

Group B

Chelsea Grindle
George Nardi
George Volichenko
Ben Gorman
Jongwoo Lee
James Bogie

A.

- Why is this topic interesting?/Why does it need to be now?

Recently the SOPA/PIPA bills proposed in Congress have received a lot of exposure in the media. Protests led by Google and Wikipedia amplified the publicity of this issue. To make the matter even more discussed, the US government shut down one of the main hubs for pirated media, Megaupload. In light of these recent developments we would like to investigate the gravity of Internet piracy. We understand that not everyone will have heard about SOPA and PIPA, however, we would still like to inquire about their attitudes toward downloading media illegally and whether they have done it themselves.

This survey could also offer us insight into other possible ways of legally acquiring music. For example, if sites like Megaupload offered free downloadable content and then shared their huge revenues from advertising with the artists themselves, that could be a compromise beneficial for everyone.

- Is there a client?

District representatives will be interested in knowing how a segment of their constituents feel about the proposed SOPA/PIPA policies. With the provided insight they will have something to base their decisions on and, thus, be better informed to make policy decisions.

Also, the results will be interesting for record labels, movie studios or agencies like the MPAA. They will have better understanding of how much revenue they are losing to piracy and how much funds they should allocate for anti-piracy campaigns and other measures for tackling this issue.

B.

- What questions do you propose to study?

We could start off by asking if a respondent is aware of the recent developments in anti-piracy policies. If they have, we can ask additional questions about their opinion of these bills and maybe their suggestions for more effective measures. Then, we will ask if they have downloaded pirated media and if so, how often. To avoid ambiguity it might be necessary to give a brief definition of "pirated content". Next, we would like to learn more about which sites are the most popular for downloading/streaming content and what types of content are the most downloaded (respondents will be advised to skip these questions if they do not engage in piracy). At the end there will be some demographic questions, including one about the respondent's annual income.

C.

Research done on the topic (each group member finds one article):

George Volichenko: Copyright infringement and enforcement in the US

(<http://piracy.ssrc.org/wp-content/uploads/2011/11/AA-Research-Note-Infringement-and-Enforcement-November-2011.pdf>)

Written by Joe Karaganis of The American Assembly, Columbia University

This research paper includes a very similar study to the one we proposed. Joe Karaganis investigates how prominent piracy is in American culture and what the attitudes toward piracy and anti-piracy policies are. His findings suggest that almost a half of US adults have acquired pirated content at least once.

Cam Bogie: Lines Drawn on Antipiracy Bills

<http://www.nytimes.com/2011/12/15/technology/lines-are-drawn-on-legislation-against-internet-piracy.html?pagewanted=all>

Written by Edward Wyatt

This article discusses the issue of internet pirating and how it is very difficult for the United States to catch pirating as most of the sites are abroad. The US is trying to create solutions to piracy as it is viewed as a crime, but the solutions are creating even more problems in other places and even fails to correct some of the original concerns.

George Nardi:

<http://theithacan.org/20835>

“Content Blocked: Students Combat Anti-piracy Bills”. Elma Gonzalez.

This is an article from Ithaca College’s newspaper; it demonstrates the extent to which this issue matters to college students. This seems to strengthen our assumption that student populations are more concerned about the proposed legislature.

Chelsea Grindle:

http://www.riaa.com/physicalpiracy.php?content_selector=What-is-Online-Piracy

RIAA, 2012, Piracy Online, RIAA

The webpage is all about online piracy and what it is and why it hurts the music community. It also explains why paid pages are created.

Ben Gorman

https://www.cdt.org/files/pdfs/SOPA_House_letter_with_PROTECT_IP_letter_FINAL.pdf

This is a letter sponsored by the Center for Technology and Democracy that outlines some of the grave concerns about SOPA/PIPA held by some prominent law professors. The letter explores the disconnect between statements of elected/appointed officials and the principles supported by the bill and offers criticisms of its structure. The letter gathered 110 signatures of law professors around the country.

Jongwoo Lee

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

An article on why SOPA/PIPA has to be stopped and deeming the new act as unconstitutional, harmful to globalization/world wide web.

D.

We are planning to use the Carnegie Mellon C-Book as the sampling frame. We can use R to create a random sample from the set of page numbers of the C-Book. Then, all the student names on those pages will be included in our sample. We plan to sample from Carnegie Mellon undergraduate students.

E.

The target population for our survey is students 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. If we only sent out survey invitations via Facebook, we would have a sampling error, as not every student has a Facebook profile. To tackle that issue, we will also send out surveys via e-mail if a person in our sample does not have a Facebook profile available for us to “message”. We anticipate a large nonresponse error because a lot of people who receive a message asking them to take a survey decide not to. We intend to ask enough people to take the survey that we will get a proper number of responses to compensate for that effect. There is also some potential for measurement error, as someone may misread a question and then answer inappropriately without us knowing how they interpreted the question. We plan on making the language clear enough and providing clarifying definitions where appropriate so that our questions are easy to understand and measurement error is minimized.

F.

What is the mode of data collection? How do you plan to carry out the survey (e.g., by telephone, e-mail) and why?

- Self-administered web based survey administered via Facebook and e-mail.
- (Degree of Interviewer Involvement) No interviewer interaction
- (Degree of Interaction with the Respondent) Small Degree of interaction with respondent (initial email/facebook message)
- (Degree of Privacy) High Degree of Privacy; no interviewer or presence of other person will affect answers/behavior and no chance of overhearing answers. the survey is anonymous.
- (Channels of Communication) Visual text only survey (no audio)
- (Technology Use) The respondent interacts with the survey via the internet on their own computer and software (Programming error)
- Reduced cost, increased timeliness, improvements in measurements

G.

What variables do you propose to measure?

We want to measure the extent of Internet users’ piracy engagement (what proportion have done it at least once, regularity of piracy acts for those who have). Also, we want to measure the awareness of proposed government policies SOPA and PIPA (proportion of people who are aware of it and of those who feel positively about it). Finally, we want to know more about the respondents’ preferences in terms of the sources of pirated content (usage rates for popular websites that are notorious for piracy) as well as for each content type what proportion of those who engage in piracy download that content (music, movies, TV shows etc.)

J.

We plan to protect the privacy of the respondents by not knowing their names. The survey will be taken online and will be totally anonymous. There won’t be any way to link a specific respondent’s survey answers to the respondent themselves. If a respondent does not feel comfortable answering a question they are not forced to answer, although all answers are assuredly confidential. There will not be any disclosure more than the analysis of the answers as whole. No individual answers will be sought out specifically.

K.

Survey Scheme: We plan on using an SRS without replacement. We are going to generate random numbers corresponding to the number of pages in the C-Book. For every number we will go to that page and contact every Undergraduate on that page. If possible, we will contact those people via Facebook. If they do not have a Facebook, we will use the email address provided in the C-Book.

L.

Questions for the survey to find out about major issues:

SOPA/PIPA Awareness and Opinions

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

Yes or No

Then present summary information:

The SOPA Bill involves the following:

- Expand the ability of U.S. law enforcement to fight online trafficking in copyrighted intellectual property and counterfeit goods
- Expand existing criminal laws to include unauthorized streaming of copyrighted content, imposing a maximum penalty of five years in prison
- Provisions include the requesting of court orders to bar advertising networks and payment facilities from conducting business with infringing websites, and search engines from linking to the sites, and court orders requiring Internet service providers to block access to the sites

The PIPA Bill involves:

- 'Preventing Real Online Threats to Economic Creativity and Theft of Intellectual Property Act'
- Stated goal of giving the US government and copyright holders additional tools to curb access to "rogue websites dedicated to infringing or counterfeit goods", especially those registered outside the U.S.

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

'Strongly oppose', 'Oppose', 'Neutral', 'Support', 'Strongly support'

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

Yes, No, Maybe, No Opinion

Learning about their piracy behavior

4) Have you ever downloaded/streamed pirated content? ('Pirated content' being copyrighted material used/reproduced without the consent of the copyright holder. Examples: Download a movie from a torrent, stream a TV show from Megavideo)

Yes or No

If the answer is no, skip to question 7.

5) How often do you download pirated content?

Daily, Weekly, Biweekly, Monthly, Once a year, Few times a year, Only once

6) If the US adopted a tougher stance on internet-piracy enforcement, would your attitudes

about using pirated content change: not at all, a little, somewhat, or a lot?

7) Have you ever been banned from the CMU network or faced negative repercussions for downloading pirated content?

Yes or No

8) What type of content do you, or would you, download most frequently?

Music, Movies, TV Shows, Books, Other

9) Do you use streaming sites (e.g. sidereel.com)/torrent trackers (thepiratebay.org)/Peer-to-peer file-sharing clients (e.g. Dtella, LimeWire)/direct download sites (e.g. rapidshare.com)? (check all that apply)

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

Yes or No

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

- Acceptable
- Illegal
- Unethical
- Both illegal and unethical
- I don't have an opinion/have never thought about it

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

- Acceptable
- Illegal
- Unethical
- Both illegal and unethical
- I don't have an opinion/have never thought about it

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

- Acceptable
- Illegal
- Unethical
- Both illegal and unethical
- I don't have an opinion/have never thought about it

Potential Scenario

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

14) 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
- Both illegal and unethical
- Neither illegal nor unethical
- You do not have an opinion on the ethics or legality of this action

15) 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
- Both illegal and unethical
- Neither illegal nor unethical
- You do not have an opinion on the ethics or legality of this action

16) After watching the movie on YouTube, you rip 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
- Both illegal and unethical
- Neither illegal nor unethical
- You do not have an opinion on the ethics or legality of this action

17) 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 or No

18) Do you have any other comments or views on internet piracy that this survey has not yet addressed? Please express them here.

Demographic Questions

19) Gender: Male or Female?

20) What ethnicity are you? (Please select the closest option)

White/Caucasian, African American, Hispanic/Latino, Asian/Pacific Islander, Native American Indian, Other

21) What class year are you?

Freshman, Sophomore, Junior, Senior, Super Senior

22) Do you live on-campus or off-campus?

On or Off

23) What is your political affiliation?

Republican, Democrat, Independent, Other, None, Do not wish to share information.

24) 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

25) 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 not please write down where you consider yourself from.

Northeast, Midwest, West, South, not from the US/other: _____

M.

To estimate the gravity of the piracy issue on campus we are going to ask our respondents an important question:

4) Have you ever downloaded/streamed pirated content?

Start by trying a very small margin of error for the proportion p like 0.01:

Then $n_0 = (z(0.0025) * \sqrt{p(1-p)}) / 0.05)^2 = (1.96 * \sqrt{0.5 * 0.5}) / 0.01)^2 = 9,604$

Above I picked the “most conservative” value for p , the one that results in the highest SD.

if N = number of undergrads at CMU = 6,178

$n = N * n_0 / (N + n_0) = (9604 * 6178) / (9604 + 6178) = 3,760$ so we would need an SRS without replacement of at least 3,760.

This is not a feasible sample size for our purposes.

If we set our desired margin of error a little higher, at 0.05 (95% CI):

Then $n_0 = (z(0.0025) * \sqrt{p(1-p)}) / 0.05)^2 = (1.96 * \sqrt{0.5 * 0.5}) / 0.05)^2 = 384.2$

$n = N * n_0 / (N + n_0) = (368.8 * 6178) / (368.8 + 6178) = 361.7$ so we would need an SRS without replacement of at least 362.

$n = 362$ seems much more reasonable and we will aim for this sample size.