A Discussion on the Ethical Issues in Peer-to-Peer Network Monitoring

Nicolas Small, James Meneghello, Kevin Lee, Nazanin Sabooniha, Raymond Schippers
School of Information Technology, Murdoch University,
Murdoch 6150, Western Australia, Australia

***Abstract*—Peer-to-peer networks have gained a significant amount of popularity since their inception, being used in a wide variety of contexts; both legal and illegal. The demand for monitoring techniques able to analyse performance, identify copyright infringers and detect the presence of peer-to-peer traffic on networks has similarly risen. Along with them ethical issues have surfaced, awareness of which is essential when planning research requiring peer-to-peer monitoring. This paper discusses the collection of data through the monitoring of peer-to-peer networks and identifies areas of particular ethical concern in its use.**

I. Introduction

Peer-to-peer networks constitute a significant amount of Internet traffic present on today's networking infrastructure [1], with a number of different uses across personal, business and academic fields. Interest in peer-to-peer and the behaviour of peers participating in these networks is leading to an increased desire for techniques that can accurately monitor these communications for a variety of reasons: legal, academic, and analytical.

Although peer-to-peer networks tend to suffer from stereotypes relating them wholly to illegal activity, there are a number of legitimate and beneficial uses for the architecture, particularly in the domain of content distribution. Several other uses such as distributed computing, collaboration and distributed databases all favour peer-to-peer due to its highly scalable and redundant properties [2].

Data collected by monitoring techniques could be used for unintended purposes, requiring the careful design of research studies in this area. Like all research involving data collection from uninformed participants, issues that present ethical dilemmas for researchers arise [3].

This paper motivated by the fact that, while a significant amount of research has been conducted on the ethics behind peer-to-peer file-sharing [4], [5], [6], little has been established regarding the ethics of monitoring peer-to-peer networks. It is hoped that with increased awareness of the various issues present in monitoring methods and the possible consequences of improperly safeguarding data collected during this monitoring, some measure of caution will be applied to investigations in the area.

Section II begins with a summary of peer-to-peer networks and their applications, then describes a subset of techniques used to monitor peer-to-peer networks and participating users. Section III describes the various approaches to Peer-to-Peer monitoring. In Section IV the paper discusses the ethical implications of researching and performing peer-to-peer network monitoring. Section V discusses these issues in the context of a case study of the monitoring of several Peer-to-Peer protocols. Finally, Section VI presents some conclusions.

II. Peer-to-Peer Networks

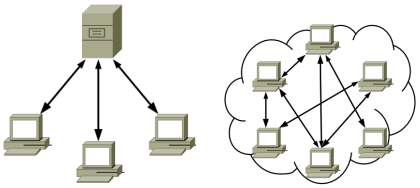
Peer-to-Peer (P2P) is a style of network architecture in which participants each share a part of their resources (storage capacity, bandwidth, processing power, etc.) for the purpose of operating in collaboration to achieve a specific task. Each node in the network should be capable of supplying services or content directly with other nodes, without communications being required to pass through intermediary servers [7]. Network membership in a P2P network is ad-hoc and dynamic, with peers playing the roles of both resource providers and consumers - often simultaneously.

In contrast to more traditional client-server architectures, P2P distributes information directly amongst participants rather than concentrating information in a centralised server cluster. Therefore, when a peer desires information from the network it must initiate any communications that take place by locating nodes that have desired data available and sending a request. This differs from client-server, in which a client is able to simply query the central server for any desired data.

A. Operation

P2P networking was developed to address certain shortcomings of existing networking architectures in specific circumstances; particularly content distribution, which traditionally required a one-to-one connection (and all bandwidth attributable) between every client and the server, represented in Fig. 1a. In doing this, massive upstream bandwidth was needed by the server in order to complete file requests in a timely fashion, and user-side upstream was only minimally used. In order to simultaneously reduce strain on dedicated servers and utilise user-side upstream, the P2P architecture was proposed has been effectively used in a number of internet protocols ranging from file transfers to distributed computing [8].

Unlike the client-server architecture which establishes a virtual connection between server and client on which to transfer data, P2P networks operate in a swarm of clients; as seen in Fig. 1b, each user becomes a node in a cloud of users. In order to effectively co-ordinate nodes, some level of hybridisation usually exists; for example, BitTorrent uses



(a) Client-Server (b) Peer-to-Peer

Fig. 1. Internet Architectures

.torrent files stored on web servers and Folding@Home uses dedicated servers to assign workloads. It is worth noting that some of these procotols have often been entirely decentralised and purified through various methods due to the legal threat an unauthorised index file hosted on a server can pose. Napster was considered a hybrid P2P protocol that used centralised servers for file indexing and search purposes, making those servers (and their owners) a primary target for anti-piracy lawsuits [9] and Distributed Denial of Service attacks - a problem generally not faced by fully-decentralised systems, as publishers or assailants would have to pursue individuals.

B. Applications

Peer-to-peer networks, whilst frequently envisioned as only used for file-sharing, have a number of other uses that aim to leverage their highly redundant, scalable nature; an attribute attractive to a number of system architectures. A survey by Androutsellis-Theotokis [2] provides some indication of how widespread possible applications for peer-to-peer are: Voice over IP [10], resource discovery for grid computing [11], video streaming [12], file-sharing [13], web portal systems such as Osiris [14] and social communications systems such as Diaspora [15]. Of particular interest to this paper is the use of P2P networking for file-sharing.

File-sharing networks represent a significant portion of P2P networks in use, with P2P files-sharing being considered to make up approximately 50 percent of all Internet traffic [16]. This file-sharing is done over a variety of different P2P protocols (e.g. BitTorrent, Gnutella, eDonkey, Ares), which usually differ slightly in terms of architecture and communication format, but operate on a set of base principles; primarily, each peer in a swarm maintains a set of local files and associated meta-data for the purpose of content distribution to other peers within the swarm.

P2P networks are particularly well-suited to file-sharing and content distribution due to several inherent attributes; redundancy, speed of distribution, cost and availability [17]. Due to their highly dynamic nature, P2P networks are almost completely redundant; the network only degrades once file availability decreases to below 100 percent, which in turn only occurs when no remaining peer contains a complete copy of the file [13]. As long as a single seed remains, the file is still available for distribution.

As a result of its highly-distributed nature, P2P also encourages substantial increases in distribution speed when compared to client-server [18], as every peer is able to distribute pieces of data to any node that does not yet have the complete file. The use of peer upstream bandwidth to distribute data also reduces the bandwidth use of the initial content provider, lowering related costs.

One of the ways in which routing and availability information was decentralised in file-sharing networks was to correctly utilise Distributed Hash Tables [19] - a method of distributing file availability and routing information over a swarm of users. There are a number of DHT algorithms in use by file-sharing systems such as Bittorrent [20], as well as Content-Addressable Networks [21] and others, as detailed by Androutsellis-Theotokis [2]. Hashes are generated for files being shared and mapped to nodes, then distributed over the user cloud so that all available files can be located by querying nodes. Users that wish to join the swarm are usually directed there by a specially-crafted hypertext link, such as a Magnet URIs. As nodes join and leave the swarm the hash tables are redistributed to retain availability, and any node that wishes to download a piece of the available files can progressively query nodes in the table for a node ID to retrieve the piece from [22]. By contrast, Napster had a centralised search server that would keep an index of nodes and file availability [23], and peers would directly query the search servers in order to determine file piece location (though the transfers themselves were still peer-to-peer).

Peer-to-peer file-sharing is not limited to unauthorised distribution of copyrighted content - a number of large organisations are utilising peer-to-peer for the purpose of content distribution when file demand reaches high levels. A prime example of this is Blizzard Inc., who now use a BitTorrent-based system to distribute patches for their software. Due to the extremely large quantity of users (approximately 11.4 million for World of Warcraft alone [24]) that require access to patches in order to play online, new methods of file delivery were trialled in order to find a more scalable technology that could cope with the amount of requests. In the end, a proprietary client leveraging BitTorrent as a distribution protocol was selected and is currently being used to distribute patches for all recent Blizzard games.

C. Anonymity

Anonymity can be difficult to achieve when creating connections over the internet, as there is no way to transfer data without some kind of point-to-point connection existing. Trusted intermediate proxy servers can be used to prevent a direct connection from peer to peer, but these servers present a form of centralisation that can be taken advantage of to identify users [25]; as well as a significant expense to users due to an increase in bandwidth quota utilisation.

Anonymous pure peer-to-peer connections are possible under certain circumstances, but are relatively rare due to a reliance on multicast connections that tend to be unavailable on the majority of remote networks without the intervention of internet service providers (who are unlikely to provide such a service). One possible way of allowing anonymous peer-to-peer connections is through use of Onion Routing [26], a

technique that involves the passing of requests between peers in a chain. Each link in the chain is unable to determine whether the request originated from the node before it or whether it is another link in the chain, so the request remains essentially anonymous. However, this approach can result in compromised anonymity if attackers perform a prolonged attack [27]. To counter this strategy, a method of file-sharing involving only unicast (point-to-point) connections was designed to allow anonymous peer-to-peer connections and transfers to take place. This approach has two disadvantages, however; if multicast connections are not utilised, there must be an untrusted proxy to initialise the sharing session that represents a single point of failure, and transfer speeds are degraded significantly due to the extra transfers between peers required [27].

Identification of a user in any network is usually only able to be traced as far as the user's nearest router (generally a home router), as any addresses identified only show that a particular router was used to access information, and rarely includes any particular details on which specific device or user accessed the information. Extensive auditing protocols are required to be able to determine precisely which user was accessing specific data at any time and are still sometimes unreliable [28]. The use of public wireless access points is a particular problem for user identification, as it generally becomes impossible to prove that a specific user was accessing information at a certain time [28].

III. Peer-to-Peer Monitoring

A. Rationale

Monitoring of peer-to-peer networks can be accomplished for a variety of reasons; copyright enforcement [28], general research purposes [29], cost/benefit analyses and discovery of peer-to-peer presence [30] are some examples of relatively common monitoring rationales.

- **Copyright Enforcement:** One of the primary reasons to monitor peer-to-peer networks is for the purpose of copyright enforcement. Often contracting to specialised companies, copyright owners direct their resources into directly monitoring and identifying the IP addresses of users on swarms that share files owned by the organisation. Once a list of infringing IP addresses has been compiled, the organisation has a number of options: notifying the users' ISPs that infringement has taken place and letting the ISP handle it, subpoenaing the ISP with warrants requesting user identification and taking their own action, or sending infringement notices directly to the user, often with a settlement notice. However, this process encounters problems stemming from potential inaccuracies in user identification below the router level (See Section II-C) [28].
- **General Research Purposes:** Monitoring of P2P networks can reveal interesting details about peers in a swarm such as statistical data (e.g. average speeds), geographical location, and the behaviour of peers, such as their habits with regards to sharing information [29] [30]. Similar studies are also useful in establishing the efficiency and performance of newly-designed protocols.
- **Benefit Analyses:** Companies planning on utilising P2P technology have reason to conduct a cost/benefit analysis on a test platform. In this case monitoring would be performed on the network to determine benefits derived from using peer communications over traditional client-server architectures.
- **P2P Presence and Security:** Peer-to-peer file-sharing networks operating within organisational networks can be problematic for a number of reasons. Employees utilising P2P file-sharing networks can create degradation of network performance, security threats like the accidental sharing of confidential data [29], and legal issues. As one study discovered, an estimated 3-4 percent of Gnutella traffic was attributed to business networks [30].

B. Monitoring Methods

Peer-to-peer monitoring and tracing commences with traffic categorisation at a number of possible levels. Three primary levels are defined as network-level, passive application-level and active application-level [1]; each providing benefits and shortcomings depending on the specific situation they are utilised in.

- **Network-level:** Examines traffic at a packet level and identifies P2P traffic by matching packets with the known characteristics of various P2P systems. Deep Packet Inspection can be deployed at this level, though real-time processing can be strenuous for processing servers depending on volume bandwidth passing through [31]; the majority of studies are not likely to require real-time analysis, however. The trace is relatively transparent, but requires network access in order to deploy it effectively.
- **Passive application-level:** Operates by running a modified P2P client and logging any routing information and/or file requests that are communicated with the client. Also considered relatively transparent, a passive client does not require administrative network access - however, useful data gathering is limited.
- **Active application-level:** Similar to passive monitoring, active monitoring also involves running a modified P2P client on an existing swarm and logging any routing and connection information. The key contrast lies in the activity of the application; an active trace will proactively seek out and discover peers and the resources they contain. As such, the data gathered is extremely plentiful and useful - but the client is extremely visible, which may be undesirable for some purposes.

Additionally, difficulties with monitoring are omnipresent - identifying peer-to-peer traffic can be difficult if it is encrypted, and thorough methods of data analysis such as Deep Packet Inspection can quickly become overloaded or result in severe performance degradation if extremely large amounts of incoming data need to be analysed in real time. The presence of monitoring agents in a peer-to-peer swarm can

also be discovered when using active monitoring, which can be undesirable [28], [32]. Regardless, the monitoring of peer-to-peer networks can still provide interesting and valuable results for academic purposes.

IV. Ethical Issues

During the course of monitoring a peer-to-peer network, several ethical issues can be encountered; though these tend to vary depending on research purpose and the network involved. For example, ethical issues with monitoring live file-sharing networks tend to be related to the legality of content being shared, whereas monitoring peer-to-peer communications networks could reveal private details about individuals. Without consent supplied by individuals who participate in the networks under observation, users' privacy is solely under the protection of the researcher [33]. It is important to note that obtaining consent in the first place is extremely difficult, as gathering valid usert identities and contact information is problematic at best.

Marx [34] presents a framework for evaluating surveillance techniques from an ethical perspective that emphasises the responsibility of the observer to avoid harm to individuals being monitored. The framework is almost wholly applicable to evaluating the ethics of peer-to-peer monitoring, due to similarities in methodology and sample groups. This framework proposes three possible sources of ethical issues in a surveillance context, briefly described below through a non-exhaustive set of associated questions:

- **The Means:** Can the collection technique cause harm? Does the technique break the trust of the subject?
- **The Data Collection Context:** Are the participants aware of the monitoring? Was a consent form signed?
- **Uses:** What is the data being used for? Could this use cause harm to the subject?

A. The Means

Given the relative passivity of P2P monitoring techniques (See Section III-B), and the fact that none of these are performed in the physically, it is unlikely that ethical concerns relating to the physical wellbeing of the subjects could be triggered by the monitoring itself.

Issues of service degradation are more topical, as demonstrated by the attempts made by a number of Internet service providers to use peer-to-peer presence sensing technology so as to throttle peer connections. The use of this monitoring in order to degrade user experience, dictated solely by the presence of P2P traffic on a connection has been met with resistance. Indeed, the few Internet service providers who implemented such throttling methods found they were negatively impacting legitimate uses of P2P, resulting in a public outcry of sufficient size to force a reversion of policy [35]. This brings the issue of invalidity to the fore; Can the technique(s) used for monitoring produce invalid results? [34]. If the detection of P2P traffic is so imprecise as to make it impossible to differentiate legal from illegal uses, it surely is unethical to arbitrarily disrupt all P2P traffic on the off-chance it might be illegal. In addition, from a copyright enforcement perspective, proper identification of peers in a P2P network can be challenging (See Section II-C), and can lead to cases of false identification, which can then cause psychological harm to the falsely accused. This makes invalidity a definite ethical concern when performing this type of monitoring.

Some peer-to-peer protocols suffer from degraded anonymity if monitoring is performed for sustained periods of time [27]; with this in mind, if monitoring 100 users behaviour for a week would result in a similar level of data quality to monitoring 10 users for 10 weeks, it may be worth using the former method: choosing the method that guarantees increased anonymity is the most ethically sound decision. Population and sample size must be taken into account when designing potential alternatives, but devising a more ethical solution with no or little loss in accuracy would be considered more acceptable than a situation that is potentially dangerous for users.

B. The Data Collection Context

From the outset, there are ethical issues arising from the context of monitoring P2P networks. From a research perspective, involving users unaware of the ongoing monitoring could be considered unethical; the decision hinging on whether users participating in a network that is part of the public domain actually have any right to privacy. It is not an unreasonable assumption to expect users to realise their actions in publicly-accessible communities may have a negative impact on their own personal privacy - if they do not wish their behaviour and actions to be monitored, they have the option of leaving the network. It is entirely possible that, disregarding the difficulties already highlighted relating to entering in contact with participants of the P2P network, informing these participants that monitoring is taking place could cause the data obtained to lose its value, especially in cases where illegal activity is being monitored. This once again leads researchers towards the performance of an ethical balancing act, weighing the need for accurate data versus proper participant awareness and consent.

In cases like these, where the ethical choice is not necessarily clear cut, Marx asserts that answering a simple question can provide a valuable look at whether or not a particular study is ethical; Would those in charge of the monitoring agree to being monitored in the same fashion? [34]. While more obviously applied to more extreme forms of surveillance, differences in ethical opinions between researchers and those under study can vary substantially; users are not always as eager to be participants to research as academics might like to assume. It is worth considering any potential divergences in these points of view.

C. Uses

Foremost amongst our ethical concern should be determining whether the data collected during monitoring could cause physical or psychological harm when put to its intended use. If data is collected from a live network that could include copyright-infringing material, publishing that data could result

in the identification of users - possibly leading to litigation. Considering the extreme penalties that tend to be handed down to infringers, successful litigation is likely to ensure significant psychological and financial harm to the user - who likely never gave their consent to begin with. This is an issue familiar to criminologists, who have codes of ethics regarding the divulging of data which could lead to repercussions on their sources. For example, the British Society of Criminology's code of ethics clearly states that the safekeeping of the participant's physical, social and psychological wellbeing is the responsibility of the researcher [36]. With this in mind, all datasets should be anonymised prior to publication in order to avoid harming the wellbeing of the participants. This then raises other issues; What if the data required for a study requires some measure of identity amongst users, or it becomes worthless? Ensuring that data collected is only used for its stated and intended purpose is paramount in this case. The possibility of stored data intended for research being transmitted to other parties and used to initiate litigation against users is a real possibility if the data is not kept secured and under the management of an ethical researcher.

As a rule, an ethically justifiable study should be able to answer the following questions with the affirmative [34]:

- Does this monitoring serve broad community goals?
- Have other ways to achieve similar ends, even if at a higher cost, been considered but found insufficient?
- Is the goal worth the cost (direct or collateral) of the means?
- Is the information collected to be used to satisfy goals other than the causing of harm or disadvantage to the subjects?

V. Monitoring Peer-to-Peer File Sharing Networks: A Case Study

A. Overview

To investigate the issues highlighted in Section IV, a study was performed by monitoring a number of different file-sharing protocols at various times over the period of two years.

B. Collection

Because file-sharing protocols operate significantly differently at a low-level (and some at a conceptual level), a solution to parse each protocol was developed individually. Each script has a similar objective; to record sightings of connections from external addresses, as well as other metadata such as the time seen, country of origin and file accessed. All network addresses collected were anonymised after analysis, to prevent storage of identifying data.

1) Gnutella: The Gnutella protocol takes advantage of peers for use as routing nodes within the network. By positioning a modified client within a network, traffic can easily be monitored and logged for later use. After network addresses are collected, they are compared to a predefined database of network owners to determine the suspected affiliation of the device. The Gnutella network was monitored for the period of one month, with 260,000 hits recorded.

2) BitTorrent: Unlike the Gnutella protocol, BitTorrent monitoring requires an active scraping mechanism to seek out torrents and determine peers connected. The scripts to handle this activity are effectively in two components; the first is a torrent scraping program written for individual search indexes that finds and retrieves torrent files, while the second accesses the trackers listed in each collected torrent file and determines seeds/peers for logging. To avoid polluting the resulting data, the script host's address is removed from the logs. A number of the most popular BitTorrent trackers were scraped and addresses monitored for a period of 15 months, with some 778 million hits recorded across 465,140 files.

3) OpenNap: Hearkening back to the earlier years of file-sharing, the Napster protocol does not provide a list of available files that can be scraped. Instead, a predefined list of search terms is used to actively seek out clients sharing files. Once a peer is found, details are collected and logged for future use. The OpenNap network was monitored for the period of one month, with 3.1 million hits recorded.

4) Sopcast: SopCast conveniently provides a channel list accessible in XML that can be parsed to determine content and peers currently accessing the content. Addresses were then queried for owner affiliation. SopCast was monitored for the period of one month, with 434,000 hits recorded.

C. Discussion

A number of serious ethical issues come into play when collecting sensitive data over a substantial period of time, as this study did. The possible damage that could result from a leak of a dataset of this type in the event of compromised anonymity is catastrophic. A basic analysis of the dataset collected suggests that, assuming the maximum penalty of USD$30,000 per infringement, damages claimed by litigious means could total an estimated USD$24 trillion. Despite the absurdity of such a sum and the likelyhood that some files are legitimate downloads or outside legal jurisdiction, circumstances dictate that care is taken in collection and anonymisation.

The data collected is not only useful to rights-holders - a number of ethically-gray entities could use the information contained within to coerce affected parties under the threat of revealing related copyright infringement. As most enterprises own whole IP address block allocations, removing the last octet of an IP address will likely not anonymise the connection wholly, as even the truncated address is likely to fall within a publically-listed subnet. It may be worth tracking individual user sightings and one-way hashing their addresses, which would maintain consistency of user connections while eliminating the storage of addresses, avoiding many issues. However, such an approach also removes the possibility of some avenues of further analysis that depend on specific network addresses, such as geographic or business mapping.

Avoiding infringement while monitoring these networks can also be an issue. As part of the swarm, a modified peer will still respond to other monitors that may be operating within the network (such as those operated by rights-holders), although

it is unnecessary to take any part of the file itself to retrieve a list of peers that have made it available. This limits our data retrieval from the swarm to merely meta-data, which does not infringe copyright. With advances in decentralisation techniques (such as peer exchange and DHT) there may be a requirement for some part of the file to be accepted in order to monitor peers - in such an instance, limiting data downloaded to less than a single piece may avoid infringement depending on local legislation.

It is important to note that data collected in this study cannot be generalised to the whole internet population. By actively scraping torrent index files from particular hosts in a particular order, there is a recognisable pattern of torrents being processed that could interfere with any complete analysis. In addition, firewalls such as PeerGuardian that prevent connection to certain recognised hostile hosts could blacklist any observers, limiting the population further. Care must be taken when drawing generalising conclusions based on data gathered with these techniques.

VI. Conclusion

This paper presents a number of key issues that should be considered when designing studies involving the monitoring of users participating in peer-to-peer networks, as well as an overall summary of P2P applications and monitoring techniques. Ethical issues include: compromised user anonymity, difficulties with and impact of obtaining user consent, and use or misuse of the data collected. Ensuring that one's experimental methodologies are designed to favour an ethical approach should therefore be a priority before engaging in the monitoring of peer-to-peer networks.

References

[1] D. Hughes, J. Walkerdine, and K. Lee, "Monitoring challenges and approaches for P2P File-Sharing systems," in *Internet Surveillance and Protection, 2006. ICISP '06. International Conference on*, 2006, p. 18.

[2] S. Androutsellis-Theotokis, "A survey of peer-to-peer file sharing technologies," 2002.

[3] L. Roberts and D. Indermaur, "Signed consent forms in criminological research: Protection for researchers and ethics committees but a threat to research participants?" *Psychiatry, Psychology and Law*, vol. 10, no. 2, p. 289–299, 2003.

[4] A. M. Levin, C. D. Mary, and K. Rhee, "Money for nothing and hits for free: the ethics of downloading music from peer-to-peer web sites," *Journal of Marketing Theory and Practice*, vol. 12, no. 1, pp. 48–60, 2004.

[5] R.-A. Shang, Y.-C. Chen, and P.-C. Chen, "Ethical decisions about sharing music files in the p2p environment," *Journal of Business Ethics*, vol. 80, no. 2, pp. pp. 349–365, 2008.

[6] M. Cenite, M. Wanzheng Wang, C. Peiwen, and G. Shimin Chan, "More than just free content: Motivations of peer-to-peer file sharers," *Journal of Communication Inquiry*, vol. 33, no. 3, pp. 206–221, Apr. 2009.

[7] W. Kellerer, "Dienstarchitekturen in der Telekommunikation-Evolution, methoden und vergleich," Technical Report TUM-LKN-TR-9801, Tech. Rep., 1998.

[8] G. Fox, "Peer-to-peer networks," *Computing in Science & Engineering*, vol. 3, no. 3, p. 75–77, 2001.

[9] J. U. Blackowicz, "RIAA v. napster: Defining copyright for the Twenty-First century," *BUJ Sci. & Tech. L.*, vol. 7, p. 182, 2001.

[10] S. A. Baset and H. Schulzrinne, "An analysis of the skype peer-to-peer internet telephony protocol," *Arxiv preprint cs/0412017*, 2004.

[11] R. Ranjan, L. Chan, A. Harwood, S. Karunasekera, and R. Buyya, "Decentralised resource discovery service for large scale federated grids," *e-science*, p. 379–387, 2007.

[12] B. Li, S. Xie, Y. Qu, G. Y. Keung, C. Lin, J. Liu, and X. Zhang, "Inside the new coolstreaming: Principles, measurements and performance implications," in *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*, 2008, p. 1031–1039.

[13] J. Pouwelse, P. Garbacki, D. Epema, and H. Sips, "The bittorrent p2p file-sharing system: Measurements and analysis," *Peer-to-Peer Systems IV*, p. 205–216, 2005.

[14] Osiris, "Osiris serverless portal system," 2010. [Online]. Available: http://www.osiris-sps.org/

[15] D. Grippi, M. Salzberg, R. Sofaer, and I. Zhitomirskiy, "DIASPORA* ALPHA," 2010. [Online]. Available: https://joindiaspora.com/

[16] H. Schulze and K. Mochalski, "Internet study 2008/2009," *Research Report, ipoque*, 2009.

[17] M. Ripeanu, I. Foster, and A. Iamnitchi, "Mapping the gnutella network: Properties of large-scale peer-to-peer systems and implications for system design," *Arxiv preprint cs/0209028*, 2002.

[18] R. Kumar and K. W. Ross, "Peer-assisted file distribution: The minimum distribution time," in *Proc. IEEE Workshop on Hot Topics in Web Systems and Technologies*, vol. 6, 2008.

[19] B. Wiley, "Distributed hash tables, part i," *Linux Journal*, vol. 2003, no. 114, Oct. 2003.

[20] J. A. Pouwelse, P. Garbacki, D. Epema, and H. J. Sips, "An introduction to the BitTorrent Peer-to-Peer File-Sharing system," Citeseer, Tech. Rep., 2004.

[21] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Shenker, "A scalable content-addressable network," in *Proceedings of the 2001 conference on Applications, technologies, architectures, and protocols for computer communications*, 2001, p. 161–172.

[22] P. Maymounkov and D. Mazieres, "Kademlia: A peer-to-peer information system based on the xor metric," *Peer-to-Peer Systems*, p. 53–65, 2002.

[23] S. M. Lui and S. H. Kwok, "Interoperability of peer-to-peer file sharing protocols," *ACM SIGecom Exchanges*, vol. 3, no. 3, p. 25–33, 2002.

[24] Blizzard Entertainment, "Blizzard entertainment: Blizzard FAQ." [Online]. Available: http://eu.blizzard.com/en-gb/company/about/legal-faq.html

[25] O. Berthold, H. Federrath, and M. K\öhntopp, "Project "anonymity and unobservability in the internet"," in *Proceedings of the tenth conference on Computers, freedom and privacy: challenging the assumptions*, 2000, p. 57–65.

[26] P. F. Syverson, D. M. Goldschlag, and M. G. Reed, "Anonymous connections and onion routing," *sp*, p. 0044, 1997.

[27] V. Scarlata, B. Levine, and C. Shields, "Responder anonymity and anonymous peer-to-peer file sharing," in *Network Protocols, 2001. Ninth International Conference on*, 2001, pp. 272–280.

[28] M. Piatek, T. Kohno, and A. Krishnamurthy, "Challenges and directions for monitoring P2P file sharing networks-or: why my printer received a DMCA takedown notice," in *Proceedings of the 3rd conference on Hot topics in security*, Berkeley, CA, USA, 2008, p. 12:1–12:7.

[29] M. E. Johnson, D. McGuire, and N. D. Willey, "Why file sharing networks are dangerous?" *Communications of the ACM*, vol. 52, p. 134–138, Feb. 2009, ACM ID: 1461962.

[30] K. Lee, D. Hughes, and J. Walkerdine, "On the penetration of business networks by P2P file sharing," in *Internet Monitoring and Protection, 2007. ICIMP 2007. Second International Conference on*, 2007, p. 23.

[31] T. Porter, "The perils of deep packet inspection," *on-line article at www. securityfocus. com*, 2005.

[32] G. Siganos, J. Pujol, and P. Rodriguez, "Monitoring the bittorrent monitors: A bird's eye view," *Passive and Active Network Measurement*, p. 175–184, 2009.

[33] H. E. Keller and S. Lee, "Ethical issues surrounding human participants research using the internet," *Ethics & behavior*, vol. 13, no. 3, p. 211–219, 2003.

[34] G. T. Marx, "Ethics for the new surveillance," *The Information Society*, vol. 14, no. 3, p. 171–185, 1998.

[35] M. Masnick, "Rogers traffic shaping making it difficult for users to use secure email," 2007. [Online]. Available: http://www.techdirt.com/articles/20070405/201336.shtml

[36] British Society of Criminology, "The british society of criminology. code of ethics." [Online]. Available: http://www.britsoccrim.org/codeofethics.htm