

Socio-Technical Issues in P2P File Sharing Systems

Danny Hughes
Computing Department,
South Drive,
Lancaster University,
Lancaster, UK. LA1 4WA
danny@comp.lancs.ac.uk

James Walkerdine
Isis Forensics,
PO Box 793,
Lancaster, UK.
LA1 9ED
j.walkerdine@isis-forensics.com

Kevin Lee
School of Computer Science,
University of Manchester,
Manchester, UK.
M13 9PL
klee@cs.man.ac.uk

1. Background

Peer-to-Peer (P2P) systems emphasise co-operation between entities known as peers at the edge of the Internet that are essentially equal and can both require and provide services [1]. Since the release of Napster in 1999, Peer-to-Peer file-sharing has enjoyed a dramatic rise in popularity. As early as 2000, a study by Plonka [2] on the University of Wisconsin campus network found that file-sharing accounted for a comparable volume of traffic to web applications, while a 2002 study by Saroiu et al. [3] on the University of Washington campus network found that file-sharing accounted for more than treble the volume of web traffic observed. Today it is widely accepted that P2P file sharing is responsible for more internet traffic than any other application, though the migration of users to more anonymous systems, together with the proliferation of P2P file sharing protocols makes exact measurements difficult [4]. As might be expected given the significance of P2P systems to the modern Internet, a number of large-scale studies of P2P traffic such as [2] [3] and [4] have been performed. These studies typically use a classical network monitoring approach and tend to focus upon the low level properties of deployed P2P systems such as bandwidth consumption, file availability and so on. We argue that existing studies neglect the social aspects of P2P file sharing systems and that empirical evidence shows that socio-technical issues are critically important on such systems. This abstract provides an overview of our work in this field, along with directions for future research.

2. Socio-technical Issues in P2P File Sharing Communities

Socio-technical issues are uniquely critical on P2P systems, which are largely anonymous and wherein all resources are provided and consumed by the users themselves. In such environments, undesirable behaviour by users has a very significant impact upon all users of the network, and due to the scale of P2P networks also gives rise to significant social issues. Our work to date has focused upon three significant socio-technical issues in P2P systems: free-riding, the distribution of illegal pornography and the use of P2P file sharing in the work-place.

2.1 Free Riding

Peer-to-peer file-sharing networks embody a social dilemma wherein the behaviour of rational users is at odds with the common good. The dilemma for each individual is whether to contribute to the common good by sharing files, or to maximise their personal experience by 'free riding' (i.e. downloading files while not contributing to the network). As there is no personal benefit gained by uploading files (in fact it is inconvenient), it is 'rational' for users to free ride. However, significant numbers of free riders degrade the utility of the entire system, an effect known as 'the tragedy of the digital commons' [5]. The first analysis of Free Riding was published by Adar in 2000 [5], and found that even at this early stage more than 66% of users on the Gnutella network (one of the first fully decentralised P2P file sharing systems) were free riders. We revisited Adar's study in 2005 and found that free riding had increased to over 85% of network users [6]. We speculated that this drop was driven primarily by the greater perceived risk of sharing files caused by increased copyright enforcement activity. Revisiting these results again this year (2007) reveals that free riding has risen to a staggering 97% of Gnutella users [7]. These results point to the imminent collapse of the Gnutella network, unless some way to alter the perceived balance of the free-riding social dilemma cannot be found.

2.2 Distribution of Illegal Pornography

As P2P networks are largely anonymous, lack any form of central authority and often support millions of users, they have aroused considerable concern in broader society. One obvious example is the use of P2P file sharing systems to circumvent copyright, which has led to a number of high profile prosecutions and lively debate. Another, perhaps more critical example, is the use of P2P file sharing

systems to distribute illegal pornography, and in particular images of child abuse. This has led on the one hand to major law enforcement efforts to target the distributors of such material and on the other has been used to justify legal attempts to ban P2P systems altogether. In 2006, in collaboration with psychologists at York St. John College, we performed an extensive analysis of Gnutella to determine the extent and characteristics of the distribution of illegal pornography [8]. We found that, on average, between 1% and 3% of resource discovery traffic references illegal pornography, the equivalent of hundreds of thousands of searches per day, however, we also found that this material was distributed by a small but very active, and essentially distinct sub-community within the network, which distributed nothing but illegal pornography. More than 60% of those users who were observed sharing illegal pornography shared nothing else. The implication of this finding is that, while P2P networks are a major vector for the distribution of illegal pornography, the 'average' P2P user finds this behaviour as unacceptable as wider society.

2.3 P2P in the Workplace

P2P file sharing causes a number of problems for businesses, including: unpredictable network usage, increased security threats and the risk of legal action, as described in [7]. To avoid the problems associated with P2P file sharing, businesses have made great efforts in recent years to secure their networks. This has included on the one hand technical attempts to block P2P applications and on the other prohibitions against the use of P2P in the work-place. We recently analysed business traffic on the Gnutella network and found that despite such efforts, workplaces still account for an average of between 3% and 4% of total P2P traffic. Furthermore, we find that small businesses account for the majority of this traffic and that large businesses are significantly more successful at securing their networks. In particular, our most recent research suggests that the success of large businesses in preventing the use of P2P systems is due to prohibitions against P2P in the workplace, as the release of popular viral media significantly increases the number of large businesses observed on P2P networks. It would appear that the release of viral media provides sufficient temptation to make users break such prohibitions against file sharing.

3. Directions for Future Work

To date, our work on monitoring and evaluating P2P file sharing has primarily been based upon results inferred from the analysis of traffic data. We are particularly interested in extending our studies beyond this, to validate our results using other methodologies. We are also particularly interested in re-using our significant body of trace data to support new analyses, particularly of the social and psychological factors which drive user behaviour on P2P file sharing networks.

4. References

- [1] Shirky C., Listening to Napster, in *Peer-to-Peer: Harnessing the Power of Disruptive Technologies*, Oram A. (ed), O'Reilly and Associates, 2001, pages 21-37.
- [2] Plonka D., Napster Traffic Measurement, Univeristy of Wisconsin-Madison, available online at: <http://net.doit.wisc.edu/data/Napster>, March 2000.
- [3] Saroiu S., Gummadi K., Dunn R. J., Gribble S. D., Levy H. M., An Analysis of Internet Content Delivery Systems, published in the proceedings of the 5th International Symposium on Operating Systems Design and Implementation (OSDI) Boston, USA, 2002
- [4] Karagiannis T., Broido A., Brownlee N., Faloutsos M., Is P2P Dying or Just Hiding?, In the Proceedings of Globecom 2004, Dallas, U.S., December 2004.
- [5] Adar E., Huberman B., First Monday, October 2000, available online at: <http://www.firstmonday.dk/issues/issue5.10>
- [6] Hughes D., Coulson G., Walkerdine J., Free Riding on Gnutella Revisited: the Bell Tolls?, published in *IEEE Distributed Systems Online*, vol. 6, no. 6, June 2005. <http://csdl2.computer.org/comp/mags/ds/2005/06/o6001.pdf>
- [7] Lee K., Walkerdine J., Hughes D., On the Penetration of Business Networks by P2P File Sharing, published in the proceedings of the 2nd International Conference on Internet Monitoring and Protection (ICIMP'07), Santa Clara, California, USA. July 2007.
- [8] Hughes D., Gibson S., Walkerdine J., Coulson G., Is Deviant Behaviour the Norm on P2P File Sharing Networks?, published in *IEEE Distributed Systems Online*, vol. 7, no. 2, February 2006. <http://csdl.computer.org/comp/mags/ds/2006/02/o2001.pdf>