Distribution of Digital Games via BitTorrent

Anders Drachen
Department of Communication
Aalborg University
drachen@hum.aau.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. 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 the first large-scale, open-method analysis of the 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. The study provides findings that reveal the magnitude of game piracy, the time-frequency of game torrents, which genres that get pirated the most, and the relationship with review scores and ESRB-ratings.

Categories and Subject Descriptors

K.8.0 [Games]; K.7.m [The Computing Profession]: Miscellaneous – Ethics.

General Terms

Economics, Security, Human Factors, Legal Aspects.

Keywords

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

1. INTRODUCTION

Game piracy, which involves the illegal copying and distribution of digital games [28], is a complex phenomenon that occurs across multiple channels and has a magnitude that is difficult to estimate [9,17,18], not the least due to the lack of clarity as to what constitutes illegal copying and copyright infringement internationally. 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 often comes from industry organizations or operators of peer-to-peer (P2P) networks [e.g., 13,15,17], but lacks objectivity and transparent methodology. The purpose of this paper is to address the need for objective information on game piracy, covering two important areas: 1) An overview of the state-of-the-art of the debate and the issues related to monitoring BitTorrent; commonly ignored in reports on piracy [5].

2) We analyze game piracy data obtained by tracking BitTorrent P2P file sharers. BitTorrent was chosen because it is regarded as one of the main channels for online piracy and the de facto standard for distribution of digital files via P2P networks [26]. For our study, monitoring was carried out for 173 game titles over a three month period running from late 2010 to early 2011, spanning most types of games and multiple hardware platforms (e.g., PC, X360, PS3, Wii, DS, PSP). Our data set of P2P activity is among the largest analyzed to date, with over 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" and "Role Playing Game" games are individually more popular than the major action titles. Additionally, 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 (averaged across multiple reviews) were found to be positively correlated with the games popularity on BitTorrent, in terms of number of sharing peers ($p < 0.05$). This indicates that game quality, as indicated by review scores, is related to piracy activity, such that higher quality games get pirated more frequently.

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]) as well as branch organizations such as the Entertainment Software Association (ESA) and the Business Software Alliance (BSA) [9,17,18]. The data reported by the industry are potentially biased, partially due to the interest of the industry to reduce piracy and thus potentially over-estimate the problem. Also, industry reports often lack methodological transparency [15].

2.1 Previous Work

Although the research literature on digital game piracy is limited, 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 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, whereas work focused on country-level variations in software piracy 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 [32]. Both sides of the piracy debate are, however, more or less in agreement when it comes to the conclusion that the phenomenon of piracy is

common, although the specific numbers vary between reports. For example, the ESA claimed that 9.78 million "illegal" downloads of roughly 200 digital games had occurred in December 2009 alone [17]. TorrentFreak.com, one of the most influential websites on P2P sharing via BitTorrent - but openly in favor of P2P networks - 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 [30]. However, neither of these studies employs an open methodology, leading one to question the reliability and validity of the data.

The credibility of industry-based or government-solicited reports is also hindered by the common methodological problems in such work, which foster suspicion of bias [14,15,30]. 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%), a pattern also noted by Ipoque [26]. However, these conclusions are based on a "snapshot" methodology, i.e. gathering data over a very small temporal interval, which runs the risk of 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. Moreover, as noted by Anderson [1]: "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]. 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 digital copies are distributed (e.g., physical copying or peer-to-peer networks) are almost impossible to monitor effectively [2,5,6,17]. Distribution channels for digital material include physical as well as networked solutions, with the most common online being P2P protocols [17], "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 over a decade ago, and are thus covered in less detail here:

Usenet: Usenet is a decentralized network launched in the 1980s to permit the sharing of conversations before the development of web forums. Usenet facilitates piracy by allowing users to upload 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 does not exist today.

File Transfer Protocol: The File Transfer Protocol (FTP) 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.

Internet Relay Chat: The Internet Relay Chat (IRC) 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 over IRC can be technically complicated compared to other methods, and may involve long waiting periods.

Physical distribution: Pre-dating the Internet, the 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 revolved around the mass duplication of pirated media, and distribution/sale at below-market prices [17]. In some cases, 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/areas, such as Italy, who are on the watch list of the International Intellectual Property Association [17].

File-hosting Services: So-called "one-click" 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 (some servers are open). 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 a wide variety of copyrighted 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 copied content because they require only very limited technical knowledge to access it. For example, standard web links to specific files can be shared and searched 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) and hosting services.

P2P protocols: P2P protocols enable end-users to share content with one another directly, eliminating the need for uploading digital content to centralized servers 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 consumers of digital content.

Developing reliable estimates of the piracy activity that occurs via any channel of distribution is challenging, however, the BitTorrent protocol is generally viewed as the major channel for game piracy today [17,18], and so can be used to estimate the extent of piracy across games types and platforms. 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% [26].

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 can be obtained by querying the trackers used to provide information about specific torrents [7] or by crawling the BitTorrent Distributed Hash Tables (DHT). 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 to track content on torrent networks 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. DATA AND 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 digital games; 2) Information about the products. The process of obtaining these data is comprised of a series of steps, as follows:

3.1 Obtaining Unique Peers from BitTorrent

The BitTorrent protocol works by breaking down files that peers (users) seek to share into many pieces of a specific size and distributing them across a network. 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 users who make the file available to others (peers) [5,6] describe each file. 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. Participation in BitTorrent is made public via the peer´s IP-address. Importantly, in the process of obtaining a peer list from the server, the peer 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 game 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 for these platforms since November 17th 2010 until February 6th 2011, the end of the tracking period (note that games can appear on BitTorrent prior to the official launch date). Due to the lack of centralized repositories of information about game releases (see below), more than a dozen of the major game websites (e.g. gamasutra.com, game developer.com, ign.com, gamespy.com, vgchartz.com, gamestats.com, mobygames.com) were mined regularly to develop the list of games released during this interval. However, it is possible that some minor/indie titles, too small commercially to appear on the sites mentioned, were not included. No games legally distributed via BitTorrent were included unless their corresponding torrents could be identified as being versions not currently permitted to be freely shared by the publisher (whether this formally constitutes illegal sharing or not is a subject of the legislation of the countries where the peer is situated).

The analysis consolidates the total number of peers for each game across torrents for different platforms (e.g. Xbox 360, PS3). Multi-platform releases count as only one title in the analysis presented here, and this is also the case for regional releases (e.g. European and US versions of a game). 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, 127 were located on BitTorrent, indicating these games had been cracked of any copyright protection (Digital Rights Management), and released on BitTorrent. During the period of tracking, 12.7 million unique peer IP addresses were identified for all these games, making this the largest study of BitTorrent-based game piracy to date, 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 peer IP addresses).

In order to obtain BitTorrent metadata files, a custom web crawler was developed to periodically issue queries to a popular BitTorrent search engine (which collates data from 500+ sites and 150,000+ trackers) website for each title, extracting the metadata files. Having located the metadata files, the web crawler obtains the tracker server URIs. Next, 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 content.

When searching for game torrents, 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 (only game executable in the torrent, not the entire game), expansion content, game soundtracks, graphical material etc. 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. Additionally, it is a common practice for copyright investigators to spread false information on BitTorrent, for example by distributing torrents which do not contain the actual game files, in an attempt to make it difficult for peers to access copyrighted material [6]. In order to avoid including false torrents and filter out the false IP addresses, standard publicly

available blacklists were applied to filter out IP addresses 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 [33]. About 200.000 IP-addresses were removed via this process (leaving the dataset used here).

3.2 Obtaining Game Data

Unlike the movie and music industry, there are no centralized reporting systems for information such as release dates, sales figures and producer details. This makes it challenging to acquire reliable information on games to contextualize analyses of piracy data. This form of business intelligence data are available from a few vendors servicing the interactive digital entertainment sector, but at prices out of scope of most academically-based projects. Alternative solutions must therefore be sought.

Release dates: 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. As noted earlier, there are no central repositories for information on digital games. 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.

Genre: Games come in great variety and there are similarly many different systems for categorizing games into "genres" or "types". Game genre systems are nebulous at best, and therefore an aggregation approach was adopted here to build a genre system based on majority consensus. In order to obtain as robust a framework as possible, a variety of recognized websites (e.g., mobygames.com, ign.com, gamespy.com, metacritic.com) were mined and genre definitions for the individual titles based on majority consensus. This led to the definition of 17 genres (Figure 3). Some of these genre categories form natural frames where it is fairly obvious which games belonged to it – e.g. "sports" and "racing". Others are more difficult to work with because some games, rather than forming natural clusters with specific features, vary across a spectrum. Therefore, it can be challenging to categorize a game into, e.g., into "action-adventure" or "adventure".

A typical genre definition found on a site such as Wikipedia is the "action game". This is a good example of a problematic definition: Across the previously 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 interaction with computer-controlled entities (NPCs). Examples include games such as Kane & Lynch, Grand Theft Auto, 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 could therefore be argued to be classified as "shooters" – another genre definition used on various game websites. Notably, the "action game" classification generally includes most of the major commercial titles for both PC and consoles (outside of sports/singing/fitness etc. games). Examples include Call of Duty: Black Ops, Bioshock 2, Darksiders, Medal of Honour, and Splinter Cell. In the current project, the "action game" classification was not used, and genres were divided at a more detailed level. For example, action games with a first-person vs. a third-person camera view formed distinct categories. Similarly, games focusing on melee-based combat were labeled "beat ´em up", and games featuring platform game mechanics as a main component labeled "platformers". "Shooters" here are action games that focus mainly on shooting, but which do not employ a FPS/TPS perspective (e.g. arcade-style space shooters).

Adventure games were separated from action-adventures on the basis of the level of combat involved – adventure games are more explorative in nature than action-adventures, which combine the aggressive gameplay of the FPS/TPS-style games with exploration. The label "RPG" (role-playing game, e.g., Fallout: Las Vegas, Divinity 2: The Dragon Knight Saga) was used for any game that the publisher labeled RPG, and is characterized by the player-controlled character developing in physical abilities during the playing of the game (e.g. the gaining new abilities via increases in class level).

3.2.1 Aggregated review scores

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 70.13 and Std. Dev. = 15.67 (n=117).

3.3 Assumptions and Limitations

The dataset presented and analyzed 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:

Sample of games: Roughly 1500 digital games are launched on a yearly basis [24], but it is unknown how many of these are cracked and released as torrents on P2P networks. This means that it is difficult to estimate how representative our sample of 173 is. This is a subject for future research.

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.

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.

Virtual Private Networks and Tor: Peers who wish to operate anonymously can utilize commercial Virtual Private Network (VPN) services [8] and techniques to introduce uncertainty into the tracker lists [4,10,11]. Alternatively, anonymous networks such as Tor can be used [29]. While there is evidence that some fraction of users participate anonymously [22], but the exact number is unknown.

Sample duration: The activity of peers in downloading files is not homogenous, but rather varies over time and across game titles (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).

Game feature 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. ANALYSIS AND RESULTS

Following data collection and pre-processing, the final dataset with BitTorrent activity and game-feature information comprised 127 games. Thirty-two were single-platform releases, and the rest were multi-platform (X360, PS3, PC a common combination). The BitTorrent activity data contained some noticeable characteristics. The frequency distribution of unique peers per game was highly asymmetrical (Figure 1); the majority of the game titles had relatively limited activity on BitTorrent (i.e., less than 50,000 unique peers observed).
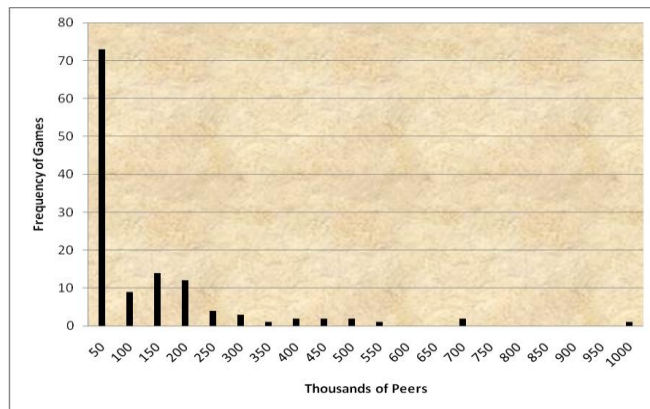
**Figure 1: Frequency distribution of the number of unique peers associated with the games in the dataset.**

Time-frequency distribution: The time-frequency distribution of game torrents has not been the subject of previous research; con-

trarily, earlier reports generally assume equal distribution across time. In fact, a variety of temporal frequency distributions are evident in the BitTorrent data presented here (Figure 2a and 2b), 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 (~60% of the examined titles). The decline can be either roughly linear (e.g., TRON Evolution, Figure 2a) or exponential (e.g., Alien Breed 3: Descent, Figure 2a). Other games have a more plateau-like structure (e.g. Lionheart: Kings Crusade, Figure 2b) whereas Tom Clancy´s Ghost Recon reaches three peaks before leveling out. In contrast, Dreamworks Megamind: Ultimate Showdown has a protracted period of low activity followed by a small peak before it disappears in mid-December 2010. Sports Island Freedom follows a somewhat similar pattern, working up to a protracted peak with a following decline. In general, the curves pan out after a maximum of 60 days, from which point on BitTorrent activity is a fraction compared to the initial highs.

This finding is important, because snap-shot type analyses, where data are collected over a short interval will tend to under-estimate or over-estimate the BitTorrent activity for a specific game title. If for example the BitTorrent traffic for a specific game is tracked right after it is made available on BitTorrent, the reported numbers will be very high, provided the game follows the typical peak-then-decrease pattern. Contrarily, measuring BitTorrent activity during the tail end of the distribution will result in low numbers. In order to fully evaluate the BitTorrent traffic for a game, an extended period of monitoring is needed to produce accurate numbers. This observation means that one should keep these limitations in mind when reading reports on BitTorrent traffic with less than 60 consecutive days of monitoring [e.g. 17,18] or an insufficiently high sample size to warrant assumption of every point on the different types of distribution curves being equally represented in the study in question. Note that this does not invalidate estimates of total torrent traffic over a given period of time; however, such results are limited in the analytical depth they provide as they do not consider the temporal (and geographic) dynamics of game torrents. Future work will investigate the time-frequency distribution of game torrents in more detail, however, two conclusions can derived from the current study: 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.

Genre distribution: In terms of genre distribution, the most common genre in the sample was "puzzle" games (n=14, 11.02% of the titles in the sample; Figure 3), closely followed by RPGs (n=13), TPS (n=12), Action-adventure and Strategy games (n=11) and FPS (n=10). The genre distribution was however well mixed (SD = 4.12). The pattern changed substantially when considering the total number of unique peers recorded per genre (Figure 4). RPG (18.9%), Action-adventure (15.9%), TPS (12.7%) and Racing (9.3%) games comprised the most popular genres. In comparison, there were fewer recorded peers for the Family, Shooter, Music, Fitness and Arcade genres, which were also the most under-represented genres in the sample in terms of number of games (Figure 3). The genres that account for the majority of the unique peers are also the most popular on a per-game basis, although RPGs (13 games in sample) and Action-adventure games (11 games in sample) are notably popular, with each game accounting for 1.5% and 1.45% of the dataset respectively (Figure

5). Racing games account for 1.33% of the dataset each on average (with Need for Speed: Hot Pursuit being a marked factor in driving up the average), followed by Simulation games (1.08%, 7 games in sample), and TPS (1.06%, 12 games in sample).
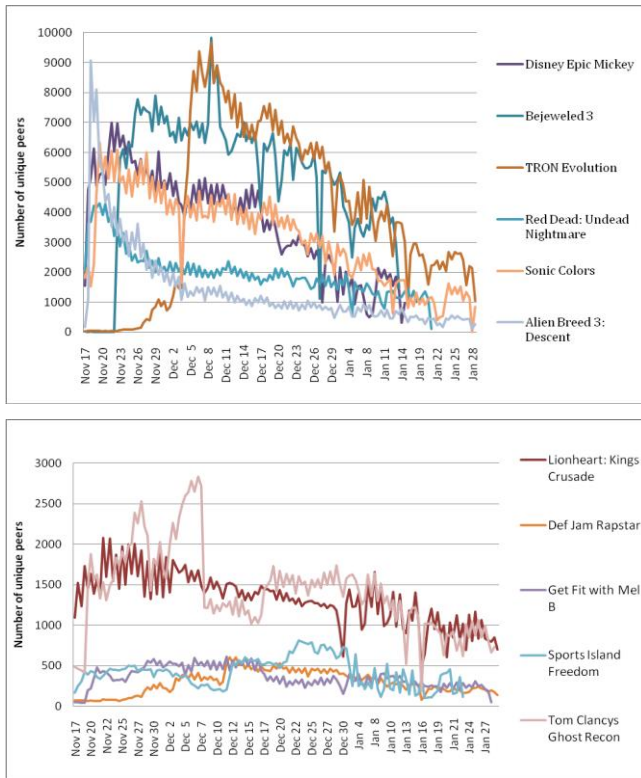

Figure 2a (top); 2b (bottom): Aggregated time-frequency structure for selected game titles (12 hour time bins). Data for TRON evolution has been divided by a factor of two.
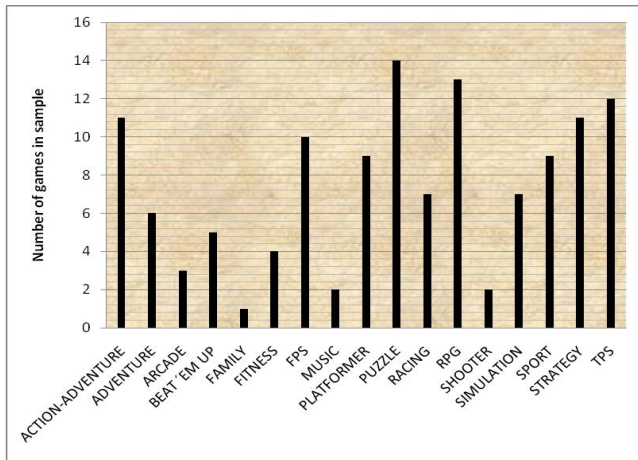

Figure 3: The number of games in each genre in the sampled games (n=127).

Aggregate review scores vs. popularity: The 10 most popular games in the sample accounted for 5.37 million unique peers (41.8%) alone (Table 1), all of which were major commercial titles. The aggregated review score of digital games is generally related to 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 shared games in the sample had aggregated review scores over 75 (on a 0-100 scale, a score of 75+ is considered "generally favorable" by metacritic.com for the games category). In order to explore a possible relationship between torrent activity and aggregated review scores, a Pearson's Product-Moment Correlation Coefficient for Metacritic Scores (mean = 70.13, Std. Dev. = 15.67) and number of unique peers per game (mean = 99894.43; Std. Dev. = 156028.6) was calculated (r= 0.28; p<0.05 significance (two-tailed) given df = 115 (n-2)) [27]. Please note that log(unique peers) was used due to the non-normal distribution of the peers data. The result indicates a moderate, 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, average review score 90, 1,056 peers), 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 review score will be even stronger for a dataset consisting of games that been tracked over a longer period. Future research will investigate this hypothesis. In general, casual games and indie games were less frequently pirated, with a few exceptions, e.g. Bejeweled 3 with over 250,000 unique peers recorded.
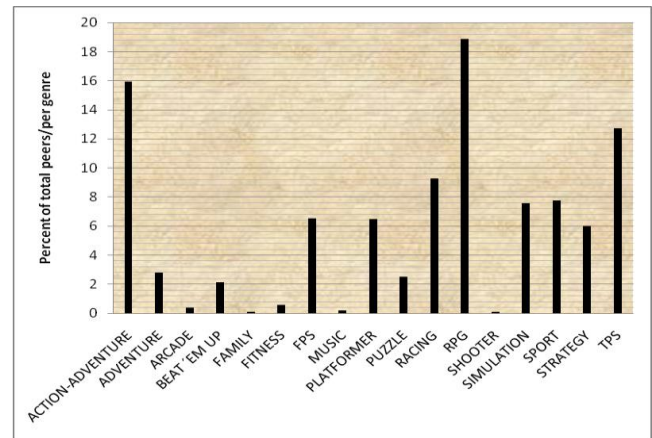
**Figure 4: Percentage distribution of the peers recorded for games within each genre (n=127).**
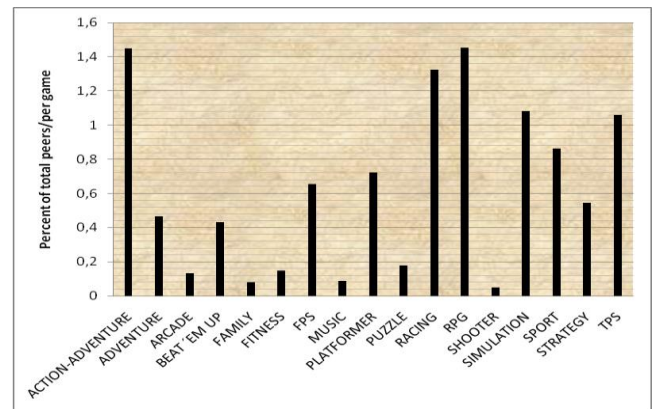
Figure 5: Percentage of the aggregated unique peers (n=127) that each game within a specific genre encompasses.

Rating vs. popularity: Of the 127 games, 36 carried an ESRB rating "E" (Everyone), 16 "EE10+" (Everyone 10+ years), 33 "M" (Mature), 39 "T" (Teen) and 3 had ESRB rates pending at the time of writing. However, "M" rated games were popular in terms of the number of unique peers (37%) (Figure 6), corresponding to 1.12% of the total dataset per game, compared to 0.39% for "E"-rated games, 0.69% for "T"-rated games but 1.5% for "E10+" rated games. This result is somewhat surprising as it is usually the "M" and "T" rated games that are mentioned on "most downloaded" lists released by torrent sites [e.g. 30], and indicates that piracy is not limited to games of a particular rating, although it remains the M-rated games that are the overall most popular and therefore the most distributed via BitTorrent.

1. 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 [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 over a three month period, generally regarded as the major channel for game piracy and the standard for P2P distribution [5,29]. The work presented provides a quantitative basis for the game piracy debate, answering key questions about the scale of BitTorrent-based distribution of game files, and the relationship with aggregated review scores and game genre. Additionally, it provides the basis for beginning to address the "why"-questions in the debate, e.g. why particular games are pirated more than others.
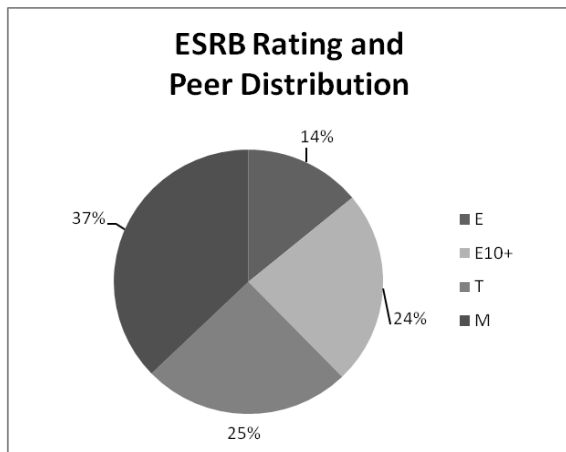
Figure 6: Distribution of unique peer numbers games according to ESRB rating.

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 (averaging close to 100,000 peers per game). 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. The analysis presented here also provides the first publicly available analysis of how BitTorrent traffic is distributed across game titles, genres and ESRB ratings.

This analysis reveals that it is a few titles, typically major commercial titles, that are the most heavily distributed on BitTorrent (Table 1). The ten most pirated titles encompass 41.8% of the total dataset. It also reveals that there is a positive correlation between aggregate review scores, such as those obtained from Metacritic.com, and BitTorrent popularity (p<0.05). This means that games with good review scores are more likely to be heavily distributed on BitTorrent. Additionally, ESRB rating also appears to hold an influence. Thirty-seven percent of the dataset was "M" rated games, with "T"-rated games comprising 25% of the dataset. Whether this leads to people getting access to games that, according to ESRB, they should not be exposed to, is an open question. In terms of genre or game type, RPGs and Action-Adventure games are by far the most popular, followed by TPS and Racing games. These are also the genres most popular on a per-game basis, with a consistently high BitTorrent activity for these genres.

Future research will focus on exploring questions such as the relationship between game piracy and additional product features. 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.

Table 1: The 10 most torrented game titles encompass 5.37 mio. unique peers, averaging 536,727 peers per game and an average review score of 74.5 (on a scale from 0-100).

Title	Genre	Unique Peers	Avg. Review Score	Developer
Fallout: New Vegas	RPG	962793	83.7	Obsidian Entertainment
Darksiders	Action Adventure	656296	82.7	Vigil Games
Need for Speed: Hot Pursuit	Racing	656243	88	Criterion Games
NBA 2k11	Sports	545559	86.7	Visual Concepts
TRON Evolution	Action Adventure	496349	59.5	Propaganda Games
Call of Duty: Black Ops	FPS	469864	83.8	Treyarch
Starcraft 2	Strategy	420138	89.5	Blizzard Entertainment
Star Wars the Force Unleashed 2	Action Adventure	415021	61	Lucas Arts
Two Worlds II	RPG	388236	73.3	Reality Pump
The Sims 3: Late Night	Simulation	356771	77.5	The Sims Studio

2.   ACKNOWLEDGMENTS

3.   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 *Proc. of the 9th ACM SIGCOMM Conf. 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 assess-ment. *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 Anony-mous Communications*, ACM, 1-8.

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

[6]   Bauer K., McCoy, D., Grundwald, D. and Sicker, D. 2009. Bitstalker: Accurately and Efficiently Monitoring BitTorrent Traffic. In *Proceedings of the First IEEE Workshop on Information Foren-sics 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., Bustamante, F. E. and Amaral, L. 2009. *SwarmScreen: Privacy Through Plausible Deniability in P2P Systems*. Northwestern Uni-versity EECS Technical Report. March, 2009.

[11]   Choffnes, D.R., Duch, J., Malmgren, D., Guiermà, R., Bustamante, F.E. and Amara, L. 2010. Strange bedfellows: commu-nity identification in bittorrent. In *Proceedings of the 9th interna-tional 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 ICT, the Netherlands. Availa-ble 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: Identify-ing and Profiling Content Providers and Big Downloaders in Bit-Torrent. 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 Sick-er, D. 2008. Shining Light in Dark Places: Understanding the Tor Network. In *Proc. of the Proceedings of the 8th international sym-posium on Privacy Enhancing Technologies*, ACM, 63-76.

[23]   Piatek, M., Kohno, T. and Krishnamurty, A. 2008. Chal-lenges 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 Digi-tal 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 Bit-Torrent DHTs for Fun and Profit. In *Proc. of the 4th USENIX Work-shop 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. Jour-nal of Business Ethics 87, 269-283.

[33]   IP Blocklist. 2010. Available from: http://www.iblocklist.com/list.php? list=bt_level1