Internet Censorship in Italy: A First Look at 3G/4G Networks

Giuseppe Aceto[1,2], Antonio Montieri[2(✉)], and Antonio Pescapè[1,2]

[1] University of Napoli Federico II, Naples, Italy
{giuseppe.aceto,pescape}@unina.it
[2] NM2 srl, Naples, Italy
montieri@nm-2.com

Abstract. The techniques used to enforce Internet Censorship vary, and as a consequence the users can experience different results while accessing the same censored content in different contexts. While the corpus of Internet censorship studies is growing, to the best of our knowledge we are the first to focus on censorship detection on 3G/4G (hereafter *mobile*) network operators. After an introduction on the censorship detection platform and tests we adopted, we report the preliminary results of an experimental campaign we performed in Italy using the five major mobile operators. Our analysis shows that there is no homogeneity of treatment for a censored resource across different mobile operators, with 99.5 % of resources showing at least two different treatments, and the pairs of operators differing in the treatment of 32.5 % up to 99.5 % of censored resources. These results have significance regarding the transparency and precision of censorship, and the possibilities for circumvention and detection strategies.

Keywords: Internet censorship · Censorship detection · Active measurements · Mobile networks · Italy

1 Introduction

The regulatory action of governments over the access to online information has fostered the practice of Internet Censorship, i.e. the intentional impairing of a client application in reaching a requested resource or service, enforced by a third party (neither the user, nor the server operator) [5]. Such action can produce different effects, depending on the censoring technique, and often directly or indirectly causes a communication error, giving the user the false impression that an outage of some kind is the cause of inaccessibility. Moreover, the effectiveness, the side effects, and the means for circumventing the censorship are all dependent on the specific censoring technique that is applied. Finally, Internet Censorship

This work is partially funded by art. 11 DM 593/2000 for NM2 srl (Italy). This work has been also carried out thanks to a *Google Faculty Research Award* for the project UBICA (User-Based Internet Censorship Analysis).

S. Foresti and G. Persiano (Eds.): CANS 2016, LNCS 10052, pp. 737–742, 2016.
DOI: 10.1007/978-3-319-48965-0_53

varies over time, national borders, and network infrastructure (access provider, backbone networks). We have researched Internet Censorship in previous measurement campaigns [3,4], and a corpus of experimental studies on this topic is growing [6–10], often including detection methods and tools, but to the best of our knowledge none has focused on censorship detection from mobile phones, before this work. We refer to [5] for the definitions and an in-depth analysis of the state-of-art of Internet Censorship detection. In this poster we present the platform used, our methodology, and the preliminary results of our analysis of censorship as enacted by five major 3G/4G *Mobile Network Operators (MNOs)* in Italy, during a measurement campaign. More specifically, we characterize the results of the tests according to four different parameters, whose combination (or *aggregated behavior*) affects both the final outcome that a mobile user would experience, and the circumvention method that is effective in that case. In the preliminary results we report how variably a censored resource is managed, varying the operator, and a pairwise comparison of MNOs in terms of targets with the same aggregated. The results clearly show how there is a significant variation across different MNOs. We are performing further analyses on the dataset (not shown in this abstract), namely:

- detailing the most common aggregated behavior;
- reporting the distribution of behaviors per operator;
- deriving the circumvention techniques that are most likely to succeed with each MNO;
- evaluating the stability over time of the observed behaviors.

2 UBICA

UBICA (User-Based Internet Censorship Analysis) is a platform that provides users with a *censorship monitoring* system. Figure 1 shows the main components of UBICA architecture. The platform leverages a globally distributed deployment of *probes* belonging to different kinds (router-based, headless client, GUI-client) that are orchestrated by a central *Management Server*. The platform provides: (i) dynamically updated censorship tests; (ii) dynamically updated targets to be verified; (iii) support for different types of probing clients; (iv) automatic censorship detection and censorship technique identification. The client has been designed to be highly portable, composed of a core measurement-related part and leverages standard UNIX utilities and mature network diagnostic tools.

The probes perform *active measurements* to collect evidences of censorship, periodically retrieving a list of test requirements (i.e. target lists and code) from the Management Server. After the evidence collection, each probe packs all the results in a report file and uploads it back to the Management Server. The reports are asynchronously parsed by such server and the significant information is stored in a SQL database. The *Analysis Engine* periodically processes data in the database, performing the censorship detection analyses through the following measurements.

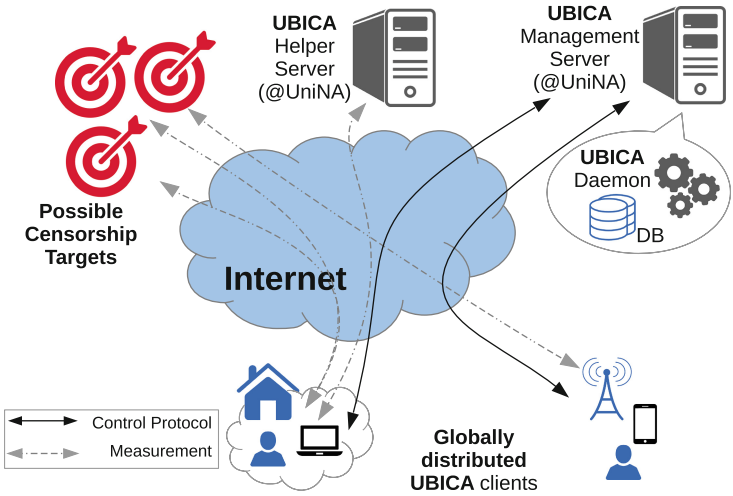

Fig. 1. UBICA architecture diagram.

DNS resolution. Given a fully qualified domain name, a DNS request of `type A` is issued from the probe towards its default resolver. The tool used to issue the request is `nslookup`. To distinguish among different DNS tampering techniques [5], the same request is issued also towards one or more open resolvers, used as control resolvers from inside the censored network.

TCP reachability. This test tries to set up a TCP connection to verify a possible filtering triggered by *IP:port*, starting a three-way handshake with a given timeout. The input parameters are *targetIP:port* and a timeout value in seconds.

HTTP reachability. An HTTP GET request is issued by this test: the response—or lack of it—and additional application level values are collected from the server. The HTTP header field *User-Agent (UA)* is conveniently set choosing it from a list previously defined (see Table 1). The tool used to issue the request and collect application level information is `curl`. The report from this test includes several values, such as content type, HTTP response code, number of redirects, etc., not reported for the sake of brevity.

3 Methodology

Experimental campaigns conducted in this work by mean of UBICA leverages headless clients equipped with Kubuntu 14.04 and connected to Internet through smartphones—tethering USB—acting as gateways. In Table 1 a summary of the factors taken into account is provided. Notably, selected MNOs account for the 96.6 % of the Italian market, and PosteMobile owns the 52.1 % of the Mobile Virtual Network Operators' market share [1]. An up-to-date list of possibly censored

targets has been obtained from [2], containing websites both blocked from judicial authority and suggested by the community of users. Since *DNS-tampering* is a widely used censorship technique, name resolutions are performed through both the MNO-provided (*default*) and the Google Public DNS (*open*) resolver. Finally, the list of UAs has been conveniently chosen for testing both mobile and desktop agents.

Table 1. Summary of factors and considered values.

Factor	Values
MNO	H3G, PosteMobile, TIM, Vodafone, Wind
Target	200 censored targets from [2]
DNS	Default (MNO-provided), Open (Google Public DNS)
User-Agent (UA)	Safari 5.1 (iPhone - iOS 5.0),
	IEMobile 7.11 (HTC Touch 3G - Windows Mobile 6.1),
	Google Chrome 41.0 (Desktop - Windows 7)

When a user requests a resource from a target, he experiences a number of different behaviors depending on the specific combination of the factors reported in Table 1. In order to facilitate their description, the possible outcomes a user can experience have been clustered into *aggregated behaviors*. In more detail, for each combination of MNOs and targets taken into account: (i) default and open DNS resolutions can be equal or different, (ii) the redirections a request is possibly subjected to, can be dependent on the UA or not, (iii) default and (iv) open DNS resolutions can return various outcomes.

Indeed, when the default resolver is leveraged, the DNS server could reply with a *forged response* not corresponding to the legitimate DNS database entry (i.e. *DNS hijacking*). More specifically, a forged response is a Resource Record of `type A` containing an IP address that does not correspond to the actual IP address obtained from the legit resolution of the requested resource [5].

Since in this case the DNS resolver acts as the *censoring device*, changing the default resolver with an open resolver will bypass the censoring device and thus allow open access to the Internet. Even though a user can correctly obtain the requested content leveraging the Google Public DNS, he could also experience a number of erroneous outcomes. Instead of the expected Resource Record, the open resolver might return an error response `NXDOMAIN` of type "no such domain". Moreover the request could incur a *connection timeout*, a connection termination by *TCP reset*, or an *HTTP error* response (i.e. 4xx and 5xx status codes).

4 Preliminary Results

In this section we provide an overview of the factors that mostly influence the browsing experience of a general user requesting a resource from a censored target. The dataset introduced in this work has been collected through preliminary

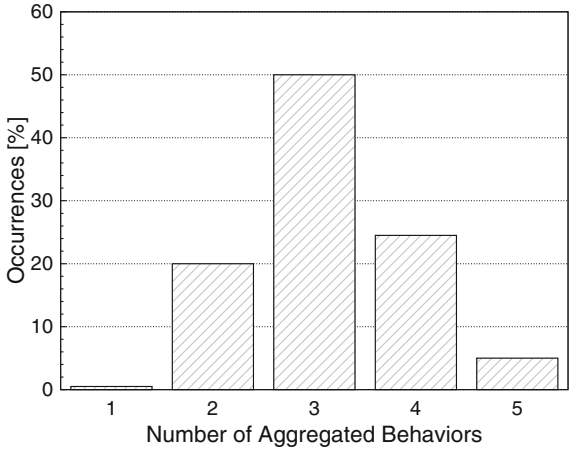

Fig. 2. Percentage of targets exhibiting a different number of aggregated behaviors when MNO is changed.

experimentations conducted in February 2016. An interesting result stemming out from these experimentations is the relationship existing between the MNO leveraged to access to the Internet and the aggregated behaviors observed. More in details, given a target, the number of different behaviors experienced by a user when he changes the MNO used to connect to the network, is an index of how differently each MNO treats various targets possibly censored. As shown in Fig. 2, 0.5 % of the targets (i.e. only 1 target) have the same aggregated behavior for all the MNOs, whilst the majority of them (100 out of 200 targets) exhibits 3 different behaviors. These results confirm that varying the MNO that offers connectivity, a user might experience distinct outcomes even in the case he wanted to retrieve the same content. However, 95 % of the targets exhibit at most 4 aggregated behaviors, showing at least 1 behavior in common between 2 MNOs.

Table 2. Pair-wise variation in censorship application between MNOs. A 100 % variation means that all targets have different behaviors between considered MNOs.

MNO	PosteMobile	TIM	Vodafone	Wind
H3G	92.5 %	32.5 %	94 %	75 %
PosteMobile		99 %	60 %	95 %
TIM			99.5 %	65.5 %
Vodafone				65 %

Table 2 summarizes the variation in the aggregated behaviors obtained between analyzed MNOs. Lowest pair-wise variation has been observed for H3G and TIM, that show different behaviors only for 32.5 % of the targets. On the

contrary, TIM and Vodafone have almost always distinct aggregated behaviors (99.5 %). Notably, although PosteMobile is a *Mobile Virtual Network Operator (MVNO)* and offers its services leasing the radio spectrum and network infrastructures from Wind, they exhibit the same aggregated behaviors for only 10 out of 200 targets.

References

1. MVNO News - Osservatorio MVNO, July 2016. http://www.mvnonews.com/osservatorio-mvno/
2. Osservatorio sulla censura di Internet in Italia, July 2016. https://censura.bofh.it/
3. Aceto, G., Botta, A., Pescapé, A., Awan, M.F., Ahmad, T., Qaisar, S.B.: Analyzing internet censorship in Pakistan. In: IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (IEEE RTSI), Bologna, Italy, September 2016
4. Aceto, G., Botta, A., Pescapè, A., Feamster, N., Faheem Awan, M., Ahmad, T., Qaisar, S.: Monitoring internet censorship with UBICA. In: Steiner, M., Barlet-Ros, P., Bonaventure, O. (eds.) TMA 2015. LNCS, vol. 9053, pp. 143–157. Springer, Heidelberg (2015)
5. Aceto, G., Pescapé, A.: Internet censorship detection: a survey. Comput. Netw. **83**, 381–421 (2015)
6. Burnett, S., Feamster, N.: Encore: lightweight measurement of web censorship with cross-origin requests. SIGCOMM Comput. Commun. Rev. **45**(4), 653–667 (2015)
7. Chaabane, A., Chen, T., Cunche, M., De Cristofaro, E., Friedman, A., Kaafar, M.A.: Censorship in the wild: analyzing internet filtering in Syria. In: Proceedings of the 2014 Conference on Internet Measurement Conference, IMC 2014, pp. 285–298. ACM, New York (2014)
8. Di Florio, A., Verde, N.V., Villani, A., Vitali, D., Mancini, L.V.: Bypassing censorship: a proven tool against the recent internet censorship in Turkey. In: 2014 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW), pp. 389–394. IEEE (2014)
9. Ensafi, R., Knockel, J., Alexander, G., Crandall, J.R.: Detecting intentional packet drops on the internet via TCP/IP side channels. In: Faloutsos, M., Kuzmanovic, A. (eds.) PAM 2014. LNCS, vol. 8362, pp. 109–118. Springer, Heidelberg (2014). doi:10.1007/978-3-319-04918-2_11
10. Jones, B., Lee, T.-W., Feamster, N., Gill, P.: Automated detection and fingerprinting of censorship block pages. In: Proceedings of the 2014 Conference on Internet Measurement Conference, IMC 2014, pp. 299–304. ACM, New York (2014)