Transparency into the Causes of Website Inaccessibility

Sadia Afroz (ICSI & UC Berkeley), Huilin Chen (ICSI & UC Berkeley), Mobin Javed (LUMS), Marc Juarez (ICSI & KU Leuven), Vern Paxson (ICSI & UC Berkeley), Shoaib Asif Qazi (LUMS), Shaarif Sajid (LUMS), Michael Carl Tschantz (ICSI)

The Need for More Transparency

There are many reasons for a page being accessible in one location but not in another. It is easy to confuse these reasons and draw wrong conclusions from inaccessibility. Our goal is to bring more transparency on why pages are inaccessible by enumerating all the possible reasons and find ways to tease them apart.

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

Network
- IP address
- Paths (ISPs & ASs)
- Type e.g., educational
- Transient errors

Server-side
- Misconfiguration
- Transient errors
- Localization
- Geo-blocking
- CDNs

Geo-blocking
- GDPR-motivated geo-blocking: to evade liability temporarily.

Other Confounding Variables
- Client: e.g., browser
- Network: e.g., PlanetLab variance
- Server: e.g., local misconfigs

Successful visits to top Alexa 500 (3 samples/site) most popular sites in US using different User-Agent strings.

Measure Impact of Confounding

Isolate variable in question (e.g., network) by pinning down the rest of variables.

Estimate the Censor’s Location
- traceroute can be used to distinguish between the server or an intermediate router blocking the page.
- Increments TTL in ICMP packets until destination is unreachable.
