Internet Regulation, a New Approach: Outline of a System Formed to Be Controlled by the Internet Users

Nikolaos Koumartzis and Andreas Veglis

Media Informatics Lab, Dept. Journalism & Mass Communication, Aristotle University of Thessaloniki, Thessaloniki 54 625, Greece

Received: October 07, 2011 / Accepted: November 01, 2011 / Published: January 25, 2012.

Abstract: This paper describes an outline for the proper design of a Fair Internet Regulation System (FIRS), i.e., a system that will be implemented in a national level and encourage the participation of Internet users in enriching and correcting its “behavior”. Authors aim to design a system that will be operated in some extent by the Internet users, and so it will be easier to be accepted by Western democracies willing to implement a fair Internet regulation policy. Last, the authors state the importance of using well-designed surveys prior to the implementation of FIRS, announce the launch of an online tool (WebObserver.net) and invite researchers to be part of this international effort.

Key words: Content filtering, content blocking, internet, regulation, blacklist.

1. Introduction

Since the beginning of the World Wide Web in 1990 (with the introduction of HTML, short for HyperText Markup Language [1]), the way Internet users access the web changed dramatically. At the beginning, users were able to access websites via a simple and direct procedure (as described in Fig. 1), but from 1996 and onwards the first blocking techniques were introduced [2].

Today, Internet regulation is on the rise: in 2006 OpenNet Initiative stated that at least 26 countries were using content blocking systems [3], while in 2009 Reporters Without Borders stated that “some sixty countries experienced a form of Web censorship, which is twice as many as in 2008” [4]. Moreover, according to related surveys conducted on massive scale, Internet users are split in the half regarding where they stand as far as Internet regulation is concerned [5], while smaller surveys (focused on a highly educated sample) showed that the majority of Internet users prefer the implementation of some short of “open” Internet regulation system (i.e., a system that they will be able to interact with, enriching or correcting its database) than no Internet regulation at all [6-8].

As stated by John Palfrey, executive director of the Berkman Center for Internet and Society at Harvard Law School, this issue is an open question, as “Some people would say that certain kinds of information should be banned” [9]. There is a worst case scenario though, as described by S.N. Hamade in which “filtering creeps into the system in an ad hoc way without formal evaluation of the standards […]” [10].

Fig. 1 Usual way of accessing the Internet.
In order to avoid such a scenario, an Internet regulation system must be developed fair enough to be accepted by the majority of the Internet users. This has to be done via collaboration between those who regulate (i.e., governments) and those who are being regulated (i.e., Internet users, Internet Service Providers etc.), a cooperation that can be proved profitable for both sides (as clearly stated in Ref. [11]). Moreover, the implementation in national level of such a system must be cost-effective in order to be feasible [12-13].

The authors supports that even if the problem of Internet regulation is not of a technical nature, the solution can be found via a proper technical design. They propose a blueprint for a “Fair Internet Regulation System” (FIRS) based on four main factors: (1) no need for excessive resources; (2) fast function; (3) interaction between users and system; and (4) “discretion” while handle specific kinds of illegal online content.

The paper is organized as follows: It begins with a brief presentation of the most-in-use blocking mechanisms today (section 2), and continues with the analysis of two Internet regulation systems already implemented in national level that can be used as a “guide” (section 3). Afterwards, it describes how an effective FIRS can be designed with the aid of surveys (see section 4.1), via a careful categorization of the targeted content (see section 4.2) and via encouraging Internet users to participate in the whole procedure (see section 4.3). Last, a blueprint for FIRS is being presented in technical terms (section 4.4) along with its advantages and disadvantages (see section 4.5 and 4.6) while possible future work is being discussed (section 5).

2. Mechanisms for Internet Regulation

Starting this analysis from a technological point of view, there are two main categories (as described in [14]) of Internet regulation mechanisms: (1) IP (stands for Internet Protocol) and URL (stands for Uniform Resource Locator) blacklists, and (2) real-time filtering.

Systems based on IP and URL blacklists are easier to be implemented and faster to function on a massive scale. Their main disadvantage is that they are easier to be circumvented in case they are noticed (by quite few methods, as those described in Refs. [7, 13, 15]) and keener to overblock content (i.e., a whole website can be censored because of a single webpage) [7].

On the other hand, systems based on real-time filtering are still on an embryonic stage. When effective Artificial Intelligence (AI) techniques will be developed, those systems can be extremely precise on what they block. Until then, there are many issues to be solved such as fast and intelligent operation, enough processing power without excessive financial resources, etc. [14].

For all the above, every Internet regulation system design willing to be used in a massive scale should be based mainly on IP and URL blacklists techniques (as described in Ref. [10]).

2.1 Packet Dropping (IP Blocking)

Packet Dropping is the simplest technique and its function is based on a list consisted of websites’ IP addresses to be blocked (see Fig. 2). Users’ requests for these IP addresses are discarded which lead to no connection with the requested server. Its most important advantage is that it can identify the type of IP and thus implement selective filtering, i.e., blocks HTTP (stands for Hypertext Transfer Protocol) packets for a particular IP address but leave email unblocked. Its crucial disadvantage is that it blocks all the web content of a particular IP address, as it is a massive and not accurate blocking system [7].

![Packet dropping system](image-url)
For example, past experiments have shown that there is a significant risk of overblocking with systems based exclusively on IP addresses [16].

2.2 Content Filtering (URL Blocking)

Content filtering is used to block very specific items (images, video, etc.) in a website. The system (as described in Fig. 3) is based on URL examination and so it is very accurate on blocking exactly what is in a list of blocking URLs [7].

Its main advantage is that there are no overblocking issues. All the traffic must be passed through a web proxy, something that needs the appropriate equipment to handle the load. Moreover, to avoid a system failure, the equipment must be replicated and that increases the final cost [13]. So, even if it is the most precise mechanism, it is the most expensive too.

3. Existing Internet Regulation Systems as a “Guide”

In order FIRS to be designed properly, there are two main issues to be considered: (1) What kind of mechanisms and in which way should be used, and (2) What kind of interaction the system should have with the Internet users. For the above, there are systems already in use around the world that can be considered as a “guide”.

3.1 Hybrid Systems: The Paradigm of CleanFeed in UK

In order to overcome the crucial disadvantages discussed above, various hybrid systems have been developed. UK’s CleanFeed (implemented in the massive British Telecom’s network, also known as BT) is a two-stage hybrid system that can be used as a “guide” in some extent for designing FIRS [7] (see Fig. 4).

In brief, the first stage resembles a packet dropping system that doesn’t discard requests but redirects them to the second stage where a web proxy resembles a content filtering system (as described in Ref. [13]). Those stages use two different lists of banned websites in order to decide what to do with the requests (the first stage use an IPs list, while the second a URLs list), both of them extracted by a blacklist provided by an non-governmental body (Internet Watch Foundation).

CleanFeed’s main advantage is that it is fast, accurate and can function at a low cost (only 5 server run the whole BT network, according to [7]). Its hybrid design though can be circumvented easily. Moreover, CleanFeed considered a “silent” system, as it doesn’t inform the user when he tries to visit a blocked website (it serves a “404 error message” that can mean quite many different issues). This feature is the main reason why many UK-based IT experts and academics opposed to its use [17-20].

3.2 Saudi Arabia System

In our search for the ideal interaction between Internet users and an Internet regulation system, Saudi Arabia’s paradigm can be considered as a “guide”. In brief, when Internet users in this country try to visit a prohibited website, a webpage is presented explaining that this site is blocked, providing the option to submit a complaint, and giving a link to state’s Internet regulation
policy (see Fig. 5) [21]. By a technical point of view, the system uses content filtering mechanisms to block websites included in two blacklists maintained by Internet Services Unit (department of King Abdulaziz City for Science and Technology organization) [22].

In conclusion, it is an “open” Internet regulation system that encourages users to be actively engaged with enriching and correcting the blacklists. The main reason behind general public opposition is that it is implemented in a monarchical state, and even if the system gives the option to Internet users to become part of the Internet regulation policy, this is not done due to the given political environment.

4. Designing a Fair Internet Regulation System

Taking into consideration those examples, FIRS must be highly adaptable to each country special political needs in order to be accepted by the general public. The aim of the blueprint discussed below is the development of an effective, fast and low cost system that will encourage Internet users to participate in the whole procedure, giving them the opportunity to enrich and correct its “behavior”. At the same time, this system must be able to handle specific kinds of online illegal content with “discretion”. A detailed outline of the proposed F.I.R.S is included in Fig 6.

4.1 The Importance of Well-Designed Surveys

In order public opinion to accept the implementation of an Internet regulation system, each society’s special needs must be taken into consideration from the very first steps of the designing procedure.

These needs differ from country to country, and so in each society must be found via well-designed surveys conducted on a great proportion of population or (in cases there are no funds for such surveys) on small highly educated and familiarized with the use of the Internet samples. Such surveys can provide researchers with answers regarding: (1) whether the general public is keen to accept and participate in some form of Internet regulation; (2) who is believed to be more suitable to operate such a system; (3) what kind of content should be targeted etc.

This paper will use Greece as an example, and a survey conducted at Aristotle University of Thessaloniki by the authors (June 2010) on a small but highly educated sample [23] consisted mainly by MA and PhD students plus teaching stuff of the Department of Journalism and Mass Communication. It produced valuable data that should be analyzed in depth, but in brief showed that 37.9 percent prefer such a system to be operated by university-based institutes (contrary to only 11.5% who trust a governmental service based inside a ministry) and regarding what content should be targeted participants pointed: (1) pornographic websites (35.1 percent); (2) hate speech content (34 percent); (3) defamation content (13.4 percent); (4) multimedia illegal sharing websites (7.2 percent).

4.2 Targeted Content and Categorization

Based on the survey’s results, a body must be chosen in order to be responsible for providing lists of websites to be blocked (The lists will be partially based on Internet users feedback). In the example of Greece, a combination of university-based institutes and non-governmental organizations seems to be a highly acceptable solution for the general public. This preference to non-government controlled bodies is crystal clear to many countries, according to many field’s experts too [24].
Next, it has to be decided what illegal content will be targeted. For example, child pornography content is considered to be worldwide a general accepted choice [10], while other kinds of online content must be evaluated separately in each country.

This paper proposes the further categorization of the targeted content to a. unquestionable must-be-filtered content and b. contradictable must-be-filtered content. In Greece example it is clear that child pornography goes to category a., while the other three kinds of content should fall in category b. in case it is decided to be targeted. As a result, two different URLs lists are formed: Blacklist A (including unquestionable must-be-filtered content) and Blacklist B (including contradictable must-be-filtered content).

4.3 System-User Interaction

Regarding the interaction between the system and the users, the authors propose a hybrid Internet regulation system not only in technical terms but also in terms of content management (and feedback for the users).

In theory, the system will be able to identify if the user tries to access: (1) no-blocked websites; (2) contradictable must-be-filtered content; (3) unquestionable must-be-filtered content.

In the first situation, it will give free access to the user. In the second situation, it will block the website by stating that it is blocked and by providing an option to submit a complaint (and giving a link to state’s Internet regulation policy). Last, in the third situation it will treat the content with “discretion”, which means that it will not inform the user for the blocked website and will serve him a “404 error message”. This will be done in order not to give to the determined users the ability to identify child pornography URLs and try to use circumvention tools to access them.
4.4 Technical Aspects

Regarding the technical aspects of the FIRS, it will be a three-stage hybrid system, using (quick but not accurate) IP blocking mechanisms for the first two stages and a (precise but time-consuming) URL blocking technique for the third stage. Moreover, it will use four different lists produced by the Blacklists A & B discussed in section 4.2:

1. IP1 List: A list of IPs that comes up from the URLs of both the Blacklist A and Blacklist B. It is used in the Stage 1 Check (see Fig. 7);
2. IP2 List: A list of IPs that comes up from the URLs of the Blacklist A. It is used in the Stage 2 Check (see Fig. 8);
3. URL1 List: A list with URLs identical to Blacklist A. It is used in the Stage 3a Check;
4. URL2 List: A list with URLs identical to Blacklist B. It is used in the Stage 3b Check.

The reason behind the four lists model should be clear for anyone who understands the relation between IP and URL. Quite simplified, IP address is the numerical name that computers use in order to locate other computers on the internet [25], while URL is a descriptive name which describes how a computer can “fetch” a resource (text, image, etc.) on the internet [26]. So, a URL may contain an IP address in order to define in which computer the resource is present. The most important thing to understand is that many URLs can have the same IP and not the opposite [7].

The three stage checking procedure is straightforward (see Fig. 6).

In Stage 1, the system defines if the user tries to access a “suspicious” website by using a quick packet dropping (IP blocking) mechanism. If the IP the user tries to visit in not in the IP1 List, then the system serves the website to the user without further checking. If the IP is in the IP1 List, then the system proceeds to Stage 2.

In Stage 2, the system determines if the user tries to access a “suspicious” website of category a or b. (as described in section 4.2) by using a quick packet dropping (IP blocking) mechanism. If the IP the user tries to visit is in the IP2 List, then the system proceeds to Stage 3a. If not, the system proceeds to Stage 3b.

In Stage 3, the system determines whether the user tries to access a blocked URL by using a precise and time-demanding content filtering (URL blocking) mechanism. So, in Stage 3a the FIRS checks if the requested URL is in the URL1 List. If it is, the system shows the user a “404 error message” webpage. If not,
it serves the website without further checking. In stage 3b, the system checks if the requested URL is in the URL2 List. If it is, the system shows the user a block-webpage informing him for the blockage of the website, giving him an option to submit a complaint and providing him a link to state’s Internet regulation policy. If not, it serves the website without further checking.

4.5 Advantages and Disadvantages

The aim of this system is to be very accurate and with low operational cost, and for this reason the FIRS combines the speed of IP blocking mechanism with the precision of URL filtering technique. Moreover, it proposes the use of four different lists in order to encourage Internet users’ participation in the Internet regulation project and at the same time it is able to handle with “discretion” specific kinds of content. Those are the main advantages of the proposed FIRS model.

On the other hand, FIRS main advantage (the three-stage design) is its greatest disadvantage too as it is vulnerable in technical terms and can be overpassed.

4.6 Circumventing FIRS and Possible Solutions

An Internet regulation system that is designed to be used in massive scale in a democratic society can’t be totally invulnerable, but extensive participation by Internet users can help it become gradually more effective.

Some of the circumvention techniques that have to be addressed are the use of web proxies (that can circumvent entirely the system) and the use of URL variations (that can circumvent the FIRS’s Stage 3) by the Internet users, and website mirroring, multiple URLs and the use of another port by the provider of the illegal content.

There are many ways to tackle these issues to some extent. For example, the use of web proxies can be tackled by adding web proxies URLs to the Blacklists A & B, or even better by programming FIRS to discard all source routed packets (i.e., packets that are trying to access a website via proxies) [13].

Moreover, the URL variations technique can be tackled thanks to the Internet users participation, or if the system is programmed to use algorithms to produce different kinds of variations for the already included URLs in Blacklists A & B.

All the above are subject to further examination.

5. Conclusions and Future Work

In this paper, the authors propose an Internet regulation system (called FIRS) that can operate fast and with low cost, and at the same time is “open” to participation and evaluation by the Internet users. Moreover, it states the importance of conducting well-designed surveys in each country prior to the FIRS design stage, in order for each system to be adapted to regional political environment and eventually be accepted by the general public. Based on all the above, a blueprint for this system is described.

Currently, the authors promote the use of surveys as a tool to the design procedure of such systems. For this reason they have already developed an online portal (see http://www.WebObserver.net) welcoming international participation, and thanks to it, related surveys are already running in some countries such as Germany and Russia.

Moreover, a beta version of the FIRS will be developed and be used in limited-scale experiments, in order to better define system’s advantages and disadvantages. Furthermore, FIRS design will be adapted in order to include new and promising techniques of effective Internet regulation (such the one described in Ref. [27]).

To be part of the international project WebObserver.net visits the website www.webobserver.net or send an email to the project’s coordinator at nkoumart@jour.auth.gr.

References

Internet Regulation, a New Approach: Outline of a System Formed to Be Controlled by the Internet Users


