Analysis of the Wireless Covert Channel Attack: Carrier Frequency Selection

Geir Olav Dyrkolbotn
Norwegian Information Security Lab, Gjøvik University College
geirolav.dyrkolbotn@gmail.com

Abstract

The wireless covert channel attack (WCCA) was suggested to see what it would take to make attacks on smart cards feasible outside a lab environment. Subversive code hidden on the card during manufacturing would manipulate the emitted electromagnetic energy during normal operation. Experiments on modern smart cards confirmed that the attack is feasible and that remote access of secret information resembling wireless skimming is possible. The performance of the channel however, has not previously been analyzed. In this article the performance of WCCA in terms of probability of error is presented. This allows a comparison of different choices of carrier frequency. The data collection necessary is also improved to allow pattern recognition techniques to be employed. This approach looks very promising when selecting multiple carrier frequencies to improve the performance of the covert channel.

1 Introduction

Combining the efforts of different fields, electromagnetic side channel attacks, covert channels and subversion, the wireless covert channel attack (WCCA), by Dyrkolbotn and Snekkenes [1] has been suggested. WCCA exploits subversive code hidden on all cards during manufacture, to launch an attack, without physical access, when infected cards are used. Experiments on modern smart cards confirmed that an insider with the opportunity to hide subversive code could potentially broadcast the card's internal secrets to a nearby receiver. However, except for a few estimates of the channel capacity, the performance of the covert channel has not been investigated. Performance in terms of probability of error (i.e. misinterpretation of symbol 0 and 1) on the communication channel can be used to compare implementation choices such as what carrier frequency to choose. The measurements used in [1] is average power density spectrums (PDS) of the emitted power from smart cards, obtained with a spectrum analyzer. Information about the actual waveform, as a function of time, is not available. The maximum signal to noise ratio, within a receivers bandwidth, is used to calculate Shannon's channel capacity in [1]. The probability of error however, is decided by the actual waveform transmitted. Even though the waveform as a function of time is not available, using theory from binary communication systems, we will show how to estimate the probability of error for WCCA

This paper was presented at the NIK-2007 conference; see http://www.nik.no/.

based on the average power density spectrums of the received waveform. The probability of error will then be used to evaluate the method of choosing the carrier frequency. We will also show the potential of using pattern recognition techniques to choose multiple carrier frequencies. This requires data to be captured as several individual PDS rather than one average PDS. Methods for choosing the optimal carrier frequencies (lowest possible probability of error) is work in progress.

In section two, the wireless covert channel attack (WCCA) is introduced. WCCA is modeled as a binary system and necessary equations to calculate the probability of error is presented. In section three, pattern recognition is introduced along with the revised experiment allowing this approach. In section 4 the results are presented. Finally a conclusion and future work is found.

2 The Wireless Covert Channel Attack

The wireless Covert Channel Attack (WCCA) by Dyrkolbotn and Snekkenes [1] relies on a highly skilled insider to undermine the security mechanisms by hiding subversive code in the smart card's software (SW). This is done during an early stage (design or compile) of its life cycle (figure 1) and will affect all cards produced.

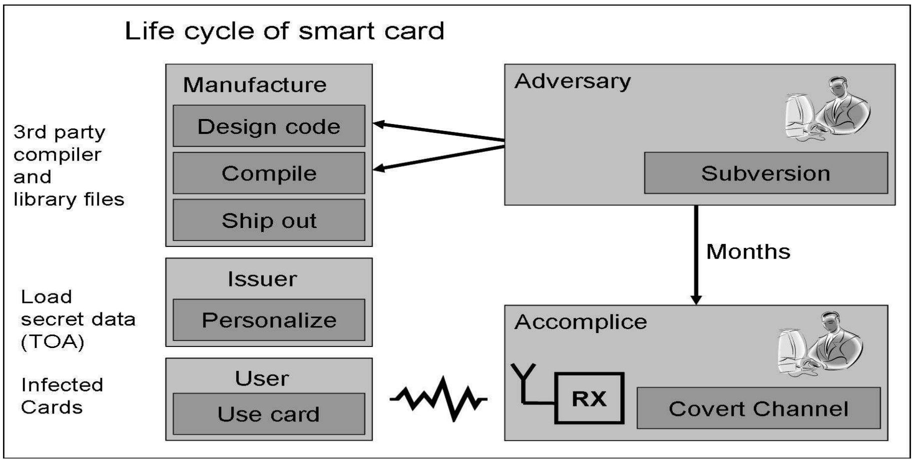

Figure 1: Scenario: The adversary hides subversive code on smart cards during an early stage the cards life cycle. Later, secret information is loaded to the card. When the infected card is used, the subversive code is activated and broadcasts secret information to the accomplice, over a covert communication channel.

These infected cards may be used e.g. in the banking industry as credit cards, loaded with personal information (cryptographic key, PIN code, account number etc.) when issued to a customer. The adversary is interested in retrieving the secret, personal information and has an accomplice involved at the use stage of the life cycle. The accomplice is typically somebody with access to a smart card terminal, e.g. a store owner or maintenance personnel. When a manipulated card is inserted into any terminal, by the owner, the subversive code exploits characteristic electromagnetic emanation (signatures) from the microprocessor, during execution of instructions, to broadcast secret information over a wireless covert channel. The success of this attack is ensured by the large number of cards infected. If a whole generation of smart cards to the banking industry is infected, there will be enough cards randomly used in the rigged locations to make the attack worth the effort. Experiments in [1] have shown that by manipulating the energy leakage from a smart card a covert channel can be created that will give access to secret information when the card is used and that the attack will work on modern smart cards equipped with countermeasures against side channel attacks.

The importance of considering the complete life cycle of a trustworthy device, with high level of security requirements, was also brought to the attention by Pankaj Rohatgi, [2] in an invited talk at CHES 2007. Work has also been done on detecting tampering with a product during manufacturing. Agrawal et al [3] approaches the problem of how to detect insertion of trojans when outsourcing part of your manufacturing. This work looks promising as a countermeasure to prevent the hiding of the subversive code necessary in WCCA.

Optimal Digital Receiver

WCCA can be modeled as a binary communication system. A basic binary system, as explained by Peebles [4], is shown figure 2. A message m is sent in each symbol interval, T_b. This message m can have of 2 possible values, $m = m_1$ if a binary 0 is transmitted and $m = m_2$ if a binary 1 is transmitted. The probability of sending each message is P_1 and P_2. The waveform actually transmitted is denoted $s(t)$ and $s(t) = s_1(t)$ if $m = m_1$ and $s(t) = s_2(t)$ if $m = m_2$. The input to the receiver, $r(t)$, is assumed to be the sum of $s(t)$ and an added noise $n(t)$. Based on measurements of $r(t)$ the receiver must make a decision on which message m_1 or m_2 that is received in each symbol interval.

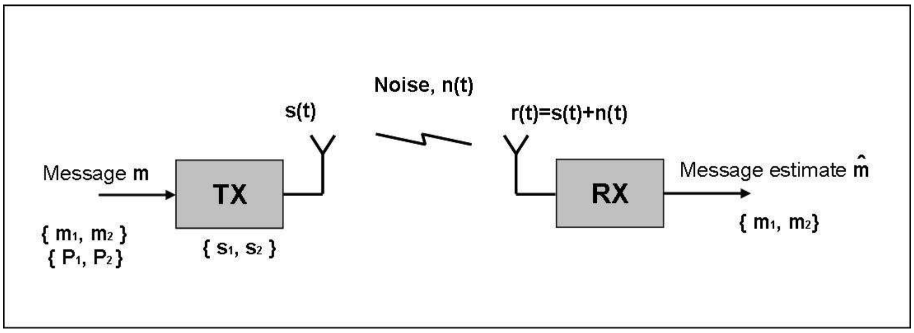

Figure 2: Basic Binary System

An optimal digital receiver, according to [4], is one that makes this decision based on maximizing the a posteriori probabilities and in this way minimized the probability of error. The decision rule can be written as:

$$if \quad \int_0^{T_b} r(t)[s_2(t) - s_1(t)]dt > V_T \quad choose\, m_2 \tag{1}$$

otherwise choose m_1, where the threshold value V_T is given by

$$V_T = \frac{E_2 + E_1}{2} + \frac{\mathcal{N}_0}{2}(\frac{P_1}{P_2}) \tag{2}$$

$\mathcal{N}_0$ is the white noise power density and E_1 and E_2 the energies of s_1 and s_2 at the receiver's input, given by

$$E_i = \int_0^{T_b} s_i^2(t)dt\ ,\ i = 1,2 \tag{3}$$

It is assumed that the receiver knows the form of s_1 and s_2 (coherent). The correlation receiver or matched filter are two possible implementation of the optimum digital system.

Waveform and carrier frequency selection

When designing WCCA, the waveform $s(t)$, and a carrier frequency, f_c, must be wisely chosen. Usually these parameters are design choices and depends on the communication system implemented. In WCCA the only design freedom is which activity (i.e instruction) that is activated on the microprocessor. The challenge is therefore to select two instructions, that will result in suitable waveforms $s_1(t)$ and $s_2(t)$ and choose the best possible carrier frequency for the given waveforms.The choice made in [1] was to look for one carrier frequency where the energy emitted could be turned on and off, by executing two different instructions.

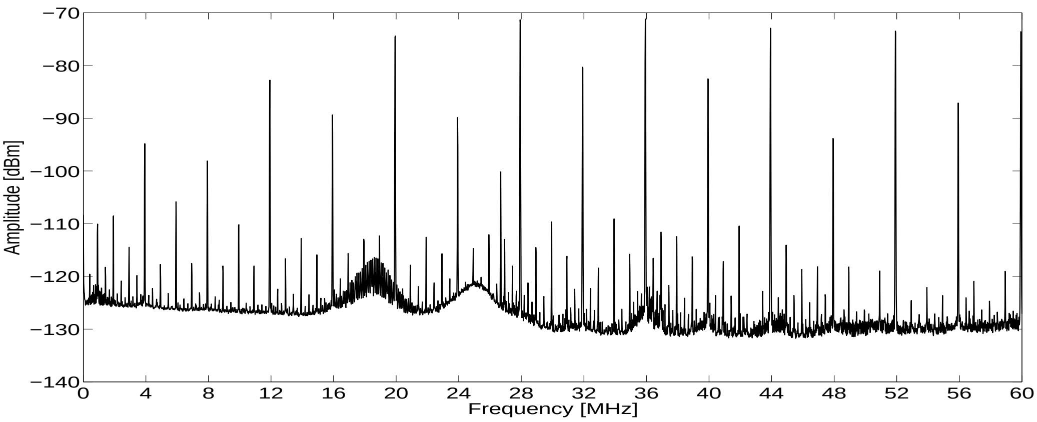

Figure 3: The electromagnetic signature of one instruction executed indefinite, represented as the average power density spectrum. 4206 sample points from DC to 60 MHz

Fundamental for choosing the instructions and the carrier frequency are recorded signatures of the electromagnetic energy emitted from the microprocessor. A signature, as defined in [1], is the average power spectrum densities (PDS) recorded from DC to 60 MHz when executing individual instructions, with fixed argument, in infinite loops, see figure 3. Averaging is done by the spectrum analyzer and variance of the individual amplitudes is therefore not available. Therefore, the carrier frequency of the covert channel in [1] is chosen as the frequency with maximum average amplitude difference between two signatures (distance of means). An exhaustive search is necessary to find which pair of instructions returns the largest distance of means. These two instructions form the symbol alphabet of the communication channel. For further detail refer to the WCCA article [1].

The obvious question know is: How to evaluate and compare the performance of the chosen symbol alphabet and carrier frequency? A common method in other communication system is to calculate the probability of error for the channel.

Performance - Probability of error

For the decision rule in equation 1, it can be shown that the probability of error for $P_1 = P_2 = 0.5$ can be written as:

$$P_e = \frac{1}{2} erfc\{\sqrt{\frac{E_1 + E_2 - 2\gamma\sqrt{E_1 E_2}}{4\mathcal{N}_0}}\} \tag{4}$$

where γ is the correlation between s_1 and s_2 and is given by:

$$\gamma = \frac{1}{\sqrt{E_1E_2}} \int_0^{T_b} s_1(t)s_2(t)dt \tag{5}$$

Calculating the probability of error using equation 4 requires that values for E_1, E_2, γ and $\mathcal{N}_0$ can be obtained from the available PDS. Values for E_i and $\mathcal{N}_0$ are found by taking the average of all amplitudes of the PDS within the receivers bandwidth. The waveforms are assumed to be uncorrelated such that $\gamma = 0$; The probability of error by this method is denoted P_{eopt}.

Further simplifications can be made by observing that choosing the carrier frequency according to the distance of mean method, usually finds a frequency where the PDS of $s_1(t)$ has a large peak and the PDS of $s_2(t)$ has a very low peak (close to noise). This basically turns the energy emitted at that frequency on and off. This is also the basic idea of Amplitude Shift Keying (ASK), also known as on-off keying. The probability of error for ASK is given by:

$$P_e = \frac{1}{2} erfc(\sqrt{\frac{E_1}{2\mathcal{N}_0}}) = \frac{1}{2} erfc(\sqrt{\frac{\varepsilon}{2}}) \tag{6}$$

where ε is the average energy per bit divided by 2 times the channel noise density. The probability of error by this method is denoted P_{eask}. Probability of error is usually plotted vs. ε for better comparison between systems.

The average PDS in [1] was good enough to show the feasibility of the attack and to find a carrier frequency by calculating the difference of means. A less naive approach, will require to take the variance of the individual amplitudes into consideration. This calls for a revised experiment and opens up for pattern recognition techniques, as explained next.

3 Pattern Recognition Approach

Pattern recognition is described by Duda et al. [5] as:

> The act of taking in raw data and making an action based on the "category" of the pattern

The design cycle for pattern recognition presented in [5] is illustrated in Figure 4. Each step in the design cycle, for the revised experiment, is introduced next.

Collect data

Instead of measuring the average PDS as in WCCA [1], several single traces are measured. The amplitude of each sample will then vary due to added noise in the channel. Even though an accurate noise model may be important, the first approach assumes gaussian distributed noise and that the noise at different samples are independent. Each sample can then be considered as a random variable with gaussian distribution $X \sim N(\mu, \sigma^2)$. The signatures used in the WCCA attack are simply the mean, μ, of all the single traces collected in this experiment. The activity considered is, as in WCCA, the execution of individual instructions, with fixed argument, in infinite loops. Terminology from pattern recognition has been adapted. The classes, ω_i, correspond to different activities on the microprocessor that we would like to classify. For this experiment 5 different instructions, and noise were measured.

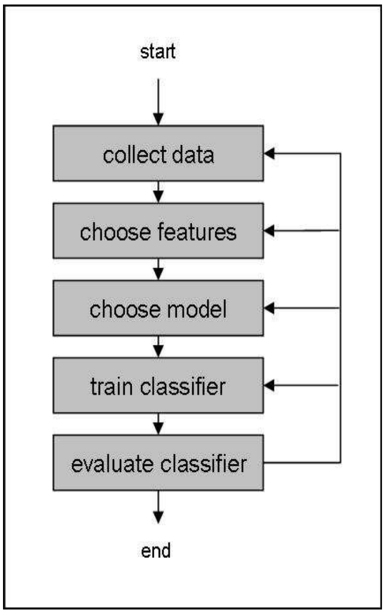

Figure 4: The design cycle of a pattern recognition system as described by Duda et al. [5]

$$\begin{array}{lll} class\ \omega_1: & bcf\ 03h,5 & switch\ to\ bank\ 1 \\ class\ \omega_2: & goto & empty\ loop \\ class\ \omega_3: & movlw\ 0xaa & move\ binary\ 10101010\ into\ w \\ class\ \omega_4: & nop & no\ operation \\ class\ \omega_5: & sublw\ 0xaa & subtract\ w\ from\ binary\ 10101010,\ store\ result\ in\ w \\ class\ \omega_6: & noise & no\ activity\ (power\ off) \end{array} \tag{7}$$

Each sample (i.e. frequency) of the measured traces is considered as a feature, such that each class ω_i is represented by $d = 3006$ features in a column vector $X_i = \{x_{i1}, x_{i2}, ..., x_{id}\}$. A total of $n = 440$ measurements was collected of each class. Each measurement resulted in a trace of 3006 features from DC to 60 MHz, as can be seen in figure 5. Comparing this trace to a trace from the original WCCA (figure 3), notice that the noise floor is about 20 dB higher and only 3006 samples are available. This is a limitation of the spectrum analyzer used recently (Rhode and Schwarz FSL-6)

Choose Features

A feature is the carrier frequency for the covert channel. How to choose the carrier frequency depends on the optimization criteria. Maximizing the range should look for the best signal to noise ratio, but this does not necessarily give the lowest probability of error. A large number of distance measures exist, see [6]. For the one dimensional case (one carrier frequency), considered here, it is computational feasible to calculate the performance of all 3006 possibilities. First the carrier frequency will be chosen by the, naive difference of means, approach of WCCA. The result will be compared to results from an exhaustive search for the lowest probability of error.

For the pattern recognition case, it is straight forward to extend to more than two instructions (M-ary symbol alphabet) and more than one carrier frequency. This comes at a complexity cost. The curse of dimensionality quickly becomes a reality and smarter methods for feature selection must be designed. This is work in progress.

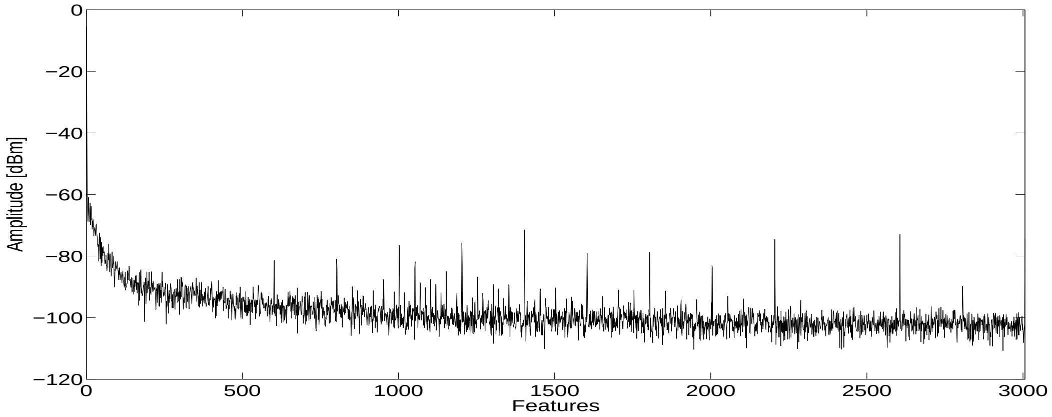

Figure 5: One trace of class 1 (bcf), represented by 3006 features. Each feature is the amplitude in dBm of frequencies from DC to 60 MHz

Choose Model

The model used is simply: Activity on a microprocessor can be classified based on energy emitted on a small finite set of frequencies. For the special case of WCCA this can be states as: Execution of two different instructions on a microprocessor can be classified based on the energy emitted on one frequency.

Train Classifier

In this approach Bayes classifier is used. Fundamental for the classification and the calculation of the decision boundaries are the a priori probabilities $P(\omega_i)$ and the conditional densities $p(x|\omega_i)$. The process of estimating $P(\omega_i)$ and $p(x|\omega_i)$, based on available data, is called training.

Bayes principle minimizes the probability of error by choosing the maximum a posteriori probability $P(\omega_i|x)$, related to a posteriori probability and density function by Bayes rule [5]. The case considered in this article is two classes and one feature. With 0/1 loss function (i.e wrong decision weighted zero, right decision weighted one) the decision rule can be written as:

$$if \quad p(x|\omega_1)P(\omega_1) > p(x|\omega_2)P(\omega_2) \quad choose\ \omega_1 \tag{8}$$

otherwise choose ω_2. Assuming the distribution of observations X to be gaussian, $X \sim N(\mu, \sigma^2)$ and a priori probabilities to be equal, P(ω_1)=P(ω_2), the decision rule reduces to:

$$if \quad \frac{(x-\mu_1)^2}{2\sigma_1^2} - \frac{(x-\mu_2)^2}{2\sigma_2^2} < \log\frac{\sigma_2}{\sigma_1} \quad choose\ \omega_1 \tag{9}$$

otherwise choose ω_2.

This is equivalent to, choosing ω_1 if $h(x) < 0$ and ω_2 otherwise, where $h(x)$ is given by:

$$\begin{array}{ll} h(x) = ax^2 + bx + c & where \\ a = \sigma_2^2 - \sigma_1^2 & second\ order\ (quadratic)\ term \\ b = 2(\mu_2\sigma_1^2 - \mu_1\sigma_2^2) & first\ order\ (linear)\ term \\ c = \sigma_2^2\mu_1^2 - \sigma_1^2\mu_2^2 - 2\sigma_1^2\sigma_2^2 \log\frac{\sigma_2}{\sigma_1} & constant \end{array} \tag{10}$$

The roots of $h(x) = 0$ defines our decision boundary between decision regions. Two roots means that the decision regions are not simply connected.

The goal of training is to take collected data of known classes and calculate a decision boundary that can be used on future measurements of unknown classes. Since only 440 traces are available, 220 traces are randomly taken out to be used for testing. Based on the remaining 220 data sets, the mean, μ and the variance σ^2 are estimated, using maximum likelihood estimators. Then the constants a, b and c of $h(x)$ are calculated according to (10) and stored. The decision boundary , and A posteriori densities $P(\omega_1|x)$ and $P(\omega_3|x)$ for class 1 and 3, and feature 1070 (i.e carrier frequency 21,3 MHz), using this method are shown in figure 6.

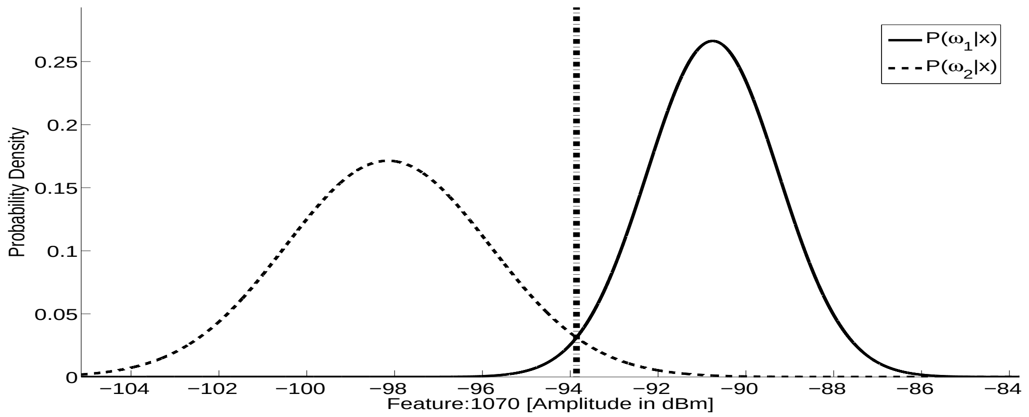

Figure 6: A posteriori probability densities for classes 1 and 3. Gaussian distribution assumed and maximum likelihood estimator used to estimate parameters. The decision line that minimizes the probability of error is located at the interception between the densities

Evaluate Classifier

For each trace reserved for testing, the set of amplitudes for the chosen feature (i.e carrier frequency) is used to calculate $h(x)$, using a, b and c from training. The trace is then classified according to $h(x) > 0$ or $h(x) < 0$. Since the correct class of every trace is known, the probability of error is simple the ratio of wrong classifications to the total number of test traces. The random split of available data does influence the result a little bit. The final probability of error, denoted P_{eBayes} is therefore the mean of repeating the splits, training and testings 10 times.

4 Results

WCCA Approach

First the carrier frequency is chosen according to the difference of means method suggested by [1]. The largest difference between any of the signatures was found between the goto instruction (class 2) and the nop instruction (class 4). The difference was 10,4 dB for carrier frequency $f_c = 23,3$ MHz. Since the resolution of the PDS's are about 19 kHz, a narrowband receiver with bandwidth of 15 kHz would receive one carrier frequency at the time. Under this assumption, using equation 4 and 6 gives the following results:

$s_1(t)$	$s_2(t)$	f_c [MHz]	P_{eopt}	P_{eask}
goto	nop	23,3	0.0078	0.0103

The probability of error for 50 carrier frequencies is shown in figure 7. The two curves indicate what probability of error that can be expected. The horizontal line indicates the probability of error of the carrier frequency, $f_c = 23,3$ MHz, chosen by the maximum difference of means method. The result shows, not surprising, that the method to choose carrier frequency in WCCA is not optimal in terms of low probability of error. Several of the 49 carrier frequencies ranked after 23,3 MHz return a lower probability of error. The lowest probability of error in Figure 7 is $P_{eopt} = 10^{-9}$.

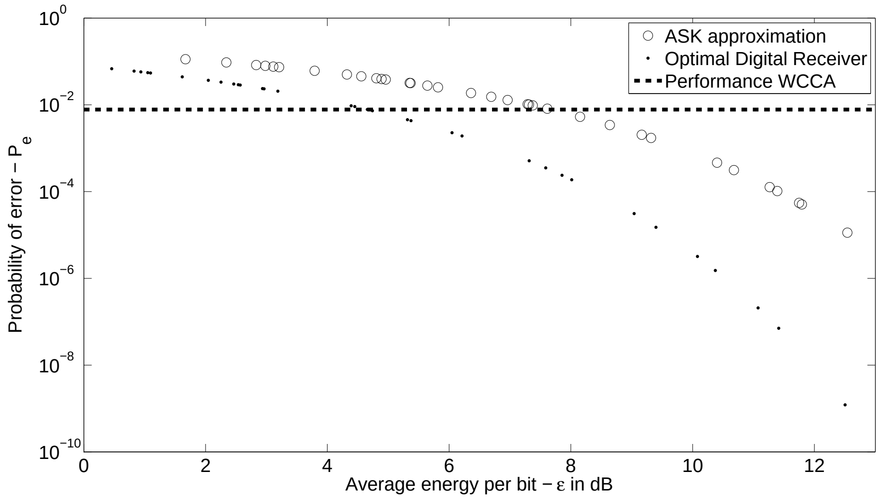

Figure 7: The probability of error calculated for the 50 carrier frequencies, using equation 6 for ASK approach, equation 4 for the optimal digital receiver approach. The horizontal line indicates the probability of error for the carrier frequency suggested by difference of means method

Pattern Recognition approach

The results in figure 7 clearly show that the approach in WCCA to choose the carrier frequency is not optimal in terms of probability of error. In the one dimensional case (i.e. one carrier frequency), it is computational feasible to run through all the possible carrier frequencies to choose the one returning the lowest probability of error. Doing this actually returns 9 frequencies with $P_{eBayes} = 0$ and 21 with $P_{eBayes} < 0.001$, all perform better than the frequency chosen by the WCCA method, which returned $P_{eBayes} = 0.0023$. The perfect classification by 9 features is probably due to a limited test set of 220 traces.

Feature 1404 (i.e. carrier frequency $f_c = 28,0$ MHz) is an excellent example of the difference between the two approaches. The amplitude difference between any classes for feature 1404 is less than 2 dB. The WCCA approach is looking at the amplitude difference of two classes in dB and will therefore dismiss feature 1404 as to low. It turn out though that there is a very small variance of the amplitude at feature 1404 such that for a limited set of 220 traces return $P_{eBayes} = 0$.

Higher dimensions

An obvious improvement of the attack would be to exploit more of the available energy by taking advantage of more than one carrier frequency and use a symbol alphabet larger than 2 instructions.

WCCA only uses one out of 3006 carrier frequencies. It should be possible to achieve better performance by utilizing more of the available energy emitted from the card. It is therefore natural to look at what happens if more than one carrier frequency is used. Pattern classification easily extends to several features. Figure 8 shows how two dimensions can be used to reduce P_e. Classifying classes 2 and 3 based on feature 1170 alone gives $P_{eBayes} = 0.2$. Using feature 1170 together with 1019 gives $P_{eBayes} = 0.02$. However, unless careful choices are made, the classification can be made worse by increasing the dimensionality. The question is therefore: What is the best way of choosing features? There is no trivial answer to this question and the challenge is work in progress.

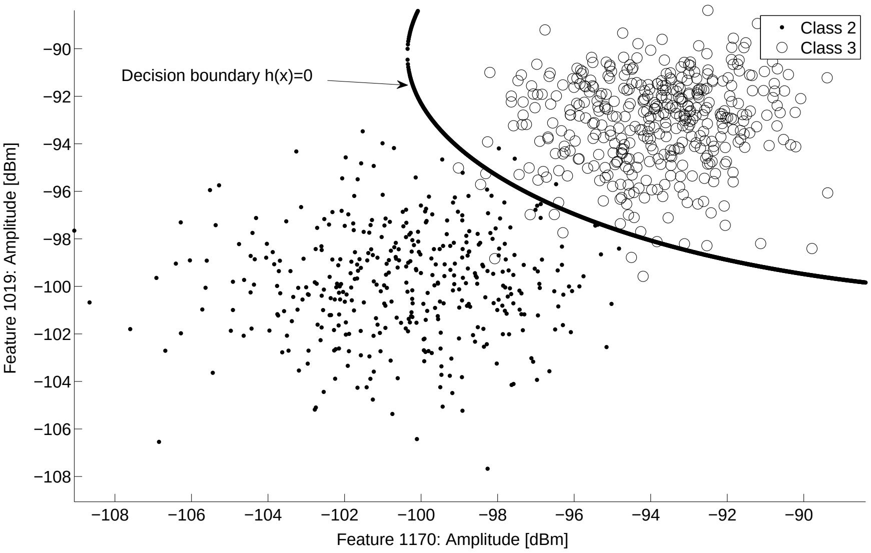

Figure 8: Two dimensional classification performs better if features are carefully chosen

Finally it is possible to extend the symbol alphabet. For a classifier this is straight forward. Preliminary results with Bayes classifier, using 5 instructions can be seen in the table below.

The results show that the probability of error can be made very small, by using enough carrier frequencies, if you select them correctly. As mentioned before, this is work in progress.

5 Instructions								
Feature	1404	1804	1153	1136	1019	1170	1604	1203
Number of features: d	1	2	3	4	5	6	7	8
Performance: d features	0,239	0,110	0,051	0,019	0,007	0,003	0,002	0,001

Table 1: Classifying 5 instructions based on an increasing number of features (i.e carrier frequencies)

5 Conclusion and Future Work

The wireless covert channel attack relies on proper choices of instructions and carrier frequency to implement a covert communication channel. The performance of the choices made, can be evaluated in terms of probability of error. Theory from a basic binary system has been used to calculate the probability of error, based on available average power density functions. The Results show that it is possible to achieve probability of error as low as 10^{-6}. It is also obvious, if not surprising, that just looking at the distance of means, without taking the variance into consideration is far from optimal. That said, when only one carrier frequency is used, it is feasible to do an exhaustive search for the optimal frequency. The performance of the channel can easily be improved, at the cost of complexity, by increasing the symbol alphabet and using more than one carrier frequency. A pattern recognition approach, as illustrated by Bayes classifier in this article, looks like a promising tool to explore this improvement. Work is already in progress on how to optimally (i.e. minimizing probability of error) choose the number of instructions to use and the number of carrier frequencies to use.

References

[1] G.O. Dyrkolbotn and E. Snekkenes. A wireless covert channel on smart cards (short paper). In *Information and Communications Security*, volume 4307 of *Lecture Notes in Computer Science*, pages 249–259. Springer Berlin / Heidelberg, 2006.

[2] P. Rohatgi. Trustworthy hardware. IBM T. J. Watson Research Center, CHES 2007.

[3] D. Agrawal, S. Baktir, D. Karakoyunlu, P. Rohatgi, and B. Sunar. Trojan detection using ic fingerprinting. In *IEEE Symposium on Security and Privacy*, pages 296–310, 2007.

[4] Jr. P.Z. Peebles. *Digital Communication Systems*. Prentice Hall, 1987.

[5] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification*. John Wiley and Sons, Inc, 2001.

[6] V. Perlibakas. Distance measure for pca-based face recognition. volume Volume 25 of *Pattern Recognition Letters*, pages Pages 711–724. Elsevier B.V., 19 April 2004.