Piracy Activity vs. Product Features in Digital Games

Anders Drachen
Department of Informatics
Copenhagen Business School
ad.inf@cbs.dk

Kevin Bauer
Cheriton School of Computer Science
University of Waterloo
k4bauer@cs.uwaterloo.ca

Robert W. D. Veitch
Department of Informatics
Copenhagen Business School
rv.inf@cbs.dk

ABSTRACT

The practice of illegally copying and distributing digital games is at the heart of one of the most heated and divisive debates in the international games environment, with stakeholders typically viewing it as a very positive (pirates) or very negative (the industry, policy makers). Despite the substantial interest in game piracy, there is very little objective information available about its magnitude or its distribution across game titles and game genres. This paper presents a large-scale analysis of the illegal distribution of digital game titles, which was conducted by monitoring the BitTorrent peer-to-peer (P2P) file-sharing protocol. The sample includes 173 games and a collection period of three months from late 2010 to early 2011. With a total of 12.6 million unique peers identified, it is the largest examination of game piracy via P2P networks to date. Analysis of the data shows that games of the "Action" genre, which include the majority of major commercial AAA-level titles, comprise 45% of the unique peers in the dataset, although games from "Racing", "Role-Playing Game" and "Simulation" genres have higher numbers of unique peers on average than "Action" games. The ten most pirated titles encompass 5.27 million aggregated unique peers alone. In addition to genre, review scores were found to be positively correlated with the logarithm of the number of unique peers per game ($p < 0.05$).

General Terms

Economics, Security, Human Factors, Legal Aspects.

Keywords

Digital games, game piracy, BitTorrent, game factors, economics of game piracy.

1.   INTRODUCTION

Game piracy, which involves the illegal copying and distribution of digital games [28], is a complex phenomena that occurs across multiple channels and has a magnitude that is difficult to estimate [9,17,18]. It is the cause of heated debate, with pirates on one side and game developers, game publishers and legislators/policy makers on the other. Despite the interest in game piracy and the controversy surrounding the activity, there is only limited information available on the subject that spans across game titles. The information that does exist often comes from industry organizations such as the Entertainment Software Association (ESA) or operators of Peer-2-Peer (P2P) networks [e.g., 31], but lacks objectivity and a transparent methodology.

The purpose of this paper is to address the need for objective information on digital game piracy, covering two important areas. Firstly, we provide a much needed overview and analysis of the current state-of-the-art of the game piracy debate and a thorough evaluation of the BitTorrent protocol and the issues related to monitoring it, which are commonly ignored in reports on digital piracy [5]. Secondly, we report analyses on game piracy data obtained via tracking of the BitTorrent P2P protocol. The BitTorrent network was chosen because it is regarded as one of the main channels for game piracy and the de facto standard for distribution

of digital files via P2P networks [5,31]. Monitoring was carried out for 173 game titles over a three month period running from late 2010 to early 2011, spanning all genres and multiple hardware platforms (e.g., PC, X360, PS3, Wii, DS, PSP), and the recording of 12.7 million unique peers. The analysis of these data indicates that the major commercial, AAA-level, action-oriented titles account for the highest proportion of activity on the BitTorrent network, although games from other genres such as "Racing", "Role Playing Game" and "Simulation" games are individually more popular than the major action titles. The distribution of the torrent activity across game titles was highly asymmetric. For example, the 10 most popular titles comprised 41.5% of the total number of unique peers in the dataset. Finally, aggregated review scores were found to be positively correlated with the logarithm of the number of unique peers per game ($p < 0.05$). This indicates review scores may be one determinant of BitTorrent activity, irrespective of the hardware platform.

2.   GAME PIRACY: AN OVERVIEW

A key problem in the game piracy debate is the lack of comprehensive and objective information about the nature and magnitude of the piracy activity and its root causes, such as its economic and behavioral drivers. The majority of the data available on game piracy originate from the industry (e.g., individual publishers or developers [12,19]) but more commonly branch organizations such as the ESA and the Business Software Alliance (BSA) [9,17,18]. The data reported by the industry are somewhat problematic, partially due to the interest of the industry to reduce piracy and thus potentially exaggerate the problem, as well as a lack of transparency about the specific procedures for collecting the reported data [15].

2.1 Previous Work

Although the research literature on digital game piracy is fragmented, the information available on other digital products is better established. For example, software piracy has been investigated at both the individual- and country-level, and informed by theory from various fields, including economics, social psychology, criminology, business ethics and marketing. Individual-level research indicates the importance of intentions, attitudes, perceived risks, and price of legal alternatives (e.g., [Peace & Galletta, 1996]. Work focused on country-level differences using aggregate indices to gauge national differences indicates that gross domestic product per capita, investment in information communications technologies and civil liberties are negatively related to software piracy rates [e.g., 32].

Both sides of the piracy debate are however more or less in agreement when it comes to the size of the problem. For example, ESA claimed that 9.78 million illegal downloads had occurred of roughly 200 digital games in December 2009 alone [17]. TorrentFreak.com, one of the most influential websites on P2P sharing via the BitTorrent protocol, reported 18.14 million downloads for the five most downloaded PC games on BitTorrent in 2010. The five most downloaded console games add a further 5.34

million downloads [31]. The lack of methodological scrutiny by objective third parties is also a problem for government-solicited research.

The credibility of industry-based or government-solicited reports is also harmed by the common methodological problems in such work, which foster suspicion of bias [e.g., 14,15,30]. As noted by Anderson [2011]: "When a TV/movie company like NBC Universal funds a P2P study from a company that specializes in antipiracy work, the end result is hardly a disinterested piece of data" [para. 4]. For example, Huygen et al. [15] examined music, film and game piracy in the Netherlands, and via an online survey of non-randomly selected Internet users in the Netherlands (n=778), extrapolating their findings to make conclusions about the entire Dutch population. Similarly, Envisional [13], a piracy research firm, concluded in a technical report commissioned by NBC Universal that among other things music piracy is virtually gone from BitTorrent, with only 2.9% of the 10,000 "most popular torrents" examined being music files (console games comprising 2.8% and PC games 3.9%). However, these conclusions are based on a somewhat flawed snapshot methodology, which is likely to overestimate the popularity of larger files, such as games, which take much longer to download and underestimate the popularity of smaller files, such as music, which take much less time to download.

The potential impact of digital piracy on industry is notoriously hard to estimate reliably [14,15,30]. However, according to the BSA [9] piracy of digital products is on the rise, with global software piracy in 2009 rising two percent, representing a total 51.5 billion USD in lost revenue. A reliable figure for digital games is – to the best knowledge of the authors - unknown.

2.2 Distribution Channels
A key challenge for investigations of game piracy is that the channels through which illegal copies are distributed (e.g., physical copying, peer-to-peer networks) are almost impossible to monitor effectively [2,5,6,17].

Distribution channels digital material include physical as well as networked solutions, with the most common online being P2P protocols [17] and "one-click" file hosting services and the copying and distribution of digital material on physical media offline. Distribution channels such as Usenet, File Transfer Protocol [FTP] and Internet Relay Chat (IRC) were arguably important previously, but have become less frequently used since the emergence of P2P protocols a bit over a decade ago [6], and are thus covered in less detail here.

Developing reliable estimates of the piracy activity that occurs via any channel of distribution is challenging, however, the BitTorrent file exchange protocol is generally viewed as the major channel for game piracy today [17,18]. This conclusion is confirmed by reports highlighting that P2P-based traffic comprises a sizeable fraction of the traffic on the Internet, with estimates varying from 40-60% [e.g., 26].

2.2.1 Usenet
Usenet is a decentralized network launched in the 1980s to permit the sharing of conversations before the development of web forums. Usenet forms a source of piracy via uploading of files rather than messages to newsgroups. The files are retrieved using newsreader clients. While Usenet in the past may have played a role in file distribution, it is likely not a factor in the contemporary game piracy.

2.2.2 FTP
The File Transfer Protocol is specifically designed for serving files over the Internet. FTP servers are centralized and therefore prone to being shut down by legal entities if discovered to host illegal content. FTP access is usually restricted to a small group and hence not a common piracy method for the general public.

2.2.3 IRC
The Internet Relay Chat method was developed in the late 1980s to facilitate real-time communication, before the development of instant messaging clients. IRC can be used to share files, but downloading files over IRC can be technically complicated compared to other methods, and may involve long waiting periods.

2.2.4 Physical Distribution
Pre-dating the internet, the illegal physical copying, distribution or selling of software, including digital games, is the oldest form of digital piracy and has existed for as long as digital content has been available on portable media. Physical piracy involves the distribution – selling, giving or swapping –of unauthorized physically copied media, including game DVDs and CDs. The distribution networks employed vary from small circles of friends to organized crime where businesses revolve around the mass duplication of pirated media, and distribution/sale at compretitive prices [17]. In some cases, illegally duplicated software is hard to distinguish from the original (legitimate) versions. The magnitude of this piracy channel is difficult to estimate with any degree of accuracy, but is common in certain countries, such as Italy and Hong Kong, who are on the watch list of the International Intellectual Property Association [17].

2.2.5 File-hosting Services
File-hosting services (e.g., RapidShare.com, MegaUpload.com) consist of servers to where digital content can be uploaded to and downloaded from, provided that the user has access rights to do so. While the use of file hosting services can be legitimate, this type of service can also be used to host and share illegally copied digital material, as shown by Antoniades et al. [1] who exposed how "one-click" hosting services offer copyright-protected content. Users of file hosting services can access such material simply by searching for content of interest. The use of one-click hosting services can be appealing to people wishing to distribute illegally copied content because they require only very limited technical knowledge to access. For example, standard HTTP web links to specific files can be shared and search for using standard web browsers. Recently, Maier et al. [21] noted a shift in the distribution of Internet traffic, claiming that the majority of Internet traffic by volume is a result of streaming media websites (e.g., youtube.com) as well as file hosting services.

2.2.6 P2P Protocols
P2P protocols enable end-users to share content with one another directly, eliminating the need for uploading digital content to a central server for mass distribution, as is the case with file-hosting services. Various P2P protocols have been developed over the past decade, including Gnutella, FastTrack and BitTorrent. Contrasting with the server-client dissemination models, these protocols allow users to act as hosts of digital content as well as users of digital content. This means that the amount of bandwith available for other peers in the network scales with the number of peers (or users) who host the entire piece of digital content or just part of it.

2.3 Legal Entities Monitoring BitTorrent
The BitTorrent protocol publicly shares information about the peers that access the network (or "swarm") and engage in file

sharing. Thereby, the IP addresses of the participants [20] can be obtained by querying the trackers used to provide information about specific torrents [23] or via crawling the BitTorrent Distributed Hash Tables (DHT) [30].

In contrast to the relatively simple legal process involved in removing content from "one-click" services, P2P protocols make the task responding to reports of piracy more difficult, as each individual peer who is participating in the sharing must be identified and contacted. From a legal standpoint, if copyright-protected content can be shown to have been uploaded to a service, the operators can be forced (through appropriate legal channels) to comply with requests to remove the material. Despite the challenge of finding and contacting peers hosting specific illegal digital content, entities acting on behalf of copyright holders have attempted to monitor BitTorrent file transfers on a massive scale [23]. This has led to an arms race, where P2P network operators and copyright holders respectively attempt to circumvent the opponent's methods for baffling and penetrating torrent networks [30]. However, Piatek et al. [23] showed how the techniques commonly employed are prone to a wide variety of errors, leading to highly inaccurate estimates of piracy activity and, worse, in some cases, falsely accusing innocent people of violating copyright laws.

Irrespective of the attempts by copyright investigators to employ techniques such as the above to identify users engaged in copyright violation via P2P networks, the inherent public nature of the BitTorrent participants makes it an ideal measurement platform for obtaining concrete empirical data on game piracy. Notably, with the recent evolution of BitTorrent, which has seen the incorporation of additional mechanisms for peer identification (in addition to the centralized tracker servers), such as DHTs, as well as a gossip-based mechanism called Peer Exchange (PEX) [see e.g. 7]. These features make it easier to identify peers.

3. METHOD

In order to obtain the data necessary for the analysis of game piracy, two data streams are necessary: 1) BitTorrent data on the online distribution of illegal copies of digital games; 2) Information about the products. The steps towards obtaining these data are comprised of a series of steps, as follows:

3.1 Obtaining Unique Peers from BitTorrent

The BitTorrent protocol basically works by breaking down files that peers (users) seek to share into many pieces of a specific size. Each of these is described via cryptographic hashes contained in a metadata file, together with additional information such as a Uniform Resource Identifier (URI), and shared with a tracker server that lists all participating peers [5,6]. The metadata file is distributed to the users via the tracker server. BitTorrent search engines (e.g., thepiratebay.org and isohunt.com), host the metadata files and provide a search capability for peers. Upon having obtained the metadata for a particular file (or "torrent"), peers can access the file via client-side BitTorrent software contacting the tracking server to obtain a randomly selected subset of the users currently sharing the file. Importantly, in the process of obtaining a peer list from the server, the peer implicitly registers itself with the tracker, enabling other peers to contact it and request parts of the file (when these are available, i.e. have been downloaded).

For the current study, a list of 173 game titles was compiled across genres and hardware platforms including Xbox360, PlayStation 3, Nintendo Wii, PC, Nintendo DS (DS) and PlayStation Portable (PSP). This list included a series of games released in the fall 2010, and every game released since November 17th 2010 until 29th January 2011, the end of the tracking period (note that games can appear on BitTorrent prior to the official launch date). The sample consisted of games from all genres (irrespective of the specific definition system), ranging from AAA-level major commercial titles (e.g., Bioshock 2, Need for Speed: Hot Pursuit, Civilization 5, Little Big Planet 2, Fallout: New Vegas) to casual and indie games (e.g., Auditorium, Super Meat Boy, Majin and the Forsaken Kingdom). Of these 173 titles, within the period of tracking (November 17th 2010 to February 6th 2011), 127 were located on BitTorrent, indicating these games had been cracked of any copyright protection (DRM, Digital Rights Management), and released on BitTorrent. During the period of tracking, 12.6 million unique peers were identified for all these games, making this the to date largest study of BitTorrent-based game piracy, surpassing even the report of the Entertainment Software Association (ESA) from 2009, who reported 9.58 million downloads for about 200 unspecified titles produced by members of the ESA during one month in late 2009 (no detailed information has been revealed, including methodology and the specific titles involved). Forty of 127 titles made their first appearance on BitTorrent during the period of tracking (comprising 1.16 million unique peers).

In order to obtain the BitTorrent metadata files for the 173 game titles, a custom web crawler was developed. The web crawler periodically issues queries to a popular BitTorrent search engine website for each title, extracting the metadata files. Having located the metadata files, the web crawler obtains the tracker server URIs. Having obtained the URIs, an HTTP GET request is issued to each URI over periodic intervals (the tracker servers are queried every few minutes) to obtain a list of IP addresses for the peers who are currently participating in sharing the specific file.

When searching for game torrents, a number of false positives can occur as torrents related to a game or with a similar name may not contain the full game. This is notably the case with key generators, .exe-files, expansion content, game soundtracks, graphical material or material of another type with similar titles. In order to eliminate such torrents, all torrent lists for each game title were manually inspected and filtered, leaving only the torrents that contain the full game. It is a common practice for copyright investigators to spread false information on BitTorrent in an attempt to make it difficult for peers to access copyrighted material [6]. In order to filter out the false IP addresses, standard publicly available blacklists were applied to filter out IP address blocks controlled by well-known copyright investigators, thus ensuring that the final set of unique peers consists only of real peers who are actively engaged in file sharing (Available from: http://www.iblocklist.com/list.php? list=bt_level1).

3.2 Obtaining Game Feature Data

A substantial challenge for the monitoring of newly released digital games on P2P networks and other piracy channels is that the game developers and publishers often do not adhere to announced release dates. This practice makes it very difficult to determine when tracking of a specific game title should begin. Given the propensity for digital games to be available on BitTorrent before the official release date, the best approach is to start tracking the game as soon as a title is reported nearing completion (e.g., beta-testing stage). This is important when game publishers choose to release a game on different dates in different regions of the world (e.g., North America first, then Europe and Asia). The time delay gives hacker groups time to crack a game's copyright protection, and upload the game to BitTorrent networks. Unlike media types such as movies and music, there are no central reposi-

tories for information on digital games, covering release dates, sales figures, product features etc. Instead, a wide variety of websites attempt to provide parts of this information with greater or lesser degrees of accuracy (e.g., Metacritic.com, gamerankings.com, gamestats.com, vgnchartz.com, gamespy.com, ign.com, gamespot.com). Mining these sites and aggregating the information derived from them forms the current best approach towards obtaining the most reliable information possible.

3.2.1 Genre
Another challenge towards defining product features is the variety of digital games and the many different systems for categorizing games into genres. Game genre systems are nebulous at best, and therefore a similar aggregation approach was adopted here to build a genre system based on majority consensus. In practice, a variety of websites (e.g., mobygames.com, ign.com, gamespy.com, metacritic.com) were mined and genre definitions for the individual titles based on majority definitions. Of the genre categories, "action games" is perhaps the most problematic. Of the genre categories, "action games" is perhaps the most problematic. Across the mentioned websites, "action games" (or derivates thereof, e.g., "action adventure", "action shooter" etc.) are those that employ a First-Person or Third-Person camera perspective and where shooting at entities and objects forms a main element of the game, in addition to navigation, some puzzle solving and NPC interaction. Examples include games such as Kane & Lynch, Grand Theft Auto, and Metro 2033. In contrast, games such as Monday Night Combat and Team Fortress 2, revolve around shooting the avatars of other players and not much else, and can therefore be more reliably classified as "shooters". Another way to distinguish the "action" game genre from other genres is that it is typically in this category that the major commercial titles for both PC and consoles (outside of sports/singing/fitness etc. games), are to be found. Examples include Call of Duty: Black Ops, Bioshock 2, Darksiders, Medal of Honour, and Splinter Cell. Based on information mined, game titles in the study were divided into 12 different genres, with "action" being by far the best represented (Figure 2).

3.2.2 Aggregated Review Score
In order to obtain a measure of the quality of a game title, aggregated review scores were obtained from several recognized metacritic sites (metacritic.com, gamerankings.com, gamestats.com). Not all the games in the sample were available on all three of these sites. Aggregated scores could not be found for 15 of the 127 torrented games. 10 of these were commercially small titles (e.g., Stardrone and Brain Puzzles 2). For one of these titles, aggregated review scores could be built manually by recording review scores from game sites such as gamespy.com and ign.com. The remainder was eliminated from any analysis involving review scores. Average review scores range from 26 (Deca Sports Freedom) to 94.67 (Mass Effect 2), with a mean score of 70.44 and Std. Dev. = 15.83 (n=113).

3.3 Assumptions and Limitations
The dataset presented here represents a comprehensive 3-month snapshot of BitTorrent activity for the game titles. However, a few assumptions are inherent in the dataset, as follows:

3.3.1 Sample of Games
Roughly 1500 digital games are launched on a yearly basis [24], but it is not unknown how many of these are cracked and released as torrents on P2P networks. This means that it cannot be estimated how large a percentage that the current sample of 173 games comprise in relation to the 1500 titles, and thus the repre-

sentative strength of the data presented here. This is a subject for future research.

3.3.2 BitTorrent
The BitTorrent protocol is generally regarded as the standard for distribution of files via P2P-networks [5]. Furthermore, it forms a main channel for online piracy [17] and a useful basis for investigating game piracy. However, BitTorrent is just one of several channels of piracy, and estimates developed from P2P-network activity of course underestimates the true scale of overall game piracy activity.

3.3.3 Dynamic IP Addresses/Network Address Translators
In building the list of unique peers, it is assumed that each IP address listed by the tracker servers corresponds to one participating peer. Dynamic IP addresses and Network Address Translators (NATs) may however be employed in some places, which leads to an underestimation of the number of peers participating (e.g., multiple peers operating behind a NAT). Conversely, a single peer utilizing dynamic IP addresses can appear to be several different IPs over the period of data recording. Therefore, it should be emphasized that the data set presented offers a best estimate of the BitTorrent activity for the games sampled.

3.3.4 Virtual Private Networks and Tor
When downloading ("leeching") or uploading ("seeding") torrents via the BitTorrent protocol, participation is made public through the various peer discovery mechanisms. Peers who wish to operate anonymously can utilize commercial Virtual Private Network (VPN) services [8] techniques to introduce uncertainty into the tracker lists [4,10,11]. Alternatively, anonymous networks such as Tor can be used [The Tor Project 2010]. While there is evidence that some fraction of users participate anonymously, such users are considered relatively uncommon and are frowned upon by tracker server operators [22].

3.3.5 Sample Duration
The activity of peers in downloading files is not homogenous over time, but rather varies (see below). Additionally, torrents will only be available for a specific amount of time. This variation potentially biases measures that aggregate data across files. However, assuming that the variance is randomly distributed across time and torrent files, the bias is in effect noise given a large enough sample set. To the best knowledge of the authors, there is no published research studying the time-frequency behavior of torrents, and the standardized approaches of assuming random distribution is therefore adopted here (but see below).

3.3.6 Game Features Information
Issues such as invalid release dates, genre definitions, game information etc. may occur on the websites mined for information about the games. Aggregation of information across multiple websites forms an attempt to avoid including erroneous game product information.

4. DATA, ANALYSIS AND RESULTS
Following data collection and pre-processing, the final dataset with BitTorrent activity and game-feature information comprised 127 games. The BitTorrent activity data contained some noticeable characteristics. The frequency distribution of unique peers per game (Figure 1) was highly asymmetrical; the majority of the game titles had relatively limited activity on BitTorrent (i.e., less than 50,000 unique peers observed). The 10 most popular games

in the sample accounted for 5.37 million unique peers alone (Table 1), all of which were major commercial titles.

Table 1: The 10 most torrented game titles encompass 5.37 million of the unique peers in the dataset, averaging 536,727 peers per game and an average review score of 74.5 (on a scale from 0-100). Note: Genre definitions here from www.metacritic.com, not the system used in Figure 1.

Title	Genre	Unique Peers	Avg. Review Score	Developer
Fallout: New Vegas	Role-Playing	962,793	83.7	Obsidian Entertainment
Darksiders	Action Adventure	656,296	82.7	Vigil Games
Need for Speed: Hot Pursuit	Racing	656,243	88	Criterion Games
NBA 2k11	Basketball	545,559	86.7	Visual Concepts
TRON Evolution	Action Adventure	496,349	59.5	Propaganda Games
Call of Duty: Black Ops	First-Person Shooter	469,864	83.8	Treyarch
Starcraft 2	Real-Time Strategy	420,138	89.5	Blizzard Entertainment
Star Wars the Force Unleashed 2	Action	415,021	61	Lucas Arts
Two Worlds II	Role-Playing	388,236	73.3	Reality Pump
The Sims 3: Late Night	Virtual Life Games	356,771	77.5	The Sims Studio

In general, casual games and indie games were much less frequently pirated, with a few exceptions, notably Bejeweled 3 with over 250,000 unique peers recorded.

In terms of genre distribution, the majority of the games in the sample were classified as "action" games (n=48, 37.21% of the titles in the sample; Figure 2). The high number of games in this genre is primarily an effect of the releases occurring during November 2010-February 2011. In essence, more action games are released than titles in any other genre. The "action" category also encompasses a major part of the torrent activity (45.61%), while "role-playing games" accounted for 15.58% of the torrent activity, despite these games only comprising 10 titles (Figure 3). Further analysis reveals that while the "action" games are overall the most popular genre in terms of the number of unique peers accessing the related BitTorrent files, on a per game basis, they are far from the most popular (each encompassing 0.95% of the total number of unique peers on average). To the contrary, games within the "Racing" (5 titles), "Role-Playing Game" (10 titles) and "Simulation" (7 titles) genres are more popular, with each game in these categories encompassing 1.6%, 1.43% and 1.08% of the dataset (Figure 4).

The aggregated review score of digital games is generally expected to be linked with the financial success of a game, although this is not always the case [24]. Similarly, it is possible that review scores are also related to how much a game is distributed on BitTorrent – which is indicated by the high proportion of major commercial titles in the 127 game sample, as well as the observation that 7 of the 10 most torrented games in the sample has ag-

gregated review scores over 75 (on a 0-100 scale). In order to explore a possible relationship between torrent activity and aggregated review scores, Pearson's Product-Moment Correlation Coefficient for Metacritic Scores (mean = 70.44, Std. Dev. = 15.83) and number of unique peers per game (mean = 99894.43; Std. Dev. = 156028.6) was calculated (r= 0.2; p<0.05 significance (two-tailed) given df = 125 (n-2)) [14]. Please note that log(unique peers) was used due to the non-normal distribution of the peers data. The result indicates a statistically significant positive relationship between the number of unique peers and aggregated review scores. Put differently, Metacritic Scores explain 10% of the variance in the unique peers per game on BitTorrent.
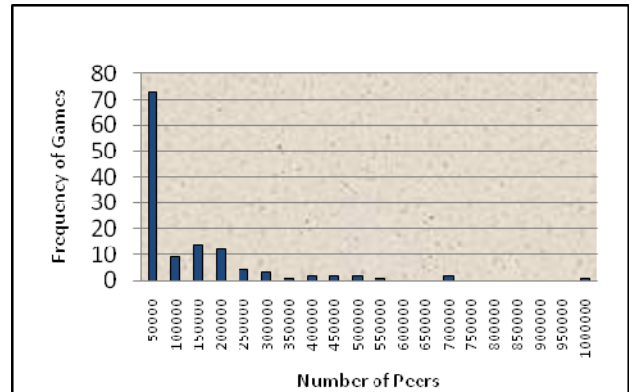

Figure 1: Frequency distribution of the number of unique peers associated with the games in the dataset.
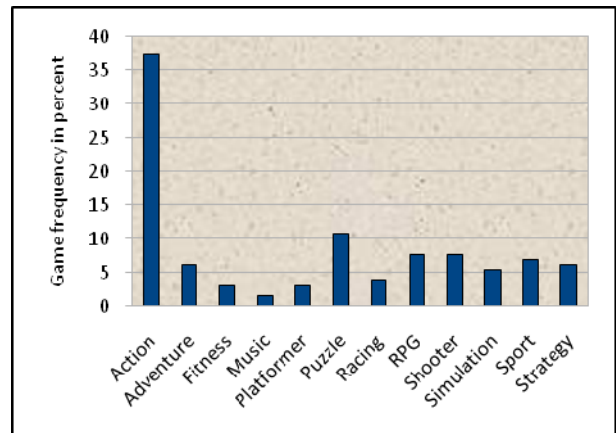

Figure 2: The percentage frequency of unique peers in the sampled games (n=127). RPG = role-playing game.

The result indicates a positive relationship between the number of unique peers and aggregated review scores. However, it should be noted that there are notable exceptions in the higher end of the aggregated review scores. For a few of these, such as Little Big Planet 2 (released on Jan 18th 2011 for PS3, aggregated average review score 90, unique peers 1,056), this may partially be an artifact of a late release date during the period of tracking (i.e., a short period where tracking of the game title was carried out).

It can thus be hypothesized that the correlation between Unique Peers and Metacritic score will be even stronger for a dataset consisting of only games that have all been tracked over a longer period. Future research will investigate this hypothesis.
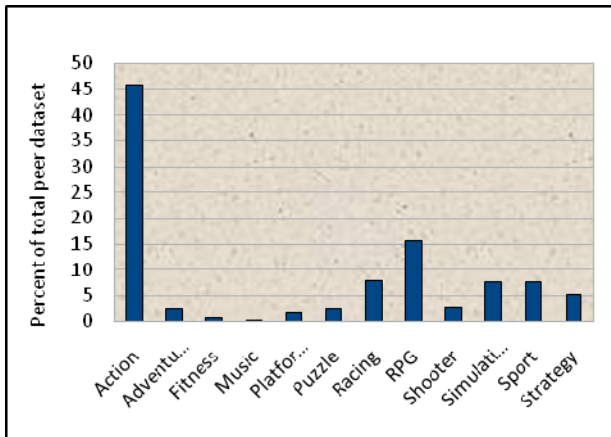
5

Figure 3: Percentage distribution of the peers recorded for games within each genre (127 games). RPG = role-playing game.
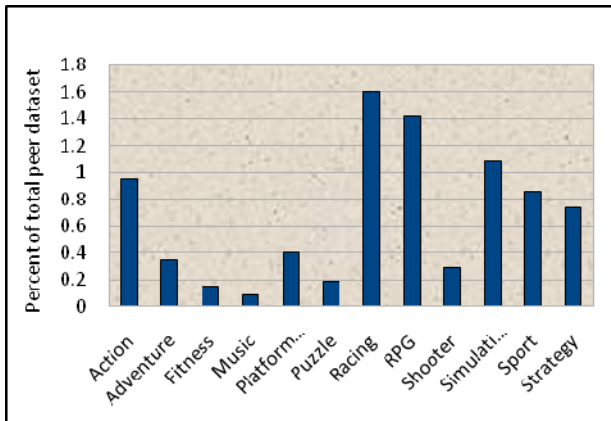

Figure 4: Distribution of the percentage of the 12.6 million aggregated unique peers (across 127 game titles) that each game within a specific genre encompasses. Highest ranking are "Racing" games, where each game contains 1.6% (corresponding to 201600 unique peers) of the total number of peers recorded. RPG = role-playing game.

The time-frequency distribution of game torrents is not homogenous, but rather varies over time. This phenomenon does not appear to have been a subject of previous research. Twenty games were selected randomly and their time-frequency pattern examined (time vs. number of unique peers, summed across all hardware platforms; Figure 6). This reveals a variety of temporal frequency distributions, with most titles following a distinctive pattern: initial rapid increase (often in the first day or few days since the first appearance of a torrent for the game), with a marked peak, and a slow following decline (12 of the 20 games). The decline can be either linear (e.g., TRON Evolution) or exponential (e.g., Alien Breed 3: Descent). Other games have a more plateau-like structure (e.g., World of Warcraft: Cataclysm). Lionheart: Kings Crusade follows a similar structure, whereas Tom Clancy´s Ghost Recon reaches three peaks before leveling out. In contrast, Dreamworks Megamind: Ultimate Showdown has a protracted period of low acitivity followed by a small peak before it disappears in mid-December 2010. Sports Island Freedom follows a somewhat similar pattern, working up to a peak with a following decline. Finally, the number of unique peers accessing the minor game titles The Mysterious Case of Dr. Jekyll and Mr. Hyde and Tomb of Sojir are small and there is no discernible peak structure.

Future work will investigate the time-frequency distribution of game torrents in more detail, however, two conclusions that can be derived from the current study is that: 1) When working with small samples of files in BitTorrent research, the time-frequency distribution of the corresponding torrents needs to be considered to avoid biasing results; 2) Torrent activity for digital games varies substantially over time, and not according to any one pattern.

5. DISCUSSION AND CONCLUSION

The illegal copying and distribution of digital games stands at the heart of one of the central controversies in the international interactive entertainment environment. Despite the substantial interest in the problem, the wealth of industry-based reports of piracy rates [e.g., 9,17,18], and the size of the industry, there is only minimal objective information available about the magnitude of game piracy and its distribution across game titles or genres. In this paper, a first step has been taken towards addressing this knowledge gap, via the analysis of a 12.6 million unique peer dataset obtained from BitTorrent, generally regarded as the major channel for game piracy and the standard for P2P distribution of [5,29].

The first and perhaps most important contribution of this paper is to provide objective documentation of the magnitude of distribution of digital game files via BitTorrent. Out of 173 game titles in the study, released during the fall 2010 or early 2011, 127 were found on BitTorrent networks. Approximately 12.6 million unique peers accessed these files, indicating the prevalence of game piracy via BitTorrent-based distribution. Unlike previous work, the data reported here are objective, quantitative and developed using state-of-the-art techniques and with a public and open methodology. How the number of unique peers translates into lost sales is a contested issue [9,15,25,17], and one that future research will investigate. Additionally, analysis of the peer data indicate that it is the major commercial titles that are the most heavily distributed games on BitTorrent, with some exceptions from the more casual form. The ten most pirated titles encompass 5.27 million aggregated unique peers. While games of the "Action" genre comprise about 45% of the unique peers in the dataset, on per game basis titles from the "Racing", "Role-Playing Game" and "Simulation" genres display higher numbers of unique peers on average than "Action" games. While there are fewer of these games, the ones that are there are more popular. Finally, the evidence presented indicate that review scores and torrent activity for game titles correlate positively (p<0.05).

Future research will focus on exploring questions such as the relationship between game piracy and additional product features (e.g., ESRB rating). Furthermore, the time-frequency distribution of piracy rates is of interest in order to examine if specific patterns and cycles in BitTorrent activity can be defined. Future research will also explore whether marketing strategies and differences in international release dates have an effect on piracy activity. With the data in place, explanations for the root causes of the patterns observed can be investigated.
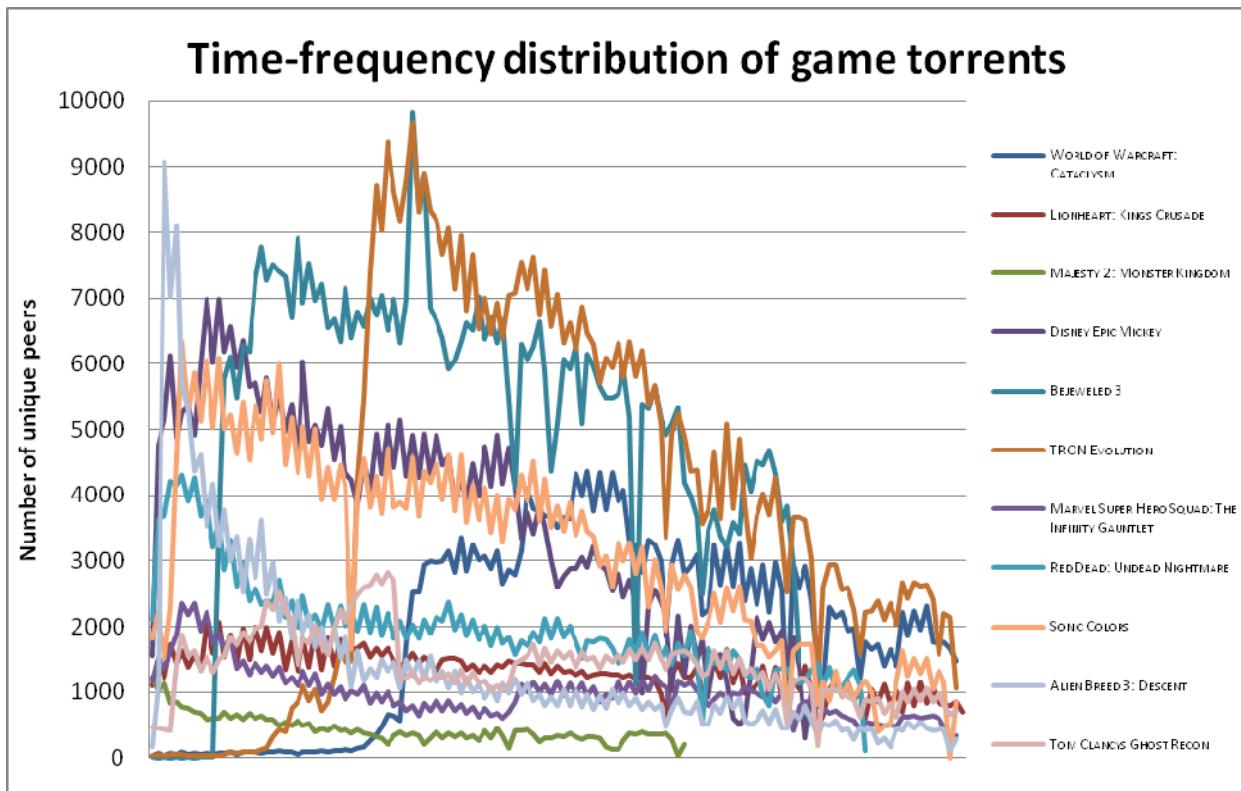
6. ACKNOWLEDGMENTS

Figure 5: Aggregated torrent time-frequency structure for selected game titles (12 hour time bins). The data for TRON evolution has been divided by a factor of two in order to fit into the diagram.

7. REFERENCES

[1]   Anderson, N. 2011. Where have all the music pirates gone? *Ars Technica*, February 4th 2011. Available from: http://arstechnica.com/tech-policy/news/2011/02/where-have-all-the-music-pirates-gone.ars

[2]   Antoniades, D., Markatos, E.P. and Dovrolis, C. 2009. One-click hosting services: a file-sharing hideout. In *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement,* ACM, 223-234.

[3]   Banerjee, D., Khalid, A. M. and Sturm, J.-E. 2005. Socio-economic development and software piracy. An empirical assessment. *Applied Economics 37*, 2091 - 2097.

[4]   Bauer, K., McCoy, D., Sicker, D. and Grunwald, D. 2008. BitBlender: Light-Weight Anonymity for BitTorrent. In *Proceedings of the 4th ACM Workshop on Applications of Private and Anonymous Communications*, ACM, 1-8.

[5]   Bauer, K., Grundwald, D. and Sicker, D. 2009a. The Arms Race in P2P. In Proceedings of the 37th Research Conference on Communication, Information and Internet Policy, TPRC, 1-18.

[6]   Bauer K., McCoy, D., Grundwald, D. and Sicker, D. 2009b. Bitstalker: Accurately and Efficiently Monitoring BitTorrent Traffic. In *Proceedings of the First IEEE Workshop on Information Forensics and Security,* IEEE Publishers, 181-185.

[7]   *BitTorrent Protocl Specification*. 2010. Available from: http://wiki.theory.org/BitTorrentSpecification.

[8]   *BTGuard: Anonymous BitTorrent Services*. 2011. Available from: http://btguard.com/.

[9]   Business Software Alliance. 2010. *Seventh Annual BSA and IDC Global Software Piracy Study*. Available from: http://www.bsa.org/country/Research%20and%20Statistics.aspx

[10]   Choffnes, D. R., Duch, J., Malmgren, D., Guierm, R., Busta-mante, F. E. and Amaral, L. 2009. *SwarmScreen: Privacy Through Plausible Deniability in P2P Systems*. Northwestern University EECS Technical Report. March, 2009.

[11]   Choffnes, D.R., Duch, J., Malmgren, D., Guiermà, R., Busta-mante, F.E. and Amara, L. 2010. Strange bedfellows: community identification in bittorrent. In *Proceedings of the 9th international conference on Peer-to-peer systems,* USENIX Association, 13-13

[12]   Ellison, B. 2008. World of Goo Co-Creator Claims 90% Piracy Rate. *Shacknews,* November 13, 2008. Available from: http://www.shacknews.com/onearticle.x/55906.

[13]   Envisional, 2011. *An Estimate of Infringing Use of the Internet*. Technical report (v. 1.8). Available from: http://documents.envisional.com/docs/Envisional-Internet_Usage-Jan2011.pdf

[14]   Ernesto. 2011. Arrr! The Music Pirates Are Still Here. *TorrentFreak*, February 7th 2011. Available from: http://torrentfreak.com/arrr-the-music-pirates-are-still-here-110207/

[15]   Goldacre, B. 2009. Illegal downloads and dodgy figures. *The Guardian*, June 5 2009. Available from: http://www.guardian.co.uk/commentisfree/2009/jun/05/ben-goldacre-bad-science-music-downloads.

[16]   Huygen, A., Eijk, N. V., Poort, J. and Rutten, P. 2009. *Ups and downs – Economic and cultural effects of file sharing on music, film and games*. TNO Information on Communication Technology, the

Netherlands. Available from: http://www.tno.nl/content.cfm?context=thema&content=inno_publicatie&laag1=897&laag2=918&item_id=473.

[17] International Intellectual Property Alliance (IIPA). 2010. *2010 Special 301 Report on Copyright Protection and Enforcement*. Available from: www.iipa.com.

[18] KIRK, R. 2010. *Special 301 Report 2010.* Published by the Office of the United States Trade Representative, United States of America. Available from: http://bangkok.usembassy.gov/root/pdfs/2010_special_301_report.pdf.

[19] Klepek, P. 2008. Bethesda Is Tired Of Spending Money Supporting Software Pirates. *MTV Multiplayer*, October 13 2008. Available from: http://multiplayerblog.mtv.com/2008/10/13/bethesda-deals-with-pirates/.

[20] LeBlond, S., Legout, A., Lefessant, F. Dabbous, W. and Kaafar, M. A. 2010. Spying the World from Your Laptop: Identifying and Profiling Content Providers and Big Downloaders in BitTorrent. In *Proceedings of the 3rd USENIX Workshop on Large-Scale Exploits and Emergent Threats,* USENIX Association, 4-4.

[21] Maier, G., Feldmann, A., Paxson, V. and Allman, M. 2009. On dominant characteristics of residential broadband internet traffic. In *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement*, ACM, 90-102.

[22] Mccoy, D., Bauer, K., Grunwald, D., Kohno, T. and Sicker, D. 2008. Shining Light in Dark Places: Understanding the Tor Network. In *Proceedings of the Proceedings of the 8th international symposium on Privacy Enhancing Technologies*, ACM, 63-76.

[23] Piatek, M., Kohno, T. and Krishnamurty, A. 2008. Challenges and directions for monitoring P2P file sharing networks-or: why my printer received a DMCA takedown notice. In *Proceedings of the Proceedings of the 3rd conference on Hot topics in security*, USENIX Association, 1-7.

[24] Rabin, S. 2010. *Introduction to Game Development*. Charles River Media.

[25] Sanchez, J. 2008. 750,000 lost jobs? The dodgy digita behind the war on piracy. *Ars Technica*, October 7th, 2008. Available from: http://arstechnica.com/tech-policy/news/2008/10/dodgy-digits-behind-the-war-on-piracy.ars

[26] Schulze, H., and Mochalski, K. 2009. *Internet Study 2008/2009*. Available from: http://www.ipoque.com/study/ipoque-Internet-Study-08-09.pdf.

[27] Sigle, D. 2010 *Critical Values of the Pearson Product-Moment Correlation Coefficient*. Neagle School of Education - University of Connecticut. Available from: http://www.gifted.uconn.edu/siegle/research/correlation/corrchrt.htm

[28] Stryszkowski, P. and Scorpecci, D. 2009. *Piracy of Digital Content*. OECD Publishing.

[29] The Tor Project: *Anonymity Online*. 2010. Available from: http://www.torproject.org/

[30] TorrentFreak. 2010. Call of Duty Black Ops Most Pirated Game of 2010. *TorrentFreak*. December 28 2010. Available from: http://torrentfreak.com/call-of-duty-black-ops-most-pirated-game-of-2010-101228/.

[31] Wolchok, S., and Halderman, J. A. 2010. Crawling BitTorrent DHTs for Fun and Profit. In *Proceedings of the 4th USENIX Workshop on Offensive Technologies*, USENIX Association, 112-113.

[32] Yang, D., Sonmez, M., Bosworth, D. and Fryxell, G. 2009. Global Software Piracy: Searching for Further Explanations. Journal of Business Ethics 87, 269-283.