Portals, the deep web and the dark web

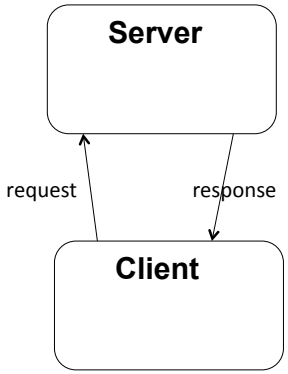
CSC9UB2

Content

- The surface web vs. the deep web
  - Or what you can get by searching using Google
    Vs.
  - What's there, but you can't it get by Googling
- Why the deep web exists
  - And why the deep web is *neither* scary *nor* illegal
- Why the dark web exists
  - And why maybe it is illegal and scary!

The surface web

- The original way of using the web
- Web pages held on a server
- Searchable, findable by searching robots
  - AKA spiders
  - Simply traversing the web by following links



Hiding things on the web

- Lots of what's on the web is not findable by searching robots
  - Note that we are talking about material that *is* web-accessible in some way
- Dynamic content
- Private Web
- Scripted content
- Non-HTML/text content
- Unlinked content
- Actively hidden content: the Dark web

Dynamic content

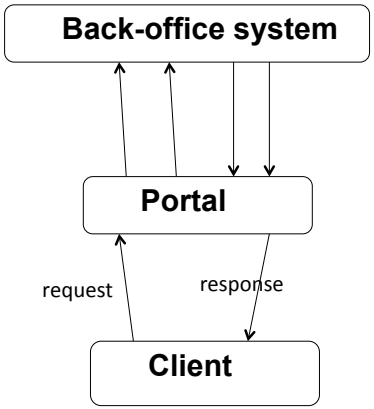
- Many web sites actively (dynamically) create the content that is displayed
- Using JavaScript, PHP and other client and server-side software.
  - CSCU9B2 pages, for example
- Why?
  - Generating appropriate lists for e-Commerce
    Personalised pages
  - Responding to queries
- Clearly, such material is not searchable by search engines.
  - Nothing secretive about this at all: how could an e-Commerce site keep all the pages it might need?

Private Web

- Private web sites are websites that have controlled access
  - Usually password protected
- Clearly, not accessible to web-searching spiders
- Many examples exist
  - Repositories of data, catalogues of libraries, sites with academic papers, thousands of others.
  - Very large amount of data

Portals

- Portals
  - May create responses dynamically
  - Often give access to large volumes of data
  - Generally require registration
    - Username & password protected
    - Not searchable



Portal examples

- Someone generally used (in the past weeks)
  - Stirling university student/staff portal
  - CARMEN Neuroscience portal
  - EPSRC Extranet
  - JISC Web of Science
  - JES (research councils website)
    - Plus some e-Commerce sites
- Very large volumes of data
  - Petabytes. Much bigger than the surface internet!

Scripted content

- Pages that are only accessible
  - through links produced by JavaScript
  - as well as content dynamically downloaded from Web servers via Flash or Ajax solutions.
- Generally used in conjunction with dynamic web content.

Non-HTML/text content

- Search engines rely on being able to read the content of pages
- So even if they can find the page,
  - Pages that are unreadable
    E.g. encoded, or non-textual, or sound or images etc.
  - … are not searched
- This can be quite reasonable
  - Would you want to search an mp3 file for text?
- But can be abused
  - Code something to make it look like an mp3 file
    - But it actually contains something else

Steganography

- See http://en.wikipedia.org/wiki/Steganography
- Concealing a message inside another file
  - E.g. text in a video, or in an mp3 file
- Can be very effective
  - If you have no idea that anything is hidden in a file then you won't look for it there
  - Security through obscurity
- Actually quite easy to add extra bits on to certain types of files
  - .wav files, uncompressed image files are good candidates.
    - Vary the least significant bit of a colour, or a loud sound

Encrypted text

- … or of course, the text (or images or sounds) can simply be encrypted.
- But then it looks like encrypted text
  - So someone might try to decode it!
- Or one can mix encrypting with steganography
  - Hard to find, and then hard to decode!

Unlinked content

- If you create a web page
  - But ensure that nothing ever links to it
    - Orphan page
- … then no search engine can find it.
- Doesn't need to have a name
  - URL might be http://<N>.<N>.<N>.<N>:<pno>/
- That is, an IPv4 address, and a port number
  - Something similar can be done for IPv6 addresses as well

Actively hidden content: the Dark web

- Why so people want it?
- The surface internet is easily traced
  - IP numbers used can be traced to specific machines, or networks
- If companies/people/organisations do not want to be traced
  - For economic, legal or political reasons
- … but do want to use the web as a technology…

The Dark web

- (some use the dark net = deep net: here I am considering the dark net as a small part of the deep net)
- Connections are made only between trusted peers
  - Private peer-to-peer networks
  - Sometimes called "friends" (F2F)
    - Peer-to-peer network, password protected
    - Users make connections to people they know
    - There may be others on the net as well …
  - Not necessarily nefarious.
  - Often using non-standard protocols and port numbers
  - See https://en.wikipedia.org/wiki/List_of_TCP_and_UDP_port_numbers

Software for the dark net

- Freenet is a popular darknet (friend-to-friend) by default
- GNUnet is a popular darknet if the "F2F (network) topology" option is enabled
- Retroshare run as a darknet (friend-to-friend)
- Routing:

  The onion router: TOR

The onion router (1)

- Free software to enable anonymous communication
- Core principles go back to the 1990's
  - U.S. Naval Research Laboratory employees, mathematician Paul Syverson and computer scientists Michael Reed and David Goldschlag,

To protect U.S. intelligence communications online.
Onion routing was further developed by DARPA in 1997.

The onion router (2)

- Tor directs Internet traffic through a free, worldwide, volunteer network consisting of more than six thousand relays
- To conceal a user's location and usage from anyone conducting network surveillance or traffic analysis.
- NSA described it as
  "the King of high-secure, low-latency Internet anonymity" with "no contenders for the throne in waiting"
- See http://en.wikipedia.org/wiki/Tor_(anonymity_network)

Searching the deep web

- Some of the deep web is searchable
  - Even if not by Google
- BrightPlanet has a technology that can be used
  - They harvest huge quantities of data
  - And use it as a business resource for companies etc.
- See
  http://www.brightplanet.com/deep-web-university-2/video/

What's on the dark web?

- Quite a lot of illegal stuff.
- See https://en.wikipedia.org/wiki/Dark_web
- Image from http://www.sickchirpse.com/deep-web-guide/