Concept of the System to Protect Children's Access to Information in Schools Using the DLP-system and RFID-technology

Ulia Trifonova, Roman Zharinov
Saint-Petersburg State University of Aerospace Instrumentation
Saint-Petersburg, Russia
{julia, roman}@vu.spb.ru

Abstract

Nowadays, education is impossible without a computer, especially for children. That's why it is necessary to protect children from information that can harm their health as well as physical or spiritual growth (ineligible information), during their work with computer. There are types of information, which dissemination is prohibited or restricted to children under Russian law. This article describes concept of the system to protect children's access to information in schools using the DLP-system and RFID-technology. Article provides an overview of popular DLP-systems, services and algorithms for the detection harmful textual and image information. Data storage protocol of RFID-tags, which is used in Russian subway, is also considered.

Index Terms: DLP-system, RFID, Information security for children, Information filtering.

I. Introduction

Russian Federal law from 29.12.2010 № 436-FZ become operative in September 2012 [1]. This act is aimed to protect children from information, which is harmful to their health as well as physical or spiritual growth. It determines the types of information, which dissemination is prohibited and restricted to children. Restriction on dissemination certain types of information depending on the age group of children. Today, there is no system that can filter information depending on the age category of user. That is why the creation of this system is actually for the Russian market. It is proposed to create content filtering system based on DLP-system. Such system will help to organize control of children's access to information, which is especially important for schools, where children work self-dependent in the computer lab.

A. Types of information, which dissemination to children is prohibited in Russia

Russian Federal law of 29.12.2010 № 436-FZ determines the types of information, which dissemination to children is prohibited, they are:

1. Information that stimulate children to commit acts that constitute a menace to their lives and health, including suicide.
2. Information that causes a desire to drug, imbibe, smoke etc.
3. Information that causes a desire to gamble, prostitute, tramp, mendicancy.
4. Information that motivate to violence and cruelty to people and animals.
5. Information, which denies family values, forms disrespect for parents and other family members.
6. Information that justify illegal behavior.
7. Information that contain obscene language.
8. Information that contain pornography.

B. Types of information, which dissemination is restricted depending on the age group of children

The law also determines the types of information, which dissemination to children is restricted depending on the age group, they are:

- Information that describes the cruelty, physical and psychological violence, crime, anti-social behavior.
- Information that makes children fear, terror or panic, diminishes human dignity, including non-violent death, illness, suicide, accident.
- Information that depicts sexual relations between man and woman.
- Information that contains bad words and expressions that do not relate to the swearing.

C. Classification of information products

Information products separate to five groups in depend on the age of children. There are label of information product for each group. Label of information products - designation of information products in accordance with the classification:

"0+" – for children under six years;
"6+" – for children over six years;
"12+" – for children over twelve years;
"16+" – for children over sixteen years;
"18+" – prohibited for children.

For each age group we'll develop its security policy.

II. INFORMATION ANALYSIS

A. Text analysis

For text filtering and restriction of access to the text information, we suggest using the Data Leak Prevention (DLP)-system. DLP system was developed as a tool of protection against leaks of confidential information, which allows to detect the accidental or deliberate use of the user's data. DLP-system supports automatic or manual analysis of the events of transferred information. However, if system can detect the presence of confidential information in data, it can also be used to analyze data for existence of inappropriate content.

DLP systems can analyze the data when it is transmitted over the network, processed on personal computers and stored in the local network (shared resources, databases and other repositories).

Currently, most used DLP-systems on the Russian market of software for corporate customers are: InfoWatch Traffic Monitor, SecurIT Zgate, Symantec DLP, Websense DSS. Comparison of these DLP-systems is given in Table I.

Due to the fact that schools cannot purchase the specialized expensive software and based on comparative characteristics, shown in Table I, we have chosen MyDLP system. MyDLP – is a data loss prevention system, its main functions are: analysis of open and secure protocols (eg. http, https, ftp, ftps), Microsoft Office documents, archives (RAR, 7zip, zip), the calculation of hash function (such as md5) of files, ability to integrate with Web-proxy Squid.

For chosen DLP-system we should modify module of analysis of text information so that it can provide ability to read the text with various encodings of the Russian language

TABLE I
COMPARISON OF POPULAR AND OPEN SOURCE DLP-SYSTEMS

Parameter	InfoWatch	SecurIT	Symantec	Websense	MyDLP	OpenDLP
Name of the system	TrafficMonitor	Zgate	DataLossPrevention	Data Security Solutions	Data Loss Prevention	OpenDLP
Availability components	No	Yes	No	No	Yes	No
Installation location	Server, client	Server, client	Server, client	Server, client	Server, client	Server
Roles	A few	Any	Any	Any	A few	No
Can analyze security protocols	Yes	Yes	Yes	Yes	Yes	Only http, ftp
Text analysis	Dictionary, linguistic analysis, transliteration	Dictionary, linguistic analysis, transliteration	Dictionary, linguistic analysis	Dictionary, linguistic analysis	Exact match, Regular Expressions, dictionary	Exact match, Regular Expressions
Image analysis	Yes	Yes	Yes	Yes	No	No
The ability to create modules	No	No	No	No	Yes: Erlang, Python, PHP	Yes: Perl
Price	90 000$	Individual price for each customer	Individual price for each customer	Individual price for each customer	Open source	Open source

(utf-8, windows-1251, cp866, etc.), use of frequency of occurrence of "banned" words and of algorithms of morphological analysis of words.

The most efficient algorithm to find the "forbidden" or potentially dangerous words for Russian language – is morphological analysis. Morphological analysis - is search for keywords in the input data. For Russian language there are many words for which there is no single analysis. There are several ways where we can take part of the word:

1. Create own morphological dictionary. This method is difficult and does not guarantee 100% success.
2. Using Stemming Algorithm. Stemming performed by converting words based on certain rules:
   a. Find and delete the pre-defined endings. Search for words that end adjective, verb or noun.
   b. If a word ends in "и", delete it.
   c. Find and delete the diversion ends.
   d. Remove the double letter "н", and if the word ends in "ейш" or " ейше" remove these endings.

B. Image analysis

For the detection potentially harmful images there are 2 ways:

- We can “cut image” by link, size, or by some "mask" from template.
- Use of algorithms for finding the naked bodies.

For blocking ads or pornographic images from Web-sites, we will use regular expressions from third-party sites.

For image filtering we can use a variety of services and algorithms. Comparative analysis is given in Table II.

TABLE II
OVERVIEW OF SERVICES AND ALGORITHMS FOR THE DETECTION HARMFUL IMAGE INFORMATION

Parameter	piFilter	LogiPik	Finding Naked People	Nudity Detection
Integration	Processing uploaded images	PHP5 library	Algorithm [2]	Algorithm [3]
Accuracy	91.5%	90%	43%	94%
Price	134$ for 60000 queries	Individual price for each customer	Open source	Open source

While creating a filtration system for schools, the main factor we need to consider is the price of resulting system. That's why, the services presented in table 2 are not suitable for us. To implement our own image analysis module we will need to integrate several open algorithms, to improve the reliability of the system.

C. Filtering of audio-video content

Today, there are no open source algorithms that can search potentially harmful information in audio-video content. Therefore, on the first stage of developing system it is proposed to filter the audio and video content through the using of white lists.

III. AUTHORIZATION

To reduce the costs of the developed system it is proposed to implement authentication mechanism using RFID-technologies based on smart cards for public transport. In the Russian subway two versions of Mifare Classic RFID-cards are used: 1K or 4K. Mifare Classic card is divided into sectors; depending on the modification (1K or 4K) there can be 16 or 40 sectors, respectively. Access to each sector is protected by two 48-it keys A and B. On these cards stored some service information and information about card holder, including: name, surname, sex and date of birth. Information about card's holder is located in sectors 13-14, which are accessed by the standard «MIFARE Application Directory (MAD)» cipher key A: a0a1a2a3a4a5 hex. Remaining sectors of the card are private, and access to them is possible only if you know cipher keys.

To retrieve the information from RFID-tags one must install RFID-reader for each workplace. The RFID-readers, suitable for a specific task, cost up to $15 a piece. RFID-card must be in the reader's field of view until user finished his work.

If no RFID-card is detected near RFID-reader, then security policy for age group 0+ is activated.

To exclude the possibility of using vicarious RFID-cards, "black list" will be implemented. In this list the unique identifiers of the personal cards that have been lost or given to other persons will be recorded.

III. CONCLUSION

This article describes the concept of filtering the information, received from the network (Internet, Local Area Network) and local storage (hard disk, CD, flash memory, etc.) based on persons age category. Such system will ensure the safe work for children with different types of information. Such system may be used in schools, where access to the Internet and data on local media can get kids of all ages.

REFERENCES

[1] Federal law of 29.12.2010 № 436-FZ "*O zashchite detey ot informatsii, prichinyayushchey vred ikh zdorovyu i razvitiyu*" (in Russian).
[2] http://www.cs.hmc.edu/~fleck/naked.html
[3] http://onebit.us/x/i/814381733331796005.pdf

[4] https://code.google.com/p/ruadlist
[5] U.V. Trifonova, "Kids. Protection from ineligible content", *XIII International Forum Modern information society formation - problems, perspectives, innovation approaches: Proceedings of the International Forum*, Saint-Petersburg, SPb.: SUAI, 2012, pp. 186-190.
[6] Roman Zharinov , "Protection technology to intercept personal information in DLP-systems", *Proceedings of the Tomsk State University of Control Systems and Radioelectronics*, 1(25), part 2, 2012.
[7] Roman Zharinov, Vladimir Bashun, "Adaptation RSA DLP product for the Russian market", *Information Security of Russian Regions*, Saint-Petersburg, Russia, 2011 (in Russian).
[8] Ulia Trifonova, Roman Zharinov, "Legal aspects and perlustration methods of using DLP-systems", *65th international students scientific conference*, SUAI, 2012.