Project: Disguiser: End-to-End Censorship Measurement

Date: 11/2022

Tuning Fine-Grained Censorship with Disguiser: Assessing the Impact of Path Diversity on the Internet Censorship

A Brief Overview of Disguiser

> Disguiser aims to measure and investigate global censorship activities and its deployment through an end-to-end framework that enables ground truth for automatic and accurate censorship detection.

Internet censorship has been widely witnessed and its severity varies from country to country. Such information control, typically placed by authority entities such as governments, ISPs, or organizations, can be achieved by various techniques such as IP-layer censorship (*e.g.*, blocking IP addresses) and application-layer censorship (*e.g.*, domain names based blocking in DNS, HTTP, and HTTPS).

Typically, in the presence of censorship, the censor devices would first identify the accessed domains (if being censored), and then intercept and terminate the connections to block the censored domains. In the meantime, the censor devices would return a blockpage to the clients, noticing that the accessed content is unavailable. However, automatically and efficiently identifying the blockages remain challenging due to the variations of blockages and lack of ground truth to detect blockages.

Research Team

Shuai Hao is currently an Assistant Professor in the Department of Computer Science at Old Dominion University (ODU), Norfolk, Virginia. Prior to joining ODU, he worked as a Postdoctoral Researcher in the Center for Applied Internet Data Analysis (CAIDA) at the University of California San Diego.

Haining Wang is currently a Professor of Electrical and Computer Engineering at Virginia Tech. He has over twenty years of research experience in cybersecurity. In particular, his works on effective defense against bots in various online services, Denial of Service (DoS) attacks, and network traffic anomaly detection have been widely acknowledged and cited. He is a Fellow of IEEE.

Chase Cotton is currently a Professor of Electrical and Computer Engineering at the University of Delaware. He is a researcher, carrier executive, product manager, consultant, and educator for the technologies used in Internet and data services for over 30 years. His earlier research involved creating new methods in bridging, multicast, traffic monitoring, transport protocols, custom VLSI, and Gigabit networking.

The high-level idea of Disguiser is to provide a static payload as ground truth, which can be used to indicate the occurrence of censorship when the static payload has been altered by network devices. The framework of our proposed system design is illustrated in the Figure 1. Our client-server instructs the vantage points to (1) craft DNS/HTTP/HTTPS requests with the test domain names embedded, (2) send the packets to our control server to trigger censorship, and (3) collect the response back for analysis. Our control server replies to arbitrary requests with a static payload for each type of protocol. Note that we do not send any requests to legitimate servers, and the accessed domains in the requests (if being censored) would still trigger the censorship since the censor devices will see the undesired domains but have no knowledge of whether the destination IP address is associated with a legitimate server of the censored domain.

With the end-to-end design of our framework, we can provide a static payload on the server-side so that the content injected by the censor (i.e., the blockages) can be straightforwardly detected when the client observed a response that is different from our static payload. In other words, this provides us a baseline by controlling what should be expected at the client-side when no censorship is involved so as to accurately recognize the censorship activities.

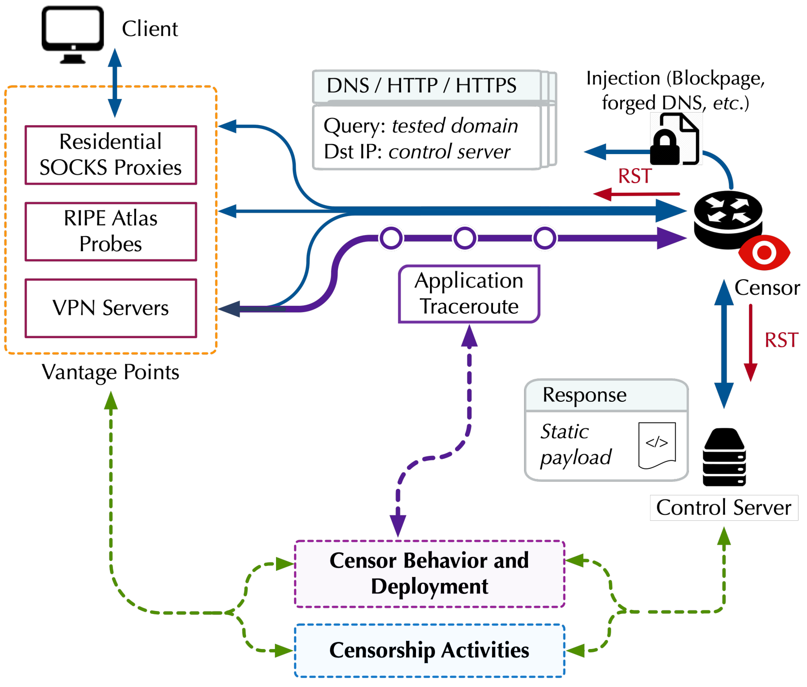

Figure 1: Illustration of Disguiser

Pinpointing Censor Devices Application Traceroute

With the end-to-end framework of Disguiser, we can further explore and identify the deployment of censor devices the manipulate the connections.

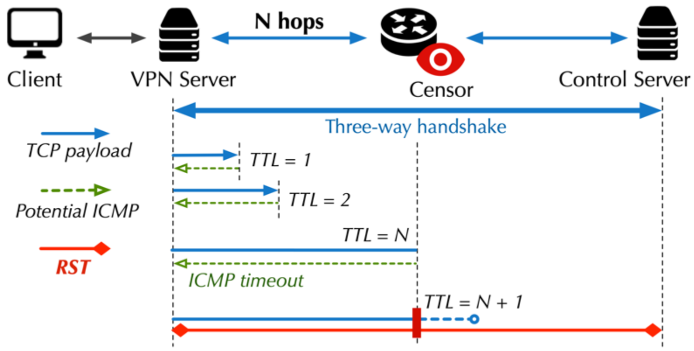

Figure 2: Illustration of Application Traceroute

To conduct an application traceroute, a vantage point will first complete a TCP three-way handshake with our control server to establish a connection. Then, it increases the TTL value of the request that contains a censored domain name. As the TTL increases, we should receive Time Exceeded ICMP packets from routers on the path. Then, when a packet reaches the censor, we will receive a sign of censorship such as injected RST packets. As shown in Figure 2, assuming that the censor's router is N hops away from vantage points and the censor uses its own default TTL value in its injected packet, we should observe a sign of censorship only when TTL is set to N + 1 or larger; otherwise, the packet will be dropped before reaching the censor router (TTL < N) or without being processed when it reaches the censor router (TTL = N).

Table 1 lists the censor deployment information that we can obtain through VPN servers. With Disguiser, we can identify both in-path and on-path censors. The "Hops to Border" column shows that the censors do tend to be deployed close to the nation's border routers. The "AS Rank" column show that the censors are often deployed in the ASes with a high rank, indicating that they can monitor a large number of Internet users in that country.

Table 1: Censor Information (From [1])

Country	Censor Router◇	ASN	ISP	AS Rank	Hops to Border	In-path	Copy TTL
Belarus	178.124.134.*	AS6697	Beltelecom	3	3	✓	
China†	202.97.84.*	AS4134	Chinanet	2	2		
	219.158.4.*	AS4837	China Unicom	3	4		
	61.152.24.*	AS4812	China Telecom	7	4		
	220.181.177.*	AS23724	China Telecom	14	4		
Egypt‡	172.17.50.*	-	-	-	0	✓	
India	125.19.50.*	AS9498	Bharti Airtel	1	1		
	116.119.44.*						
	125.18.125.*						
Iran♮	10.199.250.*	-	-	-	3	✓	✓
Kazakhstan	195.93.153.*	AS48716	PS Internet	17	1	✓	
	91.185.5.*	AS41798	Transtelecom	2	0		
	92.47.151.*	AS9198	Kazakhtelecom	4	1		
Oman	134.0.217.*	AS8529	Omantel	1	-	✓	
Pakistan	110.93.252.*	AS38193	Transworld Associates	1	0	✓	
Russia	195.239.20.*	AS3216	PJSC Vimpelcom	3	1	✓	
	31.192.111.*	AS49335	Server v arendy	111	-		
Saudi Arabia	84.235.94.*	AS39386	Saudi Telecom	1	1	✓	✓
	84.235.12.*	AS25019		5			
South Korea	112.174.83.*	AS4766	Korea Telecom	1	1		
	112.174.84.*						
Turkey	81.212.201.*	AS9121	Türk Telekom	1	2	✓	
Vietnam	113.171.45.*	AS45899	VNPT	2	3	✓	
	113.171.59.*				2		

Enhancing Disguiser for Fine-grained Censorship Investigation

Until now, existing censorship studies have been primarily focusing on country-level characterization. However, the deployment and implementation of censorship may be highly diverse at the ISP level. Leveraging the end-to-end measurement approaches, we then enhance Disguiser's original design by deploying multiple geo-distributed backend server to explore and analyze fine-grained censorship.

Internet censor is typically deployed by authorities to serve the purpose of restricting people from accessing the Internet. Until now, empirical studies mainly focus on country-level characterization. It is a reasonable compromise between feasibility and granularity. On one hand, censorship policies are typically enforced by nations' authorities, resulting in that the nation-wide censorship could be largely consistent. However, the severity of censorship may be highly diverse at the ISP level due to the difference of deployments and implementations, which makes the country-level observation too coarse to draw reasonable conclusions. Limited by the measurement methods, such diversity is mostly underestimated in previous work as researchers have no control on which transit networks or gateways the experiment packets will reach.

By leveraging end-to-end measurement approaches, we can explore multiple paths from one single vantage point. Specifically, we deploy back-end servers distributed worldwide as the destination IP. The packets which carried the same test domain but with a different destination IP will be forced to traverse different border networks. As shown in Figure 2, we then identify the censorship activity on each path for one vantage point, thereby quantifying the censorship diversity caused by the different paths and ISP networks in one country.

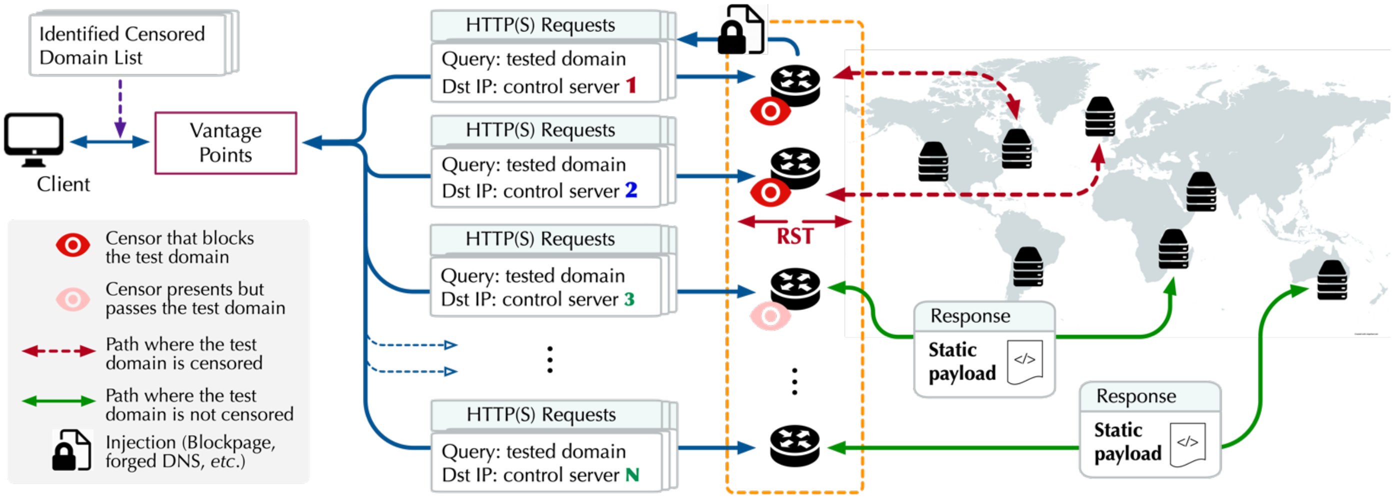

Figure 2: Illustration of exploring fine-grained censorship: identifying multiple paths for diverse censorship policies and deployments within one country

Experiment Setup

To achieve desired experimental goals, we carefully select our control servers by using two criteria: (1) control servers must be widely spread across the world, and (2) there should be fewer censor activities in the hosting countries to avoid interference. To this end, we launch six servers on Amazon AWS, each of which is hosted in a different region: Virginia (North America – East Coast), California (North America – West Coast), Sao Paulo (South America), London (Europe), Bahrain (Middle East), and Cape Town (Africa).

Observations

We here simply define the censorship diversity as the inconsistency of identified presence of censorship activities towards different destinations of control servers from one same vantage points. Table 2 lists the fraction of vantage points that have observed the censorship diversity in each country. The results show that the censorship diversity is prevalence, indicating the country-level aggregation of censorship in existing studies is far away from accurate for understanding the practices, policies, and severity of global Internet censorship activities.

Table 2: Fractions of vantage points that observe the censorship diversity

Country	Percentage
Albania, Argentina, Belgium, Bosnia Herzegovina, Brazil, China, Czechia, Dominican Republic, Ecuador, Estonia, Finland, Georgia, Germany, Ghana, India, Kenya, Macao, Mexico, Moldova, Nepal, Nicaragua, Norway, Pakistan, Panama, Peru, Poland, Serbia, Slovenia, South Africa, Spain, Sweden, Uganda, Uruguay	100%
Kuwait, Australia, Portugal, Indonesia, Colombia, Israel, Chile, Japan, France, United Kingdom, Netherlands	90% - 99%
Taiwan, Singapore, Philippines, Bulgaria, Latvia, Venezuela, Hong Kong	80% - 89%
Vietnam, Russia, Greece, Bangladesh, New Zealand, United States, Italy, Canada	70% - 79%
Ukraine, Lithuania, Thailand	60% - 69%
Turkey, Kazakhstan, FrenchPolynesia, Belarus, Benin, Morocco	50% - 59%
SouthKorea, Iran, United Arab Emirates, Hungary	30% - 39%
Malaysia, Yemen	20% - 29%
Saudi Arabia	10% -19%
Armenia, Botswana, Cambodia, Costa Rica, Cyprus, Denmark, Guam, Iraq, Ireland, Malta, North Macedonia, Paraguay, Romania, Slovakia, Qatar	0% - 9%

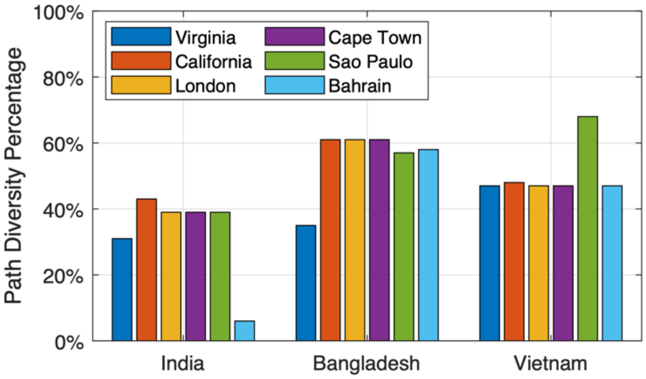

Figure 7 shows some interesting preliminary results in three countries: India, Bangladesh, and Vietnam. In India, the paths toward to the control server in Bahrain (Middle East) have experienced far less censorship than the other paths (around 8%), while the other paths experience a percentage of censorship ranging from 18.74% to 26.50%. Similarly, in Bangladesh, the paths to Virginia (U.S. East) have also experienced less censorship (36.5%) than the other paths (around 50%). In the case of Vietnam, the paths to Sao Paulo (South America) have been detected a larger percentage of censorship than the other paths.

Resources

The design, implementation, and deployment of Disguiser have been described in the research paper published in ACM SIGMETRICS 2022:

[1] Lin Jin, Shuai Hao, Haining Wang, and Chase Cotton.
Understanding the Practices of Global Censorship through Accurate, End-to-End Measurements,
In *ACM SIGMETRICS 2022*, June 2022.

The code of Disguiser and its enhancement have been open-sourced in two separate Github repositories, which include the code base for deploying the framework and performing the analysis:

https://github.com/e2ecensor/Disguiser_public
https://github.com/e2ecensor/newDisguiser

More data collected by Disguiser and its enhancement have been processed and shared with Google Drive, and its descriptions and usage have been specified in the *Readme*s of Github repositories:

https://drive.google.com/drive/u/1/folders/106F_7gkKO-zRqpdyOokGT_Gr-wonRfnk
https://drive.google.com/drive/folders/1vZ7JuQsWQYIKkT8hxX-_qRldjnSuykQy

All above information is published in our project website:

e2ecensor.github.io

.