Patterns in the distribution of digital games via BitTorrent

Anders Drachen*

Department of Communication and Psychology, Aalborg University and PLAIT Lab, Northeastern University, 360 Huntington Avenue, Boston, MA 02115, USA E-mail: andersdrachen@gmail.com *Corresponding author

Robert W.D. Veitch

Department of IT Management, Copenhagen Business School, Howitzvej 60, 4th floor, 2000 Frederiksberg, Denmark E-mail: rwdv.itm@cbs.dk

Abstract: The distribution of illegal copies of computer games via digital networks forms the centre in one of the most heated debates in the international games environment, but there is minimal objective information available. Here the results of a large-scale, open-method analysis of the distribution of computer games via BitTorrent peer-to-peer file-sharing protocol is presented. 173 games were included, tracked over a period of three months from 2010 to 2011. A total of 12.6 million unique peers were identified across over 200 countries. Analysis indicates that the distribution of illegal copies of games follows distinct pattern, e.g., that a few game titles drive the traffic – the 10 most accessed games encompassed 42.7% of the number of peers tracked. The traffic is geographically localised – 20 countries encompassed 76.7% of the total. Geographic patterns in the distribution of BitTorrent peers are presented, as well as time–frequency distributions of torrents, and additional results.

Keywords: computer games; security; piracy; game piracy; BitTorrent; economics of piracy; video games; analytics; software piracy; digital piracy; media; communication.

Reference to this paper should be made as follows: Drachen, A. and Veitch, R.W.D. (xxxx) 'Patterns in the distribution of digital games via BitTorrent', *Int. J. Advanced Media and Communication*, Vol. x, No. x, pp.xxx–xxx.

Biographical notes: Anders Drachen (PhD) is a veteran Data Scientist, currently operating as Lead Game Analyst for Game Analytics (www.gameanalytics.com), and as Associate Professor at PLAIT (Northeastern University, Massachusetts). His research is focused on game analytics, business intelligence, game development, game user research and game data mining.

Robert W.D. Veitch (MSc) is currently finalising his PhD at the Institute for Informatics at Copenhagen Business School on the antecedents of internet piracy. His research interests include digital/internet piracy, consumer

behaviour, decision making, and culture. Before joining Copenhagen Business School, he was a research analyst for a New Zealand firm specialising in stakeholder surveys and analysis. A formerly-registered psychologist in New Zealand, his educational background is in industrial, work and organisational psychology. He received an MSc from the University of Auckland in 2007 and have published academic research on the psychological contracts of temporary employees.

This paper is a revised and expanded version of a paper entitled 'Distribution of digital games via BitTorrent' presented at the *15th Mindtrek Conference: Envisioning Future Media Environments*, 3–5 October, Tampere, Finland, 2011. http://doi.acm.org/10.1145/10.1145/2181037.2181077; where initial results of the study presented here were presented. In the current paper, these results are described in context, the state-of-the-art has been substantially updated, numerous additional resources included and the geographic data and analysis presented have not previously been published.

1 Introduction

The distribution of illegally copied digital games via online or offline channels is a complex phenomenon that occurs across multiple channels and has a magnitude that is difficult to estimate (GAO, 2010), not the least due to the lack of clarity as to what constitutes illegal copying and copyright infringement internationally – i.e., what piracy actually is. As noted by Karaganis (2011): "*Piracy has never had a stable legal definition and is almost certainly better understood as a product of enforcement debates than as a description of specific behaviour*?" Nevertheless, the distribution of copies of digital games outside of the legal channels has been the cause of heated debate, with the users of illegally copied games on one side, and the game industry and legislators/policy makers on the other (Klepek, 2008; Sanchez, 2008; Andersen, 2012; Karaganis, 2011).

Both sides of the piracy debate are more or less in agreement when it comes to the conclusion that the phenomenon of digital game piracy is common, although the specific numbers vary between reports. For example, the Entertainment Software Association (ESA) claimed that it had traced 9.78 million 'illegal' downloads of roughly 200 digital games as on December 2009 (IIPA, 2010). TorrentFreak.com, one of the most influential websites on peer-to-peer (P2P) sharing via BitTorrent, 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 (TorrentFreak, 2010). However, neither of these studies used an open methodology, and therefore the conclusions cannot be verified independently.

This is a common problem: there is limited information available that originates in open sources from where results presented can be independently validated (Goldacre, 2009; Sanchez, 2008; GAO, 2010; Karaganis, 2011). This contrasts the research literature available on other digital products, e.g., software piracy, which has been investigated both at the individual- and country-level, and informed by fields such as economics, social psychology, criminology, business ethics and marketing (Yang et al., 2009). The information that does exist often comes from industry organisations or operators of P2P networks (Business Software Alliance, 2010; Envisional, 2011; IIPA, 2010; Kirk, 2010; BSA/IDC, 2009, 2010a, 2010b; OECD, 2009), but lacks objectivity and transparent

2

methodologies. This to the degree that Karaganis (2011) noted: "Criticising RIAA [Recording Industry Association of America], MPAA [Motion Picture Association of America], and BSA [Business Software Alliance] claims about piracy has become a cottage industry in the past few years, driven by the relative ease with which headline piracy numbers have been shown to be wrong or impossible to source". This problem led the United States Government Accountability Office (GAO) (GAO, 2010) to note that:

"Generally, the illicit nature of counterfeiting and piracy makes estimating the economic impact of IP infringements extremely difficult, so assumptions must be used to offset the lack of data. Efforts to estimate losses involve assumptions such as the rate at which consumers would substitute counterfeit for legitimate products, which can have enormous impacts on the resulting estimates."

The GAO (2010) report also criticises several widely cited industry reports, including BSA (BSA/IDC, 2009, 2010a, 2010b), noting that the study presented uses problematic assumptions, such as a one-to-one rate of substitution between pirated copies of software and resulting loss of sales.

The credibility of industry-based or government-solicited reports is also problematised by the common methodological problems in such work, which lead to bias or foster suspicion of bias (Ernesto, 2011; Goldacre, 2009; TorrentFreak, 2010). For example, Huygen et al. (2009) 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 (2011), a piracy research firm, used a snapshot methodology to examine 10,000 allegedly 'most popular torrents', concluding that music piracy was virtually gone from BitTorrent (2.9% of the torrents being music files, console games 2.8% and PC games 3.9%). This pattern was also noted by Ipoque (Schulze and Mochalski, 2009), but conclusions based on snapshots fail to account for file size and corresponding download times, leading to overestimation of the popularity of larger files. Anderson (2011) noted: "When a TV/movie company like NBC Universal funds a P2P study from a company that specialises in antipiracy work, the end result is hardly a disinterested piece of data". Finally, the potential impact of digital piracy on industry is notoriously hard to estimate reliably (Ernesto, 2011; Goldacre, 2009; TorrentFreak, 2010; Karaganis, 2011) and a reliable figure for digital games is unknown to the best knowledge of the authors.

2 Contribution and importance

The purpose of this paper is to address the need for objective information on game piracy. Analysis is presented on game distribution data obtained from tracking BitTorrent P2P file sharers, covering over 12.6 million unique peers from more than 250 countries/areas across a sample of 173 games, making it the first such study on computer games and one of the largest datasets on P2P activity to date. In addition, an overview of the current state-of-the-art in the game piracy debate is presented, and the challenges related to monitoring BitTorrent, commonly ignored in reports on the topic, are detailed (Bauer et al., 2009a).

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

(Schulze and Mochalski, 2009). For the present study, monitoring was carried out 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 (Xbox 360), PS3 (PlayStation 3), Wii (Nintendo Wii), DS (Nintendo Dual Screen) and PSP (PlayStation Portable).

The analyses presented indicate that the distribution of illegal copies of games follows distinct patterns, and a number of conclusions are reached.

- *Time-frequency patterns are not equally distributed (Section 5.2)*: The time-frequency distribution of torrents is generally assumed to be equally distributed across time; however, for game torrents at least this is not the case. A variety of temporal frequency distribution patterns are evident in the BitTorrent data presented here, with most titles following a distinctive pattern rapid increase-slow decrease pattern. This form of frequency distribution patterns has been previously reported for Google searches for movie torrents (Bauckhage, 2011).
- Asymmetric geographic distribution (Section 5.3): The traffic is geographically highly localised the 20 countries with the highest number of unique peers encompassed 76.7% of the total. The highest number of unique peers per country was recorded for China, the USA, Brazil, Italy and the Russian Federation, with each of these accounting for over 800,000 unique peers. When accounting for the relative population of the countries represented in the dataset, the peers/population ratio was highest for the countries Croatia, Greece, Portugal, Romania and Hungary. In addition, when investigating the peer data for the current study as a function of internet users in the specific countries, a number of countries stand out as having relatively high numbers of unique peers represented in the dataset, including Romania, Croatia, Ukraine, Greece, Poland, Italy, Armenia and Serbia. Portugal, Israel and Qatar also have more than 1% peers per internet user.
- Asymmetric distribution across titles and game genres (Section 5.4): The 10 most accessed games on BitTorrent in the sample encompassed 42.7% of the number of unique peers tracked. This conclusion is in alignment with results from other BitTorrent monitoring reports (e.g., TorrentFreak, 2010, 2011). The results presented here indicate that the major commercial (AAA-level), action-oriented titles account for the highest proportion of activity on the BitTorrent network. However, it also appears that games from other genres such as 'Racing' and 'Role Playing Game' games are individually more popular than the major action titles.
- *Review score correlates with BitTorrent popularity (Section 5.5)*: A high-aggregated review score (averaged across multiple online reviews) appears to positively correlate with popularity on BitTorrent, in terms of number of sharing peers (p < 0.05). This indicates the effect of positive recognition on BitTorrent activity for digital games; and suggests that game quality, as indicated by review scores, is related to piracy activity, such that higher quality games get pirated more frequently (if they are reviewed online).
- Entertainment Software Rating Board (ESRB) rating vs. popularity (Section 5.6): The analysis indicates that not only games with 'mature' and 'teen' ESRB ratings were popular on BitTorrent, but also games with suitable for everyone, and everyone above 10 years old, i.e., games for children and 'family' games. Although the mature-rated games are the overall most popular, this result indicates that

distribution of illegal copies of games via BitTorrent is distributed across ESRB ratings.

3 Distribution channels for game piracy

The channels through which game piracy operate are in many cases almost impossible to monitor effectively (Antoniades et al., 2009; Bauer et al., 2009a, 2009b; IIPA, 2010; Karaganis, 2011). Distribution channels include both physical and networked solutions, with the most common methods being P2P protocols (IIPA, 2010), '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 and internet relay chat have become less frequently used since the emergence of P2P protocols over a decade ago. Distribution channels for game piracy are described in more detail by Drachen et al. (2011), and for media piracy in general by Karaganis (2011), and will therefore only be briefly summarised here.

Physical distribution: The distribution of physically copies media can occur across networks that vary in size from small circles of friends to organised crime where businesses revolved around the mass duplication of pirated media, and distribution/sale at below-market prices (IIPA, 2010). The magnitude of this piracy channel is difficult to estimate accurately, but appears to be common in certain countries/areas, such as Italy (IIPA, 2010; Karaganis, 2011).

File-hosting services: File-hosting services consist of servers to where digital content can be hosted and accessed, 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 (Anderson, 2011).

P2P protocols: P2P protocols enable end-users to share content with one another directly, eliminating the need for uploading digital content to centralised servers for mass distribution. This is the only distribution channel that it is possible to monitor directly, as the BitTorrent protocol publicly shares information about the peers who 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 (BitTorrent Protocol Specification, 2010) or by crawling the BitTorrent distributed hash tables (DHTs). Recent evolution of BitTorrent has seen the incorporation of additional mechanisms for peer identification (in addition to the centralised tracker servers), such as DHTs, as well as a gossip-based mechanism called peer exchange (see e.g., BitTorrent Protocol Specification, 2010), which has made it progressively easier to identify peers. Finding and especially contacting peers remains challenging, however. Despite the challenge of finding and contacting peers hosting specific illegal digital content, in order to remove the infringing material from the network, entities acting on behalf of copyright holders have attempted to monitor P2P file transfers on a massive scale (Piatek et al., 2008). This has led to a form of arms race, where network operators and copyright holders, respectively, attempt to circumvent the opponent's methods for baffling vs. penetrating torrent networks (TorrentFreak, 2010). However, Piatek et al. (2008) showed how the techniques commonly employed to track content on torrent networks are prone to a variety of errors, leading to highly inaccurate estimates of piracy

activity and, worse, in some cases, falsely accusing innocent people of violating copyright laws.

The BitTorrent protocol is generally viewed as the major channel for game piracy today (IIPA, 2010; Kirk, 2010), and reportedly comprises a sizeable fraction of the traffic on the internet, with estimates varying from 40% to 60% (Schulze and Mochalski, 2009) it forms a starting point for investigating game piracy. Recently, Maier et al. (2009) have noted a shift in the distribution of internet traffic, claiming that the majority of internet traffic by volume is a result of hosting services and streaming media websites. On a final note, alternatives to estimating piracy rates, although bereft with an unknown degree of uncertainty, also exist specifically for game, such as tracking the percentage of calls to technical support from gamers playing with pirated copies (Ghazi, 2009).

4 Data and method

The data used here originate from BitTorrent data about the online distribution of digital games, and information about the games themselves. The process of obtaining these data comprises a series of steps, described in detail by Drachen et al. (2011), which can be summarised as follows.

4.1 BitTorrent

The BitTorrent protocol breaks down files into pieces of a specific size and distributes them across a network. Information such as a uniform resource identifier (URI) and cryptographic hashes contained in a metadata file are shared with a tracker server which lists all users making the file available (Bauer et al., 2009a, 2009b). The metadata file is distributed to the users via the server. BitTorrent search engines (e.g., thepiratebay.org and isohunt.com) host the metadata files, providing search capability for peers. A peer, having obtained the metadata for a particular file (or 'torrent'), can access the file via client-side BitTorrent software. This contacts the tracking server to obtain a randomly selected subset of the users currently sharing the file. 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).

For this project, a custom web crawler was developed to periodically issue queries to a BitTorrent search engine (which collates data from 500+ sites and 150,000+ trackers). Having located the metadata files, the web crawler obtained the tracker server URIs. Following an HTTP 'GET' request is issued to each URI over periodic intervals (a few minutes) to obtain a list of IP addresses for the peers who are currently participating in sharing.

A list of 173 game titles was compiled, including the best estimate of the authors all games released for any platform since 17 November 2010 until 6 February 2011 as well as a series of games released earlier in the fall of 2010 (note that games can appear on BitTorrent prior to the official launch date). There are no official repositories for game product information, and therefore over 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 locate games released during the period of monitoring. It is possible that titles, too small commercially to appear on the sites mentioned, were not included. Games legally being distributed via BitTorrent were

not included, unless torrents for these games which could be identified as versions not currently permitted to be shared were identified.

The raw torrent data were consolidated, giving the total number of peers for each game across torrents for different platforms and regional versions. The sample consisted of games of different genres and ranged from casual and indie titles to major commercial games. Within the period of tracking, 127 of the 173 titles were located on BitTorrent, and 12.6 million unique peer IP addresses identified. A total of 40 titles made their first appearance on BitTorrent during the period of tracking (encompassing 1.16 million unique peers).

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. Such torrents were eliminated by manually inspecting all torrents, leaving only those containing the full game. In addition, it is a known practice for copyright investigators to spread false information on BitTorrent, e.g., via distributing torrents which do not contain the actual game files (Bauer et al., 2009b). Standard publicly available blacklists were applied to filter out IP addresses controlled by well-known copyright investigators (IP Blocklist, 2010). About 200,000 IP-addresses were removed via this process.

4.2 Game information

There is no centralised reporting system for information about digital games, e.g., release dates and sales figures. Furthermore, developers and publishers often do not adhere to announced released dates. This provides a challenge in determining the relationship between release dates and appearances of a game on BitTorrent, as well as obtaining information on which platforms games are released for (such platform releases can also be staggered). However, a variety of websites provide parts of the required information with varying degrees of accuracy (e.g., vgnchartz.com, gamespy.com, ign.com, gamespot.com).

Genre: Games come in great variety, and various attempts have been made to categorise them. A typical approach is to define genres, but these are subject to some debate. Some genre categories are self-containing and games are typically easily allocated to them - e.g., 'racing games'. Game design varies, however, across a spectrum, and therefore many games share features across traditionally used genre systems, and can be difficult to categorise conclusively. In order to mitigate this problem and build as robust a framework as possible, the genre categories used here were based on majority consensus for each title individually, using genre categories aggregated across a variety of major game websites (e.g., mobygames.com, ign.com, gamespy.com, metacritic.com). A total of 17 genres were defined (Figure 6). See Drachen et al. (2011) for a detailed discussion of the genre categorisation.

Aggregated review scores: Aggregated review scores were based on data from several metacritic sites (metacritic.com, gamerankings.com, gamestats.com). Data for all games in the sample were not available from all metacritic sites, and aggregated review scores could not be found, or built manually by collecting individual reviews, for ten games, and these were eliminated from any analysis involving review scores. Aggregate 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).

4.3 Assumptions and limitations

Representativeness of sample: Roughly 1500 digital games are launched on a yearly basis via official channels (Rabin, 2010), plus an unknown number which do not reach distribution large enough to register on the international game sites. It is not known how many of these games are cracked and released as torrents on P2P networks. This means that it is difficult to estimate how representative the sample of 173 games is or calculate total BitTorrent distribution rates based on this number.

Underestimation: The BitTorrent protocol is generally regarded as the standard for distribution of files via P2P-networks (Bauer et al., 2009a), and forms a main channel for online piracy (IIPA, 2010), But BitTorrent is just one channel, and torrent activity alone underestimates the true scale of game piracy with an unknown factor.

Dynamic IP addresses/network address translators: The dataset presented offers a best estimate of the BitTorrent activity for the games sampled. This is because when building the list of unique peers used here, it is assumed that each IP address listed by the tracker servers corresponds to one participating peer. However, there may be peers who use dynamic IP addresses and network address translators (NATs) which leads to underestimation of the number of peers participating (e.g., multiple peers operating behind a NAT). Conversely, a single peer utilising dynamic IP addresses can appear to be several different IPs, notably due to the extended period of tracking.

Virtual Private Networks: Peers who wish to operate anonymously can utilise commercial Virtual Private Network (VPN) services (BTGuard: Anonymous BitTorrent Services, 2011) and techniques to introduce uncertainty into the tracker lists (Bauer et al., 2008; Choffnes et al., 2009, 2010). Furthermore, anonymous networks such as Tor can be used (The Tor Project, 2010). While there is evidence that some fraction of users participate anonymously in BitTorrent (Mccoy et al., 2008), the exact number is unknown.

Sample duration: The activity of peers in downloading files is not temporally homogeneous, and also varies across titles. In addition, torrents will only be available for a specific amount of time. This variation potentially biases measures that aggregate data across files. It is here assumed that any variance effect is randomly distributed.

Game feature information: 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 was employed in an attempt to mitigate these effects.

5 Analysis and results

Following pre-processing of the BitTorrent data and collection of associated data for each game, a subsample of 127 games remained from where torrent data and game-feature information were available.

5.1 Distribution across platforms

The sample encompasses games released for 14 different hardware platforms/OS, with the most common being PC, Xbox 360 (X360), PC, Playstation 3 (PS3) and Wii, and the more uncommon being: Android, Wii, iOS/Mac, webOS, NDS, PSP, Wp7, PS2, Java ME and N-Gage. Of the game titles 32 were single-platform releases, the rest multi-platform (X360, PS3, PC being a common combination for major commercial titles). And 12.6% were released for the Wii console, 63.8% for the X360, 49.6% for the PS3 and 64.6% for the PC. Remaining platforms, including OS' like webOS and iOS, account for a comparatively small fraction of the platforms the games in the sample are released for.

5.2 Frequency distribution

At first glance, the BitTorrent activity data contain a number of notable characteristics. Firstly, the frequency distribution of the number of unique peers per game is highly asymmetrical (Figure 1). A small number of games have been accessed a large number of times (100,000+), with the majority of the games having less than 50,000 unique peers observed, i.e., relatively limited popularity on BitTorrent during the interval measured.

Secondly, the time-frequency distribution of the torrent activity for each torrent as well for the sum of torrents for a specific game is not equal as a function of time. Earlier reports generally assume equal distribution across time, but there are a variety of frequency distributions evident in the dataset presented here (Figure 2(a)). The data presented here are aggregated across all torrents available containing a full version of the game in question. In general, at any given time, one or a few torrents are substantially more popular than the others, but the exact magnitude of the difference varies from game to game, and is most pronounced for major commercial titles ('AAA'-level games).

Focusing on the games in the sample where the initial appearance on BitTorrent lies within the period of monitoring, these typically (~60% of the examined titles) follow a rapid increase-slow decline pattern. The rapid increase in popularity typically lies on the first day or within the first days of the appearance of a torrent. The decline rate can be roughly linear (e.g., *TRON Evolution*, Figure 2(a)) or exponential (e.g., *Alien Breed 3: Descent*, Figure 2(b)). Other games display a more flat structure with a very slow decline ratio (e.g., *Lionheart: Kings Crusade*, Figure 2(b)), whereas *Tom Clancy's Ghost Recon* peaks three times before entering the decline phase, roughly correlating with new torrents becoming available. *Sports Island Freedom* works up to a protracted peak with a following gradual decline. In general, the curves for aggregated torrent activity have declined after a maximum of 60 days, from which point on activity is a fraction compared with the initial maximal values.

This variety in time-frequency patterns matches the only other source of raw torrent data for digital games known by the authors, collected by *Vigilant Defender*, a company that provides piracy protection services. The data were reportedly collected using Userenity, a variant of the Utorrent client. While *Vigilant Defender* is a commercial enterprise, and the data focused on the AAA-level games, the data, collected in 2010, are made freely available (http://www.vigilantdefender.com/TorrentWatch.php). The data from *Vigilant Defender* were collected for individual game torrents, and are presented as thus, not aggregated across all torrents for a particular game title as is the case here.

The time-frequency distribution of game torrents comprises an important finding, because they indicate that snap-shot analyses of BitTorrent activity, where data are

collected over a short interval, risk under-estimating or over-estimating the BitTorrent activity for a specific game title, depending on when the torrents lifecycle the snapshot is made. In conclusion, these results indicate that when working with small sample sizes of torrents, the time-frequency distribution of these needs to be considered to avoid biasing results; and furthermore that torrent activity for digital games – whether for individual torrents or aggregated activity varies substantially over time, and not according to any one pattern. This means that activity for a torrent cannot be estimated or predicted based on a just a single universal frequency model. While overall traffic can be estimated using a snapshot approach, the sample size would need to be very large in order to avoid problems with statistical significance.





Source: Modified from Drachen et al. (2011)

Figure 2 (a) and (b): Aggregated time-frequency structure for eight game titles (12 hour time bins). Data for TRON evolution has been divided by a factor of 2 to fit in the diagram



Figure 2 (a) and (b): Aggregated time-frequency structure for eight game titles (12 hour time bins). Data for TRON evolution has been divided by a factor of 2 to fit in the diagram (continued)



Source: Modified from Drachen et al. (2011)

5.3 Geographical distribution of peers

As noted above, the geographic location of the peers in the dataset was recorded, and aggregated according to country/area, encompassing a total of 251 countries/areas. The data indicate that the peers are highly distributed worldwide, although certain regions comprise hotspots of activity, largely correlated with the conclusions of Karaganis (2011) for media piracy in general. Aggregated across regions (Figure 3), the distribution of aggregated unique peers is highest for eastern (19.73%) and Western Europe (22.49%), respectively, with Asia (excluding Russian Federation and India) at a close third (18.61%) (Table 1). The 20 highest ranked countries encompass 9.67 million unique peers of 12.6 million total, i.e., 76.7%, indicative of a concentration in the activity of the BitTorrent peers for games. Remaining 195 countries account for only 23.3%.



Figure 3 Distribution of unique peers in the BitTorrent data as a function of geographic region

As noted above, several previous investigations into media piracy have measured piracy rates by using industry estimations which typically operationalise piracy rate as the ratio of products acquired legally to illegally. However, the raw peer numbers in themselves do not provide accurate information of the relative game piracy activity in the specific countries as they do not take into account local conditions. In order to reach a more accurate picture of the geographic distribution of the peers, accounting for population numbers provides a first step. When accounting for the relative population of the countries represented in the dataset, the peers/population ratio (P/P) was highest for the countries Croatia, Greece, Portugal and Romania (P/P ratios over 2%) (Table 2). Difference in peers per population is a factor of over 100 – from 5.2% for Croatia to 0.05% for India.

Furthermore, when investigating the peer data for the current study as a function of internet users in the specific countries (using data from the World Bank (2008 numbers), a number of countries stand out as having very high numbers of unique peers represented in the dataset, including Romania, Croatia, Ukraine, Greece, Poland, Italy, Armenia and Serbia. Portugal, Israel and Qatar also have more than 1% peers per internet user (Figure 4). It was only possible to obtain data on internet users for 125 countries (marked in Figure 4), which is, however, more than any previous cross-national investigation into national differences in piracy rates, to the best knowledge of the authors.

Table 1	Distribution of unique peers in the BitTorrent data as a function of geographic region.
	Percent of the total dataset indicated for each region (left side of table), and the
	highest ranked country within each region (right side of table)

Region	%/Peers	Country/highest rank	Peers (1000s)
Africa	4.56	Egypt	92.9
Asia (exl. Russia/India)	18.61	Philippines	175.7
Australia/New Zealand	1.41	Australia	173.3
Russian Federation	9.87	Russian Federation	1222.8
Eastern Europe	19.73	Poland	544.8
Western Europe	22.49	Italy	973.5
North America	8.35	USA	839.4
South America	10.40	Brazil	820.2
India (incl. Islands)	4.59	India	596.8

 Table 2
 Geographic distribution of unique peers in the BitTorrent data as a function of country, summarising the aggregated number of peers recorded, the population of the country and the peer/population ratio (P/P). Twenty highest ranked countries sorted in order of lowest to highest P/P ratio (rightmost column)

Country	Peers (1000s)	Population (mio.)	P/P
India	596.84	1220.20	0.0005
China	1052.74	1343.24	0.0008
Philippines	175.22	103.78	0.0017
Germany	217.83	81.90	0.0027
USA	839.37	309.05	0.0027

Patterns in the distribution of digital games via BitTorrent

Table 2	Geographic distribution of unique peers in the BitTorrent data as a function of
	country, summarising the aggregated number of peers recorded, the population of the
	country and the peer/population ratio (P/P). Twenty highest ranked countries sorted
	in order of lowest to highest P/P ratio (rightmost column) (continued)

Country	Peers (1000s)	Population (mio.)	P/P
France	242.27	62.64	0.0039
Brazil	820.20	205.72	0.0040
Spain	212.30	46.07	0.0046
UK	393.02	61.35	0.0064
Ukraine	319.95	44.85	0.0071
Australia	173.33	22.34	0.0078
Canada	265.41	34.11	0.0078
Russian Federation	1222.80	138.08	0.0096
Poland	544.79	38.19	0.0143
Italy	973.52	60.19	0.0162
Hungary	182.07	10.00	0.0182
Romania	470.81	21.85	0.0215
Portugal	305.84	10.63	0.0288
Greece	328.12	11.28	0.0291
Croatia	233.15	4.48	0.0520





Investigating the number of unique peers as a function of the total population, economic indices or even the IT-infrastructure of the countries (e.g., the number of internet users or internet Service Providers, ISPs) provide the means to evaluate piracy activity in relation to the people who actually have the opportunity and means to conduct illegal file sharing. This is a more useful indicator of the general attitude towards game piracy among a population than raw peer or download counts (e.g., reported by IIPA's (2010) for digital games based on data from the entertainment software association). Future research aims to investigate the relationship between additional economic indices and IT-infrastructure variables, e.g., the relationship between broadband availability and number of unique peers. Given the relative size of some game torrents (e.g., AAA-level games typically take up several GB of space), broadband availability may be an important constraint on network-based game piracy, but this hypothesis remains to be tested and certainly does not preclude physical piracy (IIPA, 2010; Kirk, 2010; Karaganis, 2011).

5.4 Distribution across titles and game genres

As noted above, the distribution of unique peers per game is highly asymmetric, with a few game titles being highly popular. The 10 most popular games in the sample accounted for 5.37 million unique peers (42.7%) alone (Table 3). All of these are AAA-level titles, from a variety of genres, including Role-Playing Games (RPGs), racing, action-adventure games, simulations and strategy games. This pattern matches the reports from TorrentFreak, one of the main communities for BitTorrent enthusiasts, who reported that the most downloaded games in 2011 (TorrentFreak, 2011), which comprise a similar mixture of game genres. TorrentFreak (2011) also report a difference in the most downloaded games across the four major hardware platforms (X360, PS3, PC, Wii), but it should be noted that the details of the method used by TorrentFreak to collect these numbers are not reported, apart from noting that they are estimated download numbers collected from all public BitTorrent trackers.

The games in the sample covered a variety of game genres, with most common being 'puzzle' games (n = 14, 11.02% of the titles in the sample), closely followed by RPGs (n = 13), TPS (third-person shooter) (n = 12), action-adventure and strategy games (n = 11) and FPS (First-Person Shooter) (n = 10). When considering the number of unique peers per game genre (Figure 5), it becomes evident that it is not the number of games available for a specific genre that drives BitTorrent popularity, but rather particular genres appear more popular than others. Notably, RPGs (18.9%), action-adventure (15.9%), TPS (12.7%) and racing (9.3%) games comprised the most popular genres. These game genres are also the most popular on a per-game basis, 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. Racing games account for 1.33% of the dataset each on average (with Need for Speed: Hot Pursuit being an important factor in driving up the average), followed by simulation games (1.08%, 7 games in sample) and TPS (1.06%, 12 games in sample). There were comparatively fewer recorded unique peers for the family, shooter, music, fitness and arcade genres, but these were also the most numerically under-represented in the sample (Drachen et al., 2011). This result may, therefore, be biased due to a lack of a hit game from these genres, which as noted above drives a substantial part of the BitTorrent traffic for games, during the period of monitoring.



Figure 5 Percentage distribution of the peers recorded for games within each genre (n = 127)

Source: Modified from Drachen et al. (2011)

5.5 Review scores vs. BitTorrent popularity

In order to explain what makes a game popular on BitTorrent, a number of variables can be posited, including the brand strength of the developer and publisher, the amount of marketing being done, the relative popularity in the gaming community, the availability of quality torrents compared with false torrents distributed by anti-piracy agencies and the quality of the game. Within the digital games industry, online and offline reviews are important drivers of sales and thus revenue, and the aggregated review score of digital games is generally related to the financial success of a game, although this is not always the case (Rabin, 2010), and substantial efforts leveraged towards ensuring that a company's games are reviewed and mentioned across the gamut of games-dedicated websites as well as in print magazines. As also reported in Drachen et al. (2011), in the current dataset, it can be observed that 7 of the 10 with the most unique peers recorded had aggregated review scores over 75 (on a 0-100 scale, a score of 75+ is considered 'generally favourable' by metacritic.com for the games category). Furthermore, major commercial game titles make up a high proportion of the 127 games sample, i.e., of the proportion of the 173 games that were eventually illegally distributed on BitTorrent. While there have been cases of games receiving positive reviews and still not doing well in the marketplace, receiving negative reviews is usually considered to reduce sales, notably for the AAA-level games. Here aggregate review scores are used as a proxy for the quality of a game, and tentatively commercial success, and correlating these with BitTorrent popularity indicates that these two variables are positively correlated at p < 0.05. This was calculated using 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 < 0.05significance (two-tailed) given df = 115 (n-2)) (Sigle, 2010). Please note that log (unique peers) was used due to the non-normal distribution of the peers data.

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 18 January 2011 for PS3, average review score 90; 1056 peers), although may be an

artefact of a late release date during the period of tracking (i.e., a short period where tracking of the game title was carried out). In general, casual games were less frequently pirated, with a few exceptions, e.g., *Bejeweled 3* with over 250,000 unique peers.

5.6 ESRB rating vs. popularity

ESRB rating is a system for assigning age and content ratings to computer games and mobile applications, and assists with enforcing advertising and marketing guidelines for the computer game industry (for additional information see: http://www.esrb.org/index-js.jsp). There are six categories, of which four were assigned to games in the current sample. 'M' (mature) or 'T' (teen) ratings are generally mentioned on 'most downloaded' lists released by torrent sites (e.g., TorrentFreak, 2010). However, the current dataset indicates that distribution of illegal copies of digital games via BitTorrent is not limited to games with these ratings, although the M-rated games remain the overall most popular. Specifically, 36 of 127 games in the sample carried an ESRB rating 'E' (Everyone), 16 'EE10+' (Everyone 10+ years), 33 'M' (Mature) and 39 'T' (Teen). ESRB ratings could not be located for three of the games in the sample. 'M' rated games were popular in terms of the percentage of the unique peers encompassed by these games in the dataset (37%), corresponding to 1.12% of the total dataset per game, compared with 14% total/0.39% per game for 'E'-rated games, and 25% total/0.69% per game for 'T'-rated games, but 24% total/1.5% per game for 'E10+' rated games.

6 Discussion and conclusions

The piracy debate in the digital games field remains heated and illegal copying and distribution of games is accused with driving a shift in the industry away from the PC platform over to consoles, which are somewhat harder to use on pirated games. At the same time, the industry is increasingly moving towards an online platform for both major commercial and casual games, which heavily reduces the risks of piracy. There is, however, a general lack of objective information on the magnitude of game piracy, its distribution across different channels and the motivations of the people behind it, as well as how to effectively manage - or take advantage of - piracy (Drachen et al., 2011; Karaganis, 2011). Here a step is taken towards addressing this knowledge gap, based on a 12.6 million peers large dataset obtained over a three-month period via monitoring BitTorrent, one of the main channels of online digital game piracy (Bauer et al., 2009a; The Tor Project, 2010). The work presented here, and in the preceding publications from the research project driving it (e.g., Drachen et al., 2011), provides a quantitative basis for the game piracy debate, and provides the basis for beginning to address the why? and how? questions in the debate. Unlike the majority of previous work, the data reported here are objective, quantitative and developed using state-of-the-art techniques and with a public and open methodology.

The findings present a picture of P2P distributions that confirms some existing assumptions about piracy while contradicting others. First and foremost, P2P game piracy is extraordinarily prevalent and geographically distributed. Of 173 game titles released during the fall of 2010 and early 2011, 127 were found on BitTorrent networks, and over 12.6 million unique peers accessed these files (close to 100,000 peers on average per game) from over 250 countries/areas. Second, while it is typically major commercial

titles that drive the BitTorrent traffic (Table 3), it is not only AAA-level action games that are pirated – casual games and family games but they are also distributed via BitTorrent. It appears that good review scores correlate with piracy rates (p < 0.05), although whether review scores are a causal factor which cannot be concluded based on the present analysis. It is the case, however, that a small proportion of the games drive a large fraction of the traffic: The ten most distributed titles encompass 41.8% of the total dataset. 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.

In terms of geographic distribution, Europe and Asia (excluding India and Russian Federation) comprise the regions with the most activity, but on a peers/population basis it is notably countries from South-Eastern and Eastern Europe that are the most active. These results partially contrast patterns from general software piracy, which highlight, e.g., East Asia as having high piracy rates (e.g., IIPA, 2010; Kirk, 2010; BSA/IDC, 2009, 2010a, 2010b). This is also the case that the geographical distribution of peers is held up against the number of internet users, and may indicate that due to especially major commercial game titles constituting big files (several GBs), BitTorrent-based game piracy is mostly available when enough broadband is available to effectively download these files via P2P networks. In emerging economies, physical copying and distribution maybe more effective means of game piracy (IIPA, 2010; Kirk, 2010; Karaganis, 2011).

Table 3The 10 most popular game titles on BitTorrent encompass 5.37 mio. unique peers
in the dataset, averaging 536,727 peers per game and an average aggregate review
score of 74.5 on a 100 point scale

Title	Genre	Unique peers	Avg. review score	Developer
Fallout: New Vegas	RPG	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	Sports	545,559	86.7	Visual Concepts
TRON Evolution	Action Adventure	496,349	59.5	Propaganda Games
Call of Duty: Black Ops	FPS	469,864	83.8	Treyarch
Starcraft 2	Strategy	420,138	89.5	Blizzard Entertainment
Star Wars the Force Unleashed 2	Action Adventure	415,021	61	Lucas Arts
Two Worlds II	RPG	388,236	73.3	Reality Pump
The Sims 3: Late Night	Simulation	356,771	77.5	The Sims Studio

Source: Modified from Drachen et al. (2011)

Acknowledgements

The authors extend their warmest gratitude to Ioanna Constantiou, Copenhagen Business School (CBS) and Jacob Nørbjerg (CBS) for advice and feedback; and to Copenhagen Entertainment and Copenhagen Business School for financial support.

References

- Andersen, N. (2012) Piracy Problems? US Copyright Industries Show Terrific Health, Ars Technica, 3 November, 2011, Available from: http://arstechnica.com/techpolicy/2011/11/piracy-problems-us-copyright-industries-show-terrific-health/
- Anderson, N. (2011) Where Have all the Music Pirates Gone? Ars Technica, 4 February, 2011, Available from: http://arstechnica.com/tech-policy/news/2011/02/where-have-all-the-musicpirates-gone.ars
- Antoniades, D., Markatos, E.P. and Dovrolis, C. (2009) 'One-click hosting services: a file-sharing hideout', Proc. of the 9th ACM SIGCOMM Conf. on Internet Measurement, ACM, pp.223–234.
- Bauckhage, C. (2011) 'Where the pirates are', *Proceedings of WebSci'11*, 14–17 June, Koblenz, Germany.
- Bauer, K., Grundwald, D. and Sicker, D. (2009a) 'The arms race in P2P', Proceedings of the 37th Research Conference on Communication, Information and Internet Policy, TPRC, pp.1–18.
- Bauer, K., McCoy, D., Grundwald, D. and Sicker, D. (2009b) 'Bitstalker: accurately and efficiently monitoring BitTorrent traffic', *Proceedings of the First IEEE Workshop on Information Forensics and Security*, IEEE Publishers, pp.181–185.
- Bauer, K., McCoy, D., Sicker, D. and Grunwald, D. (2008) 'BitBlender: light-weight anonymity for BitTorrent', Proceedings of the 4th ACM Workshop on Applications of Private and Anonymous Communications, ACM, pp.1–8.
- BitTorrent Protocol Specification (2010) Available from: http://wiki.theory.org/ BitTorrentSpecification
- BTGuard: Anonymous BitTorrent Services (2011) Available from: http://btguard.com/
- Business Software Alliance (2010) Seventh Annual BSA and IDC Global Software Piracy Study, Available from: http://www.bsa.org/country/Research%20and%20Statistics.aspx
- Business Software Alliance and International Data Corporation (BSA/IDC) (2009) *Sixth Annual BSA-IDC Global Software Piracy Study*, BSA, Washington DC, Available from: http://global.bsa.org/globalpiracy2008/studies/globalpiracy2008.pdf
- Business Software Alliance and International Data Corporation (BSA/IDC) (2010a) The Economic Benefits of Reducing PC Software Piracy, BSA, Washington DC, http://portal.bsa.org/ piracyimpact2010/index.html
- Business Software Alliance and International Data Corporation (BSA/IDC) (2010b) 2009 Global Software Piracy Study, BSA, Washington DC.
- Choffnes, D.R., Duch, J., Malmgren, D., Guierm, R., Bustamante, F.E. and Amaral, L. (2009) SwarmScreen: Privacy Through Plausible Deniability in P2P Systems, EECS Technical Report, Northwestern University, March.
- Choffnes, D.R., Duch, J., Malmgren, D., Guiermà, R., Bustamante, F.E. and Amara, L. (2010) 'Strange bedfellows: community identification in BitTorrent', *Proceedings of the 9th International Conference on Peer-to-Peer Systems*, USENIX Association, pp.13-18.
- Drachen, A., Bauer, K. and Veitch, R.W.D. (2011) 'Distribution of digital games via BitTorrent', Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, ACM Publishers, pp.233–240.
- 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.
- Ernesto (2011) 'Arrr! The music pirates are still here', *TorrentFreak*, 7 February 2011, Available from: http://torrentfreak.com/arrr-the-music-pirates-are-still-here-110207/
- Ghazi, K. (2009) *PC Game Piracy Examined*, TweakGuides, June, Available from: http://www. tweakguides.com/Piracy_1.html

- Goldacre, B. (2009) 'Illegal downloads and dodgy figures', *The Guardian*, 5 June 2009, Available from: http://www.guardian.co.uk/commentisfree/2009/jun/05/ben-goldacre-bad-science-music-downloads
- 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. Available from: http://www.tno.nl/content.cfm?context =thema&content= inno_publicatie& laag1=897&laag2=918&item_id=473
- International Intellectual Property Alliance (IIPA) (2010) 2010 Special 301 Report on Copyright Protection and Enforcement, Available from: www.iipa.com
- IP Blocklist (2010) Available from: http://www.iblocklist.com/list.php? list=bt level1
- Karaganis, J. (Ed.) (2011) Media Piracy in Emerging Economies, Social Science Research Council, USA, ISBN 978-0-98412574-6, Available from: http://piracy.ssrc.org
- 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
- Klepek, P. (2008) 'Bethesda is tired of spending money supporting software pirates', MTV Multiplayer, 13 October 2008, Available from: http://multiplayerblog.mtv.com /2008/10/13/bethesda-deals-with-pirates/.
- Maier, G., Feldmann, A., Paxson, V. and Allman, M. (2009) 'On dominant characteristics of residential broadband internet traffic', *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement*, ACM, pp.90–102.
- Mccoy, D., Bauer, K., Grunwald, D., Kohno, T. and Sicker, D. (2008) 'Shining light in dark places: understanding the Tor network', *Proceedings of the 8th International Symposium on Privacy Enhancing Technologies*, ACM, pp.63–76.
- Organisation for Economic Co-operation and Development (OECD) (2009) *Piracy of Digital Content*, OECD, Paris, Available from: http://www.oecd.org/document/35/0,3343, en_2649_34223_43394531_1_1_1_1_0.0.htmlTorrentFreak, most downloaded 2011: http://torrentfreak.com/top-10-most-pirated-games-of-2011-111230/
- 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', *Proceedings of* the 3rd Conference on Hot Topics in Security, USENIX Association, pp.1–7.
- Rabin, S. (2010) Introduction to Game Development, Charles River Media, Independence, KY.
- Sanchez, J. (2008) '750,000 lost jobs? The dodgy digita behind the war on piracy', Ars Technica, 7 October, 2008, Available from: http://arstechnica.com/tech-policy/news/2008/10/dodgydigits-behind-the-war-on-piracy.ars
- Schulze, H. and Mochalski, K. (2009) *Internet Study 2008/2009*, Available from: http://www.ipoque.com/study/ipoque-Internet-Study-08-09.pdf
- 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.
- The Tor Project (2010) Anonymity Online, Available from: http://www.torproject.org/
- TorrentFreak (2010) 'Call of duty black ops most pirated game of 2010', *TorrentFreak*, 28 December 2010, Available from: http://torrentfreak.com/call-of-duty-black-ops-mostpirated-game-of-2010-101228/.
- TorrentFreak (2011) most downloaded: http://torrentfreak.com/top-10-most-pirated-games-of-2011-111230/.
- US Government Accountability Office (GAO) (2010) Intellectual Property: Observations on Efforts to Quantify the Economic Effects of Counterfeit and Pirated Goods, GAO-10-423, Washington DC, 41 pp, Available from: http://www.gao.gov/new.items/d10423.pdf
- Yang, D., Sonmez, M., Bosworth, D. and Fryxell, G. (2009) 'Global software piracy: searching for further explanations', *Journal of Business Ethics*, Vol. 87, pp.269–283.

Bibliography

- Banerjee, D., Khalid, A.M. and Sturm, J-E. (2005) 'Socio-economic development and software piracy. An empirical assessment', *Applied Economics*, Vol. 37, pp.2091–2097.
- De Weck, J. and Mawad, M. (2012) Free Online Games Sink Pirates to Unlock Emerging-Market Growth, Bloomberg BusinessWeek, 27 August, 2012.
- Ellison, B. (2008) World of Goo Co-Creator Claims 90% Piracy Rate, Shacknews, 13 November, 2008, Available from: http://www.shacknews.com/onearticle.x/55906
- 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', *Proceedings of the 3rd USENIX Workshop on Large-Scale Exploits and Emergent Threats*, USENIX Association, pp.4–12.
- Orland, K. (2012) Ubisoft Boss: Free-to-Play A Natural Reaction to High Piracy Rates, Ars Technica, 22 August, 2012, Available from: http://arstechnica.com/gaming/2012/08/ ubisoft-boss-free-to-play-a-natural-reaction-to-high-piracy-rates/?comments=1#comments-bar
- Stryszkowski, P. and Scorpecci, D. (2009) Piracy of Digital Content, OECD Publishing, Paris, France.
- Wolchok, S. and Halderman, J.A. (2010) 'Crawling BitTorrent DHTs for fun and profit', *Proc. of the 4th USENIX Workshop on Offensive Technologies*, USENIX Association, pp.112, 113.