What is the Internet doing to me?
(witidtm 2022/2023 - TEU00311)

Lab Sessions

Stephen Farrell
stephen.farrell@cs.tcd.ie

https://github.com/sftcd/witidtm
https://down.dsg.cs.tcd.ie/witidtm

Initial Goals

- Sort yourselves into groups of 2-3
- Login to SCSS account on lab machine and/or get your laptop online
- Fire up a “new”/“clean” browser
- Explore settings, esp security/privacy related
- Watch HTTP traffic (shift-ctrl-I)
- Find the “worst” site you can
  - NSFW disallowed!!!
  - You define/justify “worst”
  - Report back from group

Stretch goal

- If we get the above done great, if not, that’s ok
- Could be some people get to the stretch goal today or we look at it in future...
- Stretch goal: HTTP archive (.har) file generation and a bit of analysis

Login/get-online

- Desktop logins: no “domain”, use SCSS password, not your TCD password (if those differ), so e.g. if your TCD email is bloggsj14@tcd.ie then you enter “bloggsj14” as the username
- Get online: we’ll deal with things as they arise
- After you’re done: see if anyone else needs help

I did check that username... :-)

“Your message to bloggsj14@tcd.ie couldn't be delivered.
A custom mail flow rule created by an admin at tcdud.onmicrosoft.com has blocked your message.
5.1.1 The e-mail service at tcd.ie does not know this email address.”

Go do stuff!

Play with “new” browser

- Don’t use one that has e.g. stored credentials for some account you care about – basically don’t muck up your daily-driver setup
- Lab machine browsers can probably be reset easily enough (TBC)
- On own laptop: maybe install one you’ve not used before
  - Possibles: Firefox, edge, vivaldi, brave, opera...
  - More exist, but start being careful if you go beyond the above as esp. less widely used browser downloads have been known to contain malware from time to time (but mostly on phones)

Browser settings

- Defaults are important and not always what *you* would want!
- Play about in settings and see what you find – ask if not sure or comment if you think you should tweak
- As they differ a bit, you might want do this for all browsers you use: e.g. lab machine, laptop, phone
- Things to check/set:
  - Telemetry
  - Search engine/Search suggestions
  - Locations/Camera/ Microphone/Notification permissions
  - DNS over HTTPS (DoH)
  - Cookies
  - Blockers/Tracking protection
  - Logins/passwords
  - Javascript/NoScript
  - Site data
  - Clear things on exit

Watch web traffic

- Open browser
- Type shift-ctrl-I (or equivalent) to open developer interface
- Re-size screens to taste
- Choose “network” tab in developer pane
- Try loading a few sites and watch what happens
  - DO NOT load NSFW sites!
- Look about on the web and decide which site is the “worst” from your POV and why
  - Just yell/put up hand when you have a “worst” to nominate
  - We’ll pick a winner if we’ve time – Prize == applause:-)

Shift-ctrl-I for macs...

- Macs differ:
  - Firefox: Option + Command + I
  - Safari: Option + Command + C
  - Chrome: Option + Command + C
- Access to developer tools in Safari has to be activated in the settings first. If anyone has problems with that they can find detailed information on how to do this here: https://support.apple.com/en-ie/guide/safari/sfri20948/mac
- Thanks to Luca Schäfer (2021 student) for the above

Stretch goal

- Figure out how to save an HTTP archive file (.har)
- Figure out how to view .har files
- Figure out how to diff .har files
- See what changes between seemingly identical browser sessions
  - ...any tracking?

Go do stuff!

More goals

- Find the location of an image
- Consider what facial recognition means for us
- Stretch goal: minimal image manipulation to defeat recognition

- DO NOT use any image that has a reasonable probability of upsetting anyone

Images and the Internet

- Why are details of images relevant to this module?

Why are details of images relevant to this module?

- We upload lots of images
- Automated image capture is near ubiquituous
  - CCTV, ANPR, ...
- Other people upload images of us
- Organisations with image databases analyse those

Data vs. Metadata

- Typically we talk about the “data” as being the main thing being processed or communicated or stored...
  - E.g: the bits rendered for an image or video, the content of an email, the messages in a text chat or the audio packets in a voice call
- So-called “metadata” is also data but is “about” the above rather than part of the above
  - E.g. timing of a communication, sender/receiver IP addresses, the size of data fields, etc
- Even if data is well-protected (e.g. encrypted), metadata can leak separately (or be deliberately stored/exposed) so meta-data creates risk
- For someone surveilling, metadata can be more attractive than data, e.g. law enforcement may benefit more from building a social graph of criminals compared to seeing the content of a few messages, or, facebook might learn enough from whatapp metadata that they no longer need to see the content to sell advertising
  - Metadata is also often more structured and hence easier to process for those who want to see what people are doing
- Metadata can also be a little unexpected to end-users, e.g. author information in documents, or, in images...

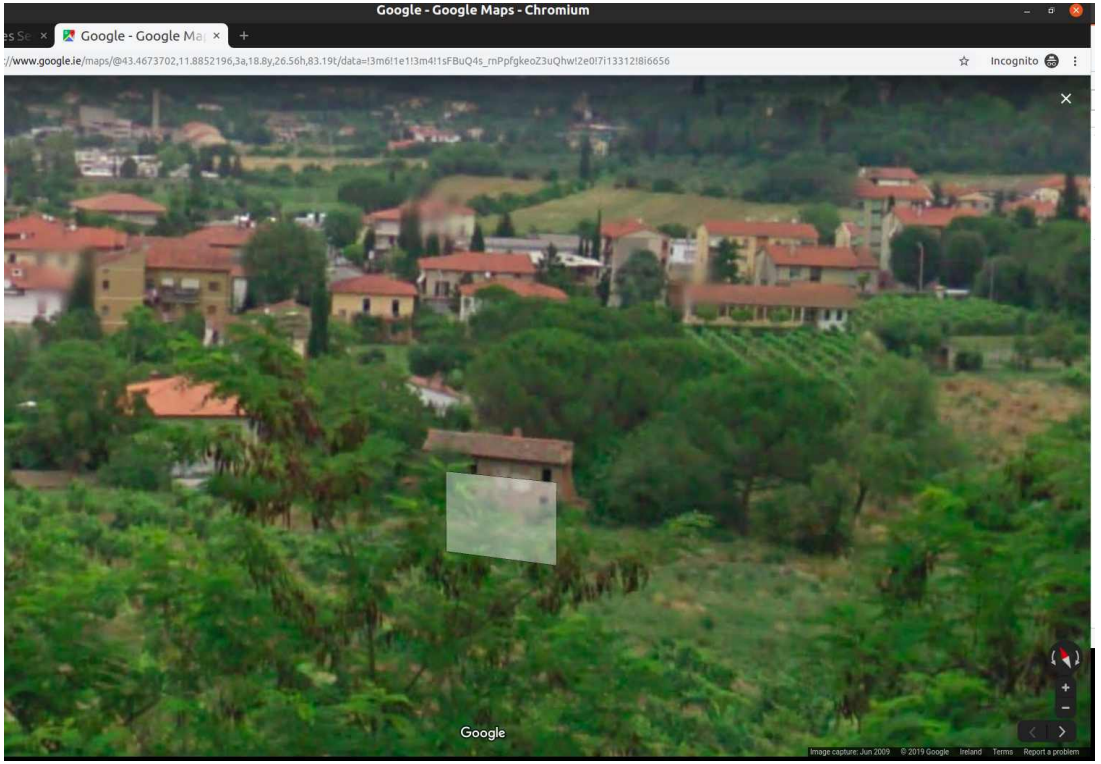
Image Metadata: EXIF

https://photographylife.com/what-is-exif-data
https://helpdeskgeek.com/how-to/how-to-remove-exif-data-from-your-photos/

```
exif DSCN0010.jpg
EXIF tags in 'DSCN0010.jpg' ('Intel' byte order):
--------------------
+-------------------------------------------------------------
Tag                |Value
--------------------
+-------------------------------------------------------------
Image Description  |
Manufacturer       |NIKON
Model              |COOLPIX P6000
Orientation        |Top-left
X-Resolution       |300
Y-Resolution       |300
Resolution Unit    |Inch
Software           |Nikon Transfer 1.1 W
Date and Time      |2008:11:01 21:15:07
YCbCr Positioning  |Centered
Compression        |JPEG compression
X-Resolution       |72
Y-Resolution       |72
Resolution Unit    |Inch
Exposure Time      |1/75 sec.
F-Number           |f/5.9
Exposure Program   |Normal program
ISO Speed Ratings  |64
Exif Version       |Exif Version 2.2
Date and Time (Origi|2008:10:22 16:28:39
Date and Time (Digit|2008:10:22 16:28:39
Components Configura|Y Cb Cr -
Exposure Bias      |0.00 EV
Maximum Aperture Val|2.90 EV (f/2.7)
Metering Mode      |Pattern
Light Source       |Unknown
Flash              |Flash did not fire, compulsory flash mode
Focal Length       |24.0 mm
Maker Note         |3298 bytes undefined data
User Comment       |
FlashPixVersion    |FlashPix Version 1.0
Color Space        |sRGB
Pixel X Dimension  |640
Pixel Y Dimension  |480
File Source        |DSC
Scene Type         |Directly photographed
Custom Rendered    |Normal process
Exposure Mode      |Auto exposure
White Balance      |Auto white balance
Digital Zoom Ratio |0.00
Focal Length in 35mm|112
Scene Capture Type |Standard
Gain Control       |Normal
Contrast           |Normal
Saturation         |Normal
Sharpness          |Normal
Subject Distance Ran|Unknown
North or South Latit|N
Latitude           |43, 28, 2.81400000
East or West Longitu|E
Longitude          |11, 53, 6.45599999
Altitude Reference |Sea level
GPS Time (Atomic Clo|14:27:07.24
GPS Satellites     |06
GPS Image Direction |
Geodetic Survey Data|WGS-84
GPS Date           |2008:10:23
Interoperability Ind|R98
Interoperability Ver|0100
-------------------+---------------------------------------------------------
EXIF data contains a thumbnail (6702 bytes).
```

https://raw.githubusercontent.com/ianare/exif-samples/master/jpg/gps/DSCN0010.jpg

52100 Arezzo, Province of Arezzo, Italy



- Took about 5 minutes to find this in Google street view
- Most of that was finding a way to map degree, minutes, seconds to fractional Lat,Long
- All **automatable**, could easily produce location history from a set of images
- How could such a “leak” be damaging to you or to someone else in your images?

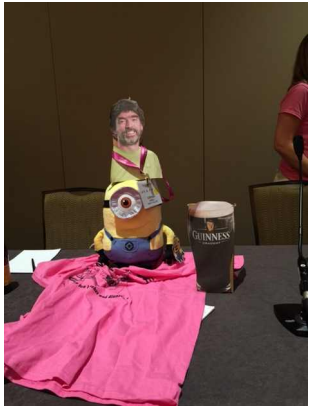
https://www.google.ie/maps/@43.4673702,11.8852196,3a,75y,23.31h,83.05t/data=!3m6!1e1!3m4!1sFBuQ4s_rnPpfgkeoZ3uQhw!2e0!7i13312!8i6656

Viewing EXIF Data

- Local: Right-click and “properties”
- Better local: install something allowing you to scan multiple images
  - E.g. “`sudo apt install exiftool`” in Linux
- On web: save image to local then GOTO above

My Pictures

- I scanned the 27,733 files in my Pictures directory and found 499 of those contained EXIF GPS lat/long values
  - Earliest was from 2009, most recent from 2019
  - Latitidue ranged from 67° N, to 3° S
  - Longitude ranged from 157° W, to 67° E
    - The 157° W pic: a conference in Hawaii I attended remotely;-(



- On the linux command line that looked like:

```
$ exiftool -gpslatitude -gpslongitude -createdate -csv -r ~/Pictures
```

Uploads?

- What happens when you upload photos to web sites?
- It varies... a 2013 test showed the results on the right
- That's probably changed for the better, but be aware that YMMV, so test what you use, if interested
- Worth noting that some people, esp. Photo enthuiasts, do want all the EXIF data preserved/shared

what do social sites do with the gps location on your uploaded photos?

	Upload Source					
	Computer			Phone		
	Remove	Hide	Display	Remove	Hide	Display
facebook	X			X		
Twitter	X			X		
Google +		X			X	
Linked in	X			X		
myspace	X			X		
Pinterest			X	X		
Instagram	X					
flickr from YAHOO!		X			X	
tumblr.			X	X		
Blogger	X			X		
WordPress			X	X		

source: GPS for Today

https://www.gpsfortoday.com/what-social-networks-protect-your-exif-and-gps-location-data-from-other-users/

Your EXIF task...

1) Find some image(s) online or locally
2) Determine if they contain EXIF location data
3) Find the location of that image in e.g. Google street view
4) As time permits: GOTO 1

- What can you infer from the above?
- What could you infer if you did the above for a number of images of related subjects?

Go do stuff!

Facial recognition

- A kind of “biometric” (more later on the imperfections of biometrics:-)
- Nice overview, including tricky issues at: https://en.wikipedia.org/wiki/Facial_recognition_system (accessed 20210927)
- Basic idea: program analyses image bits, search for pattern that looks like a face (eyes, nose, mouth, ...), classifies that (based on machine learning using image collections), compare results from two images – if close enough, declare match
  - False positives and negatives will happen
- Note: this is not my area of expertise!

Facial recognition (ab)uses

- Find a photo of “this person”
  - Find local pics of your mum, organise your image gallery
- Find people with outstanding arrest warrants in a crowd
  - Recognition of faces in moving crowd is harder than individually, but likely, not that much harder
- Determine ethnicity of people using public transport
  - Critics may say things like the above - proponents might talk about improving efficiency but build systems that have this effect

Your facial recognition task

- GOTO https://www.kairos.com/demos
  - I’ve no opinion of that system, other than that it offers the comparison I wanted for the lab
  - Hopefully it doesn’t stop working on us (e.g. because we used it too much;-)
- Play with various image pairs, with/without the same person visible, to try understand how well/badly this particular face verification works
  - Hint: a web search for images of a well known figure (politician, musician, ...) should produce a fairly good range of images of the same subject
- If you can do such comparisons some other way, great, but do tell us about it

- What do you infer about images uploaded to web sites or “the cloud”?
- What do you infer about images **you** capture or upload?

Stretch goal

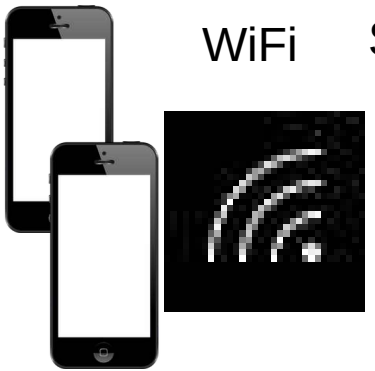
- Find an image that matches itself when tested with your facial recogniser (easy:-)
- Pick an image editor
  - My suggestion: gimp, https://www.gimp.org/
  - But there may be simpler options
- Try find the “smallest” change (not perceptible to a human) that causes matching to fail
- General topic: adversarial images
  - https://davideliu.com/2020/05/27/introduction-to-adversarial-attacks-on-images/ accessed 20210927
  - Same concepts apply to other machine learning settings, e.g. text, audio, ...

Go do stuff!

Yet Moar goals

- Join my hotspot
  - SSID: witidtm-lab
  - WPA Password: bal-mtditiw
- We’ll watch some traffic using wireshark, and chat about that
- Stretch goal: do it yourself at home if interested

Hotspot setup



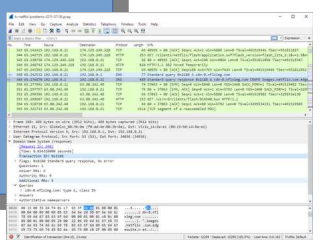
WiFi SSID: witidtm-lab WPA Password: bal-mtditiw

Turris Home Router

Doing (lab) WiFi, DHCP, DNS,
NAT and forwarding
Conncected upstream to my laptop

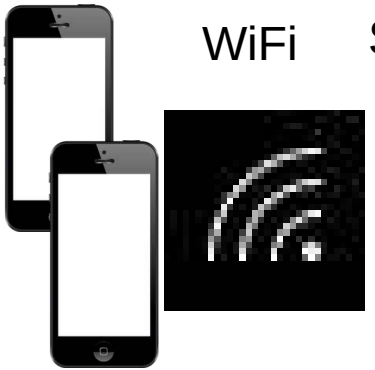
My laptop

Forwarding, DHCP for Turris
DNS for hotspot WiFi clients
Running Wireshark
Conncected upstream to college
(via college WiFi)



TCD network

Hotspot setup



WiFi SSID: witidtm-lab WPA Password: bal-mtditiw

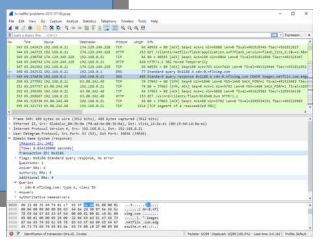
Turris Home Router

Doing (lab) WiFi, DHCP, DNS,
NAT and forwarding
Conncected upstream to my laptop

All your traffic are belong to us!

My laptop

Forwarding, DHCP for Turris
DNS for hotspot WiFi clients
Running Wireshark
Conncected upstream to college
(via college WiFi)



TCD network

Wireshark

- Fine tool to observe network traffic
  - https://www.wireshark.org/
- The network-side equivalent of what you saw in a browser with shift-ctrl-I
- But not limited to web traffic, you see it all as it is “on the wire”

Some traffic

- Display wireshark and see what’s visible

MAC Addresses

- Device-tracking often (ab)uses **long-term hard-coded identifiers** such as MAC addresses (or IMEI/IMSI in mobile n/w)
- MAC address: link layer address (mostly) hardcoded to radio or other network chip
  - Same form of address used in WiFi and most other network protocols at link layer, e.g. Bluetooth
  - Roughly: how two devices on the same local area network (LAN) identify one another
- Looks like “6C:9C:ED:87:27:60” (48 bits) - 1st half is manufacturer ID (Cisco), 2nd half device-ID (a WiFi router in TCD SCSS)
  - You can look up manuf IDs from the registry, e.g. https://www.adminsub.net/mac-address-finder/84:C7
- MAC address is often fixed for the lifetime of the device; There is now a 64-bit version, not sure how widely used yet
  - You can probably see these in the “about device” tab or similar

Randomised MACs

- MAC address randomisation is a good idea and starting to be deployed
  - Often, the MAC address only really needs to be stable for a session, so can be randomised
  - But – if you paid for the hotel WiFi that might be based on your MAC address, or an enterprise network might use MAC addresses to decide which machines are allowed on the local network, or the machine may be a switch/router/server where changing MAC address would break stuff or be inefficient
- So you can’t always randomise, and doing so well needs higher-layer controls
- HOWTO turn on varies by OS and version
- On an android 10 phone I used have:
  - Developer options/Enhanced Wi-Fi MAC randomisation
  - You may need to turn on developer options first (search for HOWTO)

Some traffic

- Display wireshark and see what’s visible

DHCP (1)

- Dynamic Host Configuration Protocol (DHCP) is (almost always) how your device gets an IPv4 address after joining a network
  - Spec is RFC2131 from 1997
- DHCP has a “hostname” option that client’s send and that has often been the same as a long-term device name, e.g. “Stephen’s iPhone”
  - That can be used for a loooong time
  - Recent phone OSes tend to send something more random looking but often don’t vary the value
  - If you migrate from an old to a new device, that setting might carry over even if your current OS would otherwise use a random value
  - Changing hostname could break stuff though so be a bit careful
- Other DHCP options clients send can also be identifying e.g. OS version

DHCP (2)

- As well as returning an IP address, the DHCP server can send many other options
  - Most aren’t widely deployed
- The DNS server IP address option is though
  - Tells clients what server to use for DNS in this network
- OSes can override that but mostly (so far) don’t
  - Unless you have DoT configured or some Apple stuff

Router Admin

- Demo OpenWRT/Foris/Luci

DNS names

- Today, we can mostly see the DNS names being queried as that's cleartext
- Starting to see more use of encrypted DNS traffic
- Two flavours: DoT and DoH
- We'll look quickly at that but chat more about it later
- On that old crappy android device enabling DoT used be:
  - Settings/Connections/More Connection Settings/Private DNS
  - BUT that won't work with college as upstream as DoT uses port 853 and college block that port – it should work fine at home and with your mobile data provider

Do53 vs DoT vs DoH

- Do53 == old style cleartext DNS
- Who do you want to/care about seeing your DNS traffic?
  - Your ISP, TCD, coffee-shop and their ISP, Cloudflare, Google ?
- Pros and cons to each of these

Brave Browser

- Not a bad browser on mobiles
- Has various “shields up/down” settings
  - Settings/Brave Sheilds and Privacy/Use Secure DNS
  - DoH – that does work in college
    - There may be a set of known services from which to choose
    - Or you can add a custom one:
      - For cloudflare try: https://1.1.1.1/dns-query
  - DoH can also be set in most other browsers these days too

Some traffic

- Display wireshark and see what’s visible

Stretch Goal

- Repeat this at home if interested
  - Can help with, but not mandatory for, my assignment
- Setup hotspot using laptop
- Install wireshark
- Inspect traffic and learn

Fake a real web site

- https://highscalability.com/ re-directs to an insecure http:// URL -- **BAD PRACTICE!!!**
- Inside my laptop I'm running a web server
- I can "easily" re-direct the traffic for the insecure URL to that but let the secure stuff go through

```
# start lighttpd
$ ./testlighttpd.sh
$ sudo sysctl -w net.ipv4.conf.all.route_localnet=1
$ sudo iptables -t nat -A PREROUTING -i enxa0cec80097d6 \
    -p tcp -d 172.67.173.147 --dport 80 \
  -j DNAT --to-destination 127.0.0.1:8099
$ sudo iptables -t nat -A PREROUTING -i enxa0cec80097d6 \
  -p tcp -d 104.21.30.199 --dport 80 \
  -j DNAT --to-destination 127.0.0.1:8099
```

Lab Conclusions...

- What do you conclude?
  - <add stuff here>