Internet Censorship in Iran: A First Look

Simurgh Aryan
Aryan Censorship Project
aryan.censorship.project@gmail.com

Homa Aryan
Aryan Censorship Project
aryan.censorship.project@gmail.com

J. Alex Halderman
University of Michigan
jhalderm@umich.edu

Abstract
The Iranian government operates one of the largest and most sophisticated Internet censorship regimes in the world, but the mechanisms it employs have received little research attention, primarily due to lack of access to network connections within the country and personal risks to Iranian citizens who take part. In this paper, we examine the status of Internet censorship in Iran based on network measurements conducted from a major Iranian ISP during the lead up to the June 2013 presidential election. We measure the scope of the censorship by probing Alexa’s top 500 websites in 18 different categories. We investigate the technical mechanisms used for HTTP Host–based blocking, keyword filtering, DNS hijacking, and protocol-based throttling. Finally, we map the network topology of the censorship infrastructure and find evidence that it relies heavily on centralized equipment, a property that might be fruitfully exploited by next generation approaches to censorship circumvention.

1 Introduction

Iran is known as one of the leading suppressors of Internet freedom. Reporters Without Borders ranks Iran as one of the “twelve enemies of the Internet” [33], and Freedom House has dubbed it the “least free” country in terms of Internet freedom [23]. Iran’s Internet censorship goes beyond simply blocking access to particular websites and services. Some conservative voices have called for the creation of a fully separate “Halal Internet,” which would contain only content allowed by their strict interpretation of Islamic law [35]. Although the government’s stated policies fall short of this extreme view [1], it recently created a Cyber Police unit, FATA [19], which monitors Iranians’ online activities and prosecutes dissidents [10]. High ranking officials have actively encouraged adoption of domestic sites for applications like blogging, email, and social media [31, 39], while censorship and connection throttling discourage use of similar services hosted abroad, which are more difficult to police.

While these developments have been widely reported, little research has been conducted on the technology and network topology behind Iran’s Internet censorship regime. Probing the network from within the country is dangerous, due to a climate of heavy government control and personal risks to Iranian citizens who take part. Despite these risks, this study seeks to narrow the gaps in our knowledge by providing a firsthand view of how Internet access is being restricted in Iran.

To conduct our study, we established a small testbed in Iran from which to perform network measurements. Our primary aim was to understand the mechanisms used for filtering in the country. To accomplish this, we analyzed traffic to blocked and non-blocked hosts at the packet level, and we used traceroutes to study hops inside the country’s infrastructure. We conducted our primary measurements in the two months leading up to the June 2013 presidential election, when we expected censorship mechanisms to be aggressively deployed [4, 11]. Our results expose details of how traffic is being monitored and modified and provide an initial understanding of the censor’s capabilities and limitations.

Although our study provides an initial technical perspective into mechanisms of censorship in Iran, it cannot be the final word on Internet freedom in the country. We probed the network from only one node at one ISP; while we observed that content blocking occurred exclusively at a centralized location in the national network, we cannot conclude that other ISPs do not apply additional layers of distributed filtering. Furthermore, our study spanned a period of only two months; Iran is known to adjust their censorship mechanisms frequently, and ongoing measurements are necessary in order to understand these changes over time. We hope to continue our probing and add additional collection points in order to present a broader perspective in future work.

*Pseudonymous authors.
Figure 1: Requests for censored sites are redirected to this page, located at http://10.10.34.34, which explains: “Access to the requested website is not possible. For complaints click here.” After a 30 second delay, the user is forwarded to another censorship website, peyvandha.ir.

2 Background

The administrative hierarchy of Internet censorship in Iran is complex and includes many players. In March 2012, the supreme leader of Iran issued a directive establishing a new centralized agency responsible for managing the country’s cyber policies known as the **Supreme Council of Cyberspace** [12]. This council controls three government bodies that are associated with censorship [20]:

- **The Committee for Determining Offensive Contents**, located at internet.ir and peyvandha.ir, which controls censorship policies in Iran. This committee is responsible for maintaining and updating lists of censored websites and also for enforcing internet communication policies.

- **The Iran Cyber Police**, or FATA Police, which is responsible for prosecuting users who are involved in illegal Internet activities as described by the Committee for Determining Offensive Contents.

- **The Revolutionary Guard Cyber Defense Command**, better known as the Iran Cyber Army, which is responsible for defending Iran against cyber attacks and implementing countermeasures.

Every ISP in Iran works under the jurisdiction of the **Communication Regulatory Authority of Iran** (CRA) [8], which enforces the censorship policies put in place by the Committee for Determining Offensive Contents.

Iran’s government has been practicing Internet censorship for more than a decade. The first initiative to limit Internet access was issued by Iran’s supreme leader in January 2002 in an order called the “Comprehensive Proclamation of Computer Information Network Policies” [22]. Initially, individual ISPs used IP address filtering to block access to certain “morally questionable” websites [29]. This system was later gradually replaced with a centralized system run by the state-run Telecommunication Company of Iran (TCI). Under this system, any web request to a blocked site is redirected to a web page owned by the censor, located at the address 10.10.34.34 (see Figure 1). This address, first established in March 2010 [2], is within private network address space as described by RFC 1918 [32] and is only accessible from inside Iran’s national network.

The government has been observed to use a variety of techniques to control Internet access in the country:

**Broadband speed limitations.** Guidelines issued by the CRA [16] limit the bandwidth of home users to 128 kb/s. It is believed that this limitation is imposed to hinder access to multimedia content such as streaming audio and video. Researchers, faculty members, and university students are exempt from this limitation upon providing the appropriate documentation [16].

**DNS redirection.** DNS queries for some sites respond with a fake local IP address (10.10.34.34) that acts as a black hole.

**HTTP host and keyword filtering.** Authorities block access to certain prohibited sites by manipulating connections based on the HTTP Host header. Access to URLs containing certain keywords is also blocked. The list of prohibited keywords originally contained terms frequently used to access adult content, but it has been expanded in recent years in reaction to events causing political and economic turmoil, including presidential elections [40].

**Connection throttling.** In addition to these techniques, Iran has been observed to deploy connection throttling, particularly during times of political and economic unrest [40]. This has sometimes taken the form of throttling speeds to specific sites or protocols and sometimes complete throttling of all traffic [4]. Following the events of the presidential election in 2009, connection speeds to webmail services such as Gmail were reported to be significantly hindered [27]. Certain protocols, including HTTPS, SSH and VPN tunnels, have also been reported to be blocked or throttled at times [25].

It is believed that the Iranian government also has the capability to conduct **SSL man-in-the-middle attacks**. In 2011, an attacker who claimed to be Iranian compromised the DigiNotar certificate authority [30] and created hundreds of fake certificates for websites including google.com [30]. These certificates are reported to have
been used to initiate MITM attacks against more than 300,000 Internet users, almost all of them in Iran [30].

The most obscure aspect of Iran’s censorship program is the network devices that allow traffic manipulation in such ways. It has been suggested that Iran relies on deep packet inspection (DPI) to selectively monitor and modify traffic [5, 26], but the locations, manufacturers, and full capabilities of these systems remain uncertain.

Iranian users have in turn used various types of anti-censorship tools to gain free access to the web [6]. Some of the widely used tools in Iran include Tor [41], Green Simurgh [15], FreeGate [13], Your Freedom [48], and Ultrasurf [43], as well as VPNs and SSH tunnels. In response, the censors have entered a cat-and-mouse game to block users’ access to them inside Iran. Many of the servers associated with these services have been blacklisted, and protocols associated with them have been blocked or throttled [25]. The Tor Project has detected DPI-based blocking being used against its service in multiple instances [28]. A malicious version of Green Simurgh infected with tracking software that records user activities has been reported in the wild [45].

3 Related Work

Most of the technical literature on Internet censorship has been focused on the Great Firewall of China (GFC). By necessity, we can give only a very brief survey here. Clayton et al. [7] address keyword filtering in the GFC, and Crandall et al. [9] claim that, contrary to previous belief, censorship in China does not occur only at the borders. Xu et al. [47] further explore the AS-level topology of China’s network and manage to uncover many local firewall nodes inside China’s infrastructure.

Other recent studies examining network topology and censorship mechanisms across many countries include discussion of Iran. In 2011, Roberts et al. mapped the autonomous systems (AS) of several different countries [36]. Their findings suggest that Internet traffic in Iran passes through a single “point of control.” The following year, Verkamp and Gupta [44] looked at the mechanics of censorship in 11 countries around the world, including Iran. They conclude that censorship in Iran is hostname based and results in a 403 response followed by a redirection. Our work lends further evidence to support these findings.

Over the past year, Anderson performed the first technical studies dedicated to Internet censorship issues in Iran. In September 2012, he examined the use of private IP address space inside Iran’s national network and found a large number of hosts in this space [3]. While our work was under review, Anderson released another report that looked at politically motivated Internet throttling in Iran [4]. He observed prolonged and significant disruptions in quality of service on dates associated with political or economic unrest. Our work complements these findings and is (to the best of our knowledge) the first peer-reviewed technical publication focused on Iranian Internet censorship.

In a less technical vein, Reporters Without Borders has addressed Internet censorship in Iran [33], as have investigative journalists [24, 34], who cite data about devices that have facilitated DPI by Iran’s government. The Iran Media Research program [21] and Small Media [37] have each issued reports on media censorship in Iran, emphasizing its social and humanitarian impacts.

4 Experiments and Results

The experiments we report below were conducted in April and May 2013, during the lead-up to the Iranian presidential election on June 14. We used a machine located inside Iran connected to one of the country’s major ISPs. This

Figure 2: Effects of Iranian Internet censorship on the top 500 websites for 18 Alexa categories.
machine ran Ubuntu 12.10 (Quantal Quetzal). We refer to it by the codename Aryan; we have obfuscated any identifiable characteristics in our work to protect parties involved with this research inside Iran. In some of our experiments, we also used a second server outside Iran running Ubuntu 12.04 LTS (Precise Pangolin), Apache 2.2.22, OpenSSH 5.9p1, and BIND9.8.1p1. This host, which we refer to by the codename Bob, allowed us to monitor packets on both ends of the connection.

4.1 The Scope of Censorship

To evaluate the extent of censorship inside Iran, we surveyed the most-visited websites based on Alexa web traffic rankings. We created a crawler that retrieved the top 500 websites in each of 18 different Alexa categories and initiated a GET request from Aryan to these websites. The results of this experiment are illustrated in Figure 2. The most censored category was, predictably, Adult websites, where more than 95% of sites were blocked. The most alarming was the overall Top 500 category, where more than 50% of the Internet’s most visited websites were censored. Surprisingly, the Art category was the third most censored, followed by Society and News.

A breakdown of our results by blocking mechanism is presented in Table 1. DNS hijacking was observed only on domains associated with facebook.com, youtube.com, and plus.google.com. A number of website connections also timed out repeatedly when accessed from Iran. These sites generally belonged to sectors that are subject to U.S. government sanctions, such as banking (e.g., bankofamerica.com) and technology (e.g., nvidia.com), and some of them appear to block access from IP addresses in Iran.

We note that our study fails to cover many popular blocked Persian websites. These sites are not represented in Alexa’s rankings for Iran as they are solely accessed using anticensorship software from inside the country and do not appear in other lists as they have little worldwide audience other than the small number of Persian speakers.

4.2 HTTP Host and Keyword Filtering

To probe the mechanism behind Iranian host-based filtering, we initiated HTTP requests from Aryan to blocked websites and examined the resulting network traffic. Figure 3 shows the typical network interactions related to this experiment. The three-way TCP connection handshake happens successfully between Aryan and the blocked website. However, when the GET request is sent from Aryan, it receives a packet containing an HTTP “403 Forbidden” error and an <iframe> containing the censorship page shown in Figure 1. We determined that this filtering is triggered by the Host header in HTTP requests. Omitting the header will circumvent host-based filtering but frequently results in a “400 Bad Request” response from the server. This behavior is consistent with earlier observations by Verkamp and Gupta [44] but differs from IP-based censorship practiced in China, which typically involves blackholing connections to the IP addresses of banned servers at the routing level [14].

To examine how keyword filtering occurs, we created an empty HTML page named “sex.htm” on our remote server Bob. While an innocent page can be fetched successfully from Bob, visitors to this page are faced with a response similar to the host filtering case. When Aryan initiates an HTTP connection to fetch this page, it can complete the TCP handshake successfully with Bob, but upon sending the GET request, it receives an HTTP “403 Forbidden” response. Simultaneously, Bob receives 5 RST packets spoofed to appear to be coming from Aryan and closes its end of the connection. In our experiments, three of these packets had identical sequence number consistent with the TCP stream, while the other two had identical but seemingly random offsets from the previous three RST packet. This keyword filtering seemed to be limited to the HTTP request URI and did not cover POST data or the HTTP response body.

One way the censor might implement the filtering we observed is using a transparent HTTP proxy. To test this, we initiated a series of connections from Aryan to our external server Bob and compared the packets at both end-

<table>
<thead>
<tr>
<th>Category</th>
<th>OK</th>
<th>Host</th>
<th>DNS</th>
<th>T/O</th>
<th>OK %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adult</td>
<td>23</td>
<td>473</td>
<td>0</td>
<td>4</td>
<td>4.6</td>
</tr>
<tr>
<td>Top 500</td>
<td>258</td>
<td>227</td>
<td>3</td>
<td>12</td>
<td>51.6</td>
</tr>
<tr>
<td>Art</td>
<td>261</td>
<td>230</td>
<td>4</td>
<td>5</td>
<td>52.2</td>
</tr>
<tr>
<td>Society</td>
<td>300</td>
<td>190</td>
<td>0</td>
<td>10</td>
<td>60.0</td>
</tr>
<tr>
<td>News</td>
<td>358</td>
<td>140</td>
<td>0</td>
<td>2</td>
<td>71.6</td>
</tr>
<tr>
<td>Regional</td>
<td>365</td>
<td>120</td>
<td>1</td>
<td>14</td>
<td>73.0</td>
</tr>
<tr>
<td>Computers</td>
<td>393</td>
<td>97</td>
<td>3</td>
<td>7</td>
<td>78.6</td>
</tr>
<tr>
<td>Games</td>
<td>404</td>
<td>85</td>
<td>1</td>
<td>10</td>
<td>80.8</td>
</tr>
<tr>
<td>Shopping</td>
<td>407</td>
<td>86</td>
<td>0</td>
<td>7</td>
<td>81.4</td>
</tr>
<tr>
<td>Sports</td>
<td>407</td>
<td>91</td>
<td>0</td>
<td>2</td>
<td>81.4</td>
</tr>
<tr>
<td>Kids &amp; Teens</td>
<td>407</td>
<td>85</td>
<td>0</td>
<td>8</td>
<td>81.4</td>
</tr>
<tr>
<td>Business</td>
<td>408</td>
<td>68</td>
<td>0</td>
<td>24</td>
<td>81.6</td>
</tr>
<tr>
<td>Recreation</td>
<td>426</td>
<td>67</td>
<td>0</td>
<td>7</td>
<td>85.2</td>
</tr>
<tr>
<td>Home</td>
<td>448</td>
<td>42</td>
<td>0</td>
<td>10</td>
<td>89.6</td>
</tr>
<tr>
<td>Health</td>
<td>449</td>
<td>44</td>
<td>0</td>
<td>7</td>
<td>89.8</td>
</tr>
<tr>
<td>Iran</td>
<td>468</td>
<td>19</td>
<td>2</td>
<td>11</td>
<td>93.6</td>
</tr>
<tr>
<td>Science</td>
<td>469</td>
<td>26</td>
<td>0</td>
<td>5</td>
<td>93.8</td>
</tr>
<tr>
<td>Reference</td>
<td>475</td>
<td>23</td>
<td>0</td>
<td>2</td>
<td>95.0</td>
</tr>
</tbody>
</table>

Table 1: Breakdown of top-500 websites’ reachability in different Alexa categories. The Host column represent websites censored by means of HTTP Host filtering; DNS represents websites censored by DNS hijacking; and T/O represents sites that did not respond to our requests.
points. We did not observe any anomalous changes in the TCP/IP headers, HTTP headers, or payloads that would indicate the presence of an intercepting proxy between the hosts, nor any suspicious patterns in connection timing. This suggests that a transparent proxy is not in use at the ISP we studied, and that the state-run filtering mechanism is based on another technology, such as DPI. However, other individual ISPs may have implemented their own transparent proxies to cache data or enforce more strict censorship policies [38].

4.3 DNS Hijacking

In our examination of Alexa top 500 websites, we encountered three domains (facebook.com, youtube.com, and plus.google.com) for which DNS responses directed the client to the censorship page IP address (10.10.34.34) instead of a valid IP address for the site.

To determine more precisely how these requests were being blocked, we set up our own DNS server and initiated DNS requests from Aryan to it. Our observations suggested that blocked DNS queries never made it to our DNS server and were intercepted on path. In their place, our DNS server received 5 TCP RST packets spoofed from Aryan’s address. In all of our experiments, three of these RST packets had an identical random sequence number while the other two had relative sequence numbers of 30 compared to the first three. This is a particularly curious result, since the original DNS queries were UDP packets and RST packets belong to TCP connections, and it may indicate that the censorship system is misconfigured. We also sent TCP DNS packets and observed no censorship on any domains. Figure 4 shows the interaction between Aryan and our DNS server.

4.4 Connection Throttling

To analyze the methods used to throttle connection speeds, we first measured the speed of file transfers from Bob to Aryan using the HTTP, HTTPS and SSH protocols. We used a file size that should have taken Aryan 96 seconds to download using its full bandwidth. We repeated this experiment multiple times for each protocol.

For HTTP and HTTPS file transfers, Aryan used 85% and 89% of its total bandwidth on average, respectively. In contrast, for SSH file transfers, only 15% was utilized on average. All of our measurements were within 5% of these averages.

To confirm that this decrease resulted from the censor’s interference, we proceeded to obfuscate our SSH file transfer (and therefore the unencrypted portion of its handshake) by XORing packet payloads with a predefined constant key. In this way, we expected to circumvent the censor’s efforts to detect and throttle our SSH tunnel. Surprisingly, all of the trials using this modified SSH tunnel exhibited even worse performance. The obfuscated connection was constantly throttled to the point that download speed dropped to near zero at around 60 seconds into the connection. This resulted in incomplete file transfers during all of our trials.

From this observation, we hypothesize that instead of blacklisting undesired protocols, the censorship system was configured to whitelist approved protocols. This approach would allow the censor to preemptively block new, unrecognized circumvention techniques. It would also work to block Tor’s obsproxy protocol [42], which seeks to obscure identifiable features of the transport protocol. To test our theory, we used the obfuscation technique described above on HTTP file transfers. These experiments yielded results similar to the obfuscated SSH tunnel, supporting our protocol whitelisting hypothesis.
Our packet traces suggest that the throttling we observed was accomplished by dropping packets and causing TCP back-offs. Understanding the exact pattern and interval of these packet drops requires further investigation that we leave for future work.

Iran is known to vary its application of censorship technologies in connection with political and socioeconomic events, and the government has admitted throttling the Internet ahead of the June 2013 election [11]. We repeated our tests shortly after the election and found that connection throttling had been lifted from Aryan. SSH, obfuscated SSH, and obfuscated HTTP connections appeared to perform equally well, using an average of 82% of Aryan’s bandwidth with a standard deviation of 5.5%.

4.5 Topology

To explore Iran’s network topology, we first used ICMP traceroutes from Aryan to detect intermediate routers inside the country. We randomly chose 3160 destination IP addresses inside 13 neighboring countries using the IP address country block tool provided at IPInfoDB [18]. In our experiment, all of the observable first hops outside of the Iranian network were preceded by a node located at the private IP address 10.10._._, which we hypothesized was a device used for censorship. In each case, this node was preceded by one of two nodes owned by Telecommunication Company of Iran (TCI), which are the two paths used by Aryan’s ISP to connect to the outside world. (We have omitted or partially redacted these addresses to prevent possible deanonymization of our ISP.) Figure 5 displays Aryan’s view of Iran network topology.

Next we conducted an experiment to determine where along the path censorship functions were occurring, using a similar methodology to Xu et al. [47]. Aryan established a connection with a blocked website and sent GET requests with different IP TTL values. Depending on the response received by Aryan, we can determine if censorship is happening at this node or further along the path.

We established in Section 4.2 that the TCP handshake is not blocked by the censor, so we varied the TTL of the packet containing the HTTP GET request. When the TTL for our GET request expired before the suspected censorship node, we received an ICMP error. By incrementing the TTL from zero and recording the first “403 Forbidden” response, we confirmed that the suspected censorship node was indeed responsible for blocking our request. Similar experiments with DNS queries indicated that the same node was blocking them. We attempted to learn more about this censorship node by probing it with nmap, but it was unresponsive.

Additionally, we used Bob to traceroute to Aryan and to other IP addresses inside Iran. From Bob's point of view, the route to Aryan was the same as Aryan’s route to Bob, except that Bob did not receive any response from the censorship node. Traceroutes to other IPs inside Iran similarly did not reveal the first hops inside Iran. Nevertheless, traceroutes to Aryan and to other Iranian IPs shared some of their Iran-based hops (after the missing hops), suggesting that a large amount of Iran’s traffic passes through a centralized facility.

5 Future Prospects

Our study provides an initial technical perspective on the mechanisms behind Iranian Internet censorship, but a more complete understanding will require probing Iran’s network infrastructure from multiple nodes inside the country, and monitoring the status of Internet access in Iran on an ongoing basis. Although we observed that at least some of Iran’s censorship infrastructure is centralized, individual ISPs may be using additional mechanisms. Obtaining access to hosts located at many ISPs is necessary to allow researchers to explore such variations.

Similarly, even with the limited duration of our study, we were able to observe changes to connection throttling behavior before and after the June 2013 presidential election. Under special circumstances, Iran’s government expands the restrictions imposed on users, including blocking access to the SSH protocol and webmail services entirely and performing SSL man-in-the-middle attacks. A longer study could lead to an understanding of the motivations and technical strategy behind such changes.

Based on Iran’s history of aggressive Internet censorship, we consider it likely that the government will continue its efforts to monitor users and block anticensorship
techniques. Future work on next-generation anticensorship tools, such as Telex [46] and Cirripede [17], will be vital for achieving free Internet access among Iranian users in the face of increasingly sophisticated censorship mechanisms employed by the government.

6 Conclusion

In this work, we examined the status of Internet censorship in Iran. We studied how and where in the network censorship happens, and we measured the scope of the censorship by analyzing results for 18 categories of popular websites. We believe our results contribute to the efforts to understand how censorship is conducted in Iran and will be useful in the development of more robust countermeasures.

Of particular interest for censorship resistance is the centralized nature of the censorship mechanisms we observed. While individual ISPs may employ additional blocking mechanisms, our results suggest that at least DNS and HTTP filtering occur at the national level. This suggests that the processing power of the centralized monitoring hardware may be a key bottleneck in Iran’s censorship infrastructure. New censorship resistance systems could explore techniques for overwhelming the central monitoring hardware with spoofed traffic, for instance, or for tunneling data past it and then further distributing it in a peer-to-peer manner within the country. We hope future work will build on our results to probe Iran’s censorship infrastructure more deeply and to develop anticensorship mechanisms that maximally exploit its limitations.

Acknowledgments

The authors thank Eric Wustrow for discussions and advice on many aspects of this work. We also thank Collin Anderson and the anonymous reviewers for their insightful suggestions and comments. This work was funded in part by NSF grant CNS-1255153.

References


