Formal verification of protocols based on short authenticated strings
Stéphanie Delaune, Steve Kremer, Ludovic Robin

To cite this version:

HAL Id: hal-01528607
https://inria.hal.science/hal-01528607
Submitted on 29 May 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Formal verification of protocols based on short authenticated strings

Stéphanie Deleaune†, Steve Kremer‡, and Ludovic Robin†
†CNRS & IRISA, Rennes, France
‡LORIA, INRIA Nancy - Grand-Est, France

Abstract—Modern security protocols may involve humans in order to compare or copy short strings between different devices. Multi-factor authentication protocols, such as Google 2-factor or 3D-secure are typical examples of such protocols. However, such short strings may be subject to brute force attacks. In this paper we propose a symbolic model which includes attacker capabilities for both guessing short strings, and producing collisions when short strings result from an application of weak hash functions. We propose a new decision procedure for analysing (a bounded number of sessions of) protocols that rely on short strings. The procedure has been implemented in the AKISS tool and tested on protocols from the ISO/IEC 9798-6:2010 standard.

I. INTRODUCTION

Modern security protocols often include the human as a participant on his own, i.e., separately from the device he may own and use. Such security protocols are sometimes called security ceremonies [18]. In this paper we are interested in protocols where a human has to copy, or compare short strings, or verification codes on two devices. This kind of situations occurs rather frequently. In multi-factor authentication protocols, such as 3D-secure [1] or Google 2-factor [21], a short verification code is sent by SMS to the user’s phone who has to copy it to another device (to prove ownership of his phone). Another example is bluetooth tethering [9], [10] where the human user has to compare codes displayed by two devices to be connected. There has also been a recent standard [23] defining several mechanisms for “manual” data transfer: these protocols use a human to ensure that two devices can exchange some data in an authentic way without sharing any cryptographic keys, nor relying on a PKI. One important aspect is that the user introduces a low bandwidth channel: he is only able to copy (or compare) short strings. The ISO/IEC 9798-6:2010 standard [23] suggests that such short strings are 16-20 bits long. As pointed out in [25], [26], traditional methods are not suitable to reason about the security of protocols over non standard channels, and the purpose of this paper is to pursue the effort towards formalizing protocols that rely on low bandwidth channels.

As an illustrative example, we may consider the simplified commitment scheme [26] which is also described below. This protocol relies on the existence of two empirical channels.

The research leading to these results has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (grant agreements No 645865-SPOOC and No 714955-POPSSTAR), as well as from the French National Research Agency (ANR) under the project Sequoia.

According to [26], an empirical channel is a low bandwidth channel that is

- asynchronous: the attacker may delay the delivery,
- authentic: the attacker may not inject messages, and
- public: the channel does not guarantee confidentiality.

Such a channel is typically implemented by an out-of-band channel to a human who copies the data. Message transmission on an empirical channel is denoted by \( \rightarrow_e \):

1. \( A \rightarrow B : \text{info}_A, \text{hash}(\text{info}_A, n_A, r_A) \)
2. \( B \rightarrow_e A : \text{ack} \)
3. \( A \rightarrow B : n_A \)
4. \( A \rightarrow_e B : r_A \)

The goal of this protocol is to ensure that \( B \) will receive the information \( \text{info}_A \) as sent to him by \( A \). Here \( r_A \) is a short random value, whereas \( n_A \) is a long random nonce. We will now explain the need of the acknowledgment (sent in message 2) and the nonce \( n_A \) (included in the hash and sent in message 3). If we remove the nonce \( n_A \) we obtain a weaker protocol. Indeed, once the attacker has obtained the modified message 1, i.e. \( \text{info}_A, \text{hash}(\text{info}_A, r_A) \), he can compute the value of the short random \( r_A \) by brute force. Once \( r_A \) is known, the attacker can send an alternative information \( \text{info}_1 \) of his choice to \( B \) by modifying the first message to \( \text{info}_1, \text{hash}(\text{info}_1, r_A) \). If we remove message 2 a similar attack is possible as the attacker can just intercept messages 1 and 3 and perform a similar brute force attack.

The above example illustrates that short strings, that are subject to brute force, or guessing attacks are extremely tricky. One of the most successful approaches for analysing security protocols is the symbolic approach for automated security protocol analysis, based on the seminal ideas of Dolev and Yao [17]. Indeed, there exist today mature tools for analysing security protocols [7], [28], [4], [19], which have been used to successfully analyse real world protocols such as, e.g., authentication standards [5]. With the exception of a few tools (which we discuss in our related work section below), existing analysis tools do however not take into account adversary capabilities for reasoning about weak secrets. Therefore, we devise in this paper a symbolic protocol verification technique that allows to take such guesses into account.

Our contributions. In this paper we extend the AKISS verification tool [11] and the underlying theory to take into account protocols that manipulate short secrets that are subject
to brute force attacks. AKISS allows automated analysis of security protocols when restricted to a bounded number of sessions. Cryptographic primitives may be defined through arbitrary convergent equational theories that have the finite variant property. This class of theories includes standard cryptographic primitives such as symmetric and asymmetric encryption, digital signatures, hash functions, …, as well as less commonly supported primitives such as blind signatures and zero knowledge proofs. As we will see the framework is actually flexible enough to model weak hash functions, i.e., hash functions whose output is short enough to be copied by a human. Termination is guaranteed for theories that are subterm convergent, but the tool may terminate in practice on protocols relying on other theories [11]. Protocols in AKISS are written in a process calculus that is similar to the applied pi calculus [2]. Our first contribution is an extension of the semantics to allow an attacker to learn a weak secret during the execution: the fact that a secret may be guessed is expressed by a test that succeeds if the attacker is given the secret to be guessed, but fails if the secret is replaced by a fresh name. We note that our definitions are stated for arbitrary equational theories. Our modeling is reminiscent of the treatment of guessing attacks in password based protocols [14] (see the related work below). Next, building on the theory underlying AKISS, we design a novel verification procedure to verify authentication properties expressed as reachability properties for our new semantics. Unlike the original theory underlying AKISS we also allow disequality tests (else branches) when verifying reachability properties.

The procedure has been implemented in the AKISS tool and used to analyze several protocols from the ISO/IEC 9798-6:2010 standard [23]. In addition to weak secrets, we also considered weak hash functions, i.e., hash functions whose output is short and that are subject to collisions. We encode the attacker’s capability to find collisions through an equational theory in a way that is similar to [12]. Our case encode the attacker’s capability to find collisions through an equational theory in a way that is similar to [12]. Our case study studies also required us to consider different kind of channels, depending on whether out-of-band channels are supposed to be synchronous or asynchronous, public or confidential and authentic or not. We propose an encoding for each of these different kinds of channels.

Related work. The works closest to ours are the ones by Roscoe et al. [26] and Delaune and Jacquemard [15]. In [26], they propose the use of the process algebra CSP and the model-checker FDR to analyze protocols relying on weak secrets. Their tool is however restricted to messages of finite length and particular equational theories, i.e., they analyze a finite state system and do not allow for user specified equational theories. In [15], they propose a decision procedure for solving a similar problem than the one studied here. However, they rely on a simpler setting, e.g. no disequality tests, only a fixed set of primitives preventing them to model weak hash functions. More importantly, they do not provide any tool support, and their procedure is actually far from being implementable. In [12], Chothia et al. use the ProVerif tool [8] to analyze commitment protocols. Their work inspired our encoding of weak hash functions, for which collisions may be found, through an equational theory. However, they do not offer support for short random secrets that may be guessed and learned by the adversary during a protocol execution.

Our definition of whether a secret is guessable is similar to the definition of dictionary attacks in password-based protocols [14], which can be encoded in terms of static equivalence and verified using tools able to check equivalence properties, such as [8], [11]. A similar definition was also proposed by Lowe [24] in the CSP process algebra. However, when verifying dictionary attacks we are only interested in knowing whether the password may be learned or not. In the protocols studied in this paper, the short secrets are often revealed anyway (see message 4 in the example above). Therefore, deciding whether a weak name is guessable is not a relevant question in the protocols studied here, and existing approaches cannot be used. Here, we need to dynamically add knowledge of the secret to the attacker knowledge. This is technically more challenging as we may need to consider different secrets that may be needed to be guessed in a same execution for a successful attack. Multiple short secrets naturally arise as soon as several sessions are considered. In particular, using symbolic techniques (that do not explicitly represent the infinite number of all possible executions), it is not sufficient to check separately whether secrets are guessable, as one may need to guess one secret to be able to guess a second one.

There has also been interesting work on modeling human behavior in protocols [6]. We do not consider this aspect, which is orthogonal, in our work.

All files related to the tool implementation and case studies are available at [3] and full proofs are provided in the long version of this paper [16].

Outline. In Section II, we recall some basic definitions for modelling messages. Then, in Section III, we explain how the attacker may extract information from the messages exchanged in the course of a protocol execution. We introduce our formalism for modelling protocols in Section IV. We restrict ourselves to a minimalistic core calculus which is sufficient to develop our main result. Later on (see Section VII), we will introduce a richer calculus and explain how this richer model is translated into the former. In Section V, we provide an abstract representation of processes expressed in our core calculus as Horn clauses. Next, we present our saturation procedure based on Horn clause resolution in Section VI. We present our algorithm, as well as its integration in AKISS in Section VII. Lastly, in Section VIII, we report the results we obtained when analysing two protocols from the ISO/IEC 9798-6:2010 standard.

II. Preliminaries

A. Term algebra

As usual in symbolic models we represent messages as terms. We consider three infinite and disjoint sets of names: $N_{prv}$ is the set of private names, which are used to represent
nonces or keys generated by honest participants, \(N_{\text{pub}}\) is the set of public names, which are used to represent identifiers available both to the attacker and to honest participants, as well as attacker nonces, and \(N_{\text{guess}}\) is the set of guessable names, which are used to represent data with low entropy such as passwords or PIN numbers. Let \(\mathcal{N} = N_{\text{pub}} \uplus N_{\text{pub}} \uplus N_{\text{guess}}\). We consider two infinite and disjoint sets of variables, denoted \(\mathcal{X}\) and \(\mathcal{Y}\). Variables in \(\mathcal{X} = \{x, y, \ldots\}\) typically refer to unknown parts of messages experienced by participants while variables in \(\mathcal{Y} = \{w_1, w_2, \ldots\}\) are used to store messages known by the attacker.

We consider a signature \(\Sigma\), i.e., a finite set of function symbols together with their arity. As usual, a function symbol of arity 0 is called a constant. Given a signature \(\Sigma\) and a set of atoms \(A\) we denote by \(T(\Sigma, A)\) the set of terms, defined as the smallest set that contains \(A\) and that is closed under application of function symbols in \(\Sigma\). We denote by \(\vars(t)\) the set of variables occurring in a term \(t\). As usual, a substitution is a function from variables to terms, that is lifted to terms homomorphically. The application of a substitution \(\sigma\) to a term \(u\) is written \(u\sigma\), and we denote \(\text{dom}(\sigma)\) its domain, i.e., \(\text{dom}(\sigma) = \{x \mid \sigma(x) \neq x\}\). We denote the identity substitution, whose domain is the empty set, by \(\varnothing\). The positions of a term are defined as usual.

We associate an equational theory \(E\) to the signature \(\Sigma\). The equational theory is defined by a set of equations of the form \(u = v\) where \(u, v \in T(\Sigma, \mathcal{X})\), and induces an equivalence relation over terms: \(\equiv\) is the smallest congruence relation on \(T(\Sigma, \mathcal{X})\), which was first introduced in [13].

**Definition 1:** Consider a rewrite system \(R\) that is convergent, and a set of terms \(T\). A set of substitutions \(\text{vars}_R(T)\) is called a complete set of variants for the set of terms \(T\), if for any substitution \(\omega\) there exist \(\sigma \in \text{vars}_R(T)\), and a substitution \(\tau\) such that \(x\omega\downarrow = \sigma x\downarrow\tau\) for any \(x \in \vars(T)\), and \((t\omega)\downarrow = (t\sigma)\downarrow\tau\) for any \(t \in T\).

The set of variants of \(t\) represents a pre-computation such that the normal form of any instance of \(t\) is equal to an instance of \(t\sigma\downarrow\) for some \(\sigma\) in the set of variants, without the need to apply further rewrite steps. A rewrite system has the finite variant property if for any sequence of terms a finite, complete set of variants exists and is effectively computable. For the sake of readability, we will often write \(\text{vars}_R(t_1, \ldots, t_n)\) instead of \(\text{vars}_R(\{t_1, \ldots, t_n\})\).

**Example 3:** Continuing our running example, let \(T_i = \{\text{proj}_i(y)\}\) for \(i \in \{1, 2\}\). Let \(\sigma_1\) be the identity substitution and \(\sigma_2 = \{y \mapsto \langle y_1, y_2\rangle\}\). The set \(\{\sigma_1, \sigma_2\}\) is a complete set of variants for \(T_1\) (resp \(T_2\)). Indeed, for any substitution \(\omega\), either we have that \(\text{proj}_i(y\omega)\downarrow\) is in normal form, and therefore \(\sigma_1\) (together with \(\tau = \omega\downarrow\)) will satisfy the requirements. Otherwise, we have that \(y\omega\downarrow\) is actually a pair \(\langle m_1, m_2\rangle\), and we conclude choosing the substitution \(\sigma_2\) (together with \(\tau = \{y_1 \mapsto m_1, y_2 \mapsto m_2\}\)).

This finite variant property is satisfied by many equational theories interesting for modelling cryptographic protocols, e.g., symmetric and asymmetric encryption, signatures, blind signatures, zero-knowledge proofs. Moreover, this property implies the existence of a complete set of unifiers, and gives us a way to compute it effectively [20].

**Definition 2:** Consider a convergent rewrite system \(R\). Let \(\Gamma = \{u_1 = v_1, \ldots, u_k = v_k\}\) be a set of equations. A set of substitutions \(\Sigma\) is called a complete set of \(R\)-unifiers for \(\Gamma\) if:

- \(u = t[\sigma]\), i.e., \(u\) is the term obtained by replacing \(t|_p\) with \(r\sigma\) in \(t\).
- \(\text{confluent: for any } t, t_1, t_2\) such that \(t \rightarrow_R t_1\) and \(t \rightarrow_R t_2\) there exists \(u\) such that \(t_1 \rightarrow_R u\) and \(t_2 \rightarrow_R u\); and
- \(\text{terminating: it does not admit any infinite sequence } t_0 \rightarrow_R t_1 \rightarrow_R t_2 \rightarrow_R \ldots\)

We denote by \(\downarrow_R\) (or simply \(\downarrow\)) the normal form of a term \(t\).

In the following we only consider equational theories \(E\) that can be represented by a rewrite system \(R\) which is convergent, i.e., we have that \(u =_E v \iff u\downarrow_R = v\downarrow_R\).

**Example 2:** Continuing Example 1, we consider the rewrite system \(R_{\text{hash}}\) given below:

- \(\text{proj}_1(\langle x, y\rangle) \rightarrow x\)
- \(\text{proj}_2(\langle x, y\rangle) \rightarrow y\)

Actually, we have that \(R_{\text{hash}}\) is convergent and it represents the equational theory \(E_{\text{hash}}\) defined in Example 1.

Given a convergent rewrite system \(R\), we now define the notion of complete set of variants, which was first introduced in [13].
1) each \( \sigma \in \Sigma \) is such that \( u_i \sigma \downarrow = v_i \sigma \downarrow \) for \( i \in \{1, \ldots, k\} \);
2) for each \( \theta \) such that \( u_i \theta \downarrow = v_i \theta \downarrow \) for any \( i \in \{1, \ldots, k\} \),
there exists \( \sigma \in \Sigma \) and a substitution \( \tau \) such that \( x \sigma \downarrow = x \sigma \tau \downarrow \) for any \( x \in \var(G) \).

We denote \( \text{csur}_{R}(\Gamma) \) such a set.

When \( R = \emptyset \), it is well-known that for any
\[
\Gamma = \{ u_1 = v_1, \ldots, u_k = v_k \}
\]
which admits a solution (i.e. there exists \( \sigma \) such that for any \( i \in \{1, \ldots, n\} \), we have that \( u_i \sigma = v_i \sigma \)), such a complete set can be chosen with cardinality one. This element is actually unique up to some renaming and we denote it \( \text{mgu}(\Gamma) \) or \( \text{mgu}(u, v) \) when \( \Gamma = \{ u = v \} \).

In the following, we consider equational theories \( E \) that can be represented by a convergent rewrite system \( R \) that has the finite variant property.

### III. Capabilities of the Attacker

At a particular point in time, after some interaction with a protocol, an attacker may know a sequence of messages \( t_1, \ldots, t_\ell \in T(\Sigma, N) \). Such a sequence is organised into a frame
\[
\varphi = \{ w_1 \mapsto t_1, \ldots, w_\ell \mapsto t_\ell \}
\]
that is a substitution with domain \( \text{dom}(\varphi) = \{ w_1, \ldots, w_\ell \} \), and size \( |\varphi| = \ell \).

**Definition 3:** Let \( \varphi \) be a frame, and \( t \in T(\Sigma, N) \).

We say that \( t \) is strongly deducible from \( \varphi \) (using \( R \)), written \( \varphi \vdash_R t \), when \( R \varphi \downarrow = t \downarrow \) for some \( R \in T(\Sigma, N_{\text{pub}} \uplus \text{dom}(\varphi)) \).

Similarly, we say that \( t \) is weakly deducible from \( \varphi \) (using \( R \)), written \( \varphi \vdash^+_R t \), when \( R \varphi \downarrow = t \downarrow \) for some \( R \in T(\Sigma, N_{\text{pub}} \uplus N_{\\text{guess}} \uplus \text{dom}(\varphi)) \).

Intuitively, an attacker is able to deduce new messages by applying function symbols in \( \Sigma \) to public names (those in \( N_{\text{pub}} \)) and terms he already knows (those in \( \varphi \)). The term \( R \) is called a recipe. Assuming that the attacker has access to names in \( N_{\text{guess}} \), this gives him more power, and allows him to (weakly) deduce more terms.

**Definition 4:** Let \( \varphi \) be a frame, and \( g \in N_{\text{guess}} \). We say that \( g \) is guessable from \( \varphi \) when there exist \( R_1, R_2 \in T(\Sigma, N_{\text{pub}} \uplus N_{\text{guess}} \uplus \text{dom}(\varphi)) \) such that:

- \( R_1 \varphi \equiv_E R_2 \varphi \), and
- \( R_1' \varphi \not\equiv_E R_2' \varphi \) where \( R_i' = R_i [g \mapsto g'] \) for some fresh name \( g' \) \((i \in \{1, 2\}) \).

The idea underlying this definition is the following. The frame \( \varphi \) represents the information known by the attacker. To see if the attacker is able to infer the value of a guessable data, we check whether the attacker can distinguish a situation in which he has access to all the correct guesses from one where we provide him with a wrong value (here \( g' \) instead of \( g \)) for the weak data that he targeted. To distinguish the situation the attacker must apply a test (here \( R_1 \) and \( R_2 \)) that holds when faced with the right value of the guess, but fails otherwise.

**Example 4:** Let \( \varphi = \{ w_1 \mapsto \langle \text{info}_{A}, \text{hash}(\langle \text{info}_{A}, r_A \rangle) \rangle \} \) with \( r_A \in N_{\text{guess}} \). The name \( r_A \) is guessable from \( \varphi \). Indeed, considering the recipes \( R_1 = \text{proj}_2(w_1) \) and \( R_2 = \text{hash}(\langle \text{proj}_1(w_1), r_A \rangle) \), we have that:

- \( R_1 \varphi = \text{E_{hash}} \text{hash}(\langle \text{info}_{A}, r_A \rangle) = \text{E_{hash}} R_2 \varphi \), whereas
- \( (R_1[r_A \mapsto r'_{A}]) \varphi \not\equiv_{\text{E_{hash}}} R_2[r_A \mapsto r'_{A}] \varphi \).

However, we may note that \( r_A \) is not guessable from
\[
\varphi' = \{ w_1 \mapsto \langle \text{info}_{A}, \text{hash}(\langle \text{info}_{A}, \langle n_A, r_A \rangle \rangle) \} \]
assuming that \( n_A \in N_{\text{pub}} \). Intuitively, without knowing \( n_A \), there is no way for the attacker to check whether he has correctly guessed the value \( r_A \).

This way of modelling guessable names allows one to take into account the fact that weak names are deducible by the attacker as soon as he has a way to verify that the guess he has performed is correct. This reflects the deduction capabilities of the attacker in a more accurate way than what happens in usual attacker models. However, this modelling comes with some limitations since it does not take into account the fact that a sufficiently large set of guessable names is as secure as a (strong) name. Indeed, in our setting, assuming that \( r_1, \ldots, r_k \) are names from \( N_{\text{guess}} \), it is possible to guess all these names from the message \( \text{hash}(\langle r_1, \langle r_2, \ldots, r_k \rangle \rangle) \), independently of \( k \).

### IV. Our Cryptographic Process Calculus

We assume that cryptographic protocols are modelled using a simple process calculus which has some similarities with the applied pi calculus [2]. Following [11], we only consider a minimalistic core calculus, additionally extended with disequality tests and guess actions.

#### A. Syntax

Let \( C \) be a set of public channel names. A protocol is modelled by a finite set of processes generated by the following grammar:

\[
P, Q ::= 0 \quad \text{null process}
\]

\[
\text{in}(c, x).P \quad \text{input}
\]

\[
\text{out}(c, t).P \quad \text{output}
\]

\[
[s = t].P \quad \text{positive test}
\]

\[
[s \neq t].P \quad \text{negative test}
\]

\[
\text{guess}(g).P \quad \text{guess}
\]

where \( x \in X, s, t \in T(\Sigma, N \uplus X) \), \( c \in C \), and \( g \in N_{\text{guess}} \).

As usual, a receive action \( \text{in}(c, x) \) acts as a binding construct for the variable \( x \). We assume the usual definitions of free and bound variables for processes. We also assume that each variable is bound at most once. A process is ground if it does not contain any free variables. For the sake of conciseness, we sometimes omit the null process at the end of a process.

We may note that our calculus does not include operators for parallel composition, and we do not consider private channels. Given that we only consider a bounded number of sessions (i.e. a process calculus without replication) and that we aim at verifying a reachability based property, parallel composition, but also private channels, can be added as syntactic sugar to
denote the set of all interleavings. To ease the presentation, we also introduce a special action $\text{guess}(g)$ in order to indicate at which step guessing on $g$ has to be done. At the price of a (potentially exponential) blow-up on the number of processes needed to model a given protocol, we may indeed assume that this kind of actions is given explicitly.

**Example 5:** As a running example, we consider the following process where $c \in C$, $\text{ack} \in \mathcal{N}_{\text{pub}}$, $r_A \in \mathcal{N}_{\text{guess}}$, and $x, y \in \mathcal{X}$:

$$P_{\text{weak}} = \text{out}(c, m_A).\text{guess}(r_A).\text{in}(c, y).$$

where $m_A = \langle \text{info}_A, \text{hash}((\text{info}_A, r_A)) \rangle$.

Intuitively, this process models one possible interleaving (among many others) of the different actions of one session of the weak version of the protocol described in Section I. In order to keep this example simple, we discard the action $\text{out}(c, \text{ack})$ (output of a public name), and we have arbitrarily chosen to perform a guess action on $r_A$ at the second step. Moreover, the purpose of the final test $\text{proj}_1(y) \neq \text{info}_A$ is to see whether $B$ can accept a message different from the expected one, i.e. $\text{info}_A$. Intuitively, an execution of this process until the end will correspond to an attack on the protocol. Of course, in order to perform a security analysis, given a protocol (as the one described in Section I), we have to explain how such interleavings are computed. This is briefly explained in Section VII-C where a richer calculus (with some convenient, additional operators) is introduced. Moreover, the way out-of-band channels are encoded is explained in Section VIII-B. However, the theoretical development explained in Sections V, VI, and VII, only rely on the syntax formally introduced in this section.

We may be interested to study a similar interleaving of actions for the strong version of the protocol (the one with $n_A \in \mathcal{N}_{\text{pub}}$). This gives us:

$$P_{\text{strong}} = \text{out}(c, m_A).\text{guess}(r_A).\text{in}(c, y).$$

For $\varphi \in \mathcal{L}$, where $t$ is either an input, an output, a test, or a guess is defined as follows:

**RECV** $\text{in}(c, x).P, \varphi \stackrel{\text{in}(c,t)}{\rightarrow} (P\{x \mapsto t\}, \varphi)$ if $\varphi \vdash t$

**SEND** $\text{out}(c, t).P, \varphi \stackrel{\text{out}(c)}{\rightarrow} (P, \varphi \cup \{w_{|\varphi|+1} \mapsto t\})$

**TEST** $\text{test}^= ([s = t], P, \varphi) \stackrel{\text{test}^=} {\rightarrow} (P, \varphi)$ if $s = t$

**TEST** $\text{test}^\neq ([s \neq t], P, \varphi) \stackrel{\text{test}^\neq} {\rightarrow} (P, \varphi)$ if $s \neq t$

**GUESS** $\text{guess}(g).P, \varphi \stackrel{\text{guess}(g)}{\rightarrow} (P, \varphi \cup \{w_{|\varphi|+1} \mapsto g\})$ if $g$ is guessable from $\varphi$

The label $\text{in}(c, t)$ represents the input of a message $t$ sent by the attacker over the channel $c$. The label $\text{out}(c)$ indicates a message sent over channel $c$, and transition rule $\text{SEND}$ records the message sent in the frame. The rule $\text{TEST}^=\text{r}$, respectively $\text{TEST}^\neq\text{r}$, checks equality, respectively difference, of $s$ and $t$ in the equational theory and is labelled by $\text{test}^=\text{r}$, respectively $\text{test}^\neq\text{r}$. The label $\text{guess}(g)$ represents the fact that the value $g$ has been guessed, and this is only possible when $g \in \mathcal{N}_{\text{guess}}$ is indeed guessable. In such a case, $g$ is added into the frame.

**Example 6:** Considering the process $P_{\text{weak}}$ given in Example 5, we have that

$$(P_{\text{weak}}, \emptyset) \stackrel{\text{out}(c)}{\rightarrow} \langle \text{proj}_1(y), \text{hash}((\text{info}_A, r_A)) \rangle \rightarrow (P_2, \{w_1 \mapsto m_A, w_2 \mapsto r_A\})$$

$$(P_3, \{w_1 \mapsto m_A, w_2 \mapsto r_A\}) \rightarrow (P_4, \{w_1 \mapsto m_A, w_2 \mapsto r_A, w_3 \mapsto r_A\})$$

where $\text{info}_A \in \mathcal{N}_{\text{pub}}$. This execution corresponds to the attack described in the introduction. The fact that $\text{guess}(r_A)$ can be fired is a direct consequence of the fact that $r_A$ is guessable from $\varphi$ (see Example 4). Note that this action would not be possible starting from $(P_{\text{strong}}, \emptyset)$. This comes from the fact that $r_A$ is not guessable from $\varphi'$ (see Example 4).

In the rest of the paper, for our technical development, we also consider an alternative semantics $\vdash_{L}$ defined as $\vdash_{L}$, except that rules $\text{TEST}^\neq$ and $\text{GUESS}$ are replaced by the two following ones:

$$(\text{guess}(g).P, \varphi) \Rightarrow (P, \varphi \cup \{w_{|\varphi|+1} \mapsto g\})$$

$$([s \neq t].P, \varphi) \Rightarrow (P, \varphi)$$

i.e., this alternative semantics does neither check whether a name is indeed guessable, nor differences of terms.

The goal of this paper is to analyse confidentiality and authentication properties of protocols that rely on weak secrets. These properties can be directly encoded as reachability properties in a process. We therefore study the Guess Reachability Problem defined as follows.

**Guess Reachability Problem**

- **Input:** A ground process $P$
- **Output:** Do there exist $l_1, \ldots, l_n$, and a frame $\varphi$ such that $(P, \emptyset) \vdash_{L} l_1, \ldots, l_n \rightarrow (\emptyset, \varphi)$?
We emphasise that the execution is defined w.r.t. $\rightarrow$ and not $\xrightarrow{l}$. A difficulty is to ensure that $g$ is guessable when a guess$(g)$ action is fired and that disequalities hold when they occur. In case such a complete execution exists, we say that the process $P$ is guess reachable.

In case, the ground process under study only contains one guess action at the end, this problem is similar to the problem of deciding whether a protocol is resistant against guessing attack. Here, the difficulty is to develop a procedure allowing us to consider several guesses. Note also that, contrary to the case of password-based protocols, guessing a data is not a goal in itself, and does not necessarily lead to an attack on a protocol. In many protocols, participants first commit to weak data and reveal it later in the execution.

V. MODELLING USING HORN CLAUSES

As the original procedure described in [11] and implemented in the AKISS tool, our decision procedure is based on a fully abstract modelling of a process in first-order Horn clauses. We give the details of this modelling in this section.

A. Predicates

We define the set of symbolic runs, denoted $u, v, w, \ldots$, as the set of finite sequences of symbolic labels:

$$u, v, w := \epsilon | l, w$$

where $l$ is of the form in$(c, t)$, out$(c)$, test$^\approx$, test$^\neq$, or guess$(g)$ with $t \in T(\Sigma, N \cup \mathcal{X})$, $c \in \mathcal{C}$, and $g \in N_{\text{guess}}$.

The empty sequence is denoted by $\epsilon$. Intuitively, a symbolic run stands for a set of possible runs of the protocol. We denote $u \subseteq v$ when $u$ is a prefix of $v$.

We assume a set $\mathcal{Y}$ of recipe variables disjoint from $\mathcal{X}$, and we use capital letters $X, Y, Z$ to range over $\mathcal{Y}$. Variables we assume that such variables may only be substituted by terms in $T(\Sigma, N_{\text{pub}} \cup N_{\text{guess}} \cup W \cup \mathcal{Y})$.

We consider five kinds of predicates over which we construct the atomic formulas of our logic. Below, $w$ denotes a symbolic run, $R, R'$ are terms in $T(\Sigma, N_{\text{pub}} \cup N_{\text{guess}} \cup W \cup \mathcal{Y})$, and $t$ is a term in $T(\Sigma, N \cup \mathcal{X})$. Informally, these predicates have the following meaning (see Figure 1 for the formal semantics).

- **reachability predicate**: $r_w$ holds when the run $w$ is executable in the alternative semantics $\rightarrow$, i.e., ignoring constraints on guesses and disequalities;
- **attacker knowledge predicate**: $k_w^<(R, t)$ holds if whenever the run $w$ is executable (w.r.t. $\rightarrow$), the message $t$ can be constructed by the attacker using the recipe $R$ that does not contain names in $N_{\text{guess}}$;
- **attacker extended knowledge predicate**: $k_w^<(R, t)$ holds if whenever the run $w$ is executable (w.r.t. $\rightarrow$), the message $t$ can be constructed by the attacker using the recipe $R$ that may contain names in $N_{\text{guess}}$;
- **identity predicate**: $i_w(R, R')$ holds if whenever the run $w$ is executable (w.r.t. $\rightarrow$), $R$ and $R'$ (that may contains names in $N_{\text{guess}}$) are recipes for the same term; and
- **reachable identity predicate**: $r_i_w(R, R')$ is a short form for the conjunction of $i_w(R_1, R'_1), \ldots, i_w(R_n, R'_n)$, and $r_w$, where $R = (R_1, \ldots, R_n)$, and $R' = (R'_1, \ldots, R'_n)$.

A (ground) atomic formula is interpreted over a pair consisting of a process $P$ and a frame $\varphi$, and we write $(P, \varphi) = f$ when the atomic formula $f$ holds for $(P, \varphi)$ or simply $P = f$ when $\varphi$ is the empty frame. We consider first-order formulas built over the above atomic formulas and the usual connectives (conjunction, disjunction, negation, implication, existential and universal quantification). The semantics is defined as expected, but the domain of quantified variables depends on their type: variables in $\mathcal{X}$ may be mapped to any term in $T(\Sigma, N)$, while recipe variables in $\mathcal{Y}$ are mapped to recipes.

**Example 7:** Let $P_{\text{weak}}$ be the process defined in Example 5, and consider the following run:

$$w_0 = \text{out}(c), \text{guess}(r_A), \text{in}(c, m_1), \text{out}(c), \text{test}^\approx, \text{test}^\neq$$

with $m_1 = (\text{info}_1, \text{hash}(\text{info}_1, r_A))$. We have $P_{\text{weak}} \models r_w$, and this is a direct consequence of the execution trace given in Example 6. We have that $P_{\text{weak}} \models i_{\text{out}}(c)(R_1, R_2)$ where $R_1 = \text{proj}_2(w_1)$ and $R_2 = \text{hash}(\text{proj}_1(w_1), r_A)$. Indeed, the only frame reachable from $P_{\text{weak}}$ through the run out$(c)$ is $\varphi$ as defined in Example 4. Let $t = \text{hash}(\text{info}_A, r_A)$, we have that $\varphi \vdash R_1 t$ (actually we have $\varphi \vdash R_1 t$).

Let $P_{\text{strong}}$ as given in Example 5, and $w_0'$ be as follows:

$$\text{out}(c), \text{guess}(r_A), \text{in}(c, m'_1), \text{out}(c), \text{test}^\approx, \text{test}^\neq$$

with $m'_1 = (\text{info}_1, \text{hash}((\text{info}_1, 1, r_A)))$. We have also that $P_{\text{strong}} \models r_w'$ since such a trace is indeed executable w.r.t. the weak semantics $\rightarrow$.

B. Seed statements

We now identify a subset of the formulas, which we call statements. Statements will take the form of Horn clauses, and we shall be mainly concerned with them.

**Definition 5:** A statement is a Horn clause of the form

$$H \iff k_w^< (X_1, t_1), \ldots, k_w^< (X_n, t_n)$$

where:

- $H \in \{r_w, k_w^< (R, t), k_w^< (R, t), i_w(R, R'), r_i_w(R, R')\}$;
- $u_0, u_1, \ldots, u_n$ are symbolic runs such that $u_i \subseteq u_0$ for any $i \in \{1, \ldots, n\}$;
- $t, t_1, \ldots, t_n \in T(\Sigma, N \cup \mathcal{X})$;
- $R, R'$ as well as elements of $\bar{R}$ (resp. $\bar{R}'$) are in $T(\Sigma, N_{\text{pub}} \cup N_{\text{guess}} \cup W \cup \mathcal{Y})$;
- and $X_1, \ldots, X_n$ are distinct variables from $\mathcal{Y}$.

Lastly, we have that $\text{vars}(t) \subseteq \text{vars}(t_1, \ldots, t_n)$ when $H = k_w^< (R, t)$ with $* \in \{-, +\}$.

In the definition above, we implicitly assume that all variables are universally quantified, i.e. all statements are ground.
By abuse of language we sometimes call $\sigma$ a grounding substitution for a statement $H \leftarrow B_1, \ldots, B_n$ when $\sigma$ is grounding for each of the atomic formulas $B_i$.

Let $P = a_1, a_2, \ldots, a_n$ be a ground process. We assume w.l.o.g. the following naming conventions:

1. If $a_i$ is a receive action then $a_i = \text{in}(c_i, x_i)$.
2. If $a_i$ is a send action then $a_i = \text{out}(c_i, t_i)$.
3. If $a_i$ is a guess action then $a_i = \text{guess}(g_i)$.
4. If $a_i$ is a positive test action then $a_i = [s_i = t_i]$.
5. If $a_i$ is a negative test action then $a_i = [s_i \neq t_i]$.

Moreover, we assume that $x_i \neq x_j$ for any $i \neq j$. For each $1 \leq i \leq n$, we define the symbolic label $\ell_i$ as follows:

$$\ell_i = \begin{cases} 
\text{in}(c_i, x_i) & \text{if } a_i = \text{in}(c_i, x_i) \\
\text{out}(c_i) & \text{if } a_i = \text{out}(c_i, t_i) \\
\text{guess}(g_i) & \text{if } a_i = \text{guess}(g_i) \\
\text{test} & \text{if } a_i = [s_i = t_i] \\
\text{test}^\neq & \text{if } a_i = [s_i \neq t_i]
\end{cases}$$

For each $0 \leq m \leq n$, let $R(m)$ (resp. $S(m)$, $G(m)$, $T^=$, and $T^\neq$) be the set of indices of the receive (resp. send, guess, positive and negative test) actions amongst $a_1, \ldots, a_m$. Moreover, we denote by $|\Gamma|$ the number of elements in such a set $\Gamma$. Formally,

- $R(m) = \{ i \mid 1 \leq i \leq m \text{ and } a_i = \text{in}(c_i, x_i) \}$
- $S(m) = \{ i \mid 1 \leq i \leq m \text{ and } a_i = \text{out}(c_i, t_i) \}$
- $G(m) = \{ i \mid 1 \leq i \leq m \text{ and } a_i = \text{guess}(g_i) \}$
- $T^=(m) = \{ i \mid 1 \leq i \leq m \text{ and } a_i = [s_i = t_i] \}$
- $T^\neq(m) = \{ i \mid 1 \leq i \leq m \text{ and } a_i = [s_i \neq t_i] \}$

Given a set of public names $\Lambda^0_{\text{pub}} \subseteq N_{\text{pub}}$, and a set of guessable names $\Lambda^0_{\text{guess}} \subseteq N_{\text{guess}}$, the set of seed statements associated to $P, \Lambda^0_{\text{pub}}$, and $\Lambda^0_{\text{guess}}$ denoted seed($P, \Lambda^0_{\text{pub}}, \Lambda^0_{\text{guess}}$), is defined to be the set of statements given in Figure 2. Intuitively, the first three kinds of seed statements allows one to express that as soon as suitable messages are deducible by the attacker to fill the receive actions occurring before step $m$, the process is indeed (weakly) executable until step $m$. Moreover, in case its $m^{th}$ action is either a send action or a guess action, the corresponding term (or guess) will become (strongly) deducible by the attacker.
If $(\mathcal{N}_{\text{pub}}, \mathcal{N}_{\text{guess}}) = (\mathcal{N}_{\text{pub}}, \mathcal{N}_{\text{guess}})$, then we write $\text{seed}(P)$ as a shortcut for $\text{seed}(P, \mathcal{N}_{\text{pub}}, \mathcal{N}_{\text{guess}})$. This set is is said to be the set of seed statements associated to $P$.

**Example 8:** Continuing our running example, among others, we have the following statements in $\text{seed}(P_{\text{weak}})$:

- $f_1 : k_{\text{out}}(w_1, m_A) \leftarrow$
- $f_2 : k_{\text{guess}}(r_A)(w_2, r_A) \leftarrow$
- $f_3 : r_w \leftarrow k_{\text{out}}(r_A)(Y, y_1, \text{hash}(\langle y_1, r_A \rangle))$
- $f_4 : k_{\text{out}}(\text{hash}(X), \text{hash}(x)) \leftarrow k_{\text{out}}(X, x)$
- $f_5 : k_{\text{out}}(\text{proj}_2(Y), \text{proj}_2(y)) \leftarrow k_{\text{out}}(Y, y)$
- $f_6 : k_{\text{out}}(\text{proj}_2(Y), y_2) \leftarrow k_{\text{out}}(Y, \langle y_1, y_2 \rangle)$

where $w = \langle \text{out}(c), \text{guess}(r_A), \text{in}(c, \langle y_1, \text{hash}(\langle y_1, r_A \rangle) \rangle)\rangle$, $\text{out}(c), \text{test}^{=}, \text{test}^{\neq}$.

Intuitively, $f_1$ represents that $m_A$ is (strongly) deducible through recipe $w_1$ once the first output has been performed, whereas $f_2$ represents that $r_A$ is strongly deducible once the two first actions $\text{out}(c)$ and $\text{guess}(r_A)$ have been performed. The statement $f_3$ says that the process $P_{\text{weak}}$ is (weakly) executable until its last action if a message of the form $\langle m, \text{hash}(\langle m, r_A \rangle) \rangle$ (for some $m$) is deducible from the information available to the attacker after executing the two first actions, i.e. knowing the first output $m_A$ and the value $r_A$.

Some statements, e.g., $f_4$, $f_5$ and $f_6$, represent the capabilities of the attacker. The two last ones ($f_5$ and $f_6$) are computed relying on variants $\text{proj}_2(Y) = \{\sigma_1, \sigma_2\}$ (see Example 3).

### C. Soundness and completeness

We shortly show that the set of seed statements is a sound and (partially) complete modelling of a process. However, we need one more definition to state this fact.

**Definition 6:** Given a set $K$ of statements, $\mathcal{H}(K)$ is the smallest set of ground facts such that:

$\sigma$ grounding for $f = (H \leftarrow B_1, \ldots, B_n) \in K$

| CONSEQ | $H \in \mathcal{H}(K)$
| EXTEND | $k_{\text{out}}^s(R,t) \in \mathcal{H}(K)$ with $s \in \{-, +\}$
| WEAKENING | $k_{\text{out}}^{\neq}(R,t) \in \mathcal{H}(K)$

We show that as far as reachability predicates and attacker knowledge predicates are concerned, the set $\text{seed}(P)$ is a complete abstraction of a process (w.r.t. $\rightarrow$).

**Theorem 1:** Let $P$ be a ground process.

- **Soundness:** $P \models f$ for any $f \in \text{seed}(P) \cup \mathcal{H}(\text{seed}(P))$;
- **Completeness:** If $(P, \emptyset) \models (Q, \varphi)$ then

  1) $\mathcal{H}(\text{seed}(P))$;

2) if $\varphi \vdash t$ with $s \in \{-, +\}$ then $k_{\text{out}}^{\neq}(R,t) \in \mathcal{H}(\text{seed}(P))$.

We will show how the completeness of $\text{seed}(P)$ can be built upon to achieve full abstraction, i.e., including also identities of the process $P$ and a procedure for checking guess reachability.

### VI. Saturation

#### A. Saturation procedure

In this section we will describe the saturation procedure. It manipulates a set of statements called a knowledge base.

**Definition 7:** Given a statement $f = (H \leftarrow B_1, \ldots, B_n)$,

- $f$ is said to be solved if for all $1 \leq i \leq n$, we have that $B_i = k_{\text{out}}^s(X_i, x_i)$ for some $s \in \{-, +\}$, $x_i \in \mathcal{X}$, and $X_i \in \mathcal{Y}$.
- $f$ is said to be well-formed if whenever it is solved and $H = k_{\text{out}}^s(R, t)$ for some $s \in \{-, +\}$, we have that $t \notin \mathcal{X}$.

A set of well-formed statements is called a knowledge base. If $K$ is a knowledge base, $K_{\text{sat}} \subseteq \mathcal{H}(K_{\text{init}})$. The set of reachable fixed points starting from the initial set $K_{\text{init}}$ is denoted $\text{sat}(K_{\text{init}})$.

**a) Generating new statements:** Given a knowledge base $K$, new statements are generated by applying the rules in Figure 3. Each rule generates a new statement $h$. Roughly, the rule $\text{RESOLUTION}$ applies the standard rule of resolution from first-order logic between an unsolved statement $f$ and a solved deduction statement $g$. We do not require $h^* = *$ but only $h^* \leq *$ which is formally defined as follows:

$h^* \leq *$ if, and only if, $\left\{\begin{array}{l}
\text{either } h^* = *; \\
\text{or } (h^*, *) = (-, +).
\end{array}\right.$

This condition reflects the intuition that any term that is strongly deducible is also weakly deducible. The rule $\text{EQUATION}$ allows us to derive new identities on recipes that may be imposed by the execution of the protocol. The rule $\text{TEST}$ allows us to derive a sequence of identities (at most one per guess action) that necessarily hold in an execution of the protocol. Once the statement $h$ is generated, we update the knowledge base $K$ with $h$ as explained below.

**b) Update:** We will now define the update operator $\mathcal{U}$ which adds statements generated by the rules of Figure 3 to the knowledge base.

**Definition 8:** Let $K$ be a knowledge base, and $f = (H \leftarrow B)$ a statement. The update of $K$ by $f$, denoted $K \mathcal{U} f$, is $K \cup \{f\}$ when $f$ is either not a solved deduction statement, or a well-formed and solved deduction statement. Otherwise, we have that $f = k_{\text{out}}^s(R,x) \leftarrow B$ with $k_{\text{out}}^s(X,x) \in B$ for some $s^*, u^*$.
Similarly, we are able to derive the solved statement and then, after several resolution steps, we are able to derive \( K = \{ H \Rightarrow B_1, \ldots, B_m \} \) by performing several resolution steps starting from \( f \), also a knowledge base.

**Lemma 1:** Given a ground process \( P \), \( N^0_{\text{pub}} \subseteq N_{\text{pub}} \), and \( N^0_{\text{guess}} \subseteq N_{\text{guess}} \), the set \( K_{\text{init}}(\text{seed}(P,N^0_{\text{pub}},N^0_{\text{guess}})) \) is a knowledge base. Moreover, given a knowledge base \( K \), any set of statements obtained through the saturation procedure is also a knowledge base.

**Example 9:** We consider the seed statements introduced in Example 8. An application of the resolution rule between \( f_0 \) and the solved statement \( f_1 \) leads to the following solved statement:

\[
h = k_{\text{out}}(\text{proj}_2(w_1), \text{hash}((\text{info}_A, r_A))) \iff \]

which is simply added to the knowledge base by the update operator. We may then decide to apply the equation rule between \( h \) and \( f_1 \) which leads to the statement:

\[
i_{\text{out}}(\text{proj}_2(w_1), \text{hash}(X)) \iff k_{\text{out}}(X, (\text{info}_A, r_A))
\]

and \( X \) (by definition of a statement). In such a case, we have that:

\[
K \cup f = K \cup \text{init}(R, X) \iff B
\]

c) **Initial knowledge base:** We finally define on which knowledge base we initiate the saturation procedure.

**Definition 9:** Given a set of statements \( S \), the initial knowledge base associated to \( S \), denoted \( K_{\text{init}}(S) \), is defined to be the empty knowledge base updated by the set \( S \), i.e.,

\[
K_{\text{init}}(S) = (((\emptyset \cup f_1) \cup \ldots) \cup f_n)
\]

where \( f_1, \ldots, f_n \) is an enumeration of the statements in \( S \).

**Lemma 1:** Given a ground process \( P \), \( N^0_{\text{pub}} \subseteq N_{\text{pub}} \), and \( N^0_{\text{guess}} \subseteq N_{\text{guess}} \), the set \( K_{\text{init}}(\text{seed}(P,N^0_{\text{pub}},N^0_{\text{guess}})) \) is a knowledge base. Moreover, given a knowledge base \( K \), any set of statements obtained through the saturation procedure is also a knowledge base.

**Example 9:** We consider the seed statements introduced in Example 8. An application of the resolution rule between \( f_0 \) and the solved statement \( f_1 \) leads to the following solved statement:

\[
h = k_{\text{out}}(\text{proj}_2(w_1), \text{hash}((\text{info}_A, r_A))) \iff \]

which is simply added to the knowledge base by the update operator. We may then decide to apply the equation rule between \( h \) and \( f_1 \) which leads to the statement:

\[
i_{\text{out}}(\text{proj}_2(w_1), \text{hash}(X)) \iff k_{\text{out}}(X, (\text{info}_A, r_A))
\]

and then, after several resolution steps, we are able to derive the solved statement:

\[
i_{\text{out}}(\text{proj}_2(w_1), \text{hash}((\text{info}_A, r_A))) \iff k_{\text{out}}(X, (\text{info}_A, r_A))
\]

Similarly, we are able to derive \( r_w \iff k_{\text{out}}(\text{guess}(r_A), Y_1, y_1) \) by performing several resolution steps starting from \( f_3 \). Therefore, we may apply the **Test** rule to derive a final:

\[
r_w(\text{proj}_2(w_1), \text{hash}((\text{info}_A, r_A))) \iff k_{\text{out}}(\text{guess}(r_A), Y_1, y_1)
\]

Intuitively, this statement means that the symbolic run \( w \) is indeed executable and leads to a test \( \text{proj}_2(w_1) = \text{hash}((\text{info}_A, r_A)) \) that holds in the resulting frame (provided the attacker knows a public value that will be used to instantiate \( y_1 \)).

**B. Soundness and completeness**

We extend \( \mathcal{H}(K) \) to establish that any \( K_{\text{sat}} \in \text{sat}(K_{\text{init}}(P)) \) is a complete abstraction of \( P \).

**Definition 10:** Let \( K \) be a set of statements. We define \( \mathcal{H}_e(K) \) to be the smallest set of ground facts containing \( \mathcal{H}(K) \) and that is closed under the rules of Figure 4.

We have that the set of solved statements produced by the saturation procedure is a sound and complete abstraction of the ground process \( P \) (w.r.t. \( \Rightarrow \)). More formally, we have the following result.

**Theorem 2:** Let \( K \in \text{sat}(K_{\text{init}}(P)) \) for some ground process \( P \). We have that:

- **Soundness:** \( P \models f \) for any \( f \in K \cup \mathcal{H}_e(K) \);
- **Completeness:** If \( (P, \emptyset) \xrightarrow{i_1, \ldots, i_n} (Q, \varphi) \) then
  1. \( r_{i_1, \ldots, i_n} \in \mathcal{H}_e(K_{\text{sat}}) \);
  2. if \( \varphi \vdash_t \downarrow r_{i_1, \ldots, i_n} \) then \( k_{i_1, \ldots, i_n}^+(R, t, \downarrow) \in \mathcal{H}_e(K_{\text{sat}}) \);
  3. if \( \varphi \vdash_t \downarrow r_{i_1, \ldots, i_n} \) then \( i_{1, \ldots, i_n}^+(R, t, \downarrow) \in \mathcal{H}_e(K_{\text{sat}}) \).

**VII. Our procedure**

In this section, we first discuss the effectiveness and termination of the saturation procedure. Then, we describe our algorithm to solve the Guess Reachability Problem as stated at the end of Section IV. We finally discuss the integration of our algorithm in the verification tool AKISS.

**A. Effectiveness of the saturation procedure**

In this section, we will discuss some issues on how to effectively compute a saturated knowledge base and also discuss termination of the saturation procedure.
A first issue comes from the fact that $K_{\text{init}}(P)$ is infinite. Indeed, the set $\text{seed}(P)$ for a ground process $P$ is infinite because $N_{\text{pub}}$ and $N_{\text{guess}}$ contain an infinite set of names. We follow [11] to overcome this difficulty. Intuitively, we do only need to consider names occurring in $P$ for its saturation; clauses representing names not occurring in $P$ do not influence the saturation of the remaining clauses. Formally, we have the following result.

**Lemma 2:** Let $P$ be a ground process, $N_{\text{pub}}^P \subseteq N_{\text{pub}}$ (resp. $N_{\text{guess}}^P \subseteq N_{\text{guess}}$) be the finite set of public names (resp. guessable names) occurring in $P$, and $K^0 \in \text{sat}(K_{\text{init}}(P,N_{\text{pub}}^P,N_{\text{guess}}^P))$. We have that $K \subseteq K^0 \cup S$ for some $K \in \text{sat}(K_{\text{init}}(P))$ and a set $S$ containing statements of the form:

- $k_i^-(n,n) \iff \forall n \in N_{\text{pub}}$;
- $k_i^+(g,g) \iff \forall g \in N_{\text{guess}}$;
- $i_i(n,n) \iff \forall n \in N_{\text{pub}} \cup N_{\text{guess}}$;
- $r_i((R_1, \ldots, R_k), (R'_1, \ldots, R'_k)) \iff B$ where $(R_{i_0}, R'_{i_0}) = (n,n)$ for some $1 \leq i_0 \leq k$, and $n \in N_{\text{pub}} \cup N_{\text{guess}}$.

The saturation procedure may itself not terminate even if the initial knowledge base is finite. As shown in [11] in a simple setting, we conjecture that our saturation procedure terminates for the class of subterm convergent equational theories. More importantly, we have implemented our procedure and tested it on several examples considering various equational theories that are not subterm convergent, and our tool did terminate on all practical examples we have tested.

**B. Description and correctness of the algorithm**

Our procedure is described in Figure 5. Let $P$ be a ground process and $K^0 \in \text{sat}(K_{\text{init}}(P,N_{\text{pub}}^P,N_{\text{guess}}^P))$ where $N_{\text{pub}}^P$ (resp. $N_{\text{guess}}^P$) is the set containing all public names (resp. guessable names) occurring in $P$. In our procedure, $P$ is represented by the set $K_{\text{solved}}^0$ of solved statements of $K^0$.

The test $\text{GUESS-REACHABILITY}(P)$ checks whether there exists a reachable identity $r_{i_1, \ldots, i_n}((R, R')) \iff B$ in $K_{\text{solved}}^0$ that corresponds to a real execution after replacement of the remaining variables with fresh constants. For guess actions, we simply check whether the recipes $R_1 \omega$ and $R'_2 \omega$ are witnesses of the fact that the $j$th guess action contains indeed a guessable name. If all these tests succeed then $P$ is indeed guess reachable. Otherwise, in case no such a reachable identity exists, then process $P$ is not guess reachable.

**Theorem 3:** Let $P$ be a ground process, $N_{\text{pub}}^P \subseteq N_{\text{pub}}$ (resp. $N_{\text{guess}}^P \subseteq N_{\text{guess}}$) be the finite set of public names (resp. guessable names) occurring in $P$. Let $K^0 \in \text{sat}(K_{\text{init}}(P,N_{\text{pub}}^P,N_{\text{guess}}^P))$.

We have that $P$ is guess reachable if, and only if $\text{GUESS-REACHABILITY}(P, K_{\text{solved}}^0)$ holds.

The soundness part of this theorem is an easy consequence of our algorithm. Completeness is more involved. It is based on Theorem 2 but some works remains to be done to ensure that disequality tests will be satisfied. More importantly, we have to ensure that enough reachable identity predicates will be generated and tested. This is needed to ensure that when a weak name is guessable, then our algorithm will find a test witnessing this fact. Intuitively, our procedure will only generate small tests and our completeness results shows that this is indeed sufficient when we are looking for an attack. Note that, we have to consider all the tests (one per guess) together, and this may require to perform instantiations that will not be needed when considering only one guess. Intuitively, this comes from the fact that the attacker may need to output messages that are quite complex to ensure that all the guessable data will become guessable when needed (on the same trace).

**C. Integration in AKISS**

We have implemented all of the before described theory in the AKISS tool. To ease the specification of the protocols, the AKISS tool supports additional operators in the process calculus for parallel composition ($P || Q$), sequential composition ($P : Q$) [11], and we also allow the use of (statically defined) private channels. All these constructions are syntactic sugar and can be translated to sets of (linear) processes eventually at the cost of an exponential blow-up.

While the original version of AKISS only allowed the verification of equivalence properties we added support for reachability properties (specified by the means of events). Similarly we allow to define authentication properties as (non-injective) correspondence properties. These correspondence properties are specified by annotating the protocols with "begin" and "end" events and are internally translated into simple reachability properties. Note that such a translation requires the use of disequality tests.

The tool also automatically annotates the protocol with the necessary guess actions, i.e., the protocol specifier is
Guess-Reachability($P$, $K^0_{\text{solved}}$)

Let $n$ be the number of actions occurring in $P$.

For all $r_1, \ldots, r_m (\mathbb{R}, \mathbb{R'}) \in K^0_{\text{solved}}$

Let $c_1, \ldots, c_k$ be fresh names in $N_{\text{pub}}$ such that $\sigma : \text{vars}(\ell_1, \ldots, \ell_n) \cup \{x_1, \ldots, x_m\} \rightarrow \{c_1, \ldots, c_k\}$ is a bijection.

Let $\omega = \{X_i \mapsto x_i \sigma \mid 1 \leq i \leq m\}$.

Check whether $(P, 0) = (P_0, \varphi_0) \xrightarrow{\ell_1 \sigma} \ldots (P_{n-1}, \varphi_{n-1}) \xrightarrow{\ell_n \sigma} (P_n, \varphi_n) = (\emptyset, \varphi)$ using the recipes $R_j \omega$ and $R'_j \omega$ as a witness for guessing the guessable name in the $j$th guess action. If this execution is valid, return \textbf{Yes}.

Return \textbf{No}.

Fig. 5. Test for checking guess reachability

not required to include these actions. Intuitively, these guess actions have to be inserted at all possible steps. To limit the number of interleavings we implement a few straightforward optimisations: guess actions are only inserted before inputs when an output occurred since the last guess of the given weak name; when generating the interleavings we always favour outputs as they increase the attacker’s knowledge which is a sound optimisation when verifying reachability properties.

VIII. Case studies

In this section we demonstrate that our tool can be effectively used to analyse protocols based on weak secrets. Weak secrets may either be low-entropy secrets (the guessable names introduced in Section II) or the result of applying a function that produces a short output, e.g., truncating the output of a hash function. The ISO/IEC 9798-6:2010 standard [23] suggests that “short” corresponds to 16-20 bits. We will first describe how we can model such weak hash functions. Weak secrets are often used in protocols relying on a (low capacity) out-of-band channel. We therefore also explain how out-of-band channels can be encoded in our formalism.

A. Weak functions

Sometimes short verification codes need to be transmitted. One way for generating these codes is to use weak hash functions whose output is short enough to be, e.g., copied or compared by a human. However, when the output of the function is short, an attacker may efficiently find collisions. As AKISS supports user specified equational theories, we encode this property in the equational theory. While, strong cryptographic hash functions are modelled as free symbols, the property that collisions may be found by the attacker on a keyed hash function is reflected using equations:

\[
\text{sh}(k', \text{bf}(k', \text{sh}(k, m))) = \text{sh}(k, m)
\]

\[
\text{sh}(\text{bf}(m', \text{sh}(k, m)), m') = \text{sh}(k, m)
\]

The first equation models that given a short hash $\text{sh}(k, m)$ and an arbitrary key $k'$, the attacker may find by brute force a message $m' = \text{bf}(k', \text{sh}(k, m))$ such that $\text{sh}(k, m) = \text{sh}(k', m')$. The second equation allows to fix the message $m'$ and find a key $k' = \text{bf}(m', \text{sh}(k, m))$ such that $\text{sh}(m, k) = \text{sh}(m', k')$. A similar modelling was also used by Chothia et al. in [12] for verifying commitment protocols in ProVerif.

The ISO/IEC 9798-6:2010 standard [23] distinguishes two types of short hash functions, which they call check value functions and message digests. While both generate short hashes, they differ on the keys they take as input: check value functions take as second argument a weak key and message digests a strong key. Our model does reflect these two types of hash functions, simply depending on whether the second argument is a weak name or not. The requested security property for short hash functions is that an attacker may not find two messages $m, m'$ such that $\text{sh}(m, k) = \text{sh}(m', k)$ for a non-negligible fraction of the key space. This property is indeed reflected by our modelling as the equation $\text{sh}(m, x) = \text{sh}(m', x)$ does not hold for arbitrary keys, unless $m = m'$.

B. Out-of-band channels

When out-of-band channels are used in protocols we may consider different kinds of channels, depending on whether the communication is supposed to be confidential, authentic, and synchronous. To model these different kind of channels, we allow the use of private channels, similar to the applied pi calculus. However, we only allow static channel names, i.e., channel names are not transmitted dynamically during the protocol execution. In that case, private channels are merely syntactic sugar for generating the appropriate traces.

We consider the following types of channels:

- Public channels: they offer no guarantee and correspond to normal output actions in our calculus.
- Private channels: communications are confidential, authentic and synchronous; this corresponds to a usual communication using a private channel.
- Asynchronous, private channels: such communications offer confidentiality and authentication, but may be delayed by the attacker; asynchrony is encoded as usual by performing the output in parallel, i.e., if $c_{\text{ap}}$ is an asynchronous private channel, we replace $\text{out}(c_{\text{ap}}, u).P$ with $P \mid \text{out}(c_{\text{ap}}, u).0$ where $c_{\text{ap}}$ is a private channel.
- Asynchronous, authentic, but non-confidential (or simply authentic) channels can be encoded similarly; if $c_{\text{aa}}$ is an asynchronous authentic channel, we replace $\text{out}(c_{\text{aa}}, u).P$ with $\text{out}(c_{\text{aa}}, u).P \mid \text{out}(c_{\text{ap}}, u).0$ where $c_{\text{pub}}$ is a public channel and $c_{\text{priv}}$ is a private channel.

Which channel is the most appropriate depends on the context and the nature of the out-of-band channel.
C. Protocols with manual data transfer

The ISO/IEC 9798-6:2010 standard [23] presents several protocols for transferring authenticated data among two devices. The mechanisms differ in the assumptions on the input/output capabilities of the devices, the used of hash functions and the actions a user has to perform.

Devices may have a standard or simple input or output interface. While standard interfaces allow to enter/display alphanumeric strings, a simple interface generally only consists in a button or a light to confirm or affirm success. Hash functions may be short hash functions, cryptographic hash functions or MACs. Humans may need to either compare or copy short strings. Some protocols also require the human to generate a short random string.

To illustrate our work, we will detail below mechanisms 2 and 4 of [23].

**a) Manual authentication mechanism 2:** The protocol assumes two devices that have both a standard output interface and a simple input interface. The protocol uses a check-value function, i.e., a short hash function which uses a weak key. The protocol assumes that both devices are in proximity of a human who is able to compare short strings displayed on both device’s output interface.

The protocol is described in Figure 6. Initially, data $d_A$ and $d_B$ are provided to devices $A$, respectively $B$. The goal of the protocol is to check whether the devices agree on the data. We assume that the data can be chosen freely by the attacker. Once they have received their data the devices signal to the user that they are ready. When both devices are ready, the user informs device $A$ that it may start, e.g., by pressing the button of the simple input interface. Device $A$ generates a fresh, weak key $k$ and transmits it (on an insecure channel) to device $B$. Both devices display the key $k$ and the short hash of their data with $k$ on their output interface. The user compares the output of both devices and enters either accept or reject depending on whether the displayed codes coincide or not.

The protocol is easily encoded in the process calculus serving as the input language of AKISS. The process modelling the authentication mechanism is entirely displayed in Figure 7 where $k$ is a weak name, and channels $ah_1$, $ah_2$, $ha_1$, $ha_2$, $bh_1$, $bh_2$, and $hb$ are private channels used for encoding authentic channels as explained in Section VIII-B. We briefly explain below the process $B$ given in Figure 7. Initially, $d_B$ is provided to device $B$. Therefore, $B$ initially waits for an input, and answers by outputting the constant ready on an authentic (but non-confidential) channel. To model these channel properties we rely on the encoding explained in Section VIII-B. Then, device $B$ is waiting for another message on a classical (i.e. insecure) channel. $B$ answers to this message by sending the key it just received as well as the short hash it has computed. Both are sent on another authentic channel (encoded in a similar way than before). Once this is done, it simply waits for an answer from the user (this is modelled through a communication on the private channel $hb$).

In case, the received value is the public constant ok (which indicated that the user has accepted this transaction) device $B$ will end this session normally (modelled through the event end). If the event end is reached authentication should hold. The two other processes $A$ and $H$ are modelled in a rather similar way. The operator :: that occurs in process $H$ is used to model the two synchronisations that appear from the user’s point of view. Indeed, the user is supposed to receive the ready message coming from both devices before he proceeds. Similarly, the user is supposed to wait for the short strings of both devices and compare them before he accepts/rejects.

We verify the correspondence property that states that whenever Device $B$ finishes the protocol with data $d$ then Device $A$ must have started the protocol with the same data before, by annotating the processes with begin and end events.

![Figure 6. Manual authentication mechanism 2 from [23]](image-url)

<table>
<thead>
<tr>
<th>scenario</th>
<th># traces</th>
<th>time</th>
<th>status</th>
</tr>
</thead>
<tbody>
<tr>
<td>one session</td>
<td>Auth. A</td>
<td>368</td>
<td>&lt; 8s</td>
</tr>
<tr>
<td></td>
<td>Auth. B</td>
<td>168</td>
<td>&lt; 4s</td>
</tr>
<tr>
<td>without start</td>
<td>Auth. A</td>
<td>360/592</td>
<td>&lt; 8s</td>
</tr>
<tr>
<td></td>
<td>Auth. B</td>
<td>337/736</td>
<td>&lt; 8s</td>
</tr>
<tr>
<td>without start + h strong</td>
<td>Auth. A</td>
<td>592</td>
<td>&lt; 4s</td>
</tr>
<tr>
<td></td>
<td>Auth. B</td>
<td>736</td>
<td>&lt; 5s</td>
</tr>
<tr>
<td>two sessions</td>
<td>Auth. A</td>
<td>48852</td>
<td>59m</td>
</tr>
<tr>
<td></td>
<td>Auth. B</td>
<td>10794</td>
<td>8m30s</td>
</tr>
</tbody>
</table>

We can also show that if the synchronisation message ‘start’ is removed we find an attack as expected: indeed the attacker may learn $k$ and forge a value $d' = bh(k', sh(k, m))$ to be sent to $A$ in the begin of the protocol. Note that the attacker can not chose an arbitrary value for $d'$ as both $k$ and the short hash are compared by the human. We have also verified that, although this scenario is unrealistic, this attack is not possible.
\[ A = \text{in}(c, \varnothing). (\text{out}(ah_1, \text{ready}) \mid \text{out}(c, \text{ready}). \text{in}(ha_1, x_s). [x_s = \text{start}]. \text{out}(c, k). (\begin{array}{l} \text{begin}(x_d). \text{out}(ah_2, (k, \text{sh}(k, x_d))). \mid \text{out}(c, (k, \text{sh}(k, x_d))). \text{in}(ha_2, x). [x = \text{ok}] \end{array})) \]

\[ B = \text{in}(c, y_d). (\text{out}(bh_1, \text{ready}) \mid \text{out}(c, \text{ready}). \text{in}(c, yk). (\text{out}(bh_2, \langle y_k, \text{sh}(yk, y_d) \rangle) \mid \text{out}(c, \langle y_k, \text{sh}(yk, y_d) \rangle)). \text{in}(hb, y). [y = \text{ok}]. \text{end}(y_d)) \]

\[ H = \text{in}(ah_1, z_d). [z_d = \text{ready}] . \mid \text{in}(bh_1, z_b). [z_b = \text{ready}] \mid (\text{out}(ha_1, \text{start}) \mid \text{out}(c, \text{start}). (\text{in}(ah_2, z'_d) \mid \text{in}(bh_2, z'_b) ) : [z'_d = z'_b]. (\text{out}(ha_2, \text{ok}) \mid \text{out}(c, \text{ok}) \mid \text{out}(hb, \text{ok}) \mid \text{out}(c, \text{ok}))) \]

Fig. 7. Manual authentication mechanism 2 from [23] (one session Auth. B)

if we use a strong hash function instead of the short hash. Finally we also checked that the protocol is still secure in the case of two sequential sessions. Given that the protocol uses a human as an intermediate we think that parallel sessions are not a meaningful scenario.

Most of those verifications have been performed in a few seconds on a standard laptop. Actually, only the last experiments (the ones reported on the last line of each table) have been done on a 20 core Intel(R) Xeon(R) CPU E5-2687W v3 @ 3.10GHz. We also indicate the number of traces (due to all the possible interleaving of the actions) that have been considered by the tool in order to conclude. Note that in case an attack has been found, the tool stops the exploration before considering all these traces, and this is also indicated in the table.

The fact that the performances degrade when considering an additional session is not surprising and is clearly correlated with the number of traces that have to be considered to carry out the analysis. This is the usual and well-known state-space explosion problem inherent of the concurrent nature of the systems we analyse. As briefly explained in Section VII-C, we have implemented a few straightforward optimisations to mitigate this issue, but more needs to be done in this direction.

b) Manual authentication mechanism 4: This mechanism is described in Figure 8. It assumes a simple input interface and standard output interface for device A and a simple input and standard output interface for device B.

![Fig. 8. Manual authentication mechanism 4 from [23]](image_url)

Table: Scenario comparison

<table>
<thead>
<tr>
<th>Scenario</th>
<th># Traces</th>
<th>Time</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>one session Auth. A</td>
<td>35</td>
<td>&lt; 1s</td>
<td>ok</td>
</tr>
<tr>
<td>one session Auth. B</td>
<td>35</td>
<td>&lt; 1s</td>
<td>ok</td>
</tr>
<tr>
<td>without start Auth. A</td>
<td>16/70</td>
<td>&lt; 1</td>
<td>attack</td>
</tr>
<tr>
<td>without start Auth. B</td>
<td>21/125</td>
<td>&lt; 1</td>
<td>attack</td>
</tr>
<tr>
<td>k weak Auth. A</td>
<td>51/175</td>
<td>&lt; 1s</td>
<td>attack</td>
</tr>
<tr>
<td>k weak Auth. B</td>
<td>51/175</td>
<td>&lt; 1s</td>
<td>attack</td>
</tr>
<tr>
<td>two sessions Auth. A</td>
<td>38290</td>
<td>2m</td>
<td>ok</td>
</tr>
<tr>
<td>two sessions Auth. B</td>
<td>43876</td>
<td>2m</td>
<td>ok</td>
</tr>
</tbody>
</table>

The protocol relies on a standard hash function and supposes a user who can copy a short random string from device A to device B. As in the previous protocol we assume that data \( d_A \) and \( d_B \) are given to the devices and that the aim is to ensure agreement on these data. Device A generates a strong nonce \( k \), a weak nonce \( r \) and sends hash(\( d_A, k, r \)) to B over an insecure network. Both devices then notify the user that they are ready and the user informs A that it may start. A displays \( r \) to the user who copies \( r \) into B. Moreover, A sends \( k \) to B over the insecure network. Device B recomputes the hash and compares it to the previously received value. If both values are identical it outputs accept and reject otherwise. The user enters the displayed result into device A.

We may note that, despite the high number of traces, answers are relatively fast. We consider here an empty equational theory, and therefore the saturation process is faster than for mechanism 2.

IX. CONCLUSION

In this paper we presented a novel symbolic model for reasoning about protocols that use short authentication strings, possibly generated as the result of a weak hash function. We propose a new procedure for verifying protocols in this model and have implemented the procedure in the AKISS tool. The new capabilities of the tool are illustrated by analysing...
several protocols from the ISO/IEC 9798-6:2010 standard for authentication mechanisms for manual data transfer.

As future work we plan to perform additional case studies, as well as checking injective correspondance properties. Another possible generalisation is to adapt our framework to verify equivalence properties. The resulting framework might be used for analysing vote anonymity in protocols relying on short vote codes, such as in Pretty Good Democracy [27] or Du-Vote [22].

REFERENCES