

Data Analysis of Chilling Effects Reports: Project Proposal

Neeraj Devadath
SBU ID: 109773735
Stony Brook University
Stony Brook, New York
ndeavadath@cs.stonybrook.edu

Nikhil Ambekar
SBU ID: 109819116
Stony Brook University
Stony Brook, New York
nambekar@cs.stonybrook.edu

Swijal Patil
SBU ID: 109887922
Stony Brook University
Stony Brook, New York
swapatil@cs.stonybrook.edu

ABSTRACT

Chilling Effects [1] is governed by the Electronic Frontier Foundation with the intent of creating a collaborative archive of recipients of cease-and-desist notices. Chilling Effects serves as a central authority that issues notices, summons, subpoena to the violating agencies and ensures that the legal rights and responsibilities of the citizens are protected under the Digital Millennium Copyright Act [DMCA]. The DMCA protects service providers from monetary liability based on the allegedly infringing activities of third parties. Analyzing these complaints shall provide a deep insight into the different aspects of complaints and any existing trends in them.

KEYWORDS

Chilling Effects, Digital Millennium Copyright Act [DMCA], Transparency reports, Copyright, Abuse, Clustering, Classification

1. PROBLEM STATEMENT

The scope includes analyzing different trends in the complaints and notices summons and subpoenas issued by Chilling Effects against the perpetrator. The scope shall also include identification of patterns that represent the geospatial and temporal or causal relationship. The scope encompasses visualizing and inferring the results of the analyzed data to present suitable interpretations for the patterns analyzed.

2. MOTIVATION

Since abiding by the laws set by a central authority is imperative, the Chilling Effects set up by the Electronic Frontier Foundation vouches for the strict implementation of the DMCA. The DMCA reports are considered to be as ground truth for the Analysis and shall be considered as the ultimate authority in issuing any copyright infringements, patent violations and other types of encroachments. There are several rhetorical questions that have posed impending questions.

2.1 Robustness of Reporting Mechanism

There have been instance where the liberty of reporting has been ill-used and has resulted in the abuse of the reporting mechanism. There can be certain complaints which are posted just for the sake of posting, and these complaints do not hold good since, there they are nullified and rejected due to their dubious nature by the authorities. It would be interesting to detect any patterns in these complaints. Categories report complaints with varying frequencies. What is the intensity of the impact of the jurisdiction on the nature of complaints? Do people approach attorneys to file complaints? Are these complaints more valid in nature?

2.2 Geospatial Analysis

It is plausible that certain regions of the world report dubious complaints more frequently than others. Is there any demographic data that can be presented that defines the average strata of such complaints? There can be some regions extensively abused, due to reasons not restricted to legal loopholes in the law of the land, governmental policies as well as awareness regarding intellectual property rights. We propose to categorize the complaints demarcated by their geographical barriers to locally analyze them as per the legal laws of that particular region. There exist certain regions of the world that are more vulnerable to these complaints. An analysis of the location versus intensity of complaint can determine the same.

2.3 Complaint Classification

As there are several organizations that issue their Transparency reports [2][3], a trend analysis shall reveal different types of complaints that occur in these different domains. Some popular categories [5] are Internet and Telecom, File Sharing, Music Audio Video Sharing etc. Based on these categories we would be able to identify the trouble areas. We would be checking if any particular kind of complaint repeatedly occurs, determining relative importance of each category.

2.3.1 Types of Complaints

In the Transparency reports issued by most of the big names in the business, there have been certain areas that have stood out. Generally these reports are categorized as National Security Demands, Emergency Demands, Location Demands, International Demands, and other

Industry-Specific Demands. These categories derive their distinctness from the parent industry with which they are associated.

An example quoted from the AT&T Report [4] states that they have introduced a specific category pertaining to Device Requests that caters to Clients raising concerns regarding devices. Such a specific categorization would enable us to include an accurate trend analysis.

2.4 Epicenter of Attack

Several times, there can be business rivalries that translate into an attack on the company's copyrighted products. It is necessary that certain patterns be derived from these reports so that it can be concluded if a particular company, state, region is under attack from a specific user group.

2.5 Temporal variation with respect to complaints

Several instances are reported where groundbreaking news that causes people to make inflammatory comments and post material online, that results in potential candidates that shall raise complaints in the near future.

It is possible that there is a certain period of the year where complaints tend to rise and a certain period that is observed as a slack period. Is it possible that the Financial Year of a country shall determine the summons and the copyrights infringement complaints it issues? The veracity of these claims can be ascertained to be genuine or propaganda intended to hide certain financial losses or gains from the tax authorities. It is possible that this mechanism can be used to tweak the system for the benefit of the individual or claimant abusing the reporting mechanism.

A timeline analysis of the data will help us provide valuable insights about the variations in the types of complaints or which types gain prominence with the time.

3. DATASET

We have considered the Transparency Report Shared by Google [2] as the primary data set. Chilling Effects also have their own API Key that extracts data, based on the Query Processing based on the user requests. We have mooted the process of requesting an API Key and have been assured by Chilling Effects that they would be providing us one shortly. We would be eventually considering the Chilling Effects data set once the technical access issues have been resolved. It is essential that certain prevalent dominant features are assigned importance as against the dormant features that tend to occur sparsely in the dataset. The features that gain prime significance shall be strong indicators of existing trends and would represent the dataset properly.

4. APPROACH

The existing data set shall be imported into a central repository like Apache Solr [6] that shall be used to extract relevant information for identifying and analyzing patterns. The data presented shall act as an input feeder to the visualization phase, where the analyzed data would be represented suitably with statistical trend lines, graphs and reports. Depending upon the dataset under consideration, appropriate classification, clustering shall be applied. Based on the nature of the dataset the techniques chosen shall vary. There shall exist certain data points that shall be less cohesive than the other data points. These points can be termed as 'outliers' or 'anomalies'. The features that tend to occur infrequently generally can be classified as certain special cases known as 'anomalies' or 'outliers'. Clustering of requests that seem to belong to the same category shall reveal a refined analysis of formation of certain focus groups or communities that present certain interesting relationships amongst the features that are analyzed.

5. TIMELINE

Date	Milestone
16 February	Project Proposal
28 February	Data Collection
15 March	Data smoothing, Data analysis and feature extraction
31 March	Intermediate result examination and midterm report
15 April	Further analysis and consolidation of data analysis
4 May	Visualization and reports generation

6. REFERENCES

- [1] Chilling Effects [<https://www.chillingeffects.org/>]
- [2] Google Transparency Report [<http://www.google.com/transparencyreport/>]
- [3] Facebook Transparency Report [<https://govtrequests.facebook.com/>]
- [4] AT&T Transparency Report [<http://about.att.com/content/csr/home/frequently-requested-info/governance/transparencyreport.html>]
- [5] Similar Web [<http://www.similarweb.com/category>]
- [6] Apache Solr – An Open Source search platform [<http://lucene.apache.org/solr/>]