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
The dunderheads at Verizon block access to myverizon.com account if you are overseas. I plan to cancel my service. Idiots.
The best CAPTCHA is the one you can't even see! Trying to log into my bank from Bangladesh ...
Types of and Motivations for Blocking
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

**Client-side**
- User-agent
- OS
- Browser config
- Hardware
- Location

**Network**
- IP address
- Paths (ISPs & ASs)
- Type e.g., educational
- Transient errors
Your requested URL has been blocked as per the directions received from Department of Telecommunications, Government of India. Please contact administrator for more information.
<table>
<thead>
<tr>
<th>Motivations</th>
<th>Server (including CDNs)</th>
<th>Middlebox (ISP, gov.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Political censorship</td>
<td>Bowing to China’s demands</td>
<td>Great Firewall of China</td>
</tr>
<tr>
<td>Economic sanctions</td>
<td>US sites blocking Cuba/Iran</td>
<td></td>
</tr>
<tr>
<td>Third-party liability</td>
<td>Blocking US due to SESTA</td>
<td></td>
</tr>
<tr>
<td>Copyright</td>
<td>Germany blocking YouTube</td>
<td>ISP blocking Pirate Bay</td>
</tr>
<tr>
<td>Other compliance</td>
<td>GDPR</td>
<td></td>
</tr>
<tr>
<td>Security</td>
<td>Fraud</td>
<td></td>
</tr>
<tr>
<td>Hosting costs</td>
<td>CDN fees</td>
<td></td>
</tr>
<tr>
<td>Revenue</td>
<td>Price discrimination</td>
<td>Net-neutrality</td>
</tr>
<tr>
<td>Unintentional</td>
<td>Slash-dotting</td>
<td>Overloaded rural links</td>
</tr>
</tbody>
</table>
Measuring Blocking
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
   ○ Trys HTTP, but follows any automatic switches to HTTPS
3. Use RegExes to recognize block pages
Country-Based Blocking
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
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

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

- Cloudflare: country or region blocked (1009)
- SonicWall Geo-IP filter
- Other blockpage mentioning geo-blocking (403)
- Other blockpage mentioning geo-blocking (200)
- Unavailable For Legal Reasons
Security-Motivated Blocking
Types of security-motivated blocking

- **USA**: Cloudflare: IP Blocked (1006/1007/1008) (403)
- **Scotland**: Cloudflare: IP Blocked (1006/1007/1008) (200)
- **Ukraine**: Cloudflare: Browser Verification (503)  
  Cloudflare: Browser Verification (200)  
  Cloudflare: CAPTCHA Challenge
- **South Africa**: OctoNet HTTP filter: VPN / TOR Block
- **Pakistan**:
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.)
Sorry, this content is not available in your region.
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

- 40 HTTP error 403 mentioning GDPR
- 27 Misc
- 7 error 451

All 47 are newspaper websites
<table>
<thead>
<tr>
<th>Study</th>
<th>Level of operation</th>
<th>Relation to location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country</td>
<td>Mechanism</td>
<td>Directly uses location</td>
</tr>
<tr>
<td>Security</td>
<td>Motivation</td>
<td>Indirectly affects locations</td>
</tr>
<tr>
<td>GDPR</td>
<td>Motivation</td>
<td>Directly uses location</td>
</tr>
</tbody>
</table>

"The Leopard" from the "Rochester Bestiary". Wikipedia.
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

"The Leopard" from the "Rochester Bestiary". Wikipedia.