Informing Protocol Design Through Crowdsourcing: the Case of Pervasive Encryption

Anna Maria Mandalari
amandala@it.uc3m.es

Marcelo Bagnulo
marcelo@it.uc3m.es

Andra Lutu
andra@simula.no

Universidad
Carlos III de Madrid

Internet Innovation

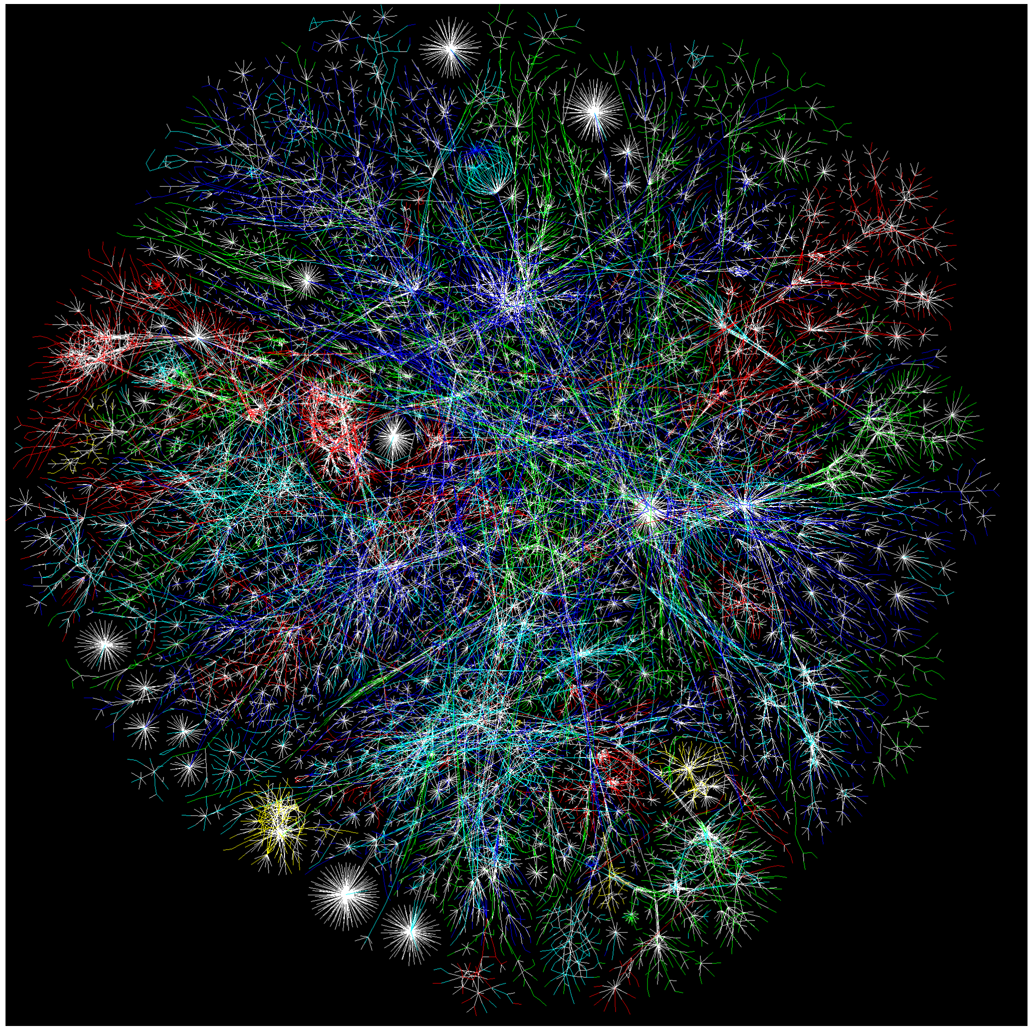
The Internet has successfully enabled multiple waves of innovation:

amandala@varpa.it.uc3m.es

- ➢ Mobility
- ➢ Heterogeneity of devices
- ➢ Video Communication
- ➢ VoIP
- ➢ …..

Internet Innovation

The Internet changes dramatically in terms of number and types of its nodes and running applications



Is the Internet Ossified?

Today,many aspects appear to be **"set in stone"**

Criticism: Middleboxes behavior

Handley, M. (2006). Why the Internet only just works. BT Technology Journal, 24(3), 119-129.

Middleboxes compatibility

Middleboxes functionalities:

- Enhancing application performance (e.g., traffic accelerators, caches, proxies);
- Traffic shaping (e.g., load balancers);
- Optimizing the usage of IPv4 address space (e.g., NATs);
- Security (e.g., firewalls).

Major criticism: they might filter traffic that does not conform to expected behaviors.

Is the Internet Ossified?

Several of the protocols standardized by IETF over the last few years face deployment challenges blamed on interference by middleboxes.

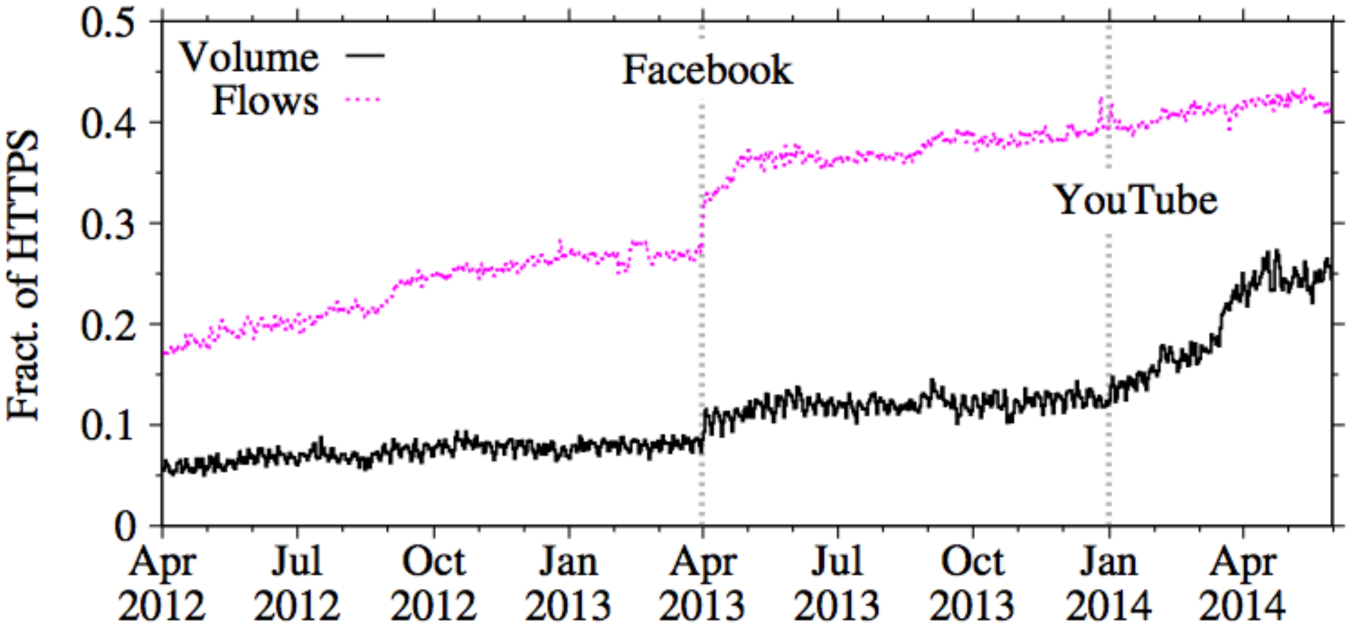
Is the Internet Ossified?

How will Internet react to a new protocol?

Understand the interaction of the new solutions with the middleboxes active along the path.

The case of pervasive encryption

Many popular applications (e.g., web, Youtube video streaming) have migrated from HTTP to the HTTPS protocol



Challenge: Provide encryption by default for all Internet communications

Naylor, David, et al. "The Cost of the S in HTTPS." Proceedings of the 10th ACM International on Conference on emerging Networking Experiments and Technologies. ACM, 2014.

The case of pervasive encryption

Understand the feasibility of pervasive encryption in the Internet.

Understand the interaction of middleboxes with the TLS across the different TCP ports that currently use plain text protocols.

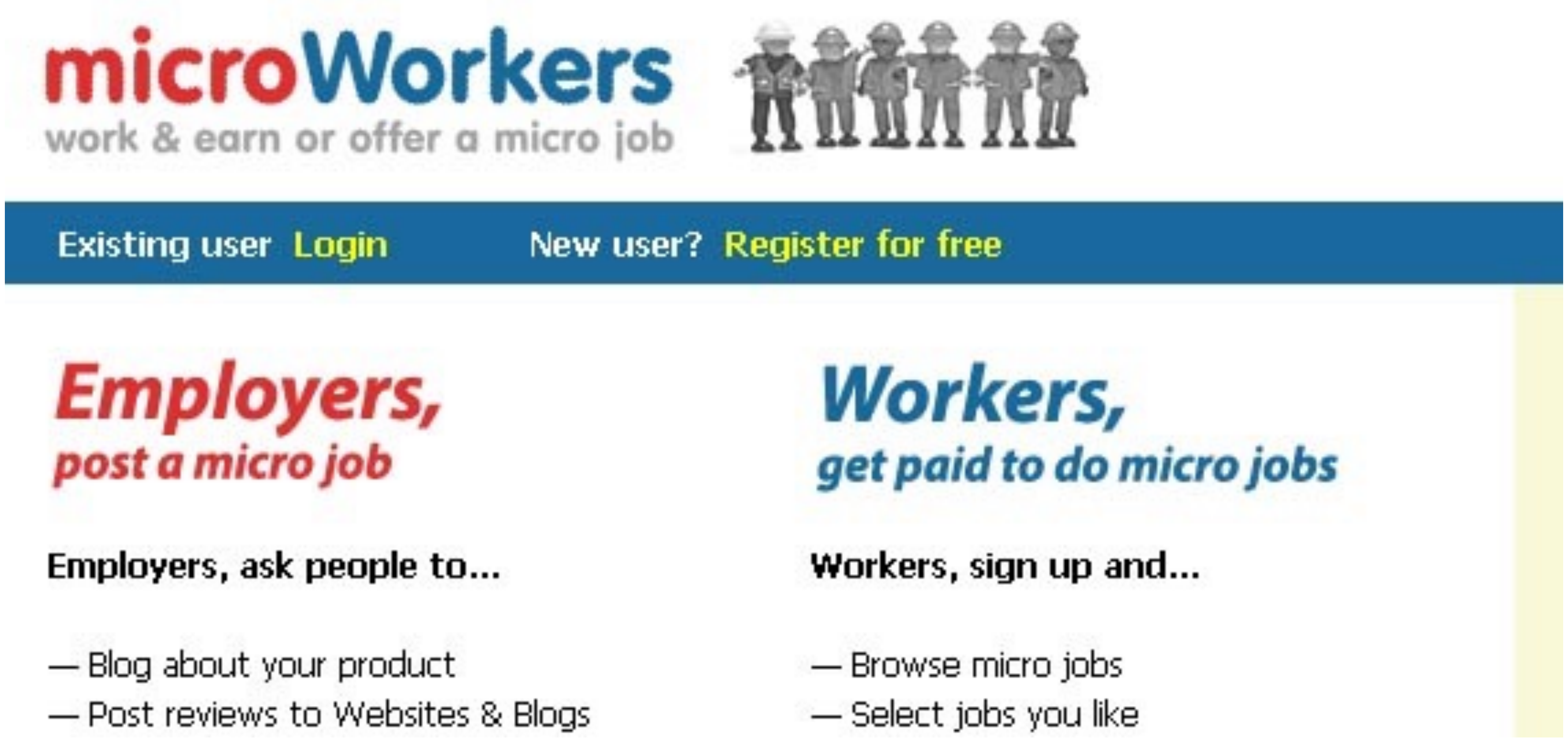
How to measure a thousand end-users?

- Be Google (or any other large Internet players)

or

- Get your code to run on a thousand users' machines through another delivery channel

Crowdsourcing platform



microWorkers

work & earn or offer a micro job

Existing user Login

New user? Register for free

Employers,

post a micro job

Employers, ask people to...

— Blog about your product

— Post reviews to Websites & Blogs

Workers,

get paid to do micro jobs

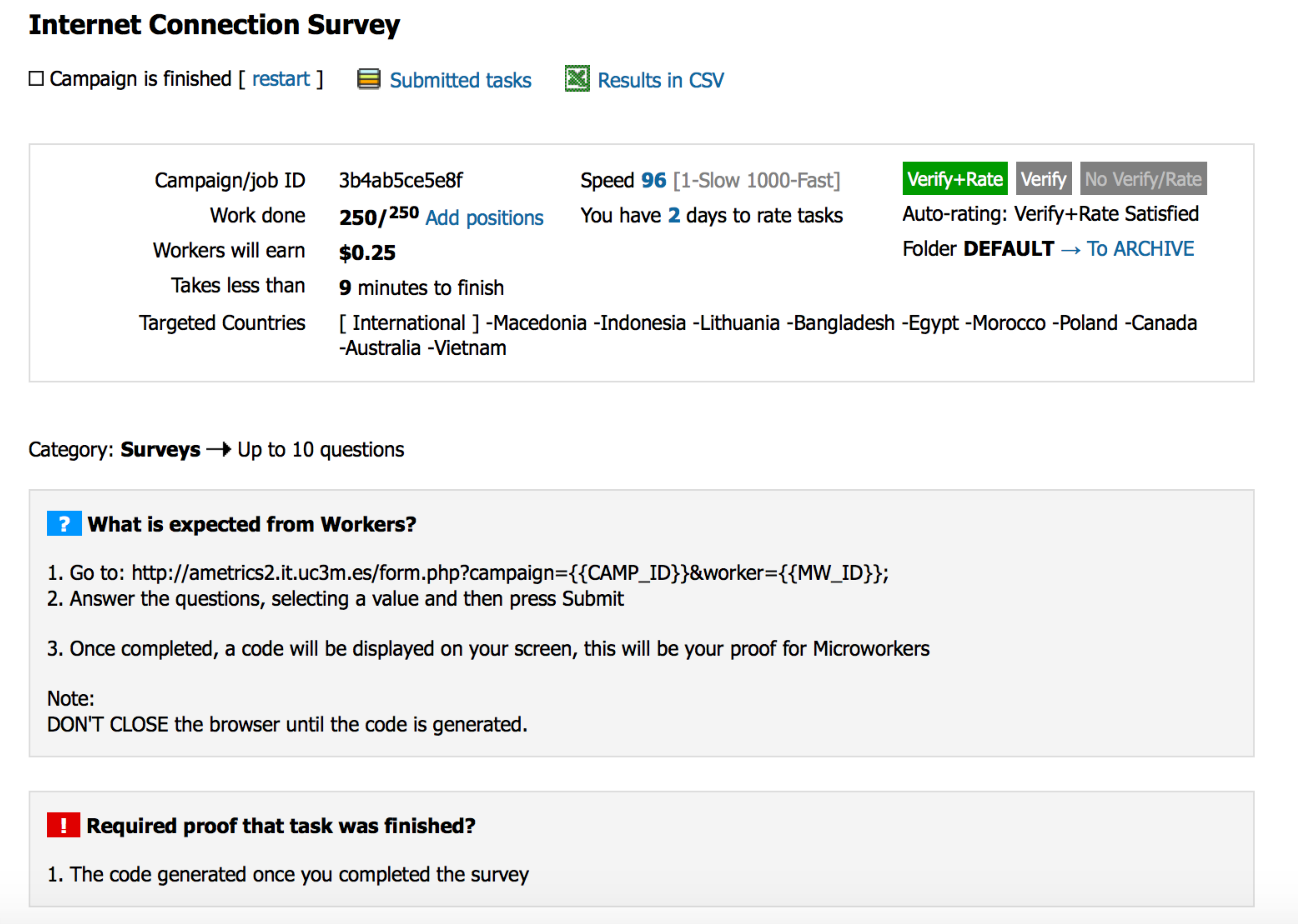
Workers, sign up and...

— Browse micro jobs

— Select jobs you like

Perform large-scale Internet measurement campaigns

Crowdsourcing platform



Internet Connection Survey

□ Campaign is finished [restart]   Submitted tasks   Results in CSV

Campaign/job ID	3b4ab5ce5e8f
Work done	**250/**^{250} Add positions
Workers will earn	**$0.25**
Takes less than	**9** minutes to finish
Targeted Countries	[International] -Macedonia -Indonesia -Lithuania -Bangladesh -Egypt -Morocco -Poland -Canada -Australia -Vietnam

Speed **96** [1-Slow 1000-Fast]

You have **2** days to rate tasks

Auto-rating: Verify+Rate Satisfied

Folder **DEFAULT** → To ARCHIVE

Category: **Surveys** → Up to 10 questions

? What is expected from Workers?

1. Go to: http://ametrics2.it.uc3m.es/form.php?campaign={{CAMP_ID}}&worker={{MW_ID}};
2. Answer the questions, selecting a value and then press Submit

3. Once completed, a code will be displayed on your screen, this will be your proof for Microworkers

Note:
DON'T CLOSE the browser until the code is generated.

! Required proof that task was finished?

1. The code generated once you completed the survey

Crowdsourcing platform

Reasons to choose Microworkers:

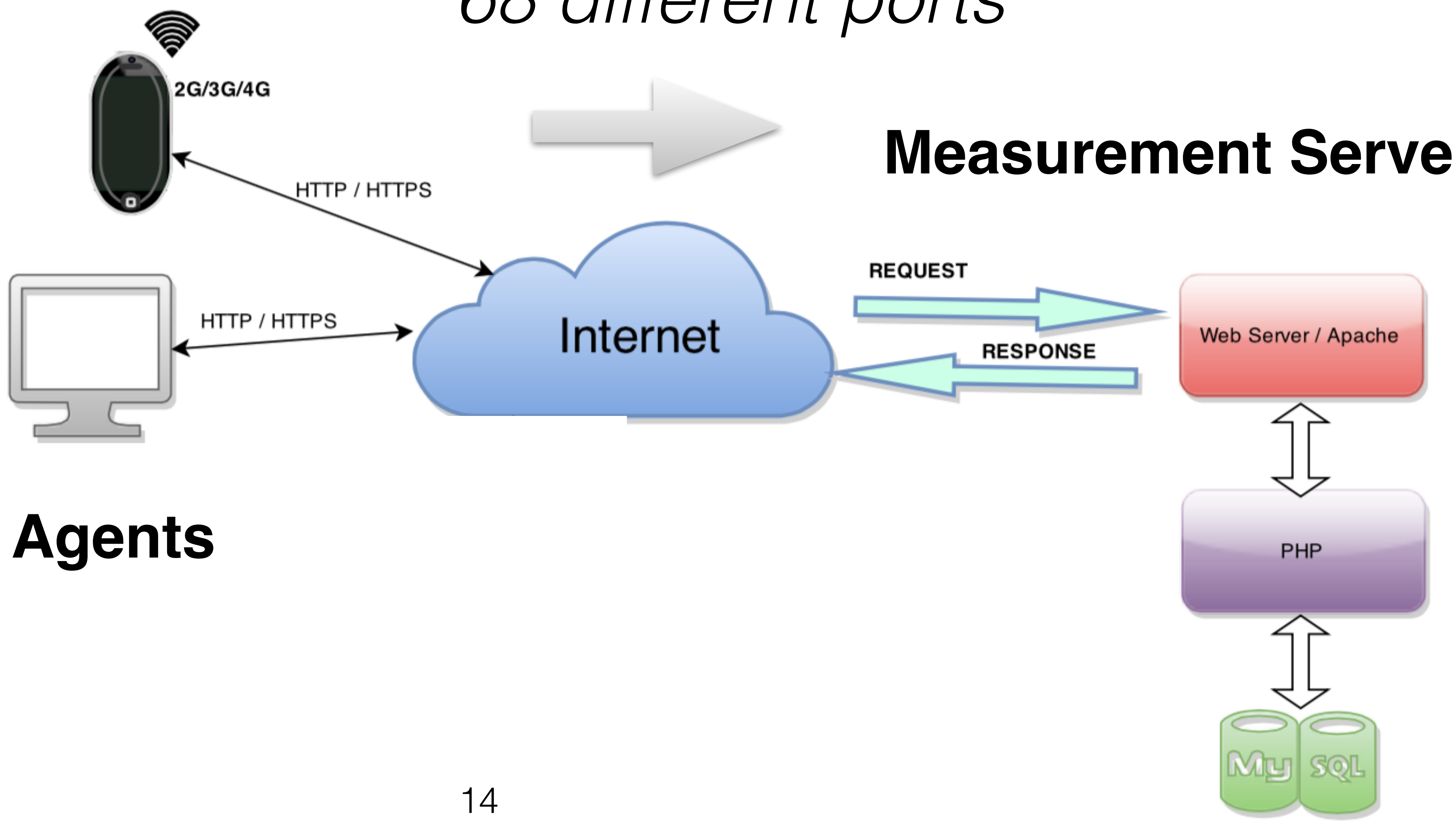
- World-wide access to employers;
- Automatic payment method based on a unique verification code;
- Possibility to select the MAs based on certain criteria, i.e. geographical location at the country level, the type of Internet access (fixed or mobile) or even the type of measurement equipment used to perform the tasks.

Experimental setup

Establish both **HTTP** and **TLS** connections to **68 different ports:**

- 10 well-known ports;
- 56 registered ports;
- 2 ephemeral ports.

TLS connections over 68 different ports

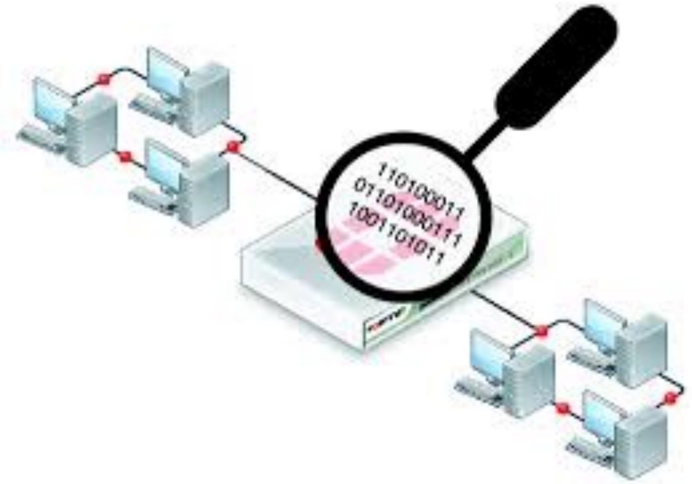


Measurement Server

Measurement Agents

Experimental setup: Measurement Server

- LAMP model (Linux, Apache Server, MySQL relational database management system, PHP);
- Packets capture.



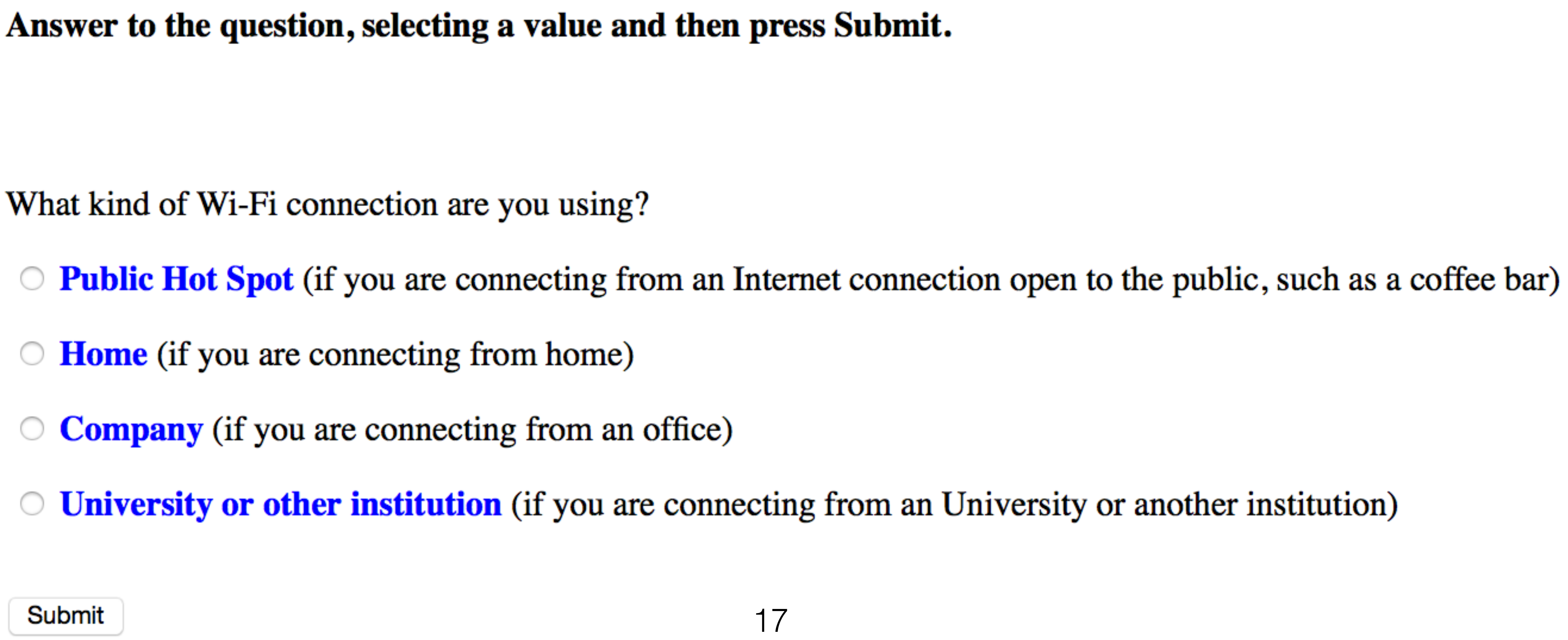
Experimental setup: Measurement Agent Common Procedure

Limit of crowdsourcing platform: some information may not be available through the platform

- Users connects using a HTTP connection in port 80 to a webpage I provide

Experimental setup: Measurement Agent Common Procedure

- Users connected from Fixed line indicate the place from where they are connecting (Home, Hot Spot, University or or other institution, Company)



Answer to the question, selecting a value and then press Submit.

What kind of Wi-Fi connection are you using?

○ **Public Hot Spot** (if you are connecting from an Internet connection open to the public, such as a coffee bar)

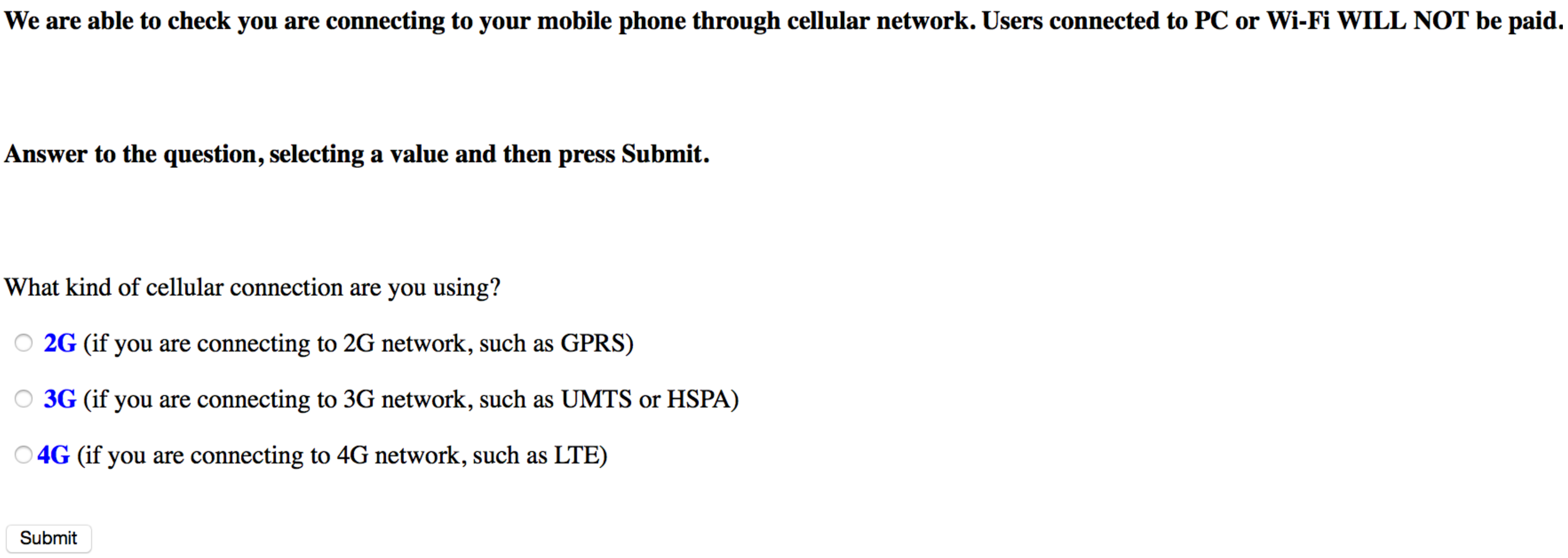
○ **Home** (if you are connecting from home)

○ **Company** (if you are connecting from an office)

○ **University or other institution** (if you are connecting from an University or another institution)

Experimental setup: Measurement Agent Common Procedure

- Users connected from Mobile line indicate the technology they are using (2G, 3G, 4G)



We are able to check you are connecting to your mobile phone through cellular network. Users connected to PC or Wi-Fi WILL NOT be paid.

Answer to the question, selecting a value and then press Submit.

What kind of cellular connection are you using?

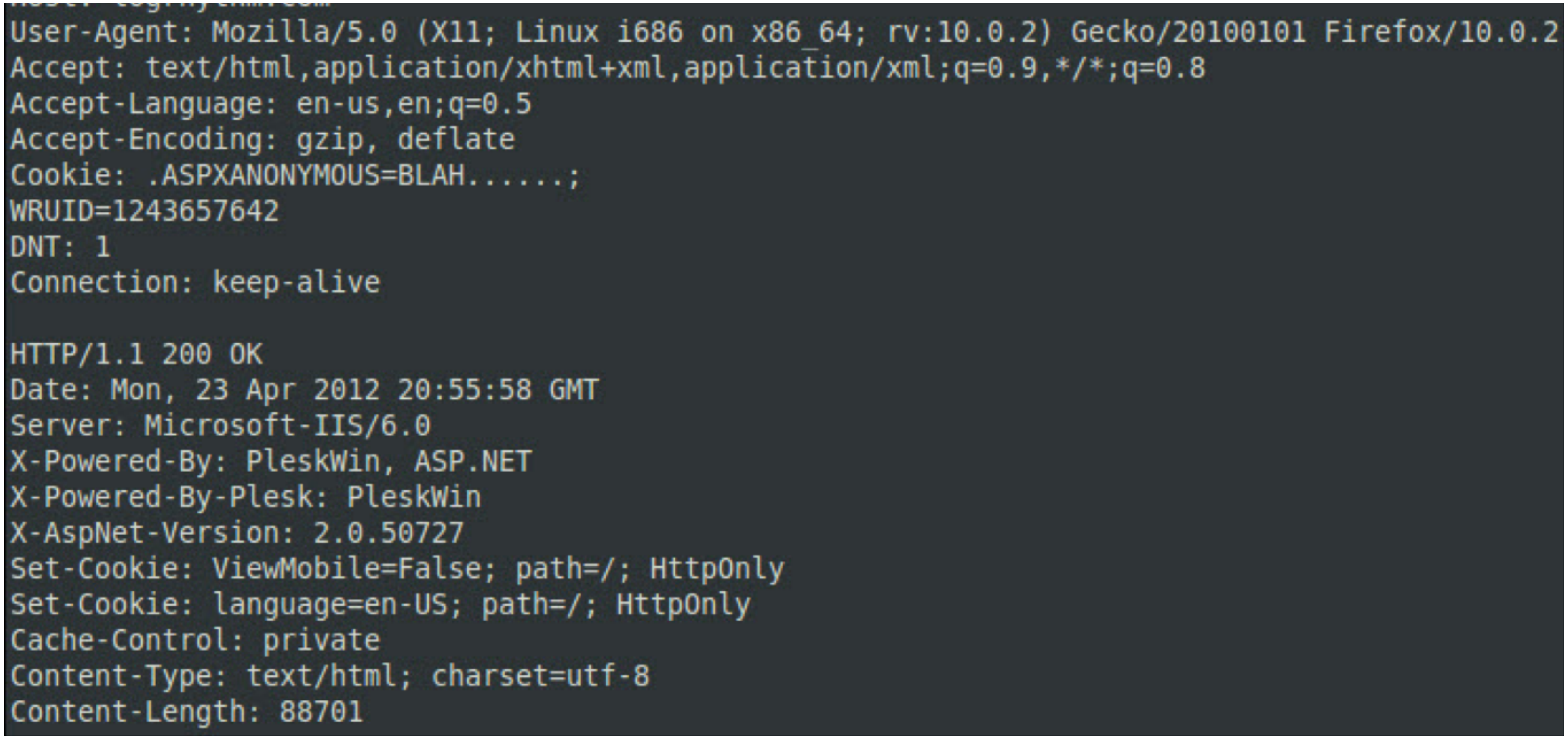
○ **2G** (if you are connecting to 2G network, such as GPRS)

○ **3G** (if you are connecting to 3G network, such as UMTS or HSPA)

○ **4G** (if you are connecting to 4G network, such as LTE)

Experimental setup: Measurement Agent Common Procedure

- I collect and store metadata on each of the MAs that connect to our servers, such as the IP address, the user agent type, the language, and any other information included in the HTTP header

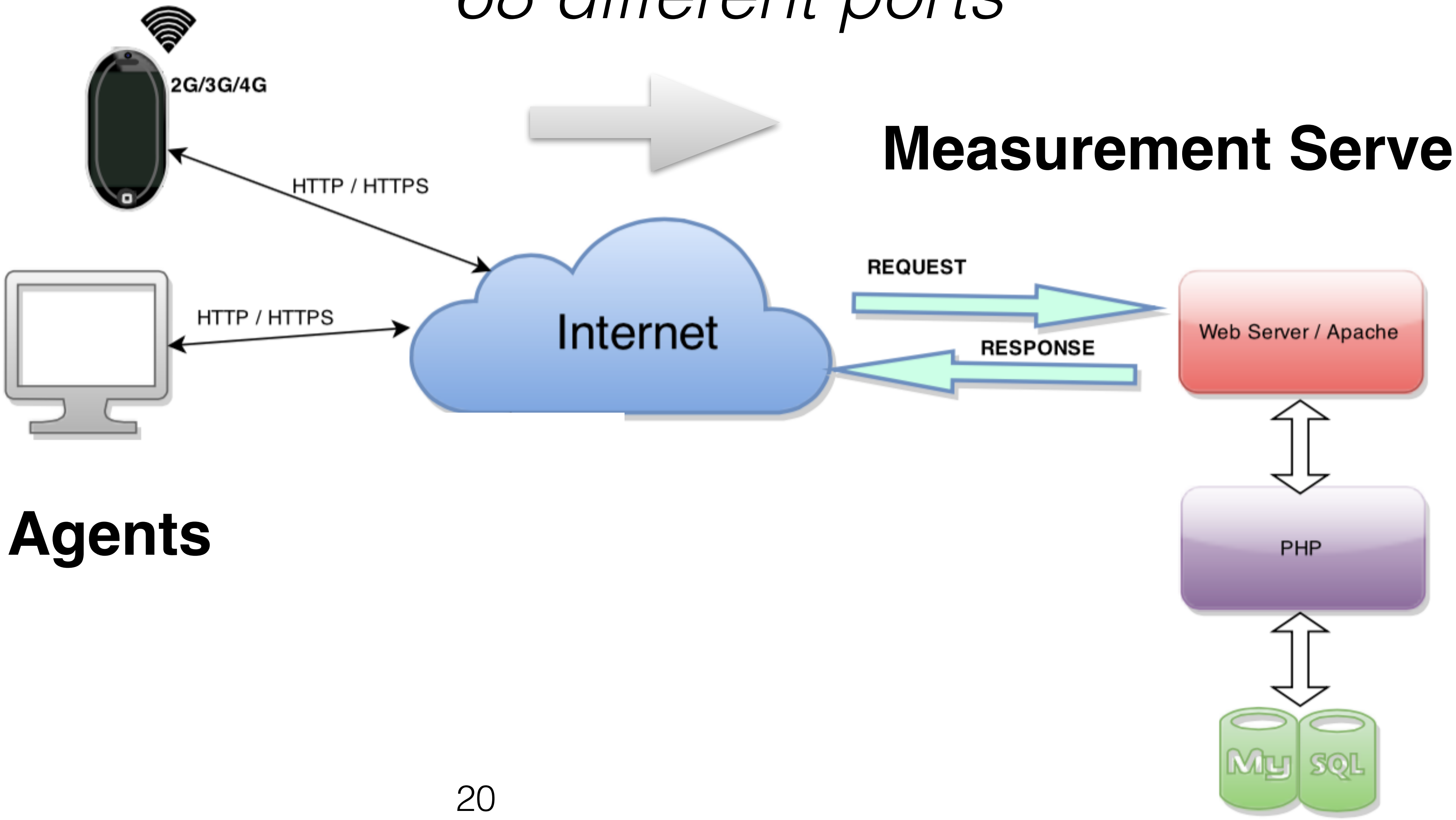


```
User-Agent: Mozilla/5.0 (X11; Linux i686 on x86_64; rv:10.0.2) Gecko/20100101 Firefox/10.0.2
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8
Accept-Language: en-us,en;q=0.5
Accept-Encoding: gzip, deflate
Cookie: .ASPXANONYMOUS=BLAH......;
WRUID=1243657642
DNT: 1
Connection: keep-alive

HTTP/1.1 200 OK
Date: Mon, 23 Apr 2012 20:55:58 GMT
Server: Microsoft-IIS/6.0
X-Powered-By: PleskWin, ASP.NET
X-Powered-By-Plesk: PleskWin
X-AspNet-Version: 2.0.50727
Set-Cookie: ViewMobile=False; path=/; HttpOnly
Set-Cookie: language=en-US; path=/; HttpOnly
Cache-Control: private
Content-Type: text/html; charset=utf-8
Content-Length: 88701
```

Experimental setup: Measurement Agent Common Procedure

- In the background, HTTP and HTTPS connections are performed from the measurement devices to our servers in all the 68 ports



TLS connections over 68 different ports

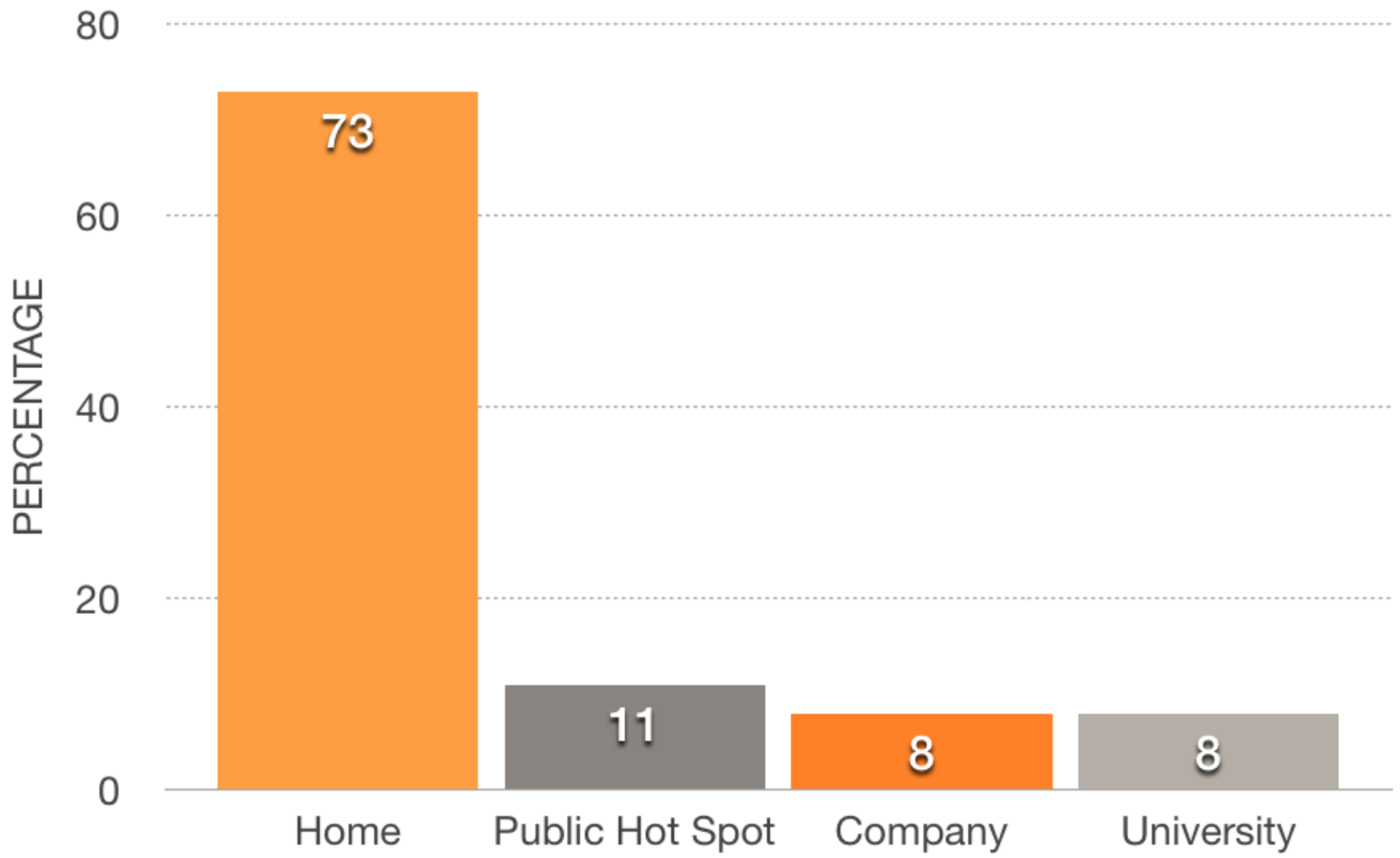
Measurement Server

Measurement Agents

Data Set

FIXED LINE:

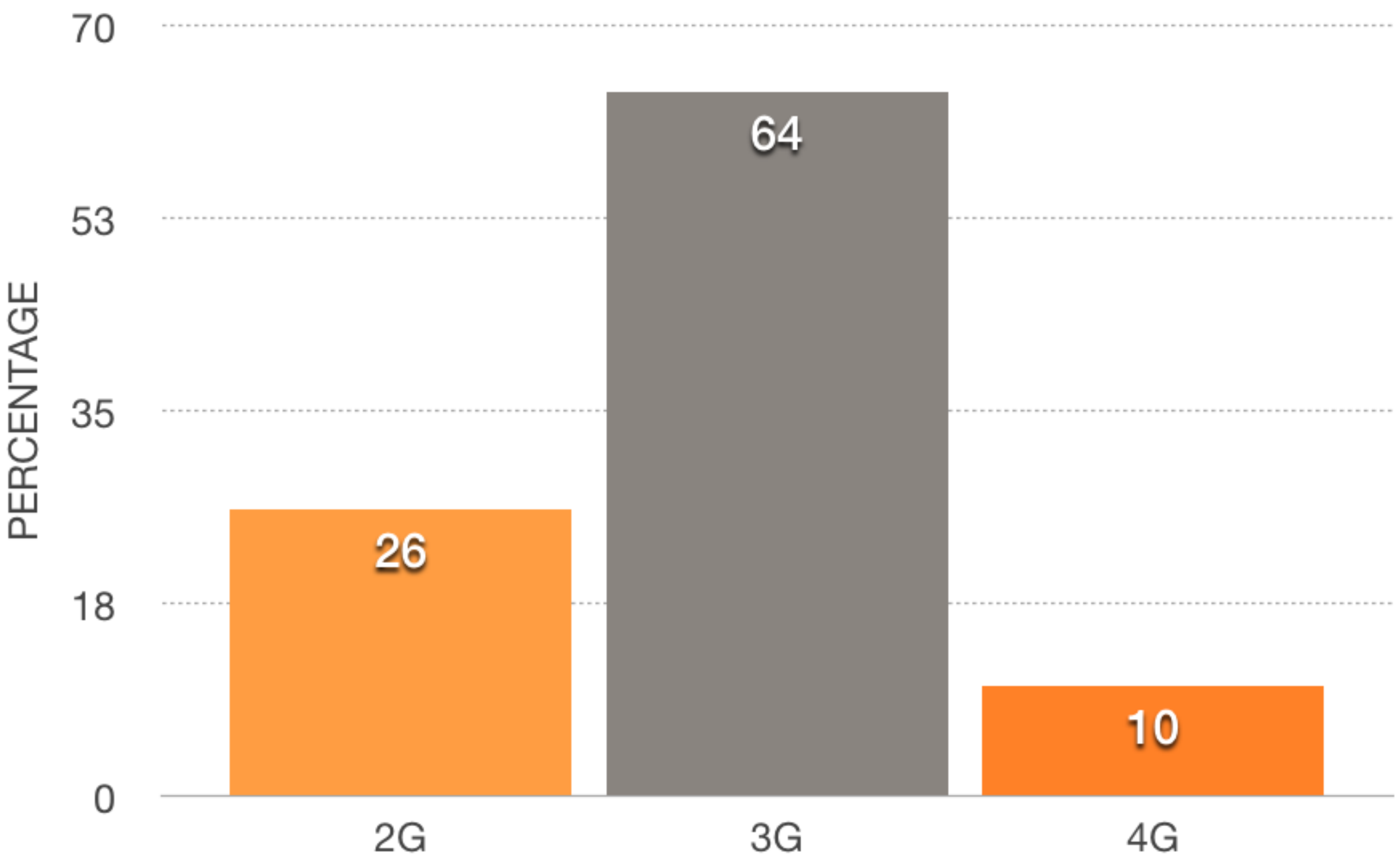
- 1,165 workers;
- 53 different countries;
- 286 ASes.



Data Set

MOBILE:

- 956 workers;
- 45 different countries;
- 183 ASes.

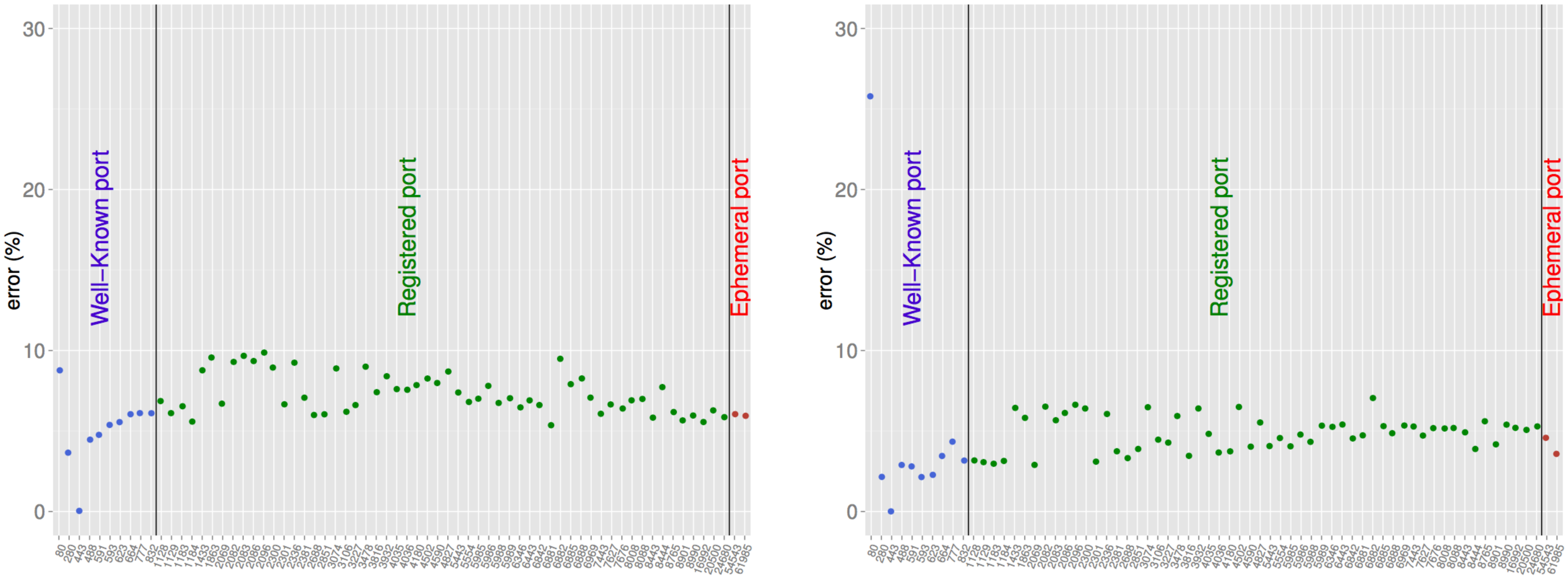


Total of 114,228 connections

The data set is freely available on http://it.uc3m.es/amandala/dataset.php

Aggregated results

ERROR = (success [HTTP] − success [TLS]) / success [HTTP]



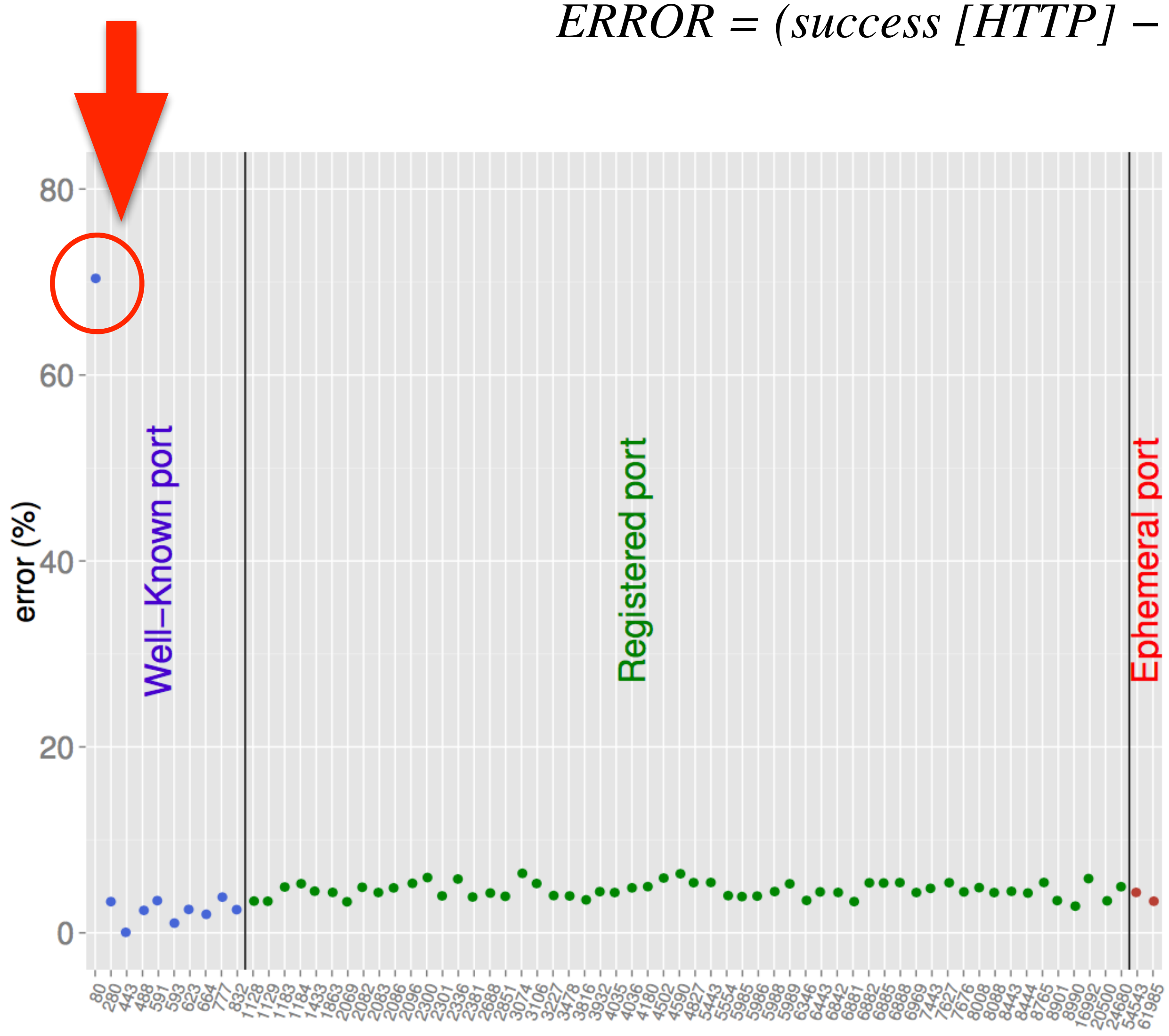
a) Fixed line

b) Mobile network

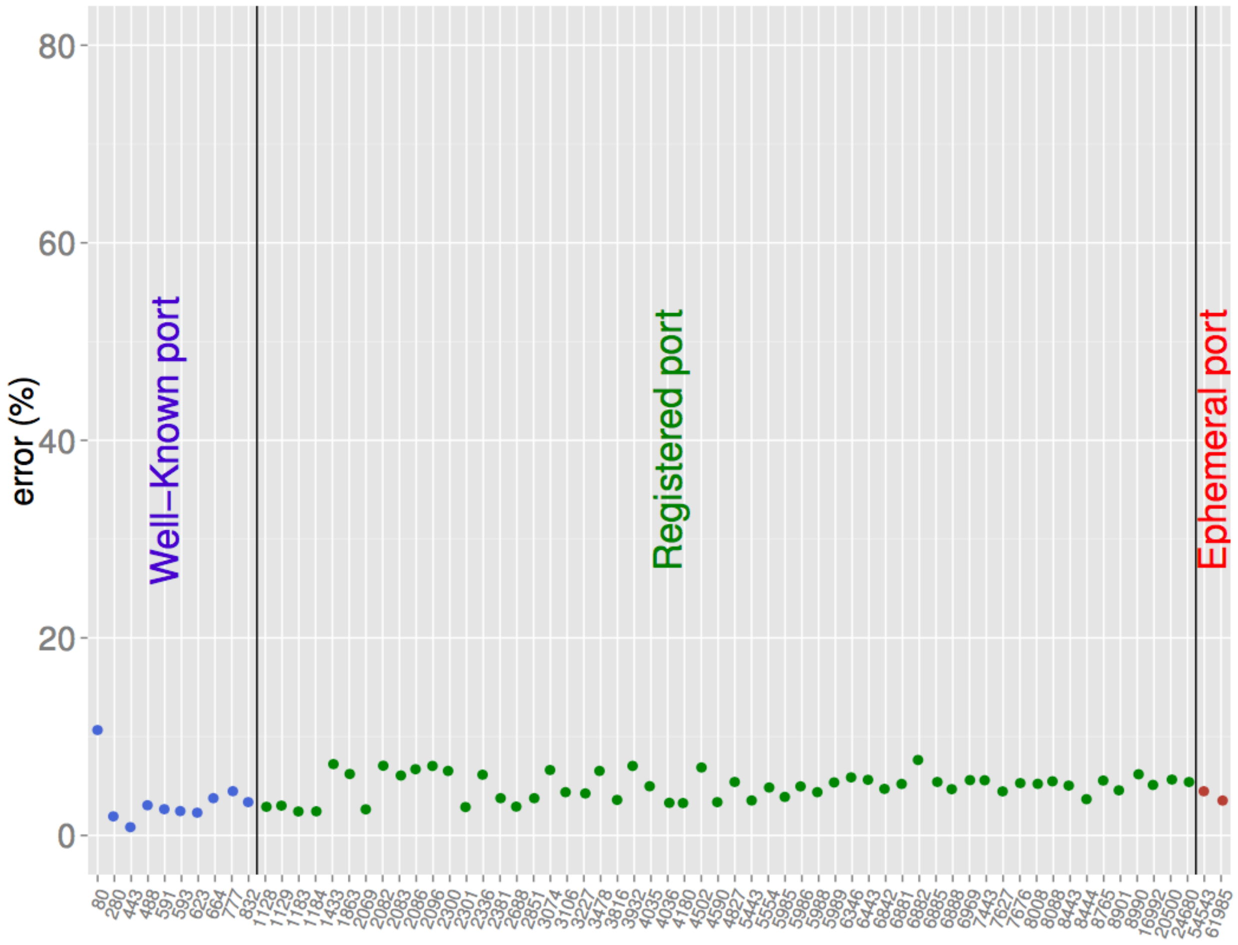
25% of the users are not able to perform a TLS connection over port 80 in mobile network.

Proxies

ERROR = (success [HTTP] – success [TLS]) / success [HTTP]



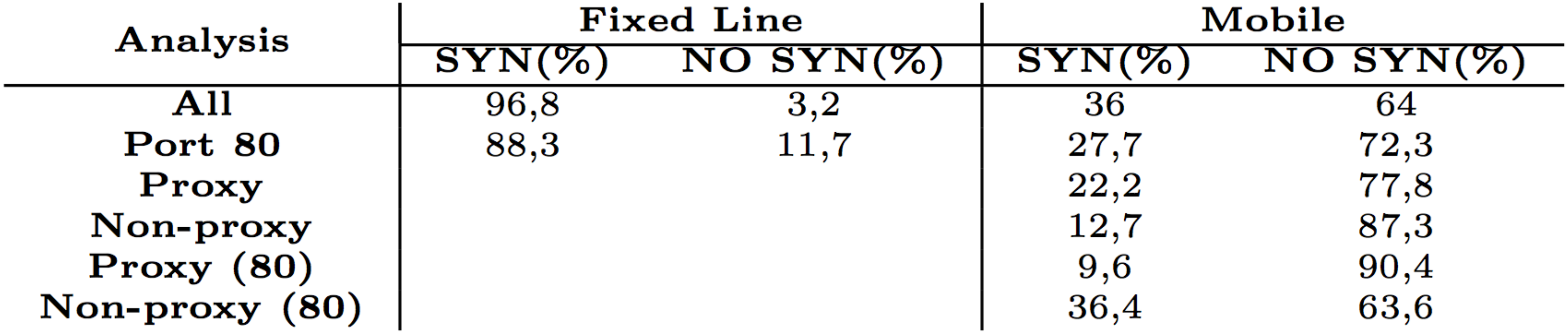
a) Mobile proxy



b) Mobile non-proxy

70% of the users that use a proxy are not able to perform a TLS connection over port 80 in mobile network.

Packets analysis

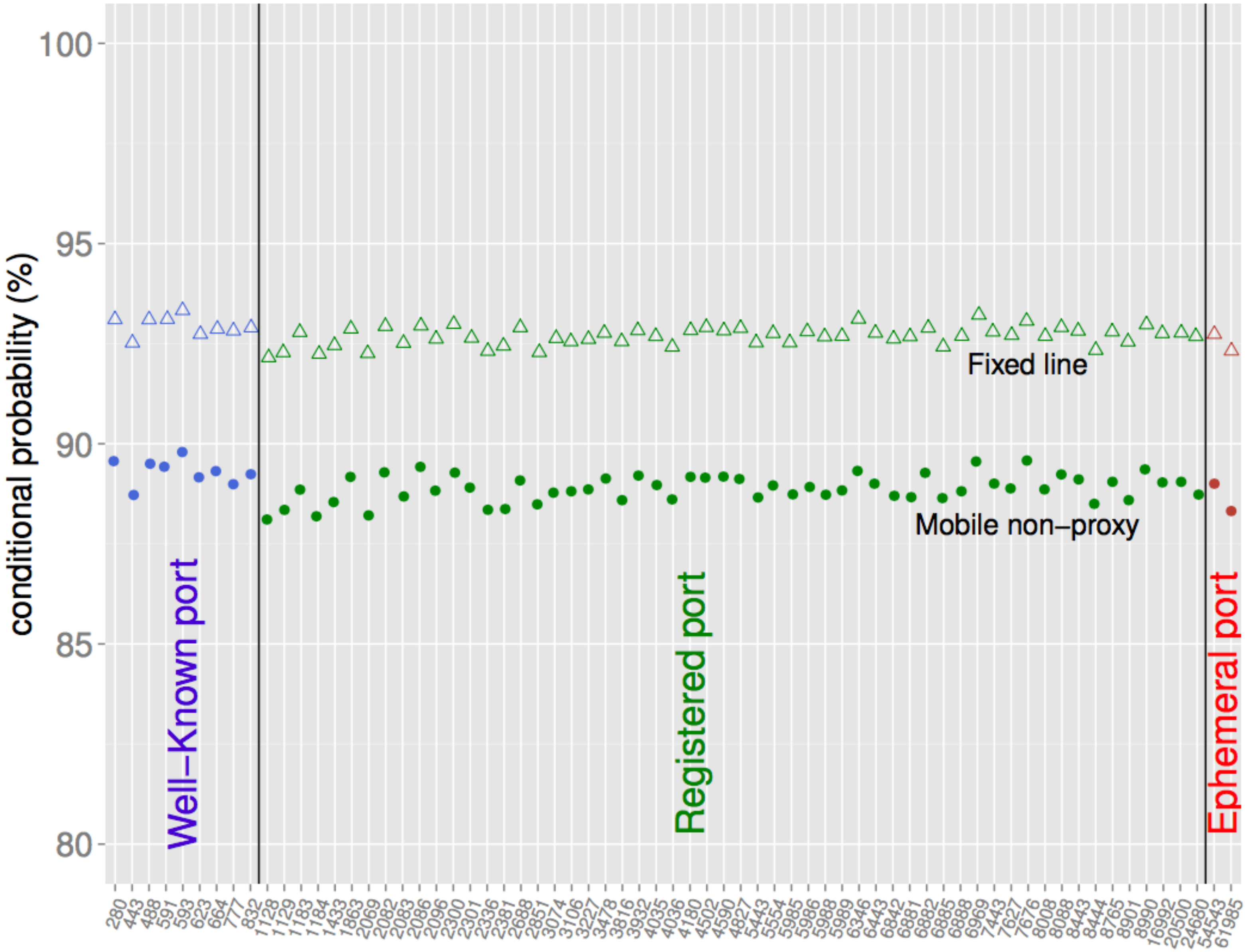


Analysis	Fixed Line		Mobile	
	SYN(%)	NO SYN(%)	SYN(%)	NO SYN(%)
All	96,8	3,2	36	64
Port 80	88,3	11,7	27,7	72,3
Proxy			22,2	77,8
Non-proxy			12,7	87,3
Proxy (80)			9,6	90,4
Non-proxy (80)			36,4	63,6

When users use a proxy, 90% of the SYN packets are missing

Consistent filtering

The percentage of errors in other ports, when an error occurs in port 80.



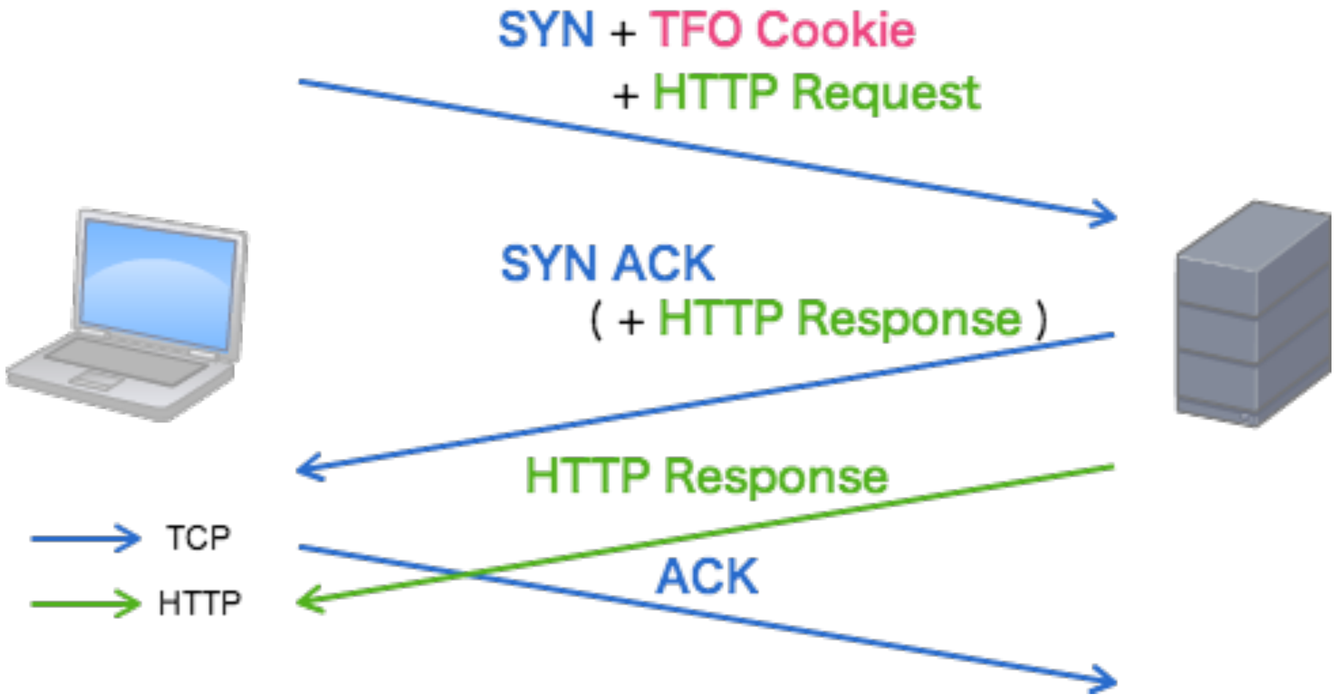
The estimated conditional probability is around 90% for both fixed line and mobile network

Conclusion

- Overcome several of the limitations of the crowdsourcing platforms
- It is probably feasible to roll out TLS protection for most ports except for port 80, assuming a low failure rate (6%)
- Our results can serve as a lower bound for the failure rate for using protocols other than expected in different ports

Work in progress

- TCP Fast Open



- HTTP/2

ANY QUESTIONS?

FAQ

www.it.uc3m.es/amandala/faq.pdf

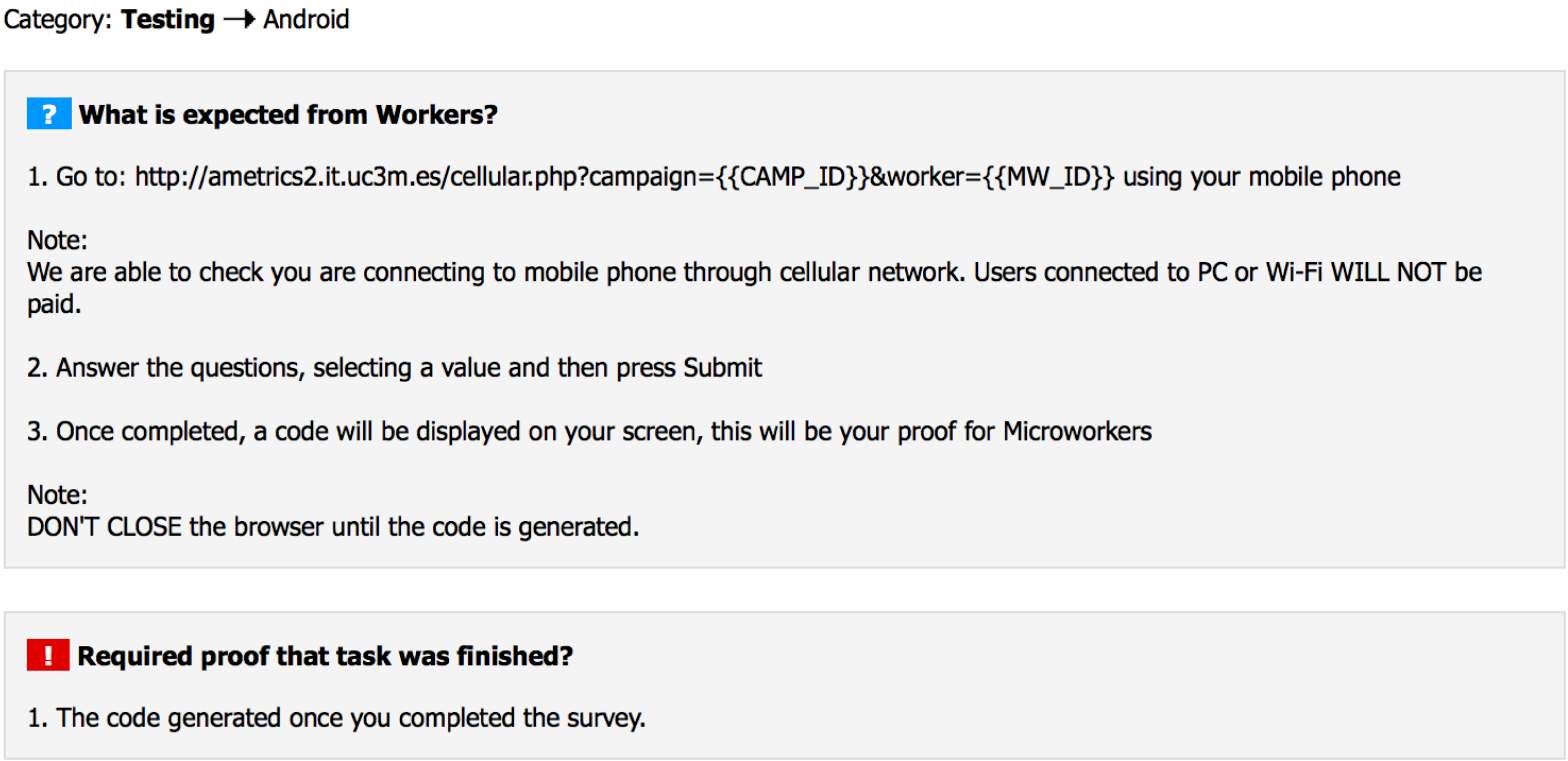
FAQ

1. How reliable are the answers you get from your online survey?
2. How do you perform the HTTPS connections?
3. How do you detect the presence of proxies?
4. What kind of devices do you select?
5. How do you select the countries?
6. Given that the 80% are at home, does that affect the results?
7. Does using IP address as an identifier for the measurements node affect the results?
8. Can you provide a list of the port numbers?

FAQ

9. What about other types of middleboxes?
10. Have you tried to find some correlation between the results and the ASes or the Countries from which the users are connected?
11. How much did the campaign cost?
12. Why don’t you use PlanetLab or other “free” platforms?
13. Is it possible to apply this method to other measurements?
14. Why do you consider the case of pervasive encryption?
15. Have you tried to randomize the order of the port numbers
16. What are the main reasons proxies block a TLS connection?

1. How reliable are the answers you get from your online survey?



Category: **Testing** → Android

? **What is expected from Workers?**

1. Go to: http://ametrics2.it.uc3m.es/cellular.php?campaign={{CAMP_ID}}&worker={{MW_ID}} using your mobile phone

Note:
We are able to check you are connecting to mobile phone through cellular network. Users connected to PC or Wi-Fi WILL NOT be paid.

2. Answer the questions, selecting a value and then press Submit

3. Once completed, a code will be displayed on your screen, this will be your proof for Microworkers

Note:
DON'T CLOSE the browser until the code is generated.

! **Required proof that task was finished?**

1. The code generated once you completed the survey.

2. How do you perform the HTTPS connections?

```
<iframe width="1" scrolling="no" height="1" frameborder="0" src="https://ametrics.it.uc3m.es:80/index.php" seamless="seamless"></iframe>
<iframe width="1" scrolling="no" height="1" frameborder="0" src="https://ametrics.it.uc3m.es/index.php" seamless="seamless"></iframe>
<iframe width="1" scrolling="no" height="1" frameborder="0" src="https://ametrics.it.uc3m.es:280/index.php" seamless="seamless"></iframe>
<iframe width="1" scrolling="no" height="1" frameborder="0" src="https://ametrics.it.uc3m.es:488/index.php" seamless="seamless"></iframe>
<iframe width="1" scrolling="no" height="1" frameborder="0" src="https://ametrics.it.uc3m.es:591/index.php" seamless="seamless"></iframe>
<iframe width="1" scrolling="no" height="1" frameborder="0" src="https://ametrics.it.uc3m.es:593/index.php" seamless="seamless"></iframe>
```

3. How do you detect the presence of proxies?

I observe two kinds of proxies in today's Internet: transparent and non-transparent.

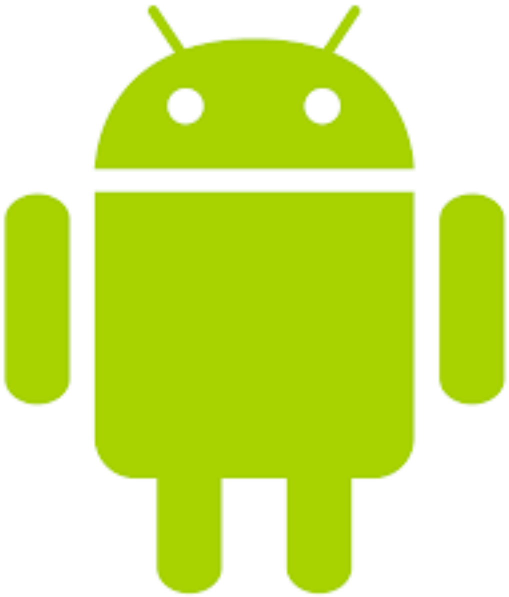
A proxy can insert into the HTTP header some standardized fields through which I are able to detect that the request has been forwarded by a proxy.

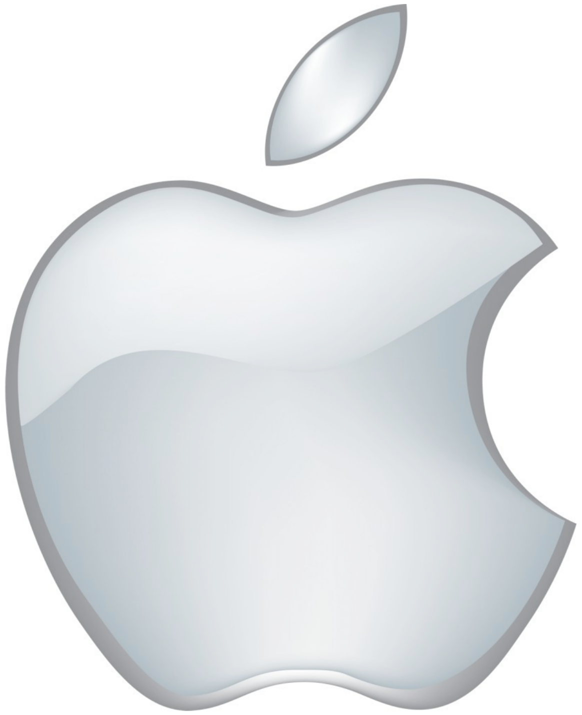
Some HTTP Header field:

- CLIENT_IP
- FORWARDED
- FORWARDED-FOR
- FORWARDED-FOR-IP
- PROXY-CONNECTION
- VIA
- X-FORWARDED
- X-FORWARDED-FOR

4. What kind of devices do you select?

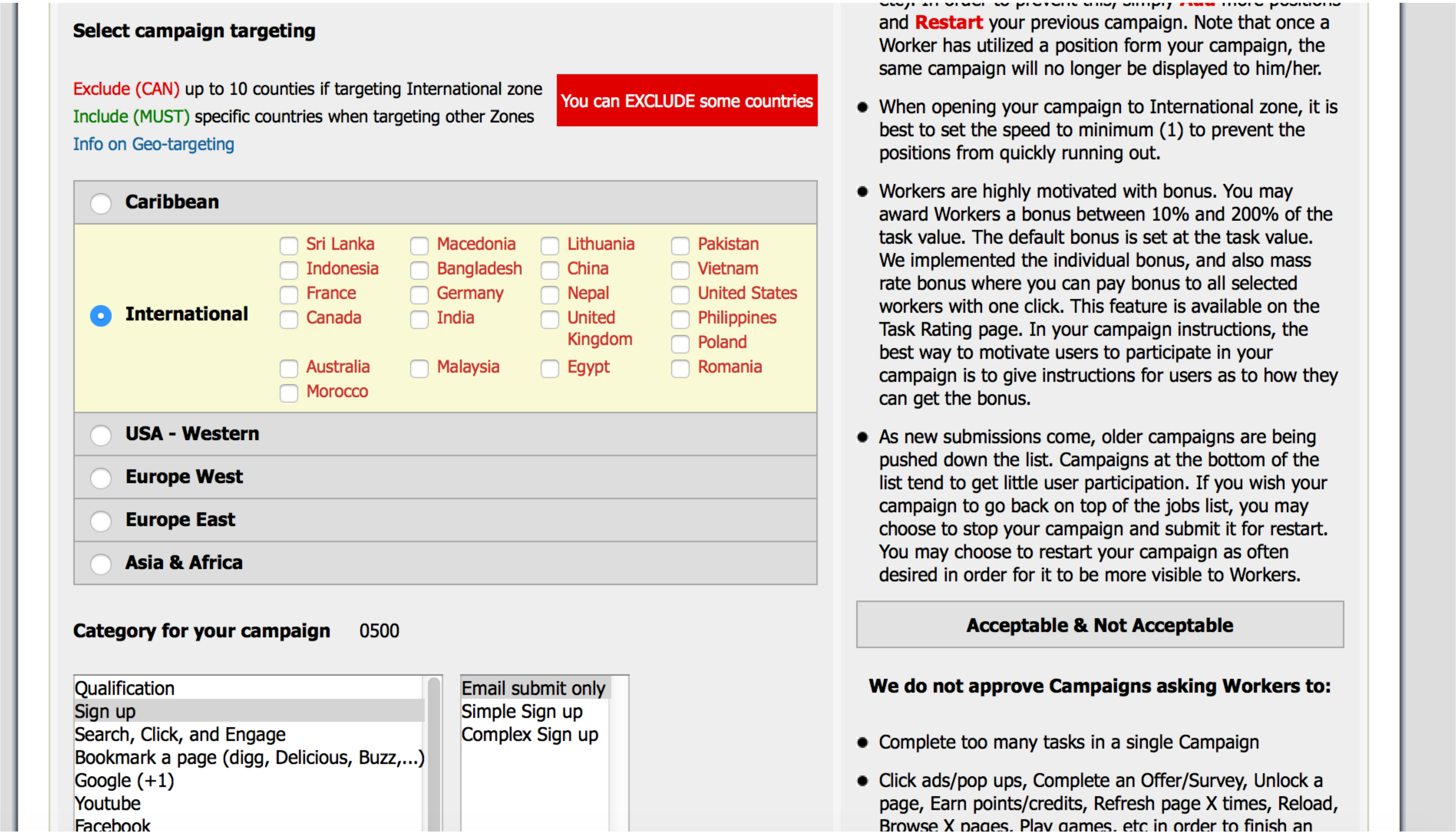
Mobile devices:





Fixed line:
All devices that have a browser

5. How do you select the countries?



Select campaign targeting

Exclude (CAN) up to 10 counties if targeting International zone
Include (MUST) specific countries when targeting other Zones
Info on Geo-targeting

You can EXCLUDE some countries

○ **Caribbean**

● **International**

□ Sri Lanka
□ Indonesia
□ France
□ Canada
□ Australia
□ Morocco
□ Macedonia
□ Bangladesh
□ Germany
□ India
□ Malaysia
□ Lithuania
□ China
□ Nepal
□ United Kingdom
□ Egypt
□ Pakistan
□ Vietnam
□ United States
□ Philippines
□ Poland
□ Romania

○ **USA - Western**

○ **Europe West**

○ **Europe East**

○ **Asia & Africa**

Category for your campaign 0500

Qualification
Sign up
Search, Click, and Engage
Bookmark a page (digg, Delicious, Buzz,...)
Google (+1)
Youtube
Facebook

Email submit only
Simple Sign up
Complex Sign up

and **Restart** your previous campaign. Note that once a Worker has utilized a position form your campaign, the same campaign will no longer be displayed to him/her.

- When opening your campaign to International zone, it is best to set the speed to minimum (1) to prevent the positions from quickly running out.
- Workers are highly motivated with bonus. You may award Workers a bonus between 10% and 200% of the task value. The default bonus is set at the task value. We implemented the individual bonus, and also mass rate bonus where you can pay bonus to all selected workers with one click. This feature is available on the Task Rating page. In your campaign instructions, the best way to motivate users to participate in your campaign is to give instructions for users as to how they can get the bonus.
- As new submissions come, older campaigns are being pushed down the list. Campaigns at the bottom of the list tend to get little user participation. If you wish your campaign to go back on top of the jobs list, you may choose to stop your campaign and submit it for restart. You may choose to restart your campaign as often desired in order for it to be more visible to Workers.

Acceptable & Not Acceptable

We do not approve Campaigns asking Workers to:

- Complete too many tasks in a single Campaign
- Click ads/pop ups, Complete an Offer/Survey, Unlock a page, Earn points/credits, Refresh page X times, Reload, Browse X pages, Play games, etc in order to finish an

6. Given that the 80% are at home, does that affect the results?

I tried to split the results considering the different scenarios I collected from the users answers.
I saw that there is not correlation between the different scenario. Also in mobile networks.

7. Does using IP address as an identifier for the measurements node affect the results?

Using IP address as an identifier for the measurement node might not work if much larger scale experiment is run or a series of tests from the same measurement node, say, every 6 hours, are run.

8. Can you provide a list of the port numbers?

80, 25, 280, 443, 488, 591, 593, 623, 664, 777, 832,
1128, 1129, 1183, 1184, 1433, 1863, 2069, 2082, 2083,
2086, 2096, 2300, 2301, 2336, 2381, 2688, 2851, 3074,
3106, 3227, 3478, 3816, 3932, 4035, 4036, 4180, 4502,
4590, 4827, 5443, 5554, 5985, 5986, 5988, 5989, 6346,
6443, 6842, 6881, 6882, 6885, 6888, 6969, 7443, 7627,
7676, 8008, 8088, 8443, 8444, 8765, 8901, 8990, 16992,
20500, 24680, 54543, 61985

9. What about other types of middleboxes?

I able only to detect proxies using this methodology, but I am expanding the method also to detect NATs or to detect other parameters. It is possible, particularly when users download for example an app we created, so that I can use Android API to get the more information about the devices or the network parameters.

10. Have you tried to find some correlation between the results and the ASes or the Countries from which the users are connected?

Yes, but we do not find any correlation.

11. How much did the campaign cost?

Fixed line: 1165 x 0,10 = 116,50 $
Mobile network: 956 x 0,30 = 286,80 $

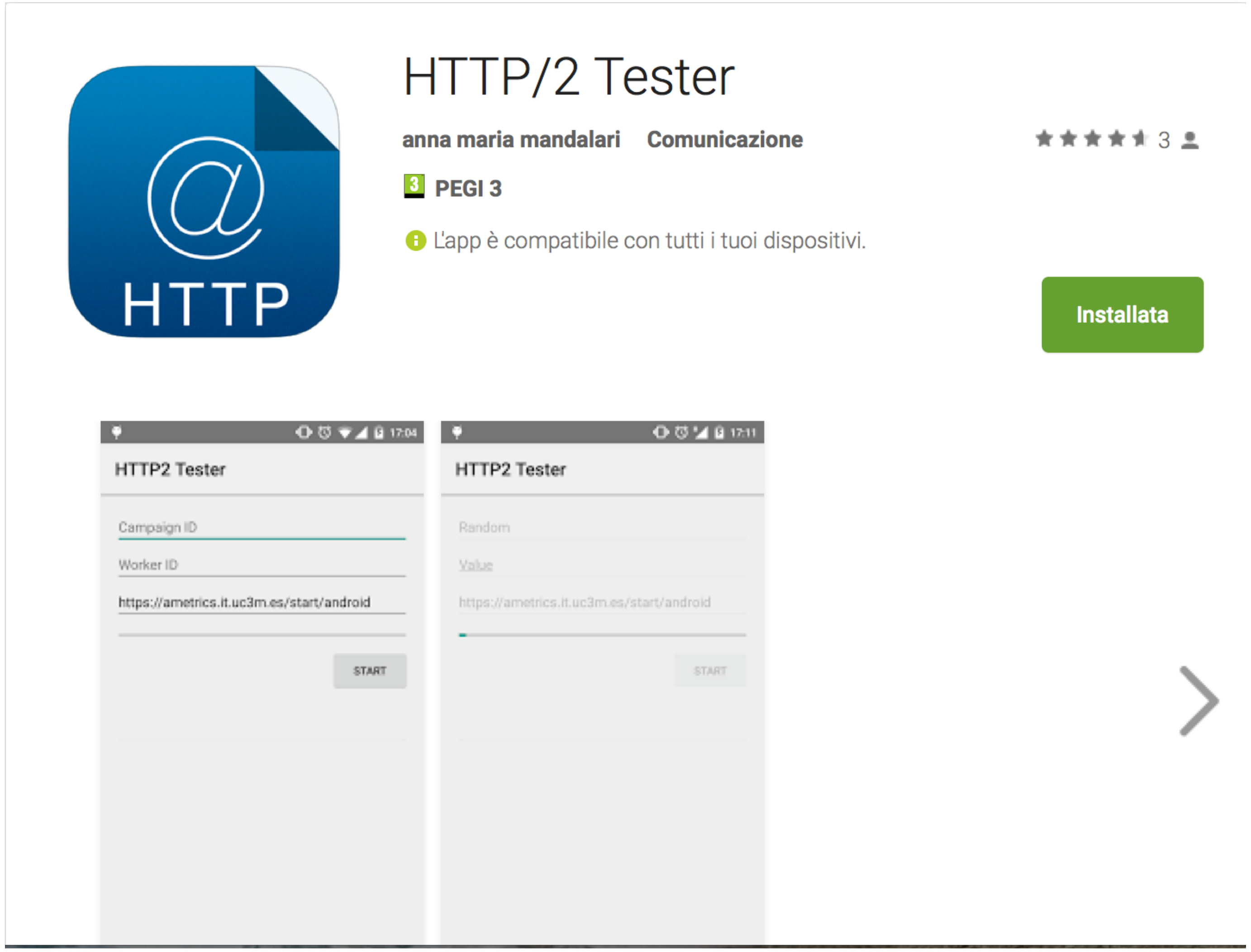
TOTAL: 403,30 $

12. Why don’t you use PlanetLab or other “free” platforms?

- The limited and often special position of testbed nodes
- No possibility to deploy your own test
- Fixed line only
- Access to the results

13. Is it possible to apply this method to other measurements?

Sure. I am using the same methodology to test other protocols like HTTP2 and TCP Fast Open.
I am improving it, deploying an Android App to get more information as possible from the devices and the network



HTTP/2 Tester

anna maria mandalari Comunicazione

3 PEGI 3

L'app è compatibile con tutti i tuoi dispositivi.

HTTP2 Tester

Campaign ID

Worker ID

https://ametrics.it.uc3m.es/start/android

HTTP2 Tester

Random

Value

https://ametrics.it.uc3m.es/start/android

14. Why do you consider the case of pervasive encryption?

After the public disclosure of the NSA global surveillance operations of foreign nationals and U.S. citizen, we observe a stronger tendency to encrypt traffic over the Internet. But, as is common knowledge, security and privacy do not come for free. HTTPS can increase the costs of a connection, significantly increasing latency, critical in mobile networks. This aims the creation of new solutions. In particular, a new protocol for security on Internet might be advantageous. For this goal many IETF Working Group (i.e. websec, ipsecme, tcpinc) has been created. In particular, the effort of tcpinc is to provide unauthenticated encryption and integrity protection at the TCP layer. However it essential to figure out how the middleboxes interact with the deployment of the proposed new protocols.

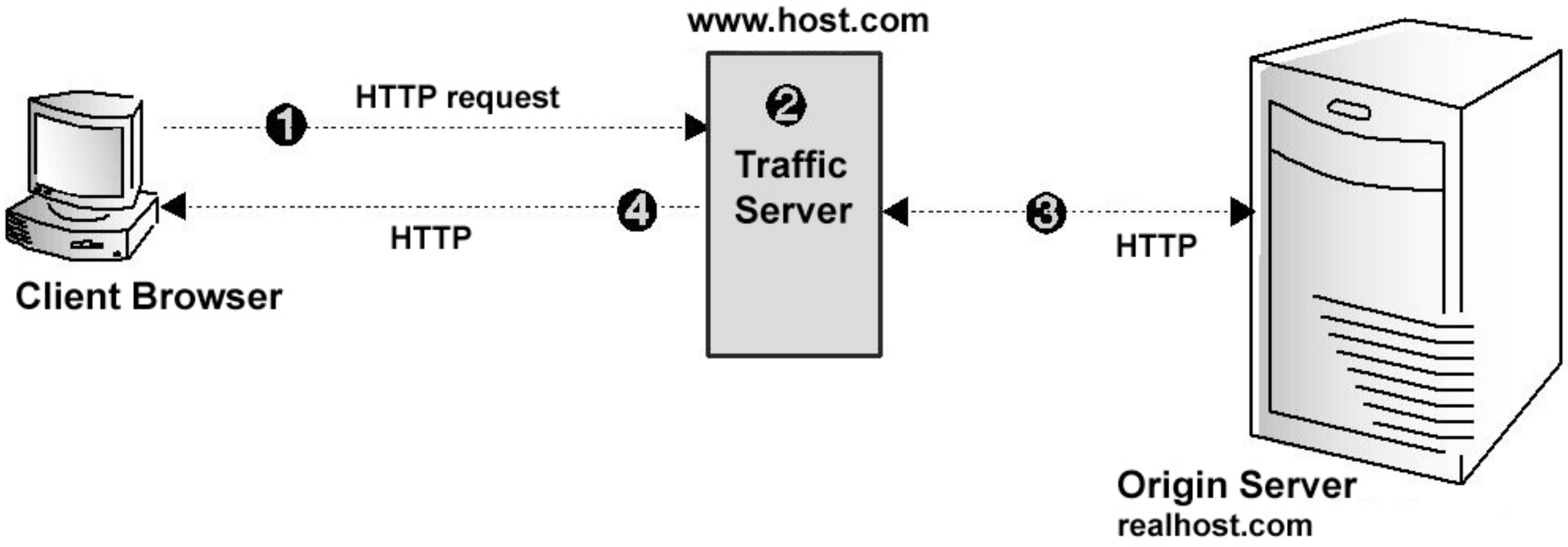
15. Have you tried to randomize the order of the port numbers?

Yes, I have. I changed the order of the i-frame, particularly for port 80.

```
<iframe width="1" scrolling="no" height="1" frameborder="0" src="https://
ametrics.it.uc3m.es:80/index.php" seamless="seamless"></iframe>
<iframe width="1" scrolling="no" height="1" frameborder="0" src="https://
ametrics.it.uc3m.es/index.php" seamless="seamless"></iframe>
<iframe width="1" scrolling="no" height="1" frameborder="0" src="https://
ametrics.it.uc3m.es:280/index.php" seamless="seamless"></iframe>
<iframe width="1" scrolling="no" height="1" frameborder="0" src="https://
ametrics.it.uc3m.es:488/index.php" seamless="seamless"></iframe>
<iframe width="1" scrolling="no" height="1" frameborder="0" src="https://
ametrics.it.uc3m.es:591/index.php" seamless="seamless"></iframe>
<iframe width="1" scrolling="no" height="1" frameborder="0" src="https://
ametrics.it.uc3m.es:593/index.php" seamless="seamless"></iframe>
```

16. What are the main reasons proxies block a TLS connection?

The proxies establish two separate connections: they terminate the TCP connection initiated by the client and they initiate a separate TCP connection between the proxy and the server.


www.host.com
1
HTTP request
2
Traffic Server
4
HTTP
3
HTTP
Client Browser
Origin Server
realhost.com