 CMC, vol.57, no.1, pp.69-80, 2018

New Method for Computer Identification Through Electromagnetic Radiation

Jun Shi[1], Zhujun Zhang[2], Yangyang Li[1,*], Rui Wang[1], Hao Shi[1] and Xile Li[3]

Abstract: The electromagnetic waves emitted from devices can be a source of information leakage and can cause electromagnetic compatibility (EMC) problems. Electromagnetic radiation signals from computer displays can be a security risk if they are intercepted and reconstructed. In addition, the leaks may reveal the hardware information of the computer, which is more important for some attackers, protectors and security inspection workers. In this paper, we propose a statistical distribution based algorithm (SD algorithm) to extracted eigenvalues from electromagnetic radiate video signals, and then classified computers by using classifier based on Bayesian and SVM. We can identify computers automatically and accurately through electromagnetic radiation by using the algorithm in our experiment environment.

Keywords: Computer security, information security, compromising emanations, electromagnetic interference, signals sources identification, SVM.

1 Introduction

Computer displays emit electromagnetic waves and eavesdroppers can intercept these electromagnetic waves and reconstruct the information [Kuhn (2006); Sekiguchi and Seto (2013); Elibol, Sarac and Erer (2012). This can be a potential information security threat as the sensitive information can be stolen from a distance without any network connection. In addition, electromagnetic emanation also leaks the hardware information of the computer itself which is more important for some attackers. For example, attackers can find and lock the target computer if they can recognize the single computer by using the computer recognition algorithm. Besides, as for protectors, the recognition algorithm has significance for prevent information from leaking. Moreover, for the security inspection workers, they need not to check the specified computer in an anechoic chamber. They can check the computer in office environment and individually recognize the emanations to determine whether the compromising emanations belong to the specified computer or not.

[1] China Academy of Electronics and Information Technology, Badachu High-Tech Park, Shijingshan District, Beijing, 100041, China.

[2] Institute of Information Engineering, Chinese Academy of Sciences, 89 Minzhuang Road, Beijing, 100093, China.

[3] Solace Corporation, 535 Legget Drive, 3rd Floor, Ottawa, Ontario, K2K 3B8, Canada.

[*] Corresponding Author: Yangyang Li. Email: liyangyang@live.com.

In 2003, Markus Kuhn demonstrated that the electromagnetic radiation signals of different graphics are different [Kuhn (2003)]. Markus Kuhn analyzed the electromagnetic radiation signals of different LCD TV sets and he found that the signals vary much between devices. This conclusion based on the reconstruction of the display image [Kuhn (2013)]. A work covering some aspects regarding the electric and electronic equipment detection and recognition by their electromagnetic emission profile is presented in Mo et al. [Mo, Lu and Zhang (2012)]. Their approach was to compare original video signal spectrum, measured on RED channel with intercepted emissions from computer. However, the RED channel of tested computer cannot be connected with attackers' devices in the practical non-cooperative attack scenario. Besides, they did not give specific measure features and recognition results. Another computer recognition-related article is Mo et al. [Mo, Lu, Zhang et al. (2013)], which proposed a method to identify the computer display electromagnetic emissions based on support vector machine (SVM). However, they did not analyze the reason that electromagnetic emissions from computer vary between devices and their method needs a large number of training data.

In this paper, we propose a statistical distribution based algorithm (SD algorithm) to extracted eigenvalues from electromagnetic radiate video signals, and then classified computers by using classifier based on Bayesian and SVM. We can identify computers automatically and accurately through electromagnetic radiation by using the algorithm in our experiment environment.

2 Modeling of electromagnetic radiate video signal

Electromagnetic radiate video signal in time domain can be represented as [Elibol, Sarac and Erer (2012)]:

$$S_m(t) = S_p(t) \otimes S_h(t) \otimes S_v(t) \tag{1}$$

where $\otimes$ denotes convolution. According to the principle of Fourier transform, the combined spectrum of the electromagnetic radiate video signal can be represented as:

$$S_m(f) = S_p(f) \cdot S_h(f) \cdot S_v(f) \tag{2}$$

where, $S_p(f)$ is pixel spectral component. $S_h(f)$ and $S_v(f)$ are spectra of the horizontal and vertical synchronization signals. While $S_p(f)$ generates spectrum lines equal to pixel clock frequency f_p and its higher harmonics, $S_h(f)$ adds components (sidebands) around f_p equal to horizontal synchronization frequency f_h and its higher harmonics. $S_v(f)$ adds further spectral lines around f_h equal to vertical synchronization frequency f_v and its higher harmonics. Therefore, the electromagnetic radiate video signal spectrum presents equal spacing distribution and the spacing are equal to horizontal synchronization frequency and vertical synchronization frequency.

The range of vertical synchronization frequency is from 40 Hz to 86 Hz, while the range of horizontal synchronization frequency is from 30 kHz to 115 kHz. In addition, the range of pixel frequency is from 31.5 MHz to 297 MHz [Elibol, Sarac and Erer (2012)]. Pixel signals change with the display image so that it cannot reflect the internal features

of computers. Thus, to prove that differences do exist among different computers, the horizontal synchronization signals are the most suitable. Thus, this paper is modeled on the waveform of horizontal synchronization signal.

A horizontal synchronization signal is periodic and there is a blank in each line. Considering the periodic property of horizontal synchronization signal, we modeled the horizontal synchronization signal as Fig. 1.

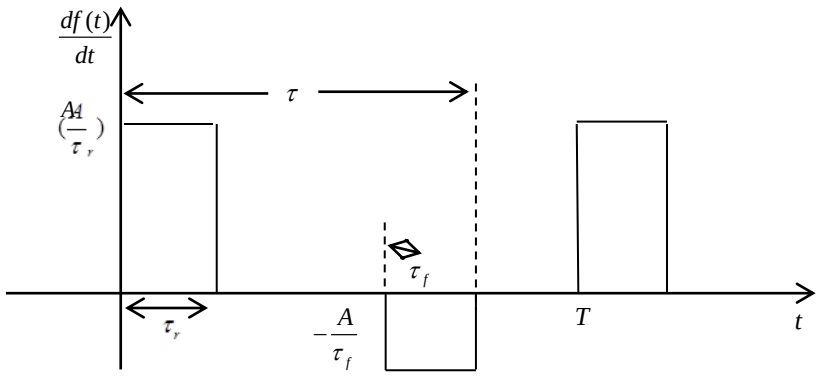


Figure 1: Model of horizontal synchronization signal

In Fig. 1, T is the period of horizontal synchronization signal. A is the amplitude of signal. τ is the scan time of each line. τ_r and τ_f are the pulse width of each signal.

$$T = \tau + b \tag{3}$$

where, b is blank time of each line.

According to the principle of Fourier transform, the frequency spectrum of this signal can be represented as:

$$F(n\omega_1)=\frac{A}{j2\pi n}e^{-jn\omega_1(\tau+\tau_r)/2}\left[\frac{\sin(\frac{n\omega_1\tau_r}{2})}{\frac{n\omega_1\tau_r}{2}}e^{jn\omega_1\tau/2}-\frac{\sin(\frac{n\omega_1\tau_f}{2})}{\frac{n\omega_1\tau_f}{2}}e^{-jn\omega_1\tau/2}\right] \tag{4}$$

If $\tau_r = \tau_f$, the single-sided (positive frequency only) spectrum is :

$$\left|F^+(n\omega_1)\right| = 2\left|F(n\omega_1)\right| = 2\frac{A\tau}{T_1}\left|\frac{\sin(n\pi\tau/T_1)}{n\pi\tau/T_1}\right|\left|\frac{\sin(n\pi\tau_r/T_1)}{n\pi\tau_r/T_1}\right| \qquad (n\neq 0) \tag{5}$$

$$F(n\omega_1) = \frac{A\tau}{T_1} \qquad (n=0) \tag{6}$$

It can be seen that, if $\tau = T/2$, the formula (6) simplifies to formula (8).

$$\left|F^+(n\omega_1)\right| = 2\left|F(n\omega_1)\right| = 2\frac{A\tau}{T_1}\left|\frac{\sin(n\pi/2)}{n\pi/2}\right|\left|\frac{\sin(n\pi\tau_r/T_1)}{n\pi\tau_r/T_1}\right| \tag{7}$$

When n is even, formula (7) is equal to 0. It means that when $\tau = T/2$, there are no even harmonics. In addition, it is easy to prove that the nearer τ approximates to $T/2$, even harmonics is smaller than odd harmonics. Thus, ratio between the scan time of each line and the period of horizontal synchronization signal influences the variation trend of harmonics. The ratio can be represented as $\tau/T = \tau/\tau + b$.

As a matter of fact, being unintentional, both scan time of each line and period of horizontal synchronization signal vary much between devices due to different production processes. Thus, it can be said that the variation trend (or shapes) of harmonics of electromagnetic radiate video signal spectrum vary between different computers. This paper proposes a new algorithm to describe the variation trend of harmonics of electromagnetic radiate video signal spectrum.

3 Algorithm

3.1 Basics of wavelet transform

Discrete Wavelet transform (DWT) is the discretization of the Continuous Wavelet Transform (CWT) through sampling particular wavelet coefficients. Sampling of CWT is achieved by letting $a = 2^{-1}$ and $b = m2^{-1}$, in $W(a,b)$. l is the discrete translation and m is the discrete dilations. DWT of a signal $f(t)$ is given by

$$W(l,m) = \int_{-\infty}^{\infty} 2^{l/2} \varphi(2^l t - m) f(t) dt \tag{8}$$

DWT [Soon, Koh, Yeo et al. (1997)] has its own advantages such as the ease of implementation and less computation time when compared to time domain. Here the signal is decomposed into approximated and detailed coefficients, where approximated coefficients consist of low frequency information and the detailed coefficients represent high frequency information. Approximated coefficients are obtained by passing the signal through a low pass filter and a dyadic down sampler. Detailed coefficients are obtained by passing the signal through a high pass filter and a dyadic down sampler.

3.2 Statistical distribution based algorithm (SD algorithm)

In this sub-section, a statistical distribution based algorithm (SD algorithm) is proposed. In addition, we use wavelet transform here because the wavelet coefficients describe the variation trend of harmonics of electromagnetic radiate video signal spectrum. We analyze the statistical distribution of wavelet coefficients by calculating the histogram of wavelet coefficients and fitting many different curves. The fit results of different distributions are given in Fig. 2. It can be observed that, the exponential distribution fits the histogram best.

Thus, the first step of the algorithm is calculating signal power spectrum.

Secondly, we calculate the wavelet coefficients of signal power spectrum. We choose two-tap Haar wavelet transform to implement our algorithm due to its simplicity.

$$X \overset{DWT}{\rightarrow} \left\{ a_L, d_j \right\} \tag{9}$$

where DWT accords to the Eq. (5). X is the signal power spectrum. $\{a_L, d_j\}$ are the wavelet coefficients.

Thirdly, we make Maximum Likelihood Estimation (MLE) of exponential distribution. MLE of exponential distribution parameter is given in formula (10) and we need to calculate μ of wavelet coefficient a_L.

$$y = \frac{1}{\mu} e^{-\frac{a_L}{\mu}} \tag{10}$$

Then, in order to realize automatic recognition, the classifier of the Bayesian and the classifier of the SVM are used.

Training data generated based on 400 signals from four different computers， which are Think Center, DELL OPTIPLEX GX520 and DELL OPTIPLEX 7020. This choice considered sampling the computers of different brands and computers of the same brand. Moreover, to analyze the individual characteristics of computers, we used two computers of the same model DELL OPTIPLEX 7020. To distinguish between the two same model computers, hereafter called DELL OPTIPLEX 7020-1 and DELL OPTIPLEX 7020-2. Each computer generated 100 signals.

As for the classifier of the Bayesian, conditional probability density functions (PDFs) of μ are obtained from training data and shown in Fig. 3. The algorithm judges to which computer the observed signal belongs. An observed signal belongs to one computer only if its conditional probability density function (PDF) f is the largest among other computers.

$$f(\mu = \mu_o | C) = \max(f(\mu = \mu_o | C_i)), \quad i = 1,2,3,4. \tag{11}$$

where, μ_O represents μ of the observed signal. C_i represents the computer source. The conditional PDFs of μ, given class label C_i can be obtained from training data as given in Fig. 3.

In conclusion, the algorithm can be divided into the following steps:

(1) Calculate the signal power spectrum.

(2) Calculate the wavelet coefficients of signal power spectrum.

(3) Calculate μ of wavelet coefficients.

(4) Look up the joint conditional PDF $f(\mu|C_i)$ obtained from training data. Calculate $f(\mu|C_1)$, $f(\mu|C_2)$ and $f(\mu|C_3)$.

(5) Compare the $f(\mu|C_1)$, $f(\mu|C_2)$ and $f(\mu|C_3)$, and find the maximum.

(6) The maximum of $f(X = X_o|C_i)$ corresponds to the right computer source.

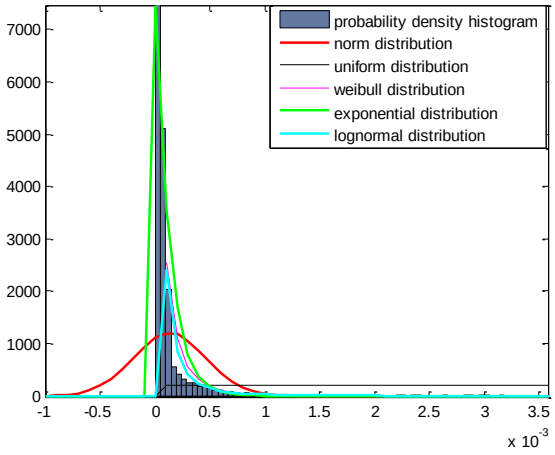


Figure 1: Fit results of statistical property

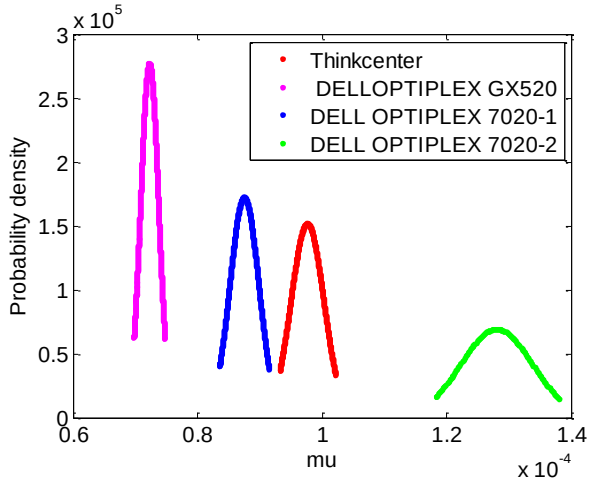


Figure 2: Conditional probability density functions (PDFs) of μ

As for the classifier of the SVM, the advantage of selecting SVM for classification is that it can map multi-dimensional feature input to high-dimensional kernel space, which is more conducive to classification.

SVM tries to find a hyperplane based on following optimization criterion [Hastie, Tibshirani and Friedman (2001)].

$$\min_{\omega,b} \|\omega\| \text{ subject to } y_k(\omega \cdot x_k + b) \geq 1, \forall k \tag{12}$$

where the margin is given by $2/\|\omega\|$. Thus, minimizing $\|\omega\|$ is equivalent to maximizing the margin. Solving this quadratic problem gives the hyperplane parameter as follows:

$$\omega = \sum_{\forall x_k \in S} \alpha_k y_k x_k \tag{13}$$

where S is a set of support vectors for both classes, and α_k is a trained weight on the corresponding support vectors. Based on this solution, one can classify an arbitrary new input x using

$$f(x) = sign(\omega \cdot x + b) = sign(\sum_{\forall x_k \in S} \alpha_k y_k x_k \cdot x + b) \tag{14}$$

The entire platform can be generalized to a nonlinear case. This generalization can be accomplished by mapping the samples to a certain high-dimensional space H:

$$\begin{aligned} &\Phi : \mathbb{R}^D \mapsto \mathrm{H} \\ &x \mapsto \Phi(x) \end{aligned} \tag{15}$$

Under such a high-dimensional space, usually called the feature space, the original overlapping data could become linearly separable. Constructing a separating hyperplane in that space yields a nonlinear decision boundary in the input space [Kim, Park, Toh et al. (2010)]. However, since the dimensionality of this new feature space could be very high (possibly infinite), a direct data mapping often becomes intractable. Nevertheless, by adopting a kernel function $k(x_i x_j)$, the nonlinear SVM can be formulated in a tractable manner without explicitly carrying out the mapping into the feature space:

$$\begin{aligned} &f(x) = sign(\Omega \cdot \Phi(x) + b) \\ &= sign(\sum_{\forall x_k \in S} \alpha_k y_k \Phi(x_k) \cdot \Phi(x) + b) = sign(\sum_{\forall x_k \in S} \alpha_k y_k k(x_k \cdot x) + b) \end{aligned} \tag{16}$$

We still use μ of wavelet coefficients calculated by formula (10) to construct a set of feature vectors and they are input into a classifier of the SVM for computer identification.

$$\mu=[\mu_1, \mu_2, \mu_3] \tag{17}$$

4 Experimental results

In this section, the proposed algorithm is applied to experimental data and the results and analysis are given.

Four computers used here are Think Center, DELL OPTIPLEX GX520, DELL OPTIPLEX 7020-1 and DELL OPTIPLEX 7020-2. The measurement setup is shown in Fig. 4. The resolution of the computer display was set at 1024×768. A log-periodic antenna (ZN30505E) designed for 30-3000 MHz was placed in front of the tested computer and its height was the same as the height of the computer display center. It is important to note that we placed the antenna 1 m-10 m from the tested computer to obtain signals. In addition, the performance of the algorithm under different antenna distance is presented in Section 5.

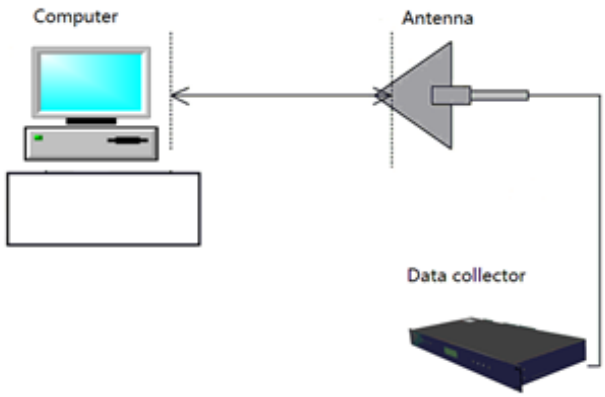
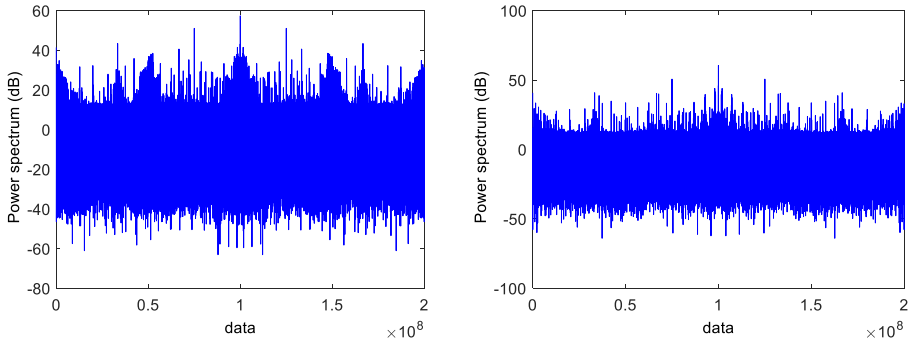


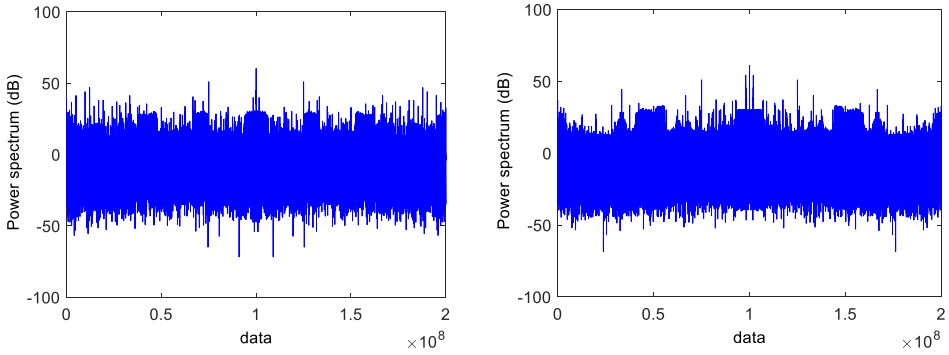
Figure 3: Measurement setup for data collection

The antenna is connected to a data collector, which can be a data acquisition card, digital oscilloscope and spectrum analyzer. A spectrum analyzer was used here. Theoretically, the noise received here is white noise which can be attributed to external noise sources as well as data collector internal noise, such as the noise figure of some filters, mixers, and semiconductors [Song and Yook (2015)]. Additionally, the antenna receives environmental white noise with many other man-made noises. As for sample frequency, according to the VESA standard, the scope of the pixel frequency is from 31.5 MHz to 297 MHz. When the resolution of the computer is 1024×768, the scope of pixel frequency is from 44.9 MHz to 94.5 MHz. Considering these video interface signals include harmonics of the fundamental signal frequency, we chose 500 MHz as the sample frequency.

To evaluate the SD algorithm, the four computers display the same images which were filled with letter "H". Fig. 5 shows the power spectrums of sub-band of four computers emanations. It can be seen that the variation trends of spectrum harmonics are different among the computers.



(a) (b)



(c) (d)

Figure 4: Power spectrums of sub-band of four computers emanations (a) Think Center (b) DELL OPTIPLEX GX520 (c) DELL OPTIPLEX 7020-1 (d) DELL OPTIPLEX 7020-2

We tested 400 sets of received signals and each computer contains 100 sets of data. It should be noticed that the test data here is different from the training data used in Section 3.

Recognition result of SD algorithm by using the classifier of the Bayesian is shown in Tab. 1.

Definitions of POD and FAR are:

$$\mathrm{POD} = TruePositives \big/ (TruePositives + FalseNegatives) \tag{18}$$

$$\mathrm{FAR} = FalsePositives \big/ (FalsePositives + TrueNegatives) \tag{19}$$

where, "Positive" labels the location that the detector judges as the true computer, and "Negative" labels the location that the detector judges as the wrong computer. In Tab. 1, True Positive (TP), False Negative (FN), False Positive (FP), True Negative (TN) and FAR of the data are summarized.

Recognition result of SD algorithm by using the classifier of the SVM is shown in Fig. 5. It can be seen that the SD algorithm has a higher POD when using the classifier of the SVM then using the classifier of the Bayesian.

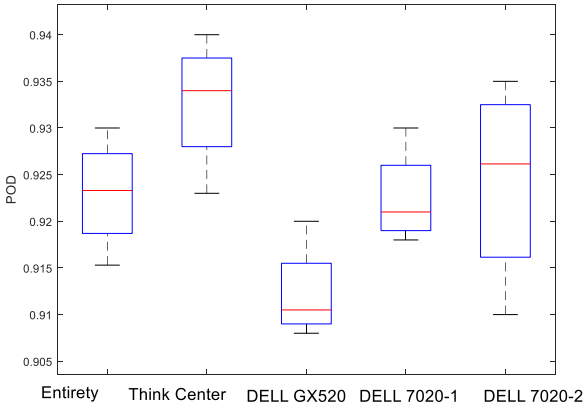


Figure 5: Recognition result of SD algorithm by using the classifier of the SVM

Table 1: Recognition result of SD algorithm by using the classifier of the Bayesian

	Think Center	DELL OPTIPLEX GX520	DELL OPTIPLEX 7020-1	DELL OPTIPLEX 7020-2
TP	93	91	90	91
FN	8	10	11	11
FP	7	9	10	9
TN	292	290	289	289
POD	92.1%	90.1%	89.1%	89.22%
FAR	2.3%	3.01%	3.34%	3.02%

5 Algorithm comparison

Considering that reference Sun et al. [Sun, Shi, Wei et al. (2016)] proposed a SP algorithm based on spectral centroid to identify the computer electromagnetic emissions, we compared SP algorithm with our proposed SD algorithm in this section.

We compared the performance of the SP algorithm with SD algorithm by using the classifier of the SVM and both algorithms use the same number of training data (400 sets of received emanation signals). The process of experimentation and the test data are the same as Section 4 which results in Tab. 1. The measurement setup is shown in Fig. 4. Four computers tested were Think Center, DELL OPTIPLEX GX520, DELL OPTIPLEX 7020-1 and DELL OPTIPLEX 7020-2.

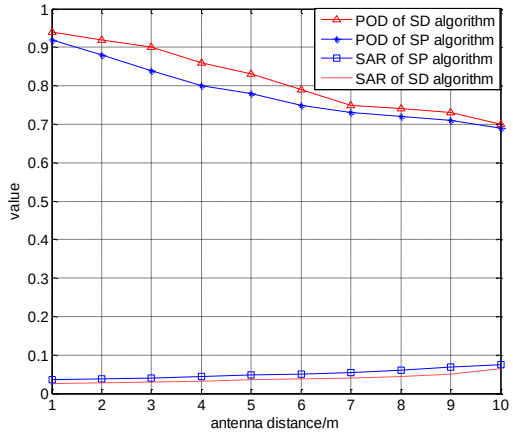


Figure 6: POD and SAR of the SD and SP algorithm as a function of antenna distance

The performance of the SP and SD algorithm under different antenna distance is presented in Fig. 6. The antenna distance is from 1 m to 10 m. It should be noticed that the POD and SAR in Fig. 6 are the average value of PODs and SARs of the four computers. It can be observed that, the accuracy of the SP and SD algorithms decrease with the increasing of antenna distance. The performance of SD algorithm is better than SP algorithm.

6 Conclusion

This paper proposed a new algorithm to realize computers recognition through electromagnetic radiate video signals. We proposed statistical distribution based algorithm (SD algorithm) to identify computers. By using the algorithm, we can automatically and accurately identify computers through electromagnetic radiate video signals in our experiment environment. In addition, the performance analysis of the SD algorithm comparing with the method proposed in Sun et al. [Sun, Shi, Wei et al. (2016)] under different antenna distance indicates that the SD algorithm has a better robustness.

The proposed method of identifying displays has practical significance. First of all, this method has significance for reconstruction of the compromising emanations. Attackers can lock onto the target computer so that they can just reconstruct the image of the objective display after the identification, especially when the attackers intercept the information in big organizations where lots of different computers are used. Secondly, in the same scene, protectors can selectively protect computers which have high compromising emanations rather than protect all computers aimlessly.

To prevent computer identification, special EMC (Electro Magnetic Compatibility) measures can be taken, such as shielding the computer and shielding cables. Then they can substantially reduce the compromising emanations and it would decrease the signal to noise ratio of received signals. The next step of our work will be combining our method with other signal processing methods to acquire a more accurate result in the low signal to noise ratio circumstances.

Acknowledgement: This work was supported by the Innovation Foundation of China Academy of Electronics and Information Technology (Grant No. 17109701). This work was also supported by the Innovation Fund of CETC (Grant No. 16105501) and the Joint Fund of CETC (Grant No. 20166141B08020101).

References

Elibol, F.; Sarac, U.; Erer, I. (2012): Realistic eavesdropping attacks on computer displays with low-cost and mobile receiver system. *20th European Signal Processing Conference*, pp. 1767-1771.

Hastie, T.; Tibshirani, R.; Friedman, J. (2001): The elements of statistical learning: Data mining, inference and prediction. *Mathematical Intelligencer*, vol. 27, no. 2, pp. 83-85.

Kim, S. K.; Park, Y. J.; Toh, K. A.; Lee, S. (2010): SVM-based feature extraction for face recognition. *Pattern Recognition*, vol. 43, no. 8, pp. 2871-2881.

Kuhn, M. (2003): *Compromising Emanations: Eavesdropping Risks of Computer Displays (Technical Report).* University of Cambridge Computer Laboratory, pp. 1-167.

Kuhn, M. (2006)**:** Eavesdropping attacks on computer displays. *Information Security Summit*, pp. 24-25.

Kuhn, M. (2013): Compromising emanations of LCD TV sets. *Electromagnetic Compatibility*, pp. 564-570.

Mo, F.; Lu, Y. H.; Zhang, J. L. (2012): Detection and identification of EM field source by using support vector machines. *International Conference on Wireless Communications, Networking & Mobile Computing*, pp. 1-4.

Mo, F.; Lu, Y. H.; Zhang, J. L.; Cui, Q.; Qiu, S. H. (2013): A support vector machine for identification of monitors based on their unintended electromagnetic emanation. *Progress in Electromagnetics Research M*, vol. 30, pp. 211-224.

Sekiguchi, H.; Seto, S. (2013): Study on maximum receivable distance for radiated emission of information technology equipment causing information leakage. *IEEE Transactions on Electromagnetic Compatibility*, pp. 547-554.

Song, T. L.; Yook, J. G. (2015): Modeling of leaked digital video signal and information recovery rate as a function of SNR. *IEEE Transactions on Electromagnetic Compatibility*, pp. 164-172.

Soon, I. Y.; Koh, S. N.; Yeo, C. K. (1997): Wavelet for speech de-noising. *TENCON'97. IEEE Region 10 Annual Conference. Speech and Image Technologies for Computing and Telecommunications*, pp. 479-482.

Sun, D. G.; Shi, J.; Wei, D.; Zhang, M. (2016): A New method to recognize computer through electromagnetic radiation based on spectral centroid. *Asia Pacific International Symposium on Electromagnetic Compatibility*, pp. 184-186.