Provably Secure Distance-Bounding: an Analysis of Prominent Protocols

Marc Fischlin Cristina Onete

Darmstadt University of Technology & CASED, Germany
www.cryptoplexity.de

Abstract. Distance-bounding protocols prevent man-in-the-middle attacks by measuring response times. Recently, Dürholz et al. [10] formalized the four attacks such protocols typically address: (1) mafia attacks, where the adversary must impersonate to a verifier in the presence of an honest prover; (2) terrorist attacks, where the adversary gets some offline prover support to impersonate; (3) distance attacks, where provers claim to be closer to verifiers than they really are; and (4) impersonation security, where adversaries impersonate provers during lazy phases. Dürholz et al. [10] also formally analyzed the security of (an enhanced version of) the construction of Kim and Avoine [14].

In this paper, we quantify the security of some other well-known distance-bounding protocols, i.e.: Brands and Chaum [6], Hancke-Kuhn [13], Avoine and Tchamkerten [4]; Reid et al. [16], the Swiss-knife protocol [15], and the very recent proposal of Yang, Zhuang, and Wong [18]. In particular, our main results show that (1) relating responses to a long-term secret key, as is the case for most protocols aiming to thwart terrorist fraud attacks, may make protocols vulnerable to so-called key-learning mafia fraud attacks, where the adversary learns a key bit-by-bit, by flipping a single time-critical response; (2) though relating responses can be a bad idea for mafia fraud, it sometimes enforces distance-fraud resistance, by thwarting in particular the attack of Boureanu et al. [5]; (3) none of the three allegedly terrorist-fraud resistant protocols, i.e. [15, 16, 18], is in fact terrorist fraud resistant; for two of these protocols this is a matter of syntax, i.e. they do not meet the strong security requirements given by Dürholz et al.; the attack against the third protocol, i.e. [18], however, is almost trivial; (4) due to the absence of a second authentication phase, the protocol of Yang, Zhuang, and Wong is vulnerable to Denial of Service attacks. In light of our results, we also review definitions of terrorist fraud, arguing that, while the strong model in [10] may be at the moment more appropriate than the weaker intuition, it may in fact be too strong to capture terrorist fraud resistance.

Keywords. distance-bounding, location privacy

1 Introduction

Designed by Brands and Chaum [6], distance-bounding protocols address man-in-the-middle (MITM) relay attacks against authentication protocols, also called mafia fraud by Desmedt [9]. Essentially, distance-bounding enhances authentication such that the verifier accepts if the prover authenticates *and* it can prove it is within a pre-set distance of the verifier (here associated with a clock). Classical distance-bounding protocols consist of lazy phases (not using a clock) and time-critical phases (where the verifier measures the time elapsed during a challenge-response phase). As relaying causes processing delays, pure MITM relay is now detected by the clock.

Though there are many distance-bounding protocols to be found in the literature, only two more thorough approaches exist towards modelling their security. Both the earlier framework of Avoine et al. [2] and Dürholz et al. [10] formalized four attacks against distance-bounding protocols, particularly for Radio Frequency Identification (RFID) systems. In this setting, provers are so-called RFID tags, and verifiers are RFID readers. The models differ slightly, though they cover the same attacks; in this paper we use the more recent framework of [10], which defines the following main attacks.

Mafia Fraud. The adversary impersonates to the reader while communicating with the genuine tag. Here the clock prevents pure relaying.

Terrorist Fraud. The tag helps the adversary authenticate by disclosing useful information in offline phases. However, the tag should not reveal trivial information like the secret key.

Distance Fraud. The (malicious) tag claims to be closer to the reader than it actually is.

Offline Impersonation Resistance. The adversary impersonates the honest tag during lazy phases only.

We note that though Avoine et al. [2] were the first to actually *formalize* security notions in distance-bounding protocols, the *terminology* itself was coined by first coined by Desmedt [9]. Furthermore, note that the fourth property, offline impersonation security was introduced by Dürholz et al. particularly for resource-constrained RFID tags, which cannot support many time-critical phases. The alternative description of Avoine et al. [2] considers impersonation resistance to be the equivalent of offline impersonation resistance combined with time-critical impersonation resistance (i.e. mafia fraud resistance).

Recently, Cremers et al. [8] also introduced a new attack, called *distance hijacking*. This attack, however, involves two provers, one honest and one malicious. Since we use the formal, single-reader-single-tag framework of Dürholz et al., this attack falls outside the scope of this paper.

1.1 RFID Distance-Bounding Protocols

One reason we choose the framework of Dürholz et al. [10] is that the model takes into account the characteristics of RFID distance bounding (though it is also suitable for general settings). This work also quantified and proved the security properties of an enhancement of the well-known protocol due to Kim et al. [14], also meant for RFID environments. On the one hand, such formal analysis can prevent security breaches like those outlined by Abyneh [1], who showed attacks against two allegedly secure schemes. On the other hand, analyzing security properties in a common, formal framework enables easier comparison between protocols.

In this paper continue the assessment work of [10] by quantifying the security of several known RFID distance-bounding protocols, with quite surprising results. We in fact show a few subtle kinks in various protocols, which compromise security.

In general, distance-bounding protocols consist of a number of slow (lazy) phases, followed by a number N_c of time-critical phases where a clock is used by the reader to time bit-exchanges between itself and the tag. Whereas lazy phases are usually computationally expensive, time-critical phases are expensive in terms of communication complexity; in fact, in RFID distance-bounding scenarios it is unclear how many time-critical phases resource-constrained devices can even support.

Since time-critical responses only typically consist of bit-exchanges, the best known mafia fraud resistance (for N_c time-critical rounds) is about $\frac{1}{2}$ per time-critical round, thus $\left(\frac{1}{2}\right)^{N_c}$ in total. This bound was first achieved by Brands and Chaum [6]. However, [6] use computationally-expensive signature schemes, unsuitable for resource-constrained RFID tags. The same resistance is less expensively achieved by the Swiss-knife protocol [15], at the cost of a second lazy phase, and it is nearly achieved by the protocol

of Avoine and Tchamkerten [4], though at the cost of exponential storage. The very efficient, well-known Hancke-Kuhn protocol [13] achieves only a resistance of $\left(\frac{3}{4}\right)^{N_c}$, being vulnerable to the so-called *Go-Early strategy*, where the mafia adversary queries the tag in advance; thus it learns about half of the correct time-critical responses and can guess the other responses (see the proof in Section 3.1). The same bound holds for the very recent protocol of Yang, Zhuang, and Wong [18], which also manages to thwart so-called *key-learning mafia fraud* attacks. By contrast, the protocol due to Reid et al. [16] *is* susceptible to key-learning attacks, and so is *not* mafia fraud resistant. It is worth noting that the more efficient symmetric distance-bounding protocols do not, in view of the recent *attack of Boureanu et al. [5]*, achieve *distance fraud resistance*. We describe this attack in detail in the following sections.

Some protocols in the literature also address terrorist fraud. An early scheme due to Reid et al. [16] using a PRF and a symmetric encryption scheme claims to achieve terrorist fraud resistance, since an adversary that knows both the PRF output and the symmetric encryption can recover the secret key. Also addressing terrorist fraud is the Swiss-knife protocol of Kim et al. [15] and the very recent proposal due of Yang et al. [18]. However, we show that, whereas the terrorist fraud proof for the latter protocol is simply *flawed*, the former two protocols are also *susceptible to a terrorist fraud attack* where the adversary learns *some* information about the long-term secret, but cannot use this information directly at full efficiency. This latter attack is more a question of *syntax*, in the sense that it uses the characteristics of the terrorist fraud model of Dürholz et al., in particular the disadvantages inherent to the simulator. Both protocols seem to intuitively achieve a measure of terrorist fraud resistance; however, it seems hard to prove them secure. We discuss the relation between model strength and intuitive security at length in Appendix A. In our terrorist fraud attack the prover hands over half of the correct responses and lets the adversary guess the other responses. This gives the adversary an advantage in sessions when the prover helps; once the prover stops helping though, the probability drops.

Notably, although these protocols also use pseudorandom functions, the PRF attack of Boureanu et al. [5] does not necessarily hold here. Thus, the protocol due to Reid et al. [16] computes only one of the time-critical responses by means of the PRF, while the other response is an encryption of the first response under a long-term secret key. Reid et al. use a generic symmetric encryption scheme, which does *not* automatically grant distance fraud resistance; however some particular instantiations of such schemes, such as one-time-pad encryption, *do* grant distance-fraud resistance, possibly under further assumptions, e.g. the fact that the secret key is chosen by the reader or by a trusted party in an honest fashion, i.e. uniformly at random, from a distribution computationally indistinguishable from the uniform random distribution (of appropriate length). Under the same assumption, the Swiss-Knife protocol [15] is distance-fraud resistant (see below). The crucial fact is that the time-critical responses are *not* both computed by means of a PRF, and they are related by means of an honestly-chosen pseudorandom value. Informally speaking, a distance-fraud adversary may now tamper with the PRF output, but not with both responses.

Finally, following the sequence of subtle attacks we present in this paper, we show that the protocol due to Yang, Zhuang, and Wong [18] is *not private*. In particular, we disprove claims by the authors that the protocol is immune to *desynchronization Denial-of-Service (DoS) attacks*.

1.2 Our Contributions

In this paper we quantify the security of the following protocols: Brands-Chaum [6], Hancke-Kuhn [13], Avoine and Tchamkerten [4], Reid et al. [16], Kim et al. [15], and Yang et al. [18]. For each protocol, we (1) show concrete security bounds for the properties that actually hold, thus (2) disproving claims of distance-fraud resistance for the protocols of Hancke and Kuhn, Avoine and Tchamkerten, and Yang et al. (the attack is a direct consequence of the generic attack of Boureanu et al. [5]), (3) disproving claims of terrorist fraud resistance for all the three allegedly terrorist fraud resistant schemes (though

the attack against [16] and [15] is rather a technical attack, specific to the simulation-based model of Dürholz et al.), (4) disproving claims of mafia fraud resistance for the protocol due to Reid et al. [16] (we show a key-learning mafia fraud attack against this protocol), (5) disprove claims of resistance to Denial-of-Service desynchronization attacks for the protocol of Yang et al.

Let us take a closer look at our result (3). Apart from the flaw in the proof of Yang et al. [18], we show that, though they are intuitively terrorist-fraud resistant, the allegedly terrorist fraud resistant schemes of [15] and [16] cannot, in fact, be *proved* to be terrorist fraud resistant in the model of Dürholz et al. The question is thus whether we are in more need of stronger protocols, or mor exact model. Our attack takes into account the syntax of the framework in [10], and uses the fact that —by yielding relatively information about a long-term secret key— the malicious tag can (1) help the adversary authenticate and, at the same time, (2) diminish its probability of later authenticating on its own. Similarly, we use the syntax of the model in [10] to break the terrorist fraud resistance of [15]. In particular, the tag can even hand over its secret key, as long as it doesn't forward its secret identifier (this will enable the adversary to authenticate with the prover's support, but not without it). The question of whether these attacks are a weakness of the protocols or of the terrorist fraud model of Dürholz et al. is discussed in the appendix.

In particular, we compare with the intuition behind terrorist fraud with the strong model of Dürholz et al. Our analysis seems to indicate that the problem lies somewhere in the middle, between the strength of the model and the security of the protocol. Our attack against the scheme of Reid et al. assumes that the malicious tag will agree to yield some *partial* information about the secret key in order to aid the adversary, but not if this information is directly useful to the adversary in future authentication sessions. However, note that the attack is in itself somewhat impractical, or flawed, in the sense that (a) when the tag helps, the adversary's success probability is not overwhelming (merely significantly larger than guessing probability); (b) when the tag stops helping the adversary has a *lower* probability to succeed than before, but the probability is not insignificant. Thus, it seems the model of Dürholz et al. is exaggeratedly strong. Note, however, that at the moment, this framework gives the only thorough formalization of terrorist fraud resistance in the literature. The initial framework and later work of Avoine et al. [2], resp. [3] seems to indicate information-theoretical hiding properties are required for the secret key shared between the reader and the tag. Note, however, that this seems too strong a restriction: intuitively, an adversary may learn some information about the key without gaining any advantage in future attacks. We leave two open questions regarding this issue: (1) Should we aim to design protocols which are provably secure in the framework of Dürholz et al.?; and (2) Is it possible to capture terrorist fraud resistance better in a different model?

The results of our protocol analysis are summarized in Table 1, where we show (only the loose) security bounds for the attacks described above, ignoring the small terms. A precise quantification can be found in Section 3.

2 Preliminaries

2.1 Model Overview

The security model in [10] considers single-verifier-single-prover distance-bounding, particularly for RFID settings; here, a single prover (tag) $\mathcal{T}$ and a single verifier (reader) $\mathcal{R}$ share a secret key sk generated by an algorithm Kg. We explicitly require that the key is selected at random from the space of all keys, which we denote K. In particular, we require that the tag does not choose its own key (this will later ensure that the protocols are distance fraud resistant). The reader uses a clock to measure the time elapsed between sending a challenge and receiving the response. Dürholz et al. consider round-based distance-bounding protocols, where rounds are *time-critical* if the clock is used, and *lazy* otherwise. We

	Mafia	Terror	Distance	Impersonation
[6] [1]	$\left(\frac{1}{2}\right)^{N_c}$	×	×	$\left(\frac{1}{2}\right)^{N_c}$
[13]	$\left(\frac{3}{4}\right)^{N_c}$	×	×	×
[4] [2]	$\frac{1}{2}(N_c+2)\left(\frac{1}{2}\right)^{N_c}$	×	$\left(\frac{3}{4}\right)^{N_c}$	$\left(\frac{1}{2}\right)^{\lvert V\rvert}$
[16] [3]	×	×	$\left(\frac{3}{4}\right)^{N_c-T}$	×
[15]	$\left(\frac{1}{2}\right)^{N_c-T}$	×	$\left(\frac{3}{4}\right)^{N_c-T}$	$\left(\frac{1}{2}\right)^{\lvert V\rvert}$
[18]	$\left(\frac{3}{4}\right)^{N_c}$	×	×	×

Figure 1: Distance Bounding at a glance. [1] This protocol uses expensive primitives. [2] This protocol requires exponential storage capacity. [3] The distance fraud proof for this protocol holds only for some instantiations of the protocol. We denote by N_c the number of time-critical rounds, by T a tolerance level for faults, and by $|V|$ is the bit length of an authentication string V sent by the verifier.

briefly review here the main possible attacks and refer to [10] for more details.

Mafia Fraud. Mafia fraud resistance is a man-in-the-middle
(MITM) attack, where pure relay is prevented by the reader's clock. Informally, the MITM consists of two adversaries: a *leech*, which impersonates the reader to an honest tag, and a *ghost*, which impersonates the tag to an honest reader. The adversary must authenticate to the reader; however, the clock detects processing delays in the MITM resulting from pure relay. Both the reader and the tag are unaware of the MITM attack.

Dürholz et al. [10] formalize mafia fraud by introducing an abstract clock denoted clock, which keeps track of the messages sent in several protocol executions called sessions. These sessions can be: reader-tag (the adversary is a passive eavesdropper), reader-adversary (the adversary impersonates the tag to the reader), and adversary-tag (the adversary impersonates the reader to the tag). Relaying is considered round-wise (actually, [10] considers phases consisting of several rounds, in case the reader measures time over many rounds). A phase is called *tainted* if the adversary *purely* relays communication between a reader-adversary and an adversary-tag session. Here, *pure* relay refers to an adversary receiving a message in a session sid and *then* relaying the exact, same message in a session sid'. Having received a response, the adversary relays it back again between sid' and sid, for all subsequent rounds in the tainted phase. If the adversary changes any of the messages in one session before it forwards them in the other session, this is not pure relay. Also, if the adversary queries one session with some message m *before* receiving the same m in the other session, this is not relaying. We refer to [10] for a precise definition of, and motivation behind tainted phases, and show the following illustration, also from [10]:

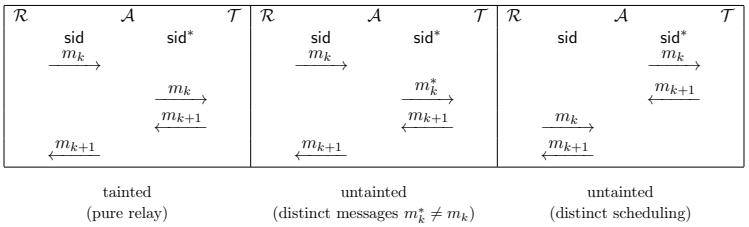

Figure 2: Tainted and Untainted Time-Critical Phases.

Terrorist Fraud. A terrorist adversary must also authenticate to the reader, this time aided by the dishonest tag, which should not give information allowing the adversary to win on its own. In [10],

the dishonest tag only interacts with the adversary in lazy phases[1]. A time-critical phase of a reader-adversary session is tainted if during this phase the adversary queries the tag.

A crucial part of terrorist fraud definition is formalizing how much a tag can help the adversary. In [10], terrorist fraud is defined in terms of a simulator: a scheme is terrorist fraud resistant if an adversary aided offline by a tag is as likely to win as a simulator having access to the adversary's internal information. To be fair, the simulator has as many impersonation attempts as the adversary, but it only begins its attack once the adversary successfully authenticates. If the tag only relays session-specific data, then this data is useless to the simulator. If, however, the data contains the secret key, then the simulator can recover it and also succeed. Formally, for every successful terrorist fraud adversary $\mathcal{A}$, there must exist a simulator $\mathcal{S}$ whose success probability is at least as large as the adversary's. As we discuss later and in the Appendix, this definition is very strong; in fact in light of our results, it seems it may be *too* strong. However, one advantage of this definition is that it outlines a formal proof technique for terrorist fraud resistance. On the other hand, other approaches in the literature, e.g. [3] seem to require that *no* information is leaked about the secret key, which seems to restrict the attack too much, i.e. the model appears to be too weak.

Distance Fraud. Distance fraud adversaries are the malicious tags themselves, who are farther than allowed from the reader, but aim to convince the reader of the contrary. Since the reader's clock measures time accurately, the adversary must anticipate the reader's challenges and respond in advance. In [10], the adversary must commit in advance to each time-critical phase response. If the adversary does not commit to the response of one phase, this phase is called tainted. For distance fraud, reader-tag and adversary-tag sessions are no longer relevant as the adversary *is* the tag.

(Offline) Impersonation Resistance. Finally, offline impersonation resistance refers to lazy-phase tag authentication. The idea, introduced by [4], is that even without the time-critical phases, the prover is still authenticated. Replays, however, are still allowed at this stage. We refer to the definition in [10] for the formalization of this notion. We note that the notion of impersonation resistance due to [10] refers *only* to lazy phases, whereas Avoine et al. [2] defines impersonation security for the entire protocol. We call the notion of [10] *offline* impersonation resistance in order to emphasize the difference between the two notions. We note that, for mafia fraud resistant protocols, the total impersonation security equals the lazy phase impersonation security of Dürholz et al. together with the mafia fraud resistance during time-critical phases.

2.2 Parameters

Apart from the upper-bound $t_{\max}$ of the roundtrip transmission time and the number of time-critical rounds N_c, [10] also allows for max. $T_{\max}$ phases with delayed responses and max. $E_{\max}$ phases with wrong responses. Though most existing protocols do not provide for erroneous/delayed communication, fault tolerance is essential in resource constrained environments, e.g. RFID.

When specifying the adversary's characteristics we consider its runtime t (including honest party processing time, where the adversary waits for honest responses) and the numbers $q_{\mathcal{R}}$, $q_{\mathcal{T}}$, resp. q_{OBS} of reader-adversary, adversary-tag, and resp. reader-tag sessions. To relate the distance bounding security levels to the security of the underlying cryptographic primitives, we often transform a successful distance-bounding adversary $\mathcal{A}$ into one or more adversaries $\mathcal{A}'$ against the primitive(s). We use standard notions for these primitives, letting $\mathbf{Adv}_{\mathcal{S}}^{\text{Exp}}(\mathcal{A}')$ denote the maximal probability that an adversary $\mathcal{A}'$ breaks a cryptographic scheme $\mathcal{S}$ in some experiment Exp. For example, $\mathbf{Adv}_{Sign}^{\text{Unf}}(\mathcal{A}')$ denotes the probability of

[1] We note that allowing the prover to only help the adversary offline is in agreement with the previous, more informal model of Avoine et al. [2] and with the general intuition behind terrorist fraud attacks.

forging signatures, $\mathbf{Adv}^{\mathrm{d}}_{\mathsf{PRF}}(\mathcal{A}')$ denotes the advantage of distinguishing a pseudorandom function PRF from random, and $\mathbf{Adv}^{\text{IND-CPA}}_{\mathcal{E}}(\mathcal{A}')$ is the advantage of breaking the IND-CPA security in encryption. The parameters of $\mathcal{A}'$ in these experiments will be specified in terms of the parameters of $\mathcal{A}$.

Flavours of unforgeability. Most distance-bounding protocols in the literature use pseudorandom functions. However, the protocol due to Brands and Chaum [6] uses signature schemes instead. A signature scheme is defined as usual as a triplet of algorithms $Sign = (\mathsf{SKg}, \mathsf{SSign}, \mathsf{SVf})$ such that, on input a security parameter (in unary) 1^k, the key-generation algorithm SKg outputs a private-public key pair (sk, pk); on input a message m and the secret key sk, the signing algorithm $\mathsf{SSign}(sk, m)$ outputs a signature σ; and on input a message m, a signature σ, and the public key pk, the verification algorithm $\mathsf{SVf}(pk, m, \sigma)$ outputs a bit indicating whether the signature verifies (the output bit is then 1), or does not verify (the output bit is 0).

The security of signature schemes is usually defined in terms of unforgeability, which intuitively captures the fact that adversaries against the signature scheme should not be able to forge signatures for "fresh" message m even if they have the possibility to query correct signatures for arbitrary messages of its choice (but these messages should be different from m for the usual notion of existential unforgeability). More formally, let $\mathsf{Sign}(sk, m)$ be an oracle that, for a secret key sk, and a message m, outputs the signature $\mathsf{SSign}(sk, m)$. We quantify adversaries against the unforgeability of the signature scheme in terms of the runtime t and the number of queries Q to the Sign oracle, and we define unforgeability as follows.

Definition 2.1 *Let* $\mathrm{Sign} = (\mathsf{SKg}, \mathsf{SSign}, \mathsf{SVf})$ *be a signature scheme as above. Let* $\mathcal{A}$ *be an adversary against the unforgeability of* Sign*, running in time* t *and making at most* Q *queries to the signing oracle* Sign *(see the experiment below). Then the forging advantage of* $\mathcal{A}$ *is defined as*

$$\boldsymbol{Adv}^{Unf}_{\mathrm{Sign}}(\mathcal{A}) = \mathrm{Prob}\left[\mathsf{Exp}^{\mathrm{Sign}}_{\mathcal{A}}(1^k) = 1\right],$$

where $\mathsf{Exp}^{\mathrm{Sign}}_{\mathcal{A}}(1^k)$ *is defined as follows.*

Experiment $\mathsf{Exp}^{\mathrm{Sign}}_{\mathcal{A}}$*:*
$(pk, sk) \leftarrow \mathsf{SKg}(1^k)$
$(m, \sigma) \leftarrow \mathcal{A}^{\mathsf{Sign}(sk,\cdot)^Q}(pk)$
The experiment outputs 1 *if (a)* $\mathsf{SVf}(pk, m, \sigma) = 1$ *and (b) the message* m *was not previously queried to* Sign*.*

For our security proof in the case of the Brands and Chaum protocol (see Section 3.0.1), we also require the property of *strong* unforgeability, where an adversary can also output a tuple (m, σ) with m previously queried to Sign, under the condition that σ is not the signature that the oracle Sign forwarded to the adversary. In fact, the definition is the same as in the case of existential unforgeability, but the security game is modified as follows.
Experiment $\mathsf{Exp}^{*\,Sign}_{\mathcal{A}}$:
$(pk, sk) \leftarrow \mathsf{SKg}(1^k)$
$(m, \sigma) \leftarrow \mathcal{A}^{\mathsf{Sign}(sk,\cdot)^Q}(pk)$
Let $L = \{m_i, \sigma_i\}_{i=1}^{Q}$ be the list of queries and responses to the signing oracle. The experiment outputs 1 if (a) $\mathsf{SVf}(pk, m, \sigma) = 1$ and (b) the tuple $(m, \sigma) \notin L$.

3 Protocol Assessment

In this section, we analyze the resulting properties of the following distance-bounding protocols: Hancke and Kuhn [13], Avoine and Tchamkerten [4], Reid et al. [16], and the Swiss-Knife protocol [15]. We note that in view of the results of Boureanu et al. [5], these protocols do not achieve distance fraud resistance. In our analysis, we show that provable terrorist fraud resistance is harder to attain than intuition indicates.

Also, most earlier distance-bounding protocols do not allow for fault tolerance. In other words, the assumption is that the prover's behavior is always constant and that both transmissions and transmission times are reliable and constant. Thus the assumption is that the verifier's challenges always arrive at the prover (in constant time), the prover's responses also always arrive at the verifier (in constant time), and the prover always has the same processing delay. This is not always the case, particularly not for resource constrained devices like RFID, where transmissions are not always reliable, and transmission times and processing delays may vary. In [10], Dürholz et al. introduce fault tolerance parameters as outlined in Section 2.2.

3.0.1 Brands and Chaum

In the Brands and Chaum construction [6], the reader and tag first exchange random bits in N_c time-critical phases, then finally signs the concatenation of these bits under the shared secret key sk — see Figure 3. Though in this case the lazy phases do not pre-date the time-critical phases, the concept of impersonation resistance is still applicable. We only require that the adversary can forward a fresh signature in a winning reader-adversary session, in the sense that the adversary has not seen it before. As the lazy phase has a single message, the only possibly relayed message here is the signature.

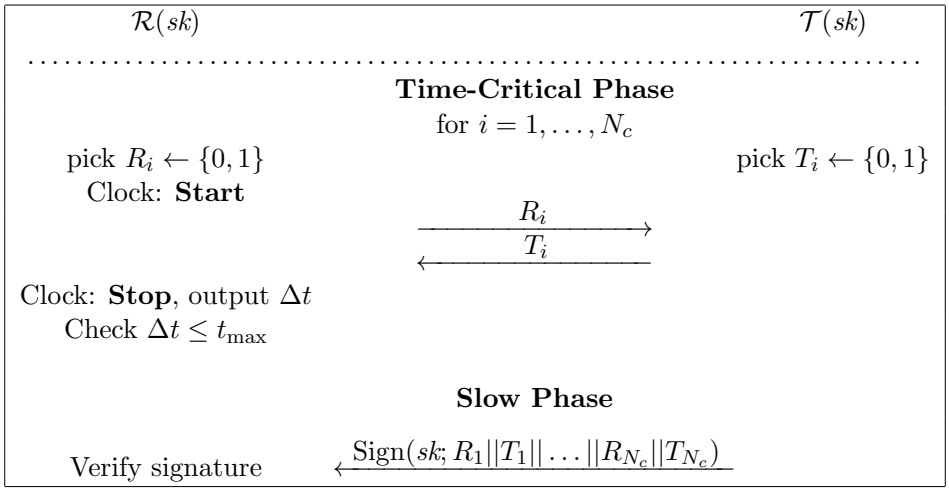

Figure 3: The Brands and Chaum protocol

Intuitively, the time-critical phases of this protocol must ensure here that the reader-to-tag distance is no greater than the one associated with $t_{\max}$. Finally, the signature in the lazy phase must ensure that the bit exchange was done by a legitimate prover. This ensures both impersonation resistance and a measure of mafia fraud resistance. Lazy phases are susceptible to pure relay (which otherwise taints time-critical rounds), thus the signature yields neither terrorist, nor distance fraud resistance. Unlike most other distance bounding protocols, however, the tag's responses are fully random, thus the difficulty in mafia fraud attacks is not to answer the time-critical rounds, but rather to make the tag generate the correct signature at the end.

We formally state the properties of this protocol below, without considering any further fault tolerance. If the parameters $E_{\max}$ and $T_{\max}$ are considered, upper-bounding the number of erroneous transmissions and respectively the number of transmissions exceeding $t_{\max}$, the security bound for mafia fraud resistance drops by a factor $2^{T_{\max}}$. Note, however, that should *any* of the transmissions go wrong, the signature will not longer verify: for scenarios where transmissions are not reliable, the tag should append the values $R_1, \ldots, R_{N_c}, T_1, \ldots, T_{N_c}$ to the signature.

Theorem 3.1 (Brands-Chaum Properties) *Let* $\text{Sign} = (\mathsf{SKg}, \mathsf{SSign}, \mathsf{SVf})$ *be the signature scheme used above. The Brands and Chaum protocol* ID *above has the following properties:*

- *It is not resistant to terrorist, nor to distance fraud.*
- *For any* $(t, q_{\mathcal{R}}, q_{\mathcal{T}}, q_{\text{OBS}})$*-impersonation adversary* $\mathcal{A}$ *against the Brands Chaum protocol, there exists an adversary* $\mathcal{A}'$ *against the strong unforgeability of* Sign *that runs in time* t *and requests at most* $q_{\mathcal{T}} + q_{\text{OBS}}$ *signatures, and that then outputs a valid forgery with an advantage* $\boldsymbol{Adv}^{imp}_{\mathsf{ID}}(\mathcal{A})$ *such that:*

$$\boldsymbol{Adv}^{imp}_{\mathsf{ID}}(\mathcal{A}) \leq \boldsymbol{Adv}^{strUnf}_{\text{Sign}}(\mathcal{A}') + q_{\mathcal{R}}(q_{\mathcal{T}} + q_{\text{OBS}}) \cdot 2^{-N_c}.$$

- *For any* $(t, q_{\mathcal{R}}, q_{\mathcal{T}}, q_{\text{OBS}})$*-mafia-fraud adversary* $\mathcal{A}$ *against the Brands Chaum protocol, there exists an adversary* $\mathcal{A}'$ *against the unforgeability of* Sign *that runs in time* t *and requests at most* $q_{\mathcal{T}} + q_{\text{OBS}}$ *signatures, and that then outputs a valid forgery with an advantage* $\boldsymbol{Adv}^{Unf}_{\text{Sign}}(\mathcal{A}')$ *such that:*

$$\boldsymbol{Adv}^{mafia}_{\mathsf{ID}}(\mathcal{A}) \leq \boldsymbol{Adv}^{Unf}_{\text{Sign}}(\mathcal{A}') + q_{\mathcal{R}}(q_{\mathcal{T}} + q_{\text{OBS}}) \cdot 2^{-N_c} + q_{\mathcal{R}} \cdot 2^{-N_c}.$$

Proof. The second statement is easily proved. Consider a reader-adversary session sid where an adversary $\mathcal{A}$ successfully impersonates to the reader. Firstly, there is at most a probability of $(q_{\text{OBS}} + q_{\mathcal{T}}) \cdot 2^{-|N_c|}$ that the challenges in this verifier-adversary session coincide with the challenges R_i in a previous reader-tag or adversary-tag session. In this case, the adversary can replay the past session (including the signature received at the end), thus winning with probability 1. We account for this probability in each of the verifier-adversary sessions, and assume from now on that the string of challenges $\{R_1, \ldots, R_{N_c}\}$ is unique amongst all the past reader-tag and adversary-tag sessions.

We consider now a single reader-adversary session sid. The adversary can now open a parallel adversary-tag session sid'. The definition of impersonation security requires that the adversary does not forward lazy phase information between the two sessions. The adversary can, however, forward the time-critical phase information between the two sessions. We show that the adversary can now win with probability at most equal to the advantage of any adversary against the strong unforgeability of the signature scheme *Sign*. Indeed, we construct such an adversary $\mathcal{A}'$ which runs $\mathcal{A}$ in a black-box way. Every time $\mathcal{A}$ runs either a prover-verifier or resp. adversary-prover session, the adversary $\mathcal{A}'$ queries its signature oracle for the transcript of the protocol, and forwards the signature to $\mathcal{A}$, storing the value of $\mathsf{SSign}(sk; R_1||T_1|| \ldots ||R_{N_c}||T_{N_c})$. Finally, assuming that $\mathcal{A}$ succeeds in impersonating the prover in a session sid, the adversary $\mathcal{A}'$ forwards the tuple consisting of the time-critical transcript of the protocol, together with the signature used by $\mathcal{A}$ in its successful attempt, to the challenger in the strong unforgeability game. In case the adversary has relayed the time-critical transcript, the adversary $\mathcal{A}$ only wins if it forwards a different signature for the same message, i.e. we must require strong unforgeability. If the impersonation adversary $\mathcal{A}$ has an advantage $\mathbf{Adv}^{\text{imp}}_{\mathsf{ID}}(\mathcal{A})$, then the unforgeability adversary $\mathcal{A}'$ has a probability at least as large to succeed in its attempt. Accounting for $q_{\mathcal{R}}$ possible replays, we obtain the specified bound.

The first statement also follows easily: the adversary $\mathcal{A}$ against distance fraud simply commits to randomly chosen values of T_i, then computes the (correct) signature with its own secret key sk, and

succeeds with probability 1. For terrorist fraud, the adversary trivially generates random time-critical responses, then forwards these values and the values of R_i to the dishonest prover, and receives the signature. The success probability of this adversary is 1. However, the simulator's success probability is upper bounded by the probability that it can forge a signature for a fresh session. For an unforgeable signature scheme, the simulator's success probability is negligible, and thus the scheme is not terrorist fraud resistant.

We now prove the last statement. Consider the adversary $\mathcal{A}$ mounting a mafia fraud attack. Consider the reader-adversary session sid where $\mathcal{A}$ successfully authenticates to the reader. As in the impersonation fraud proof, there is a probability of $(q_{\mathcal{T}} + q_{\text{OBS}}) \cdot 2^{-N_c}$ that the challenges R_i in sid are all identical to the fast phase values R_i^* in one of these sessions. In this case, the adversary can simply forward the observed/received values T_i^* to the reader in session sid; at the end, the adversary forwards the observed/received value of the signature. We now assume that the verifier-adversary session sid is the only session the adversary has access to. The adversary can start a concurrent adversary-tag session sid^*. We call a round successful if the reader input R_i and the response T_i in round i of sid are the same as the adversary input R_i^* and the response T_i^* in sid^*, *without* pure relay, i.e. the adversary must send R_i^* in sid^* before receiving R_i in sid or it must guess T_i in advance, before receiving T_i^*.

We first show that, except with negligible probability, the adversary cannot win unless all time-critical rounds are successful. Indeed if one round is not successful (i.e. R_i^* in sid is different from R_i in sid^*) and yet $\mathcal{A}$ authenticates to the reader in sid, we can build an adversary against the unforgeability of the signature scheme. This adversary's forgery is the message m containing the concatenation of all the queries and responses in session sid and the signature that $\mathcal{A}$ forwards in sid. The signature must be correct – else $\mathcal{A}$ cannot authenticate —and the message is fresh, as at least in round i the value of $R_i||T_i$ is fresh. Thus the probability that $\mathcal{A}$ wins without being successful in each round is $\mathbf{Adv}^{\text{Unf}}_{Sign}(\mathcal{A}')$.

If the adversary is successful in every round, then it wins with probability 1, since in this case the time-critical transcripts are identical between sid and sid^* and the adversary can just relay the lazy-phase response. However, since the challenges and responses are independent, the probability that the adversary is successful in every round is $\left(\frac{1}{2}\right)^N_c$ for each of the reader-adversary sessions. Accounting for all reader-adversary sessions, we obtain the bound above. □

3.1 Hancke and Kuhn

The protocol due to Hancke and Kuhn addresses mafia and distance fraud, and uses a PRF implemented as HMAC. This protocol has better distance fraud resistance than e.g. the Brands and Chaum protocol [6], but at the cost of a lower mafia fraud resistance. The protocol consists of: (i) a lazy phase, where the parties exchange nonces and pre-compute an HMAC value, divided in a left and a right half, and (ii) N_c time-critical phases where the reader forwards a random bit and the tag responds with a bit either from the left or right half of response. This is also depicted in Figure 4.

The PRF output is $2N_c$ bits, and the left and right halves of this output have equal length. Note that the value $\mathsf{PRF}(sk, N_{\mathcal{R}}||N_{\mathcal{T}})$ cannot be computed by a mafia fraud adversary, without knowledge of sk. Note that for this protocol, the lack of reader authentication leads to a decreased mafia fraud resistance; however, the random challenges provide distance fraud resistance. There is no offline impersonation resistance, as we also state more formally below.

Theorem 3.2 (Hancke-Kuhn Properties) *Let* ID *be the distance-bounding authentication scheme in Figure 4 with parameters* $(t_{\max}, N_c)$*. This scheme has the following properties:*

- *It is not offline impersonation resistant, distance fraud resistant, nor resistant to terrorist fraud.*
- *For any* $(t, q_{\mathcal{R}}, q_{\mathcal{T}}, q_{\text{OBS}})$*-mafia-fraud adversary* $\mathcal{A}$ *against the scheme there exists a* (t', q')*-distinguisher* $\mathcal{A}'$ *against* PRF *(where* $\mathcal{A}'$ *runs in time* $t' = t + O(n)$ *and makes at most* $q' = q_{\mathcal{R}} + q_{\mathcal{T}} + q_{\text{OBS}}$ *queries)*

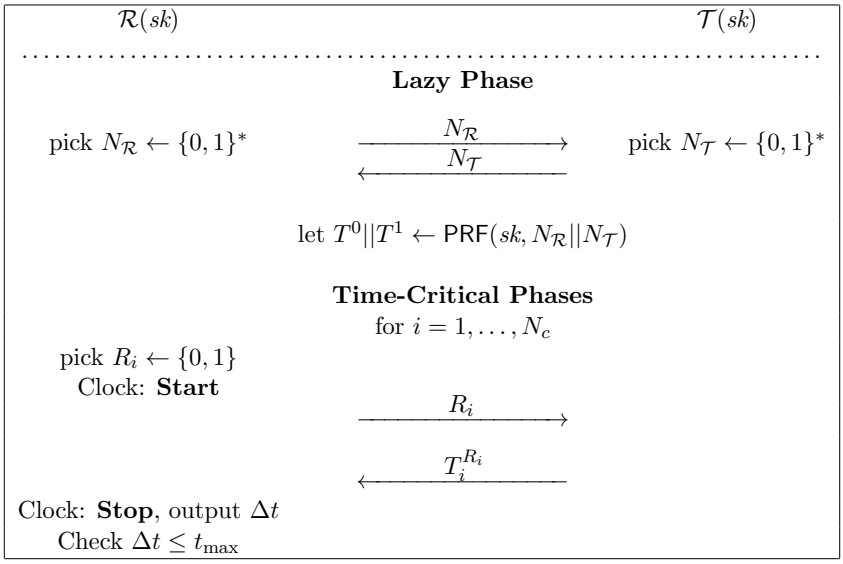

Figure 4: The Hancke and Kuhn protocol

such that

$$\begin{aligned} \boldsymbol{Adv}^{mafia}_{\mathsf{ID}}(\mathcal{A}) \;\leq\; & q_{\mathcal{R}} \cdot \left(\tfrac{3}{4}\right)^{N_c} + \binom{q_{\mathcal{R}} + q_{\text{OBS}}}{2} \cdot 2^{-|N_{\mathcal{R}}|} \\ & + \boldsymbol{Adv}^{d}_{\mathsf{PRF}}(\mathcal{A}') + \binom{q_{\mathcal{T}}}{2} \cdot 2^{-|N_{\mathcal{T}}|}. \end{aligned}$$

Proof. The first statement follows easily: an offline impersonation adversary which simply generates a random nonce during the single lazy phase of this protocol wins with probability 1 as there is no authentication during this lazy phase. For terrorist fraud resistance, a malicious tag equips the adversary during the lazy phase of session sid with the output $T^0||T^1$ of the PRF for nonces $N_{\mathcal{R}}$ and $N_{\mathcal{T}}$. This adversary succeeds with probability 1, as it can answer the queries of the honest reader in sid as though it were the legitimate tag. Now consider a simulator attempting to authenticate with the data obtained from the adversary; as the simulator's session is fresh, however, so is the reader's nonce, thus either the PRF output is different, or we can find a collision in the PRF. Finally, the attack of Boureanu et al. [5] makes the protocol not distance-fraud resistant: indeed, since the adversary in this case knows the secret key, it can choose a weak nonce, such that $T^0 = T^1$ with high probability (note that the pseudorandomness of PRF only requires that it is indistinguishable from random on the average, and not for every input). Once the distance fraud adversary forwards this nonce to the reader, it can trivially commit to the value T_i^0 for every round, irrespective of the challenge, thus winning with high probability.

The proof of the last statement consists of the following high-level steps:

1. Show that one can safely replace the PRF runs of honest parties by picking independent random strings $T^0||T^1$ and for each new nonce tuple $(N_{\mathcal{R}}, N_{\mathcal{T}}, M)$.

2. Show that nonce pairs are (almost) unique, except for possibly one adversary-tag session sid^* having the same nonce pair as a reader-adversary session sid (here the adversary relays the nonces between sessions).

3. Bound the probability that the adversary passes the time-critical phases for at most one adversary-tag interaction.

For the first step we claim that replacing the PRF-values by random (but consistent) values can at most decrease the adversary's success probability by the distinguishing advantage for PRF. This holds as we can construct adversary $\mathcal{A}'$ against PRF via black-box simulation of $\mathcal{A}$, each time applying the random or pseudorandom oracle to nonce pairs on behalf of the honest parties. Finally, $\mathcal{A}'$ checks if $\mathcal{A}$ succeeds in some reader-adversary session and outputs 1 if this happens. The distinguishing advantage of $\mathcal{A}'$ then corresponds to the decrease of the success probability of $\mathcal{A}$ when switching to random values $T^0||T^1$.

Next consider the adversary $\mathcal{A}$ mounting a mafia fraud attack and all the pairs of nonces appearing in the attack. Assume that there exist two sessions (between adversary and tag or reader, or between both honest parties) with the same pair $(N_{\mathcal{R}}, N_{\mathcal{T}})$. We claim that this can only be a reader-adversary session and an adversary-tag session, except with probability

$$\binom{q_{\mathcal{R}} + q_{\text{OBS}}}{2} \cdot 2^{-|N_{\mathcal{R}}|} + \binom{q_{\mathcal{T}}}{2} \cdot 2^{-|N_{\mathcal{T}}|}.$$

This holds as for each two executions for the reader resp. tag the nonce of this party is picked at random. If three identical nonce pairs appear in some executions then two of them are either in the at most $q_{\mathcal{R}} + q_{\text{OBS}}$ executions with the reader, or in the $q_{\mathcal{T}}$ executions with the tag. Such collisions occur with the above probability.

Declare $\mathcal{A}$ to lose if a collision appears, decreasing its success probability by this negligible term, but allowing us to consider collision-free executions. In particular, except for the matching session, all values $T^0||T^1$ in the attack are independent.

Now consider a reader-adversary session sid in which $\mathcal{A}$ successfully impersonates to $\mathcal{R}$. By assumption the same nonce pair appears in at most one other adversary-tag session. If there exists a (unique) matching adversary-tag session sid^* then we claim that this session taints sid with high probability (if there is no such session, we have the case below, where the adversary does not take advantage of a matching session). Since for this protocol it holds that $T_{\max} = 1$, this invalidates session sid. Suppose, to the contrary, that the matching session sid^* taints no time-critical phase in sid.

Consider an untainted time-critical phase of sid where $\mathcal{R}$ sends $R_i = b$ and expects $T_i^b \oplus b$. The adversary has thus successfully passed the first $i-1$ time-critical phases and can choose to do one of the following in the i-th phase:

The Go-Early Strategy. In session sid^* the adversary has sent some bit R_i^* to $\mathcal{T}$ before having received R_i (i.e., $\mathsf{clock}(\mathsf{sid}, i+2) > \mathsf{clock}(\mathsf{sid}^*, i+2)$ in the notation of [10]). Then, since R_i is random and independent of all other data, the probability of $R_i^* \neq R_i$ is $\frac{1}{2}$, in which case $\mathcal{A}$ does not receive T_i^* in sid^* and can only guess the value T_i in sid. If $b = R_i = R_i^*$, however, the adversary returns the correct reply T_i^b with probability 1.

The Go-Late Strategy. In session sid the adversary replies to R_i with some T_i^* before receiving $(T_i^b)^*$ in session sid^* (i.e., $\mathsf{clock}(\mathsf{sid}, i+3) < \mathsf{clock}(\mathsf{sid}^*, i+3)$). Now $\mathcal{A}$ succeeds only with probability $\frac{1}{2}$ for this phase.

The Modify-it Strategy. The adversary schedules the message such that it receives R_i in sid, sends some $R_i^* = b$ in sid^*, receives $(T_i^b)^*$ in sid^*, and forwards some T_i^* in sid. Hence, the scheduling corresponds to a pure relay attack, but $R_i \neq R_i^*$ or $T_i^* \neq (T_i^b)^*$. If $b = R_i^*$ is wrong, then $(T_i^b)^*$ is actually never sent by $\mathcal{T}$ in sid^* and the adversary can thus only guess T_i^* with probability $\frac{1}{2}$; if $b = R_i = R_i^*$ then $T_i^* \neq (T_i^b)^*$ makes the reader reject.

The Taint-it Strategy. The adversary taints this phase of sid through sid^*. This is equivalent here to losing in sid.

The Taint-it strategy may be ignored, as it disables sid. The Go-Late and Modify-it strategies both succeed with probability at most $\frac{1}{2}$. The Go-Early strategy succeeds with probability $\frac{3}{4}$. As all rounds are independent, and taking into account the $q_{\mathcal{R}}$ trials, this gives the claimed bound. □

3.1.1 Avoine and Tchamkerten

As noted in the previous section, the Hancke and Kuhn construction has interdependent challenges, allowing the adversary to run a complete number of time-critical exchanges with the tag before attempting to authenticate to the reader. The Avoine and Tchamkerten protocol tries to correct this flaw by providing some reader authentication. The main idea is to store the secret key in one (or more) binary tree(s); the challenges are inter-related, with the responses forming paths from the root to the leaves.

Consider now an adversary impersonating a reader in an adversary-tag session in a Go-Early strategy as in the proof of Theorem 3.2. For [13], an adversary can choose challenges R_i^* for each round $i = 1, \ldots, N_c$ and receive responses T_i^* from the tag, having a 50% probability to have queried the honest tag with the correct $R_i^* = R_i$; for these queries, the adversary will know the correct response $T_i = T_i^*$. On the other hand, once an adversary makes an incorrect guess for some R_i^*, none of the future responses will be the correct ones.

This is also depicted in Figure 5.

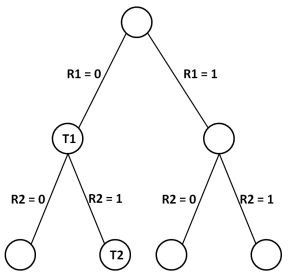

Figure 5: The Avoine-Tchamkerten protocol where the reader sends challenges $R_1 = 0$, $R_2 = 1$

Note that this protocol requires at least $2^{N_c+1} - 2$ potential response bits; at the expense of security, one could consider more than just one tree. We have, however, lazy phase authentication, leading to a total PRF output of $m + 2^{N_c+1} - 2$ bits for acceptably large m. The first m bits of the PRF output are denoted M; the last $2^{N_c+1} - 2$ bits are denoted T and stored in a tree as follows: from the top downwards and from left to right, each node is labelled with an output bit. The edge between each node and its left child is labelled 0 and the edge to the right child is labelled 1. We denote by $\mathsf{Node}(R_1, \ldots, R_i)$ the label of the node that has the path $R_1, \ldots, R_i$ from the root. With this notation, the protocol runs as shown in Figure 6. Note in particular that the responses T_i do not correspond directly to the bits of T, but are rather chosen from amongst the bits of T according to the path $R_1, \ldots, R_i$.

Intuitively, the m-bit value V offers impersonation resistance. While the inter-dependency between the challenges of the reader increases the protocol's mafia fraud resistance, this is still not optimal, and it requires a great deal of storage. We formalise the concrete security of this construction below.

Theorem 3.3 (Avoine-Tchamkerten Properties) *Let* ID *be the distance-bounding authentication scheme in Figure 6 with parameters* $(t_{\max}, N_c)$*. This scheme has the following properties:*

- *It is not resistant to terrorist fraud, nor to distance fraud.*
- *For any* $(t, q_{\mathcal{R}}, q_{\mathcal{T}}, q_{\text{OBS}})$*-mafia-fraud adversary* $\mathcal{A}$ *against the scheme there exists a* (t', q')*-distinguisher*

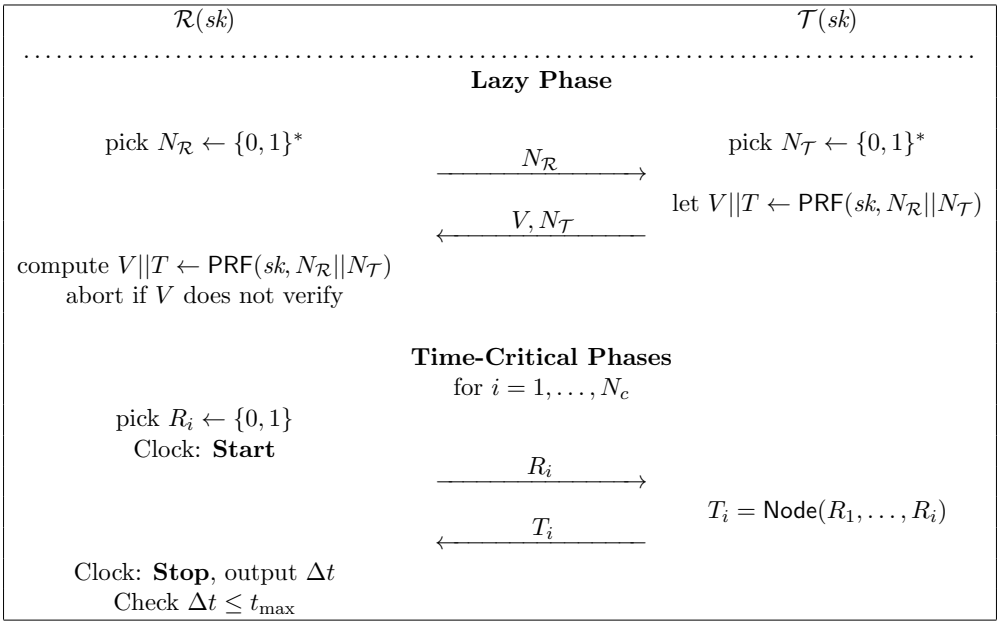

Figure 6: The Avoine and Tchamkerten protocol

$\mathcal{A}'$ against PRF *(where $t' = t + O(n)$ and $q' = q_{\mathcal{R}} + q_{\mathcal{T}} + q_{\text{OBS}}$) such that*

$$\boldsymbol{Adv}^{mafia}_{\mathsf{ID}}(\mathcal{A}) \;\leq\; \tfrac{1}{2} q_{\mathcal{R}}[N_c + 2] \cdot 2^{-N_c} + \boldsymbol{Adv}^{d}_{\mathsf{PRF}}(\mathcal{A}') + \binom{q_{\mathcal{R}} + q_{\text{OBS}}}{2} \cdot 2^{-|N_{\mathcal{R}}|} + \binom{q_{\mathcal{T}}}{2} \cdot 2^{-|N_{\mathcal{T}}|}.$$

- *For any $(t, q_{\mathcal{R}}, q_{\mathcal{T}}, q_{\text{OBS}})$-impersonation adversary $\mathcal{A}$ against the scheme there exists a (t', q')-distinguisher $\mathcal{A}'$ against* PRF *(where $t' = t + O(n)$ and $q' = q_{\mathcal{R}} + q_{\mathcal{T}} + q_{\text{OBS}}$) such that*

$$\boldsymbol{Adv}^{imp}_{\mathsf{ID}}(\mathcal{A}) \;\leq\; q_{\mathcal{R}} \cdot 2^{-|V|} + \binom{q_{\mathcal{R}} + q_{\text{OBS}}}{2} \cdot 2^{-|N_{\mathcal{R}}|} + \boldsymbol{Adv}^{d}_{\mathsf{PRF}}(\mathcal{A}') + \binom{q_{\mathcal{T}}}{2} \cdot 2^{-|N_{\mathcal{T}}|}.$$

Proof. The proofs of the first statement is identical to those of the Hancke-Kuhn protocol, as shown in Theorem 3.2. For mafia fraud resistance, we change the Hancke-Kuhn proof as follows: for each round, we denote by pass_j the event that an adversary guesses the correct R_i in a Go-Early attack. Again the Go-Early strategy is the most effective, with a success probability given by the iterative expression below:

$$\text{Prob}\left[\bigwedge\nolimits_{j=i}^{N_c} \mathsf{pass}_j \;\middle|\; \bigwedge\nolimits_{j=1}^{i-1} \mathsf{pass}_j\right] \leq \tfrac{1}{2} \cdot \tfrac{1}{2}^{N_c - i + 1} + \frac{1}{2} \cdot \text{Prob}\left[\bigwedge\nolimits_{j=i+1}^{N_c} \mathsf{pass}_j \;\middle|\; \bigwedge\nolimits_{j=1}^{i} \mathsf{pass}_j\right].$$

After summing up and iterating, we have the bound above.

Finally, impersonation security follows similarly as the mafia fraud resistance. We first account for nonce-collisions between authentication sessions, as in Theorem 3.2. The impersonation adversary's advantage is now given by the advantage of the distinguisher $\mathcal{A}'$ and the probability that the adversary knows the output of the PRF for the successful reader-adversary session sid (which only happens if it has seen, resp. generated the authentication value V in a reader-tag, resp. adversary-tag session). □

3.2 Reid et al.

The construction in [4] has better mafia fraud resistance (but greater data storage) than [13]. The scheme in [4] is additionally offline impersonation resistant. However, neither scheme is terrorist fraud

resistant. We now analyse the protocol due to Reid et al. [16], which adds a symmetric encryption scheme to the Hancke and Kuhn construction (the authors suggest a one-time-pad xor operation[2]). Thus, this protocol inherits the lack of offline impersonation resistance of [13], as well as the lower mafia fraud resistance. Interestingly enough, distance-fraud resistance might be achieved for some implementations of the symmetric encryption scheme, but not for others, since the attack of Boureanu et al. [5] cannot be extended to arbitrary schemes.

We first describe this protocol. In their paper [16] use a so-called key derivation function denoted KDF which can be viewed as a PRF for the sake of simplicity. We thus denote it PRF as for the construction due to Hancke and Kuhn. Furthermore, Reid et al. consider a symmetric IND-CPA encryption scheme denoted $\mathcal{E}$. To this scheme they associate a symmetric ephemeral secret key eph and a long term secret key sk. The notation $\mathcal{E}_{\mathsf{eph}}(sk)$ denotes the encryption under eph of the plaintext sk. We denote the corresponding decryption process by $\mathcal{D}$. Furthermore, Reid et al. [16] associate to the tag and reader some public identities $\mathcal{ID}_A, \mathcal{ID}_B$.

The main idea here is that both reader and tag compute a symmetric ephemeral key eph as the output of PRF, and then they use eph to encrypt the long-term secret key sk with $\mathcal{E}$. For each time-critical round, the reader challenges the tag with a random bit, and the tag responds with a bit, originating either from the encrypted value or from eph. Intuitively, terrorist fraud resistance results from the fact that the adversary requires either the long term secret sk(which can be later be used by the simulator to authenticate) or both eph and the encryption of sk (which can be used by the simulator to compute $\mathcal{D}_{\mathsf{eph}}(sk)$, thus also obtaining the secret sk). The scheme is depicted in Figure 7.

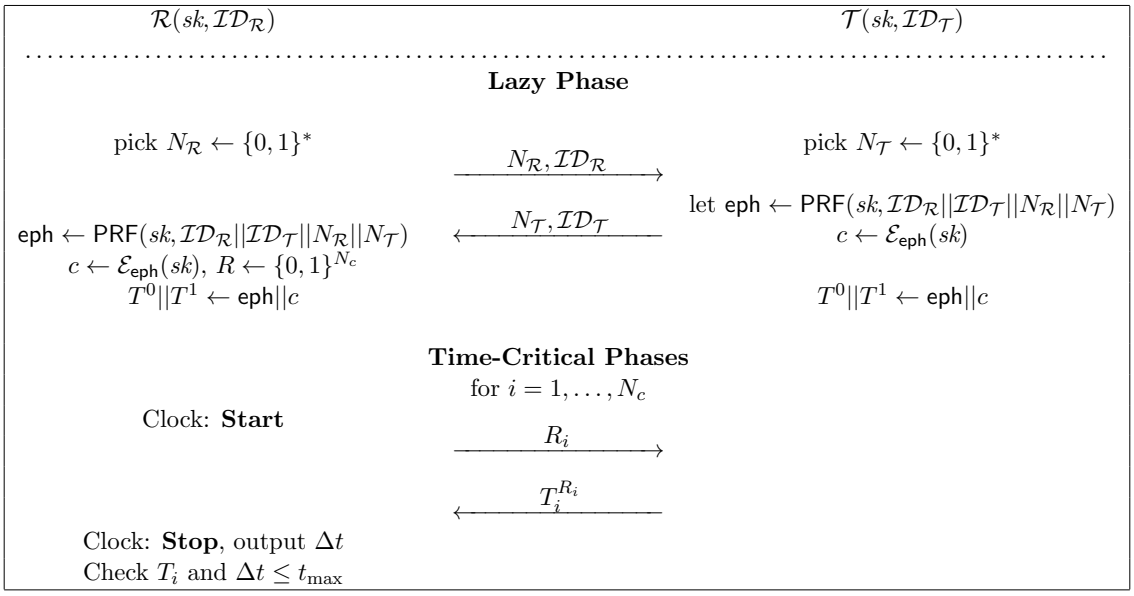

Figure 7: The Reid et al. protocol

Before stating the security properties we note that the informal definition of a successful terrorist fraud attack is that the adversary can authenticate if aided by a malicious tag in a particular session, but cannot authenticate *without* this aid. However, this informal definition is deceptive: many protocols claiming to be terrorist fraud resistant are actually resistant to the very restrictive requirement that the prover does not forward *any* sensitive information to the adversary; in other words, the adversary gains no knowledge about the secret key [3]. By tying any knowledge (or accurate guessing) of both

[2]We note that Reid et al. also suggest different versions of this protocol, where the symmetric encryption is done differently; our analysis here applies equally to generic ways of implementing the encryption scheme.

responses to the secret key, protocols can attain this form of terrorist fraud resistance. In other words, if the prover forwards even a single bit of one of the responses, this is considered trivial help, as it gives the adversary an increased success probability in future rounds. We note, however, that in practice this definition restricts an adversary very much; in fact, in many cases some information about the secret key could leak without significantly aiding the adversary in future authentication attempts.

However, the definition of Dürholz et al. [10] considers such partial attacks to be valid terrorist fraud attacks, concretely requiring that the prover's help does not give the adversary an *equal* advantage for future sessions, i.e. it only excludes attacks where the prover forwards the adversary a fragment of the *secret key itself*, exclusively. Thus, whereas the previous definition restricts the prover's strategies such that *no* information about the secret key leaks to the adversary, the model of Dürholz et al. restricts the adversary very little: only insofar the adversary does not have an equal winning probability after the prover stops helping. We discuss the merits of both definitions in Appendix A.

In the following, we show that the protocol of Reid et al. in Figure 7 does *not* attain terrorist fraud resistance in the sense of Dürholz et al.

Theorem 3.4 (Reid et al. Properties) *Let* ID *be the distance-bounding authentication scheme in Figure 7 with parameters* $(t_{\max}, N_c)$*. This scheme has the following properties:*

- *It is neither offline impersonation resistant, nor terrorist fraud resistant (assuming the pseudorandomness of* PRF*).*
- *If the symmetric encryption scheme in this protocol is instantiated as bitwise XOR, this scheme is* not *mafia fraud resistant.*
- *If the symmetric encryption scheme used in this protocol is instantiated as bitwise XOR, and furthermore if the secret key sk is generated honestly at random from a distribution computationally indistinguishable from the uniform random distribution, for any* $(t, , q_{\mathcal{R}}, q_{\mathcal{T}}, q_{\text{OBS}})$*-distance-fraud adversary* $\mathcal{A}$ *it holds that:*

$$\boldsymbol{Adv}^{dist}_{\mathsf{ID}}(\mathcal{A}) \leq q_{\mathcal{R}} \cdot \left(\tfrac{3}{4}\right)^{N_c}.$$

Proof. This protocol inherits the lack of offline impersonation resistance from the protocol due to Hancke and Kuhn [13] and so we do not show the proof here.

We first look at the distance-fraud resistance of this protocol (i.e. the second statement). The difference for distance-fraud resistance (with respect to the Hancke and Kuhn protocol) is that the responses computed here by the tag are eph and c, only one of which is computed by means of a PRF. Thus, the attack of Boureanu et al. [5] does not transfer trivially. However, the tag can, in a similar fashion, choose a convenient nonce that outputs a "weak" eph, such that, when it is later used as a key, it yields an output that has little entropy. This is possible since the IND-CPA notion applies to keys selected at random by a key-generation algorithm.

However, in case the symmetric encryption function $\mathcal{E}$ is instantiated as bitwise XOR, the Hamming distance between the two responses, eph and c is exactly sk at each execution. The proof goes as follows. First we replace sk by a uniform random value (and lose a term equalling the distinguishing advantage between the uniform random distribution and the distribution that sk is chosen from). Now we consider each round i in a distance-fraud attack. For this phase it holds with probability $\frac{1}{2}$ that the bit c_i is different from the corresponding bit of $\mathsf{eph}_i = c_i \oplus sk$. If $c_i = \mathsf{eph}_i$ (this happens with probability $\frac{1}{2}$), the adversary forwards c_i and wins the round; else, if the two values are unequal, then the adversary has to guess which value to send (essentially predicting the challenge) and is successful with probability $\frac{1}{2}$.

However, if the symmetric encryption function *is* instantiated as a one-time-pad XOR operation, i.e. $c \leftarrow \mathsf{eph} \oplus sk$, we cannot prove the protocol mafia fraud secure. A mafia fraud adversary could run

the following attack, trying to recover the key sk. We will denote the adversary's guess of sk by sk'. The adversary begins by initiating a reader-adversary session sid_1 and an adversary-tag session sid_1^*. It relays all the lazy and time-critical rounds between the reader and the tag up to the $n-1$-th round (this is possible since the definition of mafia fraud resistance only excludes time-critical relaying —up to $T_{\max}$ rounds— for the session sid where the adversary succeeds in its authentication attempt). Finally, in the last round, it receives a challenge bit $b = R_{N_c} \in \{0,1\}$ in session sid_1. Then the adversary forwards the challenge $\bar{b} = b \oplus 1$ in session sid_1^*, receiving the value $T^{\bar{b}}$. The attacker finally forwards this value to the reader in session sid and waits to see if it is authenticated by the reader; if so, it sets $sk'_{N_c} = 0$, and else it sets $sk'_{N_c} = 1$.

There are two cases:

$b = 0$. Then, in sid_1, the reader expects c_{N_c} as a response. From session sid_1^*, where the adversary has forwarded $\bar{b} = 1$, the adversary has learned $\mathsf{eph}_{N_c} = c_{N_c} \oplus sk_{N_c}$, which it forwards to the verifier; if the verifier accepts, then $c_{N_c} = c_{N_c} \oplus sk_{N_c}$, and thus $sk_{N_c} = 0$; in this case, the adversary's guess is correct, and $sk'_{N_c} = sk_{N_c}$. Else, if the verifier rejects, then $sk_{N_c} = 1$, and the adversary has again guessed correctly.

$b = 1$. In this case, the reader expects eph_{N_c}; the value forwarded by the adversary is $c_{N_c} = \mathsf{eph}_{N_c} \oplus sk_{N_c}$. The same reasoning applies as above.

The adversary continues the attack in the same way, recovering the secret key bit-by-bit. Once the adversary has the complete $sk' = sk$, the adversary finally initiates its "challenge" session sid with the verifier, and a parallel session sid^* with the prover. The adversary forwards the lazy phase communication, and then queries the prover in advance to learn eph (i.e. it sends a challenge bit $R_i = 1$ for every round $i \in 1, \ldots, N_c$ in sid^*). Since these queries are made before the adversary receives the challenge in session sid, the adversary has not tainted the round. For every round in sid, if the verifier sends a challenge $R_i = 1$, then the adversary sends eph_i; else, if the verifier sends $R_i = 0$, then the adversary forwards $\mathsf{eph}_i \oplus sk'_i$, and thus also responds correctly. Now the adversary authenticates with probability 1.

However, this attack is not extendable to arbitrary symmetric encryption schemes. What is needed in order for the scheme to achieve some mafia fraud resistance (equal to the mafia fraud resistance of the protocol of Hancke and Kuhn [13]) is for the encryption scheme to not leak any information about the key from learning both possible time-critical round responses for a given round.

Finally we show a terrorist adversary $\mathcal{A}$ for which there exists no simulator such that $\mathbf{Adv}^{\text{terror}}_{\mathsf{ID}}(\mathcal{A}, \mathcal{S}, \mathcal{T}) \leq 0$. This would therefore show that the scheme is not terrorist fraud resistant. The idea is for the malicious tag $\mathcal{T}'$ to give information that facilitates the adversary's attack, without revealing any essential information about future impersonation attempts. Indeed, let $\mathcal{A}$ receive the value eph from $\mathcal{T}'$ in each of its $q_{\mathcal{R}}$ impersonation attempts. In the subsequent time-critical phases, if the reader sends challenge $R_i = 1$, $\mathcal{A}$ sends $T_i = \mathsf{eph}_i$; else, the adversary guesses T_i. This adversary's probability to win is thus $\frac{3}{4}^{N_c} + \mathbf{Adv}^{\text{IND-CPA}}_{\mathcal{E}}(\mathcal{A}'')$ for the adversary $\mathcal{A}''$ whose advantage to win against the IND-CPA of $\mathcal{E}$ is the largest.

Now consider a simulator $\mathcal{S}$. The simulator has no access to the tag $\mathcal{T}$, but it may run $\mathcal{A}$ internally. However, under the assumption of the pseudorandomness of PRF, there is only a negligible probability that $\mathcal{A}$ knows eph for any of the $q_{\mathcal{R}}$ impersonation sessions where the simulator attempts to authenticate to the reader. Thus, the simulator's probability of winning is $\frac{1}{2}^{N_c} + \mathbf{Adv}^{\text{IND-CPA}}_{\mathcal{E}}(\mathcal{A}'')$. Thus, the adversary has an advantage over any simulator $\mathcal{S}$. □

3.3 The Swiss-Knife RFID Distance Bounding Protocol

The Swiss-Knife protocol due to Kim et al. [15] aims to achieve privacy as well as mafia, terrorist, and distance fraud resistance. Very notably, the lazy phase of this protocol is divided into two parts: the

first precedes the time-critical phase, the second follows it. In the first part, the reader and tag exchange nonces and the tag computes a pseudo-random function (PRF) on input a system constant and a tag-chosen nonce $N_{\mathcal{T}}$. The output of this function, a, is then XORed with the long-term secret key sk, thus obtaining response vectors T^0 and T^1. These values are sent in the time-critical rounds, depending on the reader's challenge.

If this protocol is run in an RFID scenario, the size of sk, which equals the number of time-critical rounds, is restricted by the tag's capacity to sustain time-critical rounds. Thus, the keys are short. Finally, after the time-critical rounds, during the second lazy phase, the tag authenticates by computing the PRF on all the received challenges, its identity, and both the reader and the tag's nonces. The reader may then also authenticate by computing the PRF on input the tag's nonce. This second lazy phase is essential in preventing the recovery of the secret key during a mafia fraud attack. Also, this second PRF computation brings the mafia fraud resistance to loosely $(\frac{1}{2})$ per round.

In the Swiss-Knife scenario, each tag is associated with an identity ID, stored by the reader in the same database that stores the secret key sk of the tag. In order to achieve anonymity, this identity is never sent, and the reader needs to search the database exhaustively to find it. This protocol also has some fault tolerance, i.e. the reader counts a total number of errors consisting of: (1) the number of faulty challenges R_i that the tag receives; (2) the number of faulty responses T_i that the reader receives; and (3) the number of rounds in which the tag's response exceeds the time threshold $t_{\max}$. The protocol is depicted in Figure 8. Note that Kim et al. also present a more efficient version, but whereas this second scheme is computationally more efficient than the simplified one, the security properties are comparable. In Figure 8, the value const is a system constant.

Theorem 3.5 (Swiss-Knife Properties) *Let* ID *be the distance-bounding authentication scheme in Figure 8 with parameters* $(t_{\max}, N_c)$*. Assume furthermore that the secret key sk output by* Kg *is chosen uniformly at random from a distribution that is computationally indistinguishable from the random distribution on bitstrings of length* $|sk|$*. This scheme has the following properties:*

- *It is not terrorist fraud resistant (assuming the pseudorandomness of* PRF*).*
- *For any* $(t, q_{\mathcal{R}}, q_{\mathcal{T}}, q_{\text{OBS}})$*-distance-fraud adversary* $\mathcal{A}$ *against the scheme, there exists an adversary* $\mathcal{A}'$ *distinguishing sk from a truly random value such that:*

$$\boldsymbol{Adv}^{dist}_{\mathsf{ID}}(\mathcal{A}) \leq q_{\mathcal{R}} \cdot \left(\tfrac{3}{4}\right)^{N_c - T} + \boldsymbol{Adv}^{d}_{sk}(\mathcal{A}').$$

- *For any* $(t, q_{\mathcal{R}}, q_{\mathcal{T}}, q_{\text{OBS}})$*-mafia-fraud adversary* $\mathcal{A}$ *against the scheme there exists a* (t', q')*-distinguisher* $\mathcal{A}'$ *against* PRF *(where* $t' = t + O(n)$ *and* $q' = q_{\mathcal{R}} + q_{\mathcal{T}} + q_{\text{OBS}}$*) such that*

$$\begin{aligned}
\boldsymbol{Adv}^{mafia}_{\mathsf{ID}}(\mathcal{A}) \quad \leq \quad & \left(\tfrac{1}{2}\right)^{N_c - T} + 2\binom{q_{\mathcal{T}}}{2} \cdot 2^{-(|N_{\mathcal{T}}| + \lceil \frac{N_c}{2} \rceil - T)} \\
& + 2\binom{q_{\mathcal{R}} + q_{\text{OBS}}}{2} \cdot 2^{-(|N_{\mathcal{R}}| + \lceil \frac{N_c}{2} \rceil - T)} \\
& + \binom{q_{\mathcal{R}} + q_{\text{OBS}}}{2} \cdot 2^{-(|N_{\mathcal{R}}| + N_c - 1 - T)} \\
& + 2\binom{q_{\mathcal{T}}}{2} \cdot 2^{-(|N_{\mathcal{T}}| + N_c - 1 - T)} q_{\mathcal{R}} \cdot \boldsymbol{Adv}^{d}_{\mathsf{PRF}}(\mathcal{A}').
\end{aligned}$$

- *For any* $(t, q_{\mathcal{R}}, q_{\mathcal{T}}, q_{\text{OBS}})$*-offline impersonation adversary* $\mathcal{A}$ *against the scheme there exists a* (t', q')*-distinguisher* $\mathcal{A}'$ *against* PRF

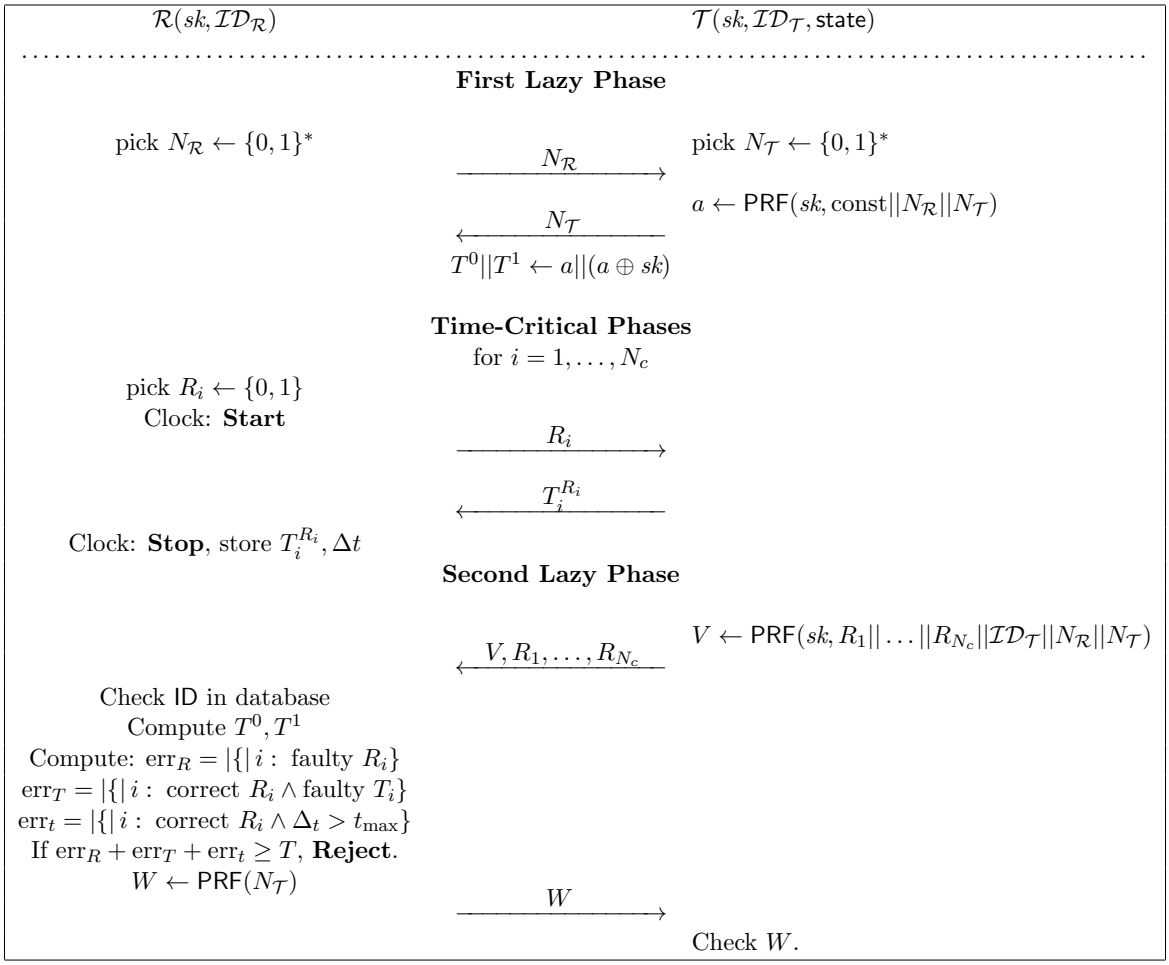

Figure 8: The Swiss-Knife protocol

(where $t' = t + O(n)$ and $q' = q_\mathcal{R} + q_\mathcal{T} + q_{\text{OBS}}$) such that

$$\begin{aligned} \boldsymbol{Adv}^{imp}_{\mathsf{ID}}(\mathcal{A}) \;\leq\; & q_\mathcal{R} \cdot 2^{-|V|} + \binom{q_\mathcal{R} + q_{\text{OBS}}}{2} \cdot 2^{-(|N_\mathcal{R}| + N_c - T)} \\ & + \boldsymbol{Adv}^{d}_{\mathsf{PRF}}(\mathcal{A}') + \binom{q_\mathcal{T}}{2} \cdot 2^{-(|N_\mathcal{T}| + N_c - T)}. \end{aligned}$$

Proof. The proof of statement 1 is trivial: the malicious tag can even send the adversary the secret key sk and the output V for the second lazy phase. However, the simulator cannot guess the value of $\mathcal{ID}_\mathcal{T}$ except with probability $2^{|\mathcal{ID}_\mathcal{T}|}$.

For the second statement, note that the tag does *not* choose the key sk, thus, if this key is chosen at random from a distribution computationally indistinguishable from the uniform random distribution, the attack of Boureanu et al. [5] is thwarted. Indeed, with great probability, whatever the value a output by the PRF for this session, the value $a \oplus sk$ is with high probability at a large Hamming distance from a. The proof goes as follows. First we replace sk by a uniform random value (and lose a term $\mathbf{Adv}^{\text{d}}_{\mathcal{D}}(\mathcal{A}')$). Now we consider each round i in a distance-fraud attack. For this phase it holds with probability $\frac{1}{2}$ that the bit $T_i^0 = a_i$ is different from the corresponding bit of $T_i^1 = a_i \oplus sk$. If $T_i^0 = T_i^1$ (this happens with probability $\frac{1}{2}$), the adversary forwards T_i^0 and wins the round; else, if the two values are unequal, then

the adversary has to guess which value to send (essentially predicting the challenge) and is successful with probability $\frac{1}{2}$. After we account for the fault tolerance level T we attain the above-stated bound.

For the third statement, the proof goes slightly differently than for previous protocols, e.g. the scheme due to Hancke and Kuhn 3.1. In particular, the response strings T^0 and T^1 are now related. The proof follows in these rough steps: (1) assuming that the secret key sk is indistinguishable from a random string of appropriate length *at the end of any mafia fraud interaction*, we can prove mafia fraud security as in the previous proofs, in particular replacing the response strings by random values. We also account, even for sessions with matching $N_{\mathcal{R}}$ and $N_{\mathcal{T}}$, for about half the challenges in time-critical rounds. Then, we show that (2) except with negligible probability the adversary cannot distinguish the key sk from a random string of corresponding length. This second step is proved as follows: note first that if the adversary merely observes the interaction between an honest prover and an honest verifier, or simply relays all messages exactly as he receives them, this does not, with great probability, reveal any information about sk. However, if two sessions share the same nonces (and the computed responses are identical), the prover may learn about $\lceil \frac{N_c}{2} \rceil$ of the bits of the secret key (wherever the challenges differ and the prover gives the correct response from the other response string). This happens with probability roughly

$$\binom{q_{\mathcal{R}} + q_{\text{OBS}}}{2} \cdot 2^{-(|N_{\mathcal{R}}| + \lceil \frac{N_c}{2} \rceil - T)} + \binom{q_{\mathcal{T}}}{2} \cdot 2^{-(|N_{\mathcal{T}}| + \lceil \frac{N_c}{2} \rceil - T)}.$$

We now assume this is not the case. Now for each of the N_c time-critical phases the adversary learns a bit from either one of the two response strings, but not from both, thus leaking no information about the secret key. Furthermore, if the adversary *does* interfere with the running of the protocol, in particular changing either a challenge or a response, there are only three possibilities: (a) the adversary changes at least one challenge or one response, in which case the matching prover computes a different authentication value in the matching adversary-prover session than the value expected by the verifier in the verifier-adversary session, and the adversary is unable to provide the correct authentication value (in this case the adversary can learn nothing beyond what it learns in a common observation of an authentication attempt); (b) the adversary has seen (or "created", by forwarding nonces) another session in which all the challenges and responses are exactly the same for all except the rounds where the adversary wishes to change the challenge and/or response (thus the adversary has seen a valid authentication string V for the altered string of challenges and responses); (c) the adversary is able to come up with a forgery for the value V. Event (b) happens with probability:

$$\binom{q_{\mathcal{R}} + q_{\text{OBS}}}{2} \cdot 2^{-(|N_{\mathcal{R}}| + N_c - 1 - T)} + \binom{q_{\mathcal{T}}}{2} \cdot 2^{-(|N_{\mathcal{T}}| + N_c - 1 - T)}.$$

Event (c) happens with probability $\mathbf{Adv}^{\text{d}}_{\mathsf{PRF}}(\mathcal{A}')$ (per verifier-adversary session). This accounts for the bound above.

The proof for offline impersonation security runs more or less as in the previous proof, only now we only account for the probability of lazy-phase authentication. □

3.4 The Protocol of Yang, Zhuang, and Wong

Similarly to the Swiss-Knife protocol, the very recent protocol due to Yang, Zhuang, and Wong [18] aims to achieve mafia, distance, terrorist, and offline impersonation security, as well as privacy. The protocol is also supposed to achieve mutual authentication. Furthermore, the authors claim to achieve this with a single slow phase, rather than two (as in the Swiss-Knife protocol). In this protocol, privacy is achieved by means of tag-id updates, where the reader is supposed to resynchronize with the tag at every authentication attempt. In the distance-bounding part, the reader authenticates first in the lazy phase (though the tag does not do so). Instead, tag authentication is limited to the time-critical phases.

Both the reader and the tag compute in this protocol a pseudo-random function with as input two nonces interchanged by the parties, and the tag's identifier. The PRF output is split into three parts: one string called v is used for the reader authentication mentioned above, and two other parts, which we denote T^1 and T^2, are used first in order to compute a third string T^3, which is the bitwise XOR of T^1, T^2, and sk, i.e. the secret key shared by the reader and the tag. If the verifier authentication is completed, then the time-critical phases begin: in each phase $i = 1, \ldots, N_c$, the reader sends a random challenge R_i. The tag computes the response depending on the value of $R_i||v_i$, where v_i denotes the i-th bit of the verifier authentication string v. In particular, if $R_i||v_i = 0||0$, then the response bit is T_i^1; if $R_i||v_i = 1||1$, then the response bit is T_i^2; else, the response bit is T_i^3. We denote by PRF the pseudo-random function used for the generation of the time-critical responses and a different pseudo-random function H for the key update. This protocol is depicted in Figure 9.

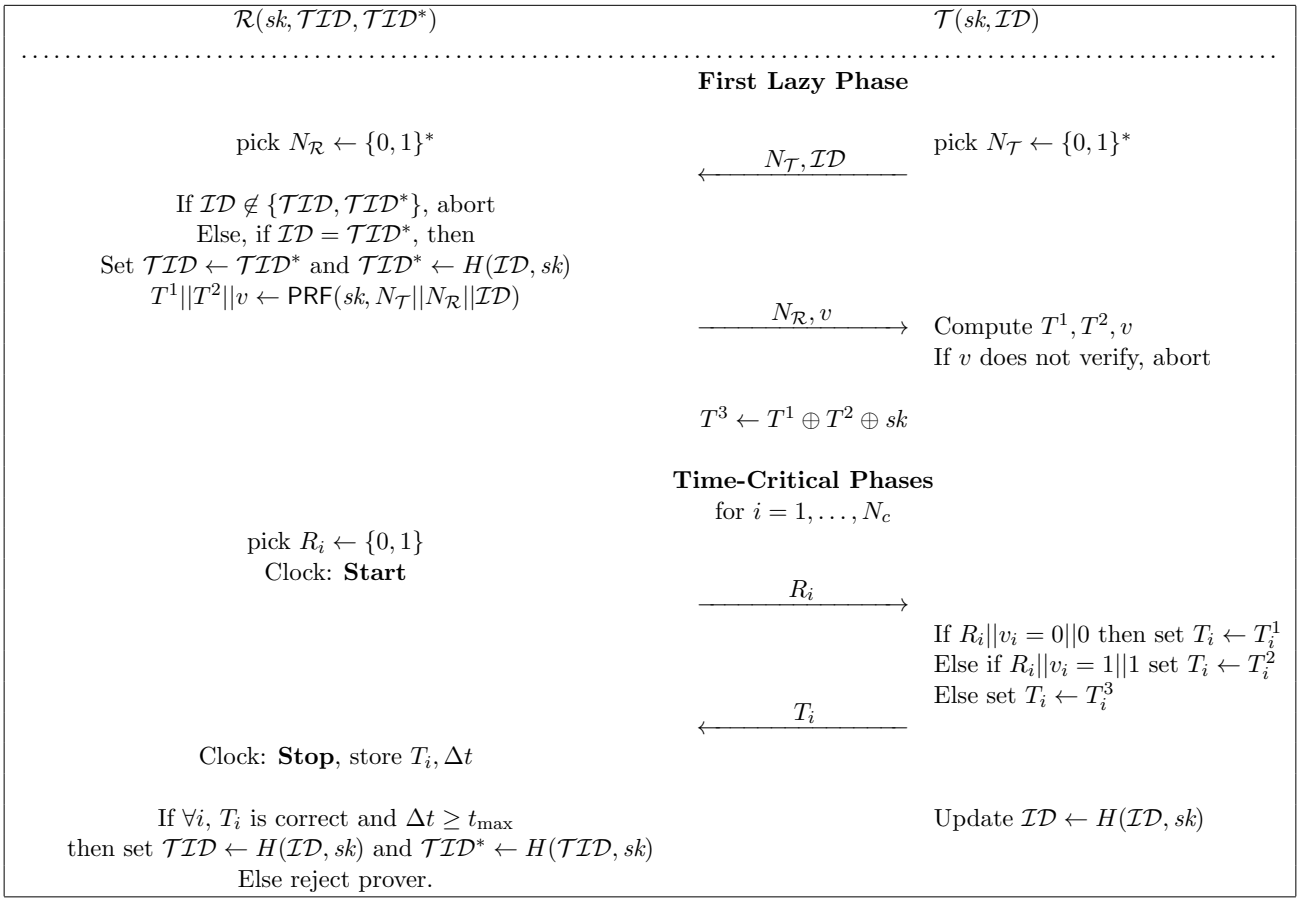

Figure 9: The Yang, Zhuang, Wong protocol

As we show in the following theorem, the protocol achieves none of its claimed properties. While the idea to make the responses depend on a couple of bits (including a randomly generated challenge) rather than on a single bit thwarts key-learning attacks (as in the case of Reid et al.), the absence of the second authentication phase enables a simple Denial-of-Service (DoS) attack, which can be used to break privacy, see e.g. [17]. Furthermore, the proof of terrorist fraud resistance in the original paper [18] is flawed, since it assumes the malicious tag would only forward the adversary (1 or 2) entire strings T^1, T^2, or T^3; however, a malicious tag can also simply forward bits from different strings, depending on the string v (which is a known value). We summarize this analysis below.

Theorem 3.6 (Yang-Zhuang-Wong Properties) *Let* ID *be the distance-bounding authentication scheme in Figure 9 with parameters* $(t_{\max}, N_c)$*. Assume furthermore that the secret key* sk *output by* Kg *is chosen uniformly at random from a distribution that is computationally indistinguishable from the random distribution on bitstrings of length* $|sk|$*. This scheme has the following properties:*

- *It is not terrorist fraud resistant, nor distance-fraud resistant, nor (tag) offline-impersonation secure.*
- *For any* $(t, q_{\mathcal{R}}, q_{\mathcal{T}}, q_{\text{OBS}})$*-mafia-fraud adversary* $\mathcal{A}$ *against the scheme there exists a* (t', q')*-distinguisher* $\mathcal{A}'$ *against* PRF *(where* $\mathcal{A}'$ *runs in time* $t' = t + O(n)$ *and makes at most* $q' = q_{\mathcal{R}} + q_{\mathcal{T}} + q_{\text{OBS}}$ *queries) such that*

$$\begin{aligned} \boldsymbol{Adv}^{mafia}_{\mathsf{ID}}(\mathcal{A}) \quad &\leq \quad q_{\mathcal{R}} \cdot \left(\tfrac{3}{4}\right)^{N_c} + \binom{q_{\mathcal{R}} + q_{\text{OBS}}}{2} \cdot 2^{-|N_{\mathcal{R}}|} \\ &\quad + \boldsymbol{Adv}^{d}_{\mathsf{PRF}}(\mathcal{A}') + \binom{q_{\mathcal{T}}}{2} \cdot 2^{-|N_{\mathcal{T}}|}. \end{aligned}$$

- *This protocol does not preserve privacy, since it does not preserve long-term completeness (it enables Denial of Service (DoS) attacks). In particular, there exists an adversary that desynchronizes the reader and tag state with probability* $\frac{1}{2}$*.*

Proof. We begin with the second statement. Note that now there are three bitstrings which, XORed together, give the secret key. This strategy effectively blocks key-learning attacks. In particular, the mafia-fraud proof for this part follows the lines of the proof in the Hancke-Kuhn protocol, so we don't give it here again.

We now focus on the first statement. First, the protocol is not distance-fraud resistant. In an attack similar to those shown by Boureanu et al. [5], a malicious prover could choose a weak nonce $N_{\mathcal{T}}$ such that $T^1 = T^2 = sk$. In this case, it holds that $T^3 = sk \oplus sk \oplus sk$, and thus the malicious prover can always respond to a randomly generated challenge, regardless of the value of v.

Second, the protocol is trivially not (tag) offline-impersonation secure in the definition of Dürholz et al., since it involves no lazy-phase authentication for the tag.

Thirdly, the protocol is not terrorist fraud resistant. The malicious tag's strategy is as follows. After computing v, it forwards the following values to the adversary: for $i = 1, \ldots, N_c$, if $v_i = 0$, then it forwards $0||T^1_i$ and $1||T^3_i$, but *not* T^2_i; and else, if $V_i = 1$, then it forwards $1||T^2_i$ and $0||T^3_i$, but *not* T^1_i. Here, the prepended bits indicate to the adversary which response to forward during the time-critical phases. Now the adversary can authenticate with probability 1; however, it doesn't learn any information about the key, which is still bitwise hidden by the third, unknown string. Thus, a simulator can't hope to compete with the adversary in this setting.

Finally, we show a Denial of Service attack against this protocol. Note that when the tag is in the reader's proximity, an adversary *can* effectively perform relays. The attack goes as follows: during an honest reader-tag authentication session, the adversary forwards all the communication, apart from the last time-critical phase. For this phase, the adversary does *not* forward the reader's challenge to the tag, which eventually drops the session. However, upon receiving the challenge from the reader, the adversary has a probability of $\frac{1}{2}$ to guess the correct response and thus authenticate, making the reader update state such that it is desynchronized from the tag. Indeed the tag state will still be some value $\mathcal{ID}$, while the reader will keep states: $\mathcal{TID} = H(\mathcal{ID}, sk)$ and $\mathcal{TID}^* = H(H(\mathcal{ID}, sk), sk)$. This directly contradicts the proof of privacy given by the protocol designers [18]. □

References

[1] Abyneh, M.R.S.: Security analysis of two distance-bounding protocols. In: Proceedings of RFIDSec 2011. Lecture Notes in Computer Science, Springer (2011)

[2] Avoine, G., Bingol, M.A., Karda, S., Lauradoux, C., Martin, B.: A formal framework for cryptanalyzing rfid distance bounding protocols. `http://eprint.iacr.org/2009/543.pdf` (2009)

[3] Avoine, G., Lauradoux, C., Martin, B.: How secret-sharing can defeat terrorist fraud. In: Proceedings of the Fourth ACM Conference on Wireless Network Security WISEC 2011. pp. 145–156. ACM Press (2011)

[4] Avoine, G., Tchamkerten, A.: An efficient distance bounding rfid authentication protocol: Balancing false-acceptance rate and memory requirement. In: Information Security. Lecture Notes in Computer Science, vol. 5735, pp. 250–261. Springer-Verlag (2009)

[5] Boureanu, I., Mitrokotsa, A., Vaudenay, S.: On the pseudorandom function assumption in (secure) distance-bounding protocols. In: Accepted at LatinCrypt 2012. pp. ??–?? (2012)

[6] Brands, S., Chaum, D.: Distance-bounding protocols. In: Advances in Cryptology — Eurocrypt'93. pp. 344–359. Lecture Notes in Computer Science, Springer-Verlag (1993)

[7] Cole, P.H., Ranasinghe, D.C.: Networked RFID Systems and Lightweight Cryptography. Springer-Verlag (2008)

[8] Cremers, C., Rasmussen, K.B., Čapkun, S.: Distance hijacking attacks on distance bounding protocols. Cryptology ePrint Archive, Report 2011/129 (2011), ePRINTURL

[9] Desmedt, Y.: Major security problems with the 'unforgeable' (feige)-fiat-shamir proofs of identity and how to overcome them. In: SecuriCom. pp. 15–17. SEDEP Paris, France (1988)

[10] Dürholz, U., Fischlin, M., Kasper, M., Onete, C.: A formal approach to distance bounding RFID protocols. In: Proceedings of the 14$^{\text{th}}$ Information Security Conference ISC 2011. pp. 47–62. Lecture Notes in Computer Science, Springer-Verlag (2011)

[11] Editors, Zhang, Y., Kitsos, P.: Security in RFID and Sensor Networks. CRC Press (2009)

[12] H-Security, T.: Chip-based ID cards pose security risk at airports. `http://www.h-online.com/security/news/item/Chip-based-ID-cards-pose-security-risk-at-airports-905662.html` (2010)

[13] Hancke, G.P., Kuhn, M.G.: An rfid distance bounding protocol. In: SECURECOMM. pp. 67–73. ACM Press (2005)

[14] Kim, C.H., Avoine, G.: Rfid distance bounding protocol with mixed challenges to prevent relay attacks. In: Proceedings of the 8th International Conference on Cryptology and Networks Security (CANS 2009). Lecture Notes in Computer Science, vol. 5888, pp. 119–131. Springer-Verlag (2009)

[15] Kim, C.H., Avoine, G., Koeune, F., Standaert, F.X., Pereira, O.: The swiss-knife RFID distance bounding protocol. In: Proceedings of the 14$^{\text{th}}$ Information Security Conference ISC 2011. pp. 98–115. Lecture Notes in Computer Science, Springer-Verlag (2009)

[16] Reid, J., Nieto, J.M.G., Tang, T., Senadji, B.: Detecting relay attacks with timing-based protocols. In: ASIACCS. pp. 204–213. ACM Press (2007)

[17] Vaudenay, S.: On privacy models for rfid. In: Advances in Cryptology — Asiacrypt'07. Lecture Notes in Computer Science, vol. 4883, pp. 68–87. Springer-Verlag (2007)

[18] Yang, A., Zhuang, Y., Wong, D.S.: An Efficient Single-Slow-Phase Mutually Authenticated RFID Distance-Bounding Protocol with Tag Privacy. In: Information and Communications Security, Lecture Notes in Computer Science, vol. 7618, pp. 285–292. Springer-Verlag (2012)

A The Case for Terrorist Fraud Resistance

In this paper, we prove that two constructions which claim to achieve terrorist fraud resistance, and which intuitively achieve some degree thereof, can be proved to be insecure in the framework due to Dürholz et al. [10]. In particular, our attack against these two schemes uses the fact that *partial* key-related information from the dishonest prover gives the adversary some advantage over a simulator, which represents the adversary when the prover stops helping.

Our results may be viewed from two separate points of view. It can be argued, on the one hand, that our model due to Dürholz et al. is too strong, and does not accurately capture the notion of terrorist fraud resistance. On the other hand, our results may be viewed as proof that terrorist fraud resistance is in fact a very powerful attack, which is difficult to counteract in practice. We present and assess both points of view in the considerations below.

Model strength. As noted in Section 3.2, the notion achieved by [16] is very weak in the sense that it excludes even prover information that significantly aids adversary authentication while disclosing a relatively insignificant *part* of the secret key. We note that previous definitions, such as the one in Avoine et al.'s framework [2], are ambiguous regarding this point. In fact, Avoine et al. require, literally, that the prover's help gives the adversary no advantage in future attempts. It is unclear, however, what "further" means in this context: does it refer to the success probability of the adversary *after* the prover helped it, compared to the adversary's success *before* the prover helped it, or rather to the notion captured by [10], i.e. the success probability of the adversary *after* the prover helped it compared to *while* the prover helped it?

In fact, the recent work of Avoine et al. [3] indicates that a construction is terrorist fraud resistant if it leaks no information about the secret key (in particular the proof of [3] looks at the conditional entropy of the secret). However, we can argue that perhaps this restriction on the adversarial capacity is too strong: in particular, the adversary could learn information about the secret *as long as this does not increase its success probability.*

If we take the intuitive notion of Avoine et al. [2], the protocol due to Reid et al. [16] is intuitively terrorist fraud resistant. However, we point out that no formal definition in the literature covers the weaker definition of terrorist fraud resistance presented informally above. Thus, it is difficult to say how secure these protocols are, nor how they compare in terms of adversarial advantage.

A further question is which definition best captures the intuition behind terrorist fraud attacks. A strong degree of terrorist fraud resistance is always more desirable, thus from this point of view the definition due to Dürholz et al. sets the standard for protocol design. On the other hand, this definition seems hard to achieve, as it enables attacks where some indirect information about the key is forwarded to the adversary (as in sections 3.2 and 3.4).

The intuition of terrorist fraud resistance is that the malicious prover is willing to assist the adversary in its authentication attempt, but wants to control his access. Thus, the adversary should not be able to authenticate without the prover's help. We note, however, that the adversary always has some (usually negligible in the number of time-critical rounds) probability of authenticating without the prover's help:

this is equivalent to the probability that he guesses the correct replies or, equivalently, that he guesses the secret key.

How far does the model in [10] cover this intuitive notion? Dürholz et al. quantify the adversary's success probability in the presence of the malicious prover, and then the simulator's probability (where the simulator does not have access to the prover, only somewhat to the adversary in the state when communicating with the tag). The scheme is considered terrorist fraud resistant if the simulator's probability of success equals (or is greater than) the adversary's probability of success. In other words, an attack is successful if the prover's help enables the adversary to succeed in one session with some probability, but this probability diminishes in future sessions, when the prover is no longer available. In this scenario the prover has the guarantee that the adversary will only be able to authenticate (afterwards) with less probability. This definition also seems too strong, in the sense that Dürholz et al. accept an attack where the prover authenticates with probability 75% (3 out of 4 times), but the simulator can only authentication with probability 50% (1 out of 2 times). This contradicts the spirit of terrorist fraud resistance as it is understood in the literature.

A middle way would be to define a so-called tolerance level for the simulator, i.e. accept attacks as long as the simulator's success probability does not exceed this tolerance level. Note, however, that the attack presented in Section 3.2 can be tweaked so that the adversary still has an advantage over the simulator, whereas the simulator succeeds with a probability within the tolerance level (instead of giving half the response, the prover would forward only a number of bits of this response, thus easing the adversary's job).

It is our opinion that the notion described by Dürholz et al., though strong, does capture the intuition of terrorist fraud resistance better than the weaker definition which these protocol seem to attain. A common approach in security is to be conservative and to ask for strong(er) security, rather than to label insecure protocols as secure.

Constructive aspects. A second perspective in which to view our result is a constructive one, i.e. if we consider that the model by Dürholz et al. captures the correct notion of terrorist fraud resistance, then clearly achieving this definition requires a stronger construction. One might argue that the strong requirement posed by the model in [10] would lead to inefficient constructions. We argue, however, that the notion of terrorist fraud resistance, is, in its own right, a very strong notion: here, the (dishonest) prover *helps* the adversary authenticate. The challenge is thus to ensure that *any* information leaked to the adversary automatically will carry over to the simulator.

We also note that there is a clear separation between distance-bounding realizations for RFID and for more powerful devices. Indeed, terrorist fraud resistance might be more easily achieved if it is possible to use, say, public key cryptography. In this sense, we could wonder how realistic a threat terrorist fraud attacks are on RFID systems and whether it is worth addressing them directly in protocol design. With RFID tags used in the pharmaceutical industry, in general logistics, and in public transport [7, 11], it seems quite likely that terrorist fraud attacks are quite likely in practice in these settings. In fact, RFID systems are also used in airport security in many German airports: impersonation MITM attacks have already been mounted on these systems by the Chaos Computer Club (CCC) [12]. Though these attacks were not real-time relay attacks, the incentive to mount mafia and terrorist fraud attacks on RFID authentication protocols is rather high. It remains an open question whether RFID systems can be efficiently protected against terrorist fraud in practice, however. The results in this paper show that terrorist fraud resistance is not trivial to achieve, and that achieving it may be inefficient for RFID devices. As terrorist fraud resistance is, however, both a very strong, and a very desirable goal, the authors of this paper interpret their results as an incentive to construct protocols that *are*, in fact, terrorist fraud resistant in the notion of Dürholz et al.