Measuring Internet Censorship in Iran from Multiple Vantage Points

Neeki Hushyar
UMass Amherst
nhushyar@cs.umass.edu

Angela Upreti
UMass Amherst
aupreti@cs.umass.edu

ABSTRACT

We analyze Internet censorship in Iran through data collected using ICLab and RIPE Atlas. We use results from HTTP GET requests, TCP connect requests, TLS handshakes and DNS queries to confirm previous findings reported by Halderman et al. in 2013. Halderman et al. used a single VP in Iran whereas we use n different VPs to determine the consistency of the results from different vantage points. Through DNS queries conducted using network measurement platform, RIPE Atlas, we measure the scope of the censorship by probing 60 different domains. Through the HTTP GET requests, TCP connect requests and DNS queries conducted using ICLab, we measure the scope of the censorship by probing more than 500 domains. These domains include a test list for Iran compiled by citizen lab and top websites for Iran reported by the Alexa website. We compare the results obtained through Iranian VPs with the results obtained from a VP located in the U.S. We conclude with the results confirming the blocked webpages either by way of HTTP filtering or DNS hijacking as previously reported by Halderman. In addition, we find that the censorship encountered at varying vantage points is consistent. In the process, we also discover some weaknesses of the ICLab platform and offer some recommendations.

CCS Concepts

•**Networks** → *Network measurement; Firewalls;* Deep packet inspection;

Keywords

Censorship in Iran; ICLab; RIPE Atlas

1. INTRODUCTION

Freedom House, the U.S. based, government-funded, nonpartisan organization which researches the levels of democracy, political freedom and human rights in countries all over the world, ranked Iran as one of the worst countries in terms of free internet, second only to China. The government of Iran blocks social media and political and social content from a wide range of categories. Over the past few years users of blocked domains in Iran have faced persecution and circumvention tools, which assisted civilians in accessing blocked domains, have been blocked through a variety of methods. The affects of internet censorship and lack of net freedom have been widely publicized. Notable cases include a complete shutdown in access to social media beginning during the 2012 presidential election campaigns in Iran, the 2014 arrests of Facebook activists who had used the social media website to motivate government reform, and unrelated arrests that same year of individuals who insulted the Supreme Leader, Imam Khomeni, over a text message application.

WOODSTOCK '97 El Paso, Texas USA

DOI: 10.475/123_4

2. BACKGROUND

Over the past 15 years, the infrastructure of the Internet in Iran has been constructed in such way to give increasing amounts of control to the government. First, Internet exchange points (IXP) have been have been sponsored throughout the country. Internal IXPs increase connectivity and reliability, but also to decrease reliance on international infrastructure to maintain Internet communication. Furthermore, any Internet traffic exiting the country must cross one of few links, all of which are operated by the government, in order to interact with the global network outside of Iran. This single point of control gives the state operators additional control over the traffic allowed outside of the country. Whether the traffic is traversing international boundaries or is exclusively remaining within the country, many method are utilized to censor the content.

The government uses an array of mechanisms to control communications. These methods, detailed in the next section, include throttling connections based on protocols, filtering traffic based on HTTP headers, filtering traffic based on payload content and hijacking DNS requests. Lastly, Internet service providers are given monetary incentives to limit the connections they facilitate to only local domains. Government representatives have stated that the network has been constructed in such way to allow authorities to police Internet traffic in order to protect "public morality" and discourage the "dissemination" of lies. It must also be considered that the network has been constructed in such way to allow the complete shutdown of connections to the global community while maintaining Internet communications within Iran.

3. RELATED WORKS

The previous experiments were coducted in 2012. These experiments pinpointed the methods used by the government to implement censorship as well as locating the sources of censorship. These experiments revealed many methods used by ISP and DNS resolvers in Iran in the effort to censor Internet users. Those methods include:

Redirection: The return of false, redirected pages in response to DNS requests for censored domain names. The previous experiments reported this redirect IP as 10.10.34.34.
HTTP Filtering: Access to specific websites blocked, based either on the content of the GET request itself, or the content in the response packet[s].
Connection Throttling: Iran reportedly throttles the connection of specific connections, namely those running over SSL or virtual private networks implemented using IPSec, among others methods of encryption. Furthermore, attempts to access specific websites, independent of protocol were reported to cause connection throttling.
Broadband Speed Limitations: The maximum bandwidth afforded to home user's is limited. This was previously believed to be enforced as a method of censorship, however more recent reports reveal this may be forced by limited network capabilities.

Our experiment simulates the HTTP filtering and DNS redirection described above in order to determine the consistencies of the filters and redirections. The previous experiment, conducted in 2012, was the first to pursue the technical aspects of the Internet censorship in Iran. Most of the research was done from a single vantage point and our experiment seeks to expand upon them.

4. EXPERIMENTS AND RESULTS

We conducted network measurements in Iran using ICLab and RIPE Atlas.
ICLab is a research platform used to measure global Internet censorship. The ICLab platform consists of Raspberry Pis and VPNs as vantage points(VPs) located all around the word. This platform allows researchers to run self-written experiments on the chosen VPs. ICLab was used to conduct tests targeting over 500 domains. The domains tested fall in a variety of categories, including the top Alexa domains as accessed by people within Iran as well as several hundred additional domains, which were previously reported to be blocked. Figure 1 shows the distribution of our test domains across different categories.
RIPE Atlas is a global network of probes that measure Internet reachability and connectivity. RIPE was used determine several vantage points from which to send DNS queries. Next, we discuss our methods and results for each platform.

4.1 ICLab

ICLab platform was used to conduct DNS, traceroute, TCP connect, TLS handshake and HTTP GET measurements. All except the traceroute experiments were successful.

4.1.1 DNS

DNS queries for each domain on our list were sent from the Iranian VP to two different resolvers. A local resolver of the VP's choosing and the google resolver, 8.8.8.8 were used. A python script was then used to compare the DNS responses from a U.S. VP to the responses obtained in Iran. In Iran, all queries using the google resolver yielded a null response. We hypothesize the DNS packets destined to 8.8.8.8 are being dropped. Some queries using the local resolver also yielded a null response even though a valid IP is obtained for the query sent by the U.S-based VP. On the flip side, a few domains in our test list got null response in the U.S. and a valid IP in Iran. Many domains that got a null response in the U.S. but not in Iran are Persian websites, though not all. To our surprise, we are able to reach some of the websites that give a null DNS response for the centinel tool when we type the domain name directly in the browser. We are uncertain what might be causing this. Figure 2 shows a comparison of DNS responses from a US VP and a Iranian VP. A large portion of the DNS queries resulted in the same response both in Iran and the U.S. As we report later, some of the websites with a valid DNS response are eventually blocked via HTTP filtering.
DNS queries from Iran for forty-eight different urls, most of them belonging to blogspot.com, in our test list elicited a response of 10.10.34.36, a private IP. This makes a major portion of the differenent DNS responses seen between the U.S. and Iran. We later report similar findings about private IPs when using RIPE atlas.

4.1.2 Traceroute

We setup our experiment to conduct traceroute measurements before we handed it over to the person responsible for ICLab. Our intention was to pinpoint the locations of the censors within Iran's network. However, we got "traceroute not found or not installed" error for all of our traceroute measurements. We see this as one of the limitations of the ICLab platform when compared to RIPE Atlas. Researchers do not control the remote machines located in the test countries which means they lack a clear expectations of what kinds of tests will be successful.

4.1.3 TCP connect

TCP connection to majority of the websites in our test list succeed form our vantage points both in the U.S. and in Iran. There were only a handful of cases where TCP connection only succeed in one of the countries or failed in both. Private IPs such as 10.10.34.36 were only returned by Iranian resolvers and TCP connection to private IPs was successful in Iran. Based on our results and as previously studied by Halderman et al., we find that censorship in Iran doesn't occur at the transport layer. It occurs at the application/HTTP layer or via DNS.

4.1.4 TLS handshake

Both the Iranian and the U.S. VPs attempted a TLS handshake with the domains in our test list, regardless of whether the domain name started with an 'https' or not.
Among the domains that were able to attempt the TLS handshake, there was a fifty-fifty split between the number of websites that were able to successfully complete the handshake and those that were not. It should be noted that the majority of the domains in our list come from Citizen lab

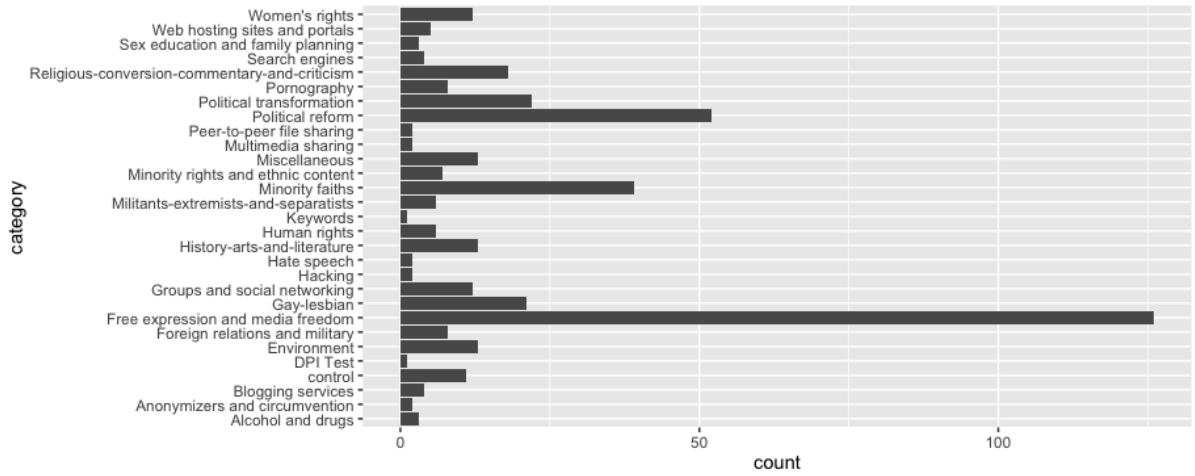

Figure 1: Test Domains by Category

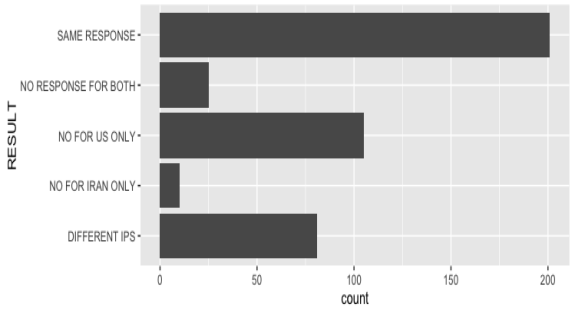

Figure 2: Comparision of DNS responses between a US-based VP and an Iranian VP

test list and have a greater probability of being censored. Table 1 sums up our results from the TLS handshake. Some errors such as 'Connection refused' occurred exclusively in Iran whereas errors such as 'TLSV1_ALERT_INTERNAL_ERROR' occcured with equal frequency in both countries.

4.1.5 HTTP

HTTP GET requests were sent from VPs in both the United States and Iran to use as control and experiment results, respectively. The same test list of more than 500 urls was used. HTTP GET requests to invalid domains were used to verify Deep Packet Inspection(DPI). To verify DPI, paths to Alexa top websites in Iran were modified with a trigger word such as 'porn'. The expectation is that an invalid domain should return a 404 Page Not Found Error.
Based on our measurements, we find that HTTP GET requests is the most prominent mode of censorship in Iran. Many domains with successful TCP handshake gave 4xx response for a GET request. We believe that the TCP connect succeeds because the collateral damage of IP-based blocking has deterred Iran from conducting IP-based blocking. Figure 3 shows this comparison of successful HTTP requests and successful TCP connect requests as seen from a VP in Iran. Our results show a significantly larger number of successful TCP connections compared to a successful HTTP

Table 1: Errors and their number of occurrences for TLS Handshake

Error Type	Iran	U.S.
[Errno 111] Connection refused	36	0
unknown protocol (_ssl.c:590)	19	15
sslv3 alert handshake failure (_ssl.c:590)	4	4
tlsv1 alert internal error (_ssl.c:590)	21	21
EOF occurred in violation of protocol (_ssl.c:590)	2	2
[Errno -2] Name or service not known	1	0
[Errno 110] Connection timed out	93	0
[Errno 104]Connection reset by peer	7	0
[Errno 101]Network is unreachable	21	0
[Errno 113] No route to host	2	0

connection. We consider an HTTP connection to be successful if the status code of the response is either 2xx or 3xx.
While a mix of status codes 200, 400, 3xx, 403 and 404 were obtained from the Iranian VP, the same test list of URLs resulted in a negligible number of failures when run from a U.S-based VP. A handful of error responses gotten for the US VP were usually 404 Page Not Found or 403 Forbidden. 404 error occurred mostly when the domain name was intentionally invalid like in the case of path modification to check for deep packet inspection. Some websites such as http://ladysun.net- a Women's rights website, http://www.alrased.net/site- a minority faith website, http://www.jawanan.org and http://www.jebhemelli.org- two political reform websites and a few other websites elicited a 403 response even from a VP located in the US. A traceroure measurement and geoIP lookup located the destination servers of these sites to be in countries such as Japan and the U.S. The reason for this censorship is unknown to us and would be a good thing to look at in future studies.
We successfully verified the use of DPI by Iran. An invalid domain should return a 404 Page Not Found Error, however, the invalid domains which had a trigger word in the

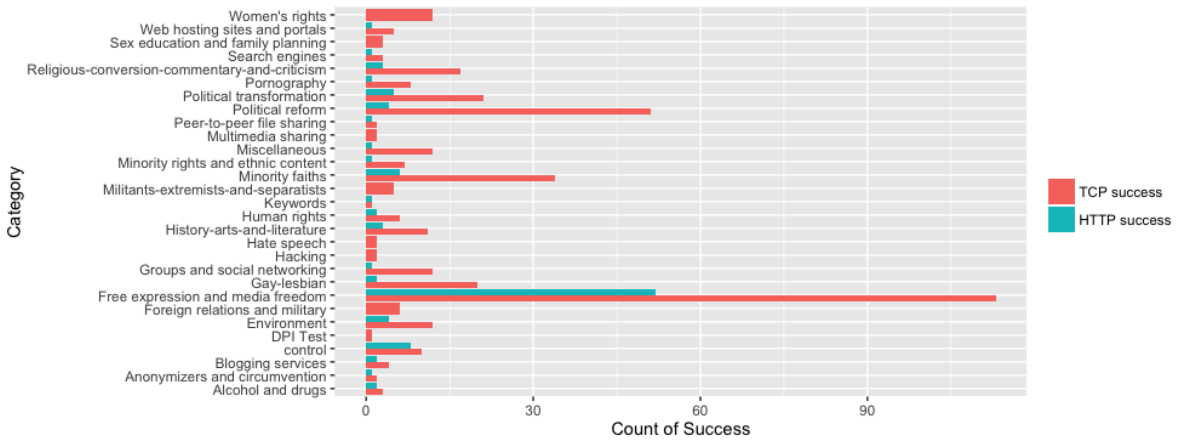

Figure 3: Comparision of successful TCP and successful HTTP requests by category

path resulted in a 403 Forbidden response. Adding '/porn' to a path always resulted in a 403 Forbidden response from Iranian VP and 404 Not Found response from a U.S-based VP. Some of our path modifications such as adding '/adult' gave a 404 response in Iran instead of the expected 403 response. This might be because 'adult' is not a trigger word in Iran. Our choice of possible trigger words was obtained through online research.

4.2 RIPE Atlas

4.2.1 DNS:

As reported in the previous study, the method used to conduct DNS hijacking is false resolutions. In these cases, a request to resolve a domain name returns an incorrect IP address which would redirect the user to an incorrect webpage. We encountered several domains for which the DNS resolutions returned a false redirect address, 10.10.34.36, instead of the domainâĂŹs correct IP address.

RIPE Atlas allowed us to set specific parameters for DNS requests. The parameters we specified in our testing include source probe and DNS resolver. First, we used a total of three different source probes from which our DNS requests were sent. Parties affiliated with RIPE Atlas operate these probes. Second, we specified the IP addresses of specific resolvers in a variety of locations within Iran. To select specific resolvers, we considered three characteristics. First, we narrowed our search to resolvers with a high reliability. All resolvers chosen have 78% or higher reliability according to the Public DNS Server List. Second we considered the location of the resolvers. Location was considered in order to make a claim about the consistency, of DNS resolvers across Iran. In total, the experiments used a series of three source probes and five DNS resolvers.

Not only did we want to make a claim regarding the consistency of the resolvers, but also one regarding the consistency probes. To do this we had two classes of test cases. One class requested DNS resolutions from a single source probe, while rotating the resolvers. A second class requested DNS resolutions by rotating source probes, and targeting the DNS resolution requests at a single resolver. For both test cases, a multitude of domains were used as the subject of the DNS requests. These domains were pulled from the top 40 most often accessed websites in Iran from the Alexa website as well as 20 additional domains previously reported to be targets of DNS hijacking.

Our results revealed that the source probe from which the DNS query was sent from had no effect on the accuracy of the result from any one resolver. That is to say, given two separate source probes, resolving the same domain name at the same resolver, always resulted in the same response. This suggests that all of the DNS requests were sent as is, and never tampered with on the way to the intended resolver. On the other hand, we determined that the resolvers were not centralized in terms of where they retrieved their rules from. In other words, the different resolvers could be inconsistent. We saw several cases of some resolvers returning false resolutions for the same domains that other resolvers returned accurate IP addresses for. As a result of this inconsistency, we compiled three categories in which our tested domains fall in.

Always Resolves: The domain names in this category were all accurately resolved. The resolutions were verified by scripts that ran the whois command to confirm the organization name matched the intended domainâĂŹs organization name. The initial tests executed 60 domain name resolution requests from three probes to three targeted resolvers. Thus each domain was tested from three probes, targeted at three resolvers, for a total of 9 resolutions per domain. A domain was assigned to this category if all of the resolutions were accurate. Domains in this category include many local news and government websites among others.

Sometimes Resolves: Our initial test of 60 domain names over three probes and three target resolvers revealed inconsistencies in resolutions. Specifically, after testing the first three resolvers, located in Tehran (2) and Isfahan, the results revealed that one resolver in Tehran, would consistently return accurate resolutions for domains which returned the 10.10.34.36 redirect pages by the other two resolvers. To determine whether or not this anomalous resolver was one of a kind, or if there exists true diversity in

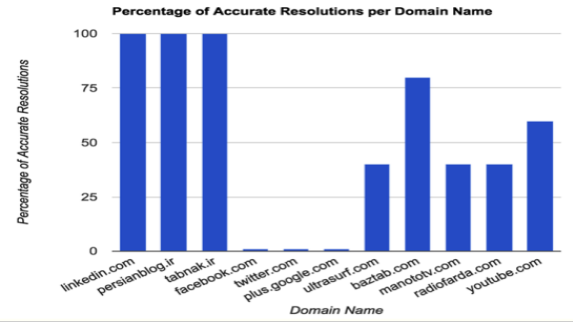

Figure 4: Resolution rate of some queried domains.

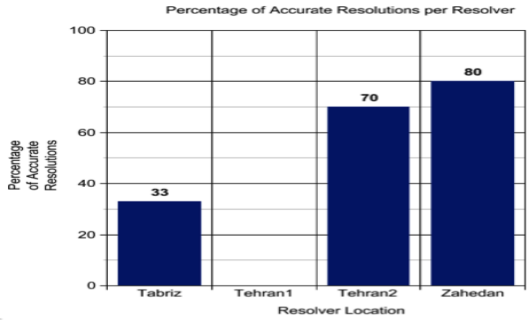

Figure 5: Percentage accuracy per DNS resolver of domains blocked by at least one resolver

replies depending on the chosen DNS resolver, we extended further testing. Our additional tests utilized the same three probes with two additional resolvers in Tabriz and Zehadan. We tested every domain that returned the false redirect address in the previous round of resolutions against these additional resolvers. We found further inconsistencies. The domain names in this category were accurately resolved by some resolvers and falsely resolved by others. This category contained a wide variety of domains whose contents covered topics including: human rights, local blogs, national news and censorship circumvention.

Resolves: The domain names in this category were never accurately resolved. This applies to all five resolvers. The category contained domains from topics including: political reform, gay+lesbian, international news, and western social media. The domain names in this category returned, consistently resolved as either the redirect IP, 10.10.34.36 or did not resolve at all and resulted in a timeout.

5. CONCLUSION

In this work, we examined censorship in Iran using two different platforms: RIPE Atlas and ICLab. We successfully conducted DNS, HTTP, TLS and TCP measurements for a test list of over 500 domains. The results from our measurements are consistent with the findings of Halderman et al. Similar to Halderman et al., We verify deep packet inspection and HTTP host-based filtering via HTTP measurements. We observe polluted DNS responses via both ICLab and RIPE. Using RIPE Atlas, we find several inconsistencies in the DNS response depending on the resolver and the location of the resolver. In addition, we also find a larger number of errors in TLS handshake for the same test list of domains in Iran compared to the U.S.
A larger scale measurement can help us understand the topology of network infrastructure as well as understand methods such as HTTP filtering in a greater detail.

6. FUTURE WORKS

We plan to conduct a larger scale experiment to test a larger set of domains. Besides citizen lab test list, we plan to pull top 500 Alexa websites for Iran pertaining to 18 different categories as done by Halderman et al. This means we will be looking at a test list of more than 10000 domains. An interesting project for the future will be to establish a list of keywords that trigger the DPI censors in Iran. We have a method in mind that uses q binary search and a large test list of domains and sub-domains. This method will be similar to the work presented in an anonymous study of DNS-based censorship in China [1].
Future experiments from a VP capable of traceroute measurements will help us get a better idea of network infrastructure for Iranian censorship. In addition, writing additional experiments for the ICLab tool, centinel, that are capable of taking measurements such as protocol based connection throttling is also of interest to us.

7. ACKNOWLEDGMENTS

We would like to thank Dr. Philipa Gill and Abbas Razagapanah for helping us run our experiments on the ICLab platform. We also want to thank Rachee Singh for her guidance in using RIPE atlas.

8. REFERENCES

[1] *Towards a Comprehensive Picture of the Great Firewall's DNS Censorship Anonymous.*