The Filtering Matrix

Interrogating Internet Filtering
and Surveillance Practices Worldwide

Nart Villeneuve
Director of Technical Research
Citizen Lab, University of Toronto

CACR 2004

The Filtering Matrix

- Technical & non-technical filtering measures targeting multiple levels of access:
  - **Self-censorship** / state-directed ‘encouragement’ to use commercial filtering products
  - **Port Blocking** / Throttling (directed at file sharing services & VOIP)
  - **Content Removal**: take down notices, forum post removals, website closures
  - **Geolocation filtering**: content accessible or inaccessible by geographic location
  - **Internet Filtering**: Internet café’s, Schools & Libraries, Businesses, ISP’s & at centralized location near the backbone connections

_OpenNet Initiative

- Mission: investigate and challenge Internet filtering and surveillance practices
  - Internet Filtering
  - Monitoring & Surveillance
  - Circumvention Technology

University of Toronto
Munk Centre
for International Studies
at Trinity College

Berkman

University of
Cambridge

Research Network

- Human-based Network
  - Established relations of trust with partners on the ground (H2H Networks)
- Technological Network
  - Developed a testing network to enumerate Internet filtering
    - Technologies to determine what content and services are blocked, where and with what technology

Filtering: National Level

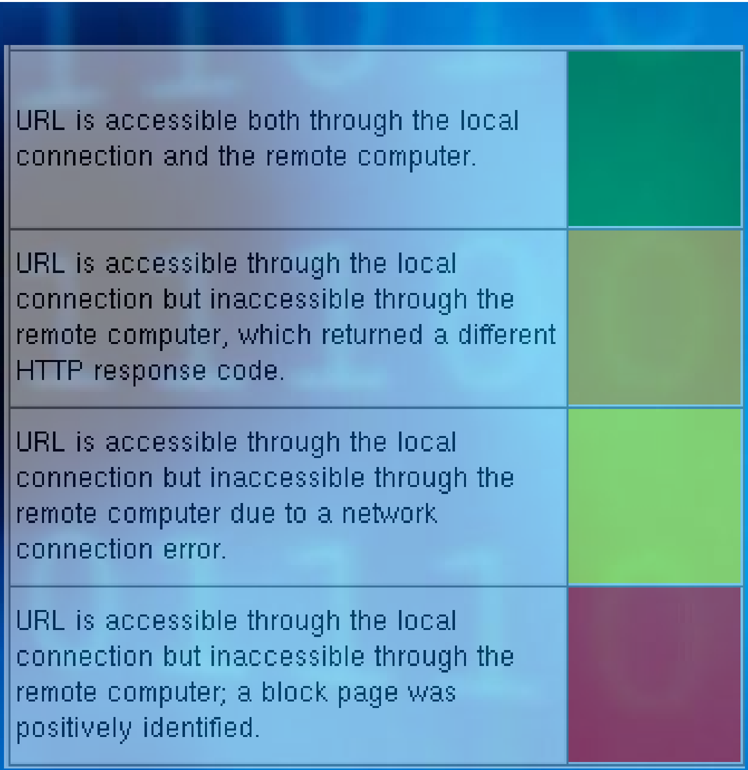
- Limited: Access restricted to a small number of websites.
- Distributed: Access is restricted to a significant number of sites, but sporadically implemented by different ISP's
- Comprehensive: Access is restricted to a number of sites within a comprehensive national framework.
  - State-directed encouragement of filtering products
  - Filtering targeted towards child pornography
  - Filtering targeted towards hate speech

Methodology

- Contextual Research
  - Background: Law & Politics
  - Reported filtering behavior
  - URL/Domain List & Keyword Generation
- Network Access
  - Proxy server
  - Long distance dial-up
  - Distributed Application
  - Dedicated Server
- Testing Phase

Analysis

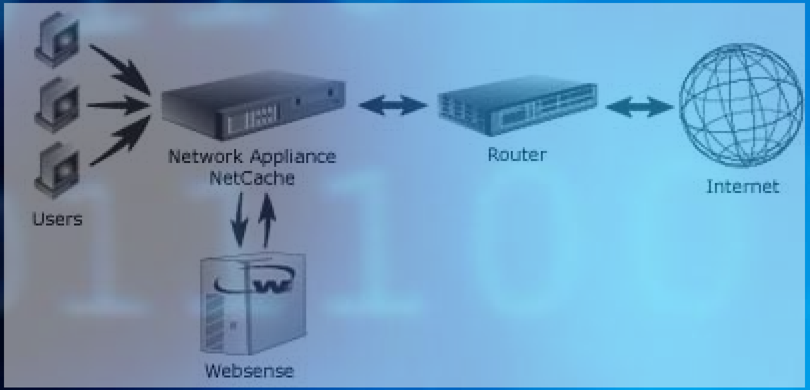
- HTTP Headers
- Blocking Behavior
  - Blockpage
  - Timeout/Redirect
  - DNS Spoofing
  - Keyword Filtering
  - Entire domain? Or specific URL path
- Block list “finger print”
- Network Interrogation

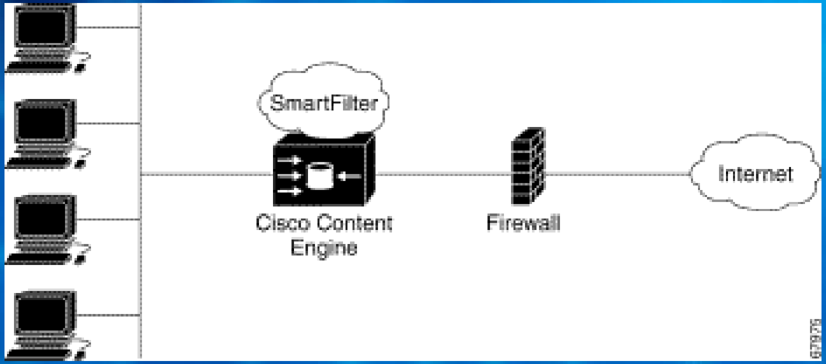


URL is accessible both through the local connection and the remote computer.	🟩
URL is accessible through the local connection but inaccessible through the remote computer, which returned a different HTTP response code.	🟫
URL is accessible through the local connection but inaccessible through the remote computer due to a network connection error.	🟩
URL is accessible through the local connection but inaccessible through the remote computer; a block page was positively identified.	🟪

Filtering Technology

- Blacklist: Deny access to categorized URL/Domains
- Whitelist: Allow access to approved sites, deny all others
- Content Analysis: Dynamically analyze requested content and block by key word (in domain, URL, or body content)





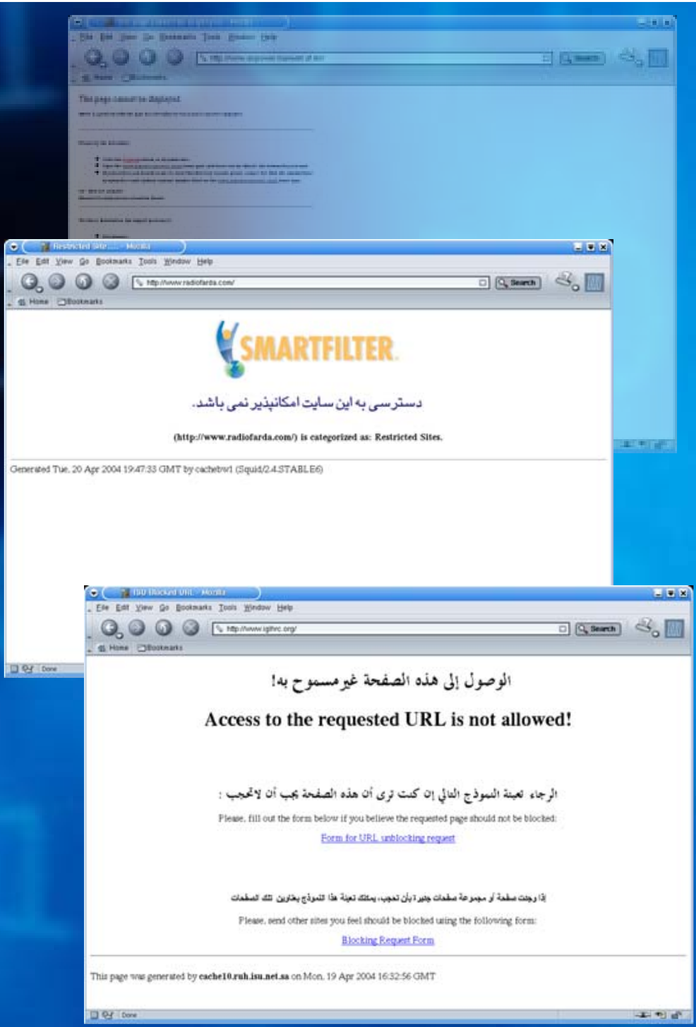
Accessing Filtered Content

Conspicuous

- Block Pages: Indicate that the site is intentionally blocked
- Often combined with block/unblock forms & contact information

Inconspicuous

- Generic timeout, 404 & error pages
- Redirection (possibly to false, look-alike sites)



The page cannot be displayed

http://www.radiofarda.com/

SMARTFILTER

دسترسی به این سایت امکانپذیر نمی باشد.

(http://www.radiofarda.com/) is categorized as: Restricted Sites.

Generated Tue, 20 Apr 2004 19:47:33 GMT by cachebw1 (Squid/2.4.STABLE6)

http://www.iglhrc.org/

الوصول إلى هذه الصفحة غيرمسموح به!

Access to the requested URL is not allowed!

الرجاء تعبئة النموذج التالي إن كنت ترى أن هذه الصفحة يجب أن لاتحجب :

Please, fill out the form below if you believe the requested page should not be blocked:

Form for URL unblocking request

إذا وجدت صفحة أو مجموعة صفحات جديرة بأن تحجب، يمكنك تعبئة هذا النموذج بعناوين تلك الصفحات

Please, send other sites you feel should be blocked using the following form:

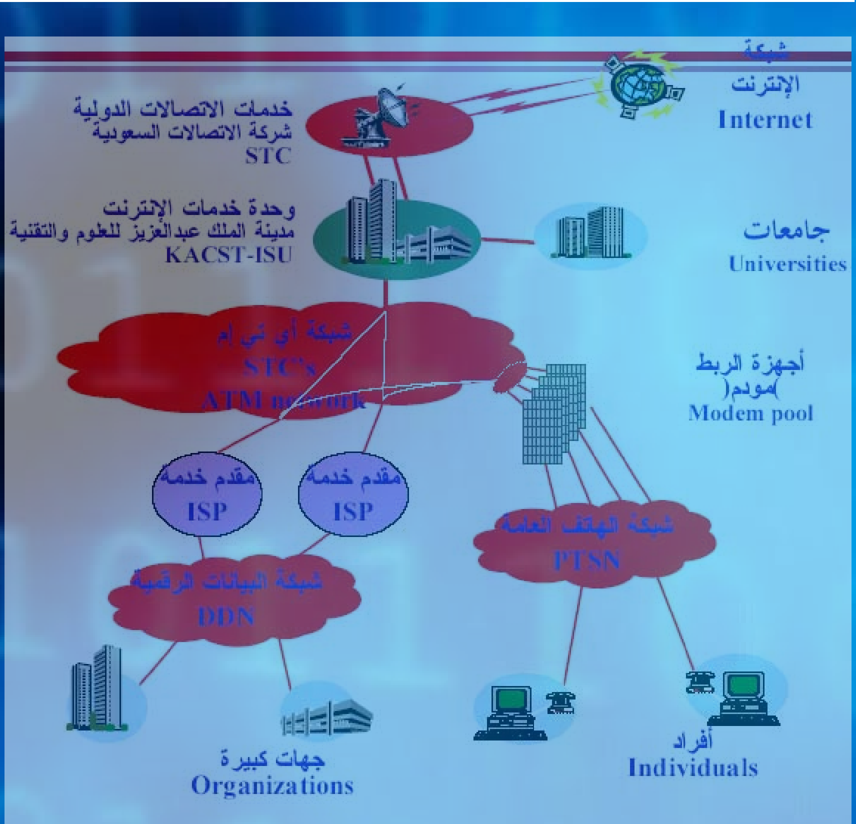
Blocking Request Form

This page was generated by **cache10.ruh.isu.net.sa** on Mon, 19 Apr 2004 16:32:56 GMT

Restricted Site ... - Mozilla
ISU Blocked URL - Mozilla
File Edit View Go Bookmarks Tools Window Help
Search
Home
Bookmarks
Done

Saudi Arabia

- Saudi Arabia
- Centralized Filtering
- Blockpage
- List “fingerprint”
- Secure Computing / SmartFilter
- Collateral Blocking



Source: http://www.isu.net.sa/library/CETEM2001-Zoman.pdf

Collateral Filtering

GET http://www.teenpregnancy.org/teen/ HTTP/1.1

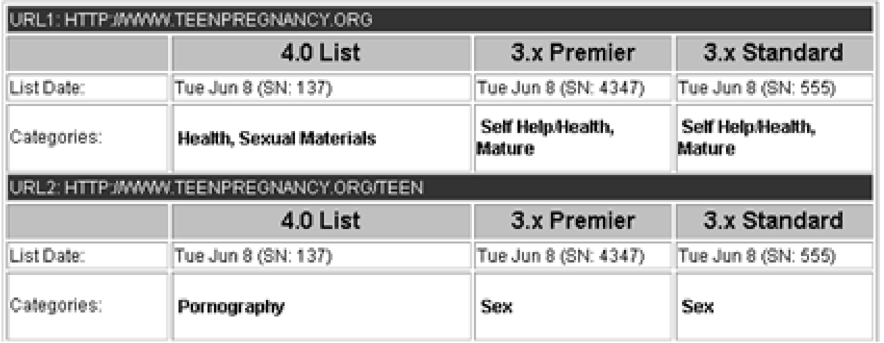
HTTP/1.x 403 Forbidden

<ISUTAG filter="sf" url="http://www.teenpregnancy.org/teen/" date="Thu, 10 Jun 2004" time="18:49:02">

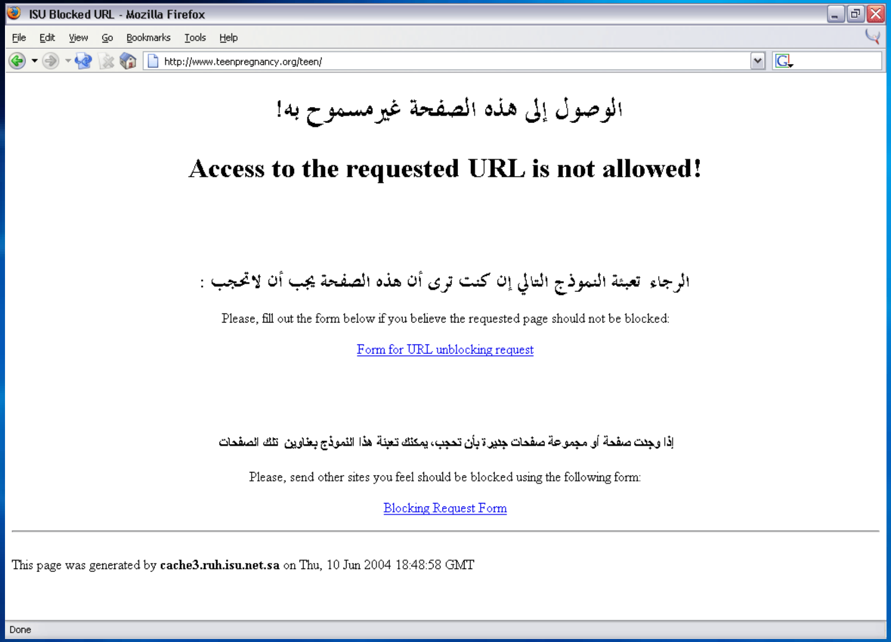
GET http://www.arabtimes.com/ HTTP/1.1

HTTP/1.x 403 Forbidden

<ISUTAG filter="local" url="http://www.arabtimes.com/" date="Thu, 10 Jun 2004" time="18:47:23">



URL1: HTTP://WWW.TEENPREGNANCY.ORG			
	4.0 List	3.x Premier	3.x Standard
List Date:	Tue Jun 8 (SN: 137)	Tue Jun 8 (SN: 4347)	Tue Jun 8 (SN: 555)
Categories:	**Health, Sexual Materials**	**Self Help/Health, Mature**	**Self Help/Health, Mature**
URL2: HTTP://WWW.TEENPREGNANCY.ORG/TEEN			
	4.0 List	3.x Premier	3.x Standard
List Date:	Tue Jun 8 (SN: 137)	Tue Jun 8 (SN: 4347)	Tue Jun 8 (SN: 555)
Categories:	**Pornography**	**Sex**	**Sex**


ISU Blocked URL - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://www.teenpregnancy.org/teen/
Done

الوصول إلى هذه الصفحة غيرمسموح به!

Access to the requested URL is not allowed!

الرجاء تعبئة النموذج التالي إن كنت ترى أن هذه الصفحة يجب أن لاتحجب :

Please, fill out the form below if you believe the requested page should not be blocked:

Form for URL unblocking request

إذا وجدت صفحة أو مجموعة صفحات جديرة بأن تحجب، يمكنك تعبئة هذا النموذج بعناوين تلك الصفحات

Please, send other sites you feel should be blocked using the following form:

Blocking Request Form

This page was generated by **cache3.ruh.isu.net.sa** on Thu, 10 Jun 2004 18:48:58 GMT

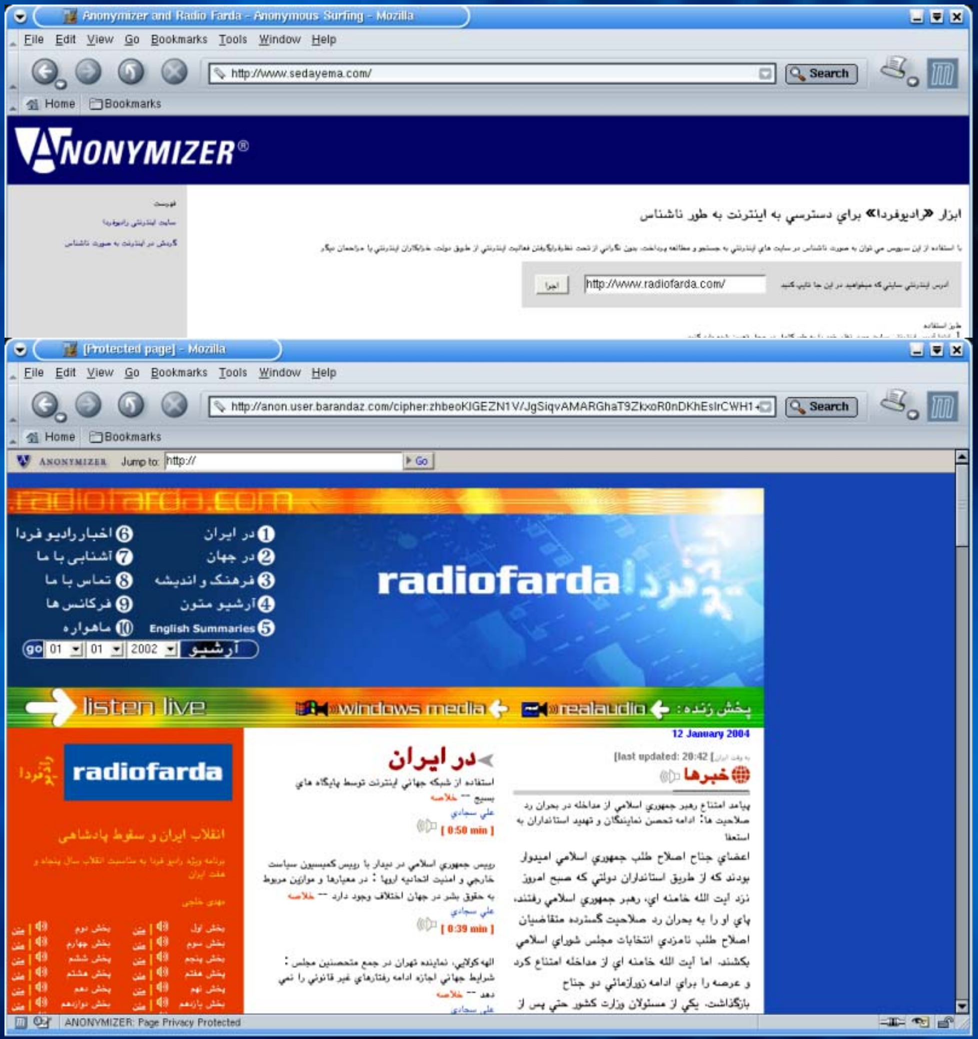
Unintended Consequences

- IP blocking & Virtual hosts
  - USA: Under Pennsylvania state law 1.5 million legitimate websites were blocked while trying to block approximately 400 websites suspected of containing child abuse images.
  - India: Blocked access to all Yahoo Groups in an attempt to block one group
  - China: blocks access to all of Geocities (including premium accounts that have unique domain names)

Circumvention Technology

- Development
  - Psiphon: personal proxy application targeted towards users that have at least one trusted point of contact in a non-filtered country
- Resource
  - On-line clearinghouse project that archives, tests, and assesses anti-censorship, privacy/anonymity, security and encryption software.

Iran: IBB/Anonymizer


Anonymizer and Radio Farda - Anonymous Surfing - Mozilla

http://www.sedayema.com/

ANONYMIZER®

ابزار «راديوفردا» براي دسترسي به اينترنت به طور ناشناس

http://www.radiofarda.com/

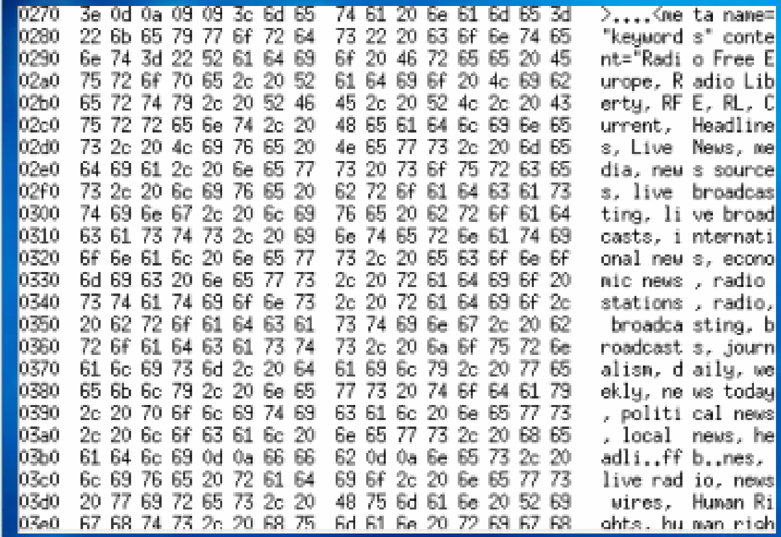
[Protected page] - Mozilla

http://anon.user.barandaz.com/cipher:zhbeoKIGEZN1V/JgSiqvAMARGhaT9ZkxoR0nDKhEslrCWH1

ANONYMIZER Jump to: http://

ANONYMIZER: Page Privacy Protected

- Porn-filtering by keyword in domain
- Although circumvention is successful, the content can still be intercepted (No SSL)



```
0270  3e 0d 0a 09 09 3c 6d 65  74 61 20 6e 61 6d 65 3d  >....<me ta name=
0280  22 6b 65 79 77 6f 72 64  73 22 20 63 6f 6e 74 65  "keyword s" conte
0290  6e 74 3d 22 52 61 64 69  6f 20 46 72 65 65 20 45  nt="Radi o Free E
02a0  75 72 6f 70 65 2c 20 52  61 64 69 6f 20 4c 69 62  urope, R adio Lib
02b0  65 72 74 79 2c 20 52 46  45 2c 20 52 4c 2c 20 43  erty, RF E, RL, C
02c0  75 72 72 65 6e 74 2c 20  48 65 61 64 6c 69 6e 65  urrent,  Headline
02d0  73 2c 20 4c 69 76 65 20  4e 65 77 73 2c 20 6d 65  s, Live  News, me
02e0  64 69 61 2c 20 6e 65 77  73 20 73 6f 75 72 63 65  dia, new s source
02f0  73 2c 20 6c 69 76 65 20  62 72 6f 61 64 63 61 73  s, live  broadcas
0300  74 69 6e 67 2c 20 6c 69  76 65 20 62 72 6f 61 64  ting, li ve broad
0310  63 61 73 74 73 2c 20 69  6e 74 65 72 6e 61 74 69  casts, i nternati
0320  6f 6e 61 6c 20 6e 65 77  73 2c 20 65 63 6f 6e 6f  onal new s, econo
0330  6d 69 63 20 6e 65 77 73  2c 20 72 61 64 69 6f 20  mic news , radio 
0340  73 74 61 74 69 6f 6e 73  2c 20 72 61 64 69 6f 2c  stations , radio,
0350  20 62 72 6f 61 64 63 61  73 74 69 6e 67 2c 20 62   broadca sting, b
0360  72 6f 61 64 63 61 73 74  73 2c 20 6a 6f 75 72 6e  roadcast s, journ
0370  61 6c 69 73 6d 2c 20 64  61 69 6c 79 2c 20 77 65  alism, d aily, we
0380  65 6b 6c 79 2c 20 6e 65  77 73 20 74 6f 64 61 79  ekly, ne ws today
0390  2c 20 70 6f 6c 69 74 69  63 61 6c 20 6e 65 77 73  , politi cal news
03a0  2c 20 6c 6f 63 61 6c 20  6e 65 77 73 2c 20 68 65  , local  news, he
03b0  61 64 6c 69 0d 0a 66 66  62 0d 0a 6e 65 73 2c 20  adli..ff b..nes, 
03c0  6c 69 76 65 20 72 61 64  69 6f 2c 20 6e 65 77 73  live rad io, news
03d0  20 77 69 72 65 73 2c 20  48 75 6d 61 6e 20 52 69   wires,  Human Ri
03e0  67 68 74 73 2c 20 68 75  6d 61 6e 20 72 69 67 68  ghts, hu man righ
```

A Hacktivist Lab

- Bridge the gap between political science and computer science
- Provide social science research with a technical backbone
- Develop a better understanding of the political implications of technology

www.citizenlab.org
www.opennetinitiative.net