Analyzing Tor abuse complaints

Rachee Singh
University of Massachusetts Amherst

Sadia Afroz
UC Berkeley, ICSI

1. INTRODUCTION.

We want to understand the characteristics of the network abuse originating from Tor. Understanding the abuse of Tor is important because website operators and content providers discriminate against Tor users [3], in many cases, because of the abuse originating from Tor. Akamai observed that an HTTP request from a Tor IP is 30 times more likely to be a malicious attack than one that comes from a non-Tor IP [1]. Cloudflare found that 94% of requests from the Tor network are malicious [2]. Abuse control is crucial for the Tor network itself to maximize the benefit of anonymous communication networks. To understand the nature of abuse websites receive through Tor, we look at the complaints exit operators received while running a Tor exit. We assume that every time a website receives abuse through Tor the administrator of the website complains to the exit operator. So these complaints can be a proxy for understanding the type and frequency of abuse happen through a Tor exit. In this document, we study a partial set of complaints sent to `abuse@torservers.net`. Torservers.net runs several high bandwidth Tor exits. We only have access to the complaints sent to one of the exit operators who runs a dozen of Tor exits.

Our preliminary analysis of the complaint emails shows:

1. Over 99% (around 3 million) of the complaints are DMCA complaints. `noreply@p2p.copyright-notice.com` sent over 1 million of the complaints (Figure 2).
2. Over 99% of the DMCA complaints mention the use of BitTorrent and a few mention eDonkey (Figure 1).
3. The DMCA complaints mention 1200 unique IPs. Only 323 of these IP addresses are related to the Tor network (according to the consensuses published since 2010). WHOIS IP lookup shows that 74.3% of the 323 IP addresses listed contact is `abuse@privateinternetaccess.com`.
4. The majority of the non-DMCA complaints are about brute force login attacks on WordPress (Figure 3).
5. Email spam complaints almost stopped after August 2014 (Figure 5).

Note that compared to the total traffic of the exit relays, the number of abuse complaints is negligible. The next step of the analysis is to correlate the number of complaints with the exit relay bandwidth and policy.

2. THE COMPLAINT EMAILS.

The dataset consists of a partial set of complaints sent to `abuse@torservers.net` from June 2010 to April 2016. There are approximately 3 million complaint emails in this dataset. Of these, 99% (2,971,227) of the complaints are DMCA related complaints. The remaining (approx) 13,000 emails have other complaints and responses from the side of Tor abuse servers. The corpus has 6,971 non-DMCA complaint emails.

Cluster	Topic	Top words
0	Mixed	graphicsoneinc, junkemailfilter, asthma, pascal, bob, die, suffers, helo, population, sun
1	Mixed	password, failed, sshd, whois, timezone, denmark, investigate, extracted, invalid, apologies
2	Googlegroups	groups, rubin, googlegroups, broadcast, jimenez, jew, steve, posting, satanic, conspiracy
3	Copyright	paramount, copyright, irdeto, cert, infringement, notice, pgp, compliance, voxility, material
4	Sent by nforce	mnt, logs, furanet, htdocs, login, wp, sites, query, zwiebelfreunde, nforce
5	Sent by Spamcop	ip, received, content, spamcop, abuse, mail, thank, message, manager, mso
6	Sent by ValueHost	valuehost, administrator, index, attempts, mail, noc, abuse, yor, disturb, shared
7	Sent by Webiron	abuse, issues, webiron, mail, clients, service, ip, ticket, time, blacklist

Table 1: Result of KMeans clustering

2.1 Automatically identify abuse.

We extract the nature of abuse, the time of complaint and the exit IP being complained about from each mail of the corpus. This is straightforward in emails that are DMCA violations since these emails have a format that can be easily parsed via regular expressions. To extract the relevant abuse information from non-DMCA complaint emails, we followed three steps: apply clustering, search for regular expressions, and search for relevant terms.

Apply clustering.

We consider each email as a document and then vectorise each document based on Term Frequency-Inverse Document Frequency (TF-IDF) values. We perform KMeans clustering on this set of vectorised documents to obtain 8 clusters (Table 1).

Based on the TF-IDF terms, cluster 2 contains googlegroups complaints, cluster 3 contains copyright infringement complaints, cluster 4 contains complaints filed by nforce, cluster 5 contains complaints filed by SpamCop, cluster 6 contains ValueHost complaints, cluster 7 contains Webiron complaints. Clusters 0 and 1 seem to be a mixed bag of complaints. Thankfully none of the profanity showed up in top 10 terms.

While we can make guesses about the type of abuse being complained about in the emails based on the cluster it belongs to, clustering of not perfect. In addition to this, clusters 0 and 1 seem to be mixed with emails of different types. In order to deal with the case of mis-clustered emails and emails in undefined clusters (cluster 0 and 1), we further process the corpus for extracting the type of abuse.

Search for regular expressions.

For all emails sent within a cluster, we enumerate regular expressions that capture the format of the email. For instance, if an email in the copyright complaint cluster is sent by Voxility, we use a regular expression to capture the type of abuse. The regular expressions are produced after manually observing the formats of emails received. Some senders have an email format while some sends free form text. To handle the free-form text, we use the last step. For emails in undefined clusters, we look for server logs (HTTP GET/POST request logs, SSH login fail logs etc) to infer the kind of abuse.

Search for relevant terms.

If the previous steps fail to find the kind of abuse, we look for terms like port scanning and email hacked to infer the kind of abuse that the email could be complaining about. Inferences made using this step are likely to be erroneous.

3. RESULTS.

3.1 Analysing DMCA complaints

We first look at the distribution of the DMCA complaints over time (Figure 2). The 3 million DMCA complaints refer to 1200 unique exit relay IPs. In addition, the complaints highlight the use of BitTorrent or eDonkey for accessing content illegally (Figure 1). Interestingly, most of these 1 million complaints are originated from the same sender noreply@p2p.copyright-notice.com (Figure 2).

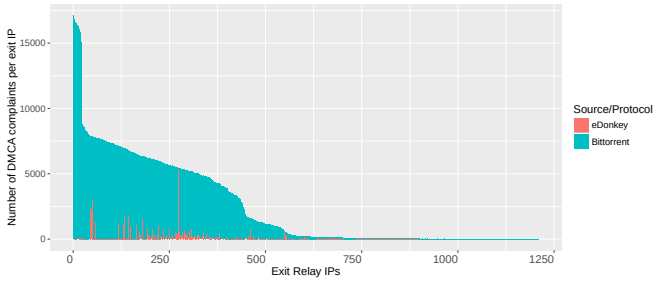

Figure 1: Number of DMCA complaints per exit relay IP and the protocol used for abuse by each.

3.2 Analysing non-DMCA complaints

These complaints include complaints from individuals and companies about Spam, intrusions an various other kinds of attacks. Based on the 3-step approach, we identified the type of abuse. From the around 7000 original emails, we were not able to classify approximately 800 emails. For these, the type of abuse is 'other' in our analysis. We classified these manually. The largest number of complaints are about brute force login attacks on WordPress followed by the general category of malicious access of servers (includes complaints of intrusion, and brute force HTTP POSTs) (Figure 3).

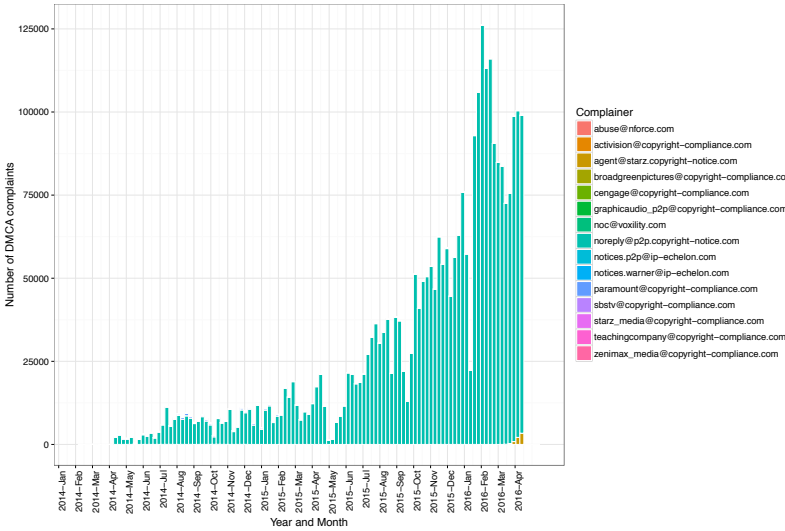

Figure 2: Number of complaints by each complainer, per week (for DMCA complaints).

We also look at the distribution of various types of non-copyright abuse over time (Figure 4) (all WordPress plugin exploits combined to one category). We noticed a spike in complaints of brute force WordPress logins at the end of 2015. Another interesting point is that email spam complaints stopped after August 2014 (Figure 5).

Acknowledgment

We thank Moritz Bartl for sharing the abuse complaints with us.

4. REFERENCES

[1] akamai's [state of the internet] / security, Q2 2015 report.

[2] The Trouble with Tor.

[3] S. Khattak, D. Fifield, S. Afroz, M. Javed, S. Sundaresan, V. Paxson, S. J. Murdoch, and D. McCoy. Do you see what i see?: Differential treatment of anonymous users. In *Network and Distributed System Security Symposium 2016*. IETF, 2016.

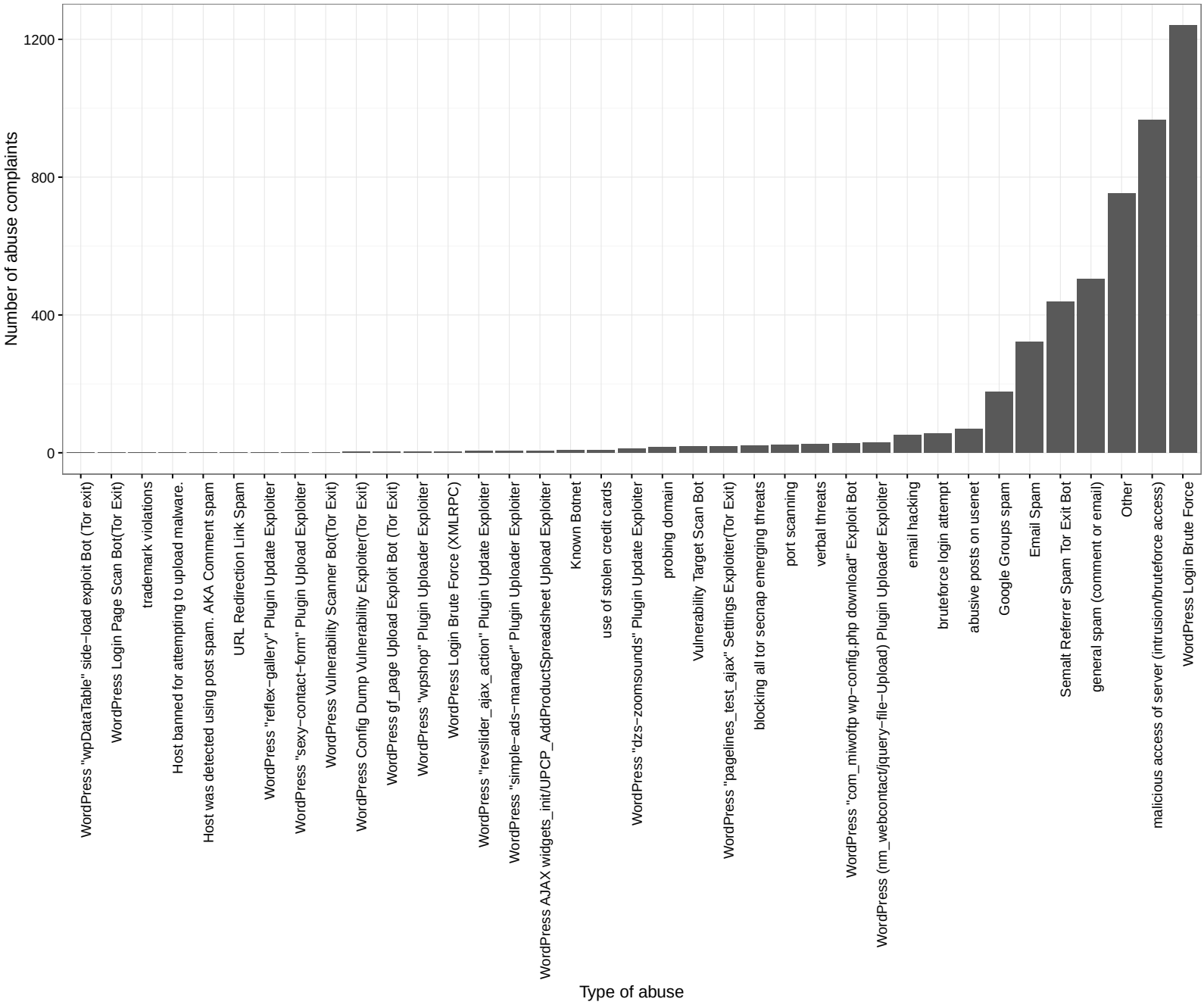

Figure 3: Types of non-copyright abuse complaints and their popularity

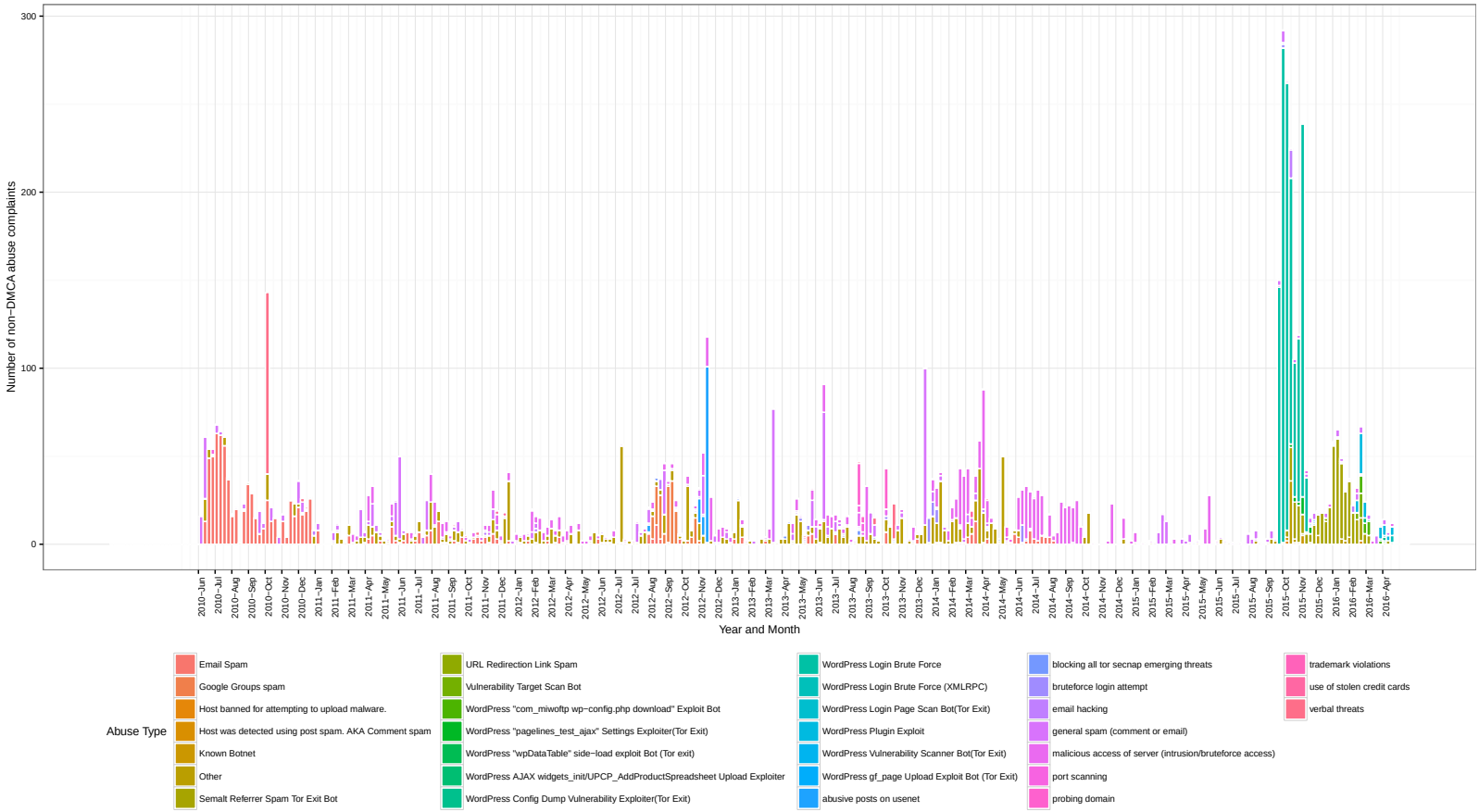

Figure 4: Distribution of different types of non-copyright complaints over time.

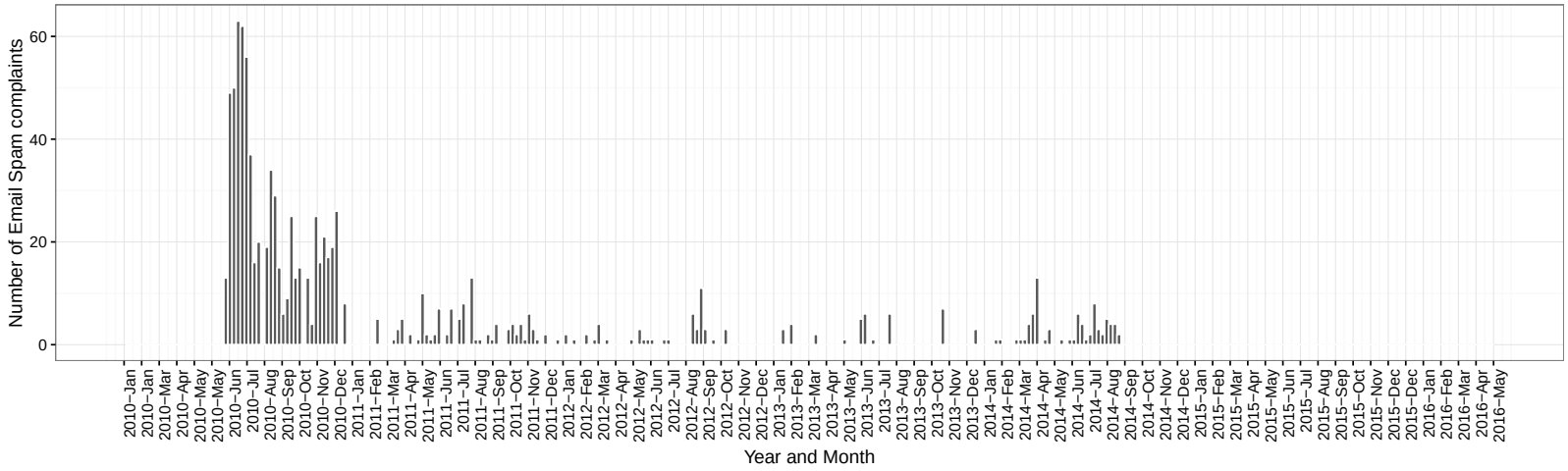

Figure 5: The end of email spam complaints from Tor