Hic Sunt Proxies: Unveiling Proxy Phenomena in Mobile Networks

Raffaele Zullo, Antonio Pescapé, Università di Napoli Federico II, Italy
Korian Edeline, Benoit Donnet, Université de Liège, Belgium

<r.zullo@studenti.unina.it>, <pescape@unina.it>, <firstname.name@ulg.ac.be>

This work has been partially funded by the European Union's Horizon 2020 research and innovation program under grant agreement No 688421. The opinions expressed and arguments employed reflect only the authors' views. The European Commission is not responsible for any use that may be made of that information. The work of Antonio Pescapé has been partially supported by the art. 11 DM 593/2000 for NM2 srl.

Roadmap

- Proxies
- Mobile Tracebox
  - Rooted mode
  - Non-rooted mode
- Measurement campaign
- Proxy detection
  - Observed scenarios
- Results
  - Proxy prevalence and scope
  - Proxy characterization
    - Transport layer behavior
    - Impact on the Receive Window
    - Impact on extensions
    - Location
    - Fingerprinting
- Conclusions

Proxies

- TCP-terminating proxies
- Packet-rewriting proxies
- Transparent
- Detection methodologies
- Extensively deployed in mobile networks

- Transport layer behavior?
- Scope limited to HTTP or a few TCP services?
- Impact on initial and maximum Receive Window?
- Up-to-date with new extensions (e.g. TCP Options, TCP ECN)?
- Location inside the network?
- Limitations of detection methodologies?
- New detection methodologies?


Mobile Tracebox

- **Rooted mode**[1]
  - **Raw sockets**
    - Every single field of IP and TCP headers can be set
  - **Server-based** probes
  - **Traceroute** probes

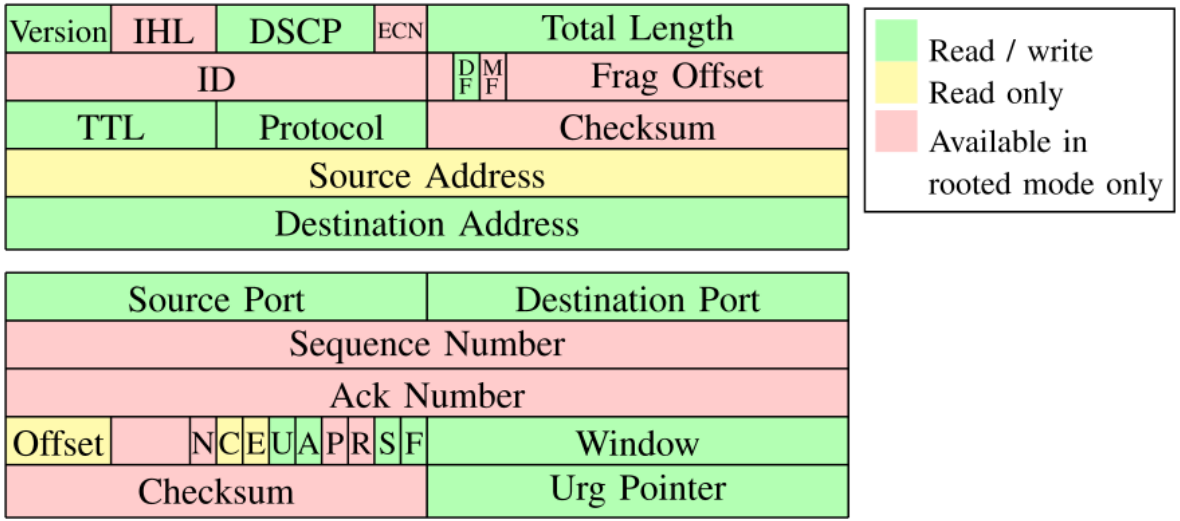
[1] Zullo et al., “Hic Sunt NATs: Uncovering Address Translation with a Smart Traceroute”, MNM 2017

Mobile Tracebox: non-rooted mode

- **Non-raw sockets**
  - `bind(), setsockopt(), connect(), etc.`
  - IP and TCP header fields
    1. **Can be directly set or manipulated**
    2. Can't be manipulated but **the default value can be retrieved**
    3. Can't be tested in non-rooted mode
- **Server-based** probes
- **Pseudo-traceroute** probes
  - Direct access to ICMP messages requires root
  - Path length can still be measured
  - `connect()` iteratively called with incremental TTL
- **Non-responding server** probes


Mobile Tracebox: non-rooted mode

IP and TCP header of TCP probe packets forged in non rooted mode



Mobile Tracebox: non-rooted mode

- TCP Options on Syn probe packets forged in non rooted mode

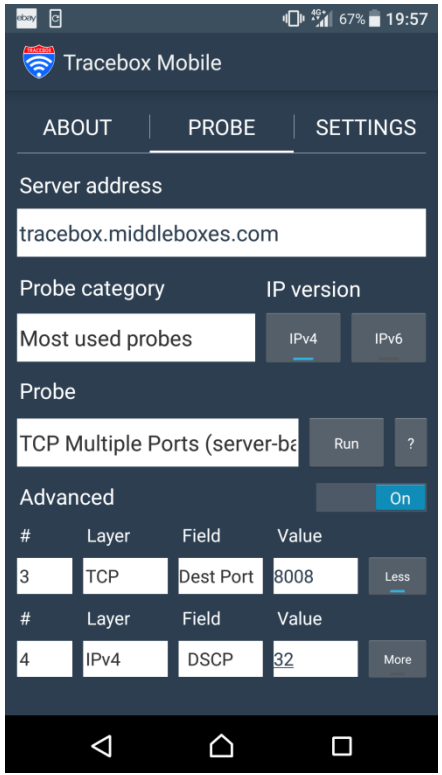
Option	Kind	Value	Note
MSS	2	Read / write	
WS	3	Read / write	
SP	4	-	
TS	8	No access	
TFO	254 (Exp.)		Android 6.0 or later later

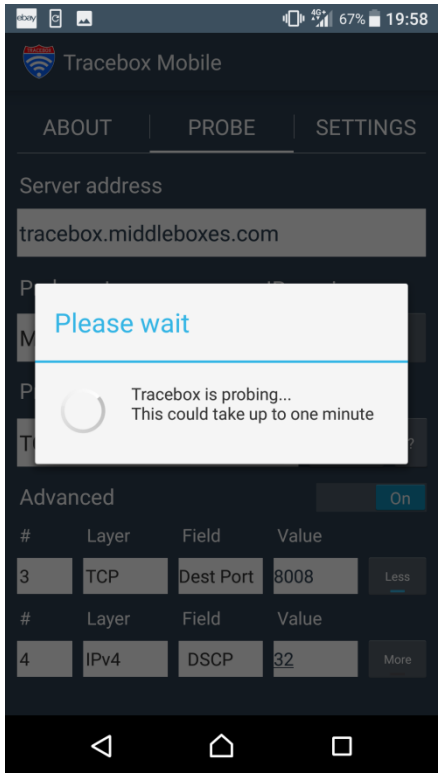
- Syn and Payload packets

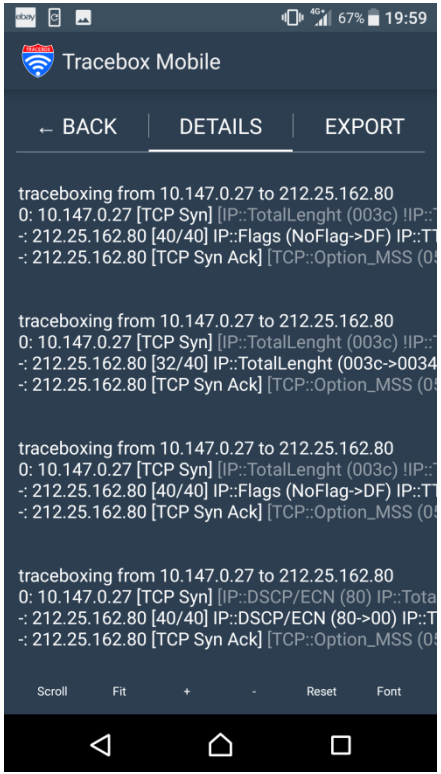
- Port tested:
  - HTTP (port 80)
  - FTP (21)
  - SMTP (25)
  - HTTPS (443)
  - SIP (5060)
  - Non-standard port (10000)
  - Custom

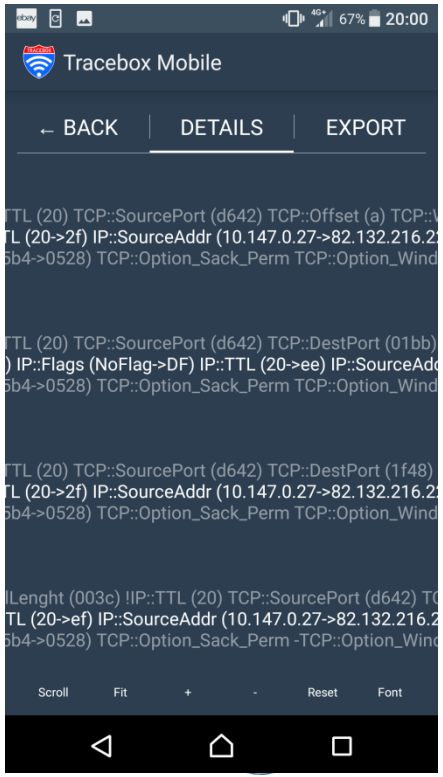

Mobile Tracebox: non-rooted mode

https://play.google.com/store/apps/details?id=be.ac.ulg.mobiletracebox









Mobile Tracebox: non-rooted mode

```
 0:  10.147.0.27     [TCP Syn]      IP::TotalLenght (003c) !IP::TTL (20) TCP::SourcePort (d642) TCP::Offset (a) TCP::Window (ffff) TCP::Option_NOP (+1) …
 : 212.25.162.80  [40/40]           IP::Flags (NoFlag->DF)  IP::TTL (20->2f)  IP::SourceAddr (10.147.0.27->82.132.216.220)  TCP::Window (ffff->7210)
                                    TCP::Option_WindowScale (08->05)
 : 212.25.162.80  [TCP Syn Ack]     TCP::Option_MSS (05b4->0528)  TCP::Option_Sack_Perm  TCP::Option_WindowScale (05)  TCP::Option_Timestamp

 0:  10.147.0.27     [TCP Syn]      IP::TotalLenght (003c)  !IP::TTL (20)  TCP::SourcePort (d642)  TCP::DestPort (01bb)  TCP::Offset (a)  TCP::Window (ffff) …
 :  212.25.162.80  [32/40]          IP::TotalLenght (003c->0034)  IP::Flags (NoFlag->DF)  IP::TTL (20->ee)  IP::SourceAddr (10.147.0.27->82.132.216.220)
                                    TCP::Offset (a->8) TCP::Window (ffff->1ffe)  TCP::Option_MSS (05b4->0528)  -TCP::Option_Timestamp  TCP::Option_NOP (+2)
 :  212.25.162.80  [TCP Syn Ack]    TCP::Option_MSS (05b4->0528)  TCP::Option_Sack_Perm  TCP::Option_WindowScale (08)  -TCP::Option_Timestamp

 0:  10.147.0.27     [TCP Syn]      IP::TotalLenght (003c)  !IP::TTL (20)  TCP::SourcePort (d642)  TCP::DestPort (1f48)  TCP::Offset (a)  TCP::Window (ffff) …
 :  212.25.162.80  [40/40]          IP::Flags (NoFlag->DF)  IP::TTL (20->2f)  IP::SourceAddr (10.147.0.27->82.132.216.220)  TCP::Window (ffff->7210)
                                    TCP::Option_WindowScale (08->05)
 :  212.25.162.80  [TCP Syn Ack]    TCP::Option_MSS (05b4->0528)  TCP::Option_Sack_Perm  TCP::Option_WindowScale (05)  TCP::Option_Timestamp

 0:  10.147.0.27     [TCP Syn]      IP::DSCP/ECN (80)  IP::TotalLenght (003c)  !IP::TTL (20)  TCP::SourcePort (d642)  TCP::DestPort (0015)  TCP::Offset (a) …
21: 212.25.162.80 [40/40]           IP::DSCP/ECN (80->00) IP::TTL (20->0c) IP::SourceAddr (10.147.0.27->82.132.216.220) TCP::Option_MSS (05b4->0528)
21: 212.25.162.80  [TCP Syn Ack]    TCP::Option_MSS (05b4->0528)  TCP::Option_Sack_Perm -TCP::Option_WindowScale -TCP::Option_Timestamp
```

Measurement campaign

- 30 months after first release non-rooted version
- 800+ downloads
- Due to the nature of our analysis, we excluded:
  - Networks tested in rooted mode
  - Networks tested in non-rooted mode but without preserving the default settings on the analyzed fields
- **96 Cellular networks**
- **385 Wi-Fi networks**
- **69 Countries**
- **6 Continents** (no Antarctica)

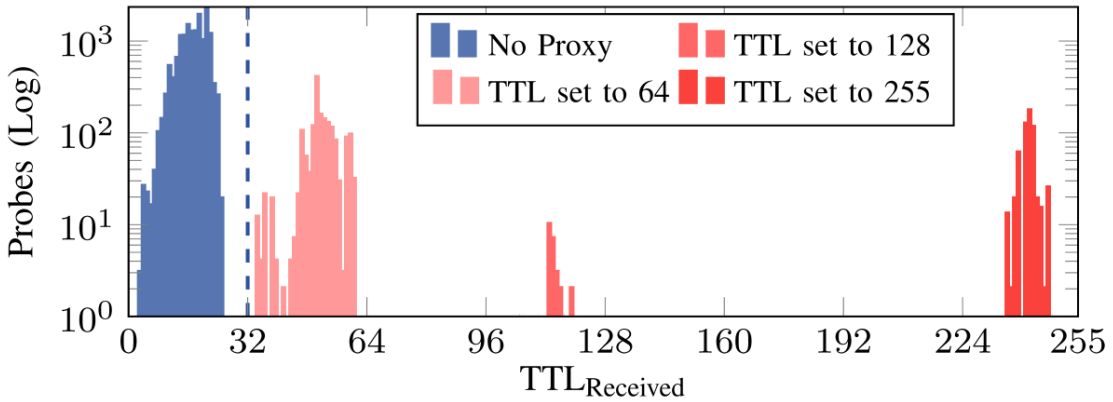
- Rooted probes from a subset of these networks (17 cellular and 32 Wi-Fi networks)
  Only to delve into a few specific facets


Proxy detection

- TCP-terminating proxy
  - Splits a TCP connection in two parts
    - Proxy's Syn carries proxy's default TTL regardless of client's Syn TTL
    - Potential inconsistence in measured path length
  - Pseudo-traceroute probes:
    - Path length measured as the number of hops till responding node
      - Anomalous values, especially ranging from 1 to 5 hops, invariant to different destination addresses
  - Server-based probes
    - Path length measured as difference between sent packet's TTL and received packet's TTL
      - Anomalous values: received packet's TTL higher than sent packet's TTL


Proxy detection through TTL rewriting

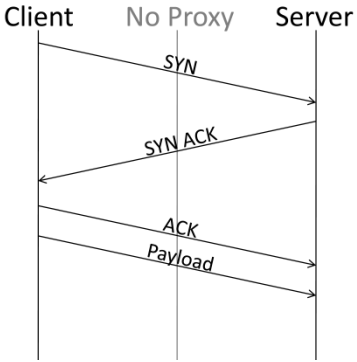
- Mobile Tracebox default TTL: 32 and 64
  - Only probes with initial TTL of 32 are eligible
- TTL distribution of received probe packets sent with an initial TTL of 32



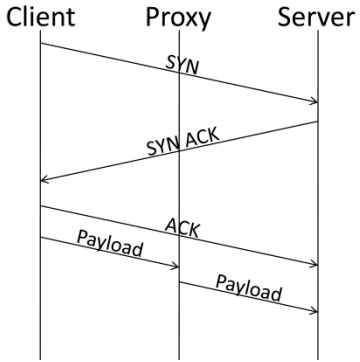
- Further validation
  - Path length distribution of traceroute probes targeting our server suggests that portion of paths with a length >32 is negligible
  - Packets sent with initial TTL of 64 were received with a TTL >= 34, sign that no proxy using a default TTL of 30-32 was observed
  - The only values outside of the 4 regions were recorded during roaming (roaming probes excluded in data cleansing phase)
  - 64 is the most common proxies' TTL

Scenarios observed

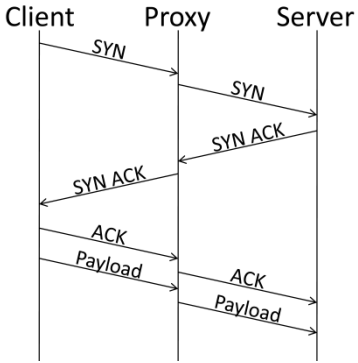
- Previous classification can be extended to Payload packets other than Syn packets
  - Crucial to detect if payload modifications are due to TCP-terminating proxies or packet-rewriting proxies
  - It reveals a class of proxies that don't alter TTL on Syn but rewrite it on payload packets
- Non-responding server test can instead discriminate whether a proxy altering Syn responds with a Syn Ack immediately or after Syn Ack reception from the destination



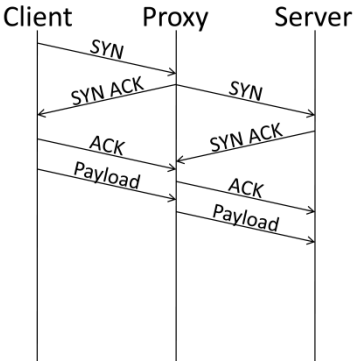
No proxy



Proxy not altering Syn



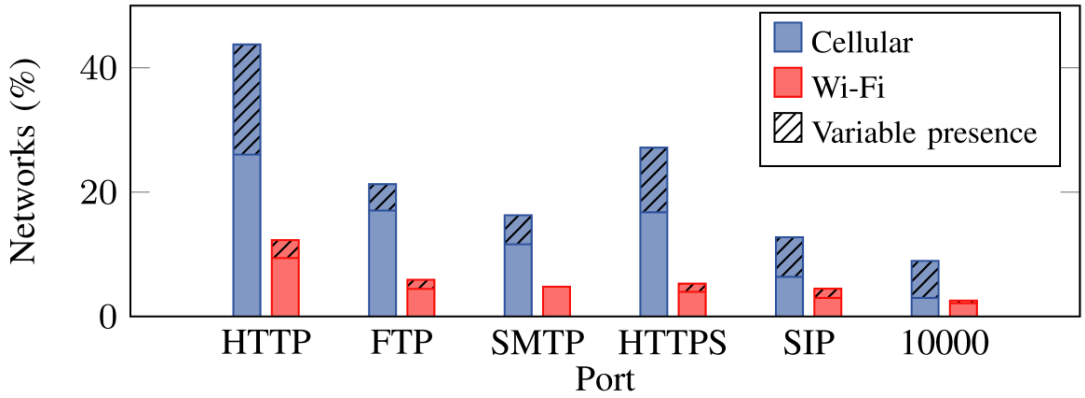
Proxy altering Syn
(no early Syn Ack)



Proxy with early Syn Ack

Proxy prevalence

- Proxy detection on different TCP Ports



- Only HTTP proxy
- Only specific non-HTTP proxy
- HTTP and other known services (e.g. HTTP, HTTPS and FTP) proxies
- Proxy revealed on all tested ports
  - Networks where **all** traffic is routed through a proxy

Proxy prevalence

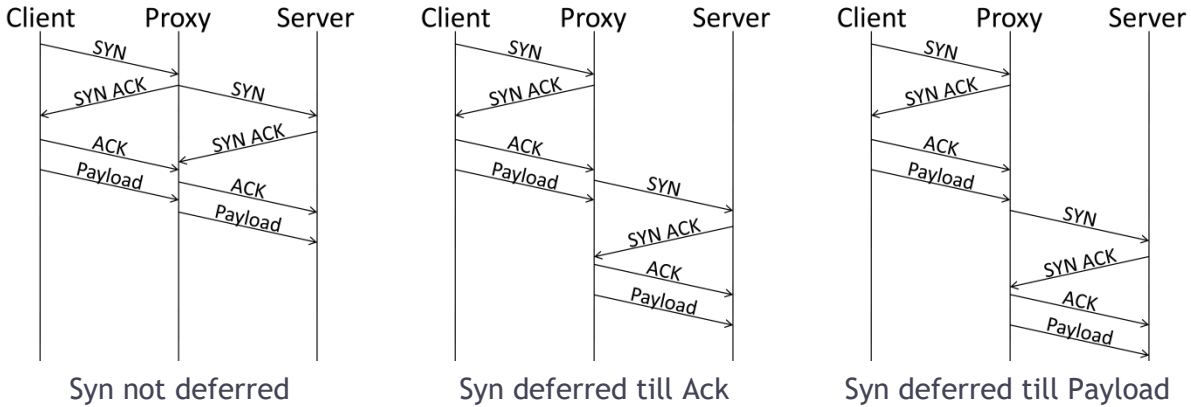
- Packet rewriting proxies
  - 1 cellular network
  - 2 Wi-Fi networks
- TCP-terminating proxies

Scenario	Cellular	Wi-Fi
Syn not altered	14%	15%
Syn altered, no early Syn Ack	4%	8%
Syn altered, early Syn Ack	82%	77%

Early Syn Ack proxy is most common typology in both categories of mobile networks

Deferred Handshake

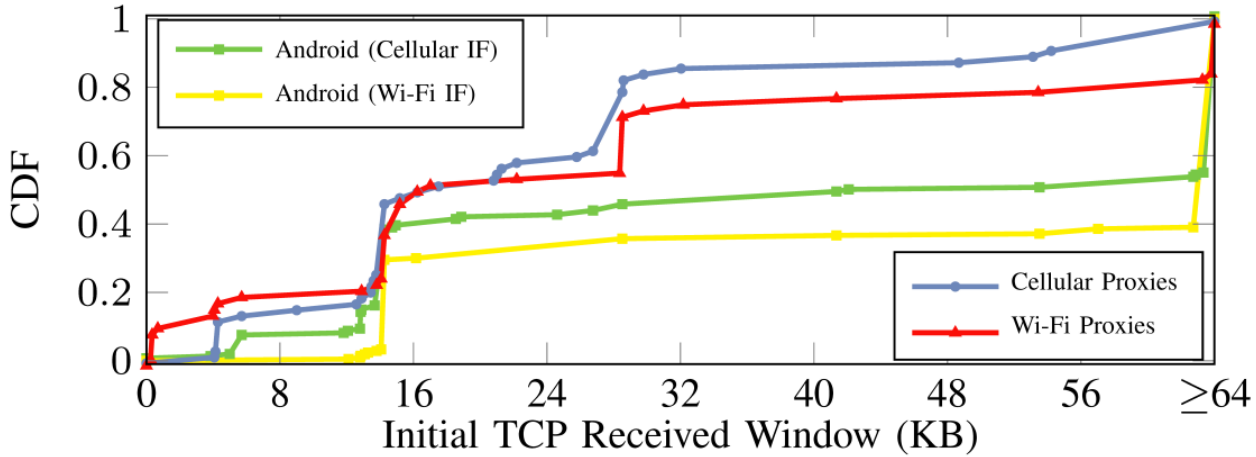
- **Early Syn Ack** decouples the handshake between client and proxy from the handshake between proxy and server
  - First figure suggests that proxy concomitantly responds with a Syn Ack to the client and forwards its Syn
  - Previously observed HTTP proxies deferring Syn till reception of actual HTTP request
  - Other than these we observed a **third scenario** in which proxy's Syn is deferred till completion of 3wHS with client



Syn not deferred | Syn deferred till Ack | Syn deferred till Payload

Initial TCP Receive Window

Initial TCP Window as set by Android devices (using cellular and Wi-Fi interfaces) and by proxies



- Most recurring values for proxies are in range between 14K and 30K
- Most recurring values for devices are around 64K and above
- In 55% of the probes initial Window Size advertised by the proxy is lower than the one advertised by the device

Initial TCP Receive Window

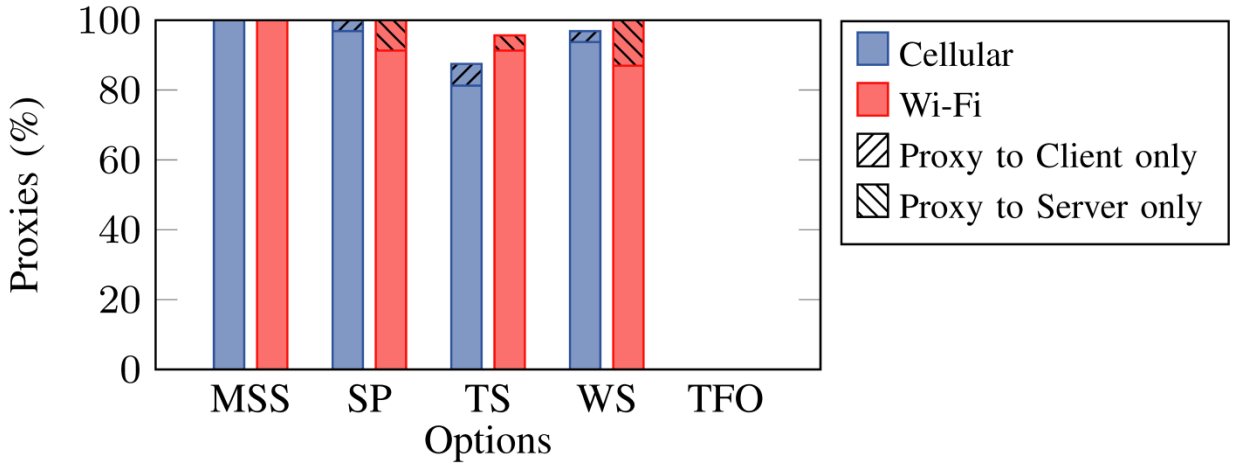
- RFC6928 (2013) recommended to increase the initial Congestion Window (IW) to 10 segments
- In 2017 85% of HTTP servers and 80% of TLS servers in Alexa Top 1m already supported this recommendation
- To benefit from IW10 client has to advertise an initial Receive Window of at least 10 segments

Window (# of segments)	Cellular		Wi-Fi	
	Device	Proxy	Device	Proxy
<10	7%	15%	0%	20%
=10	30%	31%	29%	28%
>10	63%	54%	71%	62%

- The percentage of proxies that cannot benefit from servers supporting IW10 is higher than the percentage of devices

TCP Options

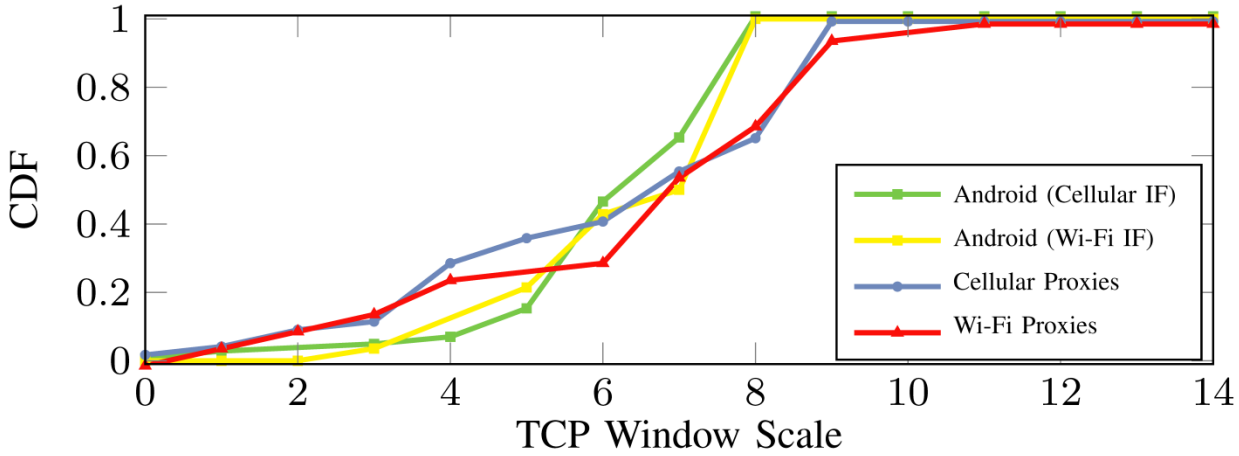
TCP Options supported by cellular and Wi-Fi proxies



- Some cellular proxies support certain Options only on client-to-proxy connection
- Some Wi-Fi proxies support certain Options only on proxy-to-server connection
- TFO probes and rooted probes with MPTCP, Unassigned Option kind or without any Option suggest that the set of Options on proxy's Syn is fixed and do not depend on client's Syn
- Other extensions: **TCP ECN**

Window Scale Factor

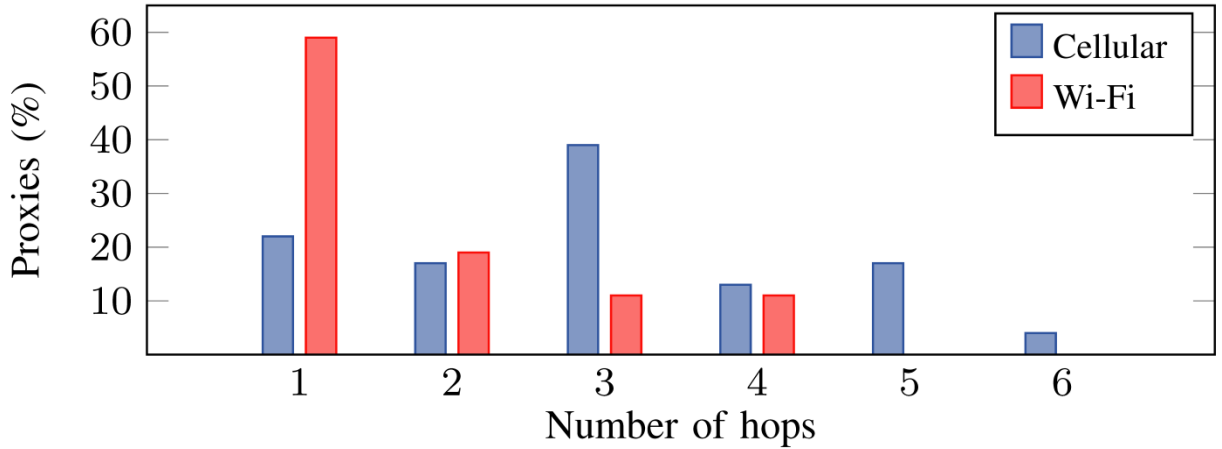
TCP Window Scale as set by Android devices (using cellular and Wi-Fi interfaces) and by proxies



- Distributions are very close
- Only in 40% of the recorded probes the proxy sets a WS value lower than the original
- 22% of proxies advertised different WS factors on the connection to the client and to the server

Proxy location

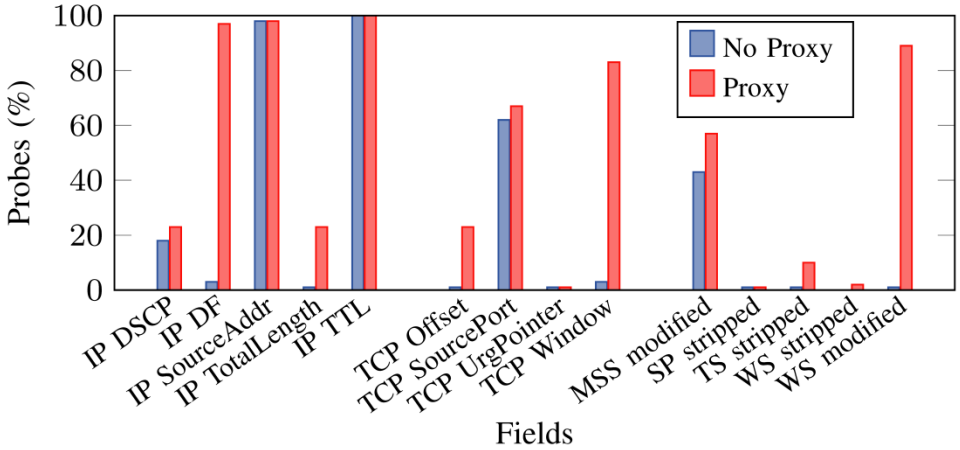
Proxy location in mobile networks



- Only (pseudo)traceroute probes preceded by a server-based probe that has highlighted the presence of a proxy altering Syn

Fingerprinting

Fingerprinting proxies by packet modifications: comparison of packet modifications detected on paths with a proxy and without



- Modifications on **IP Don't Fragment**, **TCP Window**, **Window Scale**, and to a lesser extent on TCP Offset (and consequently on IP Total Length) are almost exclusively ascribable to proxies
- Fingerprinting helps to discriminate the case of proxies not altering Syn TTL
  Does the proxy forward the original Syn or forge its own Syn keeping the original TTL?

Measurement caveats

A few caveats emerged from our study

1. Detection methodologies based on a non-responding test and more generally on a Syn only test are not capable to detect all TCP-terminating proxies, due to the presence of proxies not employing early Syn Ack and also not altering Syn
2. Detection methodologies based on fingerprinting packets received on a specific port (e.g. HTTP) and on a non-standard port do not succeed in networks where all TCP traffic is routed through a proxy
3. A measured path length up to 6 hops does not rule out the presence of a proxy
4. Testing if a certain feature (e.g. TCP Options, TCP ECN) is supported by a proxy on client-to-proxy or proxy-to-server connection does not imply that is supported or not on the other connection


Conclusions and Future Work

- Measurement tool to detect and characterize proxies in mobile networks without root privileges
- TCP-terminating and packet-rewriting proxies
- Proxy detection through TTL rewriting
- Proxy prevalence and scope
- Proxy characterization
  - Transport layer behavior
  - Impact on the Receive Window
  - Impact on extensions
  - Location
  - Fingerprinting
- Measurement caveats

- New probing scheme to test further aspects of proxies
- Continue our crowdsourced study to understand the trend of proxies in mobile Internet


Thank you

Mobile Tracebox

https://play.google.com/store/apps/details?id=be.ac.ulg.mobiletracebox

Questions, comments, etc

Raffaele Zullo

<r.zullo@studenti.unina.it>

<raffaele@erg.abdn.ac.uk>