Measurement 58 (2014) 468–475

Contents lists available at ScienceDirect

Measurement

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

RF fingerprint measurements for the identification of devices in wireless communication networks based on feature reduction and subspace transformation

J.L. Padilla [b], P. Padilla [a,*], J.F. Valenzuela-Valdés [c], J. Ramírez [a], J.M. Górriz [a]

[a] Department of Signal Theory, Telematics and Communications – CITIC, University of Granada, 18071 Granada, Spain
[b] Department of Electronics and Computer Technology – CITIC, University of Granada, 18071 Granada, Spain
[c] Department of Computer and Telematic Systems Engineering, University of Extremadura, 06800 Merida, Spain

ARTICLE INFO

Article history:
Received 9 October 2013
Received in revised form 23 April 2014
Accepted 5 September 2014
Available online 16 September 2014

Keywords:
RF Fingerprint
Subspace transformation
Wireless communications
Network identification

ABSTRACT

This document proposes a radiofrequency (RF) fingerprinting strategy for the proper identification of wireless devices in mobile and wireless networks. The proposed identification methods are based on the extraction of the preamble RF fingerprint of a device and its comparison with a set of already known device RF fingerprints. The identification method combines techniques for feature reduction such as Principal Component Analysis (PCA) and Partial Least Squares regression (PLS), both based on subspace transformation, along with a similarity-based analysis. In this work, a complete procedure for RF fingerprint data extraction and analysis is provided. In addition, some experimentation with commercial Wi-Fi devices is carried out for the methods validation.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

The technological advances in last years have enhanced the development of a variety of systems and services, based on wireless communications. Compared to wired networks and systems, the propagation channel in wireless networks is prone to suffer from interferences or security threats such as signal interception, spoofing or jamming, among others [1,2]. As a consequence, the secure identification of the forming devices of a wireless network is an issue of great concern regarding the network security. One of the most common and harmful security threats is related to jamming attack strategies [1]. Despite the considered jamming modality, affecting either the link layer or the physical one, the key fact to avoid such attacks is to properly identify the jammer. This jamming agent, acting either as a network user or as an external one must not only be identified but also blacklisted. Thus, subsequent attacks coming from this blacklisted jammer can be quickly identified and its jamming effects controlled, neglected or mitigated [2,3].

This paper provides a method for the proper identification of network devices in a network to detect jamming agents. The method is based on the analysis of the radiofrequency (RF) fingerprint of the devices of a wireless network. The premise in RF fingerprinting is that the signals transmitted by a wireless device are repeatedly extractable and unique. As a consequence, they may be used to identify the device when transmitting in the network. Recent works demonstrate that this uniqueness exists, being attributable to various factors: manufacturing, aging, environmental, etc. [4–6].

The manuscript is organized as follows: in Section 2, the basics of RF fingerprinting are provided. Section 3 is devoted to the identification methods based on feature reduction and similarity analysis. In section 4, the experimental setup is presented, along with the test procedure. In section 5, the

* Corresponding author. Tel.: +34 958248899.
E-mail address: pablopadilla@ugr.es (P. Padilla).

http://dx.doi.org/10.1016/j.measurement.2014.09.009
0263-2241/© 2014 Elsevier Ltd. All rights reserved.

evaluation and results of the proposed methods are provided. Eventually, in Section 6, conclusions are drawn.

2. RF fingerprinting

The process of identifying wireless transmitters by examining the signal RF characteristics at the beginning of transmission is commonly referred as RF fingerprinting. The analysis and extraction of relevant parameters of the RF signal let define the RF identification data of a wireless device [4]. RF fingerprinting is focused on the registration and storage of the RF features of the signal transmitted by any device in the network. This group of RF features is the fingerprint of the wireless device [4,5].

The procedure for the extraction of the RF fingerprint of a wireless device implies the capture and analysis of the initial preamble of the RF signal [5,6]. In the literature, it can be found a list of parameters to be extracted: the transient waveform, the instantaneous phase or amplitude of the signal, their evolution in time, the signal tendency (first or second order derivative), the transient ramp profiles in a temporal range, etc. [4]. Fig. 1 shows an example of the RF preamble for two different wireless network devices at 2.4 GHz (Wi-Fi). Other issues regarding RF fingerprinting may be found in [7–11].

3. Identification method based on feature extraction and similarity analysis

Once the RF fingerprint of the wireless device is available, it is possible to define a procedure for its analysis and further identification. Some procedures have been proposed in the literature, based on four steps: channel monitoring, signal transient detection (preamble starting location), feature extraction (fingerprint) and classification/identification [6,8]. In this way, the approach proposed in this work combines a reduction in the feature space dimensionality given by each fingerprint with the comparison of each resulting fingerprint sample with a set of already labeled samples.

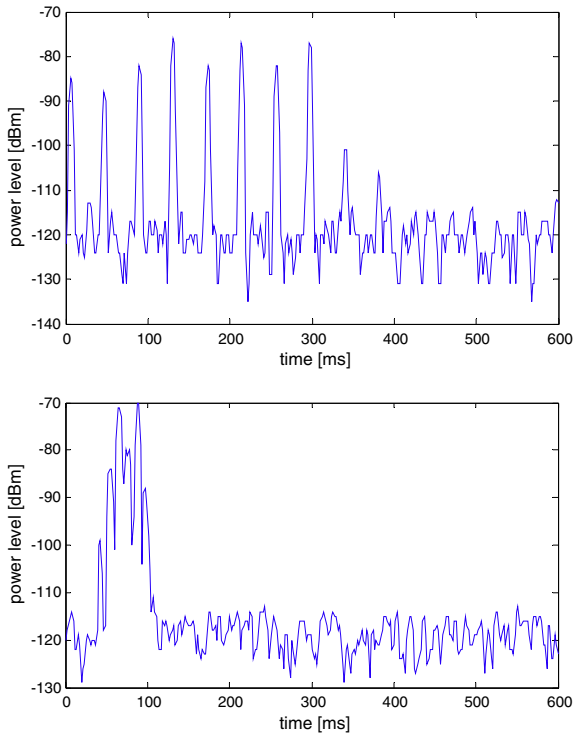

Fig. 1. Example of RF fingerprints of two different Wi-Fi devices.

3.1. Feature extraction and reduction based on PCA and PLS

In most of the cases, not all the available variables in the samples of a dataset contain relevant information. In order to extract the highest amount of information from these samples, feature reduction strategies are followed. The most common reduction methods are those based on latent structures for dimensionality reduction. Standard projection models are Principal Component Analysis (PCA) [12] and Partial Least Squares (PLS) [13]. Both of them are widely applied in data mining and data modeling.

PCA is applied to find the space of maximum variance in the M-dimensional feature space of a $(N \times M)$ zero-mean matrix $\mathbf{X}$, formed by N samples of M variables each one. The space transformation is linear, obtained using a calibration. PCA lets transform the original set of samples into a lower number K of uncorrelated features, called principal components (PCs), according to the expression:

$$\mathbf{X} = \mathbf{T} \cdot \mathbf{P}^{\mathbf{T}} + \mathbf{R} \tag{1}$$

where $\mathbf{T}$ is the $N \times K$ score matrix containing the projection of the original set in the K-subspace, $\mathbf{P}$ is the $M \times K$ matrix containing the K eigenvectors of $\mathbf{X}^T \cdot \mathbf{X}$, and $\mathbf{R}$ is residual of the transformation.

Another useful dimensionality reduction strategy consists on the least squares regression approach, PLS. The linear regression in PLS includes additional information to the original set of data (i.e. labels of the samples in the dataset). PLS is a linear algorithm for modeling the relation between two data sets $\mathbf{X}$ and $\mathbf{Y}$. The aim of the PLS regression is to estimate $\mathbf{Y}$ from a subspace of $\mathbf{X}$ which maximizes its covariance with $\mathbf{Y}$. This subspace of $\mathbf{X}$ is formed by the latent variables (LVs) in $\mathbf{X}$, in a similar way to PCA and its PCs. PLS decomposes the matrix of zero-mean variables $\mathbf{X}$ and the matrix of zero-mean variables $\mathbf{Y}$ into the form:

$$\mathbf{X} = \mathbf{T} \cdot \mathbf{P}^{\mathbf{T}} + \mathbf{E} \tag{2}$$

$$\mathbf{Y} = \mathbf{U} \cdot \mathbf{Q}^{\mathbf{T}} + \mathbf{F} \tag{3}$$

where $\mathbf{T}$ and $\mathbf{U}$ are the score matrices which contain the projections of $\mathbf{X}$ and $\mathbf{Y}$ to the latent subspace, $\mathbf{P}$ and $\mathbf{Q}$ are the regression matrices, and $\mathbf{E}$ and $\mathbf{F}$ are the matrices of residuals of $\mathbf{X}$ and $\mathbf{Y}$, respectively.

Then, the PLS regression is performed according to the expression:

$$\mathbf{Y} = \mathbf{X} \cdot \mathbf{S} \cdot \mathbf{Q}^{\mathbf{T}} + \mathbf{F} \tag{4}$$

$$\mathbf{S} = \mathbf{W} \cdot (\mathbf{P}^{\mathbf{T}} \cdot \mathbf{W})^{-1} \tag{5}$$

where $\mathbf{W}$ is a matrix of weights to be computed in the regression. As already mentioned, in most of the cases,

Table 1
Description of the experimental setup.

Description of the dataset	
Frequency	2.4 GHz
Total number of different devices (N)	20
Number of devices per manufacturer	2
Number of measured samples per device	15
Number of variables per sample (M)	310
Measurement setup	
Measuring equipment	R&S FS300 spectrum analyzer
Resolution bandwidth (RBW)	200 Hz
Span	0 Hz
Time Sweep	600 ms (40 ms/div, 15 div.)
Reference power level	−115 dBm
Video bandwidth (VBW)	None

the labels of the samples of the original dataset **X** are used to compose **Y**.

3.2. Similarity analysis based on distance to the most similar samples

The next step in the identification method implies the comparison of each processed fingerprint sample with a set of already labeled samples. The classification is achieved through the calculation of the lowest distance (d) of the fingerprint sample under test to the ones in a reference set. Under these circumstances, a list of the reference samples sorted by their distance to the sample under test can be computed. Considering the closest reference samples to the under-test one, commonly named 'neighbors', the sample under test can be classified. The classification of the sample under test (S_{test}) is done by selecting the label of the majority of the V nearest neighbors among all the I ($I = N-1$) available samples (Si) in the reference set, as in (6).

$$d_i = \sum_{m=1}^{M} |S_{test}(m) - S_i(m)| \tag{6}$$

where M is the number of variables in each sample.

Table 2
Commercial devices used in the experimentation.

Devices under test	
1. Zyxel NMD2205 (×2)	2. NetGear WNA3100 M (×2)
3. D-Link DWA-140 (×2)	4. ASUS USB-N13 (×2)
5. SWEEX LW323 (×2)	6. SiteCom 300 N (×2)
7. TP-LINK TL-WN821 (×2)	8. TRENDNET TEW-649UB (×2)
9. HTC Wildfire S (×2)	10. Intel WiFi Link 5300 (×2)

4. Experimental setup and test procedure

The proposed extraction procedure implies the capture and study of the preamble of the RF signal of the wireless device under test by means of a conventional RF spectrum analyzer with 'zero span' configuration, in order to get the time domain response of the channel [14]. In a real case, the signal can be easily extracted from the receiver and processed directly.

The identification procedure proposed in this work implies that, initially, a set of measured RF signals have to be conveniently stored and identified (trusted devices of the network, for example) in a reference dataset for further analysis of new unidentified device signals. Once the set is available, each new measured RF fingerprint is analyzed and compared with the ones in the dataset. If the new sample is identified and assigned to a known device, it is labeled as a new sample of this particular device. Otherwise, it is stored as a 'new device' signal. Thus, in both cases the sample is included in the reference dataset.

The test set-up considered in this work implies the use of an R&S FS300 with the configuration depicted in Table 1. The spectrum analyzer common setup parameters are described in [15]. This table provides additionally the description of the dataset acquired for the experimentation, along with the measurement setup. The referred dataset contains fingerprint measurements of a variety of commercial Wi-Fi devices. The description of the employed devices is provided in Table 2. Fig. 2 shows a caption of the measuring process with the spectrum analyzer.

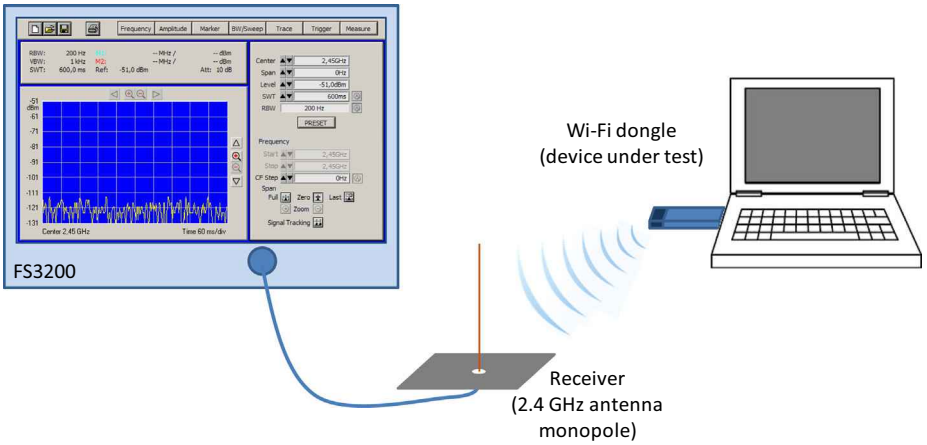

Fig. 2. Measuring scheme and processing setup with the R&S FS300 spectrum analyzer.

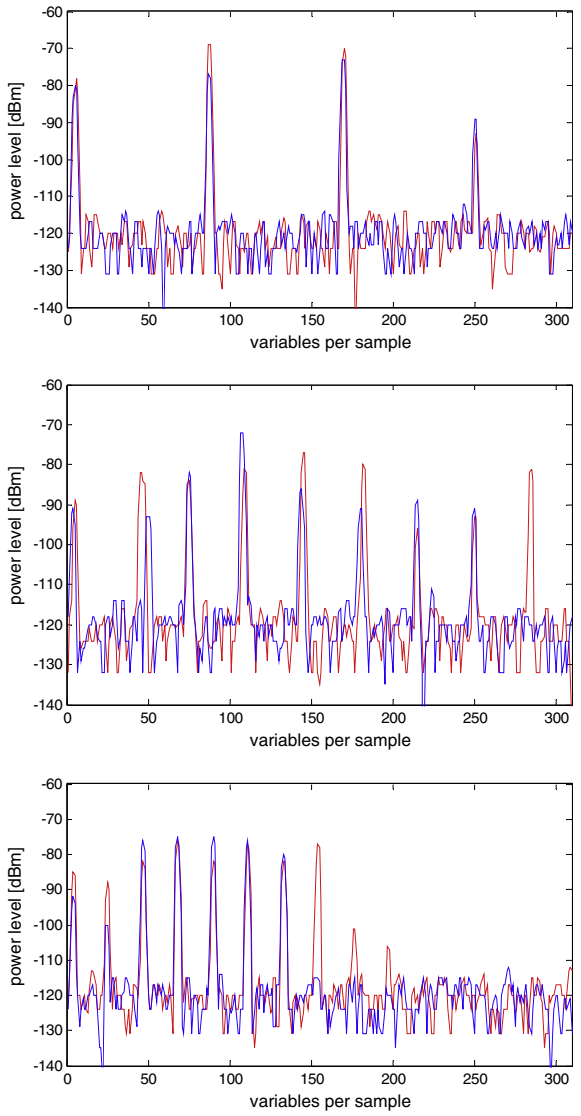

Fig. 3. Different RF fingerprints of commercial devices (two samples per each device), (a) SWEEX LW323, (b) ASUS USB-N13, (c) HTC Wildfire S.

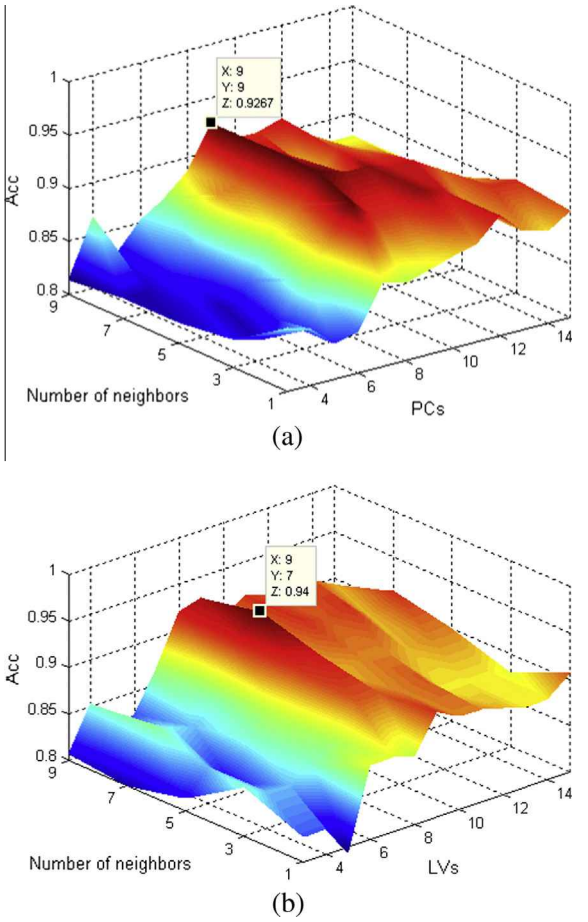

Fig. 4. Accuracy results considering the analysis of a set of devices, one per manufacturer, (a) PCA, (b) PLS.

5. Evaluation, results and discussion

Once the wireless devices are measured and the experimental dataset is conveniently labeled and stored, the performance evaluation is carried out. As already mentioned, the identification of a new device in the network is carried out with a comparison of its RF fingerprint with the list of

Table 3
Accuracy of the experimentation considering the analysis of a set of devices, one per manufacturer.

	Number of neighbors (N)									
	1		3		5		7		9	
Raw analysis	0.900		0.907		0.877		0.877		0.853	
With threshold (best: −105 dBm)	0.907		0.907		0.893		0.893		0.877	
PCA (9 *PCs*)	0.883		0.907		**0.927**		**0.927**		**0.927**	
PLS (9 *LVs*)	0.913		0.907		0.920		**0.940**		**0.927**	
Random selection	0.100		0.100		0.100		0.100		0.100	
	Number of components (*PCs* or *LVs*)									
	4	5	6	7	8	9	10	11	12	13
PCA ($V = 5$)	0.833	0.827	0.840	0.883	0.900	**0.927**	0.920	0.920	0.913	0.913
PLS ($V = 7$)	0.860	0.847	0.840	0.853	0.920	**0.940**	0.900	0.913	0.907	0.913

The best results in terms of performance are highlighted in bold.

the ones corresponding to the already known devices by means of the techniques provided in Section 3. Fig. 3 provides an example of different RF fingerprints. Each plot includes the fingerprint of two different samples of the same device.

A one-leave-out train and test strategy is selected for the performance evaluation. Thus, N evaluation iterations are carried out: sample n (from 1 to N) is extracted from the dataset and the resulting $N-1$ sample dataset is used as the reference set to be compared with. The success rate (Acc) is computed to evaluate the performance of the proposed procedure.

For the feature extraction and reduction step, four evaluation approaches are evaluated:

1. Raw analysis (all the available variables in the fingerprint sample).
2. Extraction of the variables that are over a particular signal threshold.
3. PCA analysis.
4. PLS analysis.

5.1. Identification analysis of a set of devices, one per manufacturer

In this first case, a subset of 10 devices is collected from the complete dataset, one device per manufacturer. Each device has 15 registers in the reference dataset. The accuracy results are provided in Table 3, and in Figs. 4 and 5.

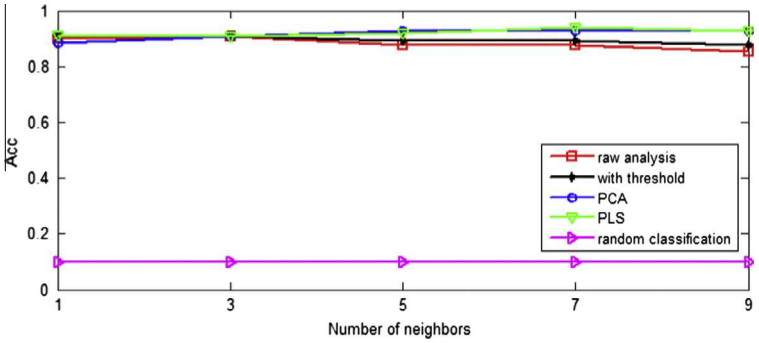

Fig. 5. Accuracy results for all the proposed approaches, considering the analysis of the set of devices, with one device per manufacturer (PCA with 9 PCs and PLS with 9 LVs).

Table 4
Accuracy of the experimentation considering the analysis of a set of devices, two per manufacturer.

	Number of neighbors (N)				
	1	3	5	7	9
Raw analysis	0.635	0.631	0.623	0.623	0.635
with threshold (best :−105 dBm)	0.659	0.620	0.631	0.623	0.651
PCA (11 *PCs*)	0.615	0.647	0.674	**0.694**	0.690
PLS (9 *LVs*)	0.690	0.694	**0.702**	**0.694**	**0.702**
Random selection	0.05	0.05	0.05	0.05	0.05

	Number of components (*PCs* or *LVs*)									
	4	5	6	7	8	9	10	11	12	13
PCA (V = 7)	0.552	0.623	0.619	0.678	0.643	0.667	0.674	**0.694**	0.674	0.651
PLS (V = 5)	0.580	0.627	0.631	0.670	0.667	**0.702**	0.698	0.690	**0.702**	0.682

The best results in terms of performance are highlighted in bold.

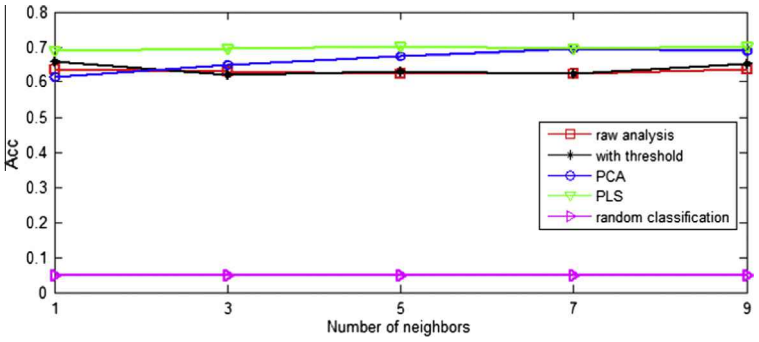

Fig. 6. Accuracy results considering the analysis of a set of devices, with two devices per manufacturer (PCA with 11 *PCs* and PLS with 9 *LVs*).

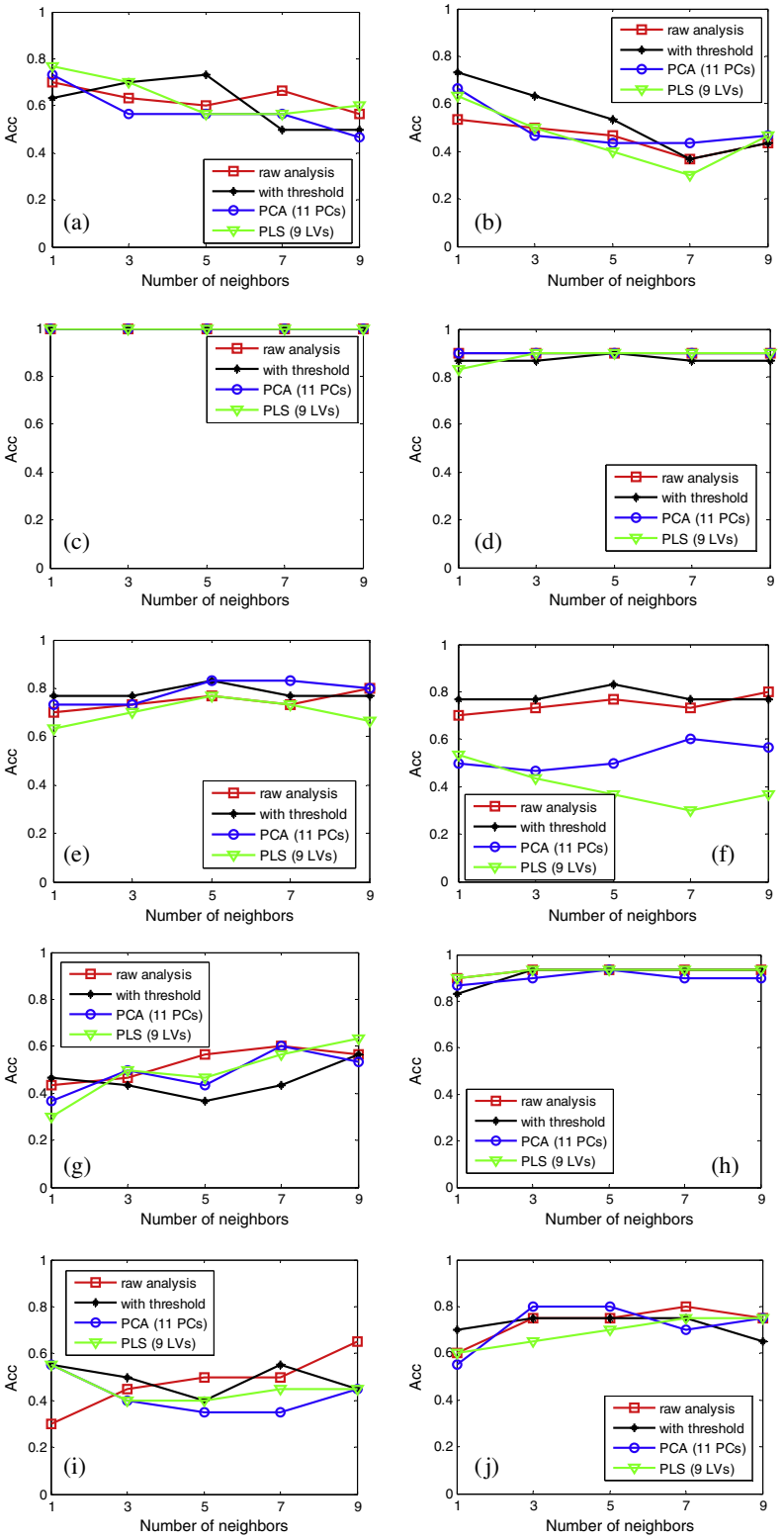

Fig. 7. Identification results for the devices of each manufacturer analyzed separately: (a) ASUS USB-N13, (b) D-Link DWA-140, (c) HTC Wildfire S, (d) NetGear WNA3100 M, (e) SiteCom 300 N, (f) TP-LINK TL-WN821, (g) Zyxel NMD2205, (h) SWEEX LW323, (i) Intel WiFi Link 5300, (j) TRENDNET TEW-649UB.

For the sake of fair comparison, a reference accuracy level is included, in order to fix the baseline level of the accuracy when random labeling is done.

As it is noticed, the recognition ability of all the proposed methods is very high, with values that are easily over 90%. The best results are obtained with PCA and PLS (the best one with PLS), considering a number of PCs or LVs around nine components. This is a common result in PCA or PLS [9,10], and indicates that these first nine PCs or LVs contain the highest amount of relevant information regarding the original data. The results of the four proposed evaluation approaches reveal that the RF fingerprint analysis is useful enough to recognize devices from different manufacturers. In addition, the use of techniques such as PCA or PLS simplifies the number of variables to be stored in the reference dataset per each sample, and yields a faster comparison step, with a lower computational cost in the identification system.

5.2. Identification analysis of a set of devices, two per manufacturer

In this case, a subset of 20 devices is collected from the complete dataset, two devices per manufacturer. Again, each device maintains 15 registers in the reference dataset. In this case, the method tries to identify not only among different manufacturers, but also among devices of the same manufacturer. The accuracy results are provided in Table 4 and Fig. 6.

It is noticed that the recognition ability of the method is reduced, as the most difficult task is to discriminate fingerprints of different devices of the same manufacturer. The result values reach 70% of accuracy, considering that the baseline value of random classification is 5%. Again, the best results are obtained with PLS.

5.3. Separate identification analysis of the two devices of each manufacturer

As revealed in the previous subsection, the presence of various devices form the same manufacturer reduces the efficiency of the method. In order to discriminate if this is a general fact, the devices of each manufacturer are analyzed separately. The identification results are provided in Fig. 7

It can be concluded that the degree of similarity between the fingerprints of different devices of the same manufacturer depends deeply on which the manufacturer is. For example, the HTC Wildfire S devices have a different enough RF fingerprint so that the identification is total; however, other devices such as the Intel WiFi Link 5300 ones have a very similar fingerprint that avoid the proper identification between devices of the same kind. In some of these cases, such as in Fig. 7g or Fig. 7i, the identification is as accurate as a random identification. Two samples of different devices of the same manufacturer are provided in Fig. 8, for HTC Wildfire S (results in Fig. 7c) and Intel WiFi Link 5300 (results in Fig. 7i).

As it is expected, the fingerprints of the HTC Wildfire S devices are severely different from one device to another. However, the contrary is the case of the Intel WiFi Link 5300, whose fingerprints are almost identical despite the device considered.

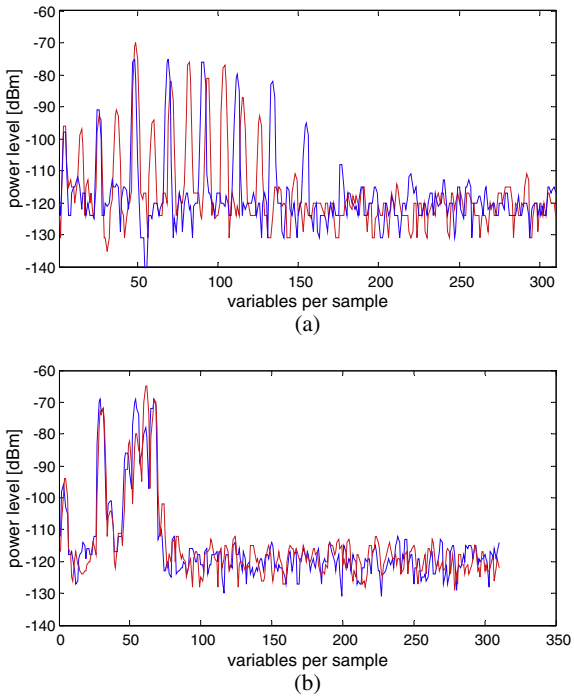

Fig. 8. Fingerprints of different devices of: (a) HTC Wildfire S and (b) Intel WiFi Link 5300.

The results provided let consider that the proposed RF fingerprinting identification method may be used in wireless networks to identify the devices connected, although it cannot provide itself a 100% identification rate. The best results may be obtained in combination with other identification methods in different layers of the protocol stack such as MAC direction identification in the link layer, considering a cross layer strategy. The problem in this link-layer identification based on the MAC direction is that, although it is quite precise in normal circumstances, it is prone to error considering the presence of nodes that can deliberately change their MAC direction. The addition of the RF identification in the physical layer prevents from this possible misidentification because of the deception in the MAC direction, as it is not possible to change the RF signal features of a particular device in the same way as the MAC layer can be changed. Therefore, there is a wide margin for the combination of the proposed method with other identification procedures of different layers for making secure the normal operation of wireless networks, being matter for future work.

6. Conclusions

This work is focused on the analysis of the RF fingerprint of wireless devices, in order to provide an adequate identification method in wireless networks. The identification method consists of the extraction of the preamble RF fingerprint of a device and its comparison with a set of

already known device fingerprints, in a similarity-based manner. The system performance is evaluated in a 20 Wi-Fi device dataset containing 15 fingerprint samples per device, with a train and test leave-one-out strategy. The best results are obtained with feature reduction techniques such as PCA and PLS, that are based on subspace transformation. More than 90% percent of proper identification is achieved in the case of using one device per manufacturer, and almost 70% when considering two devices per manufacturer. The identification between devices of the same manufacturer is really dependent of the considered one: some of them are easily identifiable, but some of them are unable to the discriminated. In addition to the obtained results, the feature reduction of such methods, PCA or PLS, reduces the computational costs of the subsequent similarity-based analysis.

Acknowledgement

This work has been carried out despite the economical difficulties of the authors' country. The authors want to overall remark the clear contribution of the Spanish Government in destroying the R&D horizon of Spain and the future of a complete generation.

References

[1] B. Kannhavong, H. Nakayama, Y. Nemoto, N. Kato, A. Jamalipour, A survey of routing attacks in mobile ad hoc networks, IEEE Wire. Commun. 14 (5) (2007) 85–91.
[2] K. Pelechrinis, M. Iliofotou, S.V. Krishnamurthy, Denial of service attacks in wireless networks: the case of jammers, IEEE Commun. Surv. Tut. 13 (2) (2011) 245–257.
[3] A. Mpitziopoulos, D. Gavalas, C. Konstantopoulos, G. Pantziou, A survey on jamming attacks and countermeasures in WSNs, IEEE Commun. Surv. Tut. 11 (4) (2009) 42–56.
[4] O. Ureten, N. Serinken, Wireless security through RF fingerprinting, Can. J. Elect. Comp. Eng. 32 (1) (2007) 27–33.
[5] R.W. Klein, M.A. Temple, M.J. Mendenhall, D.R. Reising, Sensitivity Analysis of Burst Detection and RF Fingerprinting Classification Performance, IEEE Int. Conf. Commun. 2009. ICC '09 1 (5) (2009) 14–18.
[6] W.C. Suski, M.A. Temple, M.J. Mendenhall, R.F. Mills, Using spectral fingerprints to improve wireless network security, Global Telecommun. Conf. 2008. IEEE GLOBECOM 2008. IEEE, 2008.
[7] H.L. Yuan, A.Q. Hu, Preamble-based detection of Wi-Fi transmitter RF fingerprints, Electron. Lett. 46 (16) (2010) 1165–1167.
[8] O. Ureten, N. Serinken, Bayesian detection of Wi-Fi transmitter RF fingerprints, Electron. Lett. 41 (6) (2005) 373–374.
[9] V. Lakafosis, A. Traille, L. Hoseon, E. Gebara, M.M. Tentzeris, G.R. DeJean, D. Kirovski, RF fingerprinting physical objects for anticounterfeiting applications, IEEE Trans. Microw. Theory Tech. 59 (2) (2011) 504–514.
[10] S.U. Rehman, K.W. Sowerby, C. Coghill, Analysis of impersonation attacks on systems using RF fingerprinting and low-end receivers, J. Comput. Syst. Sci. 80 (3) (2014) 591–601.
[11] R.W. Klein, M.A. Temple, M.J. Mendenhall, Application of wavelet-based RF fingerprinting to enhance wireless network security, J. Commun. Networks 11 (6) (2009) 544–555.
[12] J. Jackson, A User's Guide to Principal Components, Wiley-Interscience, England, 2003.
[13] H. Abdi, Partial least squares regression and projection on latent structure regression (PLS-Regression), Wiley Interdisciplinary Rev.: Comput. Stat. 2 (2010) 97–106.
[14] M. Bertocco, A. Sona, On the measurement of power via a superheterodyne spectrum analyzer, IEEE Trans. Instrum. Meas. 55 (5) (Oct. 2006) 1494–1501.
[15] Rohde, Schwarz, Ch 6: Using the R&S FS300, in: Spectrum Analyzer R&S FS300 Operating Manual, eighth ed., 2004.

All in-text references underlined in blue are linked to publications on ResearchGate, letting you access and read them immediately.