counter to general social norms when anonymous online if that behavior conforms to group-specific social norms.

These two approaches offer competing predictions for the likely distribution patterns of illegal pornography on Gnutella. According to the first approach (which emphasizes anonymity's generally negative effects), no clear pattern should be detectable in the way users search for and provide deviant material. Users who share such material will simply act in an individually disinhibited way. Any Gnutella user is therefore a potential provider and user of such material. In contrast, according to the second approach, you should be able to detect a pattern in the behavior of those searching for and providing deviant material. Specifically, you would expect such users to form a distinct subclass within the wider class of Gnutella users.

For example, online anonymity's effect on the behavior of someone with a sexual interest in children would be to facilitate his or her downloading of images of childhood sexual abuse to the extent that such behavior is normative for someone with such a sexual preference. Conversely, you wouldn't expect anonymity to produce such behavior in people who don't identify themselves as having such a

sexual preference. The crucial distinction is that the former approach assumes that the simple act of using Gnutella (or the Internet in general) leaves an individual potentially at risk of engaging in deviant behavior. In the latter approach, however, the anonymity that online activity affords merely facilitates behavior associated with group norms that are more or less already inscribed.

If anonymity in P2P networks generally negatively affects user behavior, then this would support those who argue for wide-ranging legal restrictions on P2P technology, as was argued in recent legal action in California.³ If, however, this behavior is due to the influence of preinscribed group norms, it might be that sharing illegal sexual material on P2P file-sharing networks merely reflects deeper societal issues, requiring more subtle approaches to discouraging such behavior.

The experiments

Our experiments were based on intercepting and analyzing `Query` and `QueryHit` messages on the Gnutella network. Essentially, analyzing `Query` messages tells us what people are searching for, and analyzing `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 that logs all the `Query` and `QueryHit` messages it routes. Using such a peer (based on the `Jtella` classes, <http://jtella.sourceforge.net>), we monitored Gnutella traffic from 27 February to 27 March 2005. We maximized the size and typicality of our sample base by connecting to the network as an ultrapeer,⁹ maintaining many incoming and outgoing connections, and periodically reconnecting to different areas of the network.

The legality of various types of pornography varies from country to country. Even within countries, disagreements exist over the precise letter of the law.¹⁰ We limited our definition of illegality to those materials depicting practices that are clearly illegal under UK and international law: rape, incest, bestiality, and the sexual abuse of children.

What proportion of Gnutella traffic relates to illegal pornography?

To answer this question, we examined samples from three Saturdays during our monitoring period (5, 12, and 19 March). We selected Saturdays owing to the relatively higher level of traffic during weekends. From these samples, we randomly extracted 10,000 `Query`s and 10,000 `QueryHit`s. We then manually classified these as relating to either illegal pornographic or other material. The samples we used in this classification, along with other raw data, are available online at <http://polo.lancs.ac.uk/p2p/deviant>.

To ensure accurate classification, two independent reviewers classified the messages according to the above criteria. We only classified messages as relating to illegal pornographic material if the reviewers could only interpret them as referring to such material. Despite this conservative approach, some level of misclassification was inevitable owing to the nature of plain-text searches. For example, although a `Query` for "young girl" might be intended to retrieve questionable material, it could also refer to legal material (for example, a song). So, we didn't select such queries. Furthermore, owing to the illegal

nature of the content being shared, it's possible that individuals use code words to avoid detection. This seems particularly likely if those sharing such material do indeed form a distinct subcommunity.

Table 2 shows how many `Query`s and `QueryHits` the reviewers classified as relating to illegal pornography. You can see that no significant difference exists between the independent reviewers' classifications ($p = 0.3$ for `Query`s; $p = 0.8$ for `QueryHits`).

Table 2. Traffic relating to illegal material.

		Querys		QueryHits	
		Number classified as relating to illegal material	Percent of total classified as relating to illegal material	Number classified as relating to illegal material	Percent of total classified as relating to illegal material
Reviewer 1	5 March	122	1.2	206	2.1
	12 March	156	1.6	295	3.0
	19 March	158	1.6	195	2.0
Reviewer 2	5 March	142	1.4	234	2.3
	12 March	184	1.8	297	3.0
	19 March	174	1.7	208	2.1

The reviewers classified an average of 1.6 percent of `Query` messages as relating to illegal pornography. The minimum value we observed was 1.2 percent on 5 March, rising to 1.8 percent on 12 March. The standard deviation between samples was 0.2 percent.

An average of 2.4 percent of `QueryHit` messages related to illegal pornography. The minimum value observed was 2 percent on 19 March, which fell from 3 percent on 12 March. The standard deviation was 0.7 percent.

The disparity between the numbers of `Query` and `QueryHit` messages is primarily because `QueryHit` messages refer to multiple files that might have matched a single `Query`.

Is this activity the result of a deviant subcommunity?

To assess whether individuals who share illegal pornography form a subcommunity of 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 Saturdays. This yielded a list of peers that we could reasonably assume to be distributing illegal material.

We selected 100 unique hosts at random from this set and determined whether they shared other material. Figure 1 shows the proportion of illegal material that these 100 peers were serving over the one-month period.

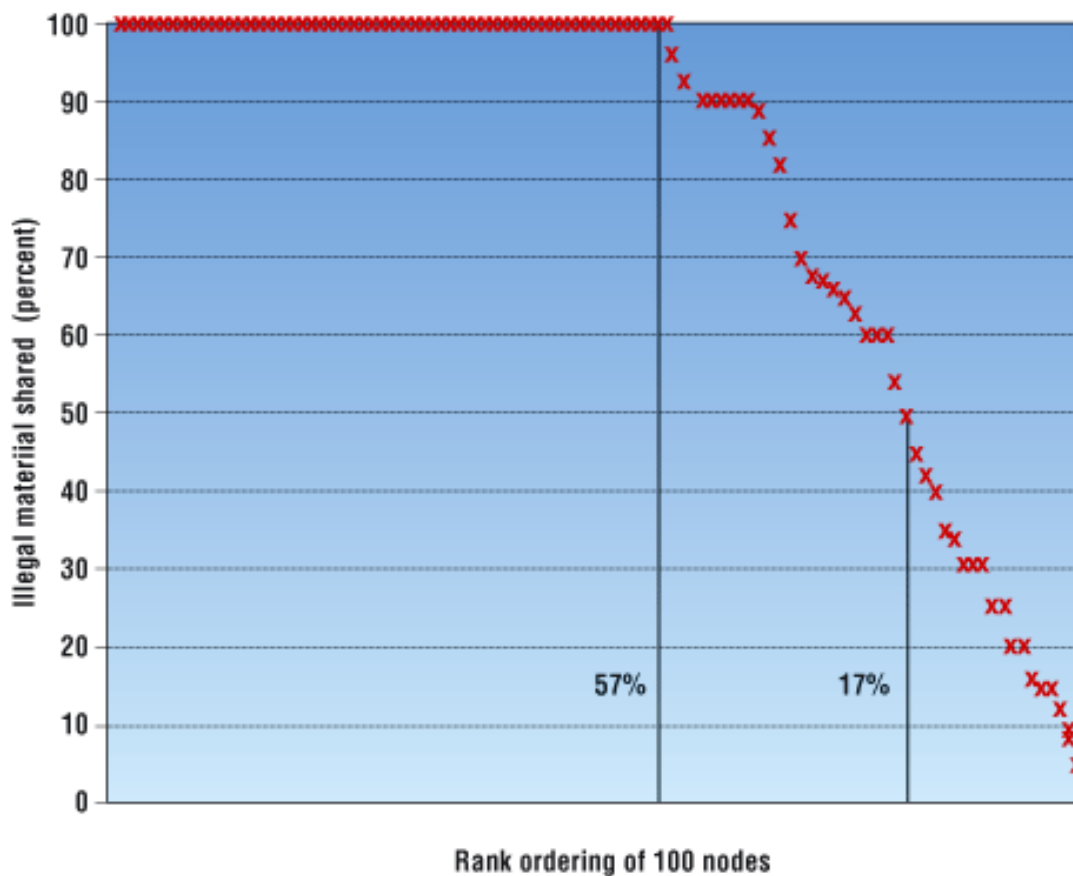


Figure 1. Percentage of illegal material shared.

Unfortunately, it's not possible to associate Query traffic with specific peers in the same way that it is with QueryHit traffic.⁴ So, we can't ascertain whether the peers serving illegal material are the same as those generating Query messages searching for this material.

Discussion

We found that 1.6 percent of search traffic and 2.4 percent of response traffic was related to illegal pornography. Although this is a small proportion, it remains significant, particularly given the Gnutella network's large size. We also found strong evidence that those peers sharing illegal pornography form a deviant subcommunity: 57 percent of peers who share such material share no other material, while only 17 percent share less than 50 percent illegal material. Table 3 shows the in-between points.

Table 3. Percentage of illegal material shared.

Percentage of illegal Query Hits	Total percentage of peers sharing
100	57
> 90	66
> 50	83
> 25	91

We can explain the second finding in terms of sociopsychological theories that emphasize the importance of group-specific social norms. The existence of a subcommunity of users sharing illegal pornography suggests that most Gnutella users don't share such material. Furthermore, although the anonymity Gnutella affords surely makes sharing such material more attractive to this subcommunity, it seems unlikely that anonymity alone is enough to trigger such deviant behavior. We might therefore speculate that, in this context, anonymity facilitates but doesn't cause deviant behavior.

Our findings have significant implications for the debate over P2P networks' legality and future survival. As we mentioned in the introduction, legal action targeting the online distribution of illegal pornography is increasing. Together with the significant amount of this material available on P2P networks, this means that such activities will likely increasingly target P2P file sharing.

Despite this, our research has shown that those responsible for distributing illegal material are a small, separate subcommunity. Our findings suggest that no action need be taken regarding P2P file-sharing networks as a whole if it's possible to effectively target this subcommunity without encroaching on the wider file-sharing community.¹¹ Furthermore, recent research suggests that a significant number of users are migrating to P2P networks that are harder to police.¹² Legal attacks such as the one proposed in California³ might simply accelerate this process, forcing the deviant subcommunity to use technologies where enforcement could be more difficult or even impossible.

Conclusion

While Gnutella is, in many respects, typical of anonymous P2P file-sharing networks, it would be beneficial to verify our results across other such networks. For example, do networks that emphasize anonymity, such as Freenet,¹³ have more traffic relating to illegal pornography? Extending our study over a longer time period and increasing the depth of analysis would also help. This study provides only a snapshot of the situation. Extending it over a longer period—for example, 12 months—would expose any underlying trends that this study missed.

A longer-term study might also illuminate other interesting phenomena. For example, do high-profile prosecutions² or public awareness campaigns (<http://www.iwf.org.uk>) actually reduce the sharing of illegal pornography? Such questions might prove important in determining effective law enforcement approaches. It's also possible to extend our study to obtain more information about the deviant subcommunity's composition. For example, does the volume of pornographic material a peer is sharing relate to its geographical location? Do cultural attitudes and the laws of a peer's host country play a significant role in shaping online behavior?

Finally, researchers should address whether users who share other materials also form distinct subcommunities. Although it appears counterintuitive, it might be that users sharing legal material (for example, jazz or punk fans) form communities that are just as isolated in the broader network as the community of users sharing illegal pornography. Anonymous file-sharing networks offer an ideal environment for evaluating questions of this type.

References

1. J. Walkerdine , L. Melville and I. Sommerville , "Dependability Properties of P2P Architectures," *Proc. 2nd IEEE Int'l Conf. Peer-to-Peer Computing (P2P 02)*, IEEE CS Press, 2003, pp. 173-174. (pdf, <http://csdl2.computer.org/comp/proceedings/p2p/2002/1810/00/18100173.pdf>)
2. Hundreds Targeted over Child Porn," <http://news.bbc.co.uk/1/hi/world/europe/4354573.stm>, BBC News, Mar. 2005.
3. Calif. Senate, Peer-to-Peer Networks: File Sharing Software," http://info.sen.ca.gov/pub/bill/sen/sb_0051-0100/sb_96_bill_20050114_introduced.html, SB96, 2005-2006 Session (14 Jan. 2005).
4. *The Gnutella Protocol Specification*, v. 0.4, 2000 (pdf, http://www9.limewire.com/developer/gnutella_protocol_0.4.pdf).
5. K.M. Douglas and C. McGarty , "Identifiability and Self-Presentation: Computer-Mediated Communication and Intergroup Interaction," *British J. Social Psychology*, vol. 40, no. 3, 2001, pp. 399-416.
6. A. Joinson , "Causes and Implications of Disinhibited Behavior on the Internet," *Psychology and the Internet: Intrapersonal, Interpersonal, and Transpersonal Implications*, J. Gackenbach, ed., Academic Press, 1998, pp. 43-60.
7. C. Demetriou and A. Silke , "A Criminological Internet 'Sting': Experimental Evidence of Illegal and Deviant Visits to a Website Trap," *British J. Criminology*, vol. 43, no. 1, 2003, pp. 213-222.
8. T. Postmes , R. Spears and M. Lea , "Social Identity, Normative Content, and 'Deindividuation'

- in Computer-Mediated Groups," *Social Identity: Context, Commitment, Content*, N. Ellemers, R. Spears, and B. Doosje, eds., Blackwell Publishers, 1999, pp. 164-183.
9. P. Kirk , *The Gnutella Protocol Specification*, <http://rfc-gnutella.sourceforge.net>, v. 0.6, 2003.
 10. Y. Akdeniz , "Computer Pornography: A Comparative Study of the US and UK Obscenity Laws and Child Pornography Laws in Relation to the Internet," *Int'l Rev. Law, Computers, and Technology*, vol. 10, no. 2, 1996, pp. 235-263.
 11. Y. Akdeniz , "Internet Content Regulation: UK Government and the Control of Internet Content," *Computer Law and Security Report*, vol. 17, no. 5, 2001, pp. 303-317.
 12. T. Karagiannis , et al., "Is P2P Dying or Just Hiding?" *Proc. 2004 IEEE Global Telecommunications Conf. (GLOBECOM 04)*, IEEE Press, 2004, pp. 1532-1538.
 13. I. Clark , et al., "Freenet: a Distributed Anonymous Information Storage and Retrieval System," *Designing Privacy-Enhancing Technologies: Proc. Int'l Workshop Design Issues in Anonymity and Unobservability*, LNCS 2009, Springer, 2001, pp. 46-66.



Daniel Hughes is a PhD research student at Lancaster University's Computing Department. His research interests include distributed systems and peer-to-peer systems. He received his master's degree in distributed interactive systems from Lancaster University. Contact him at Computing Dept., InfoLab 21, South Dr., Lancaster Univ., Lancaster LA1 4WA, UK; danny@comp.lancs.ac.uk.



Stephen Gibson is a lecturer in psychology in the School of Sports, Science and Psychology at York St John College. He is completing a doctorate at Lancaster University's Department of Psychology. Contact him at the School of Sports, Science and Psychology, York St John College, York, YO31 7EX, UK; s.gibson@yorks.ac.uk.



James Walkerdine is a research associate at Lancaster University's Computing Department. His research interests include cooperative systems, information management, human-computer interaction, and peer-to-peer systems. He received his PhD in computer science from Lancaster University. He's a member of the British Computer Society. Contact him at the Computing Dept., InfoLab 21, South Dr., Lancaster Univ., Lancaster LA1 4WA, UK; walkerdi@comp.lancs.ac.uk.



Geoff Coulson is a professor of computer science at Lancaster University's Computing Department. His main research interests are next-generation middleware and programmable networking. He received his PhD in computer science from Lancaster University. He's a member of the ACM and British Computer Society. Contact him at the Computing Dept., InfoLab 21, South Dr., Lancaster Univ., Lancaster LA1 4WA, UK; geoff@comp.lancs.ac.uk.

Related Links

DS Online's Peer-to-Peer Community, <cms:/dsonline/topics/p2p/index.xml>

"Free Riding on Gnutella Revisited: The Bell Tolls?" (pdf), <http://csdl2.computer.org/comp/mags/ds/2005/06/o6001.pdf>

"Enhancing Data Authenticity and Integrity in P2P Systems", <http://doi.ieeecomputersociety.org/10.1109/MIC.2005.124>

Cite this article: Daniel Hughes, James Walkerdine, Geoff Coulson, and Stephen Gibson, "Is Deviant Behavior the Norm on P2P File-Sharing Networks?" *IEEE Distributed Systems Online*, vol. 7, no. 2, 2006, art. no. 0602-o1001.

Another Deviant P2P Behavior: Copyright Infringement

The most obvious example of deviant behavior on P2P networks is that of copyright infringement and the resulting copyright enforcement activities by organizations such as the Recording Industry Association of America (RIAA), <http://www.riaa.com>. This issue has generated lively debate in the P2P research community and has given rise to a significant body of work devoted to both policing user behavior¹ and creating radically anonymous networks in which enforcing any form of control is impossible.² Copyright infringement has also involved the legal world in the form of recent challenges to P2P file-sharing systems' legality.³ More broadly, despite vigorous debate, society has yet to resolve fundamental questions about the nature of copyright law and its enforcement.

References

1. B. Strulo , A. Smith and J. Farr , "An Architecture for Peer-to-Peer Economies,"*Proc. 3rd IEEE Int'l Conf. Peer-to-Peer Computing (P2P 03)*, IEEE CS Press, 2003, pp. 208-209. (pdf, <http://csdl2.computer.org/comp/proceedings/p2p/2003/2023/00/20230208.pdf>)
2. I. Clark , et al., "Freenet: A Distributed Anonymous Information Storage and Retrieval System,"*Designing Privacy-Enhancing Technologies: Proc. Int'l Workshop Design Issues in Anonymity and Unobservability*, LNCS 2009, Springer, 2001, pp. 46-66.
3. Calif. Senate, "Peer-to-Peer Networks: File Sharing Software," http://info.sen.ca.gov/pub/bill/sen/sb_0051-0100/sb_96_bill_20050114_introduced.html SB96, 2005-2006 Session (14 Jan. 2005).