Content Filtering Technologies and the Case of Greek School Network

Grigorios S. Chrysomalidis, *Secondary Education Teacher, ICT, grchrysomal@yahoo.com*

Anastasia S. Chrysomalidou, *Secondary Education Teacher, Civil Engineer, anchrysomalidou@yahoo.gr*

Ioannis A. Spiliotis, *Secondary Education Teacher, Electrical Engineer, spilstam@yahoo.com*

Abstract: Today, more than ever, the Internet is no longer just an academic, scientific tool but a means of mass media, socialization and education. It’s enormous spread makes it a dynamic platform for free expression and political activism, but also an ideal environment for disseminating information related to child pornography, violent behavior, and human rights abuses. The unprecedented proliferation of cultural goods posed by the digital revolution, creates the conditions for restricting and controlling the flow of information for a variety of reasons, from protecting vulnerable social groups to enforcing public order.

The presentation of modern techniques used for Internet filtering is also of educational interest. Controlling access based on content becomes more interesting, but also necessary in the case of the Greek School Network, due to the sensitivity that exists for its target audience.

Keywords: control, censorship, content, internet, students.

Introduction

The Internet initially emerged as an important communication tool for researchers in academia, but quickly evolved, especially after the advent of the World Wide Web, into an extremely valuable medium in business, political expression, education, entertainment, consumer habits, and of the social life of modern people (Gomez etal., 2009).

In the Western world, it is generally taken for granted, based on human rights, local constitutions and legal systems, and moral values, that Internet access is free, unrestricted, and most importantly, unaltered (Wolfgarten, 2011).

The popularity of the World Wide Web, the democratic nature of the Internet, since each user can theoretically freely and easily post their own content on it, as well as its decentralized nature, create, among other things, a fertile ground for abuse and on the other hand a target for all kinds of centers of power that seek to regulate anything that may challenge them. This gives rise to the need, or the choice, depending on the case, to control the access to the content that circulates on the networks around the world on a daily basis. The demand for protection of various social groups, such as minors, from inappropriate content related to violence, pornography, drugs, etc., but also the desire of some to restrict or manipulate free expression, are the two contradictory but at the same time complementary components that contributed to the creation of an entire field of communication technology related to content filtering.

Many techniques have been proposed and implemented to protect or suppress or simply control Internet users. The study of these technologies is always a challenge, due its continuous evolution caused by the extremely dynamic form of the Internet. In a study of content control technologies, it is important always to include the legal ethical and social parameters and implications of their application.

1. Definition of Content-Based Access Control

Content control on the Internet (otherwise filtering) is not a new activity; it is something that has been applied for years. However, the term covers a wide range of policies, infrastructures, software and services. Not all types of Internet blocking are equally effective, legally equivalent, and even more one control system cannot easily be used on all types of content (Tett, 2011).

The primary objective of the control is to ensure that blocked content is not retrieved from the end-user's computer, with the help of a software or hardware product that examines all online communications and determines whether or not to block the display of specific material. For example, an email may be blocked because it is suspected of being spam, a website may be blocked because it is suspected of containing malware, or a peer-to-peer network connection may be interrupted because it is suspected of sharing child pornography content (Callanan etal ., 2009).

This is essentially a type of Internet censorship. The term censorship describes the restriction of ideas, documents, letters, photographs, films, or any other type of information and human expression. The word comes from ancient Rome, where two judges, called censors, were responsible for safeguarding public morality (Gritzalis & Mitrou, 2008).

Internet filtering started about two decades ago by blocking spam, mainly among other reasons, to prevent network congestion. It is true that after so many years spam control has not been fully achieved (Callanan et al., 2009).

1.1 Content Categories to Exclude

The first criterion that can be observed to differentiate between content exclusion approaches is the goal of the control mechanism. There are generally four different points to focus on (Callanan etal., 2009):

- Service-based approach, e.g. Email.
- Content-based approach, e.g. hate speech, child pornography, gambling.

- User-based approach, e.g. users who illegally retrieve copyrighted material or send spam.
- Search engine based approach, e.g. block search results for illegal sites.

Content that is although, among other things, the subject of Internet filtering is (Callanan etal., 2009):

- Spam.
- Erotic and pornographic material. What is usually sought is to prevent minors from accessing content that is considered harmful.
- Child pornography. It is universally condemned and related offenses are internationally recognized as criminal offenses.
- Controversial political issues, issues of hatred, xenophobia. Some countries criminalize the publication of issues of racial hatred, violence and xenophobia, while in others with particular sensitivity to issues of protection of freedom of expression, such as the US, their unimpeded publication may be allowed.
- Illegal gambling.
- Defamation and publication of false information.
- Content published by terrorist organizations. The publication of propaganda and information related to the commission of crimes and terrorist acts.
- Infringements of copyright. It includes the exchange of songs, movies, software and more generally files that are protected by copyright, through file sharing systems.

2. Categories of Exclusion Technologies and Techniques

There are two basic options to blocking content on the Internet, central blocking at the network level and personal blocking at the end-user level, often with content calibration and self-definition. There is always the possibility of implementing hybrid systems that combine the above options.

Blocking at the end user level allows him to decide the type of content to be rejected, based on criteria that can be set and regulated separately for different categories of users (parents, children, teachers, students, etc.). This option may be the most specialized and "customer-centric", but it always carries the risk of the real impossibility of blocking inappropriate content, since the user, having administrative rights, finally decides what to access and what not (Callanan etal., 2009).

With network-based blocking, the service provider (Internet Service Provider, employer, organization, etc.) can determine the type of content or activity it will prohibit for all service users (Gritzalis & Mitrou, 2008).

There are three key issues, three critical questions, in the process of determining the content to be excluded (Callanan etal., 2009):

- How the content is technologically defined?
- Who selects this content?
- How will the exclusion be carried out in the end?

2.1 Basic Content Definition Techniques

The process by which content circulated on the web is collected, examined and evaluated is complex and often resource-intensive. There are three basic strategies (Bertino et al., 2008):

- Exclusion lists are the most common method. These are "blacklists", directories that contain anything that is considered inappropriate or "white" lists that contain content to which only access will be allowed. There are different types of lists with addresses, keywords, etc. and their content is often considered confidential.
- A second method of identifying the content to be blocked involves automatically inspecting the image, text, or video being streamed.
- The use of self-determination with the help of a predefined calibration and marking by third parties.

2.2 Who Defines the Content?

In countries where the judiciary is independent of the legislature and the executive, that is in modern liberal democracies, only the judge should have the power to classify a content, a situation or an action as illegal. This issue poses one of the major challenges for information traffic control systems on the Internet. Existing national and international legal procedures are rarely compatible with the cross-border challenges of the Internet or its operation speed. As a result, it is not always the judicial authorities that decide (McCrea et al., 1998).

Focusing on the role of those who make decisions about what is illegal or unacceptable content and therefore should be excluded, leads to their recognition and categorization in (Nicoletti, 2009):

- Individuals, such as parents who want to protect their children.
- Public and private organizations and institutions, such as schools and libraries, whose main purpose is to protect the target audience.
- Businesses and commercial companies, with the main aim of protecting their own interests and the performance of their financial activities.
- Governments and state agencies, when their status as centers of power or the rights of

those they represent are threatened.

- Judges and legislators like the basically competent ones in a benevolent state.

2.3 Application of the Exclusion

After determining the inappropriate material by those who chose to implement the content control, it is excluded. Although is almost impossible to have absolutely accurate filtering, there are a wide variety of methods used to control the flow of digital information. Such methods include redirecting users to proxy servers, redirecting packets to a specific IP address blacklisted, spying on, monitoring conversations, prohibiting pricing policies, and hacking into websites and malware dissemination, denial of service (DoS) attacks, interrupted interconnection, etc. (Nicoletti, 2009).

Exclusion options include less technical and more social methods. Modifying the law to be intimidating, strict policing, physical intervention, and the frequent shutdown of content-hosting servers are some of the control mechanisms that relate more to the desire for filtering than to the technology that will be applied. (Bertino et al., 2008).

Most forms of filtering are technically difficult to detect, so the user may not even be aware that they are being censored. Most providers do not have the ability to apply individual-level or unique IP address blocking (Warf, 2014).

Although content blocking mechanisms are often quite complex, they can often be bypassed with relatively little effort. There are several reasons for this; the most basic one is that the Internet is designed to be decentralized, with built-in data flow capability beyond any barriers. It is important to note that many times control attempts work with reduced efficiency, either by blocking content that should not, showing "excessive zeal" (over-blocking), or failing to block content inappropriate based on the set criteria (under-blocking). This is both the main problem and the challenge for access control technologies (Callanan et al., 2009).

3. Consultation, Motivations and Challenges

There are many incentives for Internet censorship, and it can take many forms, including political repression of human rights activists, religious controls that inhibit the spread of ideas considered heretical or sacrilegious, protection of religious property, restrictions that exist as part of the oppression of ethnic minorities (eg refusal to allow certain languages on government websites) or sexual minorities (homosexuals). Usually, governments seeking to impose censorship do not use so much the excuse of protecting public morality from dangers as pornography or gambling, but the fight against terrorism (Warf, 2014).

The discussion about Internet filtering cannot be limited to one specific topic, it is as complex as the technology used. There are many different issues that need to be addressed as the challenges faced by control policy makers are multidimensional. There are many reasons why society today believes (or in some cases hopes) that content exclusion attempts can alleviate some important social concerns that some other approaches do not seem to address successfully. Many different entities today apply exclusion, with a wide range of material being the target of this application. Internet blocking attempts can be approached in many different ways, depending on what their desired goal will be. Several countries have already adopted such systems (Deibert et al., 2008).

The Internet is a vast, complex network of networks, with a large number of hardware, protocols, and services. The first step in trying to control access based on content is to choose where to install it. A second key concern is to determine who chooses what will be excluded, grading this feature between different users and organizations. A wide range of content can cause different concerns in different societies and each exclusion measure is necessary to describe the variety of content it targets. Blocking Internet content applies to either producers or consumers of illegal content and has different levels of effectiveness, depending on this choice (Morozov, 2011).

Internet filtering is discussed as a technical solution in relation to a wide range of illegal activities. To a large extent, without this being necessary, these acts are criminalized in the country that intends to implement or has already implemented the control technology, but they are not always criminalized in the same way in the country where the content is hosted. Child pornography is one of those categories of content that falls under the provisions of criminal law. Enforcing the block is difficult as hardware is often legally available from servers abroad. This is a direct consequence of the different national standards that exist in relation to content publishing. It is like trying to maintain national cultural standards in the age of global internet integration (Callanan et al., 2009).

3.1 Reasons for Content Blocking

One of the main reasons for content control is the lack of controls on the Internet. Because the Internet is primarily designed to be based on a decentralized architecture, it is resilient to errors and malfunctions and is resistant to external control. Content blocking is something that was definitely not taken into account in its original design.

The international dimension of the Internet is another occasion for those who from time to time want to control the flow of information on it. International cooperation, based on the principles of traditional mutual legal assistance, is often very slow and time consuming. The formal requirements and time required for foreign law enforcement agencies to cooperate often prevent the required investigations from progressing effectively. Attempts to block content can sometimes be seen as an attempt to take more effective action, even in those cases where restrictions on current international cooperation prevent action from being taken in a timely manner.

Reducing the importance of content-hosting infrastructures within the sovereignty of a country is an important incentive to impose filtering mechanisms. Publishing content that is perfectly legal in one country may be a criminal offense in another. The attempt to exclude content can therefore be described as an act of redefining territorial sovereignty and critical political, social, economic and cultural living space, where governments seek to ensure that national standards are applied in relation to globalized content available to Internet users within the country (Morozov, 2011).

The motivation for choosing to control the information circulating on the Internet is directly related to who applies it. This can be social or moral, as in the case of parents who want to protect their children. It can be political when it comes to governments that seek to maintain the rule of law, as defined each time by their own criteria. In the case of companies, the incentive may be to safeguard the productivity of their employees, reduce the waste of resources such as bandwidth, ensure the protection of sensitive corporate data and avoid legal complications when there is illegal activity of employees (Nicoletti, 2009).

3.2 Content Exclusion Target

Blocking illegal content on the Internet can not only be considered as a means of dealing with offenders who own the content (producers), but also as a means of preventing the user from downloading illegal content (consumers) (Okeke , 2012).

From the point of view of the producer or provider of illegal content, the Internet has become an important tool for distributing inappropriate material, such as child pornography, as it offers a number of benefits to perpetrators that make it difficult to deal with. The facilitation of the distribution of illegal material, for example in the case of child pornography, is proportional to the facilitation of its production, offered by the modern digital camera and digital video camera. The reason for the application of filtering technology is therefore similar to the reasons for criminalizing the exchange of child pornography material, ie to reduce criminal activity and to protect children (Callanan etal., 2009).

One of the main groups targeted by content-based access control is minors. The protection of the sensitive childhood, but also of the educational process, connected with it, are the two main parameters that determine in this case, the application of the filtering in the content of the Internet. Protecting children from being exposed to inappropriate or harmful content, but also from being part of such content, in the case of child pornography, is a global constant, accepted by all societies, whether they are under liberal democratic governments or exist of more authoritarian policies. On the other hand, the Internet, with its obvious advantages in its application in the field of education, can contribute to the intellectual development and educational integration of young people by providing free information and communication (Morozov, 2011). In Greece, the largest network directly related to young people and one of the largest in general in the country, is the Greek School Network (www.sch.gr).

4. Greek School Network

The Greek School Network is the national network of the Greek Ministry of Education and Religious Affairs, which electronically connects all primary and secondary schools, including foreign units, services and supervisors, Ministry of Education at the central and regional level, the providers of lifelong learning services, students, education executives and other teachers and institutions of the Ministry of Education. It is the largest public network in the country in number of users and has been recognized internationally as a remarkable educational network that promotes the utilization of Information and Communication Technologies (ICT) in Greek education.

In order to preserve the educational character of the network, all its users are certified persons or educational or administrative entities of the Education. Specifically, they are Schools, administrative units of the Ministry of Education, Teachers and students.

Some of the data (until September 2020) that document the widespread use and utilization of the school network are (Greek School Network, www.sch.gr, 2020):

- Connected schools: 16.061
- Connected administrative services: 2.868
- Broadband Ratio: Basic Broadband Connection (144 Kbps -30 Mbps) 75%, Fast Broadband Connection (30-100 Mbps) 14%, High Speed Broadband Connection (>100 Mbps) 11%
- Teachers with personal account: 166.949
- Students with personal account: 1.022.864

4.1 Network Architecture of Greek School Network

Greek School Network has 51 privately owned nodes (Greek School Network, www.sch.gr, 2020), which it leases inside OTE buildings, in all prefecture capitals and in the largest Greek cities and 63 points of presence with co-location of its equipment within the Metropolitan Networks of Local Authorities (or within tertiary sweat or within MAN nodes). In conclusion, the Greek School Network has a physical presence in over 114 places throughout the country. The Computer Technology Institute technical team which is responsible for the proper operation of the backbone network manages more than 200 network devices daily. The average network traffic of Greek School Network to the Internet exceeds 9Gbytes daily.

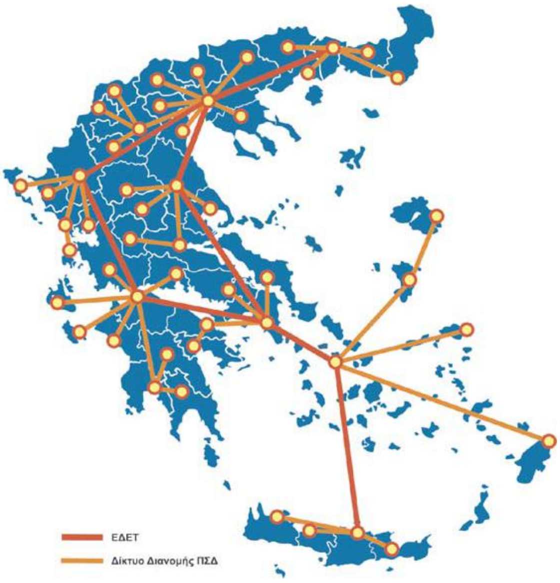

Figure 1. Greek School Network (Greek School Network, www.sch.gr, 2020)

4.2 Provided Services

A large number of online services are provided and supported today by the Greek School Network such as (Greek School Network, 2020):

- Broadband Internet access.
- Portal of the Greek School Network (www.sch.gr).
- Central user certification service
- Online Library for the Open Source Educational Software (http://opensoft.sch.gr).
- Web Hosting.
- Educational Communities and Blogs (http://blogs.sch.gr).
- Email.
- Electronic Classroom (eclass.sch.gr).
- Asynchronous Distance Learning (e-learning.sch.gr).
- Modern Distance Learning and Teleconferencing (http://conf.sch.grs).
- Video service.

and many others, among which stand out:

1. Promoting the safe use of the Internet (internet-safety.sch.gr) with instructions and suggestions to parents, teachers and students for the safe use of the Internet, because despite its undeniable usefulness it also conceals certain risks which each user gradually discovers. These risks are mainly related to children being exposed to illegal or inappropriate content,

being deceived by unknown adults who pretend to be minors, or being pressured to reveal personal information with the influence an adult may have on children.

2. Secure access to the World Wide Web, with which Greek School Network protects students from illegal and inappropriate content by providing content exclusion service. If students or teachers come across a page with inappropriate content that is not cut off by the Controlled Access service, they should contact the service administrator to request a ban on that page. If they find a page that belongs to the category of illegal content then they can contact the Greek Illegal Content Reference Line on the Internet (http://www.safeline.gr).

Internet access through the school should be done under the supervision of the teacher. The same should happen when children access the Internet from home, because the problem of Internet security is a more general problem and requires informing the whole society and especially the parents, who should be able to protect their children when these are at home (Greek School Network, 2020).

4.3 Content Control Service on the Greek School Network

Internet Access to the school laboratory, the Internet and all the services provided by the Greek School Network should be considered as a good, given to members of the educational community to promote knowledge and facilitate the educational process. Despite its undeniable usefulness, the Internet also hides some dangers. The Greek School Network in order to protect students from illegal and inappropriate content provides the Web Filtering Service on the World Wide Web. In this way access to pages is prohibited (Greek School Network, 2020):

- which propagate aggressive behavior, hatred and violence
- that promote drugs
- with gambling
- with pornographic content
- that promote racism

The access ban is made at a central point of the network and applies to all users of the Greek School Network as a whole. The set of inappropriate pages is maintained in a database and is regularly updated both by similar databases available for free on the internet, and by the Greek School Network users themselves.

4.4 Description of Current Service

The current form of the content service is implemented in the central operating node of Greek School Network in Athens. There is the border router of Greek School Network, its connection with the provider (GRNET) and all its trunk lines end (Kalogeras et al., 2012).

The border router automatically redirects the HTTP requests of Greek School Network users to an array of servers using Cisco IP policy technology. The free open source software Squid works on the servers as a transparent proxy server, intercepting users' requests in a transparent way. That is, the users of Greek School Network do not realize the existence of the proxy server, nor do they need to proceed with its configuration in their web browsers. The suitability of the content requested by the users is judged by the SquidGuard software, which maintains a database of inappropriate websites and web pages. In case the request is deemed inappropriate the user is redirected to an information page, otherwise Squid undertakes to serve the request, bringing the relevant item from the internet (Kalogeras, 2012).

An overview of the procedure followed is described in the following figure:

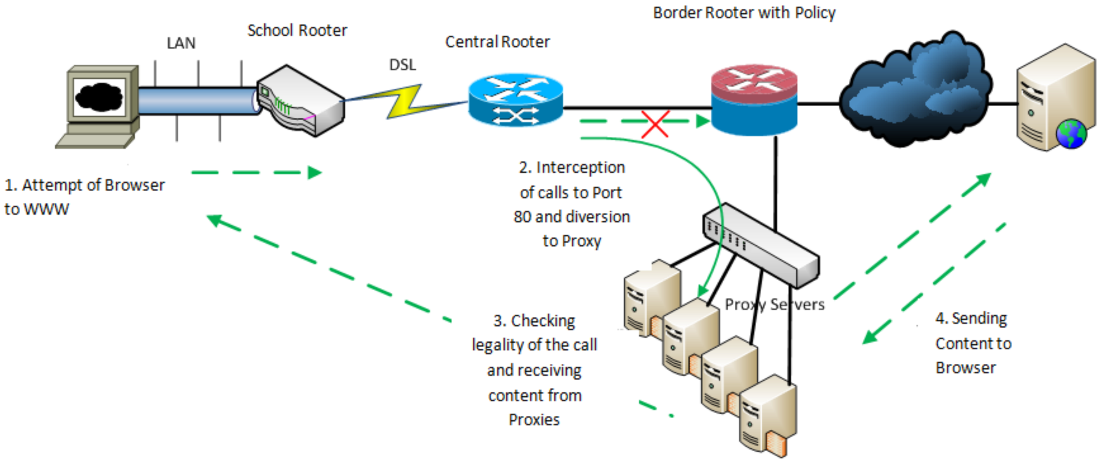
Border Rooter with Policy
LAN
School Rooter
Central Rooter
DSL
1. Attempt of Browser to WWW
2. Interception of calls to Port 80 and diversion to Proxy
Proxy Servers
3. Checking legality of the call and receiving content from Proxies
4. Sending Content to Browser

Figure2. HTTP traffic routing process (Kalogeras, 2012)

4.5 Squid

Squid is a web proxy and an automatic cache for a web cache (www.w3.org, 2017). It is Free-open source software with various uses, including the acceleration of a web server with cached function in frequent web requests, DNS and other requests when a group of users share own network resources. Squid can be used as a safety shield, filtering the traffic. Although Squid is mainly used for web transfer requests (with the http protocol) but also file transfers (with the ftp protocol), it has the ability to support (with restrictions) various other protocols such as TLS, SSL, Internet Gopher and https (www.w3.org, 2017).

Sixteen HP ProLiant DL380 G5 computer systems were supplied to meet the service needs. Squid acts as a proxy server to serve the HTTP requests of PSD users. The requests of the users are forwarded in the array by the central router as packet flows of TCP protocol with destination IP addresses outside Greek School Network and destination port 80 (Kalogeras et al., 2012).

SquidGuard, which acts as a Squid subprocess, maintains the database of inappropriate pages and takes on the role of arbitrator, allowing or not allowing access to the pages requested by users. The database is identical for all servers in the array. Each server maintains its own local copy that is automatically updated from a central point (Kalogeras et al., 2012).

Conclusions

Of all the myths associated with the Internet, the one that presents the greatest challenge is that it is an inherently liberating tool, an infrastructure that inevitably promotes democracy, giving voice to those without political power, and thus undermining authoritarian regimes and oppressive governments. Based on innovative development theories, improving the level of education and more access to information inevitably lead to a liberalization of the public sphere through a well-informed, politically and educationally self-determined audience. In Western societies in particular, it has been believed that the global community of people who browse the Internet on a daily basis should be self-governing, without the intervention of any state (Warf, 2014). The Internet has radically changed the way information is disseminated around the world, offering instant access to almost every type of digital resource. Unfortunately, this is equally true of both legal and illegal content (Gossett & Shorter, 2011).

The Internet has been built from the ground up as a decentralized network with the ability to redirect, overcoming natural disasters and control efforts. This is a major challenge for those seeking to police him, even in the case of trafficking in illegal material. The main proponents of Internet content filtering are usually governments, although some service providers have voluntarily implemented related programs. The vast majority of countries have anti-crime laws, such as child pornography, and defamation. Governments that have actively sought to prevent this type of illegal activity have encountered significant problems in trying to enforce Internet laws. The Internet has no jurisdiction, so it is often impossible for law enforcement agencies to intervene effectively against websites that violate it (Gossett & Shorter, 2011).

Content control techniques cover a wide range of technologies, as well as social and political choices. Their effectiveness varies, as do the ways to bypass them.

The Greek School Network is the largest public network in the country in number of users and is addressed to the most critical and sensitive audience, the students. Content-based access control to such a network is expected and highly desirable, in order to protect minors from the dangers of being exposed to illegal or inappropriate content and being deceived by unknown adults.

References

Bertino, E., Ferrari, E., Perego, A., Zarri G. (2008). Advanced techniques for web content filtering. Encyclopedia of Internet Technologies and Applications, Information Science reference. Hershey New York.

Callanan, C., Gercke, M., De Marco, E., Dries-Ziekenheiner, H. (2009). Internet Blocking Balancing Cybercrime Responses in Democratic Societies. Aconite Internet Solutions.

Deibert, R., Palfrey, J., Rohozinski, R., Zittrain, J. (2008). Access denied,the practice and policy of global Internet filtering. OpenNet, MIT Press.

Greek School Network, (2020). Available at www.sch.gr.

Gomez Hidalgo, H., Puertas Sanz, E., Carrero Garcia, F., DeBuenaga Rodriguez, M. (2009). Web Content Filtering. Advances in Computers, Vol. 76, (pp. 257-306).

Gossett, D. & Shorte,r J. (2011). Effectiveness of Internet Content Filtering. Journal of Information Technology Impact, Vol. 11, No. 2, (pp. 145-152).

Gritzalis, S & Mitrou, L. (2008). Content Filtering Technologies and the Law. Securing Information and Communication Systems Principles, Technologies, and Applications, Steven M. Furnell, Sokratis Katsikas, Javier Lopez, Ahmed Patel, Artech House, (pp. 243-265).

W3C Semantic Web Activity (2020). Available at http://www.w3.org/2001/sw/.

Kalogeras, D. (2012). Content control in educational networks: policies and guidelines implemented by the Greek School Network for the protection of students from illegal and offensive content on the Internet. Greek School Network Content Control Service, National Technical University of Athens.

Kalogeras, D., Chatzigiannakis, V., Christias, P. (2012). WWW Content Control Service Design Study. ΣΤΗΡΙΖΩ - Horizontal project of support schools, teachers and students on the way to the digital school, new services of the Greek School Network and Support of the digital school, ICCS.

McCrea P., Smart B., Andrews M. (1998). Blocking Content on the Internet: a Technical Perspective. CSIRO, Mathematical and Information Sciences.

Morozov, E. (2011). Whither Internet Control? The Johns Hopkins University Press, Journal of Democracy, Vol. 22, Number 2, pp. 62-74.

Nicoletti, P., (2009). Content filtering. Computer and Information Security Handbook, Elsevier, pp. 723-744.