

# A Bestiary of Blocking

The Motivations and Modes behind Website Unavailability

Michael Carl Tschantz, ICSI

Sadia Afroz, ICSI and UC Berkeley

Shaarif Sajid, LUMS

Shoaib Asif Qazi, LUMS

Mobin Javed, LUMS

Vern Paxson, ICSI and UC Berkeley





vinton g cerf

@vgcerf

Follow



The dunderheads at Verizon block access to [myverizon.com](https://myverizon.com) account if you are overseas. I plan to cancel my service. idiots.

1:58 PM - 1 Sep 2017

16 Retweets 60 Likes



10

16



60





**sadia afroz**

@sheetal57

Follow



The best CAPTCHA is the one you can't even see! Trying to log into my bank from Bangladesh ...

## Security Step

For your protection, you'll need to complete an additional security step.

Use this text code and enter the same in the text box below.



To continue, please enter the text you see above.

Continue to sign in

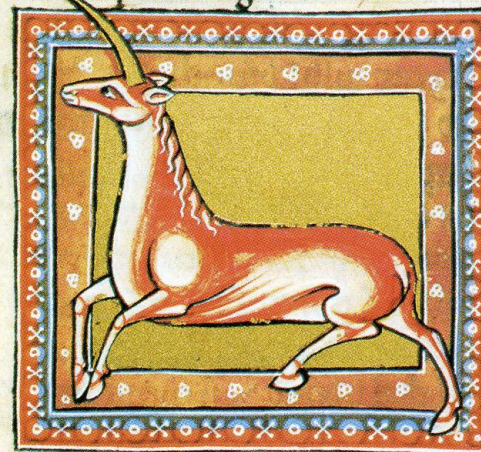
Return to home page

10:51 PM - 21 Dec 2017



# Types of and Motivations for Blocking

pluit; i munit lanat; p...  
eicit; peccem ignoscit; D F



**M**o no  
st monoc  
strum mugu  
do. equino co  
phantal ped  
da simillim  
Cornu media  
ei p tenditur  
re iurifico. a  
itudinem p  
iii. ita acit

quicquid impetrat facile ictu ei p foretur. Si u  
nit in hominum potestatem. rinctum qdam p



**U**  
dictus  
ore su  
fetus o  
sus; M  
eos in f  
geniare  
r carni  
clam r

# Block pages with or without reasons



**Sorry, our site is unavailable in your country right now.**

In the meantime, visit [CVS Health](#) to learn about our company.

The requested URL was rejected. Please consult with your administrator.

Your support ID is: 5527351751160471927

# Failing load anything

- DNS errors
- Timeouts
- Connection refused
- Connection resets

# Soft blocks: slowing page loads, CAPTCHAs



## Checking your browser before accessing vape.gg.

This process is automatic. Your browser will redirect to your requested content shortly.

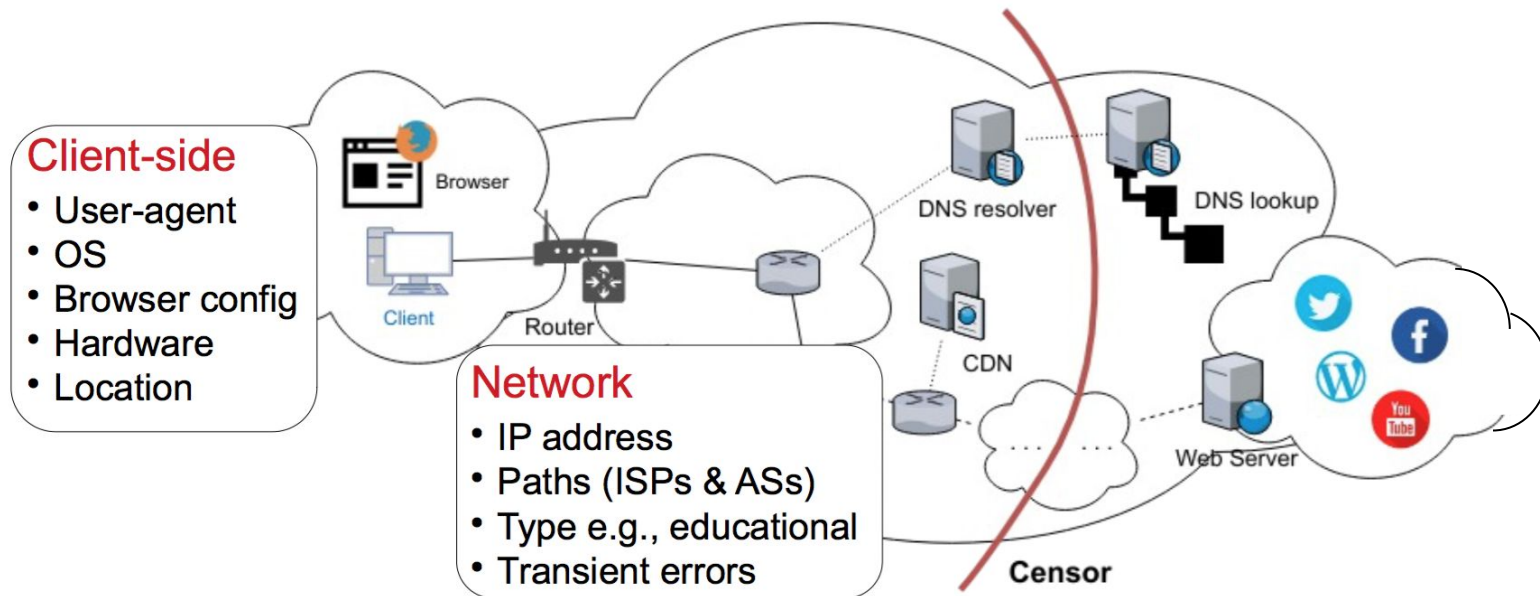
Please allow up to 5 seconds...

[DDoS protection by Cloudflare](#)

Ray ID: 421a43cf50f43e86



# A website can be inaccessible in many ways



📶 Airtel 3G

9:45 PM

📶 🔌 30% 🔋

archive.org



**Your requested URL has been blocked as per the directions received from Department of Telecommunications, Government of India. Please contact administrator for more information.**

# Motivations

---

	Server (including CDNs)	Middlebox (ISP, gov.)
Political censorship	Bowing to China's demands	Great Firewall of China
Economic sanctions	US sites blocking Cuba/Iran	
Third-party liability	Blocking US due to SESTA	
Copyright	Germany blocking YouTube	ISP blocking Pirate Bay
Other compliance	GDPR	
Security	Fraud	
Hosting costs	CDN fees	
Revenue	Price discrimination	Net-neutrality
Unintentional	Slash-dotting	Overloaded rural links

---



# Crawler

1. Attempts DNS resolution for each provided URL
  - If unresolved, try again adding “www.” prefix
2. If resolved, uses Python’s Requests package to request the webpage
  - Timeout of 30 seconds
  - Chrome user agent string from a MacBook
  - Tries HTTP, but follows any automatic switches to HTTPS
3. Use RegExes to recognize block pages



# Country-Based Blocking

JCPenney



**We Are Currently Unable to Provide a Shopping Experience for This Country**

Please try again later

## Sorry about that

Grubhub food delivery is not available in your country







SonicWALL | Network Security Appliance



**This site has been blocked by the network administrator.**

Block reason: **Gateway GEO-IP Filter Alert**

IP address: **39.53.47.73**

Connection initiated from country: **Pakistan**

# Error 1009

Ray ID: 422f2004f42f93a8 • 2018-05-30 06:25:43 UTC

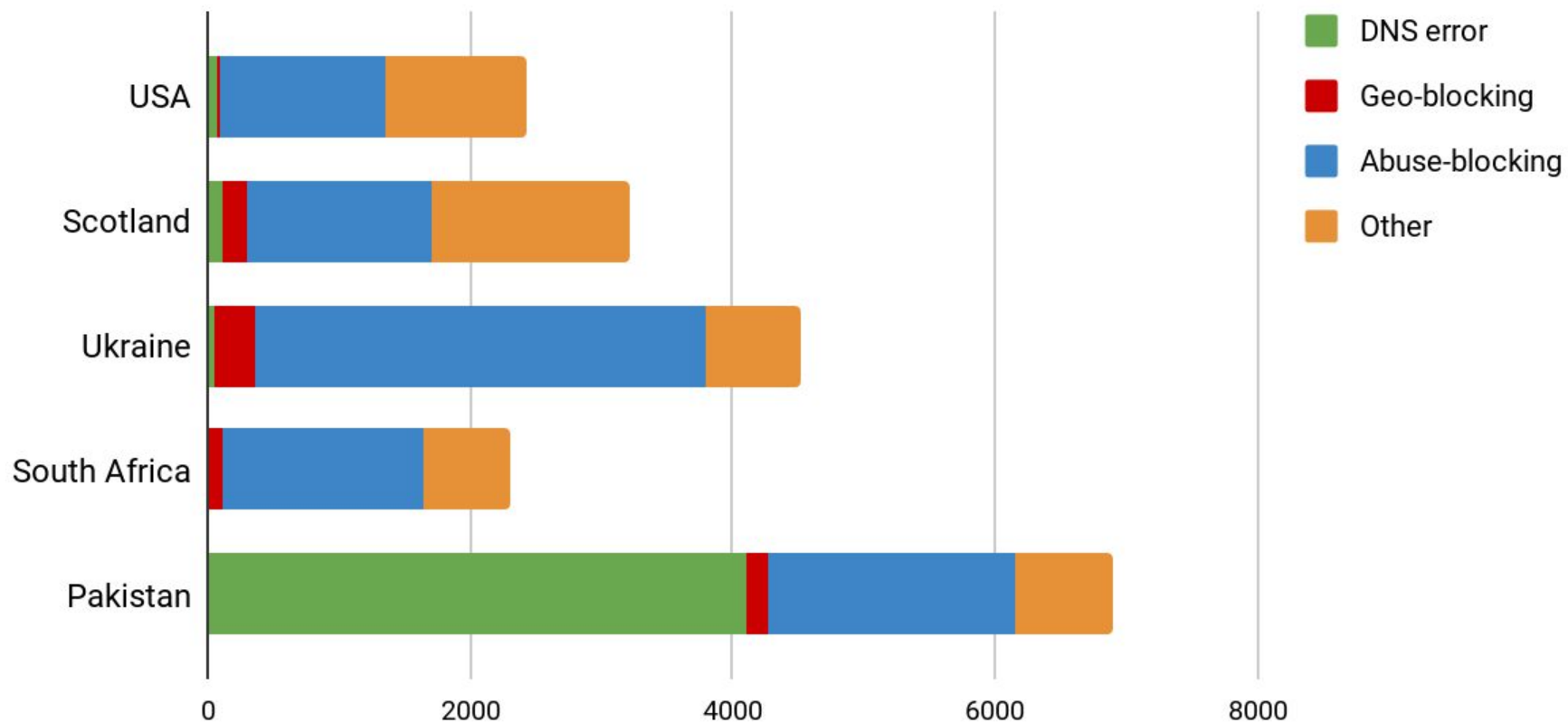
Access denied

## What happened?

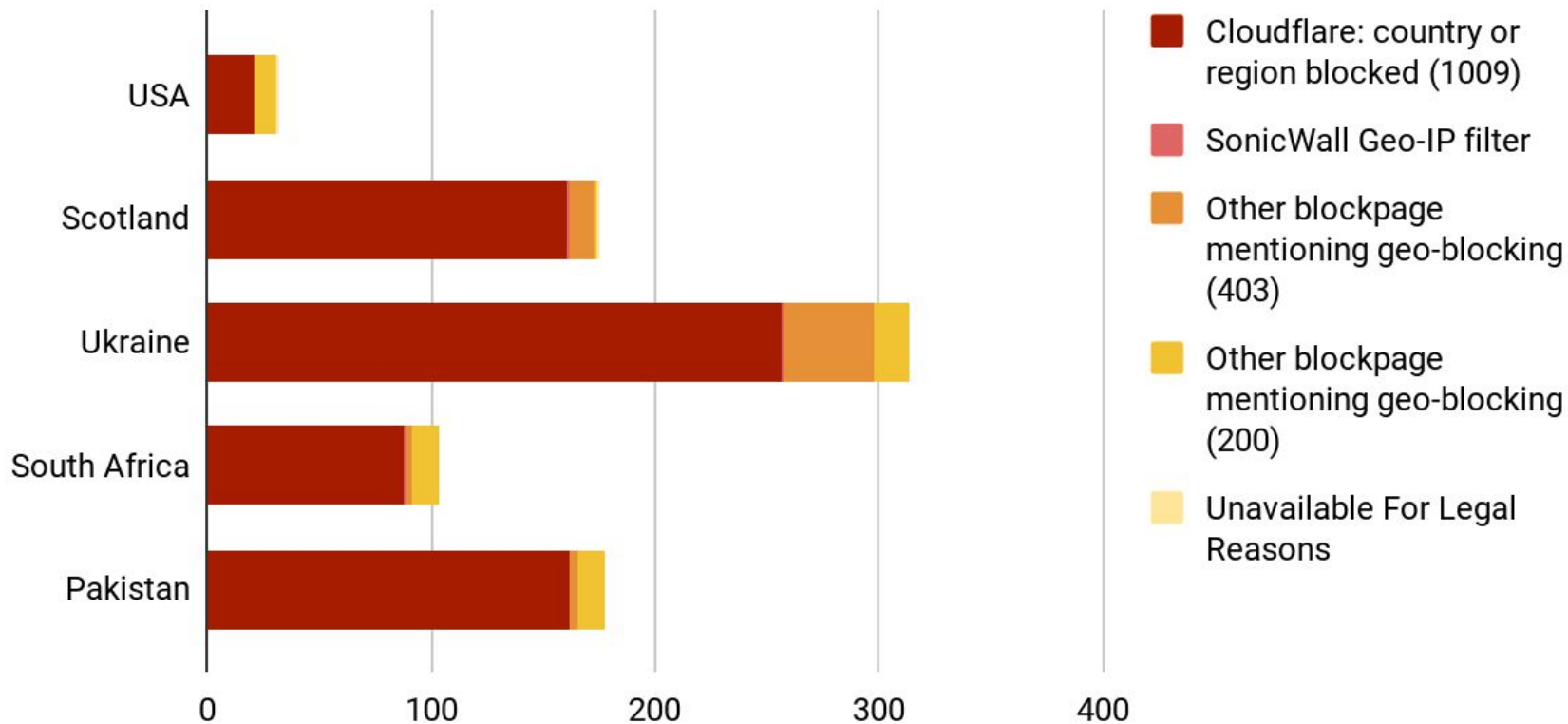
The owner of this website (stoiximan.com.cy) has banned the country or region your IP address is in (US) from accessing this website.

85,421 Cloudflare-hosted URLs from Alexa top 1M domains from





# Types of geo-blocking





# *Security-Motivated Blocking*

# Blocking mechanisms: Security

www.lendingtree.com - Access Denied

## Error code 16

This request was blocked by the security rules

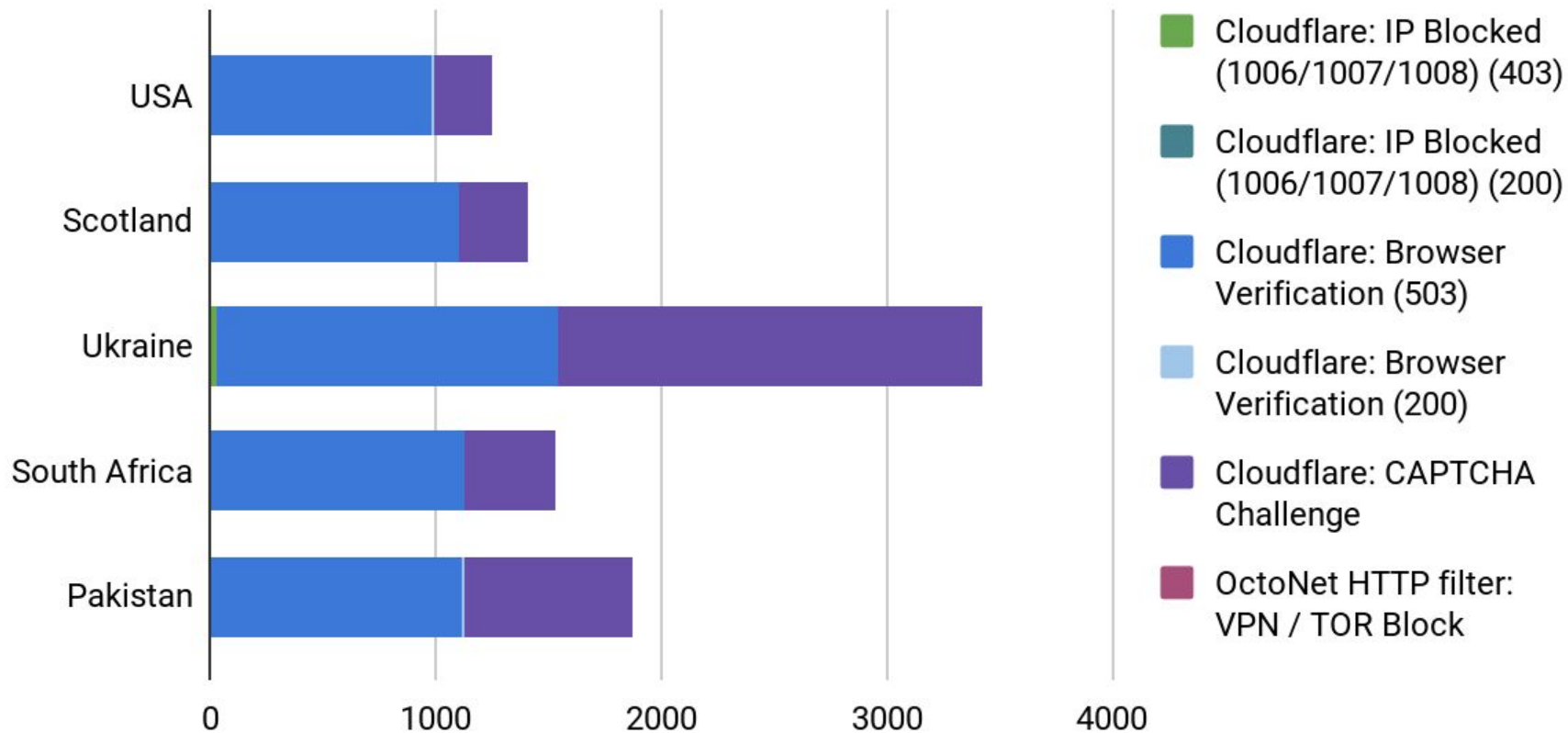
---

**2018-05-23 09:03:55 UTC**

Your IP [39.45.233.157](#) | Proxy IP [45.60.28.117](#) (ID 10407)

Incident ID: 407001860088678508-402561825777322917

# Types of security-motivated blocking







# GDPR-Motivated Blocking

# **We're sorry. This site is temporarily unavailable.**

We recognise you are attempting to access this website from a country belonging to the European Economic Area (EEA) including the EU which enforces the General Data Protection Regulation (GDPR) and therefore cannot grant you access at this time.

For any issues, [contact us](#).

(403 error.)

<https://www.newstimes.com>



451 Unavailable For Legal Reasons

**Sorry, this content is not available in your region.**

# Los Angeles Times

Unfortunately, our website is currently unavailable in most European countries. We are engaged on the issue and committed to looking at options that support our full range of digital offerings to the EU market. We continue to identify technical compliance solutions that will provide all readers with our award-winning journalism.

7081 websites from various Alexa lists before and after GDPR went into effect.

Count:

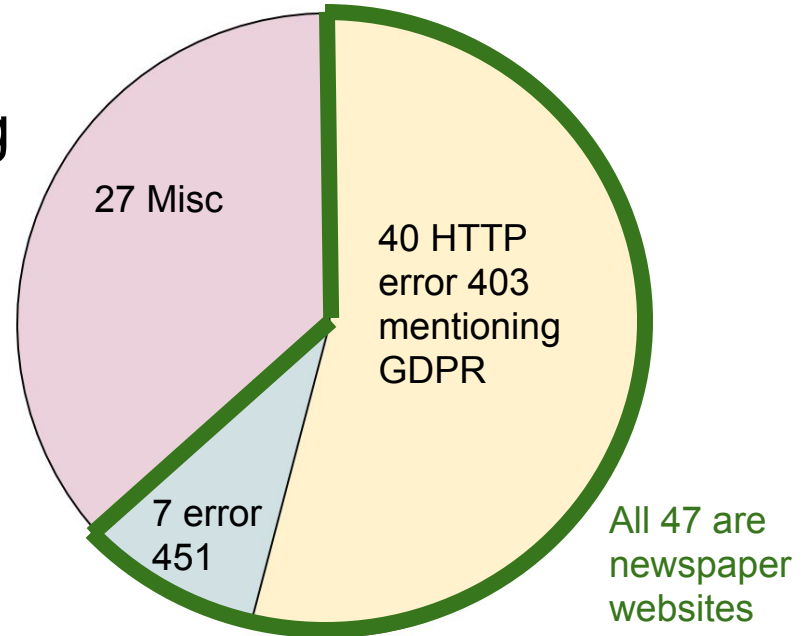
- always worked before,
- still works from the US after, but
- from none of the EU vantages after



# Measurement Results

7081 Websites tested

74 websites stopped working



Study

Level of operation

Relation to location



Country

Mechanism

Directly uses location



Security

Motivation

Indirectly affects locations



GDPR

Motivation

Directly uses location

# A Bestiary of Blocking

The Motivations and Modes behind Website Unavailability

**Michael Carl Tschantz**, ICSI

**Sadia Afroz**, ICSI and UC Berkeley

**Shaarif Sajid**, LUMS

**Shoaib Asif Qazi**, LUMS

**Mobin Javed**, LUMS

**Vern Paxson**, ICSI and UC Berkeley

