Reverse engineering blocking lists

Richard Clayton
richard.clayton@cl.cam.ac.uk

UNIVERSITY OF CAMBRIDGE Computer Laboratory

Atlanta 3rd October 2010

National Physical Laboratory
Basic idea

- Many blocking systems work by “DNS poisoning”
  - usually point at an HTTP proxy which filters exact URL
- Hence you can reverse engineer list of sites (not the URLs) by checking what all hostnames resolve to
- Simple algorithm:
  ```
  for $hostname in (list of all valid hostnames)
      if (resolve(hostname) == cache-IP-address)
          print “hostname is blocked”
  ```
- So all you need is a list of “all valid hostnames” … thanks Eric!
- c 120 million hostnames – although 40 million are DNSBLs etc
- Further clean-up (Facebook &c) reduces to c 70 million hosts
- Takes about 2 days (and 22Gbytes) over home ADSL
Initial results (May 2010, IWF list in UK)

- IWF list held about 450 URLs (says a mole)
- 40% not identified by the methodology (too obscure?)
- 35% clearly (from hostname) intentionally wicked
- Remaining 25% are legitimate “free” hosting sites (etc)

Turkey (Summer 2010)

- Checked against ~20 DNS systems run by a Turkish ISP
- Ran very slowly to avoid any impact (or any detection) so took about 30 days to go through the whole list
  - an earlier run attempting to obtain the Swedish list was barred after a few hours (probably because some DNS servers die when asked to do too much work)
- Mainly “adult” websites and gay material (both forbidden in Turkey), some political bloggers and some “religious” material (such as richarddawkins.com)
- Results passed to Turkish activists... watch this space

- PS: my list of regular expressions for “clean up” is available..
Gratuitous plug!

SATIN 2011 “Securing and Trusting Internet Names”
Workshop on DNS, DNSSEC, DNSCURVE, Passive DNS, DNS as a platform for other services, etc etc (see the CFP)
Looking for both “academic” and “industry” material

to be held at NPL Teddington (London)
Mon/Tue 4/5 April 2011 (week after IETF 80 in Prague)

http://conferences.npl.co.uk/satin/