Approaches to the Detection of Inappropriate Content in Images in the Personified System for Information Filtering

Roman Zharinov, Ulia Trifonova, Oleg Karmaleev, Alexandr Kodyakov
Saint-Petersburg State University of Aerospace Instrumentation
Saint-Petersburg, Russia
{roman, ulia}@vu.spb.ru, axelk21@yandex.ru, kod.aleksandr@gmail.com

Abstract

One of the most important development's trend of product content filtering in the part Internet traffic's control there is transition from using data bases site's categories to determine sites for it content. It stays especially actual with the development of different portals, which can contain full of different categories. Those categories can change on time or apply under client's settings.

Index Terms: DLP, Image analyze, Proxy-server.

I. INTRODUCTION

There are some approaches to data's categorizing. That methods we can separate condition to three categories:

- Using particular base of site's categories with regular updating sites and categories' list.
- Categorizing of data online with help of analyze pages' content.
- Using data about category and site's information.

II. MAIN PART

Using prepared sites' bases which separate on categories. That method is long using and is recommended itself. Nowadays those bases sell a lot of companies, such as Websense, Surfcontrol, ISS/Cobion, Secure Computing, Astaro AG, NetStar, etc. Some companies use these bases only for their own products. Other ones allow connecting their bases to third-party products. Mostly full of they are bases creating such companies like Websense, Secure Computing, SurfControl and ISS/Cobion. They contain the information about millions of sites which is very important on Internet.

Sites' sorting to categories usually produces at half-automatic mode. Firstly there is analyzed the content and determining the category with the help of specially developed software, which can include a variety of data analysis. Then the resulting information is checked by specialists for right category of that site. At the moment there two ways to use particular base of site's categories [1]:

- Using local categories' base with periodical updates. That method is convenient for large organization. Those organizations have IDS servers of filtration and which serve a lot of queries.
- Using that base which located on remote server. That method is often used in firewalls, ADLS, etc.

Using of remote categories' bases is a few increase loads on channels but it can ensure using the actual base of categories and not demand the space of storage. For advantages of using that bases started on the first query. It could defiantly decrease load on channels which transferee the data.

The main disadvantages of this method are that base will be not full and old. While the Internet is staying dynamically and development neither one of bases can't take all the Internet. As a content of sites is changed rapidly as the one stays at not that category during a few days. There are different services which the defining of category is impossible. Because the information is depend on their.

A. The ways to provide the Internet

A proxy server is an application which can sniff or forward traffic between client and some Internet resources. There five main types of proxy servers:

Forward proxies [2] – can receives requests from the internal network and transfers them to the Internet;

Transparent proxies [2] – communication scheme in which the traffic is forwarded to the proxy server implicitly (by a router);

Caching proxies – server can accelerates some requests by caching (save) copies of frequently requested resources, that can significantly reduce generated traffic and increase the performance;

Security proxies – server configured to use and execute security policies. Such proxies using for providing access to the content. Usually the client user authentication and authorization client to access certain content is controlled by the proxy server.

Reverse proxies – receives requests from the Internet and forwards them to the internal servers. Thus achieved ability of using the Internet without gain the information about the internal network or the content's location on the internal network.

For our analyze module we use several types of proxies: caching and transparent. It provides us save all incoming images and analyzes them as normal files. There are a lot of proxies which provide that functionality, the most popular presented at Table I.

For our purposes we chosen Squid proxy server for it flexible and wide functionality.

In addition to overview of services and algorithms for detection harmful image information [7] there are one more services (see Table II).

B. Detection algorithms

There are some general algorithms for detection images that can harm children health as well as physical or spiritual:

- Comparison of images in the search for characteristic features.
- Keyword filtering relating to images.
- Image analysis to determine the presence a large number of skins.

The first option was rejected due to the fact that the analysis of the images should go on the fly. This option is more suitable for the analysis of images offline. For the second – no need to analyze the image, we can filter only keywords, but now the more than 35% of Internet images and videos are not fully correspond to their description. Today the third way is the best. There are a number of algorithms, which provide a good level of filtration $(> 90\%)$ and a small number of false positives $(<5\%)$. If it started with a very slow algorithms such as the algorithm proposed by Flack [10] used only filter skin and grouping of human figures and spends processing a single image about six minutes. That WIPE

TABLE I

COMPARISON OF PROXY SERVERS

Function	AiS AliveProxy server [3]	Freeproxy [4]	Squid [5]	Nginx [6]
NAT	-	-	+	+
HTTP	+	+	+	+
DHCP	-	-	+	+
DNS	-	-	+	+
FTP	-	+	+/-*	-
POP3	-	+	+	+
SMTP	-	+	+	+
Cache	-	+	+	-
Black list	+	+	+	+
White list	-	-	+	-
Licenses	19.00$	GNU/GPL License	GNU/GPL License	BSD
OS	Windows	Windows	AIX BSDI FreeBSD Linux HP-UX IRIX Mac OS X Microsoft Windows NetBSD NeXTStep OSF и Digital Unix OpenBSD SCO Unix SunOS/Solaris	FreeBSD Linux Solaris AIX HP-UX Mac OS X Windows
Authorization methods	-	NTLM Basic Digest	DB getpwam LDAP MSNT NCSA NIS PAM POP3 RADIUS SASL SMB SSPI	POP3 IMAP SMTP LDAP htpasswd

TABLE II

IMAGE FILTERING SERVICES

Parameters	piFilter	WebPurify [8]	Twigla [9]
Photo	+	+	+
Video	-	-	+
Software	+	-	+
By hand	-	+	-
Collation	+	+	+
The analysis of skin color	-	-	+
Price	99$ for 60000 requests	2 cents for image	Individual price for each customer

system [11] contained modules such as Daubechies wavelets and vector comparison, and analyze images in 10 seconds. After the work Reigi Jones [12], which introduced the world to simple ways to determine skin and adult content, and work Bossoma [13] suggested the use of a multilayer perception (MLP). Thereby appeared algorithm analyzes the image for a split second, like, POESIA [14], [15] and InFeRno [16]. Now, let's see of our proposal of the algorithm.

We will use the combined method of analysis that is analyzed in parallel: one ICAP-server will use to analyze the text of the pages for the keywords there, while on the second ICAP-server - to analyze the image to its belonging to harmful images. Thus, combining two completely different approaches, we find a high degree of filtering of images that can harm children health. It process as follows. Decomposing image by color, and comparing the resulting histogram with the standard histograms, designed for a large volume of photos, we get a map of the skin. In other words, we obtain a three-dimensional histogram for the image with clear, where the detector found the skin. Then to get the map, we use the method of multilayer perception that can numerically evaluate the probability that a given image contains pornography. Parallel searching for keywords, we also get a numerical value. Then, compare your ratios to each other we form a final evaluation of the image of obscenity. This comparison is necessary because 25% of images and signatures are having no common. Finally, we compare result value with the threshold, sensitivity of the algorithm. If the value obtained in the analysis than the maximum, the ICAP-server not passes this image to the user. Instead, the user will see a white box the same size.

III. CONCLUSION

The detection of harmful images is important for the Internet, but there is still no effective method for preventing improper access of such information. In this article we compare some algorithms and software like proxy servers as main technical part of image detection. In future we'll try to implement offered detection algorithm and compare with other algorithms.

REFERENCES

[1] Modern trends in content filtering, Web: http://alexott.net/ru/writings/cf/index.html.
[2] Apache Module mod_proxy, Web: http://httpd.apache.org/docs/2.0/mod/mod_proxy.html.
[3] AiS AliveVPN - internet freedom since 1996, Web: http://www.atomintersoft.com/.
[4] Freeproxy, Web: http://www.handcraftedsoftware.org/.
[5] Squid: Optimising Web Delivery. url: http://www.squid-cache.org/.
[6] nginx professional services, Web: http://nginx.org/.
[7] Trifonova, U. V. и Zharinov, R. F., Concept of the System to Protect Children's Access to Information, *Proceedings of the 12th Conference of Open Innovations Association FRUCT*, 2012, pp. 142-146.
[8] WebPurify Image Moderation, Web: http://www.webpurify.com/image-moderation/.
[9] Twig.la is an automatic porn video detection software, Web: http://www.twig.la/.
[10] M.M. Fleck, et al, "Finding naked people", in: Proc. European Conf. on Computer Vision, B. Burton, R. Cipolla, Springer-Verlag, Berlin, Germany, 2:593-602, 1996.
[11] J.Z. Wang, et al, "System for Screening Objectionable Images", Computer Communications, (21)15^1355-1360, 1998.
[12] M.J. Jones, J.M. Regh, "Statistical color models with application to skin detection", Computer Vision and Pattern Recognition, 274-280, 1999.
[13] A. Bosson, et al, "Non-retrieval: blocking pornographic images", in: Proc. Intl. Conf. on the Challenge of Image and Video Retrieval, Lecture Notes in Computer Science, Springer-Verlag, London, 2383:60-70, 2002.
[14] H. Zheng, H. Liu, M. Daoudi, "Blocking Objectionable Images: Adult Images and Harmful Symbols", ICME, 2004.
[15] POESIA Project, "Poesia project", Web: http://www.poesia-filters.org/.
[16] S. Karavarsamis, N. Ntarmos, K. Blek, "InFeRno – an Intelligent Framework for Recognizing Pornographic Web Pages", ECML/PKDD (3):638-641, 2011.