

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

# Lab Session #1

Stephen Farrell  
[stephen.farrell@cs.tcd.ie](mailto:stephen.farrell@cs.tcd.ie)

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

# Goals this hour

- Introduce yourselves
- Login to SCSS account and/or get your laptop online
- Fire up a “new” browser
- Explore settings
- Watch HTTP traffic (shift-ctrl-I)
- Find the “worst” site you can
  - NSFW disallowed, otherwise you define/justify “worst”

# 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.”

# Lab: 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: 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)

# Lab: watch http 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!
- Say 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



# Lab: 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?

What is the Internet doing to me?  
(witidtm 2021/2022 - TEU00311)

## Lab Session #2

Stephen Farrell  
[stephen.farrell@cs.tcd.ie](mailto:stephen.farrell@cs.tcd.ie)

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

# Today's 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 ubiquitous
  - 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 of an image or video, the content of an email, the messages in a text chat or the audio packets in a phone call
- So-called “metadata” is also data but is “about” the above rather than part of the above
  - E.g. the timing of a communication, the sender/receiver IP addresses, the size of data, etc
- Even if data is well-protected, 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 whatsapp metadata that they no longer need to see the content to sell advertising
- Metadata can also be a little unexpected, 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/>

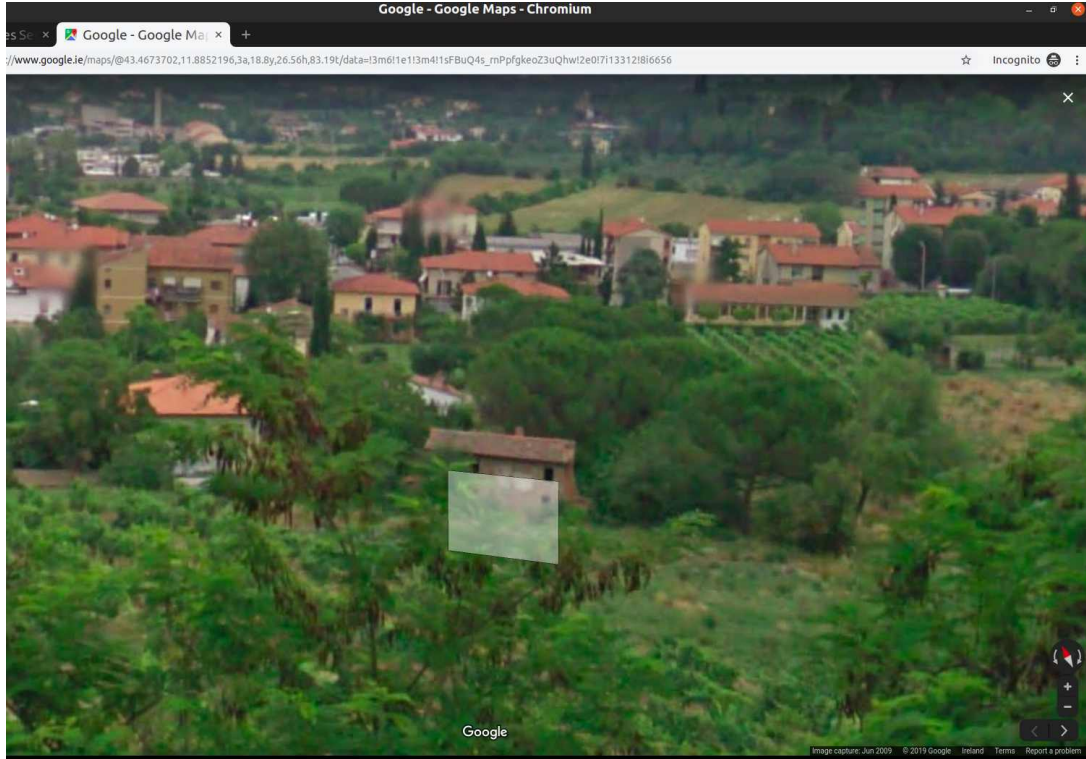


```
exif DSCN0010.jpg
EXIF tags in 'DSCN0010.jpg' ('Intel' byte order):
-----
Tag                |Value
-----
Image Description   |
Manufacturer        |NIKON
Model               |COOLPIX P6000
Orientation         |Top-left
X-Resolution        |300
Y-Resolution        |300
Resolution Unit     |Inch
Software            |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 (Orig)|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 Cl)|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?



# Viewing EXIF Data

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

# 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) If time remains: 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?

# 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](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, ...

What is the Internet doing to me?  
(witidtm 2021/2022 - TEU00311)

## Lab Session #3

Stephen Farrell  
[stephen.farrell@cs.tcd.ie](mailto:stephen.farrell@cs.tcd.ie)

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

# Today's goals

- Join my hotspot
  - SSID: witidtm-lab
  - WPA Password: bal-mtditiw
- We'll watch some traffic wireshark, and chat about a few ways to affect that
- Stretch goal: repeat this at home if interested



# Hotspot setup

SSID: witidtm-lab

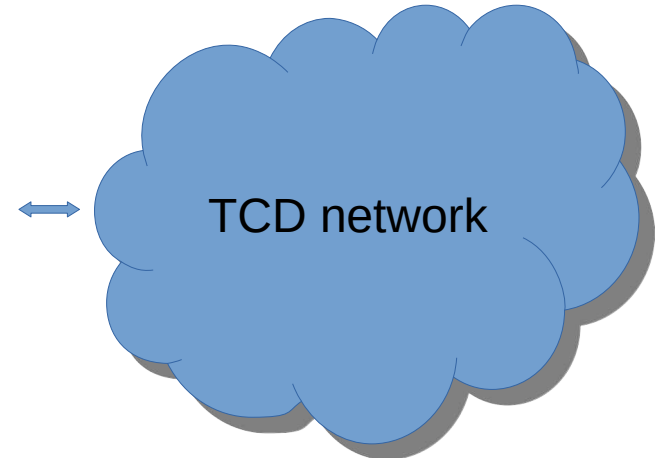
WPA Password: bal-mtditiw

My laptop

Running WiFi Hotspot  
Running Wireshark  
Connected upstream to college  
(also via WiFi)

The image shows a Wireshark packet capture window. The top pane displays a list of captured packets. Packet 77 is a DNS query from 192.168.0.21 to 192.168.0.1. The bottom pane shows the details of packet 77, which is a Standard query request for the domain 'bal-mtditiw.com'. The query is for an A record. The packet structure is as follows:

```
Frame 77: 100 bytes on wire (800 bits), 80 bytes captured (640 bits) on interface 0  
Internet Protocol Version 4, Src: 192.168.0.1, Dst: 192.168.0.21  
User Datagram Protocol, Src Port: 53, Dst Port: 53  
Domain Name System (request)  
Standard query type A  
Transaction ID: 0x12345678  
Flags: Standard query request, no error  
Questions: 1  
Answer RRs: 0  
Authority RRs: 0  
Additional RRs: 0  
Query  
  qname: bal-mtditiw.com type A, class IN  
  qtype: A  
  qclass: IN  
  qttl: 0  
  rdata: 192.168.0.1
```



# 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-l
- But not limited to web traffic, you see it all
- Let's try look at some

# 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: layer 2 address (mostly) hardcoded to radio or other network chip
  - Same form of address used in WiFi and most other network protocols at layer 2, 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) - 1<sup>st</sup> half is manufacturer ID (Cisco), 2<sup>nd</sup> half device-ID (a WiFi router in TCD SCSS)
  - You can look up the manuf ID in the registry: <https://www.adminsub.net/mac-address-finder/84:C7> gives a list that includes Sony
- 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 have:
  - Developer options/Enhanced Wi-Fi MAC randomisation
  - You may need to turn on developer options first (search for HOWTO)

# 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 my android crappy device:
  - 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

# Brave Browser

- Not a bad browser on mobiles
- Has various “shields up/down” settings
  - Settings/Brave Shields 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>

# 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

# 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



What is the Internet doing to me?  
(witidtm 2021/2022 - TEU00311)

## Lab Session #4

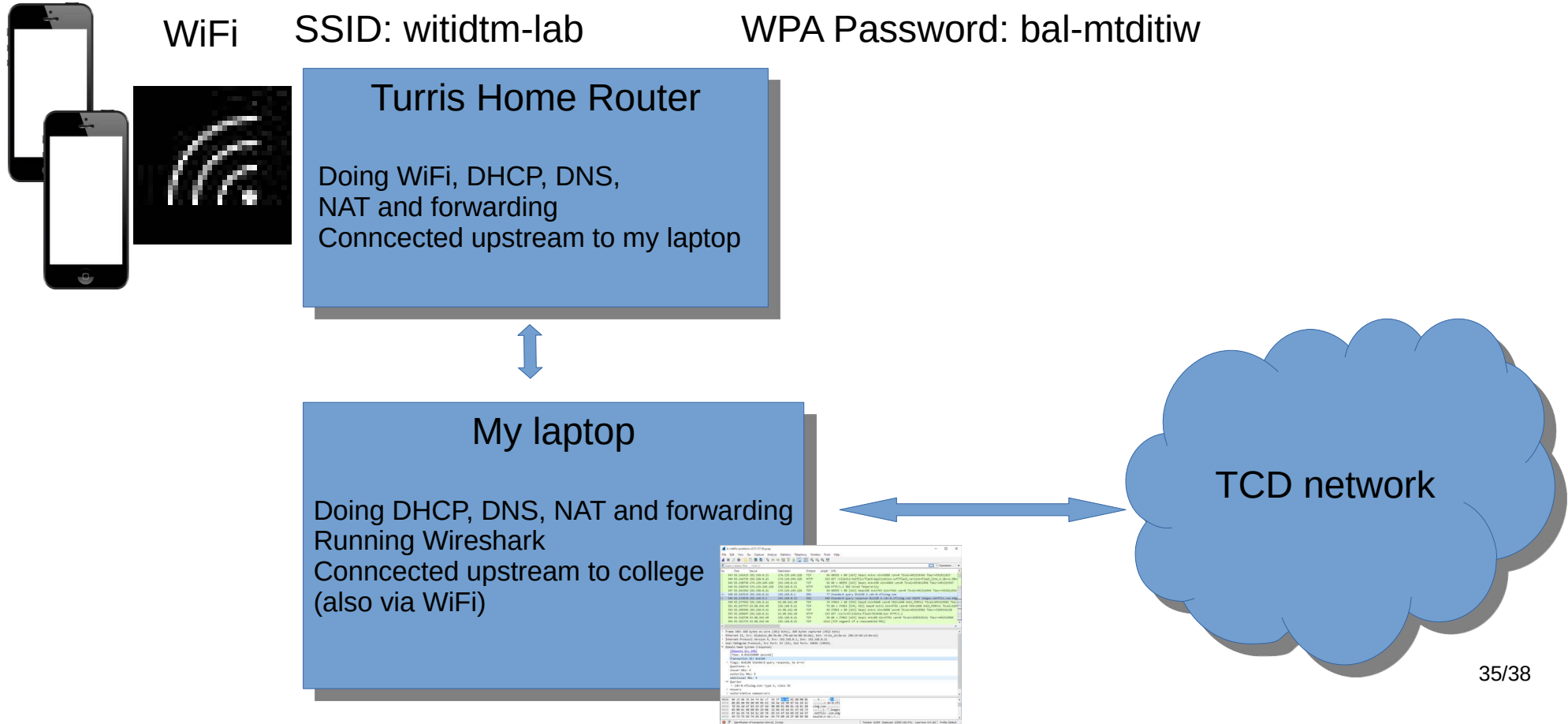
Stephen Farrell  
[stephen.farrell@cs.tcd.ie](mailto:stephen.farrell@cs.tcd.ie)

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

# Goals

- Working “hotspot” :-)
  - SSID: witidtm-lab
  - WPA Password: bal-mtditiw
- Chat about home router configuration
- Watch some traffic

# Hotspot setup



# Router Admin

- Demo OpenWRT/Foris/Luci

# Some traffic

- Display wireshark and see what's visible

# 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
```