Content filtering: methods & failures

Dr Richard Clayton

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

London

30th November 2011

How a browser works

- User types in the "URL"
  - **`http://www.example.com/page.html`**
- The hostname is translated into an "IP address"
  - **`www.example.com`** is found to be at **`172.16.17.18`**
  - this is done by a "DNS server" (at your ISP)
- A request is sent to web server address (**`172.16.17.18`**)
  - **`GET page.html`**
  - **`HOST www.example.com`**
- Appropriate page is returned; repeat for embedded images etc.
  - the web server will be at a "hosting company"
  - there may be many websites on one machine (for small sites)
  - or there may be many machines for one website (for big sites)

Simple blocking I (blackholing)

- Block all traffic to the IP address of the website
  - browser cannot connect, and user believes the website doesn’t exist
- Advantages
  - very cheap (although can’t be used for thousands of sites)
- Disadvantages
  - can result in “overblocking” (if many websites at same IP address)
  - assumes that the website has a stable IP address
    - “fast flux” phishing websites change IP address every few minutes; because the IPs they announce just relay traffic to the real website
  - assumes that the DNS tells everyone the same IP address
    - if you can identify the request made when configuring the blocking system you could tell it the wrong address to be blocked (eg the IP address of Google’s search engine)
- Evasion requires an indirect connection to the website
  - use a proxy (anonymous.com), a VPN, or “Tor”

Simple blocking II (DNS poisoning)

- Rig the DNS server so it says the website doesn’t exist
  - alternatively, user can be redirected to an explanatory page
- Advantages
  - very cheap (and scales pretty much indefinitely)
- Disadvantages
  - if done sloppily, can prevent block email for the blocked domain
  - assumes that you know all the names for a website
    - spammers have tens of thousands of names for pharmacy websites
    - using aardvark.aardvark.example.com might work
- Evasion requires using an honest DNS server
  - use 8.8.8.8 (Google’s DNS server)
  - run your own local DNS server

Simple blocking III (proxying)

- Pass traffic through a proxy which checks if URL is “bad”
  - user can be shown an explanation if URL is “bad”
- Advantages
  - can be as fine-grained as you wish (eg just specific image URLs)
  - no overblocking
- Disadvantages
  - far too expensive to send all traffic through the proxy
  - proxy disrupts authentication mechanisms that check source IP
- Evasion requires that traffic avoids inspection
  - use a proxy (anonymous.com), a VPN, or “Tor”
  - use HTTPS (encrypted connection) if website supports it
  - connect on an unusual port number (www.example.com:81)
  - mangle your URL (%70a%67e.html … just look inside email spam!)

Real blocking systems

- UK ISPs use two-stage systems
  - stage one - select traffic that might be going to “bad” site
  - stage two - pass selected traffic through the proxy
- Stage one is done by
  - inspecting the IP address (BT’s “CleanFeed” does this)
  - DNS poisoning (most large ISPs do this)
  - inspecting the traffic as it passes by (smaller ISPs do this)
- Evasion
  - as before – but you get a choice of evading stage one or two!
- The “Great Firewall of China” uses multiple mechanisms
  - blocks some IP addresses completely
  - widespread DNS poisoning
  - traffic inspection for “bad” words; connections are then reset
  - fingerprinting of destinations when traffic is encrypted

Peer to peer traffic

- Peer-to-peer not always blocked, may just be “traffic shaped”
- Originally peer-to-peer traffic used specific port numbers
  - so could tackle all traffic on “port 6000” to any IP address
  - P2P now uses random port numbers (or port 80, the HTTP port)
- Next generation of systems looked inside packets for the peer-to-peer protocol commands
  - so-called “deep packet inspection” (DPI)
  - cleverest systems could determine if payload was copyrighted!
  - so the P2P systems started to encrypt their traffic
- Latest systems look for hints that traffic is peer-to-peer
  - some parts of the protocol still occur “in the clear”
  - connection pattern can be distinctive

Email “spam”

- Email spam is detected (and blocked) by:
  - counting how many similar emails are being seen
  - considering the reputation of the sender
  - considering the pattern of words in the message
  - scoring the use of obfuscating content within the messages
  - considering the reputation of the clickable URL
- So blocking of spam is a completely different realm!
  - people say “but ISPs can block spam” ...
    - yes they can, albeit not 100% accurately
  - ... “and so they can block bad websites”
    - so they can only serve free range eggs in the canteen!
    - i.e. it’s a non sequitor

Webpage labelling

- Idea is that websites rate their content
- Doesn’t scale, and was far too expensive to get right
- ICRA.org now shut down
- DCMS still has their logo, and their tags
  - and still has one page with the word “fuck” on it, rated incorrectly
- Filtering systems actually use low-wage humans to rate pages
  - http://www.ispreview.co.uk/story/2011/10/18/students-responsible-for-deciding-which-adult-websites-uk-isps-block.html
    - "I think it’s a fairly popular job for students. The training is basically going through a number of websites and the various ratings so they get a basic idea. I’m not quite sure how exactly they work, but it would normally be one person who does a rating and one person who double checks it. You could probably start rating websites after one day of seeing various categories. It’s really not that difficult.“ (McAfee)

Blocking is a consensus activity

- ISPs can block material if
  - they concentrate on getting the details right
  - the websites don’t cheat (e.g. by moving around)
  - the users don’t try to evade the blocks!
- Blocking on end-user systems is generally more effective
  - still a consensus activity, but families run on consensus
  - can operate on the content directly
  - can be applied to different protocols (e.g. chat systems)
- BUT if there isn’t consensus
  - you don’t need to be a rocket scientist to follow instructions
  - systems “evolve” to evade blocks (lots of evidence from P2P)
  - blocking in schools has taught the new generation what a proxy is
  - blocking in corporates helps fund VPN sites
  - the “Arab Spring” has put pressure on Tor to be more robust

```
http://www.cl.cam.ac.uk/~rnc1

http://www.lightbluetouchpaper.org
```

Dr Richard Clayton

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

London

30th November 2011