Chilling Effects Response Abuse Final Project Proposal

Eric Engoron
eric.engoron@gmail.com

Jiaqi Chen
jiaqichen@cs.stonybrook.edu

Yangyang Zhu
yangyang.zhu@stonybrook.edu

ABSTRACT
In this paper, we propose a project for CSE 534 entitled The Abuse of The Response Mechanism on Chillingeffects.org. This project will aim to discover this abuse caused by human input.

Keywords
Chilling effects, Abuse, Sender, Receiver, Principle, Naïve Bayes, Machine Learning, Fake, Detection

1. PROJECT BACKGROUND

In essence, the DMCA, or Digital Millennium Copyright Act, is a United States copyright law that criminalizes infringement of digital rights management and circumvention of access control. In relation to the DMCA, chillingeffects.org is a 3rd party research website, sponsored by numerous educational institutions and companies to collect and analyze data relating to complaints about online activity, especially related to content management and removal. As part of its functionality, chillingeffects.org also allows users to submit their own copyright infringement complaints against others. Currently chillingeffects.org contains over 2.3 million records, of which approximately 2.1 million relate to DMCA (Chilling Effects).

2. OUR PROJECT GOALS

In addition to gaining general statistics as to the submitted reports such as locale, most active submitter, etc., our project will mainly focus on identifying abuse of the reporting mechanism on chillingeffects.org. In order to do so, we will define abuse in the following ways:

1. Misusing the form when submitting a report and/or submitting an incomplete form (e.g. not providing infringing URLs after providing original ones)
2. Using the reporting mechanism to personally attack another individual or company without justification
3. One user submitting multiple complaints against the same website and/or many users submitting complaints of the same website
4. Using the reporting mechanism to post links of where to download infringing materials.
5. Submitting a report that is outside of the jurisdiction of DMCA (e.g. submitting a report to delete information from your own account that you lost the password to, without going through the company that hosts the account)
6. Submitting entirely false reports

We will analyze statistics and messages related to the abuse of the reporting mechanism such as: which senders are the most active against specific websites, how many forms are incomplete and yet are still hosted on the website, which category of digital media contains the most incomplete or abusive reports against it, how many reports seem to be fake or ill-founded, how many reports cannot be acted upon by DMCA, how many reports are duplicates, which receivers receive the most reports, and so on. In addition, we will also look for a trend in an influx of complaints between senders and receivers to identify attacks on a specific website over time. For abusive entries, if the sender is not null, we assume that the entry is a personal attack on the recipient and thus their trend is a negative effect. We will also investigate if the principal company and the sender are the same. If they are not the same, we assume that the sender is working on behalf of the principal and we will gather statistics relating to relationships between senders and principals overtime. We will gather these statistics by first creating a "spam filter" of sorts to identify incomplete reports. We will then create a tagging mechanism that will create a tag for untagged reports in order to see which area of digital content has the most abuses of the system. We will also aim to identify fake posts using commonalities in already identified fake posts and a machine learning algorithm. We will compare sender and receiver in each complaint to identify if one sender is attacking a receiver using the reporting mechanism. Other statistics, as mentioned above, will also be gathered using filtering techniques.

In addition to gathering statistics on abuse of the reporting mechanism, it will be one of our goals to use these statistics to design an automated system that can identify abusive postings based on commonalities like sender, incomplete forms, and misinformation on forms.

3. PREVIOUS WORK

While Chillingeffects.org has done work to classify its reports by locale, type of complaint, receiver of complaint etc, little has been done to identify abuse of the system. Numerous media outlets have complained about chillingeffects.org being used as a breeding ground for

finding illegal materials on the Internet. Furthermore, those that were victimized by fake DMCA complaints have taken to the Internet to discuss their situation, yet said fake complaints are still listed on chillingeffects.org.

Chillingeffects.org lacks the functionality to filter out incomplete reports, reports targeting the same website, reports containing items like unavailable URLs, and specific users that have been abusing the system.

4. MOTIVATION

While the reporting mechanism of chillingeffects.org was created in order to help put a stop to copyright infringement, in reality, due to abuses of the system, its true functionality has been morphed into a system used for personal attacks, misinformation, and illegal activity. We are motivated to do this project because we would like to help chillingeffects.org remove its abusive postings and become a truly valuable tool in the fight against copyright infringement. We would like to gain a greater understanding of the word of copyright infringement and discover which areas of digital content are the most infringed.

5. TENTATIVE SCHEDULE OF WORK

Week	Task
2/9 – 2/16	Coming up with project ideas and writing the proposal
2/16 - 2/23	Identifying factors to be used in identifying incomplete forms and coding said filter
2/23 - 3/2	Identifying factors to be used to tag untagged entries and tagging said entries.
3/2 – 3/9	Parsing sender and receiver information for various complaints to determine the most active senders with entries that abuse the system. In addition we will analyze the relationship between various senders, receivers, and principal companies.
3/9 – 3/16	Identifying factors in known fake posts that can be used to identify other fake posts.
3/16 – 3/23	Coding a machine learning algorithm to attempt to identify fake posts
3/23 – 3/30	Creating a filter that will look for posts with the purpose of spreading illegal content
3/30 – 4/6	Use filters above to create a statistical report of everything mentioned thus far
4/6 – 4/13	Create a filter for posts outside of DMCA jurisdiction.
4/13 – 4/20	Creating an automated way to detect abusive postings before they are posted.
4/20 – 4/27	Creating an automated way to detect abusive postings before they are posted.
4/27 – 5/4	Analyze chronologically, the influx of abusive postings to the chilling effects website.
5/4 – 5/10	Final write-up and presentation

6. REFERENCES

[1] DCMA: http://www.copyright.gov/legislation/dmca.pdf

[2] Chilling Effects: https://www.chillingeffects.org/notices/search?utf8=%E2%9C%93&term=&sort_by=

[3] Naïve Bayes Learning: http://www4.cs.umanitoba.ca/~jacky/Teaching/Courses/COMP_4360-MachineLearning/Additional_Information/elkan-naive-bayesian-learning.pdf

[4] Article on Chillingeffects being a breeding ground for illegal activity: https://www.techdirt.com/articles/20140317/11355726599/copyright-alliance-attacks-chilling-effects-clearinghouse-argues-dmca-system-with-no-public-accountability.shtml