THE CHALLENGES OF STOPPING ILLEGAL PEER-TO-PEER FILE SHARING

Kevin Bauer, Dirk Grunwald, and Douglas Sicker
Department of Computer Science, University of Colorado

Abstract

Illegal file sharing within peer-to-peer networks has become a significant threat to the film and recording industries. Furthermore, such peer-to-peer protocols often use a significantly large amount of bandwidth relative to other protocols, complicating network management. In the past, copyright enforcement agencies have been hired to investigate users who appear to be sharing files illegally. In addition, broadband Internet service providers (ISPs) have actively throttled peer-to-peer traffic in an effort to reduce load on their networks. We observe that an "arms race" has begun between file traders and copyright holders/ISPs in which the file traders have started to develop techniques for hiding their involvement in the transfer of copyright-protected media files. In response, the copyright holders'/ISPs' investigative tactics are evolving to match the changing strategies. In this paper, we provide a survey of the current tactics used by file traders to hide their involvement in illegal file transfers and speculate about future strategies that may emerge on both sides of the arms race.

1. INTRODUCTION

Peer-to-peer (P2P) networks have recently grown in popularity for a variety of applications such as content distribution, streaming multimedia, and voice-over-IP. P2P networks are often built around a decentralized architecture to distribute data in a manner that offers high availability of content, inherent fault-tolerance, and efficiency. While P2P networks offer several important advantages over traditional client/server architectures, experience has shown that these networks are sometimes used to distribute copyright-protected media illegally.

P2P file sharing involving copyright protected content presents significant problems for network management and copyright enforcement. P2P networks utilize a large amount of bandwidth, particularly upstream bandwidth, complicating network management for broadband Internet service providers (ISPs), particularly during times of peak network utilization. In addition, the illegal dissemination of copyright-protected media is an obvious problem for the respective copyright holders that may result in a loss of revenue. As a consequence, there is ample incentive for both broadband ISPs and copyright holders to work to stop the proliferation of file sharing within P2P networks.

Our primary goal in this paper is to assume a proactive position toward understanding the current techniques for distributing and hiding copyright-protected content within P2P networks. We focus our discussion primarily on BitTorrent, since it is currently the most popular P2P protocol for file sharing. We observe that an arms race has already begun between file traders and copyright holders in which the file traders have started to develop techniques for hiding their involvement in the transfer of a copyright-protected media file. In response, the investigative tactics used by copyright holders are evolving to match these changing strategies. We provide a survey of the current tactics used by file traders to hide their involvement in illegal file transfers and speculate about future strategies that may emerge on both sides of the arms race.

The remainder of this paper is organized as follows: In Section 2, we provide an introduction to BitTorrent, the most common P2P network in use today. In Section 3, we describe the most common techniques that copyright holders have used to track the

distribution of their copyright-protected content. These strategies often include locating individual users, issuing DMCA takedown notices, or even pursuing more serious legal actions against suspected file sharers. We also discuss the past tactics used by broadband ISPs to throttle BitTorrent traffic. In response to the copyright holders' desire to protect their content, there is now significant incentive for P2P users to shed their network identities and enjoy a certain degree of anonymity. In addition, to avoid traffic shaping, P2P users have incentive to try to hide the nature of their traffic using encryption. We discuss the current tactics used to evade ISP traffic shaping practices and copyright enforcement in Section 4. In Section 5, we describe the most common methods for achieving anonymity online and present evidence from a prior study that P2P users are beginning to use BitTorrent anonymously. We also briefly outline prior proposals to incorporate anonymity mechanisms into P2P networks themselves. We also speculate about the future tactics that may be employed to distribute copyright-protected content. Finally, we provide concluding remarks in Section 6.

2. BACKGROUND ON BITTORRENT

BitTorrent has become one of the most popular peer-to-peer protocols for file sharing. A key feature of file transfers with BitTorrent is that files are not transferred sequentially, as in protocols such as HTTP or FTP. Instead, files are broken into fixed-size *pieces* and are transferred in parallel. This enables BitTorrent to transfer data very quickly and efficiently among a large number of peers. As a result, the protocol can be particularly greedy with regard to bandwidth.

To share a file with BitTorrent, a metadata file containing the piece length, a SHA1 hash of each piece to ensure integrity, and a URL to a *tracker server* is published through an

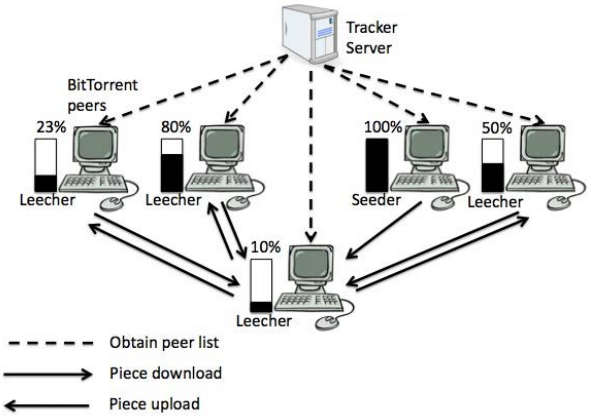

Figure 1: A file transfer with BitTorrent.

out-of-band mechanism. These metadata files are often hosted by sites such as isoHunt [4] and The Pirate Bay [8]. Once a peer obtains the metadata file for a desired file, the peer contacts the tracker server to obtain a list of other peers who are sharing the file. In the process, the peer also registers itself with the tracker. Other peer discovery mechanisms are available, including distributed trackers built upon distributed hash tables (DHTs) and gossip protocols; however, the centralized tracker server method is simple, and thus, the most commonly used. The peer finally issues requests for *blocks*, or sub-pieces (typically 16KB), from other peers. Peers who possess the complete file are called *seeders* and peers who do not are referred to as *leechers*. A file transfer using BitTorrent is illustrated in Figure 1.

The protocol's precise sequence of messages to initiate a data transfer is provided in Figure 2. First, a leecher establishes communication with another peer by exchanging *handshake messages*. The handshake consists of a plain-text protocol identifier string, a SHA1 hash that identifies the file(s) that are being shared, and a pseudo-random peer identification string. After both peers have exchanged handshake messages, the leecher sends a *bitfield message*, which contains a bit-array data structure that concisely describes the pieces of the file that the peer has already obtained. After

exchanging bitfields, the leecher knows which pieces the other peer can supply, and it proceeds by requesting specific blocks. Once a leecher has obtained a piece, it notifies other peers by sending a *have message*. More details about BitTorrent can be found in its protocol specification document [1].

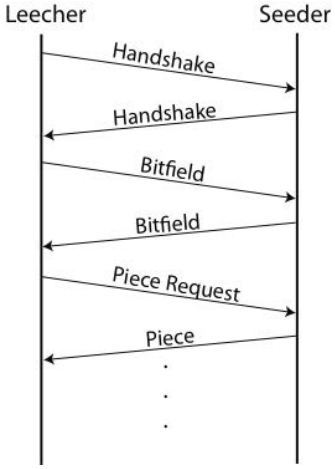

Figure 2: BitTorrent's message exchange to initiate a piece transfer.

Due to its aggressive behavior with regard to bandwidth usage (often described as *swarming* behavior) BitTorrent presents significant network management challenges for ISPs. BitTorrent is often configured to open many TCP connections simultaneously, sometimes using all bandwidth available to the user. For ISPs, this behavior may complicate network management, especially during times of peak utilization. As a result, many ISPs have actively attempted to regulate BitTorrent's bandwidth usage. In the next section, we discuss the common tactics used in the past to investigate copyright violation and to control bandwidth consumption.

3. INVESTIGATIVE TACTICS

BitTorrent is not used solely for copyright violation. There are many legitimate uses including obtaining software updates, downloading Linux ISO images, and sharing non-copyright protected movies and music. However, given the unfortunate reality that BitTorrent is being used to distribute copyright-protected movies and music, BitTorrent has caught the attention of organizations such as the Motion Picture Association of America (MPA) and the Recording Industry Association of America (RIAA). Furthermore, as a consequence of BitTorrent's aggressive network behavior, it consumes excessive amounts of bandwidth relative to other protocols, thus complicating network management for broadband ISPs. Since copyright holders and network operators *both* have incentive to curtail BitTorrent usage (though for different reasons), both have initiated campaigns aimed at slowing the proliferation of BitTorrent usage. In this section, we present an overview of the strategies employed by broadband network operators and entities representing copyright holders.

3.1 Broadband ISP Tactics

Due to the challenges that BitTorrent presents for network management, some broadband ISPs have recently adopted policies aimed at disrupting or even blocking BitTorrent traffic within their networks [3]. In particular, Comcast received extensive publicity for their use of Sandvine to specifically target BitTorrent flows with forged TCP RST (reset) packets, causing a targeted TCP connection to be prematurely and abruptly closed. This policy has been criticized by network neutrality proponents and consumer advocates in part because there was little transparency and disclosure regarding these practices. In response, researchers have produced a variety of techniques and tools [12, 21, 26] to detect this type of traffic manipulation by ISPs.

3.2 Copyright Holder Tactics

Since the tracker servers that enable illegal file transfers are often hosted in foreign countries where legal recourse against such activity is limited [11], the representatives

such as the MPA and RIAA acting on behalf of copyright holders have initiated a large scale investigative effort to identify and pursue individual users participating in illegal file transfers. Such companies as Media Defender [5] and Safenet [10] have been hired to passively monitor the tracker servers for copyright infringing file transfers to obtain the list of IP addresses of the users who are participating in the file transfers. Recall that BitTorrent's primary peer discovery mechanism requires that the IP addresses of other peers participating in the file transfer be publicly advertised.

A recent study [25] found that these investigators obtain the list of IP addresses from the trackers and send an ICMP echo (ping) message to each end-host to ensure that it is alive. These investigators often target suspected file sharers with DMCA takedown notices and even have initiated more formal legal proceedings in some cases.

However, as the authors of [25] observe, this type of investigative strategy is problematic, since it is easily prone errors, especially false positive identification. False positives occur when users are wrongly accused of actively participating in the file sharing. False positives may occur as a result of normal network activity, for example, if a user obtains a DHCP lease on an IP address that had previously participated in the file transfer. However, false positives may also occur by actively polluting a particular tracker's peer list with arbitrary IP addresses. It is possible to explicitly register arbitrary IP addresses to a tracker, thus implicating any end-host in the file sharing. The authors of [25] poignantly demonstrated the shortcomings of the current investigative tactics by registering devices such as networked printers and wireless access points to tracker lists, and subsequently receiving DMCA takedown notices for these devices' alleged involvement in illegal file transfers.

Another study has found that representatives of the copyright holders actively participate in illegal BitTorrent file transfers and attempt to launch a variety of "attacks" on leechers [19]. In particular, this study identified two distinct attack strategies: *fake-block* and *unresponsive peer* attacks.

The fake-block attack occurs when peers operated by copyright enforcers deliberately reply to piece requests with invalid blocks of data. When an entire piece is obtained, the leecher verifies the piece's integrity with a SHA1 hash. However, the hash fails due to the invalid block(s). This requires the leecher to download the entire piece again (which wastes time and bandwidth), since it does not know precisely which block is corrupt.

The unresponsive peer attack occurs when a peer completes a valid BitTorrent handshake and bitfield exchange (which is the prelude to the data transfers), but the peer refuses to send any data. This attack also causes leechers to waste time and bandwidth exchanging control messages with peers that have no intention to provide pieces of the file.

The aforementioned study found that both of these attacks are relatively common. In addition, while these attacks may cause a download to take up to 50% longer, they are ineffective at stopping BitTorrent file transfers altogether.

4. RESPONSE TO ANTI-P2P CAMPAIGNS

Given the techniques used to mitigate BitTorrent usage by network operators and copyright holders, file sharing tactics have begun to evolve to incorporate mechanisms to prevent blocking by ISPs and to avoid legal sanctions by entities representing the copyright holders. In this section, we provide an overview of the next phase in the arms race between the file sharers and ISPs/copyright holders.

4.1 Concealing BitTorrent from ISPs

In an attempt to frustrate traffic shaping or blocking by ISPs, an obfuscation technique called Message Stream Encryption has been proposed as an optional extension to the

BitTorrent protocol [6]. Message Stream Encryption requires that pairs of communicating peers perform a Diffie-Hellman key exchange to agree on a shared secret and then encrypt the BitTorrent header (and optionally the payload) using the RC4 stream cipher. This feature is available in Vuze [14], µTorrent [13], and other BitTorrent clients. In order to use the encryption feature, a peer can only communicate with other peers that support the encryption feature.

However, protocol header encryption and payload encryption are relatively ineffective at obfuscating the traffic type, since the packet size characteristics remain intact. BitTorrent traffic has a distinctive signature consisting of large bidirectional data transfers, thus it would still be relatively easy to detect despite encryption. Furthermore, sophisticated techniques based on statistical or machine learning methods could be applied to detecting if an encrypted stream is BitTorrent traffic [24, 27, 28]. Encrypting BitTorrent does, however, require the ISP to develop and deploy these types of sophisticated traffic classification techniques, which may be expensive and time consuming. Furthermore, these traffic classification techniques are not perfect and may have non-negligible classification errors. This could result in a scenario in which other protocols are misclassified as BitTorrent.

In addition to encryption, it is possible that BitTorrent may adopt a UDP transport mechanism, which is rumored to be included to a future version of µTorrent [13]. The UDP transport would render the traffic shaping practices using forged TCP RST packets ineffective.

4.2 Evading Copyright Authorities

Since the large scale investigations carried out by entities representing copyright holders have resulted in DMCA takedown notices and the potential for more serious legal sanctions, counter-strategies have emerged in an attempt to frustrate these investigations. One common strategy is to intentionally introduce randomly selected IP addresses into the tracker lists (called pollution). For instance, the Pirate Bay, a popular tracker-hosting site, has implemented this policy [9]. This tactic increases the potential for false positives to a level that may not be tolerable for the investigators. For instance, wrongly accusing innocent users of sharing files illegally could have serious consequences for the copyright holders including negative public opinion or even sanctions from government regulators.

In addition, services such as PeerGuardian [7] have emerged to provide IP address blocking capabilities for P2P applications. For instance, this service could be used to block all IP addresses that are suspected of active pollution or monitoring.

More extreme techniques to evade the copyright enforcement authorities are even starting to become common. For instance, BTGuard [2] offers a pay proxy service in which subscribing users can encrypt and tunnel their BitTorrent traffic through a proxy server hosted in a foreign country. Using such a service, when a BitTorrent client registers itself with a tracker server, the tracker server knows only the proxy's IP address, and consequently, the copyright enforcers also can observe only the proxy's IP address. Provided that the proxy service does not keep records of its clients' activity, it is difficult to determine the identity of the real client. The encrypted tunnel may also frustrate ISPs' BitTorrent traffic throttling, but as described in Section 4.1, traffic analysis techniques exist that may reveal the underlying type of traffic within the encrypted flow.

5. EMERGING STRATEGIES TO HIDE P2P

The changing tactics employed by file sharers and copyright holders/ISPs can best be described as an arms race of evolving strategies and counter-strategies. In this section, we discuss the current cutting-edge and possible future strategies that P2P users

may apply to obfuscate their activities from their ISPs to avoid traffic throttling and to hide from copyright enforcement authorities. Technologies that enable end-users to shed their network identities and enjoy anonymity while online are one line of emerging strategies. We present evidence to suggest that P2P users may be beginning to use anonymous networks to avoid traffic throttling by their ISPs and avoid identification and subsequent legal action by copyright enforcement authorities.

5.1 P2P and Anonymous Networks

The Internet and its fundamental protocols (*i.e.*, TCP/IP) were designed with no regard for anonymous network access. However, recent research in anonymous communications has provided the designs and implementations of anonymous overlay networks based on onion routing [22] and mix networks [17]. Anonymous networks are currently being used throughout the world for a variety of applications, often enabling freedom of speech and press within repressive countries.

Tor has become the most popular overlay network for anonymizing TCP-based applications [20]. Tor is able to provide a stronger form of anonymity than the proxy server approach (described in Section 4.2) because it is built around a decentralized design; therefore, no single entity knows both the source and the destination of an anonymous flow. Tor's system architecture (illustrated in Figure 3) consists of three components: *Tor routers*, *Tor proxies*, and *directory servers*. Tor routers forward TCP traffic on behalf of participating users by employing a layered encryption scheme similar to onion routing. A user running Tor proxy software creates a virtual circuit of precisely three Tor routers. First, the Tor proxy obtains a list of all available Tor routers from the set of trusted directory servers. Next, the Tor proxy establishes shared secret keys with each of the three Tor routers on the circuit and encrypts the user's data with each key in a layered fashion. Upon receiving a packet, the Tor router removes its layer of encryption and forwards the packet to the next router in the path. Once the final layer of encryption has been removed, the last Tor router forwards the payload to the destination server.

It is important to note that only the first Tor router on the path (called the *entry guard*) knows the true identity of the client, and only the last Tor router on the path (called the *exit router*) knows the identity of the destination server. Tor provides a strong degree of anonymity, subject to the assumption that it is difficult for a single entity to control both the first and last Tor routers on a user's virtual circuit [15]. However, an ISP or group of colluding ISPs *could* feasibly monitor the links entering and exiting the Tor network and perform traffic analysis to link the clients and destinations.

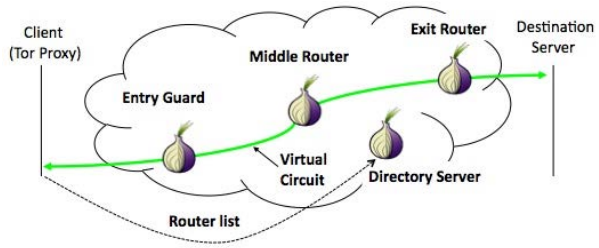

Figure 3: Tor's system architecture.

In prior work, we characterized how Tor is used in practice [23]. In particular, we analyzed the application-layer protocols that are commonly used with Tor. We discovered that individual users are starting to use Tor to conceal BitTorrent activity. While operating a Tor router for four days, we observed over 430,000 BitTorrent connections leaving the Tor network, accounting for approximately 285GB of traffic. While the number of BitTorrent connections was relatively low in comparison to other protocols such as HTTP and SSL, the amount of traffic transported over these connections was surprisingly high. However, there are plug-ins for popular BitTorrent clients (such as Vuze) that make it easy to connect the BitTorrent client to the

Tor proxy software. In addition, given the past practices of monitoring and profiling users suspected for participating in illegal BitTorrent file transfers, it is reasonable to suspect that the number of users who turn to strong anonymity mechanisms like Tor may increase in the future.

In addition to anonymizing overlay networks like Tor, it is possible that P2P users may look to other sources for anonymity. For example, the design of an anonymity layer specifically tailored for BitTorrent has been published [16]. The protocol, called BitBlender, works by introducing special peers called *relay peers* into the BitTorrent system architecture. These peers do not actively share any file(s), but merely proxy piece requests and responses on behalf of other users actively sharing the file(s).

BitBlender's primary goal is to introduce a certain degree of *plausible deniability* for peers listed by the trackers. With BitBlender, a copyright enforcement authority cannot simply examine the tracker's peer list to obtain an accurate view of the peers who are involved in the sharing. The copyright enforcer must actively participate in the file sharing and conduct sophisticated traffic analysis in order to have any chance of isolating the real active peers. However, since the relay peers exhibit many of the same protocol-level behaviors as the real peers, it may still be difficult to isolate the real peers. While BitBlender is only a proof-of-concept design (*i.e.*, there is currently no available implementation), it is possible that this relay strategy may be incorporated into popular BitTorrent clients in the future.

5.2 End Game

Until this point, we have discussed the current and emerging strategies used for hiding illegal file sharing within P2P networks. Next, we examine how the shifting strategies used to stop this type of file sharing may cause a radical shift in content hiding strategies and provide a speculative discussion of the tactics that may be used by file sharers in the future.

One potential technique for hiding content is to use a distributed and anonymous data store. Freenet [18] is a P2P network in which peers can store and retrieve files that are named by location-independent keys. To retrieve a file, a user computes a hash of the content's description - which is used as the look-up key - and forwards a retrieval request to another peer in the network. The request is forwarded through potentially many peers until the content is found, upon which, the content is sent back to the original requester through each peer that forwarded the initial request. In doing so, the replying peer does not know who actually initiated the request, and the requesting peer does not know where the data is stored. Furthermore, peers hosting files only know the hash of the file's description, so they remain agnostic regarding the content they host.

This content hiding strategy offers significant advantages over BitTorrent. The Freenet-style of content hosting and retrieval offers relatively strong deniability for both the hosts and the retrievers. Furthermore, this strategy significantly complicates investigations launched by anti-piracy agencies.

In addition to Freenet-style P2P networks, Tor offers the ability to host hidden services within the Tor network. A hidden service can be established in such a manner that the service's owner does not reveal their identity. It is difficult to shut down such a service, since its location is hidden. More details on hidden services in Tor can be found in Tor's design document [20].

Tor’s hidden services provide strong anonymity for both the service's host and those who download content, and represent perhaps the most *radical* counter-measure to anti-piracy efforts. While there is a significant performance penalty associated with using hidden services (*i.e.*, additional download time), users may be willing to cope with this limitation if there is sufficient incentive,

perhaps such as avoiding prosecution. If widespread usage of Tor's hidden services for illegal file sharing becomes a popular counter-strategy, there may be little recourse for anti-piracy authorities to stop it.

6. CONCLUSION

In this paper, we presented an overview of the current strategies for identifying illegal file sharers and a survey of the counter-measures that file sharers have employed in response. We observe that a strategic "arms race" has started as the tactics for pursing illegal file sharers and hiding evolve. In addition, if this arms race continues, we speculate about the future tactics that may be used to hide illegal file sharing and conclude that strong anonymity mechanisms and location-hidden services may be the final resort of the illegal file sharing movement. Since this implies a somewhat bleak outlook for the anti-piracy authorities, we conclude that alternative strategies – including tiered bandwidth pricing models to discourage high bandwidth usage on broadband networks and lower-cost media distribution methods – should be investigated to provide individuals with more economic incentives to obtain content from legitimate sources.

REFERENCES

[1] BitTorrent protocol specification. http://wiki.theory.org/BitTorrentSpecification.

[2] BTGuard - BitTorrent Anonymously. http://btguard.com.

[3] Comcast is using Sandvine to manage P2P connections. http://www.dslreports.com/forum/r18323368-Comcast-is-using-Sandvine-to-manage-P2P-Connections.

[4] isoHunt - the BitTorrent and P2P search engine. http://isohunt.com.

[5] Media Defender - P2P anti-piracy and P2P marketing solutions. http://www.mediadefender.com.

[6] Message stream encryption. http://www.azureuswiki.com/index.php/Message_Stream_Encryption.

[7] Peer Guardian 2. http://phoenixlabs.org/pg2.

[8] The Pirate Bay. http://thepiratebay.org.

[9] Pirate Bay tricks anti-pirates with fake peers. http://torrentfreak.com/the-pirate-bay-tricks-anti-pirates-with-fake-peers-081020.

[10] Safenet Inc: The foundation for information security. http://www.safenet-inc.com.

[11] Secrets of the pirate bay. http://www.wired.com/science/discoveries/news/2006/08/71543.

[12] Switzerland network testing tool. http://www.eff.org/testyourisp/switzerland.

[13] µTorrent - the lightweight and efficient BitTorrent client. http://www.utorrent.com.

[14] Vuze HD network. http://www.vuze.com/app.

[15] K. Bauer, D. McCoy, D. Grunwald, T. Kohno, and D. Sicker. Low-resource routing attacks against Tor. In Proceedings of the Workshop on Privacy in the Electronic Society. October 2007.

[16] K. Bauer, D. McCoy, D. Grunwald, and D. Sicker. BitBlender: Light-weight anonymity for BitTorrent. In Proceedings of the Workshop on Applications of Private and Anonymous Communications. September 2008.

[17] D. Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. Communications of the ACM, February 1981.

[18] I. Clarke, O. Sandberg, B. Wiley, and T. W. Hong. Freenet: A distributed anonymous information storage and retrieval system. In Proceedings of Workshop on Design Issues in Anonymity and Unobservability. July 2000.

[19] P. Dhungel, D. Wu, B. Schonhorst, and K. W. Ross. A measurement study of attacks on BitTorrent leechers. In Proceedings of the International Workshop on Peer-to-Peer Systems (IPTPS). February 2008.

[20] R. Dingledine, N. Mathewson, and P. Syverson. Tor: The second-generation onion router. In Proceedings of the 13th USENIX Security Symposium. August 2004.

[21] M. Dischinger, A. Mislove, A. Haeberlen, and K. P. Gummadi. Detecting BitTorrent blocking. In IMC '08: Proceedings of the 8th ACM SIGCOMM conference on Internet measurement. October 2008.

[22] D. M. Goldschlag, M. G. Reed, and P. F. Syverson. Hiding routing information. In Proceedings of Information Hiding: First International Workshop. May 1996.

[23] D. McCoy, K. Bauer, D. Grunwald, T. Kohno, and D. Sicker. Shining light in dark places: Understanding the Tor network. In Proceedings of the 8th Privacy Enhancing Technologies Symposium. July 2008.

[24] A. W. Moore and D. Zuev. Internet traffic classification using Bayesian analysis techniques. In SIGMETRICS. 2005.

[25] 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 3rd USENIX Workshop on Hot Topics in Security, July 2008.

[26] N. Weaver, R. Sommer, and V. Paxson. Detecting forged TCP reset packets. In Proceedings of NDSS. February 2009.

[27] C. Wright, F. Monrose, and G. Masson. On inferring application protocol behaviors in encrypted network traffic. Journal of Machine Learning Research. 2006.

[28] S. Zander, T. Nguyen, and G. Armitage. Automated traffic classification and application identification using machine learning. In LCN. 2005.