Content filtering: methods & failures

Dr Richard Clayton
How a browser works

• User supplies URL
  – http://www.example.com/page.html

• Domain is translated to “IP address”
  – www.example.com is found to be at 172.16.17.18

• Request is sent to web server (172.16.17.18)
  – GET page.html
  – HOST www.example.com

• Appropriate page is returned; repeat for images etc
Blocking at the ISP
(affects everyone, not just kids)
TAXONOMY

- **DNS poisoning**
  - refuse to resolve the wicked domains
  - low cost, and highly scalable

- **Blackhole routeing**
  - refuse to carry the traffic to the wicked site
  - low cost, but limits to the number of possible rules

- **Proxy filtering**
  - refuse to serve the wicked pages
  - high cost, and all traffic has to be inspected

- **Deep packet inspection**
  - spot “bad traffic passing by” and discard (or send resets)
  - expensive especially at high bandwidth (but used for Great Firewall of China and for proprietary P2P filtering)
Problems with DNS poisoning

• Apparently easy…

@ IN SOA localhost. root.localhost. ( 2004010100 86400 3600 604800 3600 )

@ IN NS localhost.

@ IN A 127.0.0.1

* IN A 127.0.0.1

• But getting it right for subdomains and for email requires some thought! Dornseif found that every German ISP he studied had made errors!
Problems with blackhole routeing

• Dropping packets will (obviously) affect every website hosted at the IP address!
  – hence useless for geocities.com or lycos.com
  – in fact useless for huge numbers of other sites as well. Edelman study found “overblocking” a significant issue: 87.3% of com/net/org sites share IP address with at least one other; 69.9% with at least 50 others (and a continuum exists at all sizes)
  – do you really want to block the “Romanian Tourist Board” website?
Problems with proxy filtering

• This method avoids overblocking (huzzah!)
• However, it can have significant costs in equipment, in customer satisfaction and in network reliability
  – economic justifications for caching proxies continue to get weaker
  – proxies often slower than going direct!
  – caching proxies obstruct many personalisation schemes for website content providers
Problems with packet inspection

• Traffic may be encrypted (or otherwise obscure)
• Resets can be just ignored
  – often hard to inspect in real time, resets can be sent when decision on acceptability of traffic known
• Deals with more than just HTML, but other protocols are far more fluid and (in case of P2P) rapidly evolve to avoid the blocking.
Avoidance for clients

• Use a different DNS server
• Use IP addresses
• Use a relay (often encrypts and anonymises)
• Encode request%73 to avoid recognition
  – look at your spam to see this raised to an art form
• Send malformed HTTP requests
  – eg: multiple HOST protocol elements
Avoidance for servers

• Move site to another IP address (easy)
• Change port number (hard to discover)
• Provide same content on many different URLs
• Accept unusually formatted requests
  – servlets at client could obfuscate or encrypt so that an intermediary has no chance of using anything short of the IP address to identify content
CleanFeed

• Part of BT “anti-child-abuse initiative”
  – two stage (hybrid) system, BT, June 2004
  – similar designs deployed by other ISPs
• First stage is IP address based
  – candidate traffic for blocking is redirected
• Second stage matches URLs
  – redirected traffic passes through a web proxy
• Best of both worlds?
  – accurate, but low cost because #2 is low volume
Design of CleanFeed
Fragility of Cleanfeed

• Evading either stage evades the system
  – all previous attacks continue to be relevant

• PLUS can attack the system in new ways
  – the credulous will fail to notice Google (or iTunes) IP addresses in DNS results for wicked sites and will flood the second stage with legitimate traffic
  – the clueless will fail to spot local IP addresses in DNS results and construct routing loops
The oracle attack

- Detect the redirection by the first stage by seeing what traffic reaches the second
- Send `tcp/80` packets with TTL set to 8, see what then comes back:
The oracle attack

BT's Network

The Internet

SYN/ACK

ICMP
The oracle attack

• Detect the redirection by the first stage by seeing what traffic reaches the second
• Send tcp/80 packets with TTL set to 8, see what then comes back:
  – ICMP time exceeded means no redirect
  – RST (or SYN ACK) means redirect to proxy
• Then use a suitable database to get domain names, eg: whois.webhosting.info
Oracle attack results

~~~.~~~.191.40    lolitaportal.****
~~~.~~~.191.42    no websites recorded in the database
~~~.~~~.191.49    samayhamed.****
~~~.~~~.191.50    amateurs-world.****
                       anime-worlds.****
                       boys-top.****
                       cute-virgins.****
                       cyber-lolita.****
                       egoldeasy.****
                       elite-sex.****
                       ... and 26 more sites with similar names

NB: missing names probably .ru or outdated database
NB: dodgy names on .41 .43 ... BUT no IWF “endorsement”
NB: It is illegal for me to check the ACTUAL contents
Politics

• Blocking was considered “impossible” until BT deployed CleanFeed

• ISPA claim 80% of consumers covered by systems that block illegal child images

• Minister now wants all of (broadband) industry to be blocking by the end of 2007
  – voluntary except: “If it appears that we are not going to meet our target through co-operation, we will review the options”
Whitehall comprehension?

- “Recently, it has become technically feasible for ISPs to block home users’ access to websites irrespective of where in the world they are hosted”

- In my view, doubtful that they understand the cost, fragility or ease of evasion of these blocking systems, let alone the reverse engineering of the blocking lists.
Other uses?

• Fratini (EU) wants Internet to be a “hostile environment” for terrorists
  – “I think it’s very important to explore further possibilities of blocking websites that incite to commit terrorist action”

• Drugs, gambling, holocaust denial…

• and don’t overlook civil cases:
  – such as, defamation, copyright material, industrial secrets, home addresses of company directors, lists of MI6 agents…
Summary

• Four basic ways of blocking content
• All have problems and can be evaded
• Hybrid systems can be lower cost, but have some extra problems as well
• Government signalling that blocking of sites on IWF list to become *de rigeur*
• Top of a very slippery slope for us all
Blocking at the end user
(can be very user-specific)
Filtering software

• Most products are for web pages and chat
• Mix blacklists and keyword detection
  – hence whitelist for when keywords fail
• Parental overrides depend on passwords...
• Australian system turned off in minutes
  – and you can just copy the tricks…
• http://www.peacefire.org/ (bit dated)
  – “you’ll understand when you’re younger”
Lemons

• Quality (and “hidden agenda”) of products not easy to determine; nor is age accuracy

• Kite mark has been in process of development for several years – which may remove some of the weaker products from the marketplace.

• HEAnet (Irish school) filtering has 85% approval from primary schools, 57% from secondary schools (+ want teacher override)
Avoidance

• Blacklists may be avoided by URL obfuscation (%73 etc) depending on software design (and quality)
• Proxy sites may avoid blocks altogether
• Keyword filtering fine for fixed pages, but useless for chat (euphemisms become as offensive as the words they replace – frak!)
Webpage labelling

• Originally based on video games ratings
• Concepts apply badly on web, and even revised they are extremely crude
• In practice, honest rating is extremely expensive and webmasters not interested
• 99.99% of web (and growing) is unrated
• DoH thinks “fuck” is not bad language!
Consent/self-censorship

• Becoming clear that the major way in which the Great Firewall of China works is that people censor themselves...

• … lesson undoubtedly applies to end-user filtering systems. Even the smartest kids may be prepared to leave system on for most of the time... (IANAP)