Crowdsourcing & split-tunneling to map and circumvent URL filtering

Walid Al-Saqaf
Örebro University - Sweden

A presentation for RIPE 68
Warsaw, May 14

Two parallel tracks using Alkasir

1) **Mapping URL filtering through crowdsourcing**

   - Reporting of suspected filtered URLs

   Manual submissions & automatic cross-validation with crowdsourcing

   -

   - Who is blocking? Autonomous Systems or Internet Service Providers?

   - Crowdsourcing to identify similar response patterns to HTTP requests

2) Circumventing filtering through SSH split-tunneling

   - Develop an effective and affordable circumvention method

   -Users to circumvent and also provide data

   - Analysis of user behavior in reporting about censorship

1) Mapping URL Filtering

- Manual entry of URL (one time process)
- If filtering identified (to be explained in the next slide)
- Automatic crowdsourcing triggered for other users in the same country

- Content analysis using statistical software (SPSS)

URL Submissions

DEMO

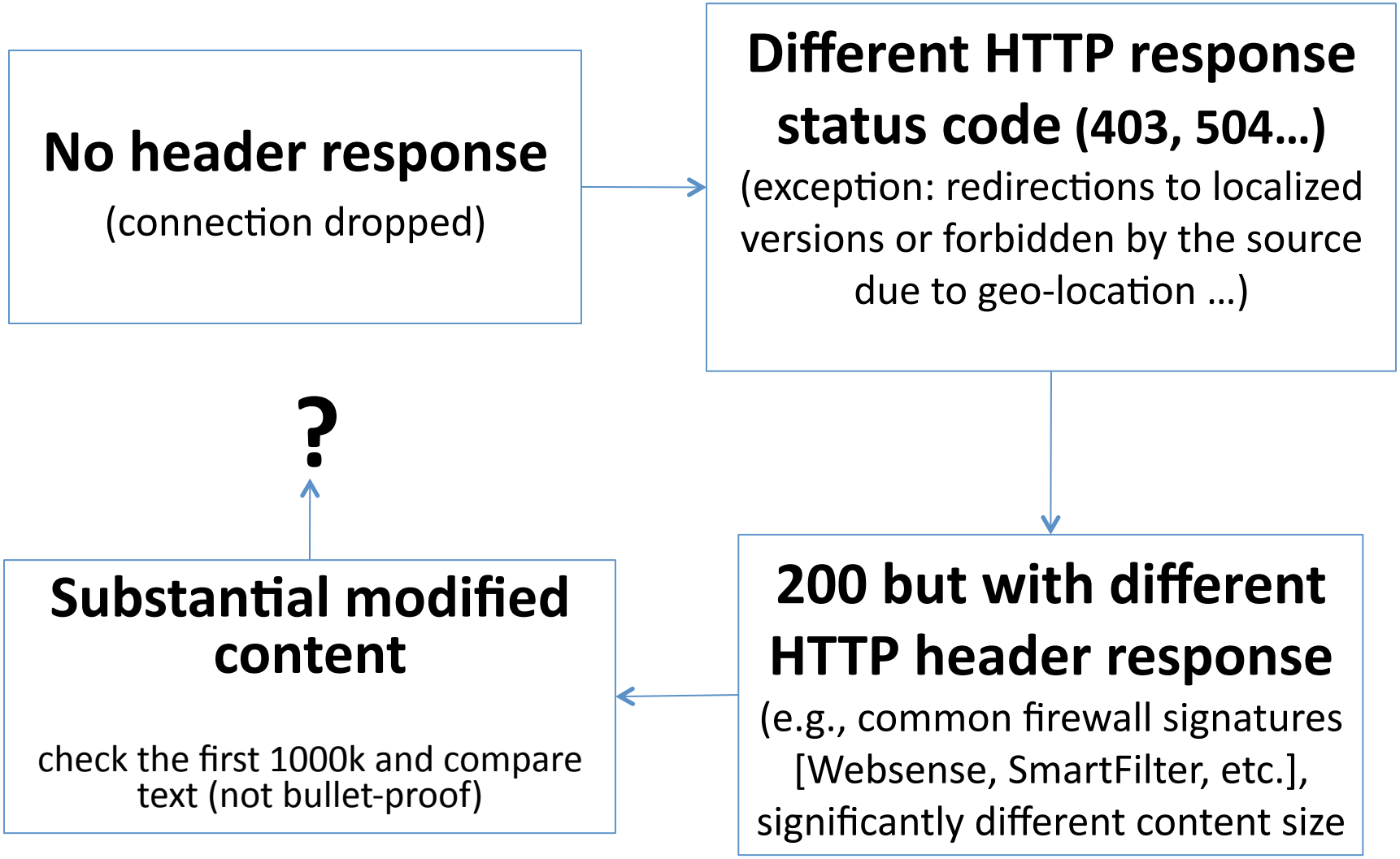
1) User reports a URL manually
2) Software verifies –using server- by comparing headers and/or content with remotely fetched version of the same URL
3) If found to be blocked, the moderator verifies if the URL is not of porn/nudity content (if approved before, it is approved automatically & vice versa)
4) Moderator is a closed group and public access is limited due to personal risk
5) Assessment of what is porn and what is not is based on meta website data
6) Entry is added to the database with relevant data accordingly
7) Requests to clients of other users using the same AS/ISP are sent to also check whether the URL is filtered (all in the background)*
8) If multiple clients confirmed the filtering, the request is sent to clients of other users in the same country but using other ISPs*

***Note:** It is also possible to check manually entered URLs in any country/ISP in background*

* Under development

How to know if a URL is filtered

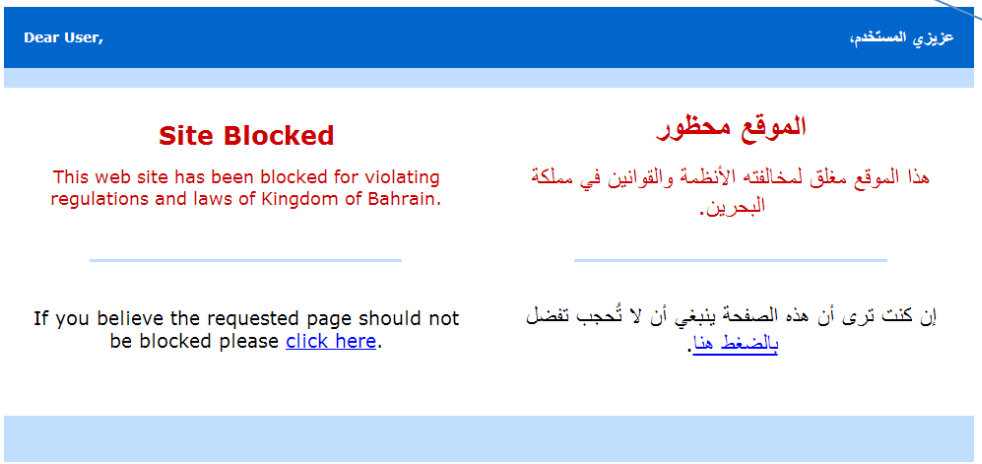
(assuming right IP address reached, i.e., no DNS tampering occurred)



Example: Bahrain, same ISP

Browse | Structure | SQL | Search | Insert | Export | Import | Operations | Triggers

url	country_code	l_header	r_header
bahrainrights.org	BH	HTTP/1.1 302 FOUNDCONTENT-LENGTH: 318 CONTENT-TYPE: TEXT/HTML LOCATION: HTTP://WWW.ANONYMOUS.COM.BH SERVER: NETCACHE APPLIANCE (NETAPP/6.0.7)	HTTP/1.1 302 FoundCACHE-CONTROL: MAX-AGE=1 CONNECTION: CLOSE CONTENT-LENGTH: 378 CONTENT-TYPE: TEXT/HTML; CHARSET=ISO-8859-1 LOCATION: HTTP://WWW.BAHRAINRIGHTS.ORG/EN SERVER: APACHE/2.2.14 (UNIX) MOD_SSL/2.2.14 OPENSSL/0.9.8L DAV/2 MOD_AUTH_PASSTHROUGH/2.1 FRONTPAGE/5.(
awaal.net	BH	NO HEADER 000 (CONNECTION FAILED)	HTTP/1.1 200 OKCACHE-CONTROL: NO-STORE, NO-CACHE, MUST-REVALIDATE, POST-CHECK=0, PRE-CHECK=0 CONNECTION: CLOSE CONTENT-TYPE: TEXT/HTML PRAGMA: NO-CACHE SERVER: APACHE/2.2.11 (UNIX) MOD_SSL/2.2.11 OPENSSL/0.9.8l DAV/2 MOD_AUTH_PASSTHROUGH/2.1 MOD_BWLIMITE TRANSFER-ENCODING: CHUNKED X-POWERED-BY: PHP/5.2.9
torproject.org	BH	NO HEADER 000 (CONNECTION FAILED)	HTTP/1.1 302 FoundCONNECTION: CLOSE CONTENT-LENGTH: 335 CONTENT-TYPE: TEXT/HTML; CHARSET=ISO-8859-1 LOCATION: HTTPS://WWW.TORPROJECT.ORG/ SERVER: APACHE/2.2.9 (DEBIAN) DAV/2 SVN/1.5.1 MOD_SSL/2.2.9 OPENSSL/0.9.8G X-PAD: AVOID BROWSER BUG
eskanaali.net	BH	HTTP/1.1 403 FORBIDDENCACHE-CONTROL: NO-CACHE CONNECTION: CLOSE CONTENT-LENGTH: 2300 CONTENT-TYPE: TEXT/HTML; CHARSET=UTF-8 PRAGMA: NO-CACHE	HTTP/1.1 200 OKCACHE-CONTROL: PRIVATE CONNECTION: CLOSE CONTENT-TYPE: TEXT/HTML; CHARSET=WINDOWS-1256 PRAGMA: PRIVATE SERVER: APACHE/2.2.14 (UNIX) MOD_SSL/2.2.14 OPENSSL/0.9.8E-FIPS-RHEL5 MOD_BWLIMITED/1.4 TRANSFER-ENCODING: CHUNKED X-POWERED-BY: PHP/5.2.11 X-UA-COMPATIBLE: IE=7



No consistency
& no transparency

Benefits of source-crowding

1) Maximizes productivity by giving users the chance to decide what filtered URLs they want to access

2) Enhances reliability through cross-validation based on location

3) Allows testing and verifying any URL virtually any time and anywhere an Alkasir client is running

4) When enabled, can track the level and depth of filtering through automatic addition of externally linked URLs in a reported document

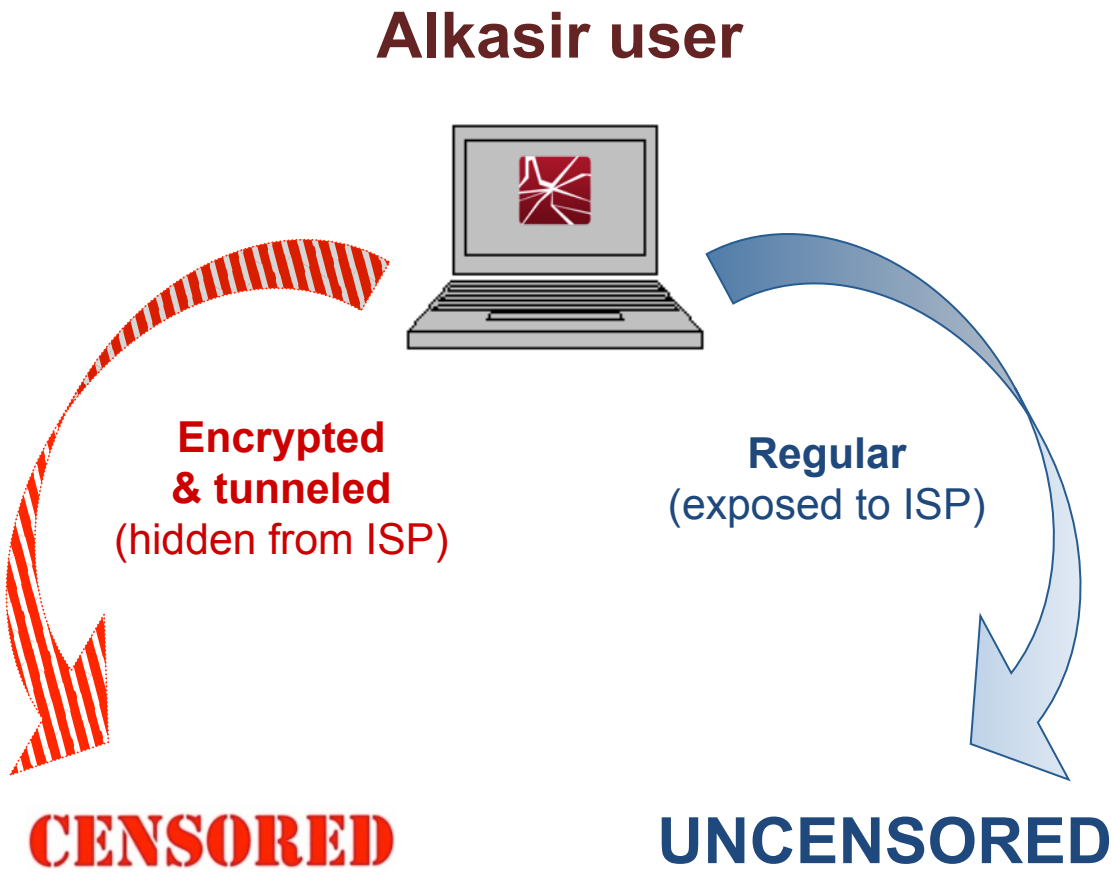
2) Circumventing URL Filtering

1) Socks local proxy activated after connecting to SSH server (SSL and obfuscated/ normal Putty) with local dynamic port (AES CBC enc for SSH2)
2) A pool of 100+ proxy IPs available (different class B addresses)
3) Proxy automatic configuration file that is retrieved when Alkasir is connected is used to divert traffic as required (split-tunneling)
4) Remote DNS resolution enabled to prevent DNS leaks
5) Alkasir.com is always accessed through the encrypted tunnel

Blank alkasir.pac file:

```
function FindProxyForURL(url,host)
        {
          if (shExpMatch(host, "*alkasir.com*"))
                {
                  return "SOCKS5 127.0.0.1:<local PORT>;
                  DIRECT";
                }   return "DIRECT";
        }
```

The split-tunneling approach



The PAC file is built based on the geo-location of the client retrieved through ip2location (paid service) with the IP as input

Benefits of split-tunneling

1) Utilizes bandwidth on the server and client efficiently

2) Allows users to access localized services, news normally

3) Reduces the risk of proxy-use detection by the ISP

4) Encourages reporting more filtered URLs, moderators to learn more about filtering methods and content

5) Can serve as an additional plug-in for other circumvention tools (utilizing the list of blocked URLs)

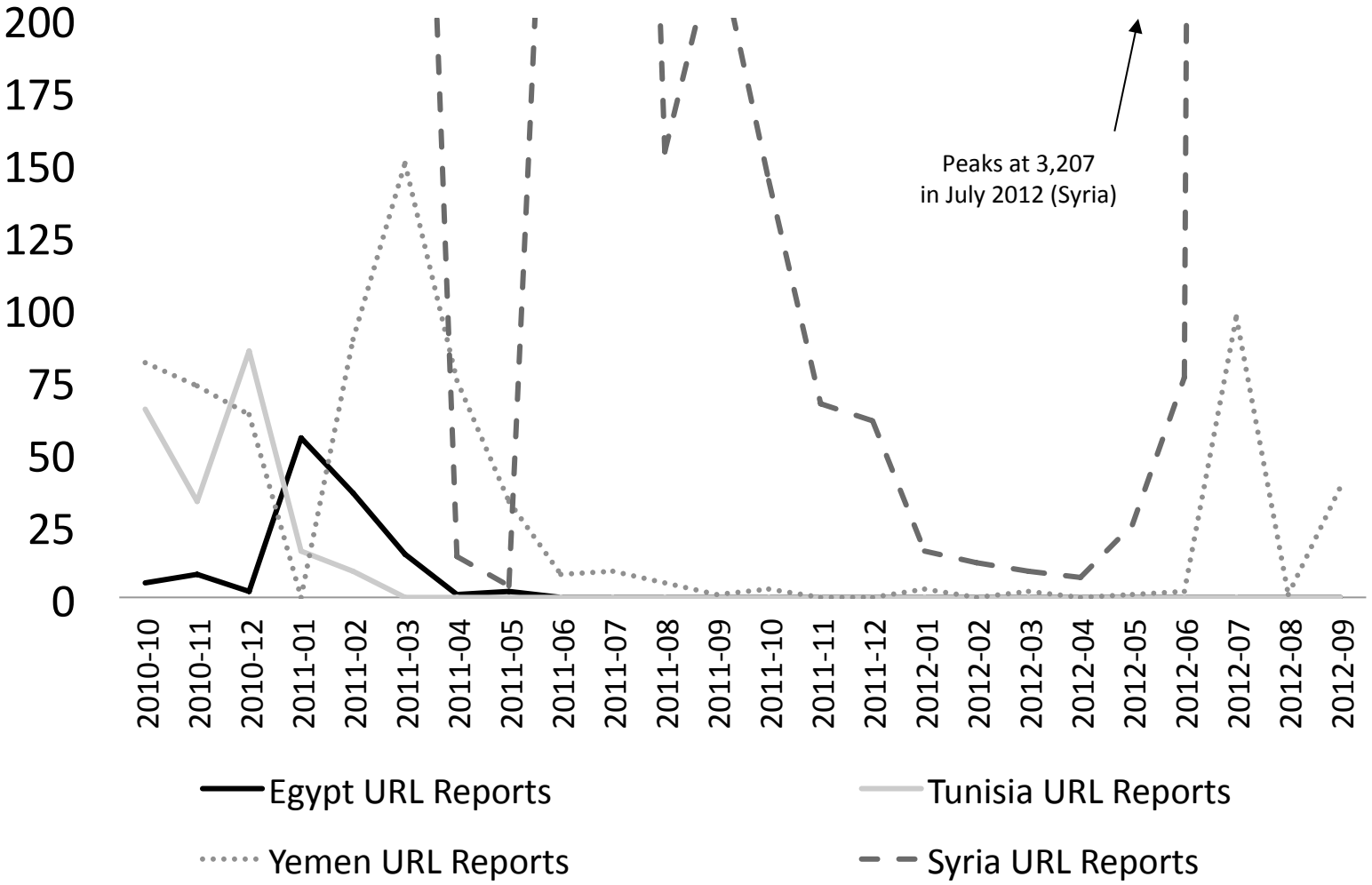
6) Encourages the updating of empirical data on filtering for use by advocacy groups, scholars, and users

Deriving the sample (1/10/2010-1/10/2012)

Country	N	S	$\langle S \rangle$	U	V	$\langle V \rangle$	τ	Φ
Syria	22,460	961,026	43	812	4,443,657	5,472	732	1321
Egypt	435	11,494	26	22	2,028	92	35	332
Tunisia	323	47,962	148	103	658	6	174	276
Yemen	3,094	123,081	40	324	193,581	597	732	169
United Arab Emirates	927	27,190	29	193	29,331	152	732	37
Libya	220	9,929	45	32	16,093	503	519	20
Bahrain	407	10,741	26	72	6,879	96	719	15
Sudan	274	8,591	31	23	664	29	732	12
Jordan	161	5,632	35	17	2,121	125	732	8
Qatar	226	4,769	21	13	239	18	693	7
Kuwait	228	4,760	21	10	446	45	732	7
Oman	221	2,904	13	22	106	5	675	4
Algeria	98	2,658	27	4	212	53	707	4
Morocco	98	1,171	12	4	82	21	732	2

Table 3: internet censorship impact factor (Φ) in Arab states

Server data shows:
Filtering patterns corresponded to political developments



A combined graph showing the number of URL reports for the four case studies (Tunisia, Egypt, Yemen, and Syria)

Survey: Most widely used circumvention tools

Circumvention solution	Users	Percentage
Hotspot Shield	257	31.0%
Ultrasurf	191	23.0%
Web-based proxy	166	20.0%
Other	72	8.7%
A VPN service	40	4.8%
Tor	25	3.0%
JonDo	22	2.7%
I don't remember	18	2.2%
DynaWeb	16	1.9%
GPass and FirePhoenix	12	1.4%
Your-Freedom	11	1.3%

Table 11: Circumvention tools used by survey informants in Arab countries

Survey: What is needed for better circumvention

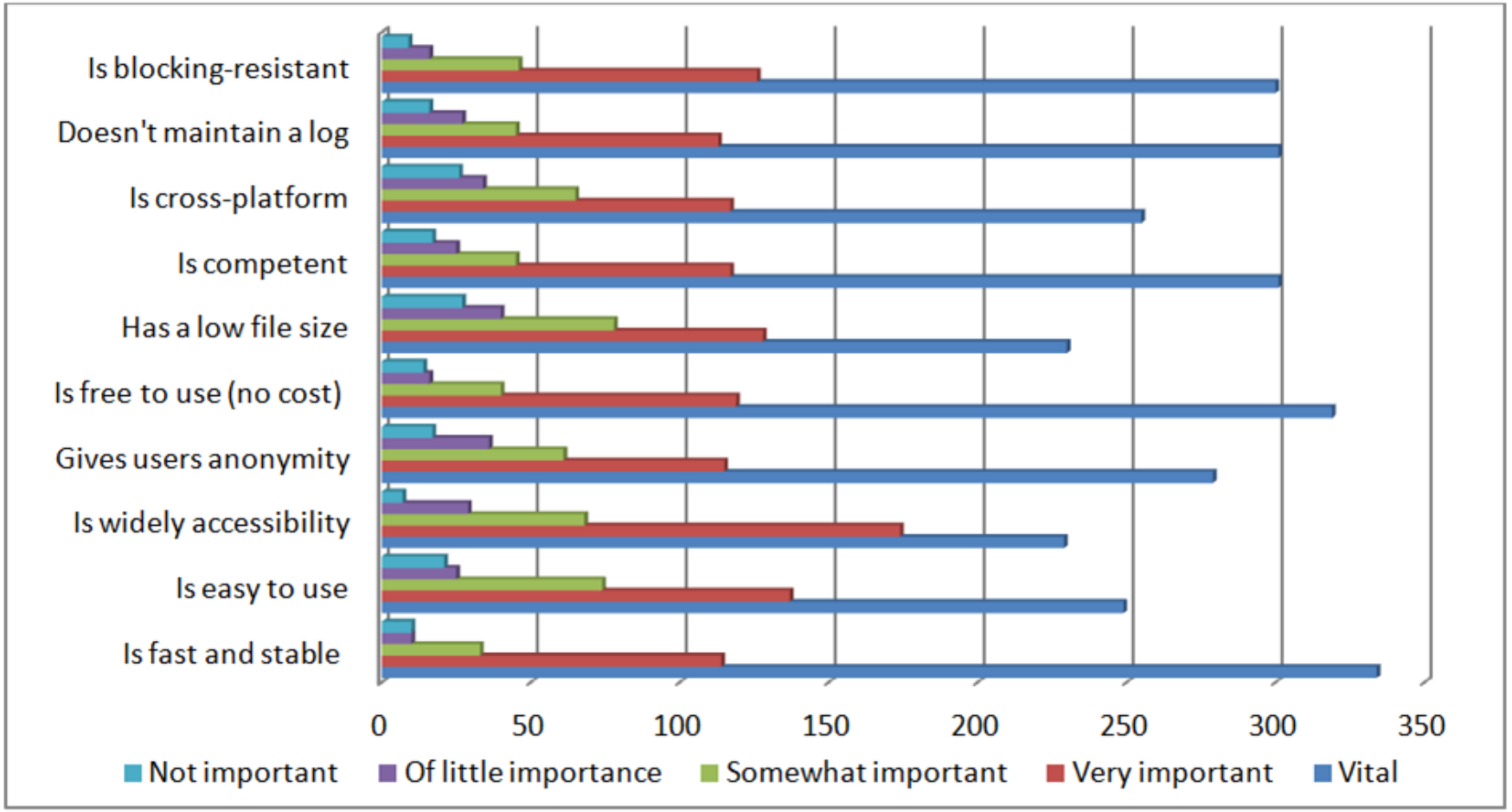


Figure 24: Graph showing factors' importance to informants in Arab countries concerning circumvention tools

Future work and ambitions

- Enhancing Alkasir's blocking-resistance
- Developing extensions for browsers and apps for mobile devices
- Reducing risks to users, ensuring lack of personal data and enhancing split-tunneling to prevent detection of proxy use
- Adding multiple circumvention methods (e.g., VPN, web-based…)
- Building an API to allow access to data (requires risk assessment)
- Defining a more concrete & systematic categorization approach
- Cooperating with like-minded projects to improve reliability of data (Herdict, OpenNet Initiative, Choke Point)
- Proposing the integration of Alkasir's data into other popular circumvention tools (Tor, Psiphon, etc.) to enable split-tunneling

Questions