Characterizing the Nature and Dynamics of Tor Exit Blocking

Rachee Singh[1], Rishab Nithyanand[2], Sadia Afroz[3,4]
Paul Pearce[3], Michael Carl Tschantz[4], Phillipa Gill[1], Vern Paxson[3,4]

[1]*University of Massachusetts – Amherst,* [2]*Stony Brook University,*
[3]*University of California – Berkeley,* [4]*International Computer Science Institute*

Abstract

Facing abusive traffic from the Tor anonymity network, online service providers discriminate against Tor users. In this study, we characterize not only the extent of such discrimination but also the nature of the undesired traffic originating from the Tor network—a task complicated by Tor's need to maintain user anonymity. We address this challenge by leveraging multiple independent data sources: email complaints sent to exit operators, commercial IP blacklists, webpage crawls via Tor, and privacy-sensitive measurements of our own Tor exit nodes. As part of our study, we also develop methods for classifying email complaints and an interactive crawler to find subtle forms of discrimination, and deploy our own exits in various configurations to understand which are prone to discrimination. We find that conservative exit policies are ineffective in preventing the blacklisting of exit relays. However, a majority of the attacks originating from Tor generate high traffic volume, suggesting the possibility of detection and prevention without violating Tor users' privacy.

1 Introduction

Anonymity systems like Tor provide a useful service to users who wish to access the Internet without revealing their intended destination to any local monitoring, or their network-layer identity to the final destination. However, as Tor has increased in scale and usage, tensions have emerged between Tor users and online service providers. Specifically, service providers claim that the anonymity provided by Tor is often used maliciously for spamming, vulnerability scanning, scraping, and other undesired behavior (e.g., [1]). As a result, Tor users now face differential treatment (e.g., needing to solve CAPTCHAs before receiving services) and even outright blocking [2].

At its core, the problem is that in return for anonymity, each Tor user shares their reputation with other users. As a result, the malicious actions of a single Tor user can lead IP abuse blacklists to include IP addresses used by Tor exit relays. Consequently, websites and content providers treat even benign Tor users as malicious. In this paper, we characterize aspects of the conflict between users desiring anonymity and websites aiming to protect themselves against malicious Tor traffic. We investigate the nature of traffic that exits the Tor network and is undesired by online service providers. We also actively measure various forms of discrimination performed against Tor users.

Challenges. We grapple with two key challenges: First, measuring Tor traffic is antithetical to the goals of the anonymity system and poses ethical challenges. Second, defining and identifying *undesired* or *abusive* network traffic is hard as opinions vary and encryption can render inspection of traffic infeasible. We address both challenges by focusing on the receivers' reactions to Tor traffic rather than the traffic itself. We consider email complaints sent to Tor relay operators (§4) and blacklisting of Tor-related IP addresses (§5), and take measurements of server responses to Tor traffic, both synthetic (§6) and user-driven (§7). These datasets not only allow us to observe the effects of undesired traffic without measuring it, but also provide an operational definition of *undesired traffic*: the traffic that leads to complaints, blacklisting, or rejecting of Tor users. This operationalization allows us to sidestep debates over what constitutes abuse and to focus on the subset of undesired Tor traffic that has affected operators and users.

Additionally, collecting and analyzing each of these four datasets presented technical challenges. Analyzing 3 million email complaints received by Tor relay operators since June 2010 required us to construct automated processing methods (§4). Understanding the inclusion of Tor-related IP addresses in IP blacklists required us to develop methods for teasing apart *reactive blacklisting*—i.e., blacklisting triggered by abuse—from *proactive blacklisting*—i.e., blacklisting due to Tor's pre-existing

reputation (§5). Measuring the prevalence of discrimination faced by users required exercising multiple aspects of websites and inspecting them for subtle forms of discrimination (e.g., CAPTCHAs and interaction-based discrimination) in addition to outright blocking. To address this issue and accurately measure discrimination against users, we go beyond the prior work of Khattak et al. and develop a crawler capable of exercising the search and login features of websites. Taking measurements of real Tor traffic required the creation and deployment of a privacy-sensitive logging approach for our own Tor exit relays. We also consider aspects of Tor exit relays that make them more susceptible to complaints, IP blacklisting, or blocking. We augment this analysis by deploying several Tor exits with varied configurations and monitoring the reactions they produced.

Key Findings. One major takeaway from our analysis is that many of the attacks originating from Tor generate high traffic volume (e.g., DDoS attacks, port scanning), raising the possibility of blocking them using privacy-sensitive techniques (§8). We believe developing, implementing, and deploying such techniques may provide online service operators a more effective means of curbing abuse than IP blacklisting while also preventing lost utility to Tor from blocking.

Our analysis of email complaints shows that, historically, the most vocal complainants about Tor traffic were a small number of copyright enforcement firms. This is no longer be the case, possibly due to Tor blocking BitTorrent's standard ports by default (Table 2 in §4). The most common non-copyright complaints were about network abuse and attempts to gain unauthorized access (Table 3 in §4).

From our analysis of commercial IP blacklists, we find that 7% of the commercial IP blacklists we analyze engage in *proactive* blocking of Tor users—i.e., blacklisting Tor exit relays soon after they are listed in the consensus. This is indicative of blacklists performing discrimination against Tor exit relays as a matter of policy, rather than based on undesired traffic (§5). Currently, 88% of Tor relays are blacklisted on one or more of the blacklists, compared to 9% and 69% of the endpoints of the VPNGate and HMA VPN services, respectively (Figure 4 in §5). We also find that conservative exit policies do not reduce Tor exit relays' susceptibility to getting blacklisted, which appears to reflect that such policies still allow for Web access, the channel most extensively used for abuse.

Finally, we find discrimination to be a pressing concern. Our synthetic experiments show that discrimination occurs on 20% of all Alexa Top 500 website front-page loads through a subset of Tor exits. Focusing on the search and login functionalities of the Alexa Top 500 websites, we see a 3.9% and 7.5% increase in observed discrimination (compared to front-page load discrimination), respectively (Table 6 in §6). We also find that real Tor users experience high fractions of failed HTTP requests (15.8–33.4%) and HTTPS handshakes (35.0–49.6%) while browsing the Alexa Top 1M websites using our deployed relays (Table 8 in §7).

2 Background and Related Work

Tensions between Tor and online services. Tor is a low-latency onion routing network with over 2M daily users and over 7K supporting servers [3]. While proponents of Internet freedom laud the anonymity provided by Tor, it can also provide a cloak for malicious network activities. Indeed, CloudFlare reported that 94% of the requests from the Tor network are "malicious", consisting of comment spam, scanning, and content scraping [1]. According to a report published by Distill networks, 48% of Tor requests are malicious, higher than non-Tor requests (38%) [4]. A study of the Sqreen application protection service found that connections through Tor are responsible for ≈30% of all attacks on their customers, including password brute force attacks, account enumerations, and fraudsters [5]. As per Akamai's State of the Internet report, an HTTP request from a Tor IP address is 30 times more likely to be a malicious attack than one from a non-Tor IP address [6]. Imperva-Incapsula found that in a period of 2.5 weeks, 48.53% of the attack requests came from Tor [7]. However, the majority of these attack sessions were originated from well-known DDoS bots and bad clients, which can be identified using approaches other than IP reputation. Not counting the attacks from well-known attackers, the fraction of attack sessions originating from Tor went down to 6.78%, which is comparable to the attacks coming from the rest of the Internet population in Ireland (5.45%).

Different services have reported similar types of attacks from Tor. The three most common attacks from Tor to Akamai's services were automated scanning (path scanning and vulnerability scanning), SQL injection, and cross-site scripting attacks [6]. IBM reports that SQL injection, automated scanning, and DDoS are the most common attacks from Tor [8]. Sqreen found authentication attacks (brute force attack on a specific user account, or accounts enumeration), path scanning, and SQL/NoSQL injections [9] are likely to originate from Tor [5]. Our analysis of the abuse complaints to a number of Tor exit relays reflects similar proportions of attack traffic (Section 4).

Despite reports claiming a higher likelihood of malicious traffic from Tor, there have been debates about the correctness of their inference methods. For instance, Perry, writing for the Tor Project's blog, ques-

tions whether CloudFlare's methods considered as malicious all traffic from an exit relay that ever sent any malicious traffic [10].

While websites might be tempted to blacklist all Tor IPs in a proactive attempt at security, doing so could cause a loss in revenue. Akamai's report highlights that Tor users are just as likely to make purchases from revenue generating websites as non-Tor users [6].

Blocking and Filtering of Tor. Many government censors around the world block access to Tor [11], the subject of numerous measurement studies [12–16]. However, such government censorship blocks access *to* the Tor entry nodes, which is different from server-side Tor blocking, which blocks access *from* the Tor exit nodes.

Khattak et al. is the only systematic measurement study of server-side Tor blocking [2]. They showed that in 2015 at least 1.3 million IP addresses blocked Tor at the TCP/IP layer, and 3.6% of the Alexa Top 1,000 websites blocked Tor at the HTTP layer. At the TCP/IP layer, the hosting services GoDaddy and Dreamhost are among the top five Tor blockers. CloudFlare blocks access at the HTTP layer. Our work extends the work of Khattak et al. by additionally measuring the blocking of login and search functionality. We find a higher rate of blocking (20.03%) than Khattak et al. (3.6%). We demonstrate that Khattak et al.'s headless crawler underestimates the blocking rate (Figure 12).

To understand the impact of blocking on Tor users, we measure the number of failed requests to Alexa Top 1M web pages at the exit level using privacy-sensitive logging on our exits.

3 Our Deployed Exits

To aid our studies of complaint emails, IP blacklisting, and discrimination, we deployed and used data from ten of our own exits in addition to current and historical records about pre-existing Tor exits.

	Max. BW	Exit Policy	Num.
Large-Default	61 MBps*	Default	2
Medium-Default	10 MBps	Default	2
Medium-RR	10 MBps	RR	2
Small-Default	2 MBps	Default	2
Small-RR	2 MBps	RR	2

Table 1: Configurations of our deployed exit relays. *The large exits' policy allows for unlimited bandwidth usage. We provide the maximum bandwidth achieved during the study period.

We vary the bandwidth and exit policy of our exits in order to understand the impact of relay characteristics on email complaints, blacklisting, and discrimination. We used bandwidth allocations for the relays of 2 MBps (small exits), 10 MBps (medium exits), and unlimited (huge exits). In total, our deployed relays handled over 3% of all Tor exit traffic during their deployment. The exit policies were varied to either be the Tor default policy or the "Reduced-Reduced" policy. The default policy [17] allows all ports except those misused for email and news spam (25, 119), network attacks (135–139, 445, 563), or peer-to-peer file sharing (1214, 4661–4666, 6346–6348, 6699, 6881–6999, plus the adjacent ports 6349–6429). The Reduced-Reduced (RR) exit policy, designed to avoid blacklisting, additionally blocks ports associated with SSH, Telnet, IRC(S), and other protocols [18]. We summarize our relay configurations in Table 1.

Analyzing the usage statistics of ports on our exit relays, we see that web-traffic accounts for 98.88% of all connections made through the RR policy exits. In contrast, traffic though the default policy exits has higher application/port diversity, with only 31.36% of observed traffic being HTTP(S). We measure this using our privacy-sensitive logging described in Section 7.

4 Email Complaints about Abuse

In this section, we look at the abuse complaints received by exit operators. We use these complaints as a proxy for understanding the type and frequency of undesired incidents happening through Tor exit relays.

4.1 The Email Corpus

In addition to our own exits, we obtained access to abuse complaints emailed to four exit relay operators, (Table 2). The largest email corpus, consisting of ≈3M emails, came from a subset of exits operated by Torservers.net (`https://torservers.net/`). Using whois queries on exit IP addresses and counting the number of exits that use Torservers.net as their abuse contact, we estimate that they run 10 to 20 exits, with the uncertainty coming from fuzzy matches.[1] According to the latest Tor consensus, Torservers is one of the largest exit operators in terms of overall bandwidth capacity. The apx exit family includes three exits: apx1 [19], apx2 [20] and apx3 [21]. The other two exits are TorLand1 [22] and jahjah [23]. TorLand1 was one of the oldest Tor exits, running since 2011 until February 2017.

Our complaints dataset lacks any complaints sent by fax or mail, or those sent to only the abuse contact of the associated autonomous system. Also, some email complaints might have been lost or deleted. For example, the

[1] The current operators of Torservers were unable to answer the exact number of exits they ran over time.

Exit Family	# Exits	% Tor Traffic	Email Dates	# Complaints	Top Complaint
Torservers.net	10–20	7.05%	2010/06–2016/04	2,987,017	DMCA Violation (99.74%)
apx	3	1.94%	2014/11–2016/05	293	Automated Scan (38.49%)
TorLand1	1	0.75%	2011/12–2016/10	307	Malicious Traffic (16.99%)
jahjah	1	0.17%	2016/1–2017/1	75	Unauthorized Login Attempts (34.15%)
Our exits	10	3.14%	2016/9–2017/2	650	Network Attack (48.68%)

Table 2: Email complaints sent to the exit operators

jahjah exit was started in 2015 but the operator was only able to provide complaints received from 2016 onwards.

4.2 Analysis

We extract the nature of abuse, the time of complaint, and the associated exit IP addresses. 99.7% of the complaints received by Torservers.net related to Digital Millennium Copyright Act (DMCA) violations, with over 1 million of the complaints sent by one IP address. These emails use a template, enabling parsing of email text with regular expressions. The majority of the non-DMCA emails also follow a template, but the structure varied across a large number of senders. To extract the relevant abuse information from non-DMCA complaint emails, we first applied KMeans clustering to identify similar emails. We manually crafted regular expressions for each cluster. We used these regular expressions to assign high confidence labels to emails. Not all emails matched such a template regular expression—e.g., one-off complaints sent by individuals. We classified these emails by looking for keywords related to types of abuse. We iteratively refined this process until manually labeled random samples showed the approach to be quite accurate, with only 2% cases of misidentification.

99.99% of all DMCA violation complaints were against the Torservers' exits. The other exits collectively received only 12 such complaints. Over 99% of DMCA complaints mentioned the usage of BitTorrent for infringement; the rest highlighted the use of eDonkey.

We categorized the Non-DMCA complaints into five broad categories enumerated in Table 3. Network abuse is the most frequent category of non-DMCA complaints. ≈15% of the complaints related to network abuse came from Icecat [24], a publisher of e-commerce statistics and product content. Icecat's emails complain about excessive connection attempts from the jahjah exit and 13 exit IP addresses hosted by Torservers. These emails were received from November 2011 until December 2012. Other exit operators also received similar complaints during the same time frame [25]. We checked a recent Tor consensus in November 2016 and found eight exits that avoided exiting to the Icecat IP address.

The second most common non-DMCA complaints are about automated scans and bruteforce login attacks on Wordpress. Automated scanning, specifically port and vulnerability scanning, accounts for 13.6% of the non-DMCA complaints across the entire time range of our dataset. Instances of Wordpress bruteforce login attacks lasted for a comparatively shorter period, September 2015 until May 2016, but constitute 12.1% of non-DMCA complaints. All of the exits in our dataset received complaints about bruteforce login attempts from Wordpress except our own exits, probably because we started our exits after the attack stopped.

Email, comment, and forum spam constitutes 9.01% of non-DMCA complaints. Note that all of the exits in question have the SMTP port 25 blocked. Our data shows a spike in the number of abuse complaints regarding referrer spam from Semalt's bots [26] towards the end 2016. 1.05% of the non-DMCA emails complain about harassment. Over 11% of the non-DMCA emails do not fall under the mentioned categories. These emails include encrypted emails and emails unrelated to abuse.

4.3 Consequences of Undesired Traffic

Along with the complaints, some emails mention the steps the sender will take to minimize abuse from Tor. 34.3% of the emails mentioned temporary blocking (19.8%), permanent blocking (0.2%), blacklisting (9.8%) or other types of blocking (4.6%). The rest of the emails notify the exit operators about the abuse. For the most frequent form of blocking, temporary blocking, the emails threaten durations ranging from 10 minutes to a week. Some companies (e.g., Webiron) maintain different blacklists depending on the severity of the abuse. The majority of the blacklists mentioned in the emails are either temporary or unspecified. Only 18 emails mentioned permanently blocking the offending IP address. A small fraction (less than 1%) of the emails ask exit operators to change the exit policy to disallow exiting to the corresponding website.

We did not find any complaint emails from known Tor discriminators, such as Cloudflare and Akamai. Among the websites we crawled to quantify discrimination against Tor, we found complaints from Expedia and

Category	Includes	Percent
Network abuse	DDoS, botnet, compromised machines	38.03%
Unauthorized access	Failed login attempts, brute-force attacks, exploits for gaining access	26.45%
Automated scan	Port scans, vulnerability scans, automated crawling	14.15%
Spam	Email, comment, and forum spam	9.01%
Harassment	Threats, obscenity	1.05%
Other	(unreadable encrypted emails, emails not reporting abuse)	11.31%

Table 3: Categories of the Non-DMCA Email Complaints (Total 8,370 emails)

Zillow. Expedia complained about an unauthorized and excessive search of Expedia websites and asked exit operators to disallow exiting to the Expedia website. Zillow's complaint was less specific, about experiencing traffic in violation of their terms and conditions.

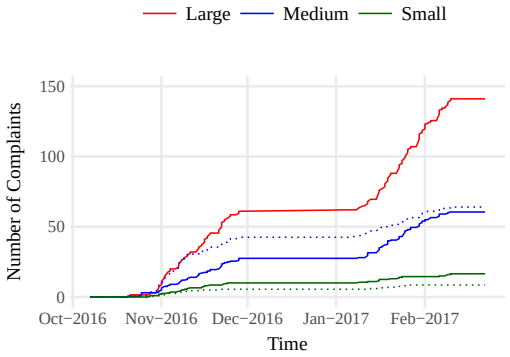


Figure 1: The cumulative number of complaints averaged over our exits sharing the same bandwidth and policy. Solid lines represent default policy exits and dashed lines represent RR policy exits.

4.4 Exit Properties and Complaints

We investigate the effects of two exit properties on the number of corresponding abuse complaints received: policy and bandwidth. For this analysis, we counted the number of email complaints that explicitly mention the IP address of our exits. We find that higher-bandwidth exits received more complaints (Figure 1). This correlation is statistically significant (Pearson's product-moment correlation = 0.98, p-value = 0.0016). However, exit policy did not have any statistically significant correlation with the number of complaints. We also did not notice any significant differences between the types of complaints that exits received.

4.5 Comparison with Average Tor Traffic

We estimate the average number of simultaneous Tor users per day through the exits getting complaints and compare it with the total amount of abuse going through the exits. Our goal is to understand how many Tor users will be affected if we block an exit because of abuse complaints. To do this, we collect the estimation of simultaneous Tor users per day from Tor Metrics [27]. Then we collect the historical Tor consensus to compute how much traffic went through an exit per day. If an exit A handles $e\%$ of the total Tor bandwidth on day d and the number of simultaneous users of Tor on d is u, then approximately $\frac{eu}{100}$ of the users used A on day d.[2] We estimate users for each exit in Table 2 from September 2011 (the beginning of the Tor metrics data). We exclude the Torservers exits because tracing the Torservers exits in the historical consensuses is difficult as those exits changed IP addresses and exit fingerprints more than once.

Exit Family	Avg. users	Avg. complaints
apx	23,082.08	0.53
TorLand1	59,284.54	0.17
jahjah	1,206.30	0.19
Our exits	3,050.81	5.55

Table 4: Avg. simultaneous users and complaints per day

Compared to the average number of Tor users, the amount of abuse is insignificant (Table 4). However, we are considering one abuse email as one instance of abuse, but in practice one email can correspond of many instances of abuse, for example, one brute-force attack can consist of thousands of visits to a website.

5 IP Address Blacklisting

We analyze how popular commercial IP blacklists treat Tor relays. IP blacklisting can be in response to malicious traffic originating from the IP, which we call *reactive blacklisting*. We also observed *proactive black-*

[2]Even though the bandwidth is one of the main factors for selecting an exit, the other factors such as the exit policy might affect which exits will be selected. For our estimation, we do not consider the effect of exit policies.

listing, blacklisting based upon a network's pre-existing reputation or the online service's policy (e.g., the video-on-demand service, Hulu, blocks access to all VPN endpoints). After discussing our data sources, we describe how we classify blacklist entries into *proactive* blacklisting of Tor simply due a policy decision to deny access from Tor, versus *reactive* blacklisting in response to abuse. We then look at the amount of blacklisting of Tor and compare it to VPN IP address spaces and the IP address space of a large university in the USA. We analyze the impact of relay uptime, consensus weight, and exit policy on blacklisting behaviour.

5.1 Data Sources

For our study we were given access to a system that gathers commercial IP threat intelligence, including blacklists, from large Web companies. Facebook's ThreatExchange [28] platform is a major contributor to the system. This system has gathered roughly 2TB of data from 110 sources since July 25, 2015. We have anonymized the names of some IP blacklists in our results.

Along with the hourly Tor consensus data, we use additional methods to gather the set of Tor exit IP addresses seen by servers. While the Tor consensus provides the IP addresses used to reach exit relays (their "onion routing" IP addresses), a significant fraction of all exit relays (6% to 10%) use a different IP address for connecting to servers. To capture these IP addresses, we also associate with each relay its exit IP address provided by Tor DNSEL [29]. Tor DNSEL gathers the IP address used by a relay for exiting traffic based on active testing.

5.2 Classifying Blacklist Entries

Given Tor's reputation of transiting undesired traffic, some blacklists proactively include Tor relay IP addresses. Since we are interested in the rate and impact of undesired traffic Tor is currently producing, we must separate *proactive* blacklisting based upon historical events from *reactive* blacklisting based upon current events.

We use several methods to classify blacklist entries into *proactive* and *reactive* ones. In the simplest case, the blacklist provides the reason behind inclusion, either on an entry or on a list-wide basis. In some cases of *reactive* listing, the blacklist even provides information about the undesired traffic leading to blacklisting.

For those entries on lists that do not provide reasons for inclusion, we look at the behavior of the list overall to infer its reason for blacklisting. We infer that lists including a large percentage of Tor IP addresses soon after they appear in the consensus data likely reflect *proactive* listing of the addresses. If more than 30% of Tor relay addresses have been enlisted on a blacklist within 24 hours of them appearing in the consensus, we consider that blacklist *proactive*. We consider the remaining lists to be *reactive*. We discuss the details of deciding the threshold of 30% in Section A of the Appendix.

Figures 2a and 2b compare the rate of blacklisting by a *proactive* and a *reactive* blacklist. These graphs show the rate of blocking Tor exit IP addresses and of non-exit Tor IP addresses, whose blocking may be superfluous. In a small number of cases, the time until blacklisting is negative since the address was blacklisted before appearing in the consensus data, presumably from the IP address's prior use or the blacklisting of whole blocks of addresses. Under our analysis, the blacklist *Paid Aggregator*, a large paid provider of threat intelligence, is a *proactive* blacklist since 76.6% of Tor IPs enlisted on it were added within 24 hours of them first appearing in the consensus (Fig. 2a). The distributions show that the majority of the listed IP addresses get listed within a few hours of them becoming Tor relays. We classify *Contributed Blacklist 12*, a data source that contributes threat intelligence to a community aggregation project, as *reactive* since only 0.06% of all Tor IP addresses were added within the first 24 hours of their appearance in the consensus or the DNSEL (Fig. 2b).

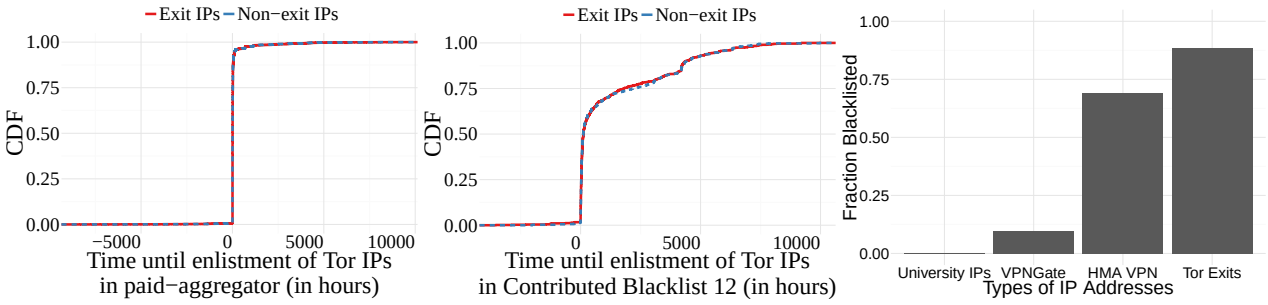
Using both methods of classifying lists, we found 84 lists that either include Tor exits proactively or reactively. Using the lists' labels and names, we classified 4 blacklists as *proactive*. We additionally classify 2 blacklists as *proactive* based on the time taken by them to enlist Tor IP addresses.

Identifying the *proactive* blacklisting of Tor exits also sheds light on the nature of Tor blocking employed by servers today. *Proactive* blacklisting implies that Tor users share fate not only with other users of their exits but also with *all* Tor users, including the ones in the distant past. We find that 6 out of 84 (7%) large commercially deployed blacklists proactively block Tor IP addresses.

5.3 Amount of Blacklisting

Figure 3 depicts the fraction of exit/non-exit relay IP addresses blacklisted by various lists during the observation time frame. From 110 blacklists that the IP reputation system gathers, 84 list Tor IP addresses in the observation time frame. For legibility, Figure 3 shows only the lists that included more than 1% of either Tor non-exit relays, Tor exit relay, or a VPN's IP addresses.

We observe that a few blacklists list a large number of Tor IP addresses, including non-exit relay IP addresses. In particular, *Paid Aggregator* (the *proactive* list shown in Fig. 2a) listed not only 48% of Tor exit addresses, but also 35% of entry and middle relay IP addresses. Blacklisting non-exit relays is surprising, since non-exit relays are not responsible for exiting traffic from the Tor net-



(a) *Paid Aggregator* (proactive) (b) *Contributed Blacklist 12* (reactive) (c) Fraction of IPs blacklisted

Figure 2: (a) and (b) provide the time (in hours) between first seeing a relay IP address in the consensus until the given blacklist enlists the IP address. Negative values indicate cases where the IP address was blacklisted before appearing in the consensus. Figure (c) compares the fraction of public IP addresses of different types of networks that are currently in any tracked blacklist.

work. Some relays have historically, at different points in time, been both exit and non-exit relays in the Tor consensus. In our analysis, we consider a relay an exit if it had the *Exit* flag at any point in its lifetime. Doing so provides a conservative estimate for the number of non-exit IP addresses that get blacklisted. In contrast, the *Snort IP* blacklist (another *proactive* list) enlists nearly 37% of exit IP addresses but less than 1% of non-exit relays.

5.4 Blacklisting of Tor vs. VPN nodes

VPN services are similar to Tor since they provide users with the option to obscure their IP addresses. In addition, like Tor exit relays, VPN nodes also egress traffic belonging to many users who could be using the VPN service for different purposes. In this section we compare blacklisting of Tor with that of popular VPN providers.

VPN providers like VPNGate [30] and HideMyAss [31] publish lists of their free-tier endpoints, making them good candidates for our study. Figure 2c shows, that in February 2017, over 88% of Tor exits are blacklisted (excluding the *proactive* blacklists) on one or more of the commercially available blacklists. In comparison, 10% of VPNGate endpoints and 69% of HMA endpoints appear on blacklists. All of these proxy services are considerably more blacklisted as compared to the IP space of a major university (three /16 prefixes used by the university campus network), of which only 0.3% IPs are blacklisted.

To have a fair comparison of the rate of blacklisting with Tor, we need a set of VPN endpoint IP addresses and a notion of when they first began to operate as VPN endpoints (similar to the notion of exit relays and their birth in the consensus). However, it is challenging to gather the IP addresses of VPN nodes over time since most VPN services do not archive such information. This is in contrast with Tor, which archives information about its relays on an hourly basis. However, the VPN provider HideMyAss (HMA) publishes a daily list of its free VPN endpoints [31]. We crawled archived versions of this list using the Wayback Machine [32] for IP addresses published between June 14, 2014 and October 27, 2016. We can then approximate the time when an IP address first served as an HMA VPN endpoint, assuming this occurs at least 60 days after the start of the time frame. In this manner we collected a set of 4,234 HMA endpoints and their *first seen* creation times. Of these, 1,581 IP addresses became HMA endpoints after our IP address reputation system started gathering blacklist data. We analyze the blacklisting of these endpoints using the IP reputation system.

Figure 3 shows the fraction of HMA endpoints blacklisted by various blacklists. Unlike for Tor relays, no particular blacklist dominates in the listing of HMA IPs. Figure 4 shows how quickly HMA endpoints get blacklisted compared to Tor exits; *reactive* blacklisting of both occurs at a similar rate.

5.5 Exit policies and bandwidth

We looked for but did not find any associations between various factors and blacklisting. In an attempt to counter IP blacklisting and abusive traffic, the Tor community has suggested that exit operators adopt more conservative exit policies [18]. Intuitively, a more open exit policy allows a larger variety of traffic (e.g., BitTorrent, ssh, telnet) that can lead to a larger variety of undesired traffic seen to originate from an exit. We analyze the exit relays that first appeared in the consensus after the IP reputation system started to gather data using the hourly consensuses of year 2015 and 2016. Since exit relays have a variety of exit policies, we find which well-known exit policy (Default, Reduced, Reduced-Reduced, Lightweight, Web) most closely matches the relay's exit policy. To

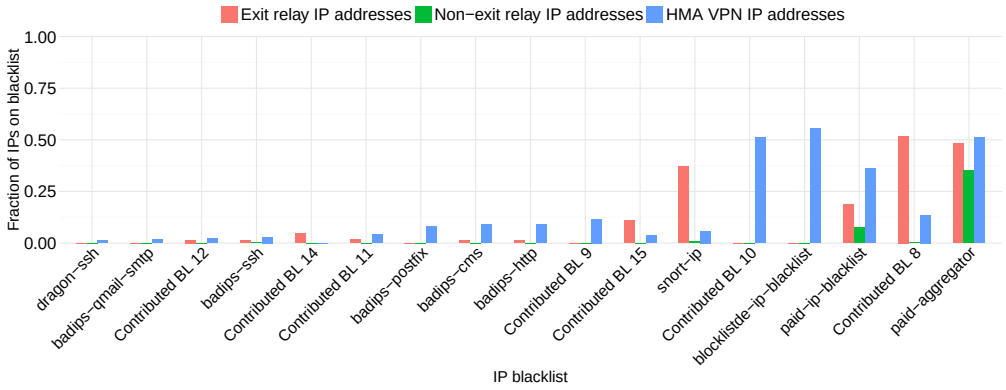


Figure 3: Fraction of Tor relay and HMA VPN IP addresses listed in IP blacklists (including *proactive* and *reactive)*. Some feed names are derived based on the broad categories of undesired traffic they blacklist: e.g., ssh (badips-ssh, dragon-ssh), content management systems/Wordpress (badips-http, badips-cms).

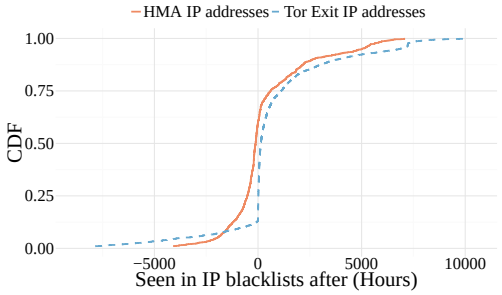


Figure 4: Comparing the time taken for Tor exit IP addresses and HMA endpoints to get blacklisted.

compute this closeness between exit policies, we calculate the Jaccard similarity between the set of open ports on a relay and each well-known exit policy. (See Appendix B). In this way, we associated approximate exit policies to 21,768 exit relays. We found that in the last 18 months, only 1.2% of exit relays have exhibited different well-known exit policies, and excluded these from our analysis. In the resulting set of exits, we assigned 81% to Default, 17% to Reduced, 0.6% to Reduced-Reduced, 0.5% to lightweight and 0.4% to Web policy.

We also compute the uptime (in hours) for each of the exit relays as the number of consensuses in which the relay was listed. In addition, we maintain the series of consensus weights that each relay exhibits in its lifetime. Higher consensus weights imply more traffic travelling through the relay, proportionally increasing the chance of undesired traffic from a relay. A high uptime increases the chance of use of a relay for undesired activities.

We trained a linear regression model on the policies, scaled uptimes, and consensus weights of exit relays, where the observed variable was the ratio of hours the IP address was blacklisted (*reactive* blacklisting only) and its overall uptime. Based on the coefficients learned by the regression model, we conclude that policy, consensus weight, and relay uptime have very little observed association on IP blacklisting of Tor relays. We provide more details about the regression model in Appendix C.

5.6 Our Newly Deployed Exit Relays

As described in Section §3, we operated exit relays of various bandwidth capacities and exit policies to actively monitor the response of the IP reputation system. In this subsection, we analyze the sequence of blacklisting events for each exit relay that we ran. Figure 5 shows the timeline of blacklisting events for each of the exit relays we operated. Each coloured dot represents an event. An event is either the appearance of a relay on a blacklist or its appearance in the consensus (an *up* event).

Prior to launching the exits, none of our prospective relays' IP addresses were on any blacklist. We see that within less than 3 hours of launching, feeds like *Snort IP* listed all our relays, supporting our classification of *Snort IP* as a *proactive* blacklist. Additionally, both *Snort IP* and *Paid Blacklist* (also classified as *proactive*) block our relay IP addresses for long periods of time. *Snort IP* enlists all of relays, and did not remove them for the entire duration of their lifetime. *Paid Blacklist* enlists IP addresses for durations of over a week. Blacklists such as badips-ssh (for protecting SSH) and badips-cms (for protecting content management systems such as Wordpress and Joomla) have short bans spanning a few days. *Contributed Blacklist 12* has the shortest bans, lasting only a few hours. We consider *Contributed Blacklist 12*'s blacklisting strategy in response to undesired traffic to be in the interest of both legitimate Tor users and content providers that do not intend to lose benign Tor traffic. On November 29, 2016, we turned off all of our relays to observe how long a *proactive* blacklist like *Snort IP* would take to de-enlist our relays. We observe that such

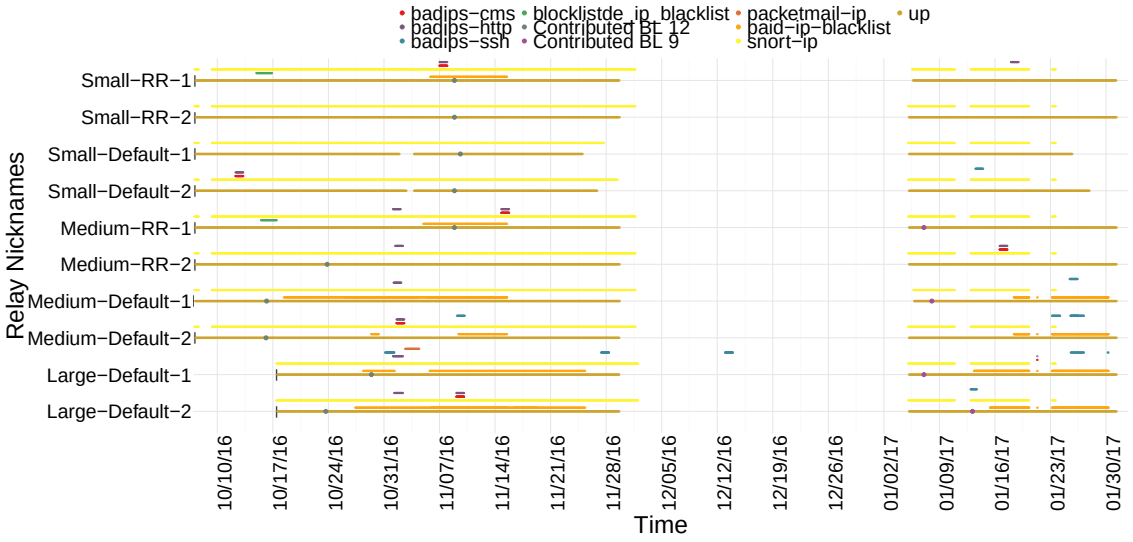


Figure 5: Blacklisting of our exit relays over time. Each coloured dot shows the instant when a relay was on a blacklist. *Snort IP* and *Paid Blacklist* have long term bans while other blacklists enlist IPs for short periods of time ranging from hours to a few days.

blacklists drop our relays just as fast as they enlist them, suggesting a policy of crawling the Tor consensus.

Note that a synchronised absence of data from any blacklist, while the relays are up, represents an outage of the IP reputation system.

6 Crawling via Tor

To quantify the number of websites discriminating against Tor, we performed crawls looking both at front-page loads, as in prior work [2], and at *search and login functionality*. We crawled the Alexa Top 500 web pages from a control host and a subset of Tor exit relays. These crawls identify two types of discrimination against Tor users: (1) the Tor user is blocked from accessing content or a service accessible to non-Tor users, or (2) the Tor user can access the content or service, but only after additional actions not required of non-Tor users—e.g., solving a CAPTCHA or performing two-factor authentication.

6.1 Crawler Design

We developed and used a Selenium-based interactive crawler to test the functionality of websites. We performed three types of crawls: (1) *Front-page crawls* attempt to load the front page of each website. We repeated the crawl four times over the course of six weeks. (2) *Search functionality crawls* perform front-page loads and then use one of five heuristics (Table 5) to scan for the presence of a "search box". Upon finding the search box, the crawler enters and submits a pre-configured search query. Our crawler found and tested the search functionality of 243 websites from the Alexa Top 500. We performed the search functionality crawl once. (3) *Login functionality crawls* load front pages and scan them for the presence of a "login" feature. Upon finding the feature, and if it has credentials for the webpage available, the crawler authenticates itself to the site (using Facebook/Google OAuth when site-specific credentials were unavailable). We created accounts on OAuth-compatible websites prior to the crawl. Since the created accounts had no prior history associated with them, we speculate that they were unlikely to be blocked as a result of unusual behavior. For example, we found that LinkedIn blocks log ins from Tor for accounts with prior log in history, but not for new accounts. Our crawler found and tested the login functionality of 62 websites from the Alexa Top 500. We performed the login functionality crawl once.

The crawler records screenshots, HTML sources, and HARs (HTTP ARchives) after each interaction. Our interactive crawler improves upon previous work in several ways. First, it uses a full browser (Firefox) and incorporates bot-detection avoidance strategies (i.e., rate-limited clicking, interacting only with visible elements, and action chains which automate cursor movements and clicks). These features allow it to avoid the (bot-)based blocking observed while performing page-loads via utilities such as `curl` and other non-webdriver libraries (`urllib`). Second, its ability to interact with websites and exercise their functionality allows us to identify cases where discrimination occurs beyond the front page — e.g., `www.tumblr.com` serves Tor users

Heuristic	Coverage
1. Visible and clickable textbox elements containing search related keywords (`q`, `query`, `querytext`, `search`) in their element `name`, `id`, `value`, or `label` are assumed to be search boxes.	98
2. The above heuristic is repeated while considering all input DOM elements.	81
3. If the DOM contains exactly one visible and clickable textbox element, it is assumed to be a search box.	22
4. If the DOM contains exactly one visible and clickable input element with a defined `max-length`, it is assumed to be a search box.	12
5. If the DOM contains exactly one visible and clickable input element, it is assumed to be a search box.	30

Table 5: Heuristics used to identify search input boxes. Heuristics are described from most specific to least specific. Coverage indicates the number of sites that were identified using the corresponding heuristic.

CAPTCHAs only after they submit a search query, and `www.imdb.com` blocks Tor users when they attempt to log in.

6.2 Relay selection

We randomly selected 100 exit relays from the set of all exit relays that supported HTTP(S) connections (i.e., the exit policy allows outbound connections to ports 80 and 443). In addition to these randomly sampled relays, we also conducted crawls through our own relays (described in Table 1) and a university-hosted control host.

Since we performed our crawls over a six-week period, several of the selected exit relays intermittently went offline, with a total of 0, 12, 19, and 28 offline during crawls 1–4, respectively. We account for the resulting page-load failures by excluding the failures from our analysis.

6.3 Identifying discrimination

In each of our experiments we simultaneously performed crawls exiting through all online sampled exits and our university-hosted control host. To identify discrimination of a selected exit relay, we first rule out cases of client and network errors through HAR file analysis. We use the HAR files to verify, for each page load, that (1) the requests generated by our browser/client were sent to the destination server (to eliminate cases of client error), and (2) our client received at least one response from the corresponding webpage (to eliminate cases of network errors). If, for a given site, either the control host or the selected exit relay did not satisfy both these conditions, we did not report discrimination due to the possibility of a client or network error.

Next, we compare the crawler-recorded screenshots of the control server and each selected exit relay using *perceptual hashing (pHash)* [33], a technique that allows us to identify the (dis)similarity of a pair of images. We report images with high similarity scores (pHash distance < 0.40) as cases where no discrimination occurred and images with high dissimilarity (pHash distance > 0.75) as cases of discrimination, while flagging others for further inspection. The thresholds were set so that only pages with extreme differences in content and structure would be automatically flagged as cases of discrimination, while similar pages were automatically flagged as cases of non-discrimination. In general, minor changes in ads/content (e.g., due to geo-location changes) do not result in flagging. We set the thresholds using data obtained from a pilot study (Figure 6).

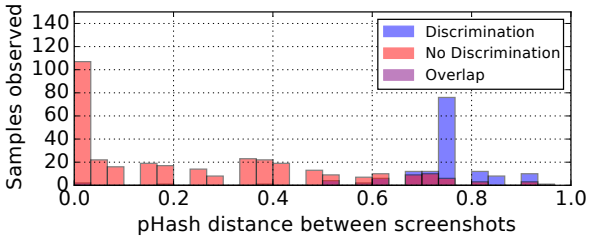


Figure 6: Results of pilot study to identify pHash distance thresholds for automatically identifying cases of (non) discrimination. We manually tagged 500 randomly chosen samples (i.e., pairs of control and exit relay screenshots of the same website) and computed the pHash distances. Based on the above distribution, we classified distances < 0.40 as "non-discrimination" and distances > 0.75 as "discrimination". Instances having pHash distances in the 0.40 to 0.75 range were manually inspected and tagged.

Then, we classified as discrimination cases where exit relays received HTTP error codes for requests that our control host successfully loaded with a 200 status. Finally, we manually tag the screenshots of remaining cases to identify more subtle discrimination—e.g., a block-page served with a 200 status.

6.4 Results

Table 6 summarizes the main results of our three types of crawls over compatible websites in the Alexa Top 500. Here, we show the fraction of interactions on which discrimination was detected. We find that 20.03% of all Alexa Top-500 (A-500) website front-page loads showed evidence of discrimination against Tor users, compared to 17.44% of the search-compatible (S-243) and 17.08% of the login-compatible (L-62) website front-page loads.

When exercising the search functionality of the 243 search-compatible websites, we see a 3.89% increase in discrimination compared to the front-page load discrimination observed for the same set of sites. Similarly, when exercising the login functionality of the 62 login-compatible websites, we observe a 7.48% increase in discrimination compared to the front-page discrimination observed for the same set of sites.

Websites	Interaction	Discrimination observed
A-500	Front page	20.03%
S-243	Front page	17.44%
	Front page + Search	21.33% (+3.89%)
L-62	Front page	17.08%
	Front page + Login	24.56% (+7.48%)

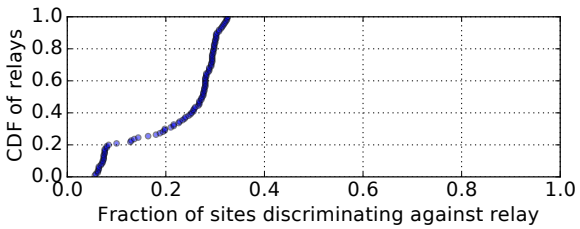
Table 6: Fraction of interactions blocked from 110 exits. A-500 denotes the Alexa Top 500 websites, S-243 denotes the 243 search-compatible websites, and L-62 denotes the 62 login-compatible websites.

Figure 7a shows the distribution of discrimination (for any interaction) faced by relays from websites in the Alexa Top 500. We find that no relay experiences discrimination by more than 32.6% of the 500 websites, but 50% of the exit relays are discriminated against by more than 27.4% of the 500 websites. Figure 7b shows the distribution of discrimination performed by websites against Tor exit relays. Here, we see that 51% of the websites perform discrimination against fewer than 5% of our studied exits, while 11% of websites perform discrimination against over 70% of our studied exits.

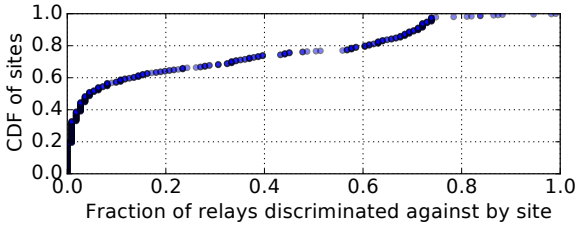
We now examine various factors associated with Tor discrimination. Since we did not (and in many cases cannot) randomly assign these factors to websites or relays, these associations might not be causal.

Hosting Provider. Figure 8 shows the fraction of relays discriminated against by websites hosted on four of the six most-used hosting platforms. We find that Amazon- and Akamai-hosted websites show the most diversity in discrimination policy, which we take as indicative of websites deploying their own individual policies and blacklists. In contrast, CloudFlare has several clusters of websites, each employing a similar blacklisting policy. This pattern is consistent with CloudFlare's move to allow individual website administrators to choose from one of several blocking policies for Tor exit relays [1]. Finally, we see 80% of China169- and CloudFlare-hosted websites perform discrimination against at least 60% of our studied relays.

Relay Characteristics. Our analysis of the association between exit-relay characteristics and the discrimination



(a) Distribution of discrimination faced by relays.



(b) Distribution of discrimination performed by websites.

Figure 7: Distribution of discrimination by Alexa Top 500 websites against 110 exit relays.

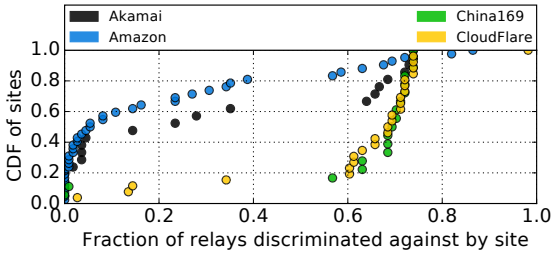


Figure 8: Distribution of discrimination performed by websites hosted on four of the six most popular hosting platforms.

faced by them found no significant correlations when accounting for relay-openness (fraction of ports for which the exit relay will service requests) or for the age of the relay. We found a small positive correlation (Pearson correlation coefficient: 0.147) between the relay bandwidth and degree of discrimination faced, but the result was not statistically significant (p-value: 0.152). Figure 9 presents these results graphically. We further analyze the impact of relay characteristics on discrimination performed by websites using popular hosting providers . We find that only Amazon has a statistically significant positive correlation between discrimination observed and relay bandwidth (Pearson correlation coefficient: 0.247, p-value: 0.015). These results are illustrated in Figure 10.

Service Category. We now analyze how aggressively four different categories of sites—search engines, shop-

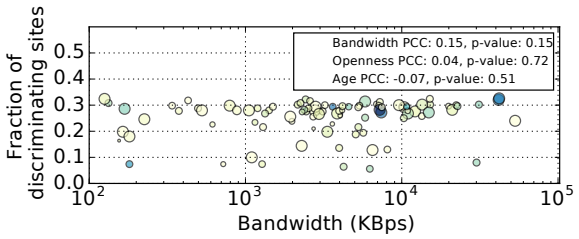


Figure 9: Relationship between relay characteristics and discrimination faced. Each circle represents a single relay. Lighter colors indicate younger relays and larger circles indicate more open exit policies. The legend shows the Pearson correlation co-efficient (PCC) and the p-value for each characteristic.

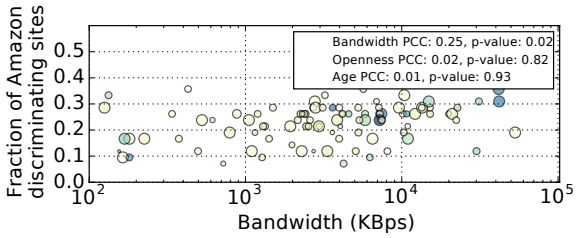


Figure 10: Impact of relay characteristics on discrimination performed by websites hosted by Amazon.

ping, news, and social networking—discriminate against Tor exit relays. We categorize sites using the McAfee URL categorization service [34]. We find that search engines are the least likely to discriminate against exit relays, with 83% of all search engines discriminating against fewer than 20% of our studied exit relays, compared to 30% of social networking sites, 32% of shopping sites, and 53% of news sites. We also find social networking and online shopping sites share similar blocking behavior. Websites in these categories are also observed to be the most aggressive—with 50% of them blocking over 60% of the chosen relays. Figure 11 illustrates the results.

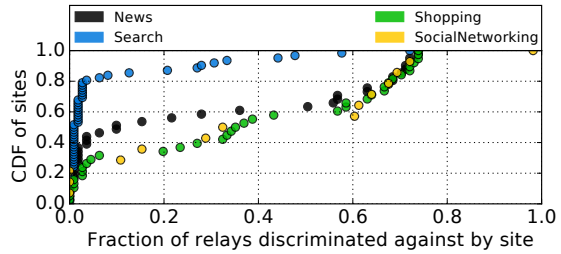


Figure 11: Distribution of discrimination performed by websites in various categories.

The Evolution of Tor Discrimination. We now focus on discrimination changes over time. For this experiment, we conducted four crawls via our own ten exit relays to the Alexa Top 500 websites. Let Day 0 denote the day when we set the relay's exit flag. We conducted crawls on Day -1, Day 0, and once a week thereafter. Table 7 shows the fraction of websites found to to discriminate against each exit set during each crawl. We observe increases in discrimination when the exit flag is assigned. We can attribute some of this can to our improved crawling methodology deployed on Day 0 (the Day -1 crawl utilized the crawler from Khattak et al.; see below), although we note that the IP addresses used by our exit relays were never used by other Tor exit relays in the past, and did not appear in any of our studied commercial blacklists before Day 0, while they immediately manifested after setting the exit flags.

Configuration	Day -1	Day 0	Wk. 2	Wk. 3	Wk. 4
Large-Default	NA	17.0	19.0	21.1	25.4
Medium-Default	9.4	20.5	24.4	25.6	24.8
Medium-RR	9.9	18.3	24.1	22.7	24.7
Small-Default	9.3	20.3	20.9	23.9	23.6
Small-RR	9.4	20.5	20.7	25.7	25.3

Table 7: Percentage of discriminating page loads for each set of deployed relays.

The high amount of discrimination observed on our Day-0 crawl for all exit relays is indicative of *proactive* discrimination against Tor exit relays. Our results do not indicate differences due to relay cateogry in the amount of discrimination experienced.

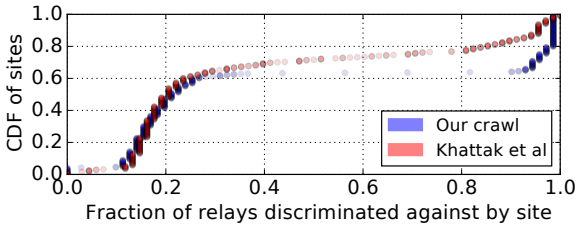


Figure 12: Impact of methodological changes on measured discrimination from data generated by a single front-page crawl.

Measurement Methodology. We now measure the impact of changes in our discrimination identification methodology compared to previous work by Khattak et al. [2]. The key differences between the methodologies are: (1) The measurements conducted by Khattak et al. are limited to identifying front-page discrimination. Our crawler also tests search and login interactions. Table

6 presents the impact of this feature. (2) Khattak et al. identify discrimination using the difference in HTTP status codes returned by the control and test nodes. This method is prone to underestimating discrimination due to the inability to detect block pages that return a HTTP 200 OK status code. Our method relies on screenshot differences and HTTP status codes as a signal for discrimination. As a result, we are able to detect discrimination performed by sites such as `livejournal.com`, `hdfc.com`, and `glassdoor.com`. (3) Khattak et al. rely on sending HTTP requests for front pages of websites using the python `urllib2` library. Although they modify the user agent of their crawler to match a regular web browser, they are easily identifiable as an irregular user since they do not load third-party objects and JavaScript. Such crawlers are blocked by many websites and bot-mitigation tools [35]. In contrast, we perform complete page loads, including third-party content and execution of JavaScript. As a consequence, our crawls are slower, requiring around 12 hours for 500 page loads (compared to 1–2 minutes required by the `urllib2` crawler).

To understand the impact of (2) and (3), we compare the discrimination results obtained from a single front-page crawl performed by both crawlers. We started both crawls on the same day, on the same set of websites, using the same set of 100 randomly sampled exit relays. The results, illustrated in Figure 12, confirm that previous work underestimates the amount of discrimination.

7 Privacy-sensitive Exit Logging

While our crawls systematically explore popular websites, they might not be typical of actual Tor usage. Thus, we performed privacy-sensitive logging on our deployed exit relays to measure how commonly users interacting with Alexa Top 1M web pages experienced failed TLS handshakes or HTTP requests This observational dataset, based on actual Tor-user web traffic distributions and user interactions, provides us with a picture of the discrimination actually encountered by users.

7.1 Logging Approach

In order to measure the number of failed TLS handshakes and HTTP requests, we developed a custom logger. Although tools such as PrivEx [36], Historε [37], and PrivCount [38] were specifically built for measuring characteristics of Tor exit traffic, they are not suitable for our study for two reasons: First, they currently do not have the capability to inspect HTTP and TLS traffic headers. Adding such functionality to the tools requires modifying the Tor relay source code—possibly introducing users of our relays to new vulnerabilities. Second, they were built with the goal of performing secure data aggregation across multiple relays. Since a single entity owned and operated all of the relays used in our study, this feature was unnecessary for our purposes.

We maintain counters for several events of interest associated with users browsing websites in the Alexa Top 1M. Our approach, designed after consultation with members of the Tor developer community, takes precautions to avoid de-anonymization of users. Since neither the Tor users nor the service operators were the subjects of our study, we were exempt from an IRB review.

First, we use bucketing and split the Alexa Top 1M websites into exponentially growing sets based on their Alexa ranks, as follows: The first set contains the top 100 websites (ranked 1–100) and the nth set for $n > 1$ contains the top $100 \times 2^{n-2} + 1$ to $100 \times 2^{n-1}$ websites. We keep a separate event counter for each set. Second, we maintain our event counters in memory and write to disk only once a day. Doing so allows our event counters to attain higher count values, increasing anonymity-set sizes. Third, to deal with the possibility of encountering cases where 24 hours does not suffice to achieve reasonably high anonymity-set sizes—e.g., if only one person visited a site during a 24 hour period—we round up each event counter to the nearest multiple of eight before writing to disk. A similar approach is used by Tor metrics [3] for reporting counts of bridge users per country.

We maintained per-bucket event counters for the number of: (1) HTTP requests to website front pages, (2) error status codes observed in their corresponding responses, (3) HTTP(S) handshakes initiated, and (4) timed-out handshakes encountered. Additionally, we also maintained a counter for the number of packets sent through each open port.

7.2 Results

Table 8 shows the percentage of failed HTTP requests and incomplete HTTPS handshakes encountered by users of our exit relays. We find that the fraction of incomplete handshakes steadily increases over time. We attribute the steep increase in HTTP error codes received during weeks four and five to our relays being (ab)used in a scraping attempt on a popular website (we received a complaint notice due to this behavior). Besides this sudden increase, we see that the fraction of HTTP errors accords with data observed through our crawls, but the fraction of incomplete HTTPS handshakes runs higher. This is likely because incomplete handshakes provide only very noisy indicators for user discrimination, with many reasons for them to occur naturally.

HTTP requests and error response codes. For exiting packets using the HTTP protocol, *iff* the URI on the HTTP request was identical (ignoring case) to

Week	1	2	3	4	5	6
HTTP	15.8	18.1	19.8	32.8	33.4	17.9
HTTPS	36.3	35.0	41.1	45.2	47.9	49.6

Table 8: The percentage of failed HTTP requests and incomplete HTTPS handshakes observed over time.

a Top 1M website, we incremented a *front-page request* event counter associated with the set containing the site. For every matching request, we maintained state to identify the corresponding response packet. If the corresponding response packet contained an error status code (4XX/5XX), we incremented an *error-status* event counter associated with the corresponding set. We break down the fraction of errors by website ranks and time in Figure 13a. We see that the fraction of error response codes is nearly evenly distributed across each set, indicating that errors are independent of website ranks.

HTTPS handshake initiation and failure. The procedure for HTTPS is similar to that for HTTP. However, we use the SNI value of *client-hello* handshake initiation packets instead of the URI of HTTP requests. Furthermore, we look for handshake failures and timed-outs instead of HTTP errors. The results in Figure 13b show a strong increasing trend in incompletion rates over time.

8 Discussion and Future Work

Limitations. Our studies each come with their own limitations, some resulting from our desire to protect the privacy of Tor users, others from the limited data sets available for study. Neither our set of emails nor our set of blacklists are complete. Given that Tor assigns traffic to exits in a mostly random fashion, we believe the emails from our sample to be representative of the complaints during their time periods for exits with similar exit policies. While there are blacklists that we were not able to observe during the period of our study, the set of blacklists used in our analysis includes numerous types from a wide range of suppliers, leading us believe that they capture all common blacklisting phenomena. Our crawls, while more in-depth than prior efforts [2], were too time-consuming to run often enough to gain statistical guarantees about discrimination by any one website. Nevertheless, taken together, they show that discrimination is common and sometimes subtle.

Implications for Tor. The large amounts of blocking and discrimination identified by our crawling and privacy-sensitive measurements suggest that Tor's utility is threatened by online service providers opting to stifle Tor users' access to their services (§6 & §7).

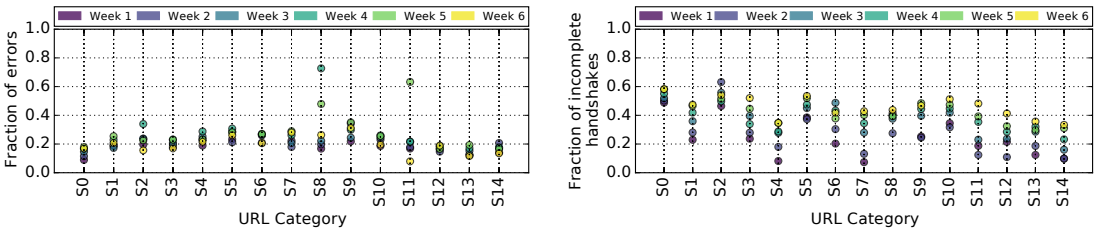
From studying blacklists we learned that some, but not all, proactively add Tor exit IP addresses (§5), presumably in response to prior undesired traffic and an expectation of more. This result highlights that Tor users fate-share with not just the Tor users sharing their current exit relay, but all Tor users—present and past. Other blacklisting appears to be reacting to undesired traffic, suggesting that blocking may decrease if Tor can reduce the amount of abuse it emits. Such a reduction may even, over time, decrease *proactive* blacklisting as Tor's reputation improves. These findings suggest the utility to implement any privacy-sensitive abuse-reduction approaches for Tor.

From the emails, we learned of the types of undesired traffic that server operators find concerning enough to warrant sending a complaint. Of the types of abuse identified in email complaints (§4), the vast majority—the DMCA complaints—appear irrelevant to blocking since DMCA violators largely use peer-to-peer services. Furthermore, at least in our sample they are no longer common (Table 2). Of the remaining complaints, nearly 90% related to large-scale abuse, such as excessive connection attempts, scanning, brute-force login attempts, and spam. While the rate of complaining might not be proportional to the rate of undesired traffic, it may provide some insights into the nature of the most troubling abuse exiting the Tor network. The exit policies have no significant impact on reducing abuse complaints and rate of discrimination against Tor users.

Given the large footprints of the observed abuse, we believe future research should seek to provide tools to curb such abuse while preserving privacy and Tor functionality. We envision Tor nodes using cryptographic protocols, such as secure multi-party computation and zero-knowledge proofs, to detect and deter users producing large amounts of traffic in patterns indicative of abuse. For example, Tor could compute privacy-sensitive global counts of visits to each threatened domain and throttle exiting traffic to ones that appear over-visited.

Implications for online services. Combining our study results, we can put the difficulties facing Tor users and online service operators into perspective: at most 182 email complaints per 100K Tor users, and over 20% of the top-500 websites blocking Tor users. Given that Tor users *do* make purchases at the same rate as non-Tor users [6], this response may be excessive and operators might wish to use less restrictive means of stifling abuse.

Operators can aid Tor in developing approaches to curb abuse or unilaterally adopt local solutions. For example, instead of outright blocking, servers could rate-limit users exiting from Tor for certain webpages (e.g., login pages). Indeed, CloudFlare is developing a cryptographic scheme using blindly signed tokens to rate limit Tor users' access to websites it hosts [39].



(a) Fraction of HTTP error codes

(b) Fraction of incomplete TLS handshakes

Figure 13: Fraction of errors encountered by users visiting the Top 1M websites over time. The URL category S1 consists of the top (1–100) websites and Sn ($n \geq 2$) consists of sites in the top [$100 \times 2^{n-2}+1$ to $100 \times 2^{n-1}$] ranks.

Ultimately, we do not view IP-based blacklisting as a suitable long-term solution for the abuse problem. In addition to Tor aggregating together users' reputations, IPv4 address exhaustion has resulted in significant IP address sharing. IPv6 may introduce the opposite problem: the abundance of addresses may make it too easy for a single user to rapidly change addresses. Thus, in the long run, we believe that online service operators should shift to more advanced ways of curbing abuse; ideally, ones compatible with Tor.

Acknowledgements

The authors would like to thank Facebook Threat Exchange for providing IP blacklists and Tor exit operators: Moritz Bartl (Torservers.net), Kenan Sulayman (apx), Riccardo Mori (jahjah), and the operator of the exit relay TorLand1 for sharing the abuse complaints they received. We are grateful to David Fifield, Mobin Javed and the anonymous reviewers for helping us improve this work. We acknowledge funding support from the Open Technology Fund and NSF grants CNS-1237265, CNS-1518918, CNS-1406041, CNS-1350720, CNS-1518845, CNS-1422566. The opinions in this paper are those of the authors and do not necessarily reflect the opinions of a sponsor or the United States Government.

References

[1] Matthew Prince. The Trouble with Tor. https://blog.cloudflare.com/the-trouble-with-tor/.

[2] Sheharbano Khattak, David Fifield, Sadia Afroz, Mobin Javed, Srikanth Sundaresan, Vern Paxson, Steven J. Murdoch, and Damon McCoy. Do you see what I see?: Differential treatment of anonymous users. In *Network and Distributed System Security Symposium 2016*. IETF, 2016.

[3] Tor Project: Anonimity Online. Tor Metrics. Available at https://metrics.torproject.org.

[4] Peter Zavlaris. Cloudflare vs Tor: Is IP Blocking Causing More Harm than Good? Distill Networks' Blog. Available at https://resources.distilnetworks.com/all-blog-posts/cloudflare-vs-tor-is-ip-blocking-causing-more-harm-than-good.

[5] Christophe Cassa. Tor – the good, the bad, and the ugly. Sqreen blog, 2016. Available at https://blog.sqreen.io/tor-the-good-the-bad-and-the-ugly/.

[6] Akamai. Akamai's [state of the internet] / security, Q2 2015 report. https://www.stateoftheinternet.com/downloads/pdfs/2015-cloud-security-report-q2.pdf.

[7] Ben Herzberg. Is TOR/I2P traffic bad for your site? Security BSides London 2017. Available at https://www.youtube.com/watch?v=ykqN36hCsoA.

[8] IBM. IBM X-Force Threat Intelligence Quarterly,3Q 2015. IBM website. Available at http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=WGL03086USEN.

[9] Aviv Ron, Alexandra Shulman-Peleg, and Emanuel Bronshtein. No sql, no injection? examining nosql security. *CoRR*, abs/1506.04082, 2015.

[10] Mike Perry. The Trouble with CloudFlare. https://blog.torproject.org/blog/trouble-cloudflare.

[11] Michael Carl Tschantz, Sadia Afroz, Vern Paxson, et al. SoK: Towards Grounding Censorship Circumvention in Empiricism. In *Security and Privacy (SP), 2016 IEEE Symposium on*, pages 914–933. IEEE, 2016.

[12] Roya Ensafi, David Fifield, Philipp Winter, Nick Feamster, Nicholas Weaver, and Vern Paxson. Examining how the Great Firewall discovers hidden circumvention servers. In *Internet Measurement Conference*. ACM, 2015.

[13] Philipp Winter and Stefan Lindskog. How the Great Firewall of China is blocking Tor. In *Free and Open Communications on the Internet*, Bellevue, WA, USA, 2012. USENIX.

[14] Roya Ensafi, Philipp Winter, Abdullah Mueen, and Jedidiah R. Crandall. Analyzing the Great Firewall of China over space and time. *Privacy Enhancing Technologies*, 1(1), 2015.

[15] Abdelberi Chaabane, Terence Chen, Mathieu Cunche, Emiliano De Cristofaro, Arik Friedman, and Mohamed Ali Kaafar. Censorship in the wild: Analyzing Internet filtering in Syria. In *Internet Measurement Conference*. ACM, 2014.

[16] David Fifield and Lynn Tsai. Censors delay in blocking circumvention proxies. In *6th USENIX Workshop on Free and Open Communications on the Internet (FOCI 16)*. USENIX Association, 2016.

[17] The Tor Project. Is there a list of default exit ports? Tor FAQ. Accessed Feb. 14, 2017. https://www.torproject.org/docs/faq.html.en#DefaultExitPorts.

[18] Contributors to the Tor Project. Reducedexitpolicy. Tor Wiki, 2016. Version 33 (May 8, 2016). https://trac.torproject.org/projects/tor/wiki/doc/ReducedExitPolicy?version=33.

[19] Details for: apx1. Atlas. Available at https://atlas.torproject.org/#details/51377C496818552E263583A44C796DF3FB0BC71B.

[20] Details for: apx2. Atlas. Available at https://atlas.torproject.org/#details/A6B0521C4C1FB91FB66398AAD523AD773E82E77E.

[21] Details for: apx3. Atlas. Available at https://atlas.torproject.org/#details/38A42B8D7C0E6346F4A4821617740AEE86EA885B.

[22] Torland1 history. Exonerator. Available at https://exonerator.torproject.org/?ip=37.130.227.133&timestamp=2017-01-01&lang=en.

[23] Details for: jahjah. Atlas. Available at https://atlas.torproject.org/#details/2B72D043164D5036BC1087613830E2ED5C60695A.

[24] Icecat: The open catelog. http://icecat.us/. Available at http://icecat.us/.

[25] Mick. [tor-relays] what to do about icecat.biz abuse complaints? https://lists.torproject.org/pipermail/tor-relays. Available at https://lists.torproject.org/pipermail/tor-relays/2012-April/001273.html.

[26] Ofer Gayer. Semalt hijacks hundreds of thousands of computers to launch a referrer spam campaign. https://www.incapsula.com/blog. Available at https://www.incapsula.com/blog/semalt-botnet-spam.html.

[27] Tor Project: Anonimity Online. Tor Metrics: Users. Available at https://metrics.torproject.org/userstats-relay-country.html.

[28] Facebook Threat Exchange. https://developers.facebook.com/products/threat-exchange.

[29] The Tor Project. Design for a Tor DNS-based exit list. Design document. https://gitweb.torproject.org/tordnsel.git/tree/doc/torel-design.txt.

[30] VPN Gate Academic Experiment Project at National University of Tsukuba, Japan. VPN Gate: Public VPN Relay Servers. http://www.vpngate.net/en/.

[31] Privax, Ltd. Free Proxy List – Public Proxy Servers (IP PORT) – Hide My Ass! http://proxylist.hidemyass.com.

[32] The Internet Archive. Internet Archive: Wayback Machine. https://archive.org/web/.

[33] Evan Klinger and David Starkweather. phash– the open source perceptual hash library. *pHash. Ανακτήθηκε*, 14(6), 2012.

[34] McAfee. Customer URL ticketing system. www.trustedsource.org/en/feedback/url?action=checklist.

[35] Distil Networks. Help Center: Third Party Plugins That Block JavaScript. https://help.distilnetworks.com/hc/en-us/articles/212154438-Third-Party-Browser-Plugins-That-Block-JavaScript.

[36] Tariq Elahi, George Danezis, and Ian Goldberg. Privex: Private collection of traffic statistics for anonymous communication networks. In *Proceedings of the 2014 ACM SIGSAC Conference on*

Computer and Communications Security, CCS '14, pages 1068–1079, New York, NY, USA, 2014. ACM.

[37] Akshaya Mani and Micah Sherr. Historε: Differentially Private and Robust Statistics Collection for Tor. In *Network and Distributed System Security Symposium (NDSS)*, February 2017.

[38] Rob Jansen and Aaron Johnson. Safely measuring tor. In *Proceedings of the 23rd ACM Conference on Computer and Communications Security (CCS '16)*. ACM, 2016.

[39] George Tankersley, Filippo Valsorda, and Alex Davidson. Solving the Cloudflare CAPTCHA (RWC2017). Real-World Crypto (RWC 2017). `https://speakerdeck.com/gtank/solving-the-cloudflare-captcha-rwc2017`.

Appendix

A Picking threshold values for *proactive* blacklisting

We classify a blacklist as *proactive* if it enlists a large fraction of Tor exit relays within the first 24 hours of them appearing in the consensus. In order to decide the threshold for the fraction of Tor exit relays that, if blocked within 24 hours, we should consider the blacklist, we analyze the value of the fraction for all blacklists. We find two blacklists: *Snort IP* and *Paid Aggregator* blacklist 37% and 31% of all Tor exits within 24 hours, respectively. All other blacklists listed fewer than 5% of Tor IP addresses within 24 hours. This large difference in the behaviour of blacklists encouraged us to pick the threshold as 30%.

B Classifying exit policies

In this section we describe our method for classifying the exit policies of all exit relays observed in 2015 and 2016. Since each relay could potentially have an arbitrary set of ports open (from the 65,535 possible ports), comparing the *openness* of exit policies is difficult. To simplify the process, we parse the exit policy of each relay to extract the set of open ports and then compute the Jaccard similarity between the relay's open ports and each of the well-known exit policies that Tor supports (Default, Reduced, Reduced-Reduced, Lightweight and Web). We classify a relay into one of the 5 categories based on the Jaccard similarity value. To ensure that the similarity in policy is large enough, we classify the relay to the category of highest similarity, provided that the similarity value is at least 0.7. Only the relays with a high enough similarity value with any of the well known exit policies are considered for further analysis.

C IP blacklisting and relay characteristics

We train a linear regression model to find the impact of relay characteristics like uptime, policy, and consensus weight on the time a relay spends on *reactive* blacklists. The observed variable is the ratio of hours spent on the blacklist to the uptime of the relay. We trained the model on 20,500 exit relays' data (with feature scaling) and found that the coefficients learned for all the factors are extremely small (consensus weight = -0.00007, uptime = 0.009, policy = -0.00001). This shows that these factors have very little impact on blacklisting of relays. It also suggests that changing to more conservative exit policies does not reduce the chances of relays getting blacklisted.