Algebraic Attacks against Linear RFID Authentication Protocols

Matthias Krause and Dirk Stegemann

Theoretical Computer Science
University of Mannheim
Mannheim, Germany

Abstract. The limited computational resources available on RFID tags imply a need for specially designed authentication protocols. The light weight authentication protocol HB^+ proposed by Juels and Weis seems currently secure for several RFID applications, but is too slow for many practical settings. As a possible alternative, authentication protocols based on choosing random elements from L secret linear n-dimensional subspaces of $GF(2)^{n+k}$ (so called linear (n,k,L)-protocols), have been considered. We show that to a certain extent, these protocols are vulnerable to algebraic attacks. Particularly, our approach allows to break Cichoń, Klonowski and Kutyłowski's CKK^2-protocol, a special linear $(n,k,2)$-protocol, for practically recommended parameters in less than a second on a standard PC. Moreover, we show that even unrestricted (n,k,L)-protocols can be efficiently broken if L is too small.
Keywords. RFID authentication, HB^+, CKK, CKK^2

1 Introduction

RFID (radio frequency identification) tags are small devices that are equipped with only little memory and computational power. Their main application is the identification of objects, e.g., items in a shopping basket, clothes in a washing machine, or containers in the cargo compartment of a ship or an aircraft. RFID tags communicate with RFID readers wirelessly with only low bandwith and over short distances. Particularly, they present identification information upon a reader's request. The most basic RFID tags contain a unique hardwired identification string, which is transmitted in plaintext. Clearly, an adversary eavesdropping on the communication can immediately impersonate (clone) the tag, and the tag (and more severely the object that it is attached to) can be traced, since it always replies with the same ID.

Dagstuhl Seminar Proceedings 09031
Symmetric Cryptography
http://drops.dagstuhl.de/opus/volltexte/2009/1957

In order to prevent cloning and tracing attacks and to preserve the tagged object's privacy, RFID tags should reveal their identities only to legitimate readers. Since most practically relevant RFID tags are too weak to execute standard authentication protocols, alternative measures are necessary. Besides technical approaches based on blocking or disturbing the communication, light weight authentication protocols and corresponding security models are intensively discussed (see, e.g., [8, 11] or the very recent paper by Blass, Kurmus, Molva, Noubir and Shikfa [1]). One of the most promising proposals is the HB^+ protocol due to Juels and Weis [9], which is based on the (NP-hard) learning parity with noise (LPN) problem and currently seems secure for many RFID applications. The most severe drawback of the protocol is that secure parameter combinations imply large amounts of transmitted data. Together with the small available bandwidth in RFID communication, this may add up to authentication times of a few seconds, which is inacceptable for many applications.

As a possible alternative to the HB^+ protocol, linear (n, k, L)-protocols, i.e., light weight symmetric authentication protocols based on methods from linear algebra, were introduced by Cichoń, Klonowski and Kutyłowski in [3]. In these protocols, the secret key (the identification information in the RFID tag) consists of the specification of L n-dimensional linear subspaces $V_1, \ldots, V_L$ of $GF(2)^{n+k}$. The prover (RFID tag) chooses a random $l \in \{1, \ldots, L\}$ and sends a random $w \in V_l$. Given a message $\tilde{w} \in GF(2)^{n+k}$, the verifier (RFID reader) accepts the proof if there is some $l \in \{1, \ldots, L\}$ such that $\tilde{w} \in V_l$. In [3], the CKK^2-protocol, a special linear $(n, k, 2)$-protocol, and the $\mathsf{CKK}^{\sigma,L}$-protocol, a special linear (n, k, L)-protocol, were suggested for practical application.

After a detailed description of linear (n, k, L)-protocols and the corresponding security model in Sect. 2, we show that linear (n, k, L)-authentication protocols are vulnerable to appropriately designed algebraic attacks. In particular, we present two very fast (polynomial time) attacks against the CKK^2-protocol (Sect. 3), which allow to recover the secret key for the proposed parameters $(n, k) = (128, 30)$ in less than a second on a standard PC, while an earlier (exponential time) attack on CKK^2 published in [5] requires a couple of hours. Concerning general linear (n, k, L)-protocols, we show in Sect. 4 that

linear $(n, k, 2)$ protocols can be broken by solving a small number of systems of linear equations in $n(n+1)/2$ unknowns that correspond to overdefined systems of quadratic equations in the keybits. This attack combines a quite obvious approach which was also used in [1] with a special symmetrization technique and several nontrivial transformations in order to get a modified linearized system with a unique solution. Our technique can easily be generalized to algebraic attacks against unrestricted (n, k, L)-protocols, in which overdefined systems of degree-L equations have to be solved. In Sect. 5, we discuss consequences of our results for the practical use of linear (n, k, L)-protocols.

We have experimentally confirmed the correctness and efficiency of our attacks with the computer algebra system Magma [2].

2 Linear (n, k, L)-Protocols

2.1 Definitions

In a linear (n, k, L)-protocol, Alice (the verifier, e.g., an RFID reader) and Bob (the prover, e.g., an RFID tag) share a common secret information (the tag's ID) from a certain keyspace. As usual, we assume that the secret key is hardwired in the RFID tag, while Alice has legal access to a database containing Bob's secret information.

For a positive integer N, we denote by $[N]$ the set $\{1, \ldots, N\}$. The secret keys of linear (n, k, L)-protocols consist of the specifications of L n-dimensional linear subspaces $V_1, \ldots, V_L$ of $GF(2)^{n+k}$, i.e., the key size is $L \cdot n \cdot k$. In particular, for all $l \in [L]$, the subspace V_l is defined by a $GF(2)$-linear mapping $f_l : GF(2)^n \longrightarrow GF(2)^k$ and a permutation $\sigma_l \in \mathcal{S}_{n+k}$ such that

$$V_l = \{\sigma_l(v||f_l(v)), v \in GF(2)^n\} \ .^1 \tag{1}$$

Note that each n-dimensional linear subspace of $GF(2)^{n+k}$ can be represented by a linear mapping f and a permutation σ in the above way (see Appendix A).

[1] For a vector $v = (v_1, \ldots, v_m)$ and a permutation $\sigma \in \mathcal{S}_m$, we define $\sigma(v) = (v_{\sigma(1)}, \ldots, v_{\sigma(m)})$.

In the mode of communication suggested by Cichoń, Klonowski and Kutyłowski [3], abbreviated by CKK-mode in the following, Alice starts the communication by sending some signal triggering Bob to compute a proof w of his identity. In particular, Bob computes $w = \sigma_l(u||f_l(u))$ for randomly (independently and uniformly) chosen $l \in [L]$ and $u \in GF(2)^n$. Alice accepts a proof $\tilde{w}$ if there is an $l \in [L]$ such that $\tilde{w} \in V_l$, or equivalently, if there is an $l \in [L]$ such that $\sigma^{-1}(\tilde{w}) = (u, f_l(u))$ for some $u \in GF(2)^n$.

Obviously, this protocol is vulnerable to replay attacks, since an adversary can store a number of proofs produced by Bob and then impersonate Bob by presenting these proofs to the verifier.

Following [9], this type of attack can be prevented using the following mode of communication, denoted by HB^+-mode in the sequel: Let $n = 2q$. Alice starts the communication by sending a randomly chosen $v \in GF(2)^q$ to Bob. Bob replies with $w = \sigma_l(u||v||f_l(u,v))$ for randomly chosen $l \in [L]$ and $u \in GF(2)^q$. Alice accepts an answer $\tilde{w}$ from Bob if there is an $l \in [L]$ such that $\sigma^{-1}(\tilde{w}) = (u||v||f_l(u,v))$ for some $u \in GF(2)^n$.

2.2 The CKK-Protocols

The protocols CKK^1, CKK^2 and $\mathsf{CKK}^{\sigma,L}$ suggested by Cichoń, Klonowski and Kutyłowski in [3] are restricted types of (n,k,L)-protocols. The basic variant CKK^1 is the $(n,k,1)$-protocol, which does not provide any security.

The protocol $\mathsf{CKK}^{\sigma,L}$ is an (n,k,L)-protocol with the restriction that $f_l = f$ and $\sigma_l = \sigma$ for a secret linear function $f : GF(2)^n \longrightarrow GF(2)^k$, a secret permutation σ, and all $l \in [L]$. Hence, the secret keys have the form (f, σ).

The protocol CKK^2 is the $(n+k,k,2)$-protocol with the additional properties that $f_1 = f_2 = f$ with f only depending on the first n inputs, and that the two permutations σ_1 and σ_2 are publicly known. In particular, σ_2 is the identity id, and σ_1 exchanges the last two blocks of length k of a word of length $n + 2k$. Hence, the secret keys have the form

$$V_1 = \{(v||a||b), v \in GF(2)^n, a, b \in GF(2)^k, a = f(v)\} \ ,$$

$$V_2 = \{(v||a||b), v \in GF(2)^n, a, b \in GF(2)^k, b = f(v)\} \ .$$

CKK^2- and $\mathsf{CKK}^{\sigma,L}$- protocols were suggested for practical application in [3], with the parameters $n = 128$ and $k = 30$.

2.3 Security of linear (n, k, L)-Protocols and our Results

We analyze the security of (n, k, L)-protocols with respect to an adversary Eve who knows that Alice and Bob communicate on the basis of a linear (n, k, L)-protocol. In contrast to the commonly assumed active adversary models (as in [1, 10], for instance), Eve is only able to eavesdrop on the communication between Alice and Bob, and may additionally draw Bob into *quality time* [12], i.e., she may force Bob to send reasonably many (usually $\ll 2^{64}$) honest proofs $w \in GF(2)^{n+k}$, but she does not have any information about the secret key shared by Alice and Bob.

Eve's aim is to produce messages which will be accepted by Alice with high probability. As a first consequence, the probability $L2^{-k}$ that a random $w \in GF(2)^{n+k}$ belongs to $\bigcup_{l=1}^{L} V_l$ should be sufficiently small, i.e., k should be large enough.

In the following, we derive an upper bound for Eve's required effort for cloning Bob, i.e., for computing a specification of Bob's secret key from a set $O = \{w^1, \ldots, w^s\}$ of *observations* (honest proofs produced by Bob). Note that if s is a little larger than Ln, with high probability the set O is *complete*, i.e., it contains a basis for each V_l, $l \in [L]$, and therefore determines Bob's secret key (see Appendix A).

There are several exhaustive search strategies for computing the secret key from a complete set of observations in $2^{\prime(n)}$ time. The interesting questions are whether there are strategies to compute the secret key from a complete set of observations more efficiently, or whether it is at least possible to efficiently distinguish a set of observations induced by a linear (n, k, L)-protocol from a truly random set of vectors from $GF(2)^{n+k}$. In this paper, we answer these questions in the affirmative.

Note that the subspaces $V_1, \ldots, V_L$ should span the whole vector space $GF(2)^{n+k}$, i.e., $V_1 \oplus \cdots \oplus V_L = GF(2)^{n+k}$. Otherwise an efficient distinction from the random case can be easily performed. This implies $Ln \geq n + k$, in particular $n \geq k$ for $L = 2$.

However, the parameter k should not be too large. Consider for instance the case $L = 2$ and $n = k$, and take a set B of $n+k$ random

observations which form a basis of $GF(2)^{n+k}$. Then B contains a basis B' of V_1 or of V_2. This implies that with probability 1/2 the expected representation length w.r.t. to B is only $n/2$, while the expected representation length of a truly random vector is n. [2] Based on this idea, we can recover the secret key efficiently.

So far, the only nontrivial cryptanalytic result concerning linear protocols is due to Gołębiewski, Majcher and Zagórski [5]. They present an attack against the CKK^2-protocol, which cannot be applied to the general case. Its running time is proportional to $\sum_{s=0}^{k-1} \binom{n}{s}$, i.e., of order $n^{\Theta(k)}$. For completeness, we describe the attack in Appendix C.

In the next section we describe two very fast attacks against the CKK^2-protocol. Their running times are dominated by the effort required for inverting k $(n \times n)$-matrices.

In Sect. 4 we describe an algebraic attack against general $(n, k, 2)$-protocols. The idea is to generate, based on a sufficiently large set of observations, for each output bit of the secret functions f_1 and f_2 an appropriate overdefined system of quadratic equations in the values of the output bits over a fixed basis. These systems will then be solved by standard linear algebra methods.

3 Two fast Attacks against CKK^2-Protocols

Let Alice and Bob share a secret key $f : GF(2)^n \longrightarrow GF(2)^k$, and let Bob be designed to prove his identity by means of the CKK^2-protocol. We describe two algorithms for Eve to compute a specification of f.

3.1 The first Attack

The attack is based on a set of observations $O = \{(v_1, a_1, b_1), \ldots, (v_m, a_m, b_m)\}$ of appropriate size. Let $(c_1, \ldots, c_m)$ denote the set of hidden bits behind O, i.e., $f(v_i) = a_i$ if $c_i = 1$ and $f(v_i) = b_i$ if $c_i = 0$. We compute the hidden bits as the unique solution of a system LES of linear

[2] Given a $GF(2)$-vector space V of finite dimension m, a basis $B = \{v_1, \ldots, v_m\}$ of V, and a vector $v \in V$, we call the unique coefficient vector $b = (b_1, \ldots, b_m)$ fulfilling $v = \bigoplus_{i=1}^{m} b_i v_i$ the representation of v w.r.t. to B, and the number of ones in b is called the representation length of v w.r.t. to B.

equations of moderate size. Observe that

$$f(v_i) = c_i a_i \oplus (1 \oplus c_i) b_i \text{ for all } i \in [m] \ . \tag{2}$$

The set $\{v_1, \ldots, v_m\}$ will be a generating system of $GF(2)^n$. Thus, a specification of f can be computed from the hidden bits by standard linear algebra methods.

1. Choose an appropriate number $m > n$ and generate a set of observations $O = \{(v_1, a_1, b_1), \ldots, (v_m, a_m, b_m)\}$, such that $v_1, \ldots, v_m$ form a generating system of $GF(2)^n$ and $\mathbf{0} \notin O$.
2. Initialize the system LES of linear equations in the variables $c_1, \ldots, c_m$ with $LES = \emptyset$.
3. **REPEAT**
   3.1 Choose a nontrivial linear dependency $D \subseteq [m]$, i.e., a set $D \neq \emptyset$ fulfilling $\bigoplus_{d \in D} v_d = \mathbf{0}$.
   3.2 Put the k linear equations defined by $\bigoplus_{d \in D} c_d a_d \oplus (1 \oplus c_d) b_d = \mathbf{0}$ into LES
4. **UNTIL** LES has rank m
5. Compute the vector $(c_1, \ldots, c_m)$ as the unique solution of LES.

The correctness of the attack follows directly from (2). Some remarks on a possible implementation can be found in Appendix D.

Our experiments show that the number of rounds necessary to make LES of full rank is only slightly larger than n. Table 1 shows the performance of our attack on a few example parameter combinations.

Table 1. Performance of the first attack on CKK^2

(n, k)	approx. number of observations	approx. attack time
$(128, 30)$	140	0.05 s
$(1024, 256)$	1039	2.95 s

3.2 The second Attack

Let the single output functions $f^1, \ldots, f^k : GF(2)^n \longrightarrow GF(2)$ denote the component functions of the secret function f, i.e., $f(v) =$

$(f^1(v), \ldots, f^k(v))$ for all $v \in GF(2)^n$. The attack is based on the simple fact that if an observation $(v, (a(1), \ldots, a(k)), (b(1), \ldots, b(k)))$ satisfies $a(r) = b(r)$ for some $r \in [k]$, which is true with probability $1/2$, then $f^r(v) = a(r) = b(r)$. The attack is defined as follows:

1. Let $e_1, \ldots, e_n$ denote the standard basis of $GF(2)^n$.
2. For $r \in [k]$ do
   2.1 Generate a set of obsevations $O_r = ((v_{r,1}, a_{r,1}, b_{r,1}), \ldots, (v_{r,n}, a_{r,n}, b_{r,n}))$ such that $v_{r,1}, \ldots, v_{r,n}$ form a basis of $GF(2)^n$ and $a_{r,i}(r) = b_{r,i}(r) = f^r(v_{r,i})$ for all $i \in [n]$.
   2.2 Compute from this $f^r(e_1), \ldots, f^r(e_n)$.

The correctness of the algorithm follows straightforwardly from the definitions. The expected number of observaions needed for constructing O_r is $2 \cdot E(n)$, with $E(n)$ defined as in Appendix A.

In contrast to the first attack on CKK^2, we can only exploit an observation (v, a, b) if $a(r) = b(r)$ for some r, which implies that the amout of data needed to recover the secret key will be higher than before. Also the computation is a little more costly since k Gaussian eliminations are performed to compute the functions f^r w.r.t. the standard basis. Nevertheless, the attack is still very efficient for practically proposed parameter choices, see Table 3.2.

Table 2. Performance of the second attack on CKK^2

(n, k)	approx. number of observations	approx. attack time
$(128, 30)$	311	0.3 s
$(1024, 256)$	2197	179 s

4 Algebraic Attacks against $(n, k, 2)$-protocols

4.1 Preliminaries

For describing the attack we need some simple facts on the structure of linear subspaces.

Given a vector $v \in GF(2)^{n+k}$, a subset $S \subseteq GF(2)^{n+k}$ and an index set $J = \{j_1 < \ldots < j_s\} \subseteq [n + k]$, we denote by $v|_J$ the

restriction of v w.r.t. J, i.e., $v|_J = (v_{j_1}, \ldots, v_{j_s}) \in GF(2)^s$. Further let $S|_J = \{v|_J, v \in S\}$. Observe that if S is a linear subspace then so is $S|_J$.

We denote by $GF(2)^J$ the set of $GF(2)$-assignments $c : J \longrightarrow GF(2)$ to J. For I, J disjoint and nonempty subsets of $[n+k]$ and $c \in GF(2)^I$, $d \in GF(2)^J$, we denote by $c \cup d$ the unique assignment z of $I \cup J$ with $z|_I = c$ and $z|_J = d$.

Let $V \subseteq GF(2)^{n+k}$ be an n-dimensional subspace of $GF(2)^{n+k}$. A nonempty index set $J \subseteq [n+k]$, $|J| = n$ is called a basis set w.r.t. V if $\dim(V|_J) = n$. For a fixed basis $B = \{v^1, \cdots, v^s\}$ of V let us denote by $M = M(V, B)$ the $n \times (n+k)$-matrix formed by the rows $v^1, \cdots, v^n$. The following Lemma can be easiliy proved by standard linear algebra arguments.

Lemma 1. *(i) A set $J \subseteq [n+k]$, $|J| = n$, is a basis set w.r.t. V if the n rows of M corresponding to the indices in J form a basis of $GF(2)^n$.*

(ii) If $J \subseteq [n+k]$ is a basis set w.r.t. V then there is a linear mapping $f : GF(2)^n \longrightarrow GF(2)^k$ such that

$$V = \{w \cup f(w), w \in GF(2)^J\}.$$

Note that (ii) implies that V can be represented by f and a permutation $\sigma \in \mathcal{S}_{n+k}$ as

$$V = \{\sigma(v, f(v)), v \in GF(2)^n\},$$

(take a permutation σ, which maps J to $\{1, \cdots, s\}$). Further we obtain as a corollary

Lemma 2. *If $J \subseteq [n+k]$ is a basis set w.r.t. V then for all $i \in [n+k] \setminus J$ the following holds. There is a linear functional $g : GF(2)^n \longrightarrow GF(2)$ such that for all $v \in V$ it holds that $v_i = g(v|_J)$.* □

4.2 Outline of the Attack

The weakness of CKK^2-protocols is that observations (v, a, b) contain the information about $f(v)$, the only problem is to decide whether $f(v) = a$ or $f(v) = b$. This is not the case for general $(n, k, 2)$-protocols; a single observation $\sigma_1(v, f_1(v))$ does not say anything

about $f_2(v)$, and vice versa. We describe an algebraic attack based on the observation that linear dependencies in a set of observations can be translated into nontrivial equations in the keybits. The difference to the attack in Sect. 3 is that we have to introduce unknowns corresponding to the values $f_1(v^i)$ and $f_2(v^i)$, and that we get quadratic equations instead of linear ones.

The attack is based on the following algorithm $ANALYZE$, which can be considered as an algebraic attack against $(n, 1, 2)$-protocols with secret key $(f_1, \sigma_1), (f_2, \sigma_2)$, $f_1, f_2 : GF(2)^n \longrightarrow GF(2)$ linear, for which $\sigma_1 = \sigma_2 = id$.

Consider a set of observations $(v^1, w_1), \cdots, (v^m, w_m)$, $v^j \in GF(2)^n$, $w_j \in GF(2)$ for $j = 1, \cdots, m$, m sufficiently large, induced by such an $(n, 1, 2)$-protocol. The algorithm $ANALYZE$ computes specifications of the two linear mappings $f_1, f_2 : GF(2)^n \longrightarrow GF(2)$ such that $f_1(v^j) = w_j$ or $f_2(v^j) = w_j$ for all $j = 1, \cdots, m$, i.e., the specifications of the two secret vector spaces $V_l = \{(v, f_l(v)), v \in GF(2)^n\}$ and $V_2 = \{(v, f_2(v)), v \in GF(2)^n\}$.

The algorithm $ANALYZE$ can be used to attack the general case $k > 1$ as follows. Let $\{v^1, \cdots, v^m\} \subseteq GF(2)^{n+k}$ denote a set of observations induced by the application of an $(n, k, 2)$-protocol with a secret key V_1, V_2 corresponding to two n-dimensional linear subspaces of $GF[2]^{n+k}$, m sufficiently large.

1. **REPEAT**
2. Choose a set $I = \{i_1 < i_2 < \cdots < i_n\} \subseteq [n + k]$ and some $i \in [n + k] \setminus I$
3. Apply ANALYZE to the transformed observations $\{\tilde{v^1}, \cdots, \tilde{v^m}\} \subseteq GF(2)^{n+1}$, where for all $v \in GF(2)^{n+k}$ we define $\tilde{v} = (v_{i_1}, \cdots, v_{i_n}, v_i)$.
4. **UNTIL** ANALYZE successfully produces specifications of two distinct n-dimendional linear subspaces $W_1, W_2 \subseteq GF(2)^{n+1}$ given by two linear mappings $f_1, f_2 : GF(2)^n \longrightarrow GF(2)$.

W_1, W_2 can be used to compute specifications of the secret n-dimensional subspaces $V_1 \subseteq W_1$ and $V_2 \subseteq W_2$ of $GF(2)^{n+k}$ as follows. The probabilities for the three possible events that a random observation belongs to $W_1 \setminus W_2$, resp. to $W_2 \setminus W_1$, resp. to $W_1 \cap W_2$ should be the same and near $1/3$. The first two events allow to determine if $v^i \in V_1$ or if $v^i \in V_2$.

The sample size number m is determined by the minimal number of observations necessary for successfully applying $ANALYZE$. Thus, m is much grater than $3 \times E(n)$ (see Appendix). This implies that with high probability, $v^1, \cdots, v^m$ contain subsets $B_1 \subseteq W_1 \setminus W_2$ and $B_2 \subseteq W_2 \setminus W_1$ such that B_1 is a basis of V_1 and B_2 is a basis of V_2.

The estimation of the success probability of the attack is based on the hypothesis that the secret subspaces V_1 and V_2 are randomly and independently chosen according to the following experiment.

1. **Repeat**
2. Choose randomly $n \times n + k$-matrix M over $GF(2)$ w.r.t. the uniform distribution.
3. **UNTIL** $rank(M) = n$.
4. Take V as the linear span of the rows of M.

Under this assumption, the probability that a fixed set $I \subseteq [n+k]$, $|I| = n$ is a basis set of V_1 and V_2 is $p(n)^2 \approx (0.2887)^2 \approx 0.083$ (see Appendix A). Experiments with small values of n and k show that on average around 12 different sets I have to be tried before finding one which is a basis set for both V_1 and V_2. In the following subsection we describe the algorithm $ANALYZE$.

4.3 The Algorithm *ANALYZE*

1. Choose a set $O = \{(v^1, w_1), \ldots, (v^n, w_n)\} \subseteq GF(2)^{n+1}$ of observations such that $B = \{v^1, \ldots, v^n\}$ forms a basis of $GF(2)^n$. For $i \in [n]$ let x_i and y_i denote the variables corresponding to $f_1(v^i)$ and $f_2(v^i)$, respectively.
2. For $b \in \{0, 1\}$ let $I_b = \{i \in [n], w_i = b\}$.
3. For all $i \in [n]$ let $t_i = x_i \oplus y_i$, and for all $i < j \in [n]$ let $t_{i,j} = x_i y_j \oplus x_j y_i$.
4. Observe that for all $i \in [n]$ the equality $(w_i \oplus x_i)(w_i \oplus y_i) = 0$ holds. This implies

$$x_i y_i = 0 \text{ if } i \in I_0 \text{ and } x_i y_i = 1 \oplus t_i \text{ if } i \in I_1. \tag{3}$$

5. Observe that for each observation $(v, w) \in GF(2)^{n+1}$, $v \notin B$, the following holds: If $v = \bigoplus_{i \in I} v_i$, (i.e., $I \subseteq [n]$ defines the unique

representation of v w.r.t. B), then

$$\left(w \oplus \bigoplus_{i \in I} x_i\right)\left(w \oplus \bigoplus_{i \in I} y_i\right) = 0 \tag{4}$$

Observe that relation (4) can be rewritten as a relation $T_B(I, w)$ in the variables t_i and $t_{i,j}$ in the following way: If $w = 0$ then relation (4) is equivalent to $\bigoplus_{i \in I} x_i y_i \oplus \bigoplus_{i<j \in I} t_{i,j} = 0$. Together with relation (3) this implies $\bigoplus_{i \in I_1 \cap I}(t_i \oplus 1) \oplus \bigoplus_{i<j \in I} t_{i,j} = 0$ for $w = 0$. Consequently, for $w = 0$ we define $T_B(I, w)$ as

$$\bigoplus_{i \in I \cap I_1} t_i \oplus \bigoplus_{i<j \in I} t_{i,j} = \begin{cases} 0 & \text{if } |I \cap I_1| \text{ is even} \\ 1 & \text{if } |I \cap I_1| \text{ is odd} \end{cases}.$$

If $w = 1$ then relation (4) is equivalent to $1 \oplus \bigoplus_{i \in I} t_i \oplus \bigoplus_{i \in I \cap I_1}(t_i \oplus 1) \oplus \bigoplus_{i<j \in I} t_{i,j} = 0$. Hence, for $w = 1$ we define $T_B(I, w)$ as

$$\bigoplus_{i \in I \cap I_0} t_i \oplus \bigoplus_{i<j \in I} t_{i,j} = \begin{cases} 0 & \text{if } |I \cap I_1| \text{ is odd} \\ 1 & \text{if } |I \cap I_1| \text{ is even} \end{cases}.$$

Note that a relation similar to relation (4) was also exhibited in [1] for designing an algebraic attack against so-called F_f-protocols.

Our attack works as follows.

1. Let initially the system LES of linear equations in the $\frac{1}{2}(n^2+n)$ variables t_i ($i \in [n]$) and $t_{i,j}$ ($i < j \in [n]$) be empty.
2. **REPEAT**
   2.1 Choose an observation (v, w), $v \notin B \cup \{\mathbf{0}\}$ and compute the unique subset $I \subseteq [n]$ with $v = \bigoplus_{i \in I} v^i$.
   2.2 Enlarge the system LES by the linear equation $T_B(I, w)$.
3. **UNTIL** The system LES has $\frac{1}{2}(n^2 + n)$ linearly independent equations.
4. Compute by Gaussian elimination the unique solution θ of the system LES.
5. Compute from θ the unique correct assignment to x_i, y_i for all $i \in [n]$.

The correct assignments to the x_i and y_i variables (step 5 of the algorithm) can be computed from $\theta = (\theta_i)_{i \in [n]}\,(\theta_{i,j})_{i<j \in [n]}$ as follows.

For $b = 0, 1$ let K_b denote the set $K_b = \{i \in [n], \theta_i = b\}$. We know that for all $i \in K_0$ it holds that $x_i = y_i = w_i$, and for all $i \in K_1$ it holds that $y_i = x_i \oplus 1$. This implies that for all $i < j$ in K_1, $\theta_{i,j}$ satisfies

$$\theta_{i,j} = x_i(x_j \oplus 1) \oplus x_j(x_i \oplus 1) = x_i \oplus x_j \ .$$

This yields a system LES^* of $1/2|K_1|(|K_1| - 1)$ linear equations in the variables x_i, $i \in K_1$, of rank $|K_1| - 1$. As it does not matter which of the two secret linear subspaces we denote by V_1 and which by V_2, we have the freedom to set $x_k = 0$ for some fixed $k \in K_1$. The system LES^* together with $x_k = 0$ yields a system of full rank and allows to compute the correct assigment to the x_i-variables by Gaussian elimination.

4.4 Analysis and experimental Results

The background for the fact that the repeat cycle of the algorithm $ANALYZE$ is left after a finite number of rounds is that the following $(2^n - (n + 1)) \times (n(n + 1)/2)$-matrix $M(n)$ over $GF(2)$ has full row rank (which is not hard to show). The row indicess of $M(n)$ are all subsets $I \subseteq [n]$ with $|I| \geq 2$, the column indices are $[n] \cup \{(i, j), 1 \leq i < j \leq n\}$. We have $M(n)_{I,i} = 1$ iff $i \in I$ and $M(n)_{I,(i,j)} = 1$ iff $\{i, j\} \subseteq [n]$.

We do not give here a theoretical analysis of the expected number of rounds of the repeat cycle. Our experiments show that the algorithm $ANALYZE$ needs only slightly more than $\frac{1}{2}(n^2 + n) + n$ observations to compute the secret functions f_1 and f_2. Particularly for $n = 128$, in order to recover the secret functions we need approx. 8390 observations and 4 minutes.

5 Summary

We have seen that the secret key of CKK^2-protocols can be computed very quickly from a sufficiently large set of observations. This kind of protocol should not be used in practice.

Our degree-2 algebraic attack against $(n, k, 2)$-protocols can be quite straightforwardly generalized to a degree-L attack against (n, k, L)-protocols for $L > 2$, which implies solving an overdefined system of

degree L equations. Using the technique of linearization, for $n = 128$ and $L \geq 4$ this means to solve a system of linear equations in more than one billion variables, which is not feasible. It is an interesting open question if the very symmetrically structured systems of degree-L equations arising during a degree-L attack can be more efficiently solved by more advanced techniques like the F4- or F5-algorithm or cube attacks [6, 7, 4]. If one could generate convincing evidence such that algorithms cannot beat linearization attacks, then linear $(128, 30, L)$-protocols could be seriously considered for practical use.

Another problem of linear (n, k, L)-protocols is the large keylength; when using a naive implementation, the hardware size of the secret key is nkL. Thus, it is an important question to look for special kinds of (n, k, L)-protocols for which there are implementations of the secret subspaces $V_1, \ldots, V_L$ which need significantly less than nkL gates. One possibility is to look for efficient hardware realizations for the $\mathsf{CKK}^{\sigma,L}$-protocols suggested in [3].

Our attack against (n, k, L)-protocols yields only one of the possibly exponentially many equivalent representations of the secret subspaces. This implies that it correctly clones the secret key only if the CKK-mode of communication is used. When the HB^+-mode is used, Eve has to compute the specification of (f_l, σ_l) correctly for $l \in [L]$. How this can be done efficiently is a subject of further research.

Acknowledgement

We are very thankful to Mirek Kutyłowski for introducing us in the topic.

References

1. E.-O. Blass, A. Kurmus, R. Molva, G. Noubir, and A. Shikfa. The F_f-family of protocols for RFID-privacy and authentication. available online at `http://eprint.iacr.org/2008/476`.
2. W. Bosma, J. Cannon, and C. Playoust. The magma algebra system. i. the user language. *J. Symbolic Comput.*, 24(3-4):235–265, 1997.
3. J. Cichoń, M. Klonowski, and M. Kutyłowski. Privacy protection for RFID with hidden subset identifiers. In *Proc. of Pervasive 2008*, volume 5013 of *LNCS*, pages 298–314. Springer, 2008.
4. I. Dinur and A. Shamir. Cube attacks on tweakable block ciphers. Cryptology ePrint Archive, Report 2008/385, 2008. http://eprint.iacr.org.

5. Z. Gołębięwski, K. Majcher, and F. Zagórski. Attacks on CKK family of RFID authentication protocols. In *Proc. Adhoc-now 2008*, volume 5198 of *LNCS*, pages 241–250. Springer, 2008.
6. J.-C. Faugère. A new efficient algorithm for computing Gröbner bases (F4). *Journal of pure and applied algebra*, 139(1-3):61–68, 1999.
7. J.-C. Faugère. A new efficient algorithm for computing Gröbner basis without reduction to zero (F5). In *Proc. of the 2002 international symposium on Symbolic and algebraic computation (ISSAC)*, pages 75–83. ACM Press, 2002.
8. A. Juels. RFID privacy: A technical primer for the non-technical reader. In *Privacy and Technologies of Identity: A Cross-Disciplinary Conversation*. Springer, 2005.
9. A. Juels and S. A. Weis. Authenticating pervasive devices with human protocols. In *Proc. of Crypto 2005*, volume 3621 of *LNCS*, pages 293–308. Springer, 2005.
10. A. Juels and S.A. Weis. Defining strong privacy for RFID. In *Proc. of PerComW'07*, pages 342–347, 2007.
11. M. Langheinrich. A survey of RFID privacy approaches. In *Workshop on Ubicomp Privacy - Technologies, Users, Policy. Workshop at Ubicomp 2007*, 2007.
12. T. van Deursen and S. Radomirovic. Attacks on RFID protocols. Cryptology ePrint Archive, Report 2008/310, 2008. http://eprint.iacr.org.

A Generating a random Basis

Let us consider the following experiment.

1. Set $B := \emptyset$.
2. **REPEAT**
   2.1 Choose a random $v \in GF(2)^n$ (w.r.t. the uniform distribution)
   2.2 $V := V \cup \{v\}$.
3. **UNTIL** V is a generating system of $GF(2)^n$.

Let $p(n)$ denote the probability that the experiment stops after n iterations (i.e., V is a basis of $GF(2)^n$), and $E(n)$ denote the expected number of iterations of the experiment. It is known that $p(n) \approx 0.2887$ and $E(n) \approx n + 1.6067$ (see, e.g., [5]).

B Some Basics on linear Subspaces

An index set $I \subseteq [m]$ is called basis set w.r.t. V if $|I| = s$ and $\dim(V|_I) = s$. The following facts can be easily dervied by standard linear algebra arguments.

Lemma 3. *Let $V \subseteq GF(2)^m$ denote an s-dimensional linear subspace of $GF(2)^m$ for some $0 < s < m$.*

(i) *If $I \subset [m]$ is a basis set w.r.t. V and $J \subseteq [m] \setminus I$, then there is a linear mapping $f : GF(2)^I \longrightarrow GF(2)^J$ such that for all $u \in V$ it holds that $u|_J = f(u|_I)$.*

(ii) *If $I \subset [m]$ is a basis set w.r.t. V and $J = [m] \setminus I$, then there is a linear mapping $f : GF(2)^I \longrightarrow GF(2)^J$ such that $V = \{v \cup f(v), v \in GF(2)^I\}$.*

(iii) *Let I, J, f be defined as in (ii) and let f only depend on a set of indices $I' \subset I$ (i.e., $f(v) = f(v|_{I'}$ for all $v \in GF(2)^I$). Then each subset of $I \setminus I'$ is not dependent.*

(iv) *If $J \subseteq [m]$ is dependent and $K = [m] \setminus J$ then there is a linear mapping $g : GF(2)^K \longrightarrow GF(2)^J$ such that $V = \{v \cup g(v), v \in GF(2)^K\}$.*

Proof. Statements (i)-(iii) can be derived by standard linear algebra arguments. In order to prove (iv) fix a basis set $I \subseteq K$ and a mapping $f : GF(2)^I \longrightarrow GF(2)^J$ like in (i) and define $g : GF(2)^K \longrightarrow GF(2)^J$ by $g(v) = f(v|_I)$ for all $v \in GF(2)^K$. □

Altogether, we obtain

Lemma 4. *Each n-dimensional subspace V of $GF(2)^{n+k}$ can be represented by a linear function $f : GF(2)^n \longrightarrow GF(2)^k$ and a permutation σ of $[n+k]$ such that $V = \{(\sigma(v, f(v))), v \in GF(2)^n\}$.*

Proof. Take a basis set $I \subseteq [n+m]$, let $J = [n+m] \setminus I$ and fix a linear mapping f as in Lemma 3, (ii). The permutation σ has to be chosen in such a way that $\sigma([n]) = I$. □

C The Attack of Gołębiewski, Majcher and Zagórski against the CKK²-Protocol

[5] contains the following attack against CKK^2 . The attack refers to a set of observations

$$O = ((v_1, a_1, b_1), \ldots, v_T, a_T, b_T))$$

generated by Bob on the basis of a secret key (i.e. $GF(2)$-linear function) $f : GF(2)^n \longrightarrow GF(2)^k$ and hidden random bit c_i, i.e.,

$$f(v_i) = c_i a_i \oplus (1 \oplus c_i) b_i \ .$$

The aim of the attack is to compute a subset $I \subseteq [T]$ such that $\{v_i, i \in I\}$ forms a basis of $GF(2)^n$, and to compute the corresponding values c_i for all $i \in I$. The idea of the attack is to

1. try to find a set $J \subseteq [T]$, $|J| \leq k$, with $\bigoplus_{j \in J} v_j = \mathbf{0}$,
2. do exhaustive search in the set of all $2^{|J|}$ $\{0,1\}$-assignments of J in order to find the set of bits $\{c_j, j \in J\}$ with

$$\bigoplus_{j \in J} (c_j a_j \oplus (1 \oplus c_j) b_j) = \mathbf{0} \ ,$$

and to repeat these steps until a sufficiently large set of v_i-vectors with the right c_i-bit is identified.

Step 1 is performed by choosing a basis B among the v_i-vectors and searching for a v_j outside B such that $v_j = \bigoplus_{i \in B'} v_i$ for a subset $B' \subseteq B$ with $|B'| < k$. The probability of finding such a v_j is

$$p(n,k) = 2^{-n} \sum_{s=0}^{k-1} \binom{n}{s} \ ,$$

i.e., the worst case running time has order $(n/k)2^k p(n,k)^{-1}$.

D Analysis and Implementation of the Attack in Section 3.1

We derive some restrictions for the choice of m, O and the linear dependencies D which should minimize the number of repetitions of the repeat cycle.

Note that the linear subsystem in step 3.2 can be written as

$$\bigoplus_{d \in D} c_d \, (a_d \oplus b_d) = \mathbf{0} \ .$$

For $D \subseteq [m]$ let $v(D) \in GF(2)^m$ denote the characteristic vector of D, i.e., for all $i \in [m]$ let $v(D)_i = 1$ if $i \in D$ and $v(D)_i = 0$ if not.

For $t = 1, 2, \ldots$ let D_t denote the linear dependencies chosen in the t-th iteration of the repeat cycle. For minimizing the number of iterations, linear dependency D_t should for all t be chosen in such a way that

– $\{v(D_1), \ldots, v(D_t)\}$ are linearly independent,
– the set of vectors $\{a_d \oplus b_d, d \in D_t\}$ form a generating system of $GF(2)^k$.

One way to achieve this is to

1. set $m = n + s$ for appropriate s,
2. to choose $v_1, \ldots, v_n$ such that they form a basis B of $GF(2)^n$,
3. to choose v_{n+t} for $t \in [s]$ in the following way. If $v_{n+t} = \bigoplus_{d \in D'} v_d$, $D' \subseteq [n]$, is the unique representation of v_{n+t} w.r.t. B, then $\{a_d \oplus b_d, d \in D_t\}$ forms a generating system of $GF(2)^k$, where $D_t = D' \cup \{n + t\}$.

On average, we can extract $v_1, \ldots, v_n$ from a set of $E(n)$ observations. Note that when choosing v_{n+t}, the expected size of D' is $n/2$. If we suppose that $n/2 \geq E(k)$, we see that only a few observations should suffice for extracting a vector v_{n+t} fulfilling the required conditions.

How large s should be? We suppose that the way we constructed LES pumps enough randomness into the linear equation, so that $E(n+s)$ equations should be enough for guaranteeing $n+s$ linearly independent equations, i.e., $sk \geq E(n+s)$, which implies $s \geq \frac{n+C}{k-1}$ for a small constant C.