Privacy leakages in Smart Home Wireless Technologies

Ignacio Sanchez*, Riccardo Satta*, Igor Nai Fovino*, Gianmarco Baldini*, Gary Steri*,
David Shaw* and Andrea Ciardulli†
Institute for the Protection and Security of the Citizen, Joint Research Centre, European Commission
Via E. Fermi 1, Ispra, 21027, VA, Italy
Email: *{ignacio.sanchez, riccardo.satta, igor.nai-fovino, gianmarco.baldini, gary.steri, david.shaw}@jrc.ec.europa.eu ;
†{andrea.ciardulli}@ext.jrc.ec.europa.eu

***Abstract*—The concept of Smart Home where appliances, sensors, actuators, displays and computing resources are connected and interact to support the life of the citizen is being increasingly researched. In this context, the Wi-Fi communication technology has grown to become the de-facto standard for data communications in Smart Home environments, with cordless telephony being dominated by the DECT protocol. Even though both technologies incorporate sets of security features aimed at securing the confidentiality and integrity of the communications, the nature and the design of both radio-frequency protocols make them vulnerable, up to a certain extent, to privacy leakages through traffic analysis attacks. In this paper we explore the information leakage vulnerabilities inherent to these technologies and their potential impact on citizens' privacy in the context of the Smart Home. We demonstrate how the websites visited by a smart device can be inferred by applying machine learning and pattern matching techniques to eavesdropped encrypted traffic.**

***Keywords*—Privacy, Wireless communications, Machine learning**

I. Introduction

The concept of Smart Home where appliances, sensors, actuators, displays and computing resources are connected and interact to provide new services raised recently huge attention within both the research and policy communities. The Smart Home scenario is part of the wider Internet of Things (IoT) paradigm and it inherits from IoT all the aspects of connectivity of the involved devices. In this particular context, Wi-Fi has become the predominant wireless technology for data communications in Smart-Home environments, with the cordless telephony being dominated by the DECT standard. While there has been considerable research activity in IoT technologies applied to the Smart Home scenario, the societal concerns of this technological evolution in relation to the privacy and security of the citizen appear to be still at an embryonic stage. We argue that the increasing use of smart-devices, their interactions within the domestic environment and with the external world, if not properly regulated, will constitute in the near future a major risk for citizens' privacy and security. To sustain our claim, in this paper we explore information leakage vulnerabilities inherent to communication technologies used in Smart Homes and their potential impacts. After presenting the state of the art in this area, we analyse the ways in which private information might leak from the wireless protocols used in Smart Homes. We demonstrate how, even when encryption mechanisms are put in place, it would be possible to profile the behaviour of the end-user and of the smart-devices operating in his house. To support our analysis, we developed a test use case where a user navigates on Internet through his smart device using an encrypted Wi-Fi connection and where a malicious neighbour is willing to profile the Internet preferences of the victim. The only precondition for our test-case is that the malicious neighbour is able to capture the encrypted traffic between the smart device and the Wi-Fi access point (operation which can be easily done using a laptop with a wireless card). It can be noted that, as the traffic is encrypted, in theory the attacker would not be able to see the final destination address of the communications. However, we demonstrate that by applying pattern matching recognition and supervised machine learning algorithms on the size and sequence of the encrypted packets, it is possible to reconstruct with high accuracy the main navigation preferences (in term of websites visited) of the target victim. On the basis of these results, we elaborate on the possible additional information which might be extracted from the formally secure and encrypted communication channels used by Smart Home devices and on the needs for better attention on Smart Home privacy issues. We conclude by describing possible future investigation in the field.

II. Related work

There has been a growing interest in security and privacy issues in the Smart Home by academic, industry and government communities. In [1] the authors provide a review of leading Smart Home projects and the associated technologies of wearable/implantable monitoring systems and assistive robotics to support elderly and disable people. The paper correctly points out that the benefits of remote monitoring must be carefully evaluated against respect for privacy, confidentiality, and security and various projects are identified, which try to address this need. A more recent survey by Komninos and others [2] has investigated and identified security challenges in Smart Homes with a specific focus on the relation to Smart Grid and Smart Meters. In fact, Smart Meters are often considered a potential privacy vulnerability in the Smart Home and an extensive number of papers have both investigated and presented mitigation solutions for privacy threats as in [3], which describes a theoretical framework to quantify the utility-privacy trade-off in Smart Meter data. While we acknowledge that the Smart Meter is a potential vulnerability, the analysis presented in this paper tries to address all the potential privacy threats which can be exploited through the wireless communication technologies used in a Smart Home. The authors in [4] have demonstrated that private activities in the home such as cooking, showering, toileting, and sleeping can be detected by eavesdropping on the wireless transmissions of sensors in a home, even when all of the transmissions are encrypted. They used a combined Fingerprint and Timing-based Snooping (FATS) attack, which is conceptually similar to the approach used in the present paper but applied to a different wireless technology.

In a similar way, in [5] the authors provide an analysis of the privacy risks and potential mitigation techniques for the sensors networks (IEEE 802.15.4 and ZigBee).

An example of eavesdropping and profiling of encrypted Wi-Fi communications is presented in [6], where it is shown that despite encryption, a side-channel information leak is a realistic and serious threat to user privacy in Wi-Fi networks.

III. Smart Home Scenario

The concept of *Smart Home* is not new; indeed it was introduced for the first time back in 1984 by the America Association of House Builders. Aldrich [7] defines a Smart Home as a residence provided with computing and information technologies allowing to anticipate or respond to the needs of the occupants. The advent of the Internet of Things (IoT) paradigm required to further extend this definition introducing the concept of "external world interaction" as not only the Smart House interacts with the occupants, but, in a broader context, it is also enabled to interact with the external world. From an engineering view-point, we can classify the elements composing a smart-house as follows:

- **Sensors and low level actuators**: light sensors, photocells, electric actuators (e.g. door motors, room lights etc.), motion sensors (such as those used for surveillance systems).
- **Smart Devices**: all the devices which are provided with a reasonable amount of power computation and which can interact directly with the end-users and the external world. Examples of these devices are smartphones, Smart TV, smart forecast stations, smart cooling systems etc. etc.
- **ICT Communication Devices**: all devices which are part of the ICT infrastructure of the Smart House dealing with both data and voice communications.
- **Smart Services**: this category groups together all the services provided within the Smart Home domain. This layer is composed of software (centralised or distributed) which might reside in the smart-object, in the cloud, or in the ICT systems of the Smart Home.

The glue that brings together the components of the Smart Home ecosystem is composed by the collection of communication protocols that allow the smart objects to collaborate and interact. In this context, the market is quickly converging toward the Wi-Fi protocol for data communications together with the DECT standard for cordless voice communications.

IV. Privacy risks, an overview

The smart-devices populating our houses can gather a huge amount of sensitive information. Unfortunately, in the enthusiastic impetus of proposing new services leveraging on the IoT paradigm without paying the due attention to the security aspects, the Smart Home digital inclusion introduced a new layer of potential threats against the privacy of the citizen. In this section we provide an overview of the possible ways in which a malicious user could leverage on Smart Home leakages to infer sensitive information on the occupants.

A. DECT

Digital Enhanced Cordless Technology, DECT, has been dominating the residential environment of cordless communications for more than a decade. With the evolution of the Smart Home concept, some of the available products have started to integrate DECT cordless phones with the networked ecosystem of smart devices that are usually interconnected via Wi-Fi.

A typical DECT installation consists of one or more cordless phones, known as Portable Parts in the standard, connected to a base station, known as Fixed Part.

The FP traditionally consisted of a dedicated device that connected the DECT network to the Public Switched Telephone Network using an analogue land line. Nowadays in the ecosystem of the Smart Home the DECT FP is often part of more complex communication devices that unify the traditional land lines with other communication lines via VoIP.

Like any other radio-frequency protocol, DECT communications are vulnerable to eavesdropping when encryption is not used. The feasibility of passive eavesdropping of DECT communications has been demonstrated by [8], using cheap dedicated hardware, and [9] employing low cost Software-Defined Radios, including the family of RTL-SDR devices. With the wide availability of these cheap SDR DVB-T dongles, the ease of monitoring DECT communications and performing privacy attacks has risen considerably. In order to protect the privacy of the communications DECT supports the usage of encryption which is mandatory in the newer revisions of the standard. When implemented properly, the DECT encryption helps to protect the phone conversation, despite some vulnerabilities documented in the encryption algorithm and associated protocol [10]. However, information about the user behaviour is still leaked despite the encryption. An attacker could remotely monitor the DECT communications and, in combination with information leaked from the WiFi communications described in the next subsection, determine private information about the individuals of the smart home such as their behavioural habits.

In DECT the Fixed Part is identified by its Radio Frequency Part Identifier (RFPI). The RFPI also identifies the DECT network and it is continuously broadcast in the clear by the FP. The RFPI is a unique 40-bit value for each FP device which acts like a MAC address for a Wi-Fi network. It is composed of the identifier of the manufacturer plus the unique identifier of this particular device within the manufacturer's production. The PPs are similarly identified by the International Portable User Identification (IPUI).

The DECT encryption protects both the control and the voice data that is transmitted over the air. However, the establishment and reception of phones calls where the session key is negotiated, before the encryption is used, can be observed by an eavesdropper. Furthermore, the transmission of voice is revealed by the presence of B-Field in the DECT packets as well as the existence of PP frames, regardless the encryption being used or not.

By remotely monitoring the DECT communications of an Smart Home, an attacker could gather useful information such as the identification and location of the DECT base and the several cordless terminals as well as the profiling of the incoming and outbound calls. Even though the voice as well as the calling and called numbers are protected when the encryption is implemented properly, the times, direction (inbound or outbound) and duration of the calls can still be determined by an eavesdropper.

All this leaked information, in combination with the information leaked by the Wi-Fi communications, pose a non-negligible threat to the privacy of the inhabitants of the smart home.

B. Wi-Fi

Wireless Local Area Network (WLAN) can be considered today the most used local network infrastructure. The typical configuration in home networks foresees an Internet router connected to the phone network on one side and to a wireless access point on the other. The access point is in charge of providing the 802.11 connectivity to all the devices in its range. 802.11 supports different security features to protect both the access to the network, via authentication mechanisms, and the traffic flowing through it, by means of encryption.

Even though the early Wi-Fi encryption, named Wired Equivalent Privacy (WEP), was found to be insecure, the newer Wi-Fi Protected Access (WPA) encryption, in both TKIP and AES variants, provides effective protection of the confidentiality and integrity of the communications, as long as a secure pass-phrase is used in the PSK configuration. However, even when secure encryption is used, a determined attacker is still able to guess useful information, which potentially can be used to infer sensitive knowledge on Smart Home occupants. Here we provide some examples of data-leakage.

1) Inventory of IoT devices: The first step to breach the privacy of the Smart Home is to know what it contains. The predominantly wireless nature of the Smart Home ICT infrastructure is a big advantage for a remote attacker. In fact, the attacker can just sit outside the perimeter with a good antenna and a wireless sniffer tool, to be able to see the data-flows generated in a target house. Even when encryption is used, not all the transmitted information is protected, as the source and destination MAC addresses of the devices still travel in clear over the air. An attacker can remotely collect the MAC addresses of the devices in order to identify the active ones operating in the Smart Home. The correspondence between the MAC addresses and vendors is regulated and maintained by the IEEE Standards Association, who release daily an updated list.[1] Taking advantage of this information, an attacker can easily determine the number and type of devices which are active in a house at a given time of the day.

2) Flows of data: The analysis of the shape of the traffic, even if encrypted, can be helpful to derive implicit information: (a) *time analysis and traffic burst correlation* might allow to identify the relation between two object in the local network (e.g. a smart-phone used as remote controller of a TV) while (b) *traffic shape analysis* can be used to infer the function of some objects (e.g. a NAS used as video repository), or the type of application installed on a device (e.g. skype, browsers etc.)

Again, also in this case, by correlating and mapping the evidence collected, an attacker can create a view (even if partial) of what is contained in the Smart Home.

3) Inference of user behaviour: A careful analysis of the previously described leaked information could lead to the profiling of the house occupants. For example, the presence of the MAC address of a given mobile phone in certain time windows during the day would allow to guess when a certain person is usually at home. The flow burst between the access point and a smart heating system, might reveal the imminent return at home of an occupant (which managed through its Internet connection to switch on the heating system). The flow between the MAC address of the Smart TV and the access point would allow to guess when in the house someone is looking the TV, and consequently, when that person, for example, goes to sleep etc. These are some easy example of the potential leaks permeating the “digital walls” of a Smart House. They are of course simplistic as the aim of this section is that of raising the attention to the problem; however the application of data-mining (e.g. clusterisation, association rule mining, decision trees etc.) and machine learning techniques on these few type of data can potentially reveal to a determined attacker a huge amount of even more sensitive information. To prove this claim, in the following section we will present the results of some experimental field tests.

V. Smart Home Privacy Leakage Use Case

As described in the previous section, Smart Home wireless communication technologies are prone to privacy leakages which, if carefully treated by a determined attacker, might lead to the inference of sensitive information about the occupants of the house. The scope of this section is demonstrating the immediacy of this risk through a set of practical experiments.

In particular, we show how it would be possible for an attacker equipped with a wireless laptop located outside the Smart House, to guess the browsing preferences (in term of web-site visited) of someone living in the target house and using a home network protected by any of the Wi-Fi encryption modes. It is worth to note that we do not rely on breaking the encryption algorithm.

Specifically, we identify newspaper websites accessed by the smart device over the encrypted network by extracting a fingerprint made up of the signal composed by the sizes and timings of the packets. In fact, elements of the HTML template static over the time, such as CSS and image files in the form of headers or buttons, are likely to show a distinctive pattern in that signal.

A. Scenario setup

We designed two different scenarios for the use case. The first one is *One-to-N traffic classification*, and consists of: i) defining a set of N candidate websites; ii) acquiring traffic by listening at the target's Wi-Fi connection; iii) classifying the acquired data as belonging to one web site out of the predefined set of N possible websites. This scenario simulates the case when the attacker has already some information on the web site the target is possibly browsing (e.g., a news site, a file sharing site), and can therefore define a set of candidate sites. In this scenario, we make use of machine learning techniques to learn a model of the traffic generated by each candidate site. This requires the prior acquisition of a *training set* [11], i.e., a set of labelled samples (acquired traffic) of each website in the candidate list, from which a statistical classifier is trained.

The second scenario is *One-to-One traffic matching*. It consists in i) choosing one probe web site p; ii) acquiring traffic by listening at the target's Wi-Fi connection; iii) determining whether the traffic acquired from the target is generated by browsing the web site p, or not. Contrary to the former scenario, here the attacker is interested only in verifying whether the target is browsing a given web site, or not. In this

[1] http://standards.ieee.org/develop/regauth/oui/oui.txt

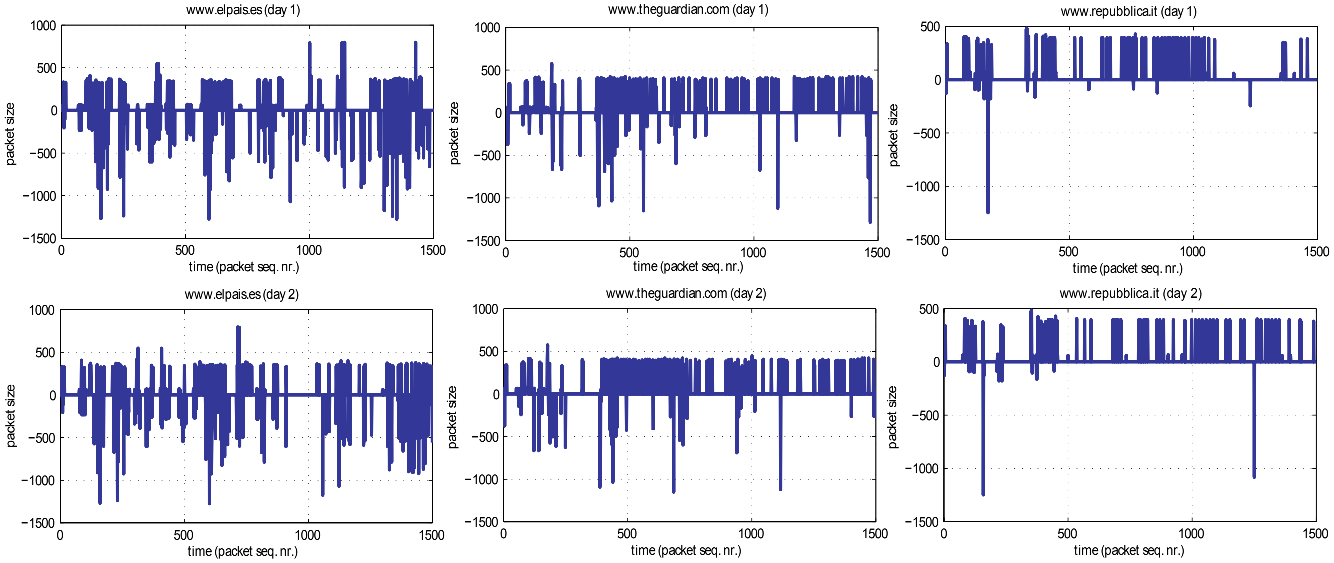

Figure 1: Example signals (two acquisitions in different days of three different web sites).

case, no training set is required, and simple pattern matching can be used instead, to compare the signature extracted from p with the signature extracted from the target's traffic.

A summary of the classification and pattern matching techniques used to implement the aforementioned scenarios is presented in the next section.

B. Pre-processing of traffic data, classification and matching

Before classification or matching, the traffic data is pre-processed so to obtain a descriptive, fixed-size vector representation. This is done in two steps, *filtering* and *vector representation*.

The packet sizes from the traffic data are first filtered by removing those sizes that were not present in all the acquisitions for each website. The idea behind this filtering is to eliminate fluctuating packet sizes that are not likely to be part of the fingerprint of the website (e.g. news of the specific day, dynamic adverts, etc.). In addition to that, we have found that the maximum values represent the MTU of the channel and the minimum values represent simple empty ACK TCP segments. These values are not relevant for our purposes and have been removed as part of the filtering of the signal.

At the end of the filtering step, the sequence of packet sizes (positive values: upload; negative values: download) is treated as a 1D signal in the time domain. Figure 1 shows the signal extracted for the download of 3 different newspaper websites over a period of 3 days. The time of arrival of the packet is not represented in the plot, just the order. One can clearly distinguish some patterns in both the download and upload channels.

The traffic signal is subdivided into 3 segments of equal length, and a histogram of the packet sizes is computed for each segment. To form the final vector representation, the three histograms are concatenated, and the vector is normalised to sum 1. One-to-N traffic classification and One-to-One traffic matching are then conducted using pre-processed data, as follows.

In the classification scenario, a statistical classifier is at first trained in a data set containing several instances of traffic acquisitions for each of the N candidate web sites (*training set*). Once trained, the classifier is used to predict the actual class of an unseen sample (in this case, the traffic data acquired from the target). To this aim, we adopted three widely used classifiers.

The first one is a k-Nearest Neighbour (kNN) classifier [12], possibly one of the simplest machine learning algorithms, which assigns to an unseen sample the most frequent class among the k nearest training samples in the training set. The parameter k is chosen as to minimise the average leave-one-out classification error in the training set.

The second classifier is a Support Vector Machine (SVM) with linear kernel [13]. Linear SVMs map the data in a higher dimensional space where the classification problem is solved linearly (i.e., by finding an optimal separation hyperplane between the classes). Since SVMs are binary (two-classes) classifiers, in multi-classes problems like the one at hand $\frac{N\cdot(N-1)}{2}$ binary SVMs are trained over all the possible combination of two classes, then the final decision is based on the highest score among all binary SVMs.

Finally, the third classifier is a Fisher's Least Square Linear Discriminant [14], which finds the linear discriminant function between the classes in the training set, by minimising the errors in the least square sense. Like SVM, a Fisher Linear Discriminant is a binary classifier, therefore N one-vs-all classifiers are trained in this case, and the final decision is based on the highest score among all of them.

In the matching scenario, the vector representations of the probe web site, and of the signal acquired from the target (denoted here respectively as p and t), are compared directly by means of *histogram intersection*, which measures the similarity of p and t and is defined as

$$S(p,t) = \sum_{i=1}^{n} \min(p_i, t_i) \qquad (1)$$

where n is the size of the vector representation, and p_i and t_i denote respectively the i-th element of p and of t. If p and t are normalised to sum 1, $S(p,t)$ is bound into $[0,1]$. The higher the value of $S(p,t)$ the more the likelihood that p and t represent the same web site. To obtain a sharp decision, $S(p,t)$

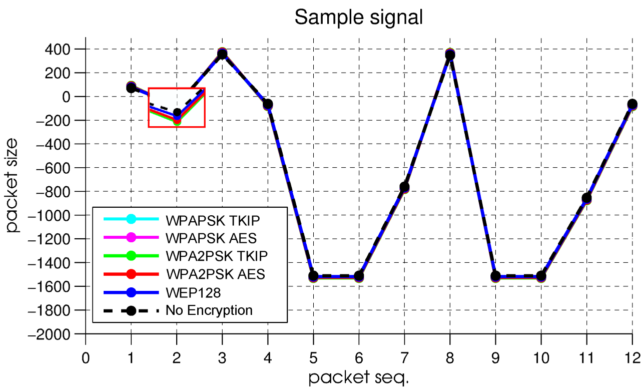

Figure 2: Signal of packet sizes for the several types of WiFi encryption

is compared against a fixed threshold Th: if $S(p,t) \geq Th$, the two vectors p and t are deemed as representing the same website.

C. Experimental Data Acquisition

In order to acquire the experimental data on which the classification techniques described above will be applied, we set up a network architecture mimicking a typical Smart Home scenario in order to simulate the action of an attacker.

The simulation involves a Smart Home client that is browsing, via Wi-Fi, a website using the Internet connection provided by a wireless Access Point with Internet connectivity, and the attacker's device which is outside the Smart Home environment but is able to capture the Wi-Fi transmissions.

We ran our simulations using websites installed in a local web server (so as not to have the interference of external links and have the possibility to precisely analyse the traffic generated by well known web pages) as well as ”live” Internet websites, in particular newspapers from different countries. In both cases, we employed different Wi-Fi clients (a Smart TV, laptops and mobile phones) to browse the websites at the same time that a laptop equipped with a Wi-Fi network card performed the attack.

This data acquisition process consisted of recording all the Wi-Fi packets exchanged between a Smart Home client and the access point. The determination of the flow to be recorded was performed by analysing the source and destination MAC addresses of the eavesdropped transmission.

For comparison purposes, the Wi-Fi connection between smart home client and access point was set up using six different configurations: *open (no encryption)*, *WEP encryption*, *WPA-PSK TKIP encryption*, *WPA-PSK AES encryption*, *WPA2-PSK TKIP encryption*, *WPA2-PSK AES encryption*. All the websites were browsed by our clients using all the six of them, allowing us to analyse the differences in the traffic generated by the same communication (i.e. same web page) as the security protocol changes. After a few tests, the acquisition process was done manually for the clients such as Smart TV and smartphones, where the automatic browsing and control of the data quality (e.g. temporary unavailability of the websites) required more efforts, whilst it was automatised for the laptop.

D. Results

In the several Wi-Fi encryption methods, the size of the encrypted packets, denoted as S_e, can be calculated as the size S_p of the equivalent non-encrypted packet, plus a fixed overhead S_h added by the headers specific to the encryption. Thus, there is no need that the data acquired as training set for one-to-N classification, and as probe for one-to-one matching, come from the same type of encrypted Wi-Fi network. Indeed, the model can be built from non-encrypted traffic; after the data is collected, one can then adapt it to the specific Wi-Fi encryption by simply compensating the fixed overhead.

TABLE I. PERFORMANCE OF DIFFERENT CLASSIFIERS IN TERMS OF ACCURACY

classifier	accuracy %
k-NN	83.33
Support Vector Machine, linear kernel	89.00
Fisher's Least Square Linear Discriminant	100.00

As a proof to that assumption, in figure 2 we plot a portion of a signal extracted from the recording of a simple website through each of the six types of Wi-Fi network. It can be seen that the sizes of the recorded packets are slightly different for each type of network and that the overhead is constant and dependent on encryption.

The 3 classification models we propose, k-NN, linear SVM and Fisher's Least Square Linear Discriminant, have been trained using the data acquired from two consecutive days for the download of ten different newspaper websites. Each newspaper website was downloaded ten times during that day, for a total of 200 instances to train the classifier with.

TABLE II. PERFORMANCE OF THE k-NN CLASSIFIER IN TERMS OF CONFUSION MATRIX

		Predicted web site									
		#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
Real web site	#1	30	0	0	0	0	0	0	0	0	0
	#2	0	30	0	0	0	0	0	0	0	0
	#3	0	0	30	0	0	0	0	0	0	0
	#4	2	0	0	0	0	0	20	8	0	0
	#5	0	0	0	0	30	0	0	0	0	0
	#6	0	0	0	0	0	30	0	0	0	0
	#7	0	0	0	0	0	0	30	0	0	0
	#8	0	0	0	0	0	0	0	30	0	0
	#9	0	0	0	0	0	19	0	0	10	1
	#10	0	0	0	0	0	0	0	0	0	30

Table I shows the performance obtained in the one-to-many classification scenario, by each of the three proposed classifiers in terms of their accuracy (average correct classification rate). Tables IV, II and III show the corresponding confusion matrices [2].

Figure 3 depicts the False Positive and False Negative ratios versus the decision threshold Th in the one-to-one matching scenario. The EER (Equal Error Rate) in the figure is the value of Th for which False Positive and False Negative ratios are equal.

Results clearly state that in the classification scenario a capture from an encrypted network can be accurately classified as belonging to a given website. Moreover, when targeting a specific web site (i.e., the matching scenario), the Equal Error Rate is as low as 6%.

VI. CONCLUSIONS

In this paper, we aimed at raising awareness regarding the limitations of current wireless communications security

[2] The (i,j) element of the confusion matrix is the number of instances of class i classified as belonging to class j.

TABLE III. PERFORMANCE OF THE SUPPORT VECTOR MACHINE CLASSIFIER IN TERMS OF CONFUSION MATRIX

		Predicted web site									
		#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
Real web site	#1	28	0	0	0	0	0	0	0	2	0
	#2	0	30	0	0	0	0	0	0	0	0
	#3	0	0	30	0	0	0	0	0	0	0
	#4	0	0	0	0	0	0	30	0	0	0
	#5	0	0	0	0	30	0	0	0	0	0
	#6	0	0	0	0	0	30	0	0	0	0
	#7	0	0	0	0	0	0	30	0	0	0
	#8	1	0	0	0	0	0	0	29	0	0
	#9	0	0	0	0	0	0	0	0	30	0
	#10	0	0	0	0	0	0	0	0	0	30

TABLE IV. PERFORMANCE OF THE FISHER CLASSIFIER IN TERMS OF CONFUSION MATRIX

		Predicted web site									
		#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
Real web site	#1	30	0	0	0	0	0	0	0	0	0
	#2	0	30	0	0	0	0	0	0	0	0
	#3	0	0	30	0	0	0	0	0	0	0
	#4	0	0	0	30	0	0	0	0	0	0
	#5	0	0	0	0	30	0	0	0	0	0
	#6	0	0	0	0	0	30	0	0	0	0
	#7	0	0	0	0	0	0	30	0	0	0
	#8	0	0	0	0	0	0	0	30	0	0
	#9	0	0	0	0	0	0	0	0	30	0
	#10	0	0	0	0	0	0	0	0	0	30

used in Smart Home environments in protecting the privacy of the inhabitants. We have shown that, despite the usage of secure encrypted Wi-Fi and DECT communications in Smart Home environments, relevant personal information such as users' presence, location and behaviour can be leaked. This is due to several factors, the main ones being that i) due to the inherent nature of these radio-frequency protocols, the transmitted information surpasses the physical boundaries of the Smart Home making it easier to eavesdrop them; ii) whilst in some cases the security mechanisms foreseen in these protocols can make a good job at protecting the confidentiality and authenticity of the information transmitted, they do not offer effective protection against information leakage attacks.

In the case of DECT voice communications, we have shown how the presence, time and duration of phone calls can be inferred by an attacker able to listen to the transmissions. As for the Wi-Fi protocol, we have demonstrated how the order and sizes of the encrypted packets, together with the ability to select one particular data flow from the analysis of the MAC addresses, can be used to detect which websites are accessed by a particular smart device via HTTP.

Countermeasures should be put in place in order to mitigate the risk of such privacy leaks. Possibilities include: injection of noisy data flows in the communication among smart devices (both between them and to the Internet); consider adding encryption in the higher communication layers; improve the resistance of existing protocols to such attacks. All these solutions come with drawbacks; our ultimate aim is to encourage the research community to provide valid answers to the forthcoming privacy challenges of Smart Home environments.

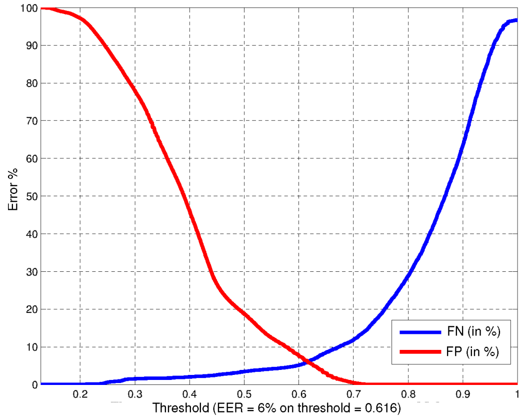

Figure 3: False Positive and False Negative ratios versus decision threshold

REFERENCES

[1] M. Chan, D. Estève, C. Escriba, and E. Campo, "A review of smart homespresent state and future challenges," *Computer methods and programs in biomedicine*, vol. 91, no. 1, pp. 55-81, 2008.

[2] N. Komninos, E. Philippou, and A. Pitsillides, "Survey in smart grid and smart home security: Issues, challenges and countermeasures," *Communications Surveys and Tutorials, IEEE*, vol.PP, no.99, pp.1,1.

[3] L. Sankar, S.R. Rajagopalan, S. Mohajer, and H.V. Poor, "Smart meter privacy: A theoretical framework," *Smart Grid, IEEE Transactions on*, vol. 4, no. 2, pp. 837–846, June 2013.

[4] V. Srinivasan, J. Stankovic, and K. Whitehouse, "Protecting your daily in-home activity information from a wireless snooping attack," in *Proc. Intl Conf. Ubiquitous Computing (UbiComp 08)*, Seoul, South Korea, 2008, pp. 202–211.

[5] K. Islam, W. Shen, and X. Wang, "Security and privacy considerations for wireless sensor networks in smart home environments," in *Proc. 16th IEEE Int. Conf on Computer Supported Cooperative Work in Design (CSCWD), 2012*, Wuhan, China, 2012, pp. 626–633.

[6] S. Chen, R. Wang, X. Wang, and K. Zhang, "Side-channel leaks in web applications: A reality today, a challenge tomorrow," in *Proc. 2010 IEEE Symposium on Security and Privacy*, Washington, DC, USA, 2010, pp. 191–206.

[7] F. K. Aldrich, "Smart homes: Past, present and future," in *Inside the Smart Home*, Richard Harper, London, UK: Springer, 2003, pp. 17–39.

[8] S. Lucks, A. Schuler, E. Tews, R.-P. Weinmann, and M. Wenzel, "Attacks on the dect authentication mechanisms," in *Proc. Topics in Cryptology, CT-RSA 2009*, San Francisco,CA, USA, 2009, pp. 48–65.

[9] I. Sanchez, G. Baldini, D. Shaw, and R. Giuliani, "Experimental passive eavesdropping of digital enhanced cordless telecommunication voice communications through low-cost software-defined radios," *Security and Communication Networks*, vol. PP, no.99, pp. 1,1, 2014.

[10] K. Nohl, E. Tews, and R.-P. Weinmann, "Cryptanalysis of the DECT standard cipher," in *Proc. Fast Software Encryption*, Seoul, Korea, 2010, pp. 1–18.

[11] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed, Hoboken, NJ: Wiley-Interscience, 2000.

[12] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *The American Statistician*, vol. 46, no. 3, pp. 175–185, 1992.

[13] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121-167, June, 1998.

[14] A. Webb, *Statistical pattern recognition*, 3rd ed. Hoboken, NJ, USA: John Wiley & Sons, 2011.