Message Time of Arrival Codes: A Fundamental Primitive for Secure Distance Measurement

Patrick Leu, Mridula Singh, Marc Roeschlin, Kenneth G. Paterson, and Srdjan Čapkun
Department of Computer Science
ETH Zurich

***Abstract*—Secure distance measurement and therefore secure Time-of-Arrival (ToA) measurement is critical for applications such as contactless payments, passive-keyless entry and start systems, and navigation systems. This paper initiates the study of Message Time of Arrival Codes (MTACs) and their security. MTACs represent a core primitive in the construction of systems for secure ToA measurement. By surfacing MTACs in this way, we are able for the first time to formally define the security requirements of physical-layer measures that protect ToA measurement systems against attacks. Our viewpoint also enables us to provide a unified presentation of existing MTACs (such as those proposed in distance-bounding protocols and in a secure distance measurement standard) and to propose basic principles for protecting ToA measurement systems against attacks that remain unaddressed by existing mechanisms. We also use our perspective to systematically explore the tradeoffs between security and performance that apply to all signal modulation techniques enabling ToA measurements.**

1. Introduction

When did the message arrive at the receiver? Can this estimate of the message arrival time be manipulated, and in particular by an attacker that controls the communication channel? In particular, can message advancement and delay attacks be prevented? This question is at the core of the problem that distance bounding protocols, secure positioning, and navigation systems are trying to solve: can we prevent the attacker from reducing or enlarging the distance that is measured between the devices? This problem is relevant in a number of application scenarios: contactless payments [1], Passive Keyless Entry and Start Systems [2]–[5], GNSS (e.g., Galileo, GPS) security [6]–[8]. If we could prevent Time of Arrival (ToA) and therefore distance manipulation attacks, we could enable many proximity-based applications, from location-based access control to secure navigation [9], [10].

As a result, many distance bounding protocols have been proposed and analyzed [11]–[13]. Implementations of distance bounding protocols have emerged that combine such protocols with distance measurement techniques [14]–[17], in particular with UWB 802.15.4 radios [18]–[20].

The main idea behind these solutions is to prevent ToA manipulation by the randomization of message content. Namely, it was commonly believed that if the attacker cannot predict the bits of the messages, then he will not be able to advance their time of arrival at the receiver. In [21] the authors argued this to be false – since bits are encoded into symbols, attackers can advance their arrival time. Different physical-layer attacks followed also validating this in practice [3], [22], [23]. This led to the conclusion that secure distance measurement systems can only be built with short symbols and using rapid bit exchange [21]. Given the limits on the output power, such a result would mean that only short-range systems could be made secure. This was shown to be incorrect in [24], which showed that longer symbols can be used if they are interleaved in transmission in a manner that is unpredictable to the attacker. This further demonstrated that secure, long-range distance measurement systems are possible. Recent works further show that, under certain conditions, distance enlargement can also be detected [25]. All these works showed that consideration of the details of how bits are encoded into symbols (i.e., modulation) is crucial in the design of secure distance measurement systems.

This discussion leads to the following questions:

Can we construct a generic message to symbol encoding that prevents any message advancement/reduction (and therefore distance delay/enlargement) for symbols of arbitrary lengths (and therefore arbitrary measurement ranges)?

Can we derive the main principles for the design of such encodings?

In this work, we show that answering these questions is indeed possible. To do so, we introduce Message Time Of Arrival Codes (MTACs), a new class of cryptographic primitive that allows receivers to verify if an adversary has manipulated the message arrival time. In a similar way that Message Authentication Codes protect message integrity, MTACs preserve the integrity of message arrival times. They are, therefore, fundamental to any protocol that relies on Time of Arrival information, such as clock synchronization [26], distance measurement [27] and positioning protocols [28]–[31].

In the same sense that bits can be encrypted with a shared key, the shape of a signal can also be hidden by masking it with a random fast-changing sequence. However, to verify a signal shape, a receiver has to aggregate the signal over a considerable time interval in order to capture enough energy. This is especially so when sender and receiver are separated by longer distances. If the attacker knows the temporal alignment of those aggregations with the signal,

DOI 10.1109/SP40000.2020.00010

he can hide his guessing errors in the null space of the (linear) aggregation function. Simple signal masking is, therefore, not sufficient for the protection against distance manipulation attacks. To address this problem, in addition to using cryptographically-secured modulation (i.e., signal generation), an MTAC also *performs cryptographic checks of the consistency of the modulation* at the receiver.

We give a formal definition of MTACs and their security. We provide the main principles for the design of these codes. We review existing secure distance measurement schemes and draft standards and show how they fit within our MTAC definitions. We then introduce a new *Variance-Based MTAC* that is inspired by our design principles. We show that adhering to these principles allows protection against physical-layer distance-reducing attacks over a wide, realistic performance region. We systematically explore the trade-off between performance and security in ranging.

The rest of the paper is organized as follows. In Section 2, we introduce physical-layer attacks against distance measurement. Section 3 then contains the formal security definitions. Section 4 explores attack strategies. In Section 5, we go over existing proposals. After that, we propose a Variance-Based MTAC in Section 6, which we underline with simulations in Section 7. We conclude in Section 8.

2. Background: Secure Distance Measurement

In this section, we introduce different RF techniques for distance measurement and highlight the challenges towards securing such systems against physical-layer attacks.

2.1. Distance Measurement Techniques and Standards

Establishing location or proximity both require estimating the physical distance between two or more wireless entities. Numerous wireless ranging and localization techniques have emerged in the last decade. Some of these observe physical properties of the signal such as RSSI [32] or phase [33] that change as a function of propagation. However, both these properties can be controlled by an attacker that relays the signal and modifies them to fit another distance claim [34], [35]. The only signal property that cannot be reliably controlled by an attacker is its time-of-flight (ToF). More precisely, an attacker cannot reduce ToF, as a signal cannot traverse space faster than the speed of light.

For ToF measurements, ultra-wideband impulse radio (UWB-IR) has emerged as a prominent technique for precise ranging. It allows high operating distances despite power constraints by transmitting multi-pulse symbols. UWB-IR ranging is in the process of being standardized in IEEE 802.15.4z and is becoming commercially available [18]–[20].

2.2. Distance-Bounding Protocols

Distance-bounding protocols that rely on ToF measurements, as provided by UWB-IR, are the cornerstone for secure proximity verification and positioning. As shown in Figure 1, the basic idea behind such applications is as follows: a prover first commits to a cryptographic nonce; when triggered by receipt of a challenge message from the verifier, the prover sends the nonce, and then sends an opening of its commitment; verification is deemed successful if the commitment opens correctly to the nonce obtained at the verifier. ToF is bounded at the verifier by the time difference between sending the trigger signal and it starting to receive the nonce.

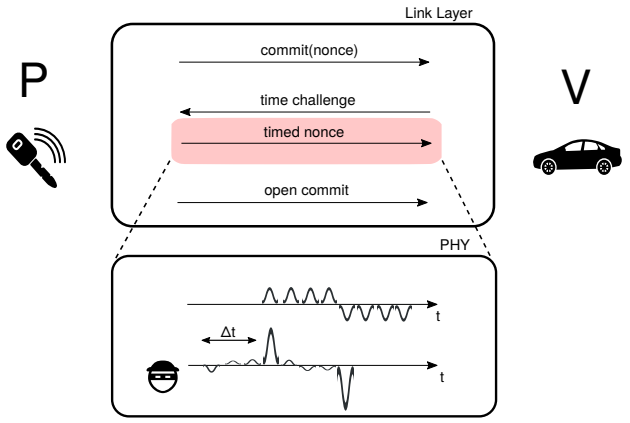

Figure 1. While distance-bounding protocols may be considered secure at the bit-level, systems can still be vulnerable at the physical layer. Distance-bounding protocols that rely on performant (i.e., long-distance), deterministic modulations are vulnerable to ED/LC attacks, as shown in the lower part of the figure. The underlying cause is time-redundant encoding for performance.

Existing vulnerabilities are related to the time-critical aspects of such a protocol, namely adjudging exactly when the nonce starts to be received at the verifier. This is relevant even when secret-dependent masked waveforms are used. It is therefore essential that both the secret information content *and* its time of arrival are carefully tested. Earlier instantiations of distance-bounding protocols relied on a rapid bit-exchange [16] to check both the content and timing of each bit of the nonce in consecutive rounds. As this requires each symbol to be short, this does not scale to longer distances. A distance commitment [16] can be used to decouple time acquisition from content verification. However, there are still doubts about the security level of the content verification, due to targeted attacks on the modulation [21].

2.3. Physical-Layer Attacks

Physical-layer attacks that target the underlying modulation cannot be addressed solely by distance-bounding protocols. In the following, we do not consider attacks that can be averted by conservative signal acquisition (Cicada attack [36]) or involve denial of service (overshadowing, jamming). Instead, we address distance manipulation attacks that exploit redundancies in the modulation, and that are not easily averted by security-aware configuration choices of existing receivers.

Early-Detect, Late-Commit (ED/LC) attack. This attack reduces the distance measured by preemptively injecting a non-committal waveform that triggers an early signal detection at the receiver [21], [22]. The goal is to cause the receiver to register an earlier time of arrival, which, however, the attacker cannot back with knowledge about actual signal content. We illustrate this attack in the lower part of Figure 1. The attacker gets away with this attack due to non-idealities of the legitimate receiver, requiring it to integrate signal power over time for each bit-wise decision, effectively limiting its resolution. To compensate for early deviations from the legitimate symbol (i.e., guessing errors), the attacker significantly amplifies his signal towards the end of each symbol. For maximum effect (distance reduction), the attacker sends the committal, information-bearing part as late as possible after the start of the injected signal. Ideally, this is done to precisely coincide with the start time of the legitimate signal so that the attacker can "copy" it's content (with amplification). An ED/LC attack can be executed fully deterministically and can lead to a distance reduction up to the product of the symbol duration and the speed of light.

Guessing attacks. If the polarity of individual pulses making up a modulated symbol does not only depend on the bit-value of the symbol, e.g., by being fully randomized as in [24], the attacker can resort to a probabilistic ED/LC attack, i.e., a guessing attack. Here, the attacker tries to guess signal components in advance in order to reduce the measured distance. As in an ED/LC attack, the attacker exploits signal redundancies that are required for robust signal reception. The basic idea is that the attacker can compensate early guessing errors by using more power towards the end of the symbol. For each symbol, the attacker can, for instance, double the power as long as his guesses are wrong and stop interfering as soon as a pulse is guessed correctly. This *power-increase* attack is discussed in more detail in Section 4.

2.4. Secure Distance Measurement Solutions

There has been a proposal addressing the outlined threats by cryptographically hiding the bit-wise aggregations in a UWB-based On-Off Keying (OOK) modulation [24]. The authors provide concrete security levels for selected attacker models. A second approach is to correlate an incoming signal, a so-called Scrambled Timestamp Sequence (STS), with the expected signal shape and locking to the peak [37]. To the best of our knowledge, there exists no estimate of the concrete security offered by the latter method. In contrast to both proposals, our work establishes the security goal of any such approach on a fundamental level and outlines a solution permitting to a more general attacker and a broader performance region. We come back to the relation between our work and these existing schemes in Section 5.

3. Message Time of Arrival Codes

In this section, we introduce Message Time of Arrival Codes (MTACs), physical-layer message codes that allow

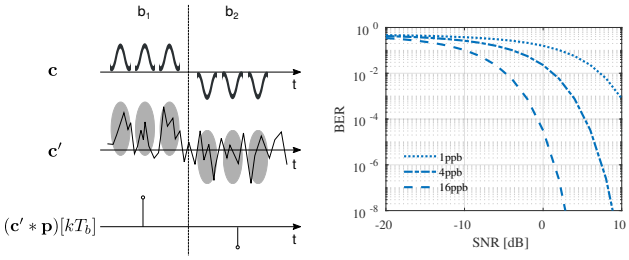

Figure 2. Under noisy conditions, the receiver has to combine multiple short-term signal contributions (samples) to retrieve information. $(\mathbf{c}' * \mathbf{p})$ denotes linear aggregation, e.g., through a matched filter.

the receiver to verify the message time of arrival securely. Such codes preserve the legitimate signal time of arrival under an adversary that tries to "shift" the signal in time, i.e., aim to create the impression of a different arrival time. Fundamentally, the adversarial behavior can be either directed towards producing the code at a time earlier than its legitimate appearance (advancement) or to erase any evidence of a signal, thus opening the possibility for a late imitation (delay). In the following definitions, we divide the problem along those two threats. Although we provide definitions for both threats, we focus on their use in preventing message advancement (i.e., distance reduction) attacks. Preventing such attacks is necessary for the security of *all* secure distance measurement systems. A system performing secure time of arrival measurement might in practice use separate codes to protect against advancement and delay attacks.

Prior to the definitions, we briefly go over boundary conditions on our proposal, stemming from requirements on precision and performance of wireless ranging signals.

3.1. System and Attacker Model

We can model any ranging signal as consisting of short-time signal contributions (i.e., pulses) that carry the information used for precise ranging. As shown in Figure 2, linear combinations of these pulses provide the statistics for the detection of information bits at the receiver. This model covers a broad range of modulation schemes.

Modulation. In the following, we state some assumptions on the modulation. Following Kerckhoffs' principle, we assume the attacker to be aware of all of these aspects of the modulation.

- The modulation consists of a series of elementary, short-time signal contributions called pulses. The effect of ED/LC attacks on such individual signal contributions is considered insignificant (say, less than 1m) in a sufficiently wideband system. We refer to the amplitude level of such a pulse as a *sample*.
- For performance-related considerations, we assume the pulses to be sufficiently spaced such that there is no inter-pulse-interference in the given channel.
- Each information bit is encoded in a symbol consisting of n_{ppb} pulses (and samples). The value

of n_{ppb} is chosen in compliance with a target performance level p within a performance region $\mathcal{P} = [0, BER_{max}] \times [0, d_{max}] \times [0, \Gamma_{max}]$, defined by intervals bounded by the maximally tolerated bit error rate BER_{max}, the maximum communication distance d_{max} as well as the maximum NLoS signal attenuation Γ_{max}.
- Bits are grouped together to form frames, and each frame consists of n_p pulses (and hence n_p/n_{ppb} bits).

Receiver Demodulation. We assume the receiver demodulates by aggregating n_{ppb} samples using correlation with a polarity[1] mask that fits the corresponding hypothesis for each possible value of the information bit. Then, a binary hypothesis test is applied to recover each bit. This is illustrated in Figure 2. We assume an AWGN[2] channel model without inter-pulse interference. In general, the bit error rate (BER) at the receiver is therefore given by the tail bound on the Gaussian distribution, i.e.

$$BER = Q\left(\sqrt{\frac{n_{ppb}P_{rx}}{\sigma_n^2}}\right), \tag{1}$$

under Gaussian thermal noise with variance $\sigma_n^2 = bW \cdot N_0$, where N_0 is the noise power spectral density at room temperature, bW is the system bandwidth and P_{rx} is the receiver-side signal power. Figure 2 highlights the effect of larger n_{ppb} (longer symbols) on BER. This is to highlight the beneficial effect of temporal diversity on performance. Although Equation 1 refers to a BSPK modulation, this effect extends to other modulation techniques. We note that, within this model, for any channel and target BER, there exists an adequate symbol length and assume that the receiver chooses the symbol length accordingly. In this work, we do not assume any (error-correcting) coding.

Attacker Model. We assume that the attacker fully controls the communication channel and has no limitations on how fast she can process messages and react to them. She is, therefore, able to detect individual samples ideally. As a consequence, the attacker's information advantage increases as the channel for legitimate communication worsens, e.g., due to increased distance. We consider two distinct attack models capturing distance reduction (message advancement) and distance enlargement (message delay). In the case of the distance reduction attacker, we pose no restriction on the attacker's abilities regarding the speed of computation,[3] location, or control of the communication channel (e.g., we give the attacker the ability to record and reactively inject messages on the channel with negligible delay). The only restriction that we pose is that the attacker cannot transmit information faster than the speed of light. The attacker's sampling rate needs to be sufficient to recover the signal.

1. Polarity refers to one of two possible phase values of the sample.

2. Additive White Gaussian Noise.

3. Although MTACs can be constructed so as to be information-theoretically secure, most practical schemes will require that the attacker is computationally restricted.

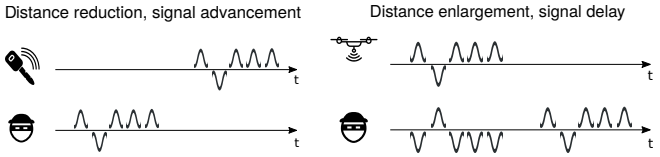

Figure 3. Ideal instantiations of distance reduction (left) and distance enlargement (right) attacks. Both attacks are less likely to be detected the better an attacker can guess the legitimate signal shape. This holds both for releasing an early version (reduction) and for covertly annihilating the valid signal (enlargement).

For an attack to be effective, we don't need to assume that the attacker has a higher bandwidth since we assume the attacker can precisely synchronize to the start of the signal. For the distance enlargement attacker, we assume that the attacker is constrained in terms of location, computation and control of the environment such that she is only able to block the reception of samples if she can anticipate their polarity. However, this includes attackers that operate with multiple (smart) antennas or increase noise levels at the legitimate receiver

Visualising Attacks. In Figure 3, we illustrate ideal instantiations of distance modification attacks. Testing for a distance reduction attack at the receiver consists of a single hypothesis test: either the signal is real (i.e., only distorted by channel) or it is attacker-generated (i.e., it is distorted in a way that indicates that many pulses were guessed wrongly). An attacker is successful if he can produce the expected signal earlier. Verifying existence or absence of a distance enlargement attack, however, involves a multi-hypothesis test in time: the receiver has to check whether a given version of the signal is the first occurrence of its kind or if there exists an earlier, potentially degraded, version sufficiently similar to the legitimate signal. Consequently, in both attacks, an attacker's success chances are higher the better he can anticipate the legitimate signal shape.

3.2. Definitions

A Message Time of Arrival Code (MTAC) is intended to allow detection of any kind of physical-layer distance-modifying attack with high likelihood.

Definition 1 *A message time of arrival code (MTAC) is a tuple of probabilistic polynomial-time algorithms (Gen, $Mtac$, $Vrfy$), such that:*

1) *The key-generation algorithm Gen takes as input the security parameter n and outputs a key k with $|k| = n$.*
2) *The code-generation algorithm $Mtac$ takes as input a key k and a message $m \in \{0,1\}^{n_b}$ and outputs a real-valued vector $\mathbf{c} = (c_1, \ldots, c_{n_p})$. Since this algorithm may be randomized, we write this as $\mathbf{c} \leftarrow Mtac_k(m)$.*
3) *The verification algorithm $Vrfy$ takes as input a key k, a real-valued vector $\mathbf{c}'$ of length n_p, and*

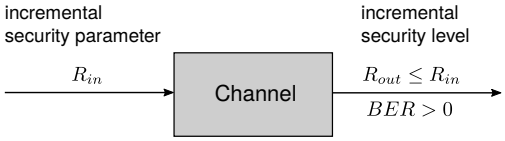

Figure 4. The wireless channel poses a fundamental indirection between the security parameter and the achievable security level. The detectable information rate at the receiver is smaller than the security parameter per second at the transmitter. The particular ratio $R_{out}/R_{in} = 1/n_{ppb}$ results from the modulation and reflects both a performance goal and the channel quality.

message m'. It outputs a bit b. We assume that $Vrfy$ is deterministic, and so write $b := Vrfy_k(m', \mathbf{c}')$.

In the above definition, we assume that m may be transmitted separately from $\mathbf{c}$; however $\mathbf{c}$ can also 'carry' m, which case we assume the existence of an efficient algorithm to extract m from $\mathbf{c}$. In this situation, we can also assume that m' can be extracted from $\mathbf{c}'$ and could choose to suppress it as an input to $Vrfy$. The value of b output by $Vrfy$ is intended to convey that message time of arrival is correct ($b = 1$) or that it cannot be securely verified ($b = 0$).

An MTAC can be seen as a keyed signal verification scheme that guarantees the integrity of the message time-of-arrival. $\mathbf{c} = (c_1, \ldots, c_{n_p})$ is a vector of samples corresponding to the digital representation of the analog signal after A/D conversion. We make no assumptions on the confidentiality or authenticity of m. We assume that these can be achieved through other means, e.g., using encryption or message authentication codes.

Before information can be verified, it has to be transmitted over a wireless channel and detected by the receiver. Strictly speaking, $Vrfy$ involves not only verification but also time-selective *detection* of physical-layer information. As highlighted in Figure 4, detection performance and the resulting security level are fundamentally connected. In general, received samples $\mathbf{c}'$ are affected by channel noise and, in consequence, not identical to $\mathbf{c}$. The detection rate R_{out}, which depends on channel and modulation, is the rate of verifiable information at the receiver. Due to temporal aggregation, it is, in general, smaller than the input data rate, i.e., $R_{out} \leq R_{in}$. Within our assumptions, the ratio R_{in}/R_{out} is given by n_{ppb}. Moreover, detection of this information over a channel is error-prone, which is reflected by a nonzero BER. Consequently, an MTAC will have a nonzero likelihood of false negatives, as well. This we address in a verification criterion that we call *robustness*.

Definition 2 *An MTAC is robust if*

1) *In the absence of an attacker, for any channel, $Vrfy$ applied on $\mathbf{c}'$ is falsely negative with probability at most $1-(1-BER)^{n_b}$, where BER is the error rate in detecting the bits carried by $\mathbf{c}$.*

This means that the false negative rate should remain bounded by the frame error rate on the bit level. Note that we will impose robustness only on detection of distance advancement. As mentioned earlier, detection of delay attacks involves a multi-hypothesis test in time and is, therefore, inherently more prone to false positives.

Distance modification can mean either distance reduction or distance enlargement. The former requires the attacker to *advance* the signal in time, the latter to *delay* the signal in time. We define two different MTAC security models, one for each type of attack (a single model would be unwieldy and difficult to use).

MTAC-A: Modelling Advancement Attacks. In what follows, $\alpha \geq 0$ denotes the *observation delay* of the adversary, measured in samples, representing how long it takes for an attacker to observe and react to a given sample.[4] On the other hand, $\delta \geq 1$ denotes the number of samples by which the adversary tries to advance the signal, quantifying its attack goal. Informally, we allow the adversary access to MTAC code values $\mathbf{c}$ for message inputs of its choice in a fully adaptive manner. Then we challenge it to produce an "advanced" signal $\mathbf{c}'$ for a message m of its choice. We model the latter by requiring the adversary to produce component $c'_{i+\delta}$ of its output before being given samples $(c_1, \ldots, c_{i-\alpha})$ of $\mathbf{c} = Mtac_k(m)$. The adversary wins if it eventually produces a vector $\mathbf{c}'$ for which $Vrfy_k(m, \mathbf{c}') = 1$. An MTAC scheme is (informally speaking) secure against advancement attacks if the probability that any efficient adversary wins is small.

We formalise these ideas in terms of a message time-of-arrival forgery experiment $Mtac\text{-}A\text{-}forge_{A,\Pi}(n)$. In this experiment:

1) The experiment sets $k \leftarrow Gen(n)$.
2) The adversary A is given oracle access to $Mtac_k()$; let Q of size q denote the set of queries made by A.
3) A outputs m, and the experiment sets $\mathbf{c} = Mtac_k(m)$.
4) A then sequentially outputs real values $c'_1, \ldots, c'_{n_p}$; however, after outputting $c'_{i+\delta-1}$ (and before outputting $c'_{i+\delta}$), A is given the samples $(c_1, \ldots, c_{i-\alpha})$ of $\mathbf{c}$.
5) Let $\mathbf{c}'$ denote $(c'_1, \ldots, c'_{n_p})$. Then the output of the experiment is defined to be 1 (and A is said to win) if and only if (1) $Vrfy_k(m, \mathbf{c}') = 1$ and (2) $m \notin Q$. Otherwise, the output of the experiment is defined to be 0.

Note that for schemes in which a message m' (possibly different from m) can be extracted from $\mathbf{c}'$, we can define a different win condition: (1) $Vrfy_k(m', \mathbf{c}') = 1$ and (2) $m' \notin Q$. Here, A still outputs a message m for which she receives a delayed version of $\mathbf{c} = Mtac_k(m)$, but she can win by "forging" a code vector $\mathbf{c}'$ for a different message m' altogether.

Definition 3 *Let $\Pi = \{Gen, Mtac, Vrfy\}$ be an MTAC-A, and let A be an adversary with observation delay α and*

4. Although in our attacker model, we pose no restriction on the adversary's abilities to reactively record and inject samples, α allows us to model weaker attackers whose reaction speed is bounded.

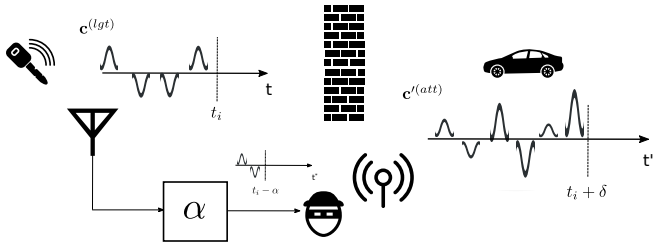

Figure 5. Distance reducing attack. The attacker sees the legitimate signal with an observation delay of α samples and sends his guess δ samples ahead of the actual signal. If successful, the attacker can reduce the measured distance between key and car by δ samples.

advancement goal δ that makes at most q queries to its MTAC oracle and that runs in time at most t (across all steps of the $Mtac\text{-}A\text{-}forge_{A,\Pi}(n)$ experiment). The advantage of A is then defined as:

$$Adv^{MTAC-A}_{A,\Pi}(n) := \Pr[Mtac\text{-}A\text{-}forge_{A,\Pi}(n) = 1].$$

We associate with Π an insecurity function $\boldsymbol{Adv}^{MTAC-A}_{\Pi}(\cdot,\cdot,\cdot,\cdot,\cdot)$, defined as:

$$\boldsymbol{Adv}^{MTAC-A}_{\Pi}(q,t,\alpha,\delta,n) := \max_{A}\{Adv^{MTAC-A}_{A,\Pi}(n)\}$$

where the maximum is taken over all adversaries with observation delay α, advancement goal δ, making at most q queries to its MTAC oracle and running in time at most t.

It is not hard to see that, with all other parameters fixed, the insecurity function is maximised w.r.t. α and δ when $\alpha = 0$ and $\delta = 1$. This corresponds to the situation where the adversary has no observation delay and is given the next sample c_i from $\mathbf{c}$ immediately after outputting its own guess c'_i. The latter corresponds to an adversary who tries to advance the signal by one pulse.

MTAC-D: Modelling Delay Attacks. In the following, we consider an adversary interested in removing all traces of the legitimate signal to perform a delay attack. Under the condition that all evidence of the legitimate signal is removed, the adversary can trivially achieve any delay goal δ without a risk of detection. As the value of δ does not help or limit the adversary, we are not using it in the model. However, by limiting the observation delay $\alpha \geq 0$, we constrain the attacker in its ability to observe (and suppress) the samples that are transmitted by the legitimate transmitter. Generally, we assume that the attacker will not be able to detect the legitimate sample, transmit an opposite sample and thus suppress the legitimate sample. Informally, we allow the adversary access to MTAC code values $\mathbf{c}$ for message inputs of its choice in a fully adaptive manner. Then, we challenge it to produce an "advanced" signal $\mathbf{c}'$ for the message m of its choice. We model the latter by requiring the adversary to produce component c'_i of its output before being given samples $(c_1, \ldots, c_{i-\alpha})$ of $\mathbf{c} = Mtac_k(m)$, i.e., the adversary needs to produce at least one sample in advance for $\alpha = 0$. The adversary wins if it eventually produces a vector $\mathbf{c}'$ for which $Vrfy_k(m, \mathbf{c}'') = 0$ for $\mathbf{c}'' := \mathbf{c} + \mathbf{c}'$. $Vrfy_k(m, \mathbf{c}'')$ outputs 0 if it does not find a trace of $\mathbf{c}$ in $\mathbf{c}''$ and is unable to detect the existence of $\mathbf{c}'$.

We formalise these ideas in terms of a message time-of-arrival forgery experiment $Mtac\text{-}D\text{-}forge_{A,\Pi}(n)$. In this experiment:

1) The experiment sets $k \leftarrow Gen(n)$.
2) The adversary A is given oracle access to $Mtac_k()$; let Q of size q denote the set of queries made by A.
3) A outputs m, and the experiment sets $\mathbf{c} = Mtac_k(m)$.
4) A then sequentially outputs real values $c'_1, \ldots, c'_{n_p}$; however, after outputting c'_i (and before outputting c'_{i+1}), A is given the samples $(c_1, \ldots, c_{i-\alpha})$ of $\mathbf{c}$. Samples c_i and c'_i arrive at the receiver at same time, resulting in the superposition $c''_i = c_i + c'_i$.
5) Let c'' denote $(c''_1, \ldots, c''_{n_p})$. Then the output of the experiment is defined to be 1 (and A is said to win) if and only if (1) $Vrfy_k(m, c'') = 0$ and (2) $m \notin Q$. Otherwise, the output of the experiment is defined to be 0.

Definition 4 *Let $\Pi = \{Gen, Mtac, Vrfy\}$ be an MTAC-D, and let A be an adversary with observation delay α that makes at most q queries to its MTAC oracle and that runs in time at most t (across all steps of the $Mtac\text{-}D\text{-}forge_{A,\Pi}(n)$ experiment). The advantage of A is then defined as:*

$$Adv^{MTAC-D}_{A,\Pi}(n) := \Pr[Mtac\text{-}D\text{-}forge_{A,\Pi}(n) = 1].$$

We associate with Π an insecurity function $\boldsymbol{Adv}^{MTAC-D}_{\Pi}(\cdot,\cdot,\cdot,\cdot)$, defined as:

$$\boldsymbol{Adv}^{MTAC-D}_{\Pi}(q,t,\alpha,n) := \max_{A}\{Adv^{MTAC-D}_{A,\Pi}(n)\}$$

where the maximum is taken over all adversaries with observation delay α, making at most q queries to its MTAC oracle and running in time at most t.

With all parameters fixed, the insecurity function is maximized for $\alpha = 0$. This corresponds to the situation when an attacker's observation delay is limited due to its position or hardware capabilities such that he cannot detect the legitimate sample and suppress them when they are already being transmitted. However, he can observe sample c_i from $\mathbf{c}$ immediately after outputting its own guess c'_i.

Practical MTAC instantiations are likely to rely on a scheme to expand some finite sequence of ideal randomness into a longer one, e.g., using PRFs. We note that, in practice, this is the component vulnerable to higher values of q and t. On the other hand, the security of the verification does not necessarily depend on q and t, i.e., is not affected by those under the assumption of ideal randomness going into signal generation. This is equivalent to stating that verification is not necessarily randomized (beyond the randomness in the signal). However, verification has to be reliable given some, within the computational model bounded, knowledge of the attacker about the PRF output used for signal generation.

4. MTAC Design Space

In this section, we shift to a statistical viewpoint on the design space of secure MTAC schemes and explain how this approach relates to the computational model presented earlier. A statistical analysis entails the advantage of summarizing the infinite number of possible attack strategies. This is particularly beneficial because legitimate as well as adversarial signals can assume uncountably many realizations due to their real-valued nature and due to the uncertainty introduced by noise. Moreover, an attacker is free to choose any amplitude level for each sample of the transmitted signal. The resulting complexity does not allow a straightforward evaluation of all possible strategies in a closed-form computational setting. Also, the security of the verification procedure itself is best analyzed in information-theoretic terms, since verification itself does not have to be randomized, i.e., its security is not necessarily limited to a bounded adversary. Therefore, we present a signal theoretic approach to evaluate different designs of MTACs and argue about the distinguishability of legitimate and attack signals in statistical terms. Although such an approach does not support explicit bounds, we can encapsulate the infinite number of attack strategies and quantify their success in a holistic way. We compare different signals using both, distance on the bit level (Hamming distance) and distance on the sample level (L2-distance), which is motivated by the fact that attack success directly depends on the receiver's inability to distinguish an attacker's guessing error from noise.

Using our statistical model, we identify the symbol-wise mean[5] and (residual) variance as the two main axes of optimization in any attack. We then derive meaningful over-approximations for these two properties that a successful attack signal needs to exhibit and define a strong attacker that will form the basis for the analysis in Section 7

4.1. Distance-reducing attacker

We ignore for a moment that the attacker has to provide a bit sequence that is accepted by the receiver and assume that the adversarial message passes bit-level verification. In that case, detecting a distance-reducing attacker means distinguishing adversarial guessing errors from benign noise on the sample level.

To formulate such a test, we model noise and attacker error as stochastic processes $\mathbf{N}$ and $\mathbf{A}$. The noise process $\mathbf{N}$ is i.i.d. Gaussian (AWGN channel), an assumption that holds as long as signal modulation places samples/pulses reasonably far apart to avoid inter-pulse interference. The attacker process $\mathbf{A}$, on the other hand, reflects the errors produced by the strategy to guess $\mathbf{c}$. An attacker can freely choose the amplitude of its signal based on any strategy, however, $\mathbf{A}$ is random w.r.t. the polarity of the adversarial samples since the attacker has to guess each sample of $\mathbf{c}$. We can capture this in the following hypothesis test:

$$\mathcal{H}_0 : \mathbf{r} \sim \mathbf{N}$$

$$\mathcal{H}_1 : \mathbf{r} \sim \mathbf{A} + \mathbf{N}$$

For each time j (corresponding to one sample), the noise process is distributed as $\mathbf{N}[j] \sim \mathcal{N}(0, \sigma_n)$, the attacker residual as $\mathbf{A}[j] \sim \mathcal{A}_j(A)$, for an attack strategy A. The best strategy is the one for which the hypothesis test distinguishing $\mathbf{A}$ from $\mathbf{N}$ fails with the highest likelihood.

Together with the bit-level requirement that we have so far ignored, we can now formulate any attacker's universal goals as:

1) **Create the correct bits:** In order to achieve correct detection of each bit, the attacker needs to shift the signal mean $\mu_{b'_i}$ w.r.t. the polarity sequence of each symbol $i \in \{1, \ldots, n_b\}$ beyond the sensitivity of the receiver.
2) **Minimize the error energy:** The attacker aims to minimize the residual energy, i.e., the variance of his error distribution $\mathcal{A}_j$ at any time j.
3) **Make the error as indistinguishable from noise as possible:** The attacker aims to hide in the noise the unavoidable[6] guessing error, i.e., to bring the distribution $\mathcal{A}_j$ close to the legitimate noise distribution $\mathcal{N}(0, \sigma_n)$.

Goal 1 targets correctness on the bit level, whereas Goals 2 and 3 are about indistinguishability of the guessed signal from the expected signal on the physical layer. As we will show, for Goal 2, there exists a clear relation to the hardness of guessing each signal sample of $\mathbf{c}$.

In the presented statistical model, achieving all three goals together represents a sufficient condition for attack success, irrespective of potential countermeasures (i.e., detection techniques). There are different ways an attacker can go about these goals: an attacker can (1) select the subset of samples/pulses she wants to interfere with, (2) choose arbitrary amplitude levels for each targeted pulse, and (3) decide how many samples need to be observed before interfering. A meaningful attack strategy will be concerned with how to make these choices in order to satisfy all three goals jointly.
We now describe two general concepts that guide any attack strategy and lead to the definition of a strong attacker by over-approximating signal mean and residual energy.

Steering the mean: Power-increase strategy. Even if the signal is fully randomized at the pulse level, an attacker can guess symbols by employing a *power-increase strategy* as shown in Figure 6. Fundamentally, pulse level randomization under sample-level feedback does not keep an attacker from steering his signal to an arbitrarily high mean under inner product with the hidden polarity sequence. An attacker starts by sending a pulse containing the entire symbol power. He will keep on doubling the power per pulse until he guesses

5. With mean we refer to the accumulated statistics per symbol after inner product with the expected polarity sequence.

6. Since being related to the underlying hardness of guessing the pulses correctly.

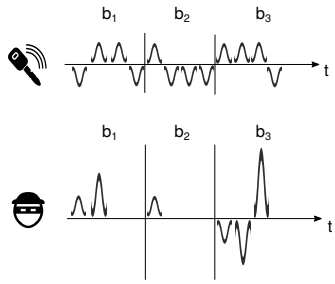

Figure 6. Even under a fully randomized pulse sequence holds: If the receiver (i.e., verifier) combines the pulses to symbols in a predictable manner, the attacker has high chances of getting a sufficiently high symbol-wise mean, by increasing the power in reaction to wrong polarity guesses.

a pulse of the symbol correctly. This attack succeeds with probability $1 - 0.5^{n_{ppb}}$ per symbol. The core takeaway from this attack is that a sample-level guessing error of the attacker does not necessarily translate to a bit-level error, due to the dimensionality reduction applied at the receiver. As long as the attacker can hide the error in the null space of this linear transformation, there is no incentive against the attacker using progressively higher energy levels to 'force' the bits. This means, Goal 1, in isolation, is easy to achieve for an attacker. However, achieving the goal with high likelihood, i.e., more attempts, is associated with higher power levels, which puts Goals 2 and 3 in increasing jeopardy.

Minimizing guessing error by learning pulse polarities. Goals 2 and 3 are directly related to the pulse-guessing performance of the attacker. Depending on how the information bits are modulated, the attacker can potentially use bit-level information to infer the signal or rely on knowledge of past pulses to anticipate the pulse polarities ahead. This would reduce the guessing error and make it harder to detect the attack. Our attacker, as introduced in Section 3.2 has full knowledge about the transmitted bits. In general, any unmasked signal redundancy in time can potentially help the attacker. An example of this is repetition coding or bit-level error-correction coding (ECC) as used in the coherent mode of IEEE 802.15.4z HRP [37]. Also, nonidealities in the underlying PRF can help an attacker.

A strong attacker. We abstract away from all possible strategies and only describe the attack signal statistically subject to an over-approximation of its properties that are linked to the attacker's success: signal mean[7] and residual energy (i.e., residual variance).

As will be motivated, residual variance emerges as observable under a maximum entropy assumption on the attacker's strategy. A result from information theory states that the Kullback-Leibler divergence (i.e., relative entropy) determines the exponent of the error in distinguishing two statistical distributions [38]. Consequently, an attacker that brings its residual closest to the legitimate signal is the strongest. Therefore, we can define the strongest attacker $\hat{A}$ as the one that is closest in the KL-sense over all times:

$$\begin{aligned}\hat{A} &:= \arg\max_{A} Adv(A) \\ &= \arg\min_{A} \sum_{j=1}^{n_p} D_{KL}(\mathcal{A}_j(A) + \mathcal{N} \parallel \mathcal{N}) \\ &= \arg\min_{A} \min_{j} n_p \; D_{KL}(\mathcal{A}_j(A) + \mathcal{N} \parallel \mathcal{N}) \\ &= \arg\min_{A} D_{KL}(\mathcal{A}(A) + \mathcal{N} \parallel \mathcal{N})\end{aligned}$$

The strategy that produces the smallest statistical distance at any j can be converted into the best strategy over the entire signal, by applying the same technique at any other time, since the noise is i.i.d. Therefore, we argue that the attacker that is locally optimal at any time is also optimal over the entire process. The strongest attacker is, therefore, the one that can produce a residual distribution $\mathcal{A} + \mathcal{N}(0, \sigma_n)$ that has smallest relative entropy compared to the legitimate noise distribution $\mathcal{N}(0, \sigma_n)$. Under the condition that the attacker's error has nonzero energy, the process $\mathcal{A}$ that minimizes relative entropy to the AWGN only is also a Gaussian.

Therefore, as an over-approximation, we can model the attacker residual signal process as normally (i.e., maximum entropy) distributed stochastic process with zero mean and a variance given by the pulse-level guessing performance, which we over-approximate. This is equivalent to assuming maximum ignorance about the attacker's process beyond the existence of some residual energy. Under these conditions, e.g. from [39], we know that the signal energy is a sufficient statistic for distinguishing two i.i.d. $\mathcal{N}(0, \sigma_1)$, $\mathcal{N}(0, \sigma_2)$-distributed processes.

Observation 1 *The signal residual variance constitutes a sufficient statistic for detection of a guessing attack with a maximum-entropy residual under AWGN noise.*

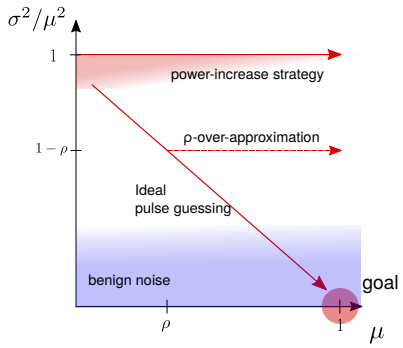

Figure 7. Attacker's strategy space. An attacker needs to exceed a certain symbol-wise mean to produce the correct bits a the receiver. This he can achieve with high likelihood using a power-increase strategy. However, there does not exist any reliable strategy for decreasing the normalized error variance. An attacker can only do so by maintaining an edge in guessing pulse polarities. This we model by over-approximating the attacker, e.g., by giving him a pulse-guessing bias ρ.

7. i.e., the inner product with the expected polarity sequence. Correct guesses contribute to it, wrong guesses diminish it.

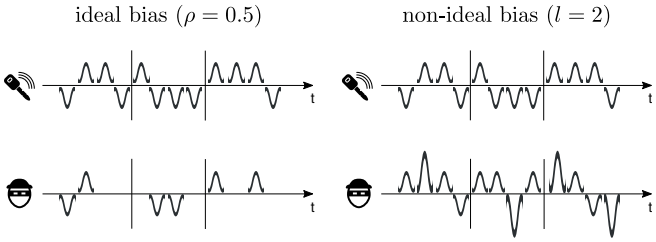

Figure 8. We model two different over-approximations for the attacker's error variance level: An ideal bias, where an attacker knows a fraction ρ of the pulse polarities and a non-ideal bias, where we give the attacker a bound l on the number of power levels for a successful power-increase attack.

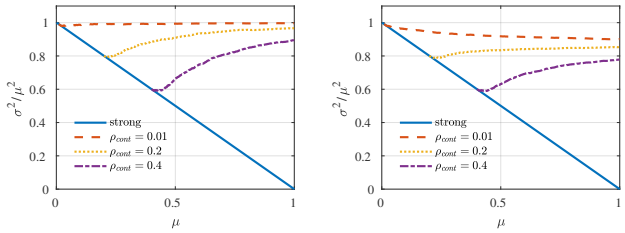

Figure 9. Normalized error variance vs. mean under over-approximation (blue) and continued interference (dashed lines). The goal of the attacker is to get mean to 1 while minimizing variance. Given an ideal bias above a certain threshold ($\rho_{cont} \approx 0.2$), an attacker has nothing to gain from continued interference. The dashed lines show the 0.1th percentile of the variance for unbiased (left) and non-ideally ($l = 2$) biased (right) guessing continuation.

Basing the classification on the residual energy is optimal if we can extract the attacker's error perfectly and within the assumptions, we can universally impose on the attacker's error process (i.e., being close to satisfying the three goals). A practical attacker will likely deviate from these assumptions, but in ways that *add* distinctive properties (i.e., non-zero higher moments) to the residual distribution. Conversely, an attacker that gets mean and variance right will win.

Observation 2 *The attacker getting the mean per bit right* and *minimizing signal residual variance together constitute a sufficient condition for attack success.*

We have seen that, without countermeasures, a power-increase strategy leads to a guessing bias in the receiver-side security parameter (i.e., the bits). As an over-approximation for the course of a power-increase strategy, we can tilt the guessing performance in the attacker's favor *on the pulse level*. For instance, we can assume that the attacker never makes a wrong guess twice in a row. This means, after at most two interferences (i.e., pulses), the attacker is guaranteed to have made a positive net contribution to the receive statistics. We refer to this attacker as having a *non-ideal bias* of $l = 2$ and illustrate it in Figure 8. There, we contrast it to an *ideally-biased* attacker, which knows a given fraction ρ of pulses.

In Figure 7, we highlight the two-dimensional nature of the attack strategy. It is easy for an attacker to steer the mean by varying his energy levels, i.e., to move along the x-axis. However, he cannot control the error variance at the same time. So, any practical attacker strategy will be concerned with trading off those two goals. Providing the attacker an ideal bias results in a diagonal towards the desired spot of high mean and low variance. In addition, as part of any over-approximation, we assume the attacker to be successful regarding the mean (e.g., through a power-increase strategy). This means the attacker can move arbitrarily on the x-axis. In the following, we motivate a specific over-approximation for the error variance, i.e., the attacker's position on the y-axis.

Observation 3 *For an attacker, reducing the signal error variance, while increasing its mean, is 'pulse-guessing-hard'. This means, without a systematic guessing bias, the (normalized) error variance is bound to increase in a guessing attack.*

In Figure 9, we display simulation results underlining this. The results show the normalized residual energy of an ideally-biased attacker (blue line) as a function of the number of interferences, as well as the effect of continued interference without bias (left) and with a non-ideal bias (right). Without bias, the normalized variance is (mostly) monotonically increasing, converging to its maximum value of 1. With a non-ideal bias, the gain that can be maintained is limited. Even with such a consistent bias, only at low values for ρ is there any incentive to continue interfering. Especially, for $\rho > 0.2$, there is no incentive to continue, even with a consistent but non-ideal bias.

Observation 4 *Once the attacker has succeeded in shifting the mean for all symbols, there is (almost) no utility in continued interference, unless the attacker has a lasting pulse-guessing bias. But even then we can find an ideal bias ρ_{cont}, such that there is no utility.*

We see in Figure 9 that persistent interference with a non-ideal bias alone (i.e., no ideal bias, red curve) results in a normalized variance of more than 0.8. We can estimate the strength of this over-approximation as $0.75^{n_p/2}$. This results from the fact that for every two pulses guessed by an attacker, we omit the possibility of two wrong guesses, an event with probability 0.25. By comparing this value to the bit-equivalent MTAC security level of 2^{-n_b}, we can see that an over-approximation with $\rho = 0.2$ is actually stronger than the bit-equivalent MTAC target security level for modulations with $n_{ppb} > 2\frac{\log(0.5)}{\log(0.75)} \approx 4.82$, i.e., at least five pulses per bit. A decrease of the relative variance to 0.8 or, equivalently, an ideal bias of $\rho = 0.2$ are, therefore, very strong over-approximations, i.e., on the order of the (receiver-side) security parameter, that become even stronger (less likely) for modulations over longer communication distances.

5. Existing MTACs

Based on our insights on the attack, we need an MTAC to verify the physical-layer integrity of a signal by measuring the (normalized) signal residual variance. To the best of our knowledge, there are three existing classes of MTACs

that, as we argue, aim for this implicitly. Each class is parametrized by a performance parameter that allows to trade off performance and security. Note that, in the following, the robustness definition does not directly apply to the first two classes since those do not entail reliable information transmission.

5.1. Sequences of single-pulse bits

To allow for longer range while using short symbols, one could encode each bit as a single pulse and tolerate up to a certain pre-configured rate of bit errors T_{BER} in the verification step, as is currently proposed in 802.15.4z LRP [37]. This results in a secure MTAC under the condition that the message m is pre-shared between transmitter and receiver. Since relying on a single pulse makes bit transmission unreliable, this is a purely physical-layer construct and does not allow for integrity-protected data transmission. In particular, a MAC will fail if there is a nonzero number of expected bit errors per message. The resulting security and performance level both depend on the BER tolerance level. For $\mathbf{c} \leftarrow m \oplus \mathbf{x}$, where $\mathbf{x}$ is an ideal random sequence, the attacker's advantage can be estimated using Sanov's theorem, provided in [38], as

$$P(\overline{X}_{n_p} \geq (1 - T_{BER})) = 2^{-n_p D_{KL}(P||S)}.$$

Here, P and S capture the empirical and theoretical binomial distributions, with $P = (1 - T_{BER}, T_{BER})$ and $S = (0.5, 0.5)$, respectively. The random variable $\overline{X}_{n_p}$ denotes the number of bits guessed correctly by the attacker. For $T_{BER} < 1$, we can achieve any concrete security goal by setting n_p appropriately. For example, under the assumption of ideal randomness and an unbounded adversary, for $\alpha = 0$ and $\delta = 1$, we achieve 32 bits of security by transmitting 116 bits while tolerating up to $T_{BER} = 20\%$ bit errors. These results directly translate to a computational setting with reduced shared randomness by replacing the ideal randomness with the output of a PRF.

5.2. Correlation sequences

A standard way of signal time acquisition is correlating an incoming signal with the expected signal shape and locking to the peak. This is suggested in current 802.15.4z HRP standardization efforts by means of a so-called Scrambled Timestamp Sequence (STS) [37]. One could argue that secure ToA verification could be achieved by checking for this peak. However, we know that tests for time acquisition and content verification should not be naively coupled [36]. If only used for content verification, the security of such an approach depends on the detailed test the receiver applies, i.e., the degree to which the relative quality of the peak (peak-to-average ratio) is considered. The power level at the peak itself can be easily steered to the desired value by an attacker relying on a power-increase strategy. Therefore, tests based on a) the existence of a correlation peak or b) the peak power level are not secure. Further exploration of this technique is deferred to future work.

5.3. Hidden encodings: UWB-PR

Secure ToA-verification based on the bit-level content can be achieved by hiding the mapping from pulses to bits as in [24]. This can be thought of a scheme that reorders the pulses belonging to each bit within the frame. This is currently the only scheme that is implemented and secure. It forces the adversary to transmit a small number of pulses, and, in case their polarity is correct, hope that they align with the bits. The authors analyze the performance and security of such an attack model. The adversary is given the same capabilities as assumed in Section 3.1. They have analyzed attack strategies for $\alpha = 0$ and $\delta = 1$, and have assumed that the communicating parties share large amounts of ideal randomness. However, a formal proof is needed to determine if the attack strategy is optimal. The results show that 16 pulses per bit are required under LoS conditions to prevent bit errors over a distance of 92m. For 32 bits of security at least 100 bits have to be sent in a message. As long as the attacker does not guess all but n_{ppb} pulses correctly, he has an advantage less than one, due to his uncertainty about the reordering. Consequently, under this assumption, we can add more bits to the frame to achieve any concrete security level.

6. Variance-Based MTAC

In the following, we propose the Variance-Based MTAC for direct variance estimation, consisting of rules for signal creation and a receiver-side verification procedure. We then embed this technique into a generic verification algorithm and address side requirements for its practical instantiation.

6.1. Tx-side signal generation (Gen, $Mtac$)

We assume each sample to follow a binary encoding, achieved either through on-off keying (OOK), frequency-shift keying (FSK) or phase-shift keying (PSK), but not pulse-position modulation (PPM). The reason is that, in PPM, the fundamental signal contribution representing each sample becomes vulnerable to ED/LC. Within our assumptions about the modulation, we can represent the transmit signal as a binary pulse sequence of length $n_p = n_{ppb} \cdot n_b$. In particular, we assume that pulses are separated by more than the channel delay spread, i.e., there is no inter-pulse interference. Without this assumption, signal degradation under benign conditions might be hard to distinguish from attacks. The bits are first encoded in a frame $\mathbf{b} = (\mathbf{s}_{b_1} \| \dots \| \mathbf{s}_{b_{n_b}})$, consisting of symbols that each represent message bit under repetition coding, either as $\mathbf{s}_1 = \{1\}^{n_{ppb}}$ or $\mathbf{s}_0 = \{-1\}^{n_{ppb}}$. Preventing an attacker from inferring pulse polarities from either the content of the message m or past samples is achieved by relying on full pulse-level randomization, i.e., by applying a secret sequence $\mathbf{x}$ on the pulses, as in

$$\mathbf{c} = \mathbf{b} \oplus \mathbf{x}.$$

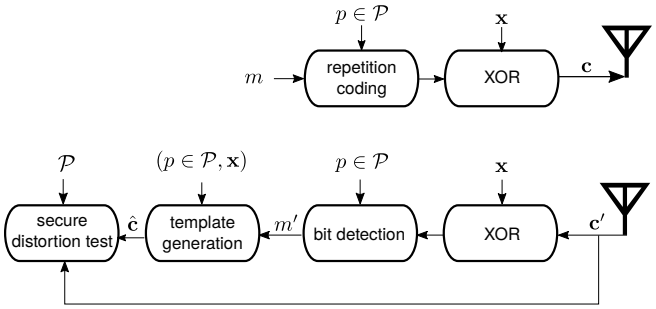

Figure 10. Tx/Rx structure of a Variance-Based MTAC. A keyed XOR and a secure distortion test are the central security components. For simplicity, we omit the modulation of each value in $\mathbf{c}$ onto a UWB pulse in the picture. Bit encoding and decoding are parametrized by a performance level p, whereas the secure distortion test applies to an entire performance region $\mathcal{P}$.

We can either idealize $\mathbf{x}$ being perfectly random, as in

$$\mathbf{x} \leftarrow \{-1, 1\}^{n_p},$$

and shared between transmitter and receiver, or being generated using a pseudorandom function that operates on a previously shared secret.

6.2. Rx-side operations (*Vrfy*)

A message time of arrival code has to combine bit detection and verification with an additional signal verification for ensuring the correct signal time of arrival. The bit-level tests are a sequence of binary hypothesis tests. The additional check is a single binary test applied to the entire signal, parametrized by the bits received. We illustrate the whole pipeline in Figure 10.

Bit detection. Each bit is carried by n_{ppb} pulses. The receiver combines the energy of those pulses subject to the bit-wise hypothesis and the XOR-mask and applies a binary hypothesis-test per bit. The outcome is a received bit sequence $m' = (b'_1, \ldots, b'_{n_b})$.

Signal residual extraction. In order to test the signal integrity on the physical layer, we need to extract the signal-level residual. We exemplify the residual extraction at the receiver in Figure 11. Under our stated assumptions about channel and modulation, the received signal $\mathbf{c}'$ consists of the actual signal $\mathbf{c}$, attenuated by path loss, as well as additive white Gaussian noise (AWGN). At the receiver, the expected pulse polarity sequence (i.e., the template) $\hat{\mathbf{c}}$ is constructed based on the detected bits b'_i and the shared XOR-sequence $\mathbf{x}$, as in $\hat{\mathbf{c}} = (\mathbf{s}_{b'_1} \| \ldots \| \mathbf{s}_{b'_{n_b}}) \oplus \mathbf{x}$. We refer to this step as *template generation* in Figure 10. The receiver-side equivalent pulse train is then given by the element-wise multiplication of the received signal with the expected pulse polarity sequence $\hat{\mathbf{c}}$, as $\mathbf{p}' = \mathbf{c}' \odot \hat{\mathbf{c}}$. The residual is then obtained by subtracting the expected value from the receiver-side equivalent pulse train, as in $\mathbf{r} = \mathbf{p}' - \mu_{\mathbf{p}'}$. A variance-based hypothesis test is concerned with whether the receive signal error consists of model error only or also contains an attacker error. As we argue in Section 4, the property to test for is the variance of the signal residual, i.e., if the variance matches the expected noise or is too large, i.e., was caused by attacker errors. However, we require some normalization since the overall receive SNR will vary.

Secure distortion test. We need to normalize the observed signal error to the overall signal energy. This way, we do not need to maintain an explicit noise estimate. The worst-case SNR is found by maximizing over the performance region $\mathcal{P}$, guiding the choice of a threshold for the legitimate distortion. We are then able to check if the observed distortion, i.e., the overall normalized signal error, is within this bound.

This involves a hypothesis test on the normalized variance of the received signal, after being XOR-ed with the expected sequence. As secure distortion function, we propose taking the ratio between the power of the signal residual and the overall received power:

$$\mathcal{D} = \frac{\mathbf{r}^2}{\|\mathbf{c}'\|^2} = \frac{\sigma^2_{\mathbf{p}'}}{\|\mathbf{c}'\|^2} = \frac{\sum_i \left(\mathbf{c}'[i]\hat{\mathbf{c}}[i] - \frac{\sum_j \mathbf{c}'[j]\hat{\mathbf{c}}[j]}{n_p}\right)^2}{\|\mathbf{c}'\|^2}$$

The distortion can be interpreted as the inverse of a receive SNR estimate, based on a hypothesis on the pulse-level structure of the received signal. A random, zero-mean process will, for instance, evaluate to a distortion of $\mathcal{D} = 1$.

Consequently, we can write the hypothesis test given by *Vrfy* as a decision between a signal containing some of the expected structure

$$\mathcal{H}_0 : \ \mathcal{D} < 1,$$

and the signal being only (attacker-induced) random noise:

$$\mathcal{H}_1 : \ \mathcal{D} = 1.$$

Performance region, decision threshold. We assume the transmitter to choose the number of pulses per bit appropriately given a previously selected performance level $p = (d', BER', \Gamma'_{nlos}),\ p \in \mathcal{P}$, i.e., such that

$$Q\left(\sqrt{\frac{n_{ppb} P_{rx}(d', \Gamma'_{nlos})}{\sigma_n^2}}\right) \overset{!}{\leq} BER'.$$

To satisfy our robustness criterion, the maximum legitimate signal distortion needs to be chosen such that the false negative rate does not exceed the underlying frame error rate, i.e.,

$$T_{\mathcal{D}}(p) = \max\left(T'_{\mathcal{D}} \in [0, 1]\right) \ \text{, s.t.} \ \ P[\mathcal{D}_{lgt} > T'_{\mathcal{D}}] \overset{!}{\leq} FER.$$

The effective threshold is then chosen as the maximum threshold over the entire performance region, i.e., $\hat{T}_{\mathcal{D}} = \max_{p \in \mathcal{P}} T_{\mathcal{D}}(p)$. As a result, $\hat{T}_{\mathcal{D}}$ results in a robust test under any performance tradeoff within the performance region $\mathcal{P}$.

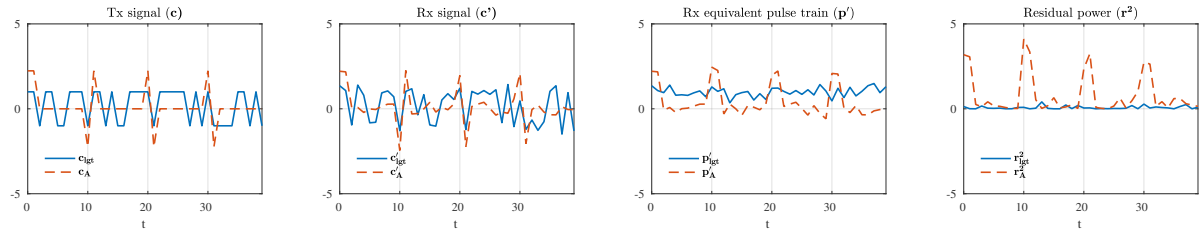

Figure 11. Legitimate (blue) and attack (red) signals in a scenario with four bits and 10 bits per symbol and repetition coding. The first plot shows the shape of the transmitted signal. The attack signal is winning since each bit contains sufficient power, despite the attacker only guessing 2 out of 10 pulses per symbol. The second plot shows the noisy signals at the receiver. The third plot shows the received signal after removing the data modulation. The residual after the expected signal component has been removed is shown in the rightmost plot. It becomes evident that the attack residual can be discerned easily from the legitimate residual, despite the attack on repetition coding (i.e., the bit level) being successful.

6.3. Variance-Based MTAC: Summary

To summarize and illustrate how to embed the Variance-Based MTAC into a distance-measurement system, we highlight the steps involved in the detection of an advancement attack by a receiver (Rx) on a signal originating from a transmitter (Tx).

(**a**) Pre-configuration

1) Rx determines the maximum accepted distortion threshold $\hat{T}_{\mathcal{D}}$ based on the maximum communication distance and maximum tolerated noise level, subject to a performance region $\mathcal{P}$.

(**b**) Key generation (Gen)

1) Tx and Rx derive a fresh pseudorandom XOR sequence $\mathbf{x}$ from some shared secret. $\mathbf{x}$ could theoretically also be secretly shared before each round[8].

(**c**) Mtac generation ($Mtac$)

1) Tx encodes the message m using repetition coding according to a chosen configuration $p \in \mathcal{P}$ and applies the XOR sequence.

(**d**) Mtac verification ($Vrfy$)

1) Rx constructs the message m' by multiplying the received pulse sequence $\mathbf{c}'$ with the expected XOR sequence and applying a bit-wise binary hypothesis test on the overall symbol energy.
2) Based on the received message m' and the XOR sequence, Rx constructs the expected pulse-level sequence $\hat{\mathbf{c}}$ (i.e., the template).
3) Rx computes the signal distortion $\mathcal{D}(\mathbf{c}', \hat{\mathbf{c}})$ between received and expected pulse sequence.
4) Rx checks if $\mathcal{D}$ exceeds $\hat{T}_{\mathcal{D}}$. If so, it declares attack.

6.4. Practical concerns

Time reference: Distance commitment. We assume the detection of an advancement attack to be limited to verification of the data relative to some established time frame. This can be achieved by a distance commitment as introduced in [16]. This means the prover is assumed to have already responded in quick fashion to the query by transmitting a deterministic preamble, i.e., is committed to certain temporal reference. Relative to this temporal reference, the prover then has to deliver the secret information (i.e., m, correctly modulated) at a pre-agreed time relative to the preamble. It is realistic to assume a channel to be coherent throughout the frame, as the duration of a UWB frame used for distance measurement is typically less than 1ms. Through a distance commitment, the vulnerabilities of a back-search [36] on the data-bearing part can be avoided.

Ranging precision. Under a distance commitment, the back-search for the acquisition of the first signal path is only necessary on the preamble of the frame. Therefore, the precision of the ranging procedure is not determined by any operation applied to the data-bearing part. Consequently, the precision of our proposal cannot be worse than that of existing schemes relying on a distance commitment. It has been shown that such a system can achieve a precision of 10cm, irrespective of communication distance [19], [24].

Bit-level security. We assume a bit-level procedure to detect if the received bits m' do not match the transmitted message m. This could be achieved by a message authentication code (MAC) appended to the frame or even transmitted on a separate, potentially ToA-agnostic channel.

7. Analysis

In the following, we explore the tradeoff between security and performance by modeling the effect of the channel and evaluating the classification performance of our Variance-Based MTAC from the previous section. The results are based on simulations, which, however, make assumptions in line with realistic UWB-based distance measurement systems. From these results, we can derive the performance region in which our proposal maintains bit-equivalent security (i.e., $Adv(\hat{A}) < 2^{-n_b}$) and how to scale to longer distances.

8. We don't have any requirements on ToA protection in this step.

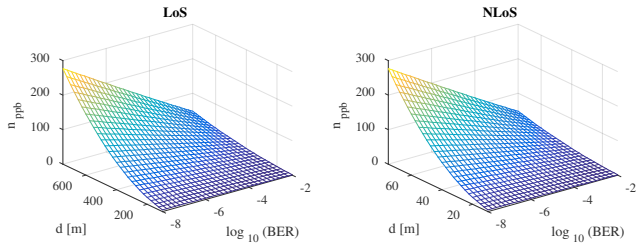

Figure 12. The number of pulses per symbol as a function of the target performance level, i.e., the target bit error rate (BER) and operating distance under FCC/ETSI constraints. These numbers refer to a LoS (left) as well as a NLoS scenario (right) with 20dB attenuation of the direct path. Lower BERs over longer distances require more pulses per symbol.

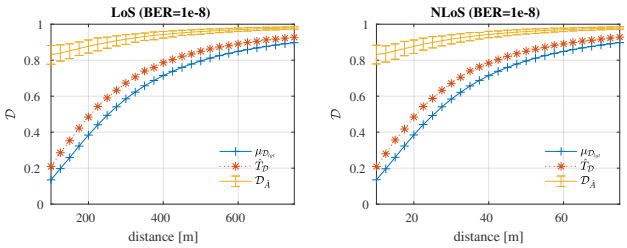

Figure 13. Over longer distances, the legitimate distortion increases. The gap between maximum legitimate distortion and minimum attack distortion becomes smaller for longer distances, eventually vanishing altogether. This means, under our strong attacker model, MTAC security can only be maintained up to some distance.

7.1. Model

Path loss model. To evaluate the impact of distance on a) the modulation required and b) the implications on security, we assume a free-space path loss model. This means the received power degrades inversely to the square of the distance, as in

$$P_{rx} = P_{tx} \left(\frac{\lambda}{4\pi d} \right)^2 \Gamma_{nlos}.$$

We assume the antennas to be operated in each other's far field, as the goal of this analysis is to understand the tension between long distance and security. As input power, we rely on the constraints put forward by the FCC and ETSI regarding UWB in licensed spectrum. This is, a maximum peak power of 0dBm within the 50MHz around the peak and an average limitation on signal power spectral density of -41.3dBm/Hz. We assume that our pulses are sufficiently spaced, such that each pulse can be sent at peak power. We assume a signal bandwidth of 620MHz at a center frequency of 6681.6MHz, which is a typical UWB channel configuration [37]. For receiver-side noise, we consider the thermal noise figure at room temperature, given by -174dBm/Hz. In a separate non-line-of-sight (NLoS) scenario, we assume an additional attenuation of 20dB which is roughly the attenuation the signal experiences when traversing the human body. In Figure 12, we show the number of pulses per symbol required under both LoS and NLoS conditions. The required number of pulses increases with longer distances and decreases if the requirement on target BER gets relaxed.

Gaussian model for variance distributions. The variance constitutes a sum of n_p independent random variables. Due to the central limit theorem, for a sufficiently high overall number of pulses, the variance distribution converges to a Gaussian, i.e.,

$$\mathcal{D}_{\hat{A}}(d) \sim \mathcal{N}\left(\mu_{\mathcal{D}_{\hat{A}}}(d), \sigma_{\mathcal{D}_{\hat{A}}}(d)\right) \quad (2)$$

$$\mathcal{D}_{lgt}(d) \sim \mathcal{N}\left(\mu_{\mathcal{D}_{lgt}}(d), \sigma_{\mathcal{D}_{lgt}}(d)\right). \quad (3)$$

In general, these distributions are a function of the communication range as well as the target BER. Through simulations, we can verify that in the area of interest (i.e., where the distributions significantly overlap), these distributions indeed fit a Gaussian hypothesis well, as we show in detail in Appendix C.

7.2. Results

We model the bit error rate of the underlying modulation according to Equation 1. We simulate this in MATLAB for a frame of 32 bits. As we detail in Appendix D, the security guarantees are maintained for longer frames. For robustness, the choice of the decision threshold should result in the same false negative rate of $Vrfy$ as under bit-wise detection, i.e., $FNR_{Vrfy} \stackrel{!}{=} 1 - (1 - BER)^{n_b}$. Under the Gaussian hypothesis for the distortion distribution, we can derive the practical decision threshold by choosing it $Q^{-1}(FNR_{Vrfy})$ normalized standard deviations above the expected legitimate distortion. The resulting threshold is indicated in Figure 13. We evaluate the probability of attacker success for a given maximum communication distance based on the attacker's best case statistics and the legitimate worst-case statistics, over a range of target BER values. This is in line with our attacker model, which does not make any assumptions about the attacker's position. For a given performance region, the upper bound of the attacker's advantage is given by

$$Adv(\hat{A}) = Q\left(\frac{\hat{\mu}_{\mathcal{D}_{\hat{A}}} - (\hat{\mu}_{\mathcal{D}_{lgt}} + Q^{-1}(FNR_{Vrfy}) \cdot \hat{\sigma}_{\mathcal{D}_{lgt}})}{\hat{\sigma}_{\mathcal{D}_{\hat{A}}}} \right),$$

whereas the statistical parameters, i.e., means and variances, are chosen in favor of forger $\hat{A}$. Specifically, we choose the attacker's parameters under minimization of the worst-case distortion and the parameters of the legitimate transmitter under maximization of the distortion, within the defined performance region. The details of those choices we provide in Appendix B. Unsurprisingly, the worst-case distance for the legitimate transmitter amounts typically to the maximum distance. The numerical values of those statistical parameters (i.e., means and variances) were obtained through simulation. We thereby modeled the attacker as having an ideal pulse-level bias of 20%, as motivated in Section 4. In the following, we are interested in the performance region in which the MTAC provides bit-equivalent security, i.e., $Adv(\hat{A}) \leq 2^{-n_b}$.

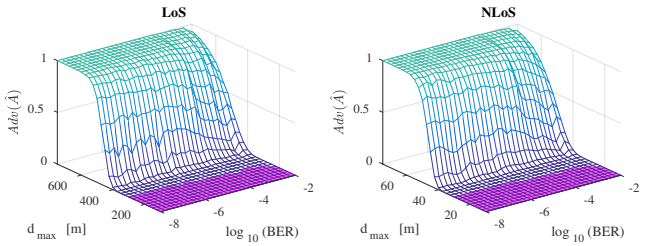

Figure 14. Attacker's advantage as a function of the performance level. We highlight the performance region within which the MTAC provides bit-equivalent security. The secure distortion test provides us with a bit-equivalently secure MTAC for distances up to 200m and 20m for LoS and NLoS scenarios, respectively.

Performance-equivalent MTAC region. Figure 14 shows the attacker's advantage as a function of the performance level. The figure highlights the performance region in which we have bit-equivalent MTAC security.

Observation 5 *Under any tradeoff between symbol length and target bit error rate: For any frame m of at least 32 bits, we can find a distortion threshold $\hat{T}_{\mathcal{D}}$ resulting in an MTAC with bit-equivalent security for distances up to 200m under LoS conditions and up to 20m under NLoS conditions.*

Extending the MTAC region. By comparing the results for LoS and NLoS conditions, we see that the MTAC region seems to degrade proportionally to the attenuation added, i.e., the results are invariant under amplification/attenuation. This means we can extrapolate to any communication range if we allocate a security link margin $\Gamma_{sec} \geq 0$ satisfying

$$\Gamma_{sec} \overset{!}{\geq} 20 \cdot \log_{10}\left(\frac{d_{max}}{200m}\right) + \Gamma_{nlos}.$$

8. Conclusion

With MTAC, we propose a physical-layer primitive for secure distance measurement. We formally define the security of its underlying algorithms. We then derive design principles for the practical instantiation of an MTAC: A randomized pulse sequence and a secure distortion test over the entire signal. The results indicate that the bit-equivalent security level can be regained over a meaningful performance region, thereby resulting in a fundamental building block preventing any physical-layer, distance-reducing attacks.

9. Acknowledgements

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 726227.

References

[1] K. Markantonakis, L. Francis, G. Hancke, and K. Mayes, "Practical relay attack on contactless transactions by using nfc mobile phones," *Radio Frequency Identification System Security: RFIDsec*, vol. 12, p. 21, 2012.

[2] A. Francillon, B. Danev, and S. Capkun, "Relay attacks on passive keyless entry and start systems in modern cars," in *Proceedings of the Network and Distributed System Security Symposium, NDSS 2011, San Diego, California, USA, 6th February - 9th February 2011*, 2011.

[3] A. Ranganathan, B. Danev, A. Francillon, and S. Capkun, "Physical-layer attacks on chirp-based ranging systems," in *Proceedings of the fifth ACM conference on Security and Privacy in Wireless and Mobile Networks*. ACM, 2012, pp. 15–26.

[4] "Mercedes 'relay' box thieves caught on cctv in solihull." http://www.bbc.com/news/uk-england-birmingham-42132689, [Online; Accessed November 10th 2018].

[5] "Radio attack lets hackers steal cars with just $20 worth of gear." https://www.wired.com/2017/04/just-pair-11-radio-gadgets-can-steal-car/, [Online; Accessed November 10th 2018].

[6] T. E. Humphreys, B. M. Ledvina, M. L. Psiaki, B. W. O'Hanlon, and P. M. Kintner, "Assessing the spoofing threat: Development of a portable gps civilian spoofer," in *Radionavigation Laboratory Conference Proceedings*, 2008.

[7] P. Papadimitratos and A. Jovanovic, "Gnss-based positioning: Attacks and countermeasures," in *Military Communications Conference, 2008. MILCOM 2008. IEEE*. IEEE, 2008, pp. 1–7.

[8] N. O. Tippenhauer, C. Pöpper, K. B. Rasmussen, and S. Capkun, "On the requirements for successful gps spoofing attacks," in *Proceedings of the 18th ACM conference on Computer and communications security*. ACM, 2011, pp. 75–86.

[9] S. Capkun, M. Cagalj, G. Karame, and N. O. Tippenhauer, "Integrity regions: Authentication through presence in wireless networks," *IEEE Transactions on Mobile Computing*, vol. 9, no. 11, pp. 1608–1621, 2010.

[10] K. B. Rasmussen, S. Capkun, and M. Cagalj, "Secnav: secure broadcast localization and time synchronization in wireless networks," in *Proceedings of the 13th annual ACM international conference on Mobile computing and networking*. ACM, 2007, pp. 310–313.

[11] S. Brands and D. Chaum, "Distance-bounding protocols," in *Workshop on the Theory and Application of of Cryptographic Techniques*. Springer, 1993, pp. 344–359.

[12] C. H. Kim, G. Avoine, F. Koeune, F.-X. Standaert, and O. Pereira, "The swiss-knife rfid distance bounding protocol," in *International Conference on Information Security and Cryptology*. Springer, 2008, pp. 98–115.

[13] A. Brelurut, D. Gerault, and P. Lafourcade, "Survey of distance bounding protocols and threats," in *International Symposium on Foundations and Practice of Security*. Springer, 2015, pp. 29–49.

[14] G. P. Hancke and M. G. Kuhn, "An rfid distance bounding protocol," in *Proceedings of the First International Conference on Security and Privacy for Emerging Areas in Communications Networks*, ser. SECURECOMM '05. Washington, DC, USA: IEEE Computer Society, 2005, pp. 67–73. [Online]. Available: http://dx.doi.org/10.1109/SECURECOMM.2005.56

[15] K. B. Rasmussen and S. Capkun, "Realization of rf distance bounding." in *USENIX Security Symposium*, 2010, pp. 389–402.

[16] N. O. Tippenhauer, H. Luecken, M. Kuhn, and S. Capkun, "Uwb rapid-bit-exchange system for distance bounding," in *Proceedings of the 8th ACM Conference on Security & Privacy in Wireless and Mobile Networks*. ACM, 2015, p. 2.

[17] A. Ranganathan, B. Danev, and S. Capkun, "Proximity verification for contactless access control and authentication systems," in *Proceedings of the 31st Annual Computer Security Applications Conference*. ACM, 2015, pp. 271–280.

[18] "3db access ag - 3db6830 proximity based access control," https://www.3db-access.com/Product.3.html, [Online; Accessed November 8th 2018].

[19] "DecaWave "dw1000 product description and applications"," https://www.decawave.com/products/dw1000, [Online; Accessed November 8th 2018].

[20] "Zebra Technologies "sapphire dart ultra wideband (uwb) real time locating system 2010."," https://www.zebra.com/us/en/solutions/location-solutions/enabling-technologies/dart-uwb.html, [Online; Accessed November 8th 2018].

[21] J. Clulow, G. P. Hancke, M. G. Kuhn, and T. Moore, "So near and yet so far: Distance-bounding attacks in wireless networks," in *European Workshop on Security in Ad-hoc and Sensor Networks*. Springer, 2006, pp. 83–97.

[22] M. Flury, M. Poturalski, P. Papadimitratos, J.-P. Hubaux, and J.-Y. Le Boudec, "Effectiveness of distance-decreasing attacks against impulse radio ranging," in *Proceedings of the third ACM conference on Wireless network security*. ACM, 2010, pp. 117–128.

[23] M. Poturalski, M. Flury, P. Papadimitratos, J.-P. Hubaux, and J.-Y. Le Boudec, "Distance bounding with ieee 802.15. 4a: Attacks and countermeasures," *IEEE Transactions on Wireless Communications*, vol. 10, no. 4, pp. 1334–1344, 2011.

[24] M. Singh, P. Leu, and S. Capkun, "UWB with pulse reordering: Securing ranging against relay and physical-layer attacks," in *26th Annual Network and Distributed System Security Symposium, NDSS 2019, San Diego, California, USA, February 24-27, 2019*, 2019.

[25] M. Singh, P. Leu, A. Abdou, and S. Capkun, "Uwb-ed: Distance enlargement attack detection in ultra-wideband," in *28th USENIX Security Symposium (USENIX Security 19)*. Santa Clara, CA: USENIX Association, 2019.

[26] S. Ganeriwal, C. Pöpper, S. Čapkun, and M. B. Srivastava, "Secure time synchronization in sensor networks," *ACM Transactions on Information and System Security (TISSEC)*, vol. 11, no. 4, p. 23, 2008.

[27] Y.-C. Hu, A. Perrig, and D. B. Johnson, "Packet leashes: a defense against wormhole attacks in wireless networks," in *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, vol. 3. IEEE, 2003, pp. 1976–1986.

[28] S. Capkun and J. Hubaux, "Secure positioning of wireless devices with application to sensor networks," in *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, vol. 3. IEEE, 2005, pp. 1917–1928.

[29] S. Capkun and J.-P. Hubaux, "Secure positioning in wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 2, pp. 221–232, 2006.

[30] N. Sastry, U. Shankar, and D. Wagner, "Secure verification of location claims," in *Proceedings of the 2nd ACM workshop on Wireless security*. ACM, 2003, pp. 1–10.

[31] D. Singelee and B. Preneel, "Location verification using secure distance bounding protocols," in *Mobile Adhoc and Sensor Systems Conference, 2005. IEEE International Conference on*. IEEE, 2005, pp. 7–pp.

[32] P. Bahl and V. N. Padmanabhan, "Radar: An in-building rf-based user location and tracking system," in *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 2. Ieee, 2000, pp. 775–784.

[33] D. Vasisht, S. Kumar, and D. Katabi, "Decimeter-level localization with a single wifi access point." in *NSDI*, vol. 16, 2016, pp. 165–178.

[34] H. T. T. Truong, X. Gao, B. Shrestha, N. Saxena, N. Asokan, and P. Nurmi, "Comparing and fusing different sensor modalities for relay attack resistance in zero-interaction authentication," in *Pervasive Computing and Communications (PerCom), 2014 IEEE International Conference on*. IEEE, 2014, pp. 163–171.

[35] H. Olafsdóttir, A. Ranganathan, and S. Capkun, "On the security of carrier phase-based ranging," in *International Conference on Cryptographic Hardware and Embedded Systems*. Springer, 2017, pp. 490–509.

[36] M. Poturalski, Mand Flury, P. Papadimitratos, J. Hubaux, and J. Le Boudec, "The cicada attack: degradation and denial of service in ir ranging," in *Ultra-Wideband (ICUWB), 2010 IEEE International Conference on*, vol. 2. IEEE, 2010, pp. 1–4.

[37] Task Group. 4z, "IEEE 802.15 WPAN "enhanced impulse radio"," http://www.ieee802.org/15/pub/TG4z.html, [Online; Accessed 22. October 2018].

[38] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.

[39] A. Lapidoth, *A foundation in digital communication*. Cambridge University Press, 2017.

Appendix A. Lessons Learned

Since our proposal provides a fundamentally secure physical-layer building block for ToA measurement, this changes prevailing assumptions on the design of higher level protocols. In [21], Clulow et al. put forward principles for distance bounding:

> "We propose a number of principles to adhere to when implementing distance-bounding systems. These restrict the choice of the communication medium to speed-of-light channels, the communication format to single bit exchanges for timing, symbol length to narrow (ultra wideband) pulses, and protocols to error-tolerant versions. These restrictions increase the technical challenge of implementing secure distance bounding."

Given the above designs, these recommendations don't seem to hold up. Also, the recommendation to single bit timing is not only not needed but also are fairly wasteful. Again, from [21]:

> "We show that proposed distance-bounding protocols of Hu, Perrig, and Johnson (2003), Sastry, Shankar and Wagner (2003), and Čapkun and Hubaux (2005, 2006) are vulnerable to a guessing attack where the malicious prover preemptively transmits guessed values for a number of response bits."

In this work, we show that these vulnerabilities are an artifact of a somewhat naively designed physical layer and modulation, and can be addressed purely on the physical layer. The problem and its solution are orthogonal to the design and security of the protocol that builds in it, which operates on a different level of abstraction, making different assumptions about the security parameter (i.e., the bits) making up the nonces. These protocols not addressing the physical layer does not mean they are necessarily vulnerable. If coupled with a physical layer that is in line with our design, they are secure, within the performance region we point out.

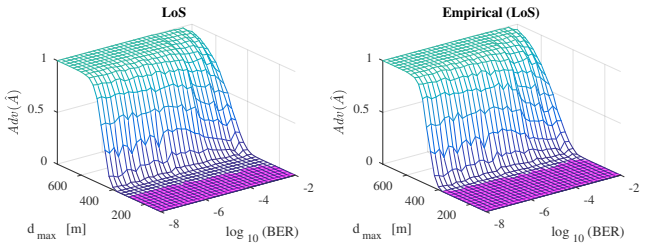

Figure 15. Attacker's advantage as a function of the performance level for LoS conditions under a Gaussian extrapolation (left) and fully empirical simulation (right). Overall, the empirical result is very similar, in particular its MTAC region is not smaller than the one resulting from the Gaussian model.

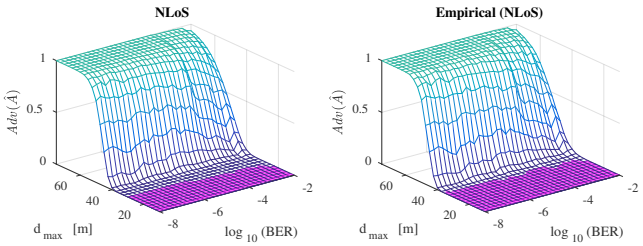

Figure 16. Attacker's advantage as a function of the performance level for NLoS conditions under a Gaussian extrapolation (left) and fully empirical simulation (right). Overall, the empirical result is very similar, in particular its MTAC region is not smaller than the one resulting from the Gaussian model.

Appendix B. Attacker Statistical Parameters

We choose the attacker's parameters under minimization of the worst-case distortion, i.e., as

$$(\hat{\mu}_{\mathcal{D}_{\hat{A}}}, \hat{\sigma}_{\mathcal{D}_{\hat{A}}}) = (\mu_{\mathcal{D}_{\hat{A}}}(d_{\hat{A},ideal}), \sigma_{\mathcal{D}_{\hat{A}}}(d_{\hat{A},ideal}))$$

$$d_{\hat{A},ideal} = \underset{d\in[0,d_{max}]}{\arg\min}\ \mu_{\mathcal{D}_{\hat{A}}}(d) - \sigma_{\mathcal{D}_{\hat{A}}}(d),$$

and the parameters of the legitimate transmitter under maximization of the distortion, within the defined performance region, i.e., as

$$(\hat{\mu}_{\mathcal{D}_{lgt}}, \hat{\sigma}_{\mathcal{D}_{lgt}}) = (\mu_{\mathcal{D}_{lgt}}(d_{lgt,worst}), \sigma_{\mathcal{D}_{lgt}}(d_{lgt,worst}))$$

$$d_{lgt,worst} = \underset{d\in[0,d_{max}]}{\arg\max}\ \mu_{\mathcal{D}_{lgt}}(d) + \sigma_{\mathcal{D}_{lgt}}(d).$$

Appendix C. Validating the Gaussian Variance Model

In the following, we motivate the Gaussian model for the distortion distribution put forward in Equations 2 and 3.

C.1. Extrapolation vs. fully empirical results

In the following, we compare our extrapolated results from Section 7 to a fully empirical (i.e., Monte-Carlo) simulation. The probability of winning as a function of the performance level is shown in Figure 15 for LoS conditions and Figure 16 for NLoS conditions. Both results refer to a frame of 20 bits. For both scenarios, we see that the the attacker's advantage evolves almost identically. We see that the fully empirical results indicate a slightly wider MTAC region, which suggest our Gaussian model to be a conservative estimate.

C.2. Variance distribution is sufficiently Gaussian

We provide quantile-quantile (QQ) plots that compare the empirical distributions against normal distributions. This allows to validate the model we use in Section 7 which serves extrapolate the empirical classification performance to small likelihoods. We provide those plots for a frame of 32 bits and a selection of communication distances, both for LoS and NLoS scenarios. Figure 17 presents those results for the attacker's variance distribution. The relevant distance for the resulting MTAC region boundary is around 100m for LoS and around 10m for NLoS. This is the distance at which the distortion for the attacker is minimal, see Figure 13. There is a slight downwards bend of the empirical value for higher quantiles. This means, a slightly bit more than expected high-variance outliers compared to the Gaussian hypothesis. This is in line with our requirements, i.e., the normal estimate being conservative regarding distinguishability. The plots for those distances show that the empirical quantiles are well aligned with the straight diagonal. Figure 18 presents those results for the attacker's distortion distribution. The relevant distance for the resulting MTAC boundary is around 200m for LoS and around 20m for NLoS, i.e., mid-range. The plots for those distances show that the empirical quantiles are well aligned with the straight line at those distances relevant for the MTAC region derived in Section 7.

Appendix D. Effect of Frame Length

Figure 19 shows the security level for one particular distance as a function of the frame length. It becomes evident that bit-level equivalence of the security level is maintianed as the length of the frame increases. We see that attacker's advantage decays faster than 2^{-n_b}.

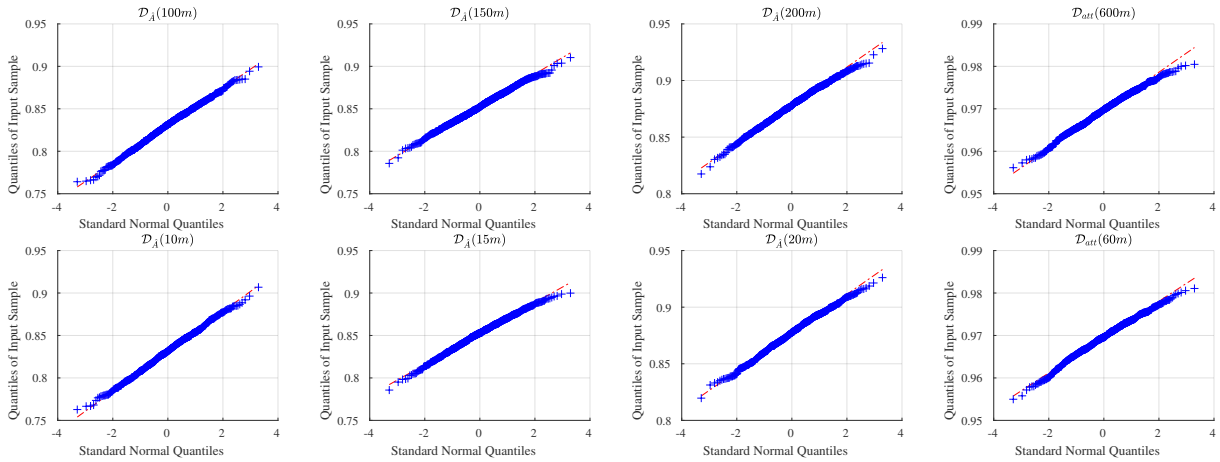

Figure 17. QQ plots comparing the attacker empirical distortion distribution for LoS (top) and NLoS (bottom) conditions for a frame of 32 bits and different distances against a normal distribution. For validity of results w.r.t. the MTAC region boundary, the attack signal distortion at a distance of 100m (LoS) and 10m (NLoS) should be close to a Gaussian. Indeed, the QQ plots of the second column are close to the diagonal.

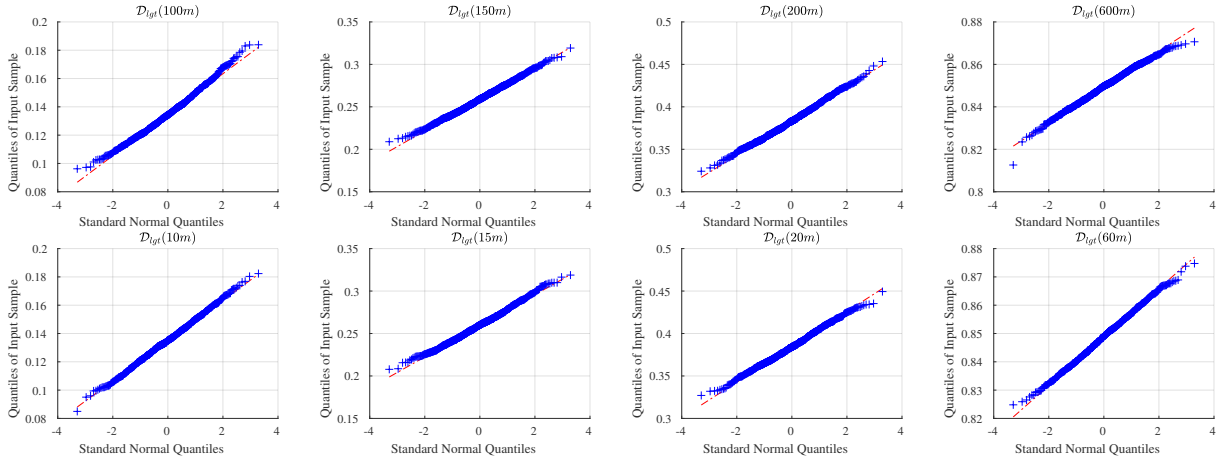

Figure 18. QQ plots comparing the legitimate empirical distortion distribution for LoS (top) and NLoS (bottom) conditions for a frame of 32 bits and different distances against a normal distribution. For validity of results w.r.t. the MTAC region boundary, the attack signal distortion at a distance of 200m (LoS) and 20m (NLoS) should be close to a Gaussian. Indeed, the QQ plots of the third column are close to the diagonal.

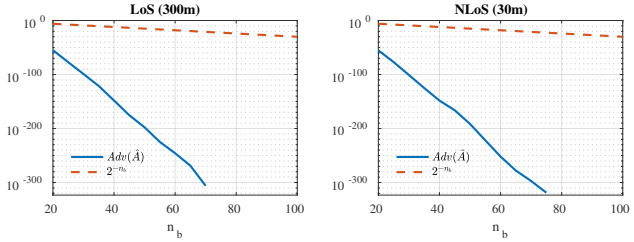

Figure 19. The attacker's advantage decreases faster than the bit-equivalent MTAC security level for longer frames. This means the security guarantees for shorter frames are maintained for longer frames.