

Internet Piracy: CMU Undergraduate Behavior & Opinions

James Bogie | Ben Gorman | Chelsea Grindle | Jong Woo Lee | George Nardi | George Volichenko
DRAFT | 4/5/12 | 36-303 | Junker | DRAFT

Section 1: Introduction

In the age of the internet, the unrestricted dissemination of information worldwide poses a real threat to intellectual property. In the U.S., intellectual property rights are protected via trademarks, trade secrets, patents, and copyrights. Copyrights are used to protect creative work designed for entertainment, such as music, books, and movies. Given the universal popularity of these goods and the ease with which they can be digitally replicated, copyrighted materials are those whose intellectual property protection has been most undermined by internet-based information trafficking. According to an August 2011 study from The American Assembly, 46% of adults have bought, copied, or downloaded unauthorized music, TV shows, or movies. Among 18-29 year olds, 70% have acquired music or video files.

Advocates of stricter copyright enforcement online have pushed hard for more stringent protection of IP, resulting in the proposed SOPA (Stop Online Piracy Act) and PIPA (Protect Intellectual Property Act) legislation. Support for the bills seemed strong in early 2012, but a flood of opposition at both the grassroots level and from internet fixtures like Wikipedia and Google curtailed that support. Google hired at least 15 lobbying firms to fight the bills, and Mozilla put a link to a petition on its homepage with the warning “Congress is trying to censor the internet”¹. Many websites also had a “blackout” to increase awareness of the proposed bills. An article from *The Ithacan*, Ithaca College’s newspaper, reported that Wikipedia’s blackout page had 162 million visitors, and Google collected more than 7 million signatures for the petition against the bills². We’re curious to know whether the opposition to regulation (at least among those in our community) stems from a concern about government regulators hastily tampering with the internet, or if the opposition is simply backlash from those of us who have gotten used to enjoying pirated content. Copyright in the age of the internet is a complex legal matter, with strong arguments from both sides. Lawrence Tribe, a law professor at Harvard, opposes SOPA and even considers it illegal because some of the would-be-removed content is constitutionally protected speech.³

As the aforementioned American Assembly survey shows, internet piracy is a rather widespread behavior among young people. However, the *extent* to which people participate in these behaviors is still unknown. Furthermore, it has yet to be explored how young people view these activities in terms of ethics and legality, and whether their views align with their behavior. The intent of our survey is to

¹ Lines Drawn on Antipiracy Bills, Edward Wyatt.
<http://www.nytimes.com/2011/12/15/technology/lines-are-drawn-on-legislation-against-internet-piracy.html?pagewanted=all>

² “Content Blocked: Students Combat Anti-piracy Bills”. Elma Gonzalez.
<http://theithacan.org/20835>

³ <http://joi.ito.com/weblog/2012/01/15/why-we-need-to.html>

measure how common internet piracy is on campus and how aware undergraduates are of recently proposed changes to internet piracy laws. Among those engaging in piracy, we want to know how often they do it, what types of content they download or stream, and what types of websites they use. People who claim to not engage in piracy are also of interest to us. We would like to learn more about their attitudes toward piracy, particularly by finding out what keeps them from pirating movies, shows, and music online. We are collecting demographic information from our respondents to both mitigate potentially confounding results and to learn about piracy behavior differences between demographics.

While we are still early in the data collection process, right now it appears that many students oppose the SOPA PIPA bills, engage in streaming and downloading, but still feel that their needs to be changes made to copyright enforcement policies.

Section 2: Methods

Target population

Our target population is the undergraduate student body at Carnegie Mellon (Pittsburgh, Pennsylvania). The only difference between the target population and the sampling frame may occur as a result of the continued inclusion of students in the C-book who have dropped out, transferred, or taken a leave of absence. We considered contacting those students via Facebook, but if we only sent out survey invitations this way, we would have a sampling error, as not every student has a Facebook profile, not to mention there might be a significant difference in online-behavior between Facebookers and non-Facebookers. So, we decided to send out surveys via e-mail, and, depending on response rates, may later supplement e-mails with friend-agnostic Facebook solicitations (i.e., send Facebook messages to users with whom we may or may not be friends, simply using Facebook's messaging capabilities as a proxy for email). We anticipate a large nonresponse error: historically, most people who receive an e-mail asking them to take a survey discard the message or otherwise do not participate. Asking enough people to take the survey should get an adequate number of responses to compensate for non-responses. There is also some potential for measurement error, as someone may misread or misinterpret a question. Thus, we plan on emphasizing unequivocal language in the survey, and providing clarifying definitions when necessary.

Sampling Frame

We are sampling from Carnegie Mellon undergraduate students, using the Carnegie Mellon C-Book as the sampling frame. We generated a random sample from the set of page numbers of the C-Book with R . Then, all the undergraduate names on those pages were recorded in the sample. Sometimes the surnames on a given C-Book page are homogeneous enough that we need to exclude that page in order to avoid over-representing a certain demographic, e.g., "Kim" represents 96% of the surnames on page 58.

Sample Design

We decided to use a self-administered, web-based survey with email as the contact method. Sending the e-mail solicitation represents the only instance of "interviewer involvement" or degree of interaction with the respondent. Results from the survey were tabulated to a third-party host site, a GoogleDocs-based survey generator. Regarding privacy, we emphasized that answers are strictly confidential. If respondents wanted a chance at winning a prize, providing an email address was optional. The survey was text-only (no audio or video). The technological requirements of respondents is

simply filling out the survey online using a computer and its software (this may cause programming error, but it is unlikely given our sample population).

Sample Size

To get an idea of an appropriate sample size for our purposes we considered one of the many important questions on our survey: "Have you ever streamed pirated content?" Ideally, we want the margin of error to be as small as possible. Using 0.01 as the margin of error, the required sample size would have to be 3,760. Since that is not feasible due to the time limits of the project, we tried a margin of error of 0.05 instead. This accuracy level requires a much more realistic sample size of 362.

Using this result and assuming a 25% response rate, we are sending the survey out to roughly 1448 undergraduate students at Carnegie Mellon. So far we have sent out an initial sample of 196 surveys to students. We have received 31 responses so far. This equates to a response rate of about $(31/196)=15.81\%$ in less than 24 hours. We will wait another day most likely and then assess our response rate. At that point, we will either continue to send out surveys via email, or use Facebook as well, depending on whether our current response rate is satisfactory. As of the 31 responses, there were no skipped questions, which we were very happy about. Only 5 did not add their andrew IDs to be entered in the raffle. This leads us to believe that the raffle idea is improving our response rate.

Survey Setup

Below are a few example questions from the different sections of our survey:

Background questions (SOPA/PIPA attitudes)

Example) Q.1 Have you heard of the recently proposed SOPA/PIPA bills?

- Yes or No

(Regardless of answer, we then provide summary information on both bills before the respondent completes the rest of the survey)

Personal piracy behavior questions

Example) Q.6 Have you ever streamed pirated content?

Example: Stream a TV show from Megavideo/SideReel.

- Yes or No (next question conditional on response)

Potential scenario questions

Example) Q.15 You download this movie from dtella (or another peer-to-peer file-sharing network) and watch it on your computer.

Do you consider this to be (check all that apply):

- Acceptable

- Illegal

- Unethical

- No opinion on the ethics or legality of this action

Demographic questions

Example) Q.26 What school are you in? (If you are in multiple schools, please select the school of your primary major.)

- CFA, MCS, CIT, HSS, Tepper, SCS

Post-survey processing

While we generated our sample population randomly from the CMU undergraduate body, there is a good chance that the actual demographics of our respondents may differ materially from those of the student body. Demographics that we suspect will be different from the student body are race, gender, and school within CMU. To compensate, we will need to apply post-stratification weights to our results. These weights will allow us to fix disproportionate sampling of post strata and disproportionate nonresponse across post strata.

Based on our results thus far, we may need to limit how many strata we choose to weight because of a lack of responses from certain demographics. For instance, right now we have no responses from black students. While we could use imputation to apply weights to races, it is likely not worth the effort. Stratifying with respect to class year will probably be an easier process and provide greater benefit to our results.

Section 3: Results

We have two main research questions. The first is “To what extent do students engage in piracy?” and the second is “What are students’ opinions on internet piracy and copyright enforcement policy?”. Based on the first 31 responses that we gathered in less than a day after sending out the first batch of emails, these are our findings:

The vast majority (90%) of respondents have heard of the SOPA/PIPA bills and 71% of all respondents somewhat oppose the bill. 13% are neutral and only 16% support them. Interestingly, while such a large percentage oppose the bills, 74% are convinced that changes need to be made with respect to copyright protection policies. These percentages answer our question concerning peoples’ views on piracy and whether or not the correct steps are being done to prevent it.

Something we were expecting to observe is that students engage in piracy via streaming a lot more than by directly downloading content. 81% of our respondents reported to have streamed pirated content, while a lower proportion of 61% have pirated by downloading.

Only 23% of respondents were unaware when they were viewing/using/listening to pirated content. This shows that a majority of respondents are aware of their habits, but still do them regardless.

In terms of downloading pirated content, most people acknowledged that it was illegal (55%) and most people agreed that it was unethical (61%). This contrasts with 39% of respondents who thought it was acceptable.

However, when the question regarded simply the streaming of pirated content, only 26% believed that to be illegal, and the majority now switched to respondents thinking this was acceptable at 58% to unethical at 42%.

For each of these questions, 13% of the respondents had no opinion on the subject. These questions were meant to address people’s attitudes toward different types of piracy and see if their opinions

would change depending on the situation. Our results closely followed our hypothesis that people would be more accepting of streaming pirated content than downloading it.

We then proceeded to ask whether or not students' attitudes would change if the US was more strict about pirated content. A surprising 52% said that their attitudes would not change at all. Of the remaining 48% that said their attitudes would in fact change, only 3% (1 respondent) said his/her attitude would change a lot. This question allows us to see what sort of reaction the students would have if the SOPA/PIPA bills would pass. It is fairly clear to us that the passing of the bills would either not change students' attitudes, or change them a little bit. No drastic measures would be taken.

Now taking a look at the demographic questions, a very large 84% of respondents were White/Caucasian. We were very intrigued to see that 52% of respondents were female, when we expected a large majority to be male.

Section 4: Discussion

So far, it appears that our research questions have been effective in gathering the information we hoped to glean from respondents. The responses we have received to our open-ended questions have not been clarifications/more detailed responses to the structured questions, but instead additional thoughts and opinions that we hoped to elicit for those specific questions. The fact that a majority of respondents have chosen to express themselves further shows that our survey is on a topic that is on people's minds (at least those that respond; obviously strong self-selection bias here).

As for specific questions, we have observed some interesting results. It is not exactly unexpected that people would feel more comfortable pirating from a site like YouTube (a Google subsidiary) rather than the back-alleys of the internet. It is interesting to ask "why", given the outcome of the survey so far. While you can argue that YouTube is built on a foundation of videos that have violated copyrights in one way or another, in its early days and still today, many of the videos are created by its users. The original content on the site provides the website with some legitimacy, relative to sites that have lost big in court like Napster. YouTube developed itself into something a "brand" and a "destination site";⁴ this could be why Google Video never did as well as YouTube: Google acquired YouTube for \$1.65B, 10/2006. With newfound deep pockets, YouTube found itself the center of attention for potential litigation. Suing a (for the most part) user-operated site remains complicated, but YouTube has had to negotiate with companies like Sony, Universal, CBS, and the NHL to give them what amounts to free advertising.⁵ There's no doubt that the online-copyright climate is changing, but it does not seem like YouTube is going away as long as it does a good job of placating its detractors. So while much of the content on the site is in a legal grey area, there it retains some legitimacy, and this is very clear in our findings: 82% of respondents indicated it was acceptable to watch what amounts to pirated material on

⁴ Niko Flores, How YouTube Works (Oct. 11, 2006), http://ondemandmedia.typepad.com/odm/2006/10/how-youtube_wor.html

⁵ Jason Breen, YouTube or YouLose: Can YouTube Survive a Copy Infringement Lawsuit? (2007) <http://igeneration.edublogs.org/files/2008/09/youtube-or-youlose.pdf>

YouTube, while just 61% of respondents considered streaming pirated content on the internet-in-general acceptable.

For quantitative data on downloaded content, most people chose "Few Times a year", perhaps because it was the most convenient to choose (if respondents were not sure how many times they had downloaded pirated content). It will be interesting to see how this response metric averages out as we get more responses.

As of yet, it appears that our survey does not have any major flaws. The main challenge going forward is getting enough respondents using our initial contact method (random emails).

The main conclusion we can make thus far is that the majority of the respondents strongly oppose the SOPA/PIPA bills, but feel that changes need to be made with respect to copyright protection on the internet. Parallel to their opinion on the bills, most respondents have downloaded pirated content. These results match our expectations coming into the project.

Appendix

Survey

CMU Internet Piracy Survey

In recent months, internet piracy and copyright enforcement have been at the forefront of legislative debates, media coverage, and corporate public statements. We are surveying Carnegie Mellon undergraduates to see how often they engage in internet piracy and what views they hold on related copyright enforcement legislation. You will be able to enter your Andrew ID on the last page of the survey if you want to participate in the raffle for a chance to win an Amazon gift card.

* Required

Have you heard of the recently proposed SOPA/PIPA bills? *

- Yes
- No

The SOPA Bill involves the following: - Expansion of the ability of U.S. law enforcement to fight online trafficking in copyrighted intellectual property and counterfeit goods - Expansion of existing criminal laws to include unauthorized streaming of copyrighted content, with a maximum penalty of a five-year prison sentence - Allowance for court orders to bar advertising networks and payment facilities from conducting business with infringing websites, block search engines from linking to the offending sites, and require Internet service providers to block access to the sites The PIPA Bill involves: - Prevention of real online threats to economic creativity and theft of intellectual property - Provisions for the U.S. government and copyright holders to utilize additional tools to curb access to online distribution and sales of pirated content and counterfeit goods, especially those registered outside the U.S.

Do you support the policies outlined in the definitions of the proposed bills?

- Strongly Oppose
- Oppose
- Neutral
- Support
- Strongly Support

Do you feel changes need to be made with respect to copyright protection in the internet era?

- Yes
- No

Have you ever downloaded pirated content? 'Pirated content' being copyrighted material used/reproduced without the consent of the copyright holder. Example: Download a movie from a torrent.

- Yes
- No

How often do you download pirated content?

- Daily
- Weekly

- Few times a month
- Few times a year
- Only once

Have you ever streamed pirated content? Example: Stream a TV show from Megavideo/SideReel.

- Yes
- No

How often do you stream pirated content?

- Daily
- Weekly
- Few times a month
- Few times a year
- Only once

Do you use ... ? Check all that apply.

- Streaming sites (e.g. sidereel.com)
- Peer-to-peer file-sharing client(s) (e.g. Dtella, LimeWire)
- Direct download sites (e.g. rapidshare.com)
- I do not use any of the above

Do you usually know whether the content you are viewing/using/listening to is being reproduced illegally or not?

- Yes
- No

Do you consider downloading pirated content to be...? Check all that apply.

- Acceptable
- Illegal
- Unethical
- No opinion on the ethics or legality of this action

Do you consider streaming/viewing/listening to unlicensed content (without downloading) to be...? Check all that apply.

- Acceptable
- Illegal
- Unethical
- No opinion on the ethics or legality of this action

For example, do you consider viewing/listening to music on YouTube that has NOT been posted by the original artist (i.e. NOT through a VEVO or official artist account) to be...?Check all that apply.

- Acceptable
- Illegal
- Unethical
- No opinion on the ethics or legality of this action

If the US adopted a tougher stance on internet piracy enforcement, to what degree would your attitudes about using pirated content change?

- Not at all
- Somewhat
- A lot
- Unsure

Scenarios

Suppose you want to watch a movie that you have anticipated for a long time.

You download this movie from dtella (or another peer-to-peer file-sharing network) and watch it on your computer. Do you consider this to be...?Check all that apply.

- Acceptable
- Illegal
- Unethical
- No opinion on the ethics or legality of this action

You cannot find the movie on dtella, but you notice it has been uploaded in multiple parts to YouTube. You watch the movie in its entirety on YouTube. Do you consider this to be...?Check all that apply.

- Acceptable
- Illegal
- Unethical
- No opinion on the ethics or legality of this action

After watching the movie on YouTube, you save it to a file on your computer and upload the file to a sharing website. You personally do not profit in any way from this activity. Do you consider this to be...?Check all that apply.

- Acceptable
- Illegal

- Unethical
- No opinion on the ethics or legality of this action

Do you think there is a way to watch the movie that is as convenient as the methods above but requires payment (potential sources include iTunes, Netflix, etc.)?

- Yes
- No

Do you have any other comments or views on internet piracy that this survey has not yet addressed?

Please express them here.

Have you had any personal experiences that caused a change in your attitude towards internet piracy and copyright enforcement? Please describe them here.

Demographic Information

Gender

- Male
- Female

What is your ethnicity? Check all that apply.

- White/Caucasian
- African American
- Hispanic/Latino
- Asian/Pacific Islander
- Native American
- Other:

What is your class year?

- Freshman
- Sophomore
- Junior
- Senior
- 5th-year senior

Do you live on campus or off campus?

- On campus

- Off campus

What is your political affiliation?

- Republican
- Democrat
- Independent
- Other
- None
- Do not wish to share to information

What school are you in?(If you are in multiple schools, please select the school of your primary major.)

- CFA
- HSS
- CIT
- SCS
- Tepper
- MCS

If you are from the United States, or have lived in the United States for the majority of your life, what region are you from?If you are not from the US, please pick Other and write down where you consider yourself from.

- Northeast
- Midwest
- South
- West
- Other:

To participate in the raffle enter your Andrew ID in the box below.We assure you that your ID will not be linked to your responses