available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cose

Computers & Security

Recognition of electro-magnetic leakage information from computer radiation with SVM☆

Zhang Hongxin[a,b,*], *Huang Yuewang*[a], *Wang Jianxin*[a], *Lu Yinghua*[a], *Zhang Jinling*[a]

[a]*School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China*
[b]*State Key Laboratory of Information Security, Graduate School of Chinese Academy of Sciences, Beijing 100049, China*

ARTICLE INFO

Article history:
Received 14 April 2008
Received in revised form
23 July 2008
Accepted 30 September 2008

Keywords:
EM radiation
Information leakage
Computer security
Support vector machine
SVMs network

ABSTRACT

This paper focuses on the far-field reception of electromagnetic (EM) radiation and the recognition of letters recovered from the EM leakage. EM radiation captured by a wideband antenna is strengthened by a pre-manipulation system with amplifiers and filters, and the useful information is extracted. After being recovered from the EM radiation with signal processing method, the text image is further recognized by support vector machine (SVM) algorithm. Here, a two-layered SVM network with 60 SVMs is constructed to recognize the letters. In the process, the text image is cut apart into single letters to meet the input requirement of letter-based SVMs. To handle those exceptions of stroke connections, an interactive method and an automatic method are proposed. The received text image is usually obtained in low resolution with characteristics of large stroke distortion, font variation and variable size. In our two applications, however, the recognition accuracy reached 99.2% for the larger font size texts and 96.4% for the smaller size texts. From this, we may draw a conclusion that the proposed SVMs network works well in recognizing textual information and emphasize the potential risk of information leakage for computer system.

Crown Copyright © 2008 Published by Elsevier Ltd. All rights reserved.

1. Introduction

TEMPEST (Transmitted Electromagnetic Pulse/Energy Standards & Testing) has become one of the most intensively studied fields of EMC (electromagnetic Compatibility) and arouses worldwide concerns on information security. It is well known that it may cause EM radiation when computer system works, which could give rise to information leakage if it is intercepted and reconstructed. Experimental data and simulation results show that computer devices, such as data wire, network cable, hard disk, CPU and CRT, are all sources of EM radiation and could cause information leakage (Durak, 2008; Kuhm and Anderson, 2008; Ling et al., 1997; van Eck, 1985; Smulders, 1990; Lackey and Upmal, 1995; Anderson and Kuhn, 2008; Kuhn, 2008). When confidential texts are shown on computer screens, the corresponding electronic signals are sorts of stochastic digital sequences which are prone to emit EM radiation with a wide frequency band, and the radiation contains extractable and recognizable information (Hongxin et al., 2004a,b, 2007). While most of the researches are based on near-field reception (Shiwei et al., 2002; Rooney, 1998), far-field reception is still one of the most challenging areas.

☆ Project was supported by NSFC (grant No 60871081, 60671055, 60771060), Specialized Research Fund for the Doctoral Program of Higher Education (grant No 20070013002, 20070013004) and Open Fund of State Key Laboratory of Information Security, Graduate School of Chinese Academy of Sciences, PR China.

* *Corresponding author*. Beijing University of Posts and Telecommunications, NO. 10, Xitu Street, Haidian District, Beijing 100876, China. Tel.: +8601062282818.
E-mail address: hongxinzhang@263.net (Z. Hongxin).

0167-4048/$ – see front matter Crown Copyright © 2008 Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.cose.2008.09.009

However, far-field reception always means greater noise and lower resolution and consequently brings higher requirements for data manipulation and textual pattern recognition.

Even after delicate data manipulation, compared with the original text on the screens, the received text characters images still have much lower resolution and greater distortion due to strong noises. Variable size, font variation, dithering and stroke distortion of the text characters all add difficulty to text reconstruction, thus we need a powerful pattern recognition algorithm with strong generalization capability. And we propose SVM to implement this work. SVM, initiated by Vapnik in the 1990s, is a supervised learning method for classification and regression. Viewing the input data as two sets of vectors in an N-dimensional space, an SVM will construct two parallel hyperplanes, each of which would "push up against" one data set so as to maximize the "margin" between the two sets. In this way, the empirical classification error can be minimized. In our application, in order to reconstruct the text, we have to be able to recognize at least 26 lower-case letters, 26 capital letters and two common punctuations (",", & "."), so a multi-SVM algorithm is to be implemented. And we will use "one-versus-the-other" multi-SVMs (Cristianini and Shawe-Taylor, 2000, Cortes and Vapnik, 1995; Burges, 1998). SVM and Kernel Methods Matlab Toolbox (Canu et al., 2005) are employed as software tool in our application.

In this paper, a tree-structure network of multi-SVMs will be set up to conquer the challenging task of discerning all the 52 distorted letters. The remaining part of the paper is organized as follows. Section 2 provides a brief description of the reception of EM radiation and pre-manipulation of data. In Section 3, an overview and other details of our proposed SVM-based recognition method will be described. Section 4 shows the experimental results of our approach. Finally, Section 5 provides concluding remarks.

2. Radiation reception and data manipulation

2.1. Reception and filtering

As proposed by Zhang Hongxin et al. our reception and pre-manipulation system is illustrated in Fig. 1. Firstly, we use a wideband antenna to capture the leakage EM radiation, and then techniques as synchronization, phase locking, averaging filtering, and matched filtering are serially used to get clearer and more recognizable text images.

With a wideband antenna receiver, we successfully extract recognizable text images from the CRT radiation of a Pentium 486 computer from a distance of 12 m (far-field). Fig. 2 shows a received text image with large font size, while Fig. 3 is another with smaller and variant font size. After the afterward manipulation process in Fig. 1, the qualities of the images are enhanced. Figs. 4 and 5 show the effect of this process of image processing and the following operations are based on them.

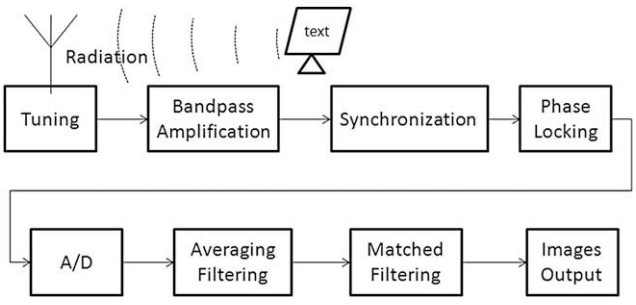

Fig. 1 – Process of radiation reception and image processing.

2.2. Cutting algorithm to attain separate letters

As we proposed a letter-based SVM recognition, in order to reconstruct the text information shown on the computer screen, the whole images (Figs. 4 and 5) has to be cut into single letters. And then separated letters are recognized one by one. Our cutting algorithm is illustrated as follows:

For the simplicity of depiction, as pictures are processed in computer, we denoted the picture as matrix P. $P(i,j)$ represents the value of the pixel in the i-th row and the j-th column. Our cutting algorithm is listed as follows and an example is illustrated in Fig. 6.

Step 1. Calculate $S(j) = \sum_i P(i,j)$, and $D(j) = S(j) - S(j-1)$, and set a threshold value D_{th}^{i} to find the vertical centre part for each text row, e.g. $M(r)$ for text row r.
Step 2. For each text row, expand upward and downward for half of the heights of their centre part, e.g. $M(r)$, to include the whole heights of the text rows.
Step 3. For each text row, calculate $Q(i) = \sum_j P(i,j)$, and set a threshold value Q_{th} to separate the letters horizontally.

2.3. Feature extraction

The separate letters are of different sizes and we have to normalize them before recognition. In our application, separate letters are normalized as $33 \times 16 = 568$ small images. We can directly take the 568 pixe values as 568 different pattern features. That is to say, input space of the SVMs is a 568-dimensional vector.

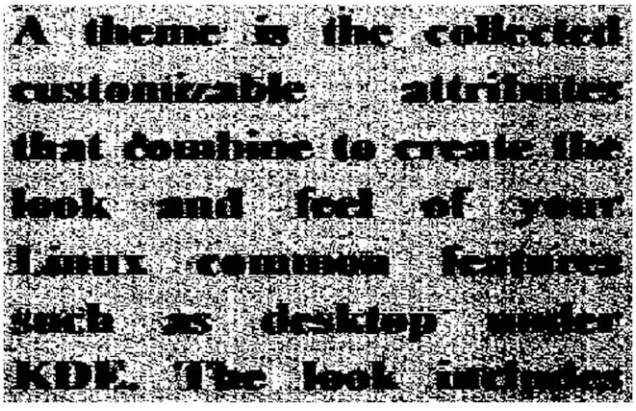

Fig. 2 – Texts before image processing (larger font size).

Download English Version:

https://daneshyari.com/en/article/454595

Download Persian Version:

https://daneshyari.com/article/454595

Daneshyari.com