



Reid, Jason F. and Gonzalez Nieto, Juan M. and Tang, Tee and Senadji, Bouchra (2007) *Detecting relay attacks with timing-based protocols*. In: Proceedings of the 2nd ACM Symposium on Information, Computer and Communications Security, 20-22 March 2007, Singapore.

© Copyright 2007 ACM

Detecting Relay Attacks with Timing-Based Protocols

Jason Reid*, Juan M. González Nieto*, Tee Tang⁺, Bouchra Senadji⁺

*Information Security Institute, ⁺ School of Engineering Systems
Queensland University of Technology
{jf.reid, j.gonzaleznieto, t.tang, b.senadji}@qut.edu.au

Abstract. Distance-bounding protocols have been proposed as a means of detecting relay attacks, also known as *mafia fraud*. In this paper we present the first symmetric key based distance-bounding protocol that is also resistant to so-called *terrorist fraud*, a variant of mafia fraud. Distance-bounding protocols require a communication channel that can exchange single bits with extremely low latency. This unconventional communication requirement has prompted Hancke and Kuhn to assert in a recent publication that ultra wide band (UWB) radio is necessary to achieve a useful distance-bounding resolution for RF security devices (contactless smart cards, RFID tags and the like). We analyse this assertion and present an alternative, novel communication approach that leverages the phenomena of side channel leakage to deliver a low latency channel. Our proposal is capable of detecting sophisticated relay attacks without resorting to the considerable expense and complexity of UWB radio. We present experimental results to support our arguments.

1 Introduction

Two recent publications [14, 9] have described practical, low cost relay attacks on ISO 14443 contactless smart cards, highlighting their vulnerability to this type of fraud. The relay attack is particularly insidious because it works without the need to circumvent any cryptographic security that may be in place. ISO 14443 cards have a short operating range of 10 cm [12]. There is an implicit assumption that if a reader is communicating with a card, that card must be within 10 cm of the reader. However this assumption may not be well founded because an attacker can simply relay the messages from the reader to a legitimate card that is far away and relay the card's responses back to the reader via interposed transceivers.

Distance-bounding protocols [2, 4, 16, 3, 10] have been proposed as a means of protecting against relay attacks. A distance-bounding protocol is an authentication protocol between a *prover* A and a *verifier* B , whereby B obtains corroborating evidence about A 's claimed identity and physical proximity at the time the protocol is run. Distance-bounding protocols can be thought of as traditional identification protocols enhanced with a distance-bounding mechanism. The former provides assurance as to the identity of the prover, while the latter allows the verifier to upper bound the distance which separates them. The distance between prover and verifier is upper-bounded by measuring the time intervals between challenges being sent and responses being received.



Fig. 1. Adversarial setting

The first distance-bounding protocol was proposed by Brands and Chaum [2] to thwart *mafia fraud* - the name for relay attacks against identification protocols first coined by Desmedt [5]. In a mafia fraud, the adversary consists of two parts: a rogue prover \bar{A} and a rogue verifier \bar{B} , sitting in between the real verifier B and prover A as shown in Figure 1. \bar{A} and \bar{B} simply relay the protocol messages between A and B . Hence what the adversary achieves is to fool B into thinking that he is directly communicating with A , when in

reality he is talking to \bar{A} . This attack does not violate the traditional security requirements of identification protocols, however it may be a concern if the verifier incorrectly makes assumptions as to the proximity of the prover. For example, consider the case where B is an RF reader enforcing access control through a door and A uses an RF proximity card to authenticate to B . A succesful mafia fraud attack would allow an adversary to open the door when A is sitting at a restaurant close by to \bar{B} who is stealthily running the identification protocol with A 's card and relaying all the information to \bar{A} , who is present near the door running the identification protocol with B using the messages received from \bar{B} . Brands and Chaum [2] described two protocols that are secure against mafia fraud attacks. The underlying identification protocols are a signature based challenge-response mechanism for one of them, and a zero-knowledge identification protocol for the other. These protocols use public key cryptographic operations, which are computationally demanding for highly resource constrained devices.

Desmedt [5] considered another type of active attack against identification protocols, which he called a *terrorist attack*. Here, unlike mafia fraud attacks where the prover is oblivious to the attack that is underway, the prover conspires with \bar{A} and \bar{B} to intentionally try to fool the verifier as to A 's location. Defending against terrorist attacks is more difficult, since A 's secret information (e.g. authentication keys) may be used in a manner which is different to what the protocol prescribes. Clearly, if A is prepared to release their secret authentication keys to \bar{A} , then the attack is trivially successful. When dealing with terrorist attacks, we preoccupied ourselves with attacks where the prover does not reveal to accomplices secret information that will allow the accomplices to impersonate A in more than a single run of the protocol. In particular, A does not reveal the long term private key. To the best of our knowledge, the only distance-bounding protocol that protects against terrorist fraud and does not require trusted functionality is the protocol of Bussard [3]. His solution is also public-key based and uses zero-knowledge techniques, making it impractical for implementation in low-cost RF computing devices. Other authors including Singl ee and Preneel [17] have noted that an efficient implementation is an open problem.

Recently, Hancke and Kuhn [10] have proposed a very efficient distance-bounding protocol which is secure against mafia fraud, but which does not protect against terrorist fraud. Hancke and Kuhn proposed the use of UWB radio to meet the demanding communications requirements of their distance-bounding protocol. The addition of UWB would add appreciable cost and complexity to contactless smart card integrated circuits, so it could only be justified in the absence of simpler, lower cost alternatives.

The contribution of this paper is twofold: firstly, we propose the first symmetric key based distance-bounding protocol which is resistant to terrorist fraud and is computationally efficient enough to be implemented in resource constrained devices; secondly, we propose a novel communication method for distance bounding based on the principle of *side channel leakage*, that has very low latency and which is comparatively inexpensive to implement.

Outline of the paper The rest of the paper is structured as follows. In Section 2 we review Hancke and Kuhn's protocol and explain why it is not resistant to terrorist fraud. In Section 3 we describe the new proposal which combines the efficiency of Hancke and Kuhn's protocol with the security of Bussard's protocol. In Section 5 we present an analysis of the communication channel requirements for distance bounding. This analysis highlights the importance of low communication latency, which is not purely a function of the channel bit rate. We propose the use of side channel information leakage as a low latency communication mechanism. In Section 6 we report on our current investigations into adapting the existing load modulation circuitry in proximity cards to our proposed communication technique. We present experimental results indicating that a modified load modulation scheme can provide sufficient timing resolution to detect sophisticated relay attacks launched by well funded attackers. Our proposed approach avoids the additional cost and complexity of adopting UWB radio. Details of the experimental method are presented in an appendix.

Notation The following notation is used in the rest of the paper.

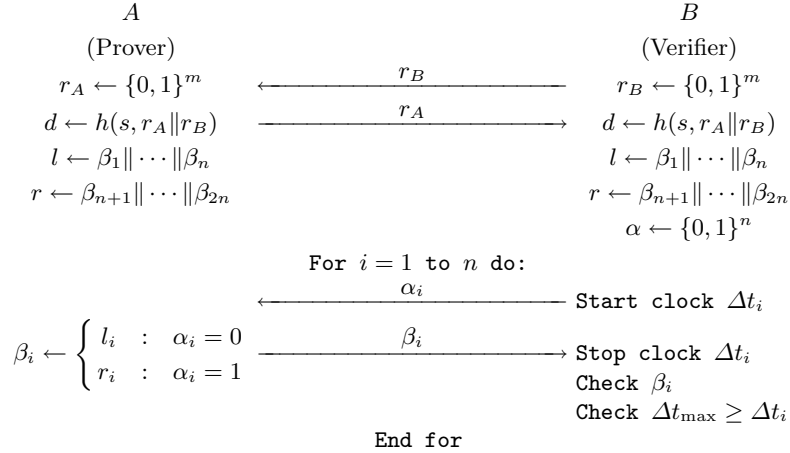
- We use \leftarrow to indicate assignment to a variable. If A is a set then $x \leftarrow A$ assigns to x a random element of A according to the uniform distribution.

- $\{0, 1\}^n$ denotes the set of all strings of bit-length n .
- Given a string s , we use s_i to denote the i^{th} least significant bit of s ;
- $time()$ is a function implemented at all parties that returns the internal clock time. To measure the time between two events we use two instructions, **Start clock** and **Stop clock**, such that **Start clock** Δt assigns $t_o = time()$ and **Stop clock** Δt , assigns $t_f = time()$ and $\Delta t = t_f - t_o$. Note that we do not require clocks at different parties to be synchronised.

2 Hancke and Kuhn’s distance-bounding protocol

Hancke and Kuhn’s [10] distance-bounding protocol is highly efficient. The protocol (see Protocol 1) is based on a symmetric-key identification mechanism, where the prover A and verifier B share a common secret value s . The distance is parametrised by the maximum challenge-response delay allowed, Δt_{\max} . The protocol starts by having A and B exchange random nonces r_A and r_B . The prover then applies a keyed hash function h to the concatenation of the nonces $r_A \| r_B$ to get d . The prover splits d into two n -bit strings l and r . A fast n -round challenge-response phase begins then. At each round, B sends challenge bit α_i , to which A must respond with the i^{th} bit of l if $\alpha_i = 0$, and the i^{th} bit of r if $\alpha_i = 1$. The verifier checks that the received response is correct. (He can do so, since he can also compute l and r .) Additionally, B measures the time Δt_i elapsed between challenge and response. B makes sure that all delays Δt_i are less than the bound Δt_{\max} . If all checks are succesful, B outputs **accept**, otherwise B outputs **reject**.

Shared Information: Secret key s .



Protocol 1: Hancke and Kuhn’s distance bounding protocol [10].

If B accepts, assuming that information cannot travel faster than the speed of light c , then the distance between A and B is upper-bounded¹ by $c\Delta t_{\max}/2$. Hancke and Kuhn [10] showed that the probability that a mafia fraud attacker can make B falsely accept is bounded by $(\frac{3}{4})^n$.

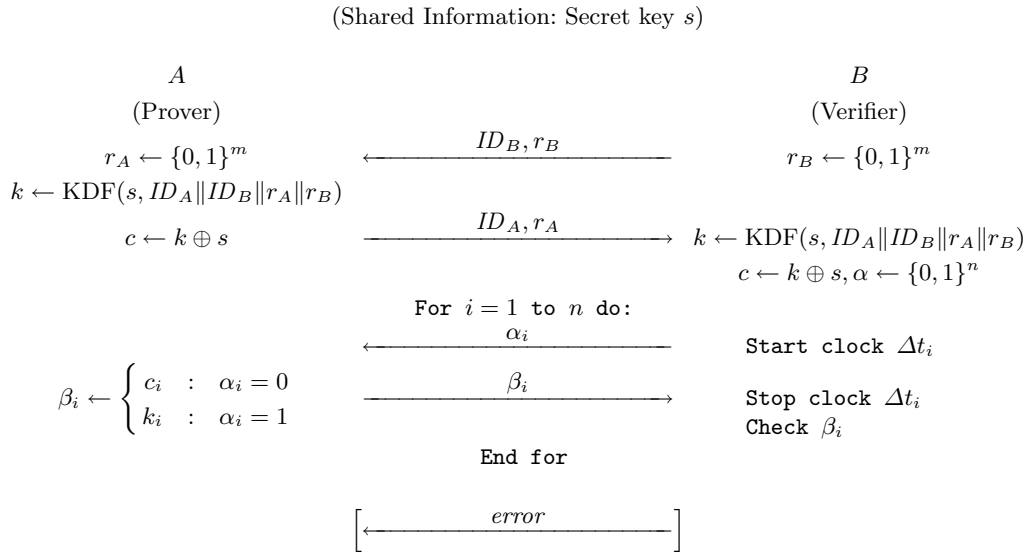
Protocol 1 is not secure against terrorist fraud attacks. A remote A can always relay r and l to a rogue prover \bar{A} who is close by to B . Note that the time-critical phase does not start until B sends the first challenge bit, and that releasing r and l does not compromise the long-term secret s .

¹ A better bound can be obtained when we know the time Δt_p that it takes for A to process a challenge. In this case, the distance is bounded by $c(\Delta t_{\max} - \Delta t_p)/2$.

The only distance-bounding protocol that protects against terrorist fraud attacks is to our knowledge the protocol of Bussard [3]. His solution is public-key based and uses zero-knowledge techniques, which makes it computationally expensive, especially for implementation in low-cost RF computing devices, such as RFID tags and proximity cards. The basic idea of Bussard's protocol is to force the prover to give away their private key in order to mount a terrorist attack. The prover computes $c = \mathcal{E}_k(sk_A)$, the encryption of her long-term (important) private key sk_A under a newly generated session key k . The verifier then sends challenge bits α_i to the prover. If $\alpha_i = 1$, the prover must respond with the bit c_i from the ciphertext. If $\alpha_i = 0$, the prover returns the bit k_i of the session key. Thus, in order to successfully and timely reply to the challenges the prover must be in possession of c and k , and therefore of sk_A .

3 New distance-bounding protocol

Bussard's protocol [3] protects against terrorist fraud attacks, but its use of asymmetric techniques makes it computationally demanding. On the other hand, the more computationally efficient protocols published, based on symmetric key authentication, do not afford terrorist fraud resistance. In this section, we propose the first symmetric key based distance-bounding protocol which is resistant to terrorist fraud attacks and is efficient enough for implementation in low cost devices. We enhance the protocol of Hancke and Kuhn [10], which is to our knowledge the most efficient mafia-fraud resistant protocol, by applying the basic idea behind the terrorist fraud resistance of Bussard's protocol. The result is shown as Protocol 2. The efficiency of the new protocol remains practically unchanged with respect to Hancke and Kuhn's [10], the main difference being the addition of a symmetric encryption (in practice, an XOR operation as discussed below).



Protocol 2: Distance bounding protocol resistant against terrorist attacks

In Protocol 2, we have made explicit the identities of prover and verifier by adding them in the initial exchange of nonces. Both A and B now use a key derivation function KDF to derive a symmetric encryption key k , which is used to encrypt the long-term shared secret s using a one-time pad. In practice f can be a message authentication code (MAC) algorithm such as CBC-MAC or HMAC [6]. The fast challenge-response phase of the protocol is similar to Hancke and Kuhn's, except that now the i^{th} bit of the ciphertext c is returned when $\alpha_i = 0$ and the i^{th} bit of the key k otherwise. Note that knowledge of k and c is equivalent

to knowing the shared secret s , since $s = k \oplus c.B$ checks the correctness of each response. If any response⁵ bit is incorrect, B will send an extra error message to A . This additional message indicates which responses were incorrect –this is required to protect against terrorist attacks as discussed below.

Noise errors In practice, as discussed by Hancke and Kuhn [10] and further elaborated in Section 5, the communications link between prover and verifier during the fast challenge-response phase is unreliable. This means that the protocol should tolerate transmission errors during that phase, by increasing the number of challenge-reponse rounds according to the expected error rate. We refer the reader to Hancke and Kuhn’s paper [10] for the quantitative analysis.

4 Security of new protocol

Here, we informally analyse Protocol 2 with respect to security against mafia and terrorist fraud. We note that it is still an open problem to provide formal definitions of security against relay attacks.

4.1 Mafia fraud

Firstly we show that the new protocol is secure against mafia fraud. The adversarial setting corresponds with the one depicted in Figure 1. A does not cooperate with the attackers \bar{A} and \bar{B} , and A is not close to B , which implies that it is not physically possible for \bar{A} and \bar{B} to pass on the challenge to A , get the response from A and relay it back to B in time.

Since f is pseudo-random, the one-time pad encryption c is also pseudo-random. This implies that it is impossible for any adversary to guess any bit k_i or c_i with probability non-negligibly different from $1/2$. Hence the best \bar{A} and \bar{B} can do is guessing the challenge bit before it is output by B and send it to A . This could be done before the challenge-response phase starts. For example, \bar{B} withholds message 2 for a time long enough to allow him to complete a run of the protocol with A , using challenges $\bar{\alpha}_i$ chosen by \bar{B} himself. \bar{B} then passes to \bar{A} the value r_A , the challenges $\bar{\alpha}_1, \dots, \bar{\alpha}_n$, and the responses $\bar{\beta}_1, \dots, \bar{\beta}_n$. \bar{A} then completes the protocol with B . Since B chooses the challenges α_i uniformly at random, on average only half of the challenges $\bar{\alpha}_i$ will coincide. When this happens, \bar{A} can send the valid response $\bar{\beta}_i$; otherwise, \bar{A} can only guess the right reponse with a probability of $1/2$. Overall, the probability that \bar{A} and \bar{B} fool the verifier into accepting is essentially $(3/4)^n$, which is negligible.

4.2 Terrorist fraud

To see that the new protocol protects against against terrorist fraud, we argue that A must release significant information about s to the accomplice \bar{A} in order to have B accepting. Firstly notice that, in order to have B accepting, someone close to B must have the right challenged bits of k and c , which only A and B can compute. Lets consider how A can help \bar{A} to make B accept in a protocol run.

Given that A is not close by, for each challenge bit α_i , A will have to either guess it in advance and pass the corresponding response to \bar{A} , or alternatively A can pass both the right c_i and k_i (and therefore $s_i = k_i \oplus c_i$). In general, let (c'_i, k'_i) for $i = 1, \dots, n$ be the values passed by A to \bar{A} , where $s'_i = k'_i \oplus c'_i$ may or may not be equal to s_i . Since A is not close by, (c'_i, k'_i) must be independent of α_i . Clearly, if (c'_i, k'_i) has no information about A ’s secrets, i.e. $Pr[s_i = k'_i \oplus c'_i] = 1/2$, then from the above discussion the probability that B accepts is at most $(3/4)^n$. Otherwise, assume that there are m instances for which $c'_i \neq c_i$ or $k'_i \neq k_i$ (but not both), i.e. in this case, A only discloses partial information about the secret s . Then the probability that B succeeds is 2^{-m} . If \bar{A} computes $s'_i = c'_i \oplus k'_i$, then m bits of the computed secret will be flipped with respect to the real secret s . \bar{A} can guess the position of these bits with probability $\binom{n}{m}^{-1}$, which is much smaller than the probability of success. If, for example, $n = 128$ and $m = 10$, then the probability that B accepts is 2^{-10} and the probability that \bar{A} guesses the correct s by flipping m bits of s' is less than 2^{-40} . However, notice that when B does not accept, B informs \bar{A} of the responses which were incorrect. Thus, for

Each β_i that was incorrect, \bar{A} can compute the i^{th} bit of the secret s ; so on average \bar{A} learns $m/2$ bits of the secret s from each protocol run. In other words, the probability that B accepts, 2^{-m} , is exactly the same that the probability that \bar{A} learns all of the bits of s .

5 Communications requirements for distance bounding

In this section we analyse requirements and implementation issues associated with the communications channel used in the time critical phase. We identify communication architecture vulnerabilities that may allow an attacker to indirectly defeat a distance-bounding protocol. With these potential pitfalls in mind, we propose a novel communication approach that exploits the underlying principle of *side channel leakage*, heretofore regarded as a security weakness, in a constructive way to provide the necessary distance-bounding resolution for constrained devices (particularly, ISO 14443 contactless smart cards).

The communication requirements for distance-bounding protocols are both demanding and unconventional - to achieve useful distance resolution they require extremely low communication latency but they do not require a correspondingly high bit rate since they exchange single bits punctuated by relatively large processing delay. Moreover, the processing overhead and variable delay associated with a reliable communication channel is unacceptable because reliability mechanisms introduce overheads; more bits need to be exchanged but more importantly, an additional and possibly variable number of processing cycles are required for a reliable channel.

The prover must calculate each response bit in a very small number of clock cycles if a distance bounding protocol is to be effective. Total round trip time for a challenge-response round comprises processing time and propagation time. For the verifier to accurately isolate the propagation component and thereby calculate the prover's distance, the processing time must be known and fixed. Consider an attacker who wants to launch a relay attack against a contactless smart card system in the style of Hancke [9] or Kfir [14]. If the system implements a timing-based relay attack detection protocol, the attacker must absorb the delay that is introduced by the relay so that the round trip time falls in the range that the verifier will accept. They may be able to do this if they can accelerate the response calculation sufficiently. The more processing cycles there are, the greater the opportunity to accelerate and thereby absorb the delay. We therefore need to consider whether the assumption of a fixed processing time is reasonable.

There are two main approaches to stop an attacker from operating a smart card at a higher than intended frequency; phase locked loop (PLL) internal clock generators and high frequency filters. Internal PLL-based clock signal generators are an increasingly popular choice among manufacturers, particularly in microprocessor cards because the frequency of the generated signal is independent of the reader-supplied frequency. This provides very effective control of the processing speed. Where the card uses a reader-supplied clock signal, overclocking protection is commonly provided by a low pass filter which resets the card when the filter threshold is exceeded. Tolerances of the order of a few percent are possible. It is therefore reasonable to assume that for appropriately designed hardware, an attacker can be limited to overclocking by no more than a few percent. For an ISO 14443 contactless smart card [12] clocked at the reader supplied frequency of 13.56 MHz, 2 % overclocking absorbs 1.5 ns of delay per clock cycle. A signal will propagate 45 cm in this time so it is clearly important to keep the number of clock cycles small.

5.1 Timing Resolution for relay attack detection

What timing resolution is required to detect relay attacks? Realistic attack scenarios on contactless smart cards may involve relay distances as small as a few metres since legitimate cards are often found in the close vicinity of a card reader. The round trip signal propagation time for a relay attack of 3 meters is 20 ns so at first glance it would appear that the ability to detect delays of the order of 20 ns is necessary. This is a technically demanding requirement. ISO 14443 contactless smart cards support a base communication rate of 106 kbits/s. At this bit rate a signal propagates 2.8 km in one bit period and Hancke and Kuhn [10] have argued that this is inadequate for distance bounding, thus motivating their proposal of UWB radio. However, we believe it is possible to achieve significantly finer distance resolution than 2.8 km without resorting to

the additional complexity of UWB. It is crucial to recognise that relay attacks introduce unavoidable delays beyond the signal propagation time. The rogue relay devices themselves each incur a circuit propagation delay known as *group delay* - the amount of time that the amplitude modulated signal is delayed by its passage through a device [13]. The relayed signal needs to pass through \bar{A} and \bar{B} in two directions so there is an additional delay component equal to four times the group delay². For example, Hancke's relay attack introduces a total delay of between 15,000 and 20,000 ns (15-20 μ s) where the round trip propagation time is only 333 ns. Thus the individual device group delay is of the order of 4 to 5 μ s. Hancke's attack was a proof of concept that used inexpensive RF relay equipment. Advanced microwave transceivers operating at gigahertz frequencies can have group delay in the order of tens of nanoseconds [1]. This type of equipment is typically found in signals intelligence applications and is both expensive and exotic³. More commonly available and less costly equipment will have a significantly higher group delay and will therefore be easier to detect.

For the purpose of our analysis we estimate a lower bound for total group delay introduced by the relay devices of 40 ns for well funded attackers. This figure will be much higher when standard, off-the-shelf equipment is used. Thus, when the signal propagation delay is included we assume the relay detection protocol must detect a minimum of 60 ns of total introduced delay for a 3 m attack.

5.2 Timing resolution for contactless card communication

In this section we analyse the timing resolution that a contactless smart card can support. We begin with a brief explanation of the card to reader communication method for ISO 14443 cards.

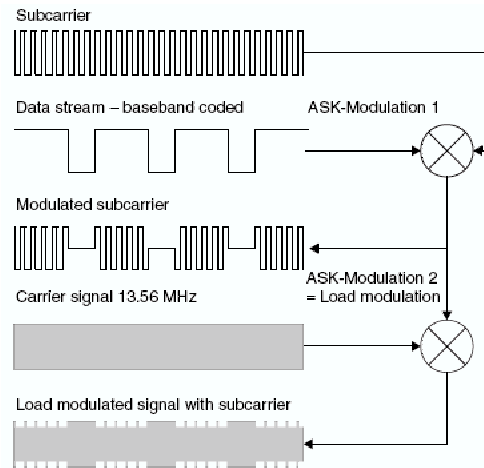


Fig. 2. Load modulation with a modulated subcarrier [7]

ISO 14443 contactless smart cards communicate with the reader at a base rate of 106 Kbits/s via load modulation. A resistor in the card's power supply circuit is switched in and out of circuit in time with the data to be transmitted according to the *subcarrier* modulation method shown in Figure 2. When the

² Kfir [14] reports simulation results indicating that a rogue verifier, \bar{B} may directly communicate with the reader B over a distance of up to 50m by directly modulating the RF sidebands. Thus it may be possible to avoid the last group delay incurred in \bar{A} for short range attacks though this is unclear since in our proposal we do not use side band modulation.

³ The Macom SMR-4820 Compact Microwave Search Receiver claims a group delay of <15ns for 10MHz output bandwidth. See <http://www.macom.com/sigint/PDF/4820.pdf>

⁸resistor is switched in, the card consumes more power and this increased consumption can be sensed as an amplitude change (as measured in the reader’s antenna circuit) in the 13.56 MHz carrier (f_c). The subcarrier generates two sidebands that contain the coded data. The sidebands appear 847 kHz either side of f_c . Since they are a reasonable spectral distance away from the main carrier frequency, a simple filter in the reader’s receive circuit can remove f_c thereby isolating one of the sidebands. Sideband modulation has better error performance in noisy environments at a cost of reduced usable bandwidth.

Hancke [10] notes that the distance resolution of a channel of bandwidth B is ‘roughly’ equal to c/B where c is the speed of light. According to this formula, the resolution is the propagation distance in one bit period which for contactless cards is two orders of magnitude coarser than we require. However, to clarify some of the roughness we will examine this further: the distance resolution is actually a function of the timing resolution - the verifier stops the clock when the bit is received. But distance-bounding protocols only exchange a single bit so the timing resolution is determined by whether the prover can *start* modulating the bit at an arbitrary point in time⁴. To make this distinction more concrete, consider that the bit period for ISO 14443 cards is 128 carrier cycles. However, because the prover is responding with a single bit sent on a clear channel, it is not necessary to wait for an arbitrary bit boundary (which occurs only once every 128 cycles) to be able to start modulating. At the instant the response is calculated the prover has the theoretical ability to *start* producing modulation peaks on the main 13.56 MHz carrier. Once chosen, the starting modulation cycle effectively determines the ending cycle which fixes when the verifier stops the timer. With some important qualifications which we will soon discuss, the timing resolution is closer to $1/f_c$ (74 ns or 22 m propagation). For ISO 14443 cards, this is a significantly finer resolution than the bit period alone would suggest (2.8 km).

This does not mean that usable timing resolution is purely a function of f_c and independent of the bit rate. Recall that an attacker avoids detection by absorbing introduced delay. If the bit period is sufficiently long with many redundant carrier cycles per bit, the attacker can begin modulating a guess of the response bit to the verifier B at the expected time via the rogue prover \bar{A} . If the delay introduced by the relay devices is not too great, then at some point part way through the modulation of the guessed bit the attacker will learn (via the rogue verifier \bar{B} which relays the signals) the correct value by monitoring the response of the real card. If the guess is wrong, the part-modulated bit can be switched to the correct modulation pattern. If this switch occurs sufficiently early in the bit period, the inconsistency will appear as noise and the new value will be accepted. To avoid this, we need to adopt a modulation scheme that is not susceptible to having a modulated bit changed part way through the bit period. The most effective way to do this is to reduce the number of carrier cycles that represent a bit. This increases the channel bandwidth but also the susceptibility to bit errors through channel noise.

We have already noted that the timing resolution for a modulated channel is *potentially* $1/f_c$ and that for ISO 14443 cards at 13.56 MHz, this represents a timing resolution of 74 ns. The qualification is necessary because the resolution depends on how quickly the card can increase the load on the reader’s antenna circuit to produce a detectable carrier amplitude change. This depends on the quality of the inductive coupling between the card and reader antenna loops. The coupling quality degrades as the distance increases and it takes more cycles to produce a detectable modulation change. We have determined through experimental observation (see Appendix A) that at closer distances, a card can produce an easily detectable modulation peak within a half cycle. This means that the timing resolution can be as low as $1/2f_c$ or 37 ns, the period between successive carrier peaks. This is smaller than the 60 ns lower limit required to detect close range attacks by well funded attackers. Therefore, if the number of processing cycles that the prover requires to receive the verifier’s challenge bit and compute and modulate the response bit can be kept small, it may not be necessary to resort to UWB communication as Hanke and Kuhn have asserted. However, a traditional layered communication stack requires too many processing cycles thus motivating our proposal.

⁴ In the discussion that follows we assume that the prover’s distance-bounding logic has direct access to the physical layer to modulate its response as a single bit. This cannot be done within the existing ISO 14443 standard for a number of reasons including bit alignment constraints associated with the frame delay time (FDT). In examining the basic principles of timing resolution, we ignore this restriction in ISO 14443.

5.3 A new approach to low latency communication

A two percent acceleration in processing speed over just 40 clock cycles at 13.56 MHz is enough to absorb the targeted lower bound of 60 ns of introduced delay that we identified in Section 5.1. If attackers can absorb this much delay they will be able to perpetrate undetected short range attacks. Therefore, the number of clock cycles required to calculate the response must clearly be very small. To address this problem, we propose a new low-latency approach to communication. The essential element of our proposal is that the verifier senses a physical side effect of the calculation process and from this, infers the result. The general principle that underlies this approach is the same one that underlies simple side channel analysis (SSCA) attacks (see for example [15]). However, communications channels of this type have previously been thought of as a serious security vulnerability, not purposely optimised and used to achieve legitimate protocol goals. A description of the technique in the context of ISO 14443 contactless smart cards follows, though it is important to note that it is equally applicable to contact cards and other devices that emit a side channel that the verifier can monitor. Indeed, we conceived the idea as a possible way of detecting network based card sharing attacks on satellite TV systems (see [8] for an attack description).

The proposed communication technique ‘leaks’ the response bit to the verifier via a side channel. SSCA is based on the following observation: computation in a microprocessor is the result of the physical process of electrons moving through semiconductor gates. This process takes a finite amount of time and consumes a measurable amount of energy. Such measurable phenomena are known as *side channel leakage*. Careful analysis of these phenomena can reveal detailed information about the internal state of the device [15]. The compelling advantage of communicating through side channel leakage is the dramatic reduction in latency - the verifier can detect the response bit as the prover calculates it, effectively being able to watch the prover in real time as it ‘thinks’. This eliminates the variable delays that would be introduced if the response bit had to traverse a layered communications stack.

To implement this approach the card needs to implement a special hardware-level instruction to calculate the response bits. The instruction should be engineered to have deliberately pronounced, output dependent leakage characteristics. One way to achieve this⁵ is by having the instruction conditionally switch a resistor into the supply circuit to increase the power consumption only if the result is for example, binary one. The load switching should be as integral to the instruction as the calculation itself, ideally occurring in the same clock cycle that the result is calculated. The key advantage of the proposed approach over the single published alternative (Hancke and Kuhn [10]) is the simplicity and modest implementation cost of the required circuitry in both card and reader. Hancke’s proposal relies on the addition of UWB radio circuitry which would increase the size of the card’s integrated circuit, (affecting its reliability) and impose appreciable cost to an extremely cost sensitive product.

Our proposal assumes that the verifier knows the small, fixed number of cycles the card requires to compute its response, allowing it to sample the amplitude of the carrier in a window starting at the nominated cycle, to detect zero or one by the absence or presence of a peak on f_c . Reader to card communication of the challenge bit uses the same approach. The card samples the carrier amplitude a fixed, short number of cycles after it detects the reader’s synchronisation marker. Note that we are proposing replacement of side band modulation with simple peak detection. Relay attacks are detected because the introduced delay will push the modulation peak onto a later cycle.

6 Experimental results

We have experimentally estimated the timing resolution that is attainable with current smart card technology by measuring the rate at which a range of cards effect amplitude changes in the reader-supplied carrier. The results are presented in Figure 3. Details of the experimental method and rationale can be found in Appendix A.

⁵ It may be more effective to sharply reduce power consumption, although we have not been able to determine this experimentally.

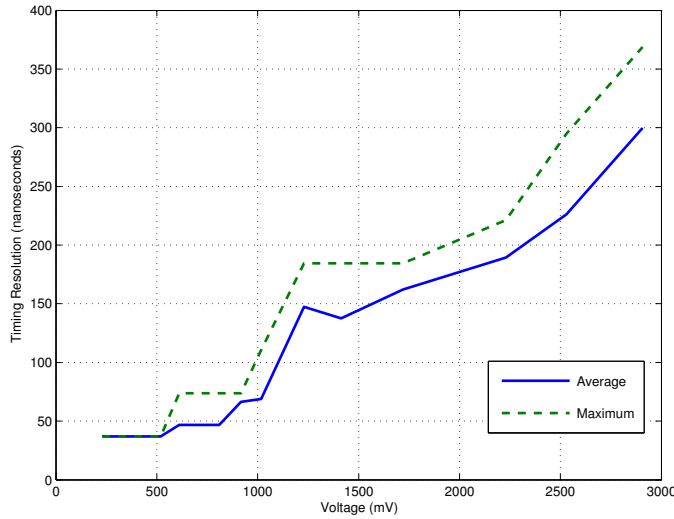


Fig. 3. Timing resolution using proposed method.

Timing resolution is presented as a function of measured antenna voltage which is itself a function of the card’s distance and orientation to the reader antenna - larger distances produce higher voltages. We found that the rate of amplitude change is a function of this voltage, irrespective of the orientation and distance that produced it so we plot timing resolution against voltage (since the combination of distance and orientation are difficult to specify with precision).

The resolution up to 800mV is conservatively sufficient to achieve our 60 ns detection target. The average timing resolution at higher voltages increases to 300 ns which is still 50 times smaller than the delay introduced by Hancke’s proof of concept attack [9]. Attaining sub-60 ns resolution places a significant restriction on the card to reader operating distance. With our reader, the card needs to be within a few millimeters to operate in this voltage range. In practice, this would mean that the user would need to touch the card on the reader. This reduction in operating range is clearly a disadvantage though it is worth noting that such fine resolution is only required to detect short range attacks using sophisticated and expensive relay equipment. Presumably, a large payoff would be required to motivate an attacker to go to such effort and expense, perhaps larger value contactless payment applications or high security physical access control. In these higher risk application scenarios it seems quite sensible to make the act of using the card more overt and deliberate by requiring very close proximity operation. With our reader, 300 ns resolution was attained in the 4-5 cm range. Better distance performance may be possible and further investigation is required to assess usable operating range. We note that the maximum distance for reliable operation that our reader could support was 7.5cm, 25% less than the 10 cm required by the standard. We suspect that with careful reader engineering, better distance performance will be possible.

We do not claim that these results prove the viability of the proposal, however they do indicate grounds for some confidence. Further investigation is needed into the impact of different RF noise environments on modulation detection. While our experiments were carried out in a busy electrical engineering laboratory, this does not characterise the broad range of possible deployment scenarios. Data needs to be gathered for a much larger range of readers and cards. Though we have not investigated it, we suspect that ISO 14443 type B cards may have better modulation performance at larger distances. Type B does not use 100% amplitude shift keying (ASK) in reader to card communications as type A does. Type A cards need a store of energy to continue operating during the short periods of carrier suspension. The availability of this reserve of power reduces the load that the card can apply to the antenna circuit to modulate the carrier. Interestingly, the

smart card manufacturer Infineon has recently applied for a patent [11] on a decoupling circuit that makes¹¹ this power reservoir unavailable when the card is modulating. They claim that the circuit improves the modulation performance at larger distances. This innovation may further improve the operating range and effective timing resolution for our proposed method.

7 Conclusion

We have proposed the first symmetric key based distance-bounding protocol that is resistant to so called terrorist fraud. In contrast to previous proposals the protocol is appropriate for implementation in resource constrained devices due to its computational efficiency. Unfortunately, our discussion of the security is informal. Providing a formal definition of security against relay attacks is still an open problem.

We have analysed the unconventional requirements that distance-bounding protocols place on the communication channel used in the time critical phase, highlighting the importance of low latency as opposed to raw bit rate. In response to these requirements, we proposed a novel approach to communication that leverages the phenomena of side channel leakage, heretofore considered exclusively as a security vulnerability. We exploit the extremely low latency of side channel leakage to address the requirements of distance-bounding protocols.

We presented experimental results indicating that a modified form of load modulation, used in the style of our proposed side channel leakage communication technique, can provide sufficient distance resolution to detect advanced relay attacks on ISO 14443 smart cards. Our technique has the disadvantage of reducing the operating range of the smart card. We have argued on the basis of these results, that it may not be necessary to incur the additional expense and complexity of implementing UWB radio for the distance-bounding communication channel, as Hancke and Kuhn [10] have asserted.

References

1. David Ballo. Measuring absolute group delay of multistage converters. In *Proceedings of 33rd European Microwave Conference*, volume 1, pages 89–92. IEEE, 2003.
2. S. Brands and D. Chaum. Distance-bounding protocols. In *Advances in Cryptology EUROCRYPT '93*, LNCS 765, pages 344–359. Springer-Verlag, 1993.
3. L. Bussard. *Trust Establishment Protocols for Communicating Devices*. PhD thesis, Institut Eurécom, Télécom, Paris, 2004.
4. Srdjan Capkun, Levente Buttyan, and Jean-Pierre Hubaux. Sector: secure tracking of node encounters in multi-hop wireless networks. In *SASN '03: Proceedings of the 1st ACM workshop on Security of ad hoc and sensor networks*, pages 21–32, New York, NY, USA, 2003. ACM Press.
5. Yvo Desmedt. Major security problems with the ‘unforgeable’ (Feige)-Fiat-Shamir proofs of identity and how to overcome them. In *SecuriCom '88*, pages 15–17, SEDEP Paris, France, 1988.
6. Y. Dodis, R. Genaro, J. Håstad, H. Krawczyk, and T. Rabin. Randomness extraction and key derivation using the CBC, cascade and HMAC modes. In *Advances in Cryptology CRYPTO 2004*, LNCS 3152, pages 344–359. Springer-Verlag, 2004.
7. Klaus Finkenzeller. *RFID Handbook*. John Wiley and Sons, Hoboken, NJ, 2nd edition, 2003.
8. Lishoy Francis, William G. Sirett, Keith Mayes, and Konstantinos Markantonakis. Countermeasures for attacks on satellite TV cards using open receivers. In *CRPIT '44: Proceedings of the 2005 Australasian workshop on Grid computing and e-research*, pages 153–158, 2005.
9. Gerhard Hancke. A practical relay attack on ISO 14443 proximity cards. Manuscript, February 2005. Available at <http://www.cl.cam.ac.uk/~gh275/relay.pdf> accessed October 2005.
10. Gerhard Hancke and Markus Kuhn. An RFID distance bounding protocol. In *Proceedings of the IEEE, SecureComm 2005*, September 2005.
11. Infineon. Device and method for supplying a data transfer unit with energy. US Patent Application 20050252972, 17 November 2005.
12. ISO/IEC. 14443 Identification cards-contactless integrated circuit(s) cards-proximity cards. International Organisation for Standardisation, Geneva, 2001.

13. Adrian Jones and Jason McManus. The measurement of group delay using a microwave system analyser. *Microwave Journal*, 43(8):106–113, 2000.
14. Ziv Kfir and Avishai Wool. Picking virtual pockets using relay attacks on contactless smartcard systems. In *Proceedings of the IEEE, SecureComm 2005*, September 2005.
15. Paul Kocher, Joshua Jaffe, and Benjamin Jun. Differential power analysis. In *Advances in Cryptology CRYPTO 99*, page pages 388397, 1999.
16. Naveen Sastry, Umesh Shankar, and David Wagner. Secure verification of location claims. In *WiSe '03: Proceedings of the 2003 ACM workshop on Wireless security*, pages 1–10, New York, NY, USA, 2003. ACM Press.
17. D. Singelee and B. Preneel. Location verification using secure distance bounding protocols. In *Proceedings of Second IEEE International Conference on Mobile Ad Hoc and Sensor Systems (MASS'05)*. IEEE Computer Society, 2005.

APPENDIX

A Investigations into modulation latency

We hypothesized that a contactless smart card could alter its power consumption using existing load modulation circuitry to communicate in our proposed style of side channel leakage, with sufficiently low latency to detect sophisticated short range relay attacks by well funded attackers. To test the latency aspect of this hypothesis⁶ we investigated the rate of change in carrier amplitude that a card could effect via load modulation, as detected in the receive circuit of the reader. As we have previously noted, the rate of change determines the effective timing resolution.

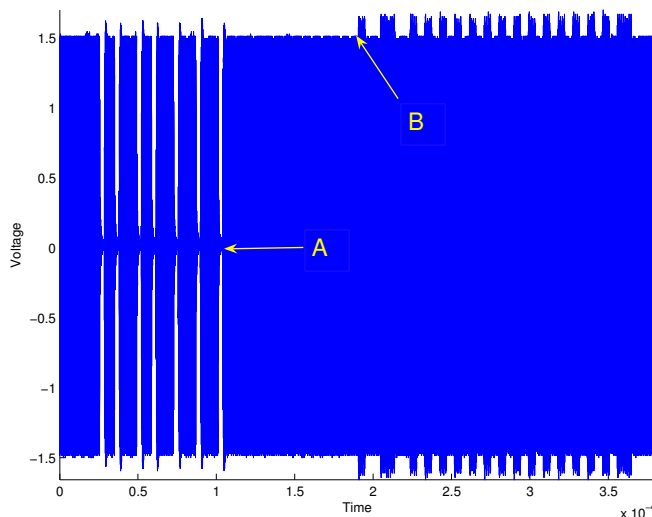


Fig. 4. REQA command/response sequence as measured in reader RX circuit.

⁶ Note that we have not implemented the proposed communication method or distance-bounding protocol in an actual smart card. Such an implementation would require very low level changes to the card operating system mask and circuitry.

A.1 Experimental setup and rationale

We used a Philips Mifare Pegoda development kit reader which conforms to the ISO 14443 type A standard. A digital oscilloscope with a 200 MHz sampling rate was directly attached to the receive circuit of the reader’s antenna. We captured multiple traces of the ‘REQA’ command/response sequence with three different models of card, all of type A (an advanced dual-interface microprocessor card, a 4K Mifare card and a ‘Ultra Lite’ low cost Mifare card). Figure 4 shows a complete sequence. REQA is the first command issued by the reader to the card when it comes into the reader’s field. It was chosen because the standard requires the card to begin modulating its response a fixed number of carrier cycles after the reader’s last command bit. This mirrors our proposal, where the card takes a fixed number of cycles to compute its response however, in the case of REQA, the response always begins with the modulation of a start bit - equivalent to a logical ‘one’ in our scheme.

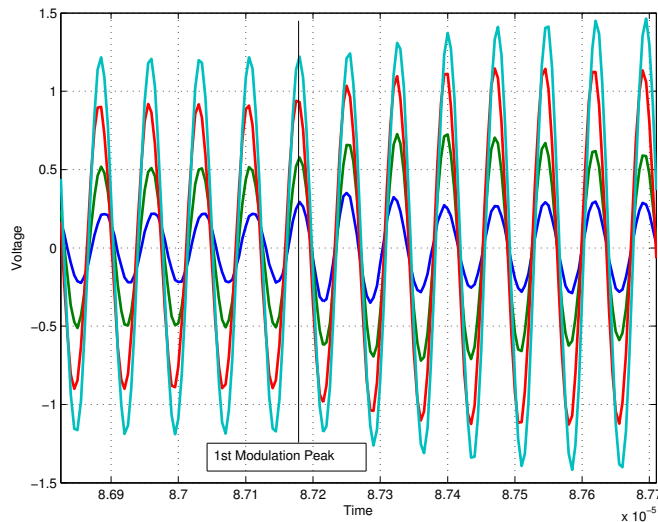


Fig. 5. Card communicates with reader via load modulation. Higher antenna voltages produce delays in effecting a detectable amplitude increase.

Traces were recorded with cards at different distances and orientations to the reader. Orientation and distance effect the voltage across the reader’s receive circuit. The larger the distance the greater the voltage and as the voltage increases, the rate of change that the card can effect on the carrier amplitude decreases. This can be seen clearly in Figure 5 which presents the carrier for four different voltages around the time marked *B* in Figure 4. The first carrier half cycle that the card attempts to modulate is marked with a vertical line. For the two smallest voltage traces an amplitude change is clearly evident. For the highest voltage trace, there is no discernible amplitude change on this cycle - the first significant change can be seen one and a half cycles later at time $0.8729 \mu\text{s}$. We found that the rate of amplitude change is a function of the voltage, irrespective of the orientation and distance that produced it, so in our analysis we consider dependent variables such as delay in detecting a modulation peak, against antenna voltage (rather than distance and orientation which in practice are difficult to specify with precision).

In ISO 14443 type A, the reader communicates with the card via 100% amplitude shift keying (short suspensions of the carrier) as can be seen in the left part of Figure 4. We aligned the traces on the first cycle that resumes the carrier in the last bit pause. This can be seen in Figure 6 which shows that the alignment

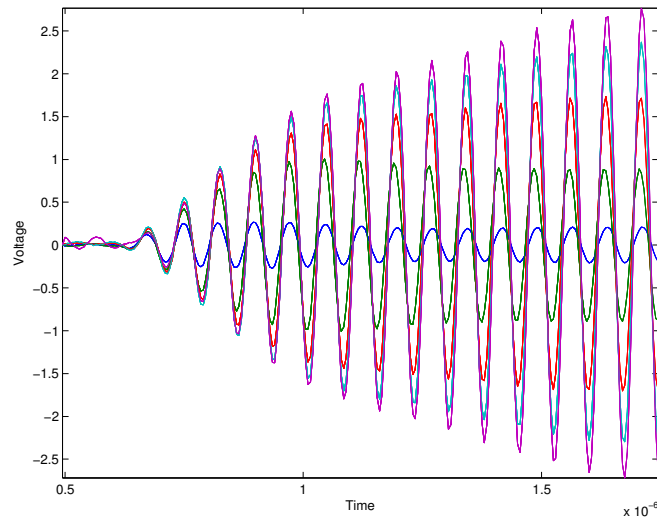


Fig. 6. Trace Alignment on resumption of carrier in last reader to card bit pause.

is precise and unambiguous. Figure 6 presents a short period of the region marked at point *A* in Figure 4 for five different trace voltages. Using this alignment as a reference point, we can confidently identify the half cycle that the card starts its load modulation on irrespective of whether an amplitude change is actually evident. Since we can identify this starting half cycle, we can measure the number of cycles required to produce a detectable amplitude change and hence characterise the timing latency at that voltage.

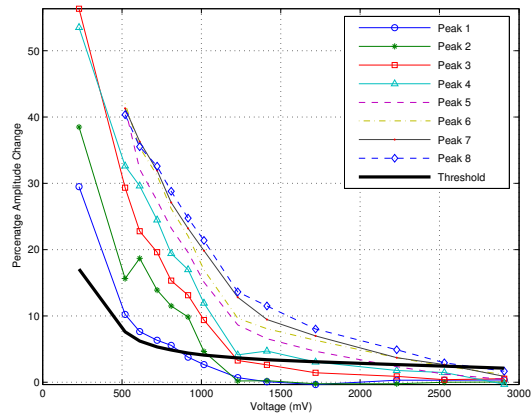


Fig. 7. Percentage change in modulation peak amplitude vs. antenna voltage.

Using Matlab, (a numerical analysis and simulation tool) we developed a model of an amplitude peak detection circuit. The detector compares each half cycle peak amplitude to an average of eight previous peaks of the same sign. If the difference exceeds a threshold value, the current peak is signaled as a modulation. The choice of threshold value is important as it determines the sensitivity to detecting true modulation peaks and also the likelihood that channel noise will be falsely interpreted as a peak. The tradeoff between such true and false positives becomes more delicate as the antenna voltage increases because changes in amplitude become progressively less pronounced, finally disappearing into the noise floor. This effect can be seen in Figure 7. Each data series represents the percentage change in amplitude for a modulation peak number versus antenna voltage. For example, the first modulation peak shows a 30% amplitude increase when the antenna voltage is 200 mV but at 1200 mV there is barely any discernible increase. By examining the amplitudes of each trace on the ‘First Modulation Peak’ line in Figure 5 in relation to the preceding peaks, it should be evident that the respective increases correspond to the Peak 1 series in Figure 7.

We derived a voltage-dependent function to generate the ‘Threshold’ values in Figure 7 in the following manner. Based on 1000 traces taken with one card type at a range of voltages, we experimentally identified a threshold for each trace (by choosing successively smaller values) that correctly identified the modulation peaks whilst keeping false positives below a maximum of 20. Lowering the threshold increases the probability of detecting the real modulation cycles as early as possible but also increases the number of false detections of non-modulated peaks that have slightly higher amplitudes due to environmental noise. The peak detector examines approximately 2300 peaks per trace (the portion of the carrier signal between the points marked *A* and *B* in Figure 4) so the somewhat arbitrary false positive count of <20 loosely approximates an false detect error rate of < 1%, based on the following rationale: in our proposal, the card signals zero by not modulating, which is what the card is doing in the 2300 half cycles between points *A* and *B* so a threshold that produces no more than 20 false positives in this region provides an approximation of a false detect error rate on any individual peak of < 1% when the card sends zero (by not modulating).

The experimentally identified threshold/voltage data pairs were fitted to a third degree polynomial and this function was used to specify threshold values for traces at a range of voltages for the other two card types. We found that the voltage dependent threshold function derived from the data for one card produced very similar true and false positive rates when used on traces for the other cards. While we do not claim our results are conclusive in this respect due to the small number of cards that we have examined, it appears possible that a single voltage dependent threshold function will work across a range of cards. If this is the case, a simple lookup table could supply the detection threshold for a given voltage. Figure 3 shows the average and maximum timing resolution as a function of antenna voltage. The first modulation peak is reliably detectable up to 500mV, so there is no detection delay, thus the timing resolution is $1/2f_c$ or 37 ns. Between 500 and 800 mV there is a slight increase in the average because the first peak is detected in only 75% of cases. We argue that the resolution up to 800mV is sufficient to detect ‘worst case’ sophisticated short range attacks using exotic and expensive relay equipment with very low group delay. The average timing resolution at higher voltages increases to 300 ns which is still a useful resolution for less sophisticated, longer distance attacks. This resolution is still 50 times smaller than the delay introduced by Hancke’s proof of concept attack [9].