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

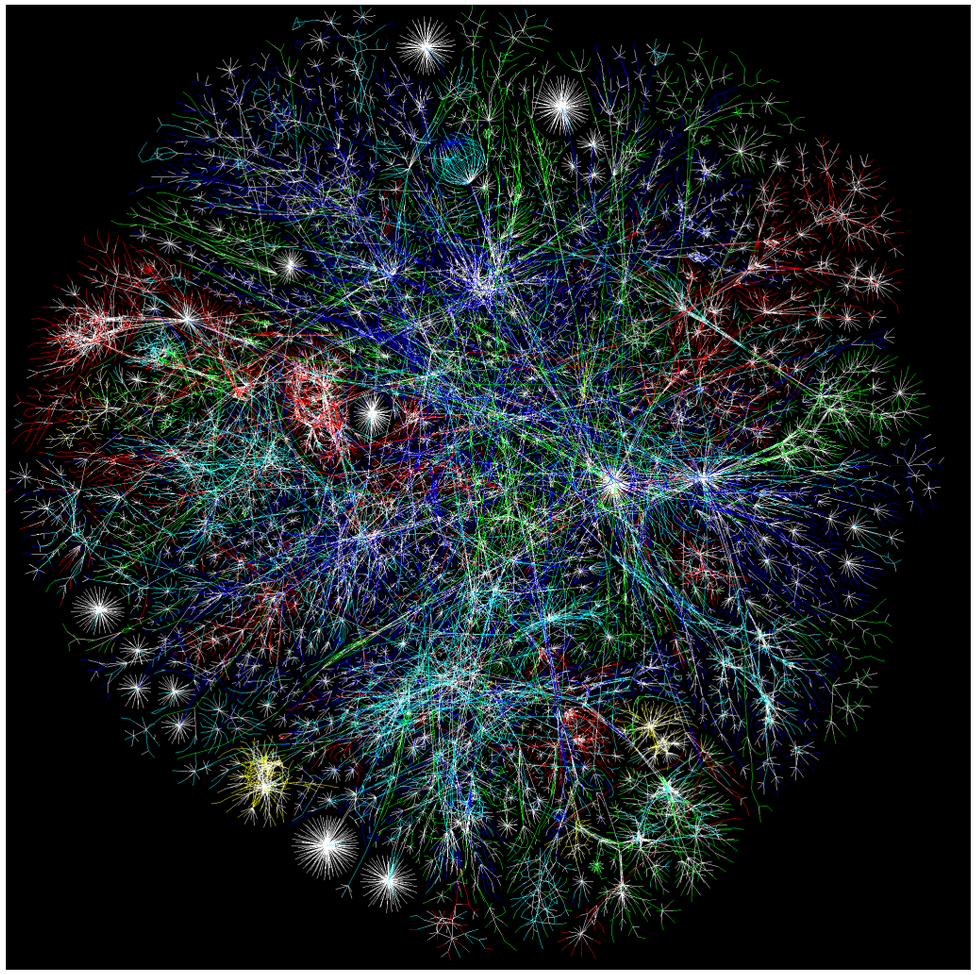
Internet Innovation

The Internet has successfully enabled multiple waves of innovation:

- 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.

Including:

- DCCP;
- UDP-lite;
- SCTP;
- Several extensions to TCP (e.g. ECN and LEDBAT).

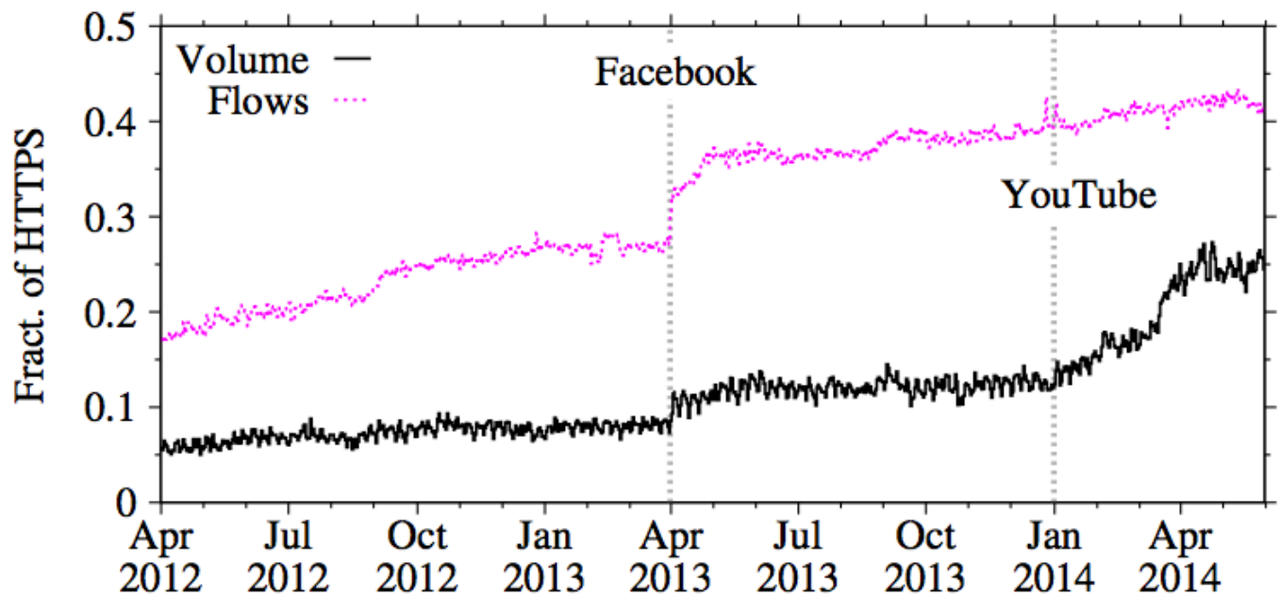
Is the Internet Ossified?

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

Perform large-scale Internet measurement campaigns

Crowdsourcing platform

Internet Connection Survey

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

Campaign/job ID	3b4ab5ce5e8f	Speed **96** [1-Slow 1000-Fast]	Verify+Rate Verify No Verify/Rate
Work done	**250/250** Add positions	You have **2** days to rate tasks	Auto-rating: Verify+Rate Satisfied
Workers will earn	**$0.25**		Folder **DEFAULT** → To ARCHIVE
Takes less than	**9** minutes to finish		
Targeted Countries	[International] -Macedonia -Indonesia -Lithuania -Bangladesh -Egypt -Morocco -Poland -Canada -Australia -Vietnam		

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:

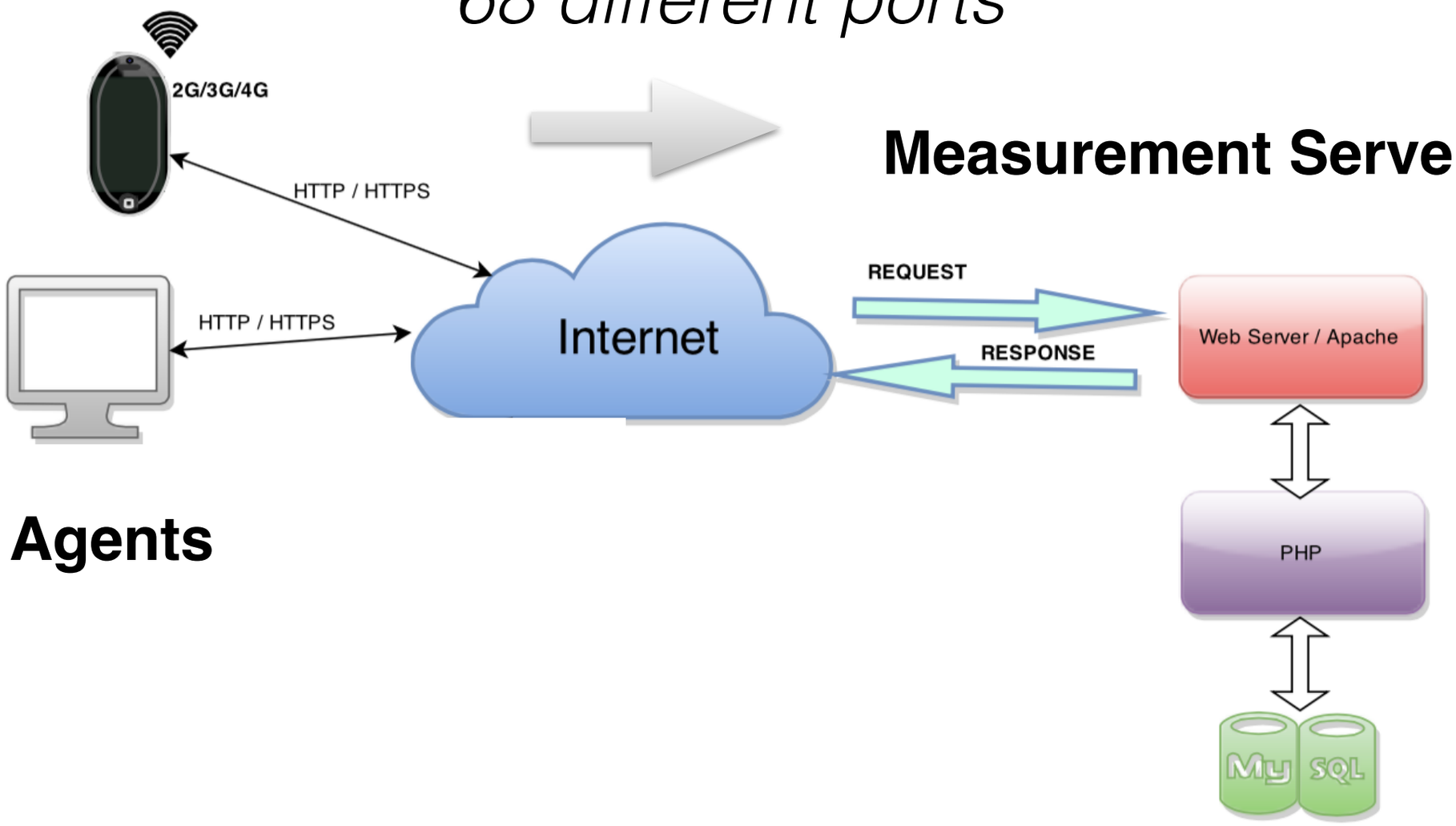
- World-wide access to employers;
- Automatic payment method based on a unique verification code;
- Possibility to select the MAs based on certain criteria, i.e. geographical location at the country level, the type of Internet access (fixed or mobile) or even the type of measurement equipment used to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

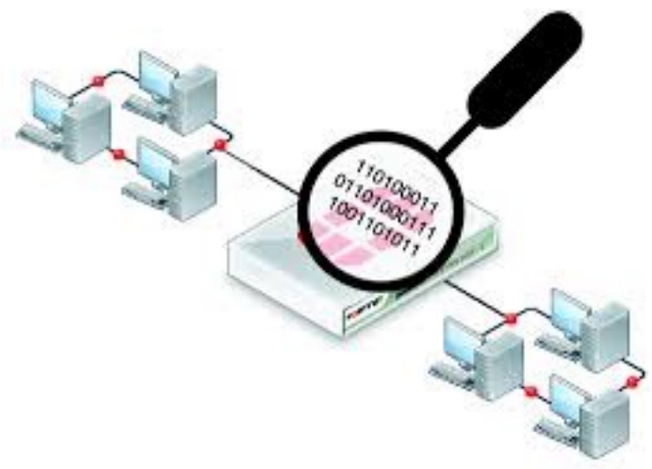


Measurement Server

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 we 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)

Submit

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)

Submit

Experimental setup: Measurement Agent Common Procedure

- We 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

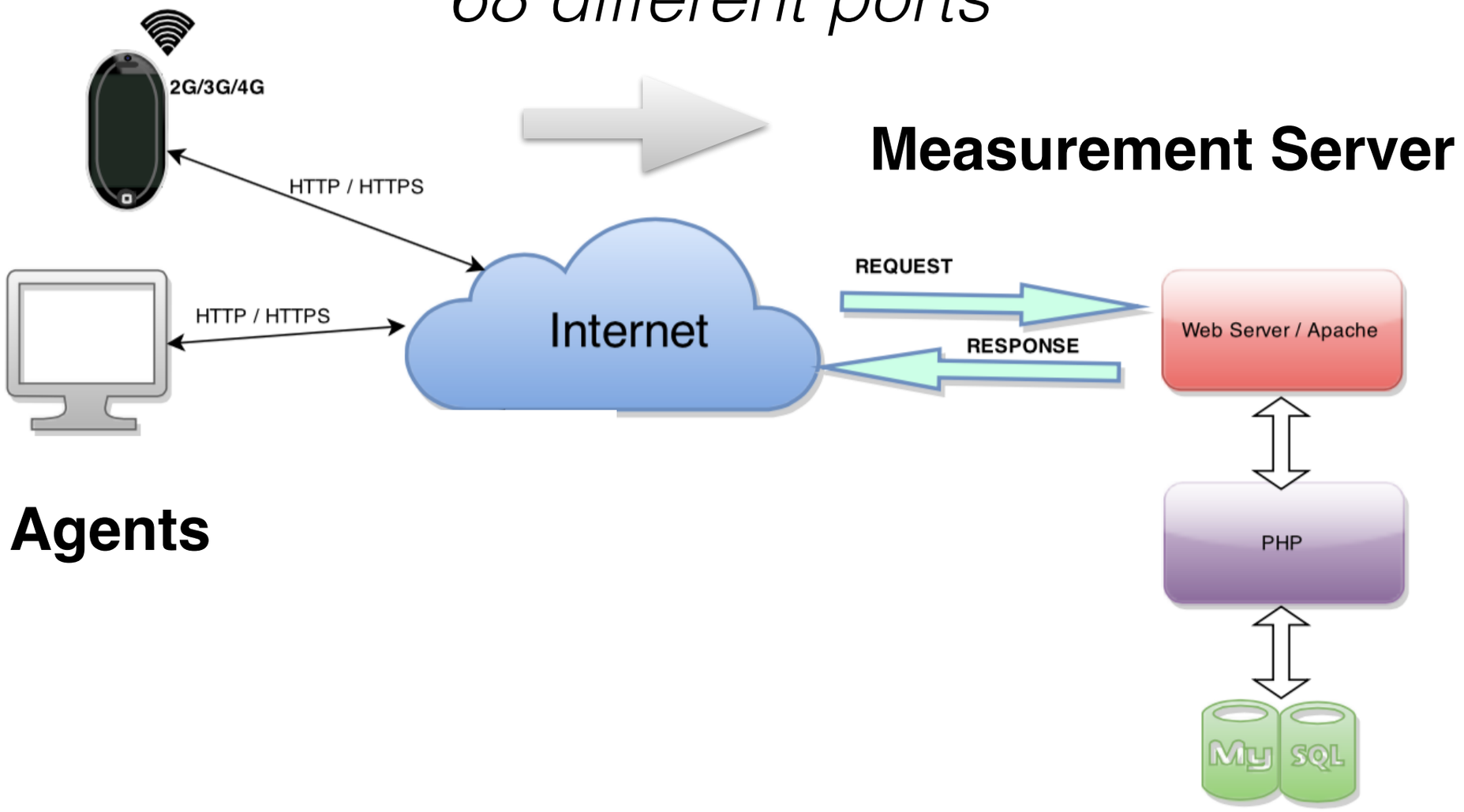
```
User-Agent: Mozilla/5.0 (X11; Linux i686 on x86_64; rv:10.0.2) Gecko/20100101 Firefox/10.0.2
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8
Accept-Language: en-us,en;q=0.5
Accept-Encoding: gzip, deflate
Cookie: .ASPXANONYMOUS=BLAH......;
WRUID=1243657642
DNT: 1
Connection: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



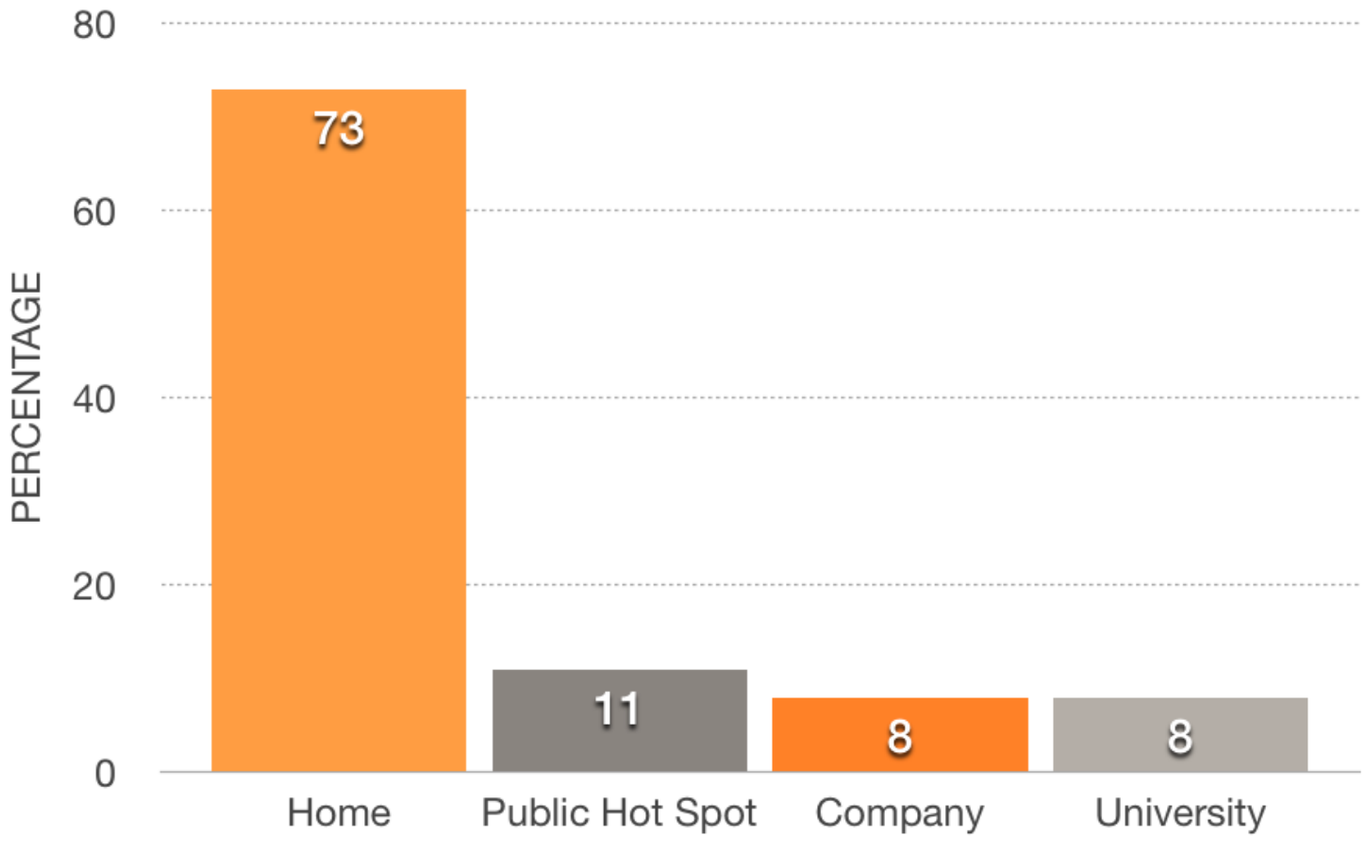
Measurement Server

Measurement Agents

Data Set

FIXED LINE:

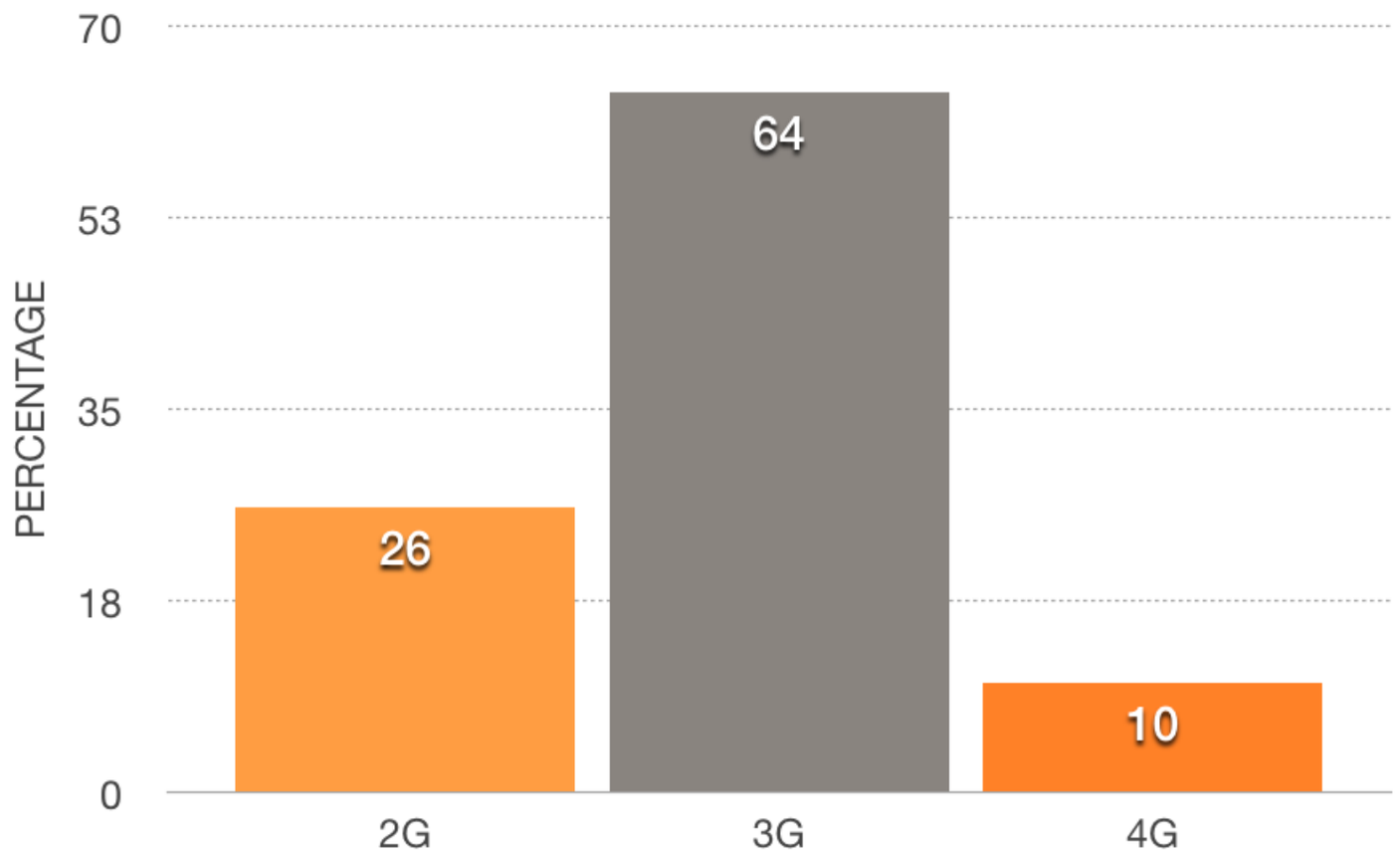
- 1,165 workers;
- 53 different countries;
- 286 ASes.



Data Set

MOBILE:

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

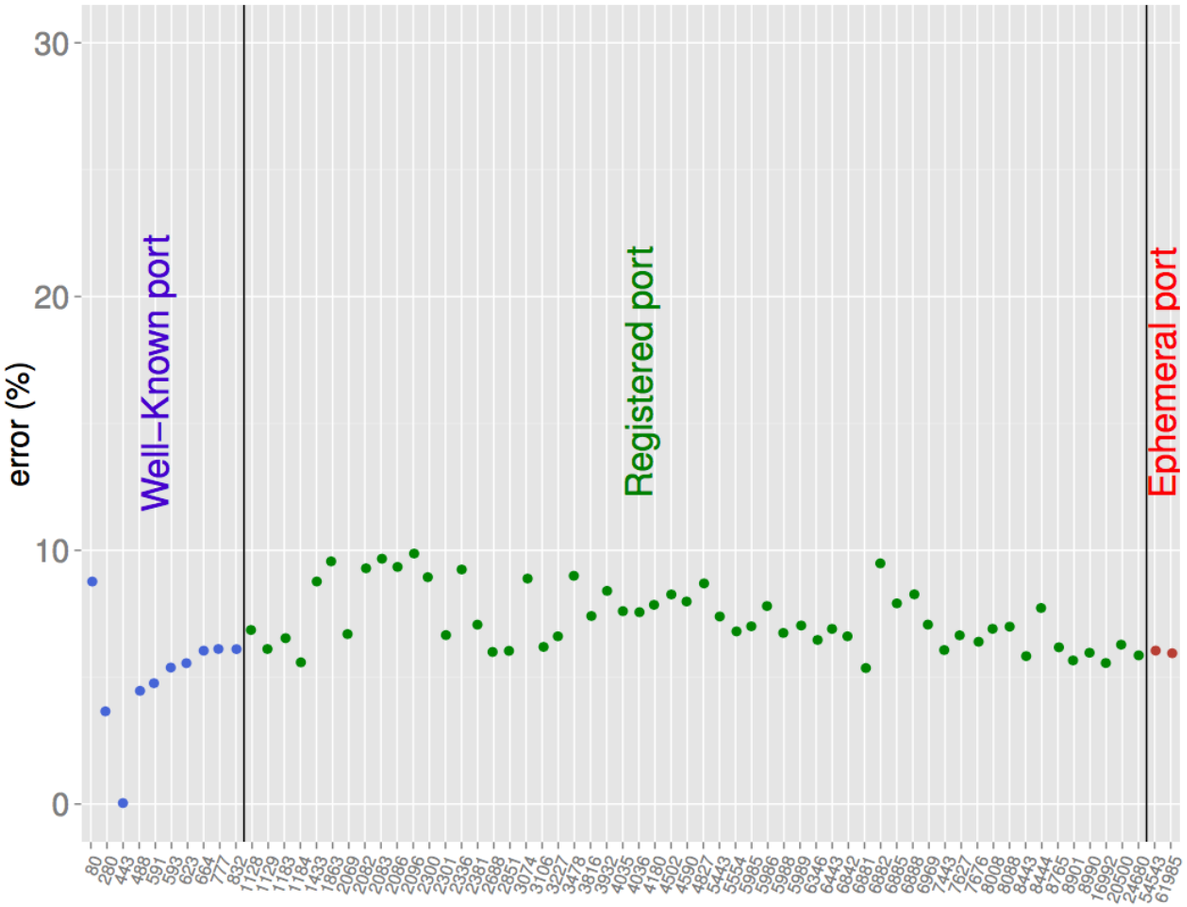


Total of 114,228 connections

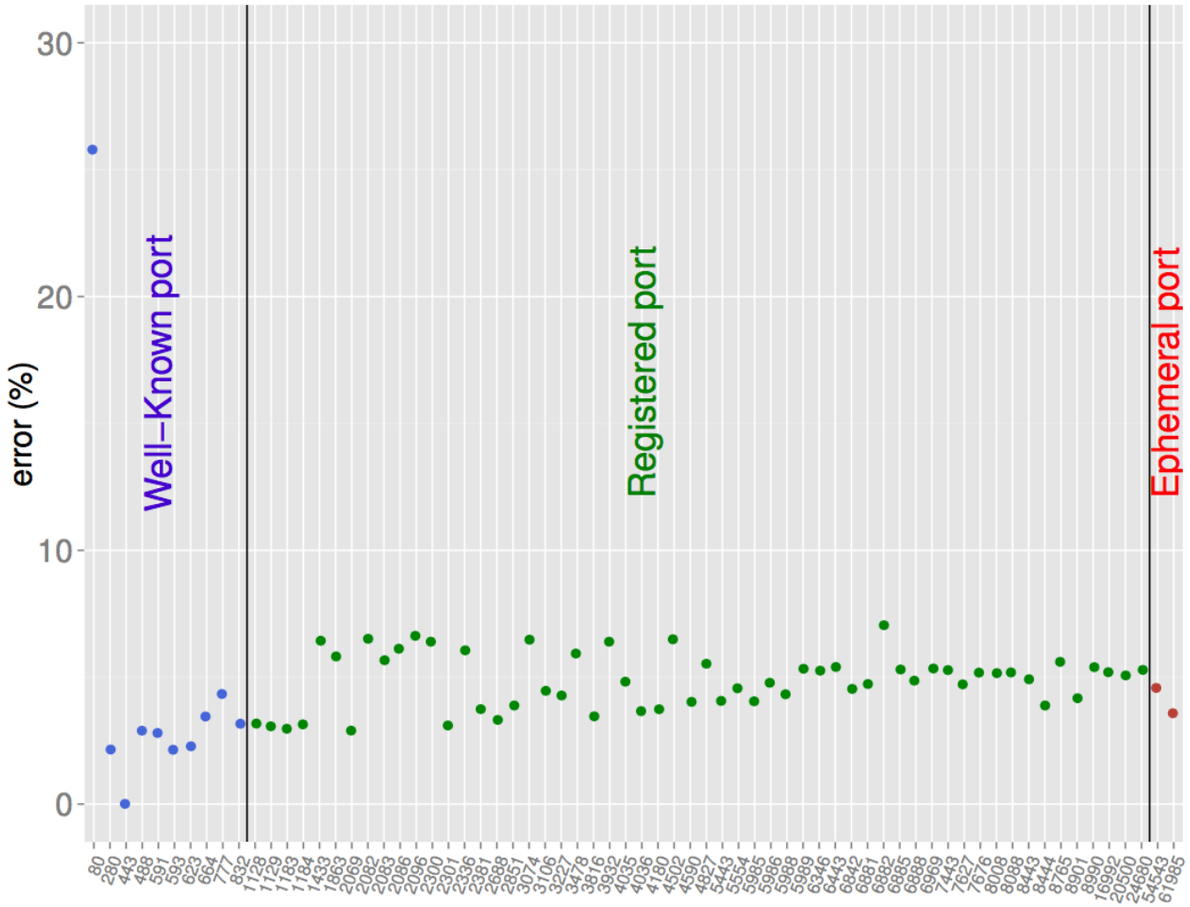
The data set is freely available on http://it.uc3m.es/amandala/dataset.php

Aggregated results

$$ERROR = (success\ [HTTP] - success\ [TLS])$$



a) Fixed line

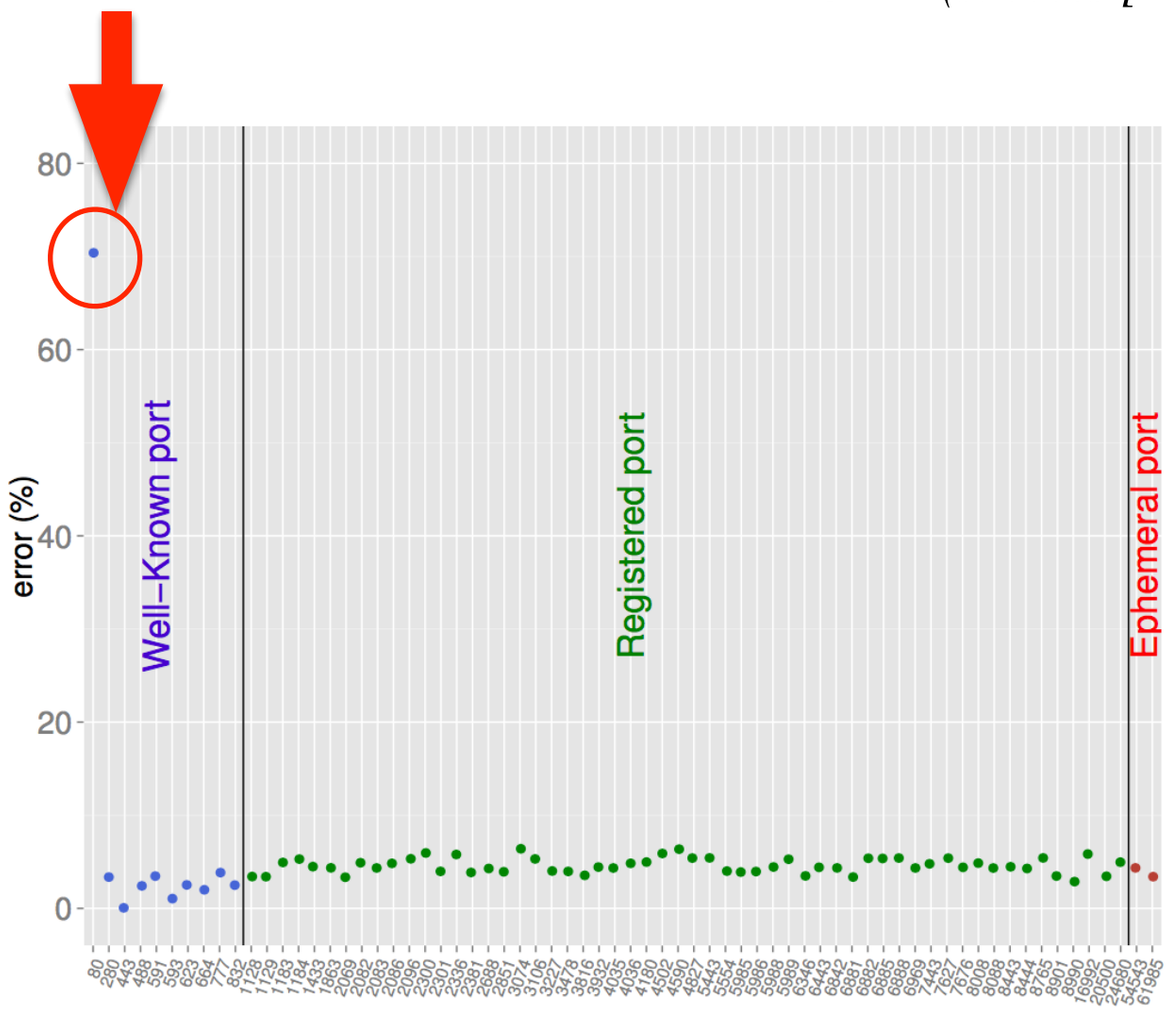


b) Mobile network

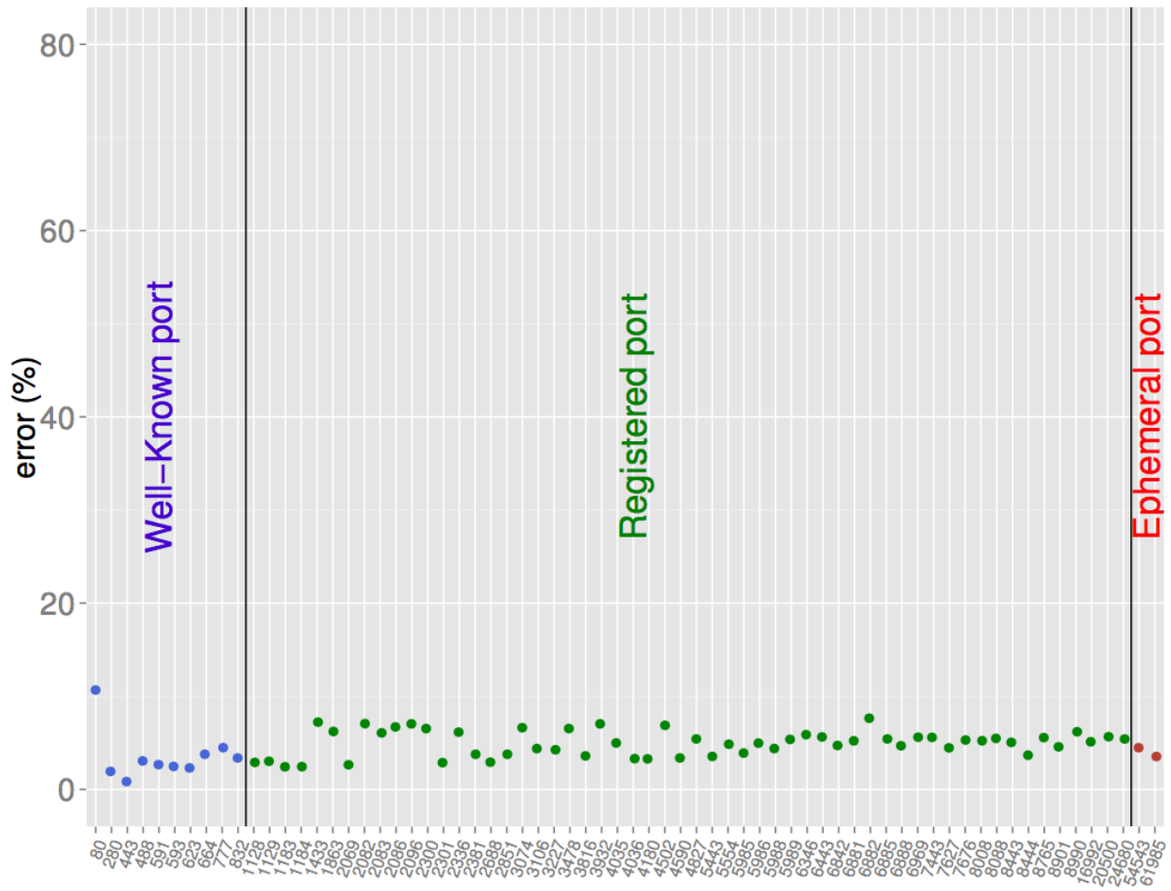
25% of the users are not able to perform a TLS connection over port 80 in mobile network.

Proxies

$$ERROR = (success\ [HTTP] - success\ [TLS])$$



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.

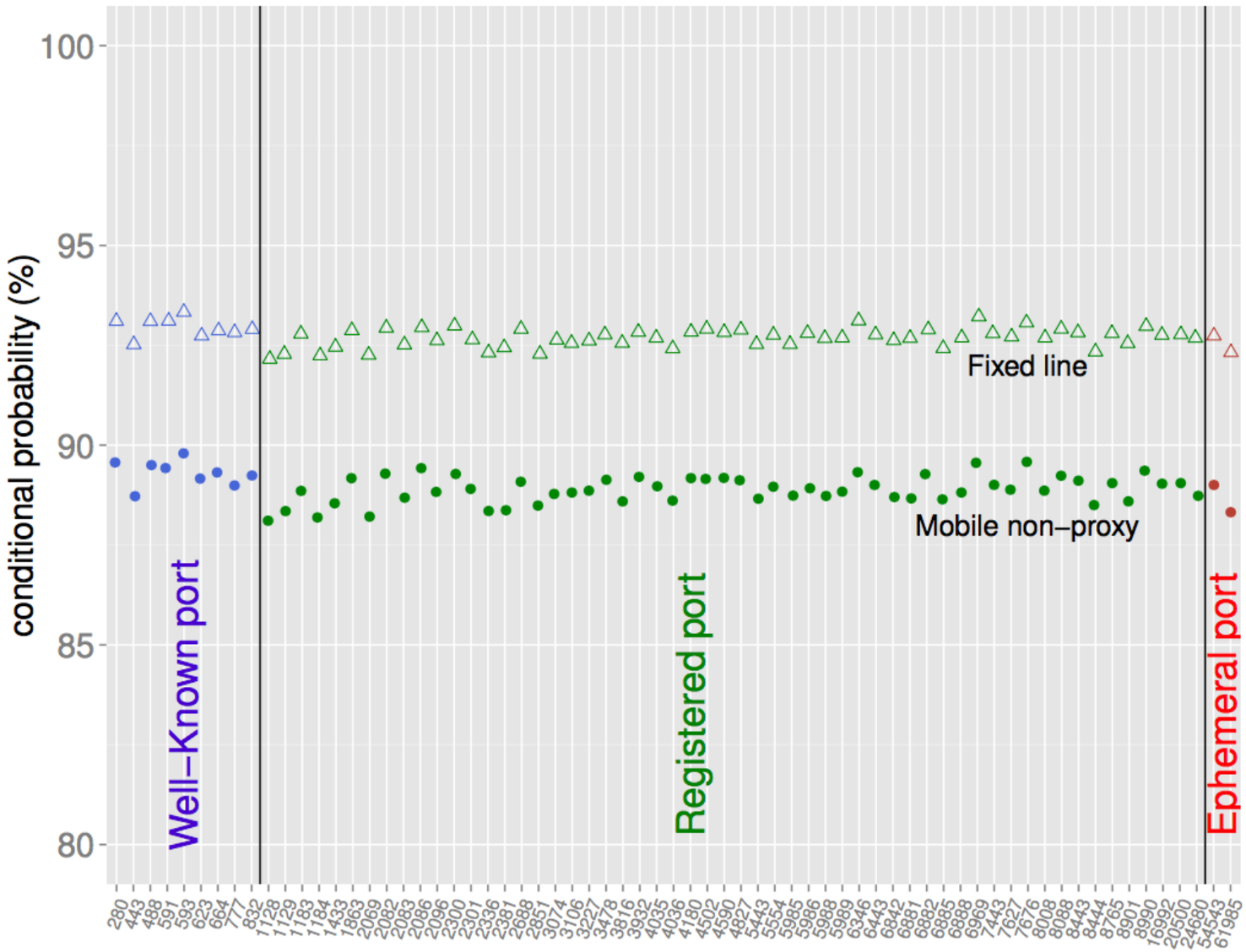
Packets analysis

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.