Abstract. We clear up confusion surrounding privacy claims about the ICAO 9303 standard for e-passports. The ICAO 9303 standard includes a Basic Access Control (BAC) protocol that should protect the user from being traced from one session to another. While it is well known that there are attacks on BAC, allowing an attacker to link multiple uses of the same passport, due to differences in implementation; there still remains confusion about whether there is an attack on unlinkability directly on the BAC protocol as specified in the ICAO 9303 standard. This paper clarifies the nature of the debate, and sources of potential confusion. We demonstrate that the original privacy claims made are flawed, by uncovering attacks on a strong formulation of unlinkability. We explain why the use of the bisimilarity equivalence technique is essential for uncovering our attacks. We also clarify what assumptions lead to proofs of formulations of unlinkability using weaker notions of equivalence. Furthermore, we propose a fix for BAC within the scope of the standard, and prove that it is correct, again using a state-of-the-art approach to bisimilarity.

1 Introduction

The Basic Access Control (BAC) mechanism for e-passports, which forms part of the ICAO 9303 standard [1], has been in operation since 2005. Since then, an improved access control mechanism, the Password Authenticated Connection Establishment (PACE) protocol [6], has been standardised in order to address known limitations with the security of BAC. However, the BAC protocol is still being implemented by a growing number of e-documents, not only e-passports. For example, many national identity cards are compliant with the BAC protocol in the ICAO 9303 standard. This means that, firstly, even a relatively minor attack on privacy is of concern to a large number of citizens internationally; and, secondly, the ICAO 9303 standard is being used in a wider range of contexts that do not necessarily have system security comparable to an airport, facilitating more sophisticated attacks.

For the above reasons, it is imperative that we clarify the existence of and nature of attacks on the privacy of BAC explained in this paper. The notion of privacy we are concerned with is a strong form of unlinkability, meaning that an e-passport that satisfies such a privacy property cannot be linked from one session to another, by a third party snooping in on wireless communications. Such a privacy issue is of concern to users carrying e-passports, who do not wish third parties to track their movements.

Unlinkability can be formulated in the following terms: an attacker cannot observe any difference between a scenario where each session with an e-passport reader is with
a new e-passport and a scenario where the same e-passport may be involved in more than one session. Strong unlinkability assumes, in addition, that the attacker has the power to make some decisions, such as feeding a challenge into a remote reader rather than a reader in the vicinity of the e-passport. We will explain that it is critical that the additional power given to the attacker by strong unlinkability is modelled by using bisimilarity as the notion of equivalence.

To understand why strong unlinkability, expressed in terms of bisimilarity, is important, we must clarify the story in the literature up until this paper. The first paper [4] formally analysing unlinkability of e-passports, using symbolic techniques, formulated weak unlinkability as a property of traces, and strong unlinkability as an equivalence problem in terms of bisimilarity. That paper mainly concerns an attack particular to the implementation of the French e-passport, exploiting distinguishable error messages from which the attacker can infer whether authentication was partially successful.

The problem with the above mentioned paper [4] is that they also make claims about e-passports implementing the ICAO 9303 standard with a single error message for all types of authentication failure, such as the UK e-passport. They make the claim that the UK e-passport satisfies the strong form of unlinkability, expressed using bisimilarity. The primary contribution we make is to clarify that their claim is false. Taking exactly the same conditions — the way they define strong unlinkability and how they model the UK e-passport — we discover a counter proof for their claims, and provide a witness in terms of a modal logic formula describing an attack on strong unlinkability.

We survey related work [19,9,11,10,12], contributing to the story behind symbolically analysing the unlinkability of BAC. With the exception of the original paper, the papers surveyed concern alternative definitions of unlinkability expressed in terms of trace equivalence rather than bisimilarity. This survey of trace-based approaches we use to emphasise the impact of using bisimilarity rather than trace equivalence when verifying unlinkability of protocols such as BAC. We also highlight other parameters impacting whether a model proves unlinkability or discovers an attack.

A secondary contribution is to propose a fix for the BAC protocol, within the scope of the ICAO 9303 standard [1]. We again showcase bisimilarity as a technique for analysing privacy properties, providing a proof that strong unlinkability holds by defining a bisimulation that is witness to our claims. Finally, we discuss implications of our analysis, for example, how our attack on strong unlinkability applies to a wide range of protocols, not limited only to PACE and a minimal example of an RFID protocol used as an illustrative example. We also touch on practical implications of our attack, which are distinct from existing practical attacks on unlinkability [14,15].

Summary. In Section 2 we investigate and refine the analysis of the BAC protocol for e-passports implemented similarly to the UK e-passport, reporting on different models and results, and identifying the fundamental modelling problems surrounding unlinkability. In Section 3 we introduce the strong unlinkability problem for a simplified authentication protocol that we will use as an example throughout the paper, we also note a fix for the protocol (encrypting the error message). Section 4 recalls background material on a state-of-the-art presentation of bisimilarity facilitating our analysis. In Section 5 we show how bisimilarity can be used to discover attacks on strong unlinkability. Finally, in Section 6 we return to the original formulation of the UK version of the
BAC protocol, demonstrating how our attack lifts to an attack on strong unlinkability, invalidating the original claim [4].

2 An investigation into unlinkability claims about BAC

In this section, we briefly survey, clarify, and expand upon the body of work symbolically analysing the Basic Access Control (BAC) protocol. The purpose of BAC is to mutually authenticate an e-passport and reader (e.g., at passport control in an airport), and establish a short-term key used in proceeding communication (e.g., transmitting personal information about the owner).

The BAC protocol is sketched informally in Fig. 1 where dashed lines (→) indicate a message transmitted via an OCR session on a page of the e-passport, and solid lines are wireless communications between a chip and reader. The reader first sends a constant message get_challenge requesting a challenge — a nonce nt sent by the e-passport — which is used during the mutual authentication of the e-passport and reader. The standard specifies that an “operating system dependent error” [1] should be sent when authentication fails. Such a failure occurs when the e-passport receives an authentication request from the reader, and either the message authentication code (MAC) is wrong, or a nonce in the message does not match the challenge nt previously sent by the e-passport.

2.1 The key paper defining strong unlinkability, but with a flawed claim

The primary contribution of this paper is to clarify that the first paper symbolically analysing the BAC protocol, as implemented by countries such as the UK (Fig. 1), contained a flawed claim. Arapinis et al. [4] define weak unlinkability as a property of traces, faithful to the ISO standard for unlinkability [2]. They then argue for a stronger property, called strong unlinkability, expressed using bisimilarity. Their work is accompanied with a trace that correctly demonstrates that the French BAC protocol violates both their definitions of unlinkability. Regarding the UK BAC protocol, they say:

Checking the bisimulation by hand, we find that SystemUK ≈ l SystemUK’ holds: A repeating tag in the SystemUK process is matched by a new tag in the idealised SystemUK’ version of the system.

Unfortunately, their statement above is false. In their work, SystemUK is a system specification in which the same e-passport can be used many times, and SystemUK’ is an
idealised specification in which each e-passport is used only once. The above statement

SystemUK ≈ SystemUK’ claims the system specification and idealised specification are
indistinguishable to an attacker, expressed in terms of labelled bisimilarity [3]. Later, in
Sections 5 and 6, we will demonstrate that there is a witness invalidating the bisimilarity
claim above, and therefore there is an attack on strong unlinkability.

Although Arapinis et al. claim, in the quote above, to have proven strong unlin-
kability by hand, no proof exists. Confusion was partly down to an old bug[1] in ProVerif.

2.2 Alternative models of unlinkability based on trace equivalence

There exist several examples of tool-supported analysis of weak unlinkability of the
BAC protocol, using trace equivalence. The PhD thesis of Cheval [9] and work on
disunification [11] is the basis of this line of work. The tool APTE [10] is the first
such tool able to directly verify finite trace equivalence properties of protocols with if-
then-else branches. To demonstrate the APTE tool, a survey is performed on a range of
protocols. However, the verification of the UK BAC protocol does not terminate after 2
days, and the authors mark it “safe?”.

The DEEPSEC [12] prover is a state-of-the-art tool for analysing finite trace equiv-
alence for security protocols. Depending on how the UK BAC protocol is modelled it
can, for two sessions, both find an attack [12], and claim that no attack exists. In our
GitHub repository[2] we provide details on both modelling scenarios. In summary, an
attack is discovered if the fixed scenario with two different e-passports is compared to
a specification where all e-passports differ. In contrast, DEEPSEC discovers no attack
whenever we consider that, in reality, for two sessions, we either have two identical
e-passports or two different e-passports. Indeed the attack discovered using a fixed con-
figuration is considered not to be practical (no trace can be executed to confirm the
presence of the same e-passport twice, and the attack is longer than necessary).

| Paper          | Equivalence Type | Model Scope | Observable | Claim made |
|----------------|------------------|-------------|------------|------------|
| Arapinis et al. [4] | Bisim.           | Unbounded   | Arbitrary  | Yes        | No         | flawed    |
| APTE [10]      | Trace            | 2 Sessions  | Fixed      | No         | ?          | N/A       |
| DEEPSEC [12]   | Trace            | 2 Sessions  | Fixed      | No         | Yes        | OK        |
| DEEPSEC (ours) | Trace            | 2 Sessions  | Arbitrary  | Yes        | No         | OK        |
| Hirschi et al. [19] | Trace          | Unbounded   | Arbitrary  | No         | No         | OK        |

Fig. 2. Comparison table of various analyses of the UK e-passport. Note all the above assume the
number of internal communications is unobservable.

A summary of the above findings is presented in Fig.2. We highlight only the most
important differences between these models, mentioned previously, namely: bisimilarity
vs. trace equivalence; and, unbounded vs. arbitrary bounded vs. fixed bounded. An-

1 This information on an old bug in ProVerif is due to Stéphanie Delaune and Vincent Cheval.
2 https://github.com/ZDSmith/bac-protocol-unlinkability
other critical modelling parameter is the choice of observables, notably the constant \texttt{get\_challenge} message in Fig. 1. This impacts whether strong unlinkability holds, by allowing an attacker to count the number of reader sessions based on the number of observed \texttt{get\_challenge} messages. This parameter does not affect weak unlinkability.

| Note on terminology: We use the term strong unlinkability in exactly the sense it was originally communicated in CSF’10 [4]. A source of potential confusion is that a paper communicated in S&P’16 [19] presents a proof of what they claim to be strong unlinkability. That claim may be misleading, since they, in fact, significantly change the definition of strong unlinkability. The most important change they make is to use trace equivalence rather than bisimilarity. If we have a proof, with trace equivalence replacing bisimilarity in the definition of strong unlinkability, then weak unlinkability follows as a corollary (this fact follows by adapting Theorem 2 in the original paper [4], since the proof of Theorem 2 does not rely on finer properties of bisimilarity). Note also that they [19] change slightly, but significantly, the observables in the model of BAC. Their forthcoming journal version [20] acknowledges and discusses this terminology mismatch. Sometimes changing definitions of terms is of little consequence; for example, differences between secrecy as a trace property and secrecy expressed in terms of bisimilarity are insignificant [15]. However, the thesis of our paper is that the same does not apply to privacy. Trace equivalence gives the attacker less power to resolve choices, and hence misses attacks, such as on the unlinkability of BAC. Related work also highlights the power of bisimilarity for discovering attacks in the context of the anonymity of the MUTE file sharing system [13], and in discussions comparing strong unlinkability, weak unlinkability and computational unlinkability games [8]. |

3 Minimal variant of the BAC authentication protocol

The analysis of the full e-passport protocol involves some large messages, which can obscure the essential problems with the protocol. Therefore, initially, we make two simplifications to the analysis for pedagogical and methodological reasons:

1. We present a minimal mutual authentication protocol that features the same problems with strong unlinkability as the BAC protocol for e-passports.
2. We show our attack can be discovered systematically by using a slightly finer notion of bisimilarity better suited to symbolic analysis.

Both of the above initial simplifications to our analysis are lifted later, in Section 6. Our use of a minimal authentication protocol also highlights, as mentioned in the introduction, that the problems with strong unlinkability in this work affect a wider class of authentication protocols, where the same key is used in different sessions.

3.1 An illustrative minimal protocol for mutual authentication

We now describe our cut-down mutual authentication protocol in Fig. 4, sufficient to explain problems with the full BAC protocol. Our protocol is similar to the Feldhoffer
protocol [18], which was proposed as a minimal mutual authentication protocol for RFID tags. A difference, compared to the Feldhofer protocol, is that we include an error message which is used by the RFID tag to signal a failed authentication session to the reader. For minimality, we also simplify the response of the tag (the Feldhofer protocol responds with \(\langle n, m\rangle_k\) rather than simply \(m\)).

Like the ICAO 9303 standard BAC protocol for e-passports, our minimal protocol achieves a strong authentication property called synchronisation [16], which is easily checked using automated tools such as Scyther [16]. The key differences, compared to BAC, is that BAC also establishes a shared session key, and uses message authentication codes to improve message integrity.

We make use of the applied \(\pi\)-calculus for modelling processes. The syntax of processes is presented in Fig. 3 along with a message theory featuring pairs and symmetric encryption (encryption using a shared secret key).

### 3.2 Modelling our minimal authentication protocol in the applied \(\pi\)-calculus

In the applied \(\pi\)-calculus, an honest reader in our minimal example in Fig. 4 can be modelled as follows.

\[
\text{Reader} \triangleq c(k).\nu m.a(x).\bar{\nu}(\langle m, x\rangle_k)
\]

Channel \(c\) is a private channel used to read a secret key (for e-passports, calculated using data read from a page using OCR). The reader receives a challenge \(x\), generates a fresh name \(m\) (the counter-challenge) and transmits the nonce and challenge encrypted together using the session key, \(\langle m, x\rangle_k\).

The tag is modelled in the applied \(\pi\)-calculus as follows.

\[
\text{Linkable} \triangleq \bar{c}(k).\nu m.\bar{a}(n).a(y).\text{if } \text{snd}(\text{dec}(y, k)) = n \text{ then } \bar{a}(\text{fst}(\text{dec}(y, k))) \text{ else } \bar{a}(\text{error})
\]

Fig. 3. A syntax for applied \(\pi\)-calculus processes with a message theory.

Fig. 4. A linkable authentication protocol.
The private channel $c$ is used to transmit a private key unique to the tag (for e-passports modelling the act of presenting a page to an OCR reader). The tag generates and sends a fresh challenge $n$. The response to the challenge $y$ is received. If the response contains the challenge, tested by $\text{snd}(\text{dec}(y, k)) = n$, then the counter-challenge $\text{fst}(\text{dec}(y, k))$ is sent. Otherwise, an error is sent. The error message signals to the reader that authentication has failed, resulting in the protocol not successfully completing.

Combining the above reader and tag, we can describe the system as follows.

$$\text{Linkable System} \triangleq \nu c.(!\text{Reader} | !\nu k.\text{Linkable})$$

Notice that channel $c$, used for sending and receiving the key of the tag, is bound, hence private. This suggests that an attacker does not have the power to intercept messages on this channel (modelling a session with an OCR reader). However, other communications take place on a public channel $a$ which an attacker can snoop over, e.g., reading using an antenna in the vicinity, and writing using a fake tag.

In the above system specification, the replicated reader, written $!\text{Reader}$, indicates that any number of sessions of the reader can be initiated in parallel. The sub-process $!\nu k.\text{Linkable}$ indicates that any number of tags can be created in parallel, each with a unique key $k$ identifying them; and, furthermore, each tag can enter any number of sessions using the same identity $k$, in parallel.

Unlinkability properties can be expressed using the above system specification and the idealised specification below:

$$\text{Linkable Spec} \triangleq \nu c.(!\text{Reader} | !\nu k.\text{Linkable})$$

Notice the only difference between $\text{Linkable System}$ and $\text{Linkable Spec}$ is the absence of replication after the generation of the key. Thus, in $\text{Linkable Spec}$, each new session is with a new tag, with a freshly generated key.

We formulate strong unlinkability as an equivalence problem by setting out to showing that $\text{Linkable System}$ and $\text{Linkable Spec}$ are equivalent from the perspective of an attacker. In principle, the idea is that if an attacker cannot tell the difference between a scenario where the same tag is allowed to be used in multiple sessions and the scenario where each tag is really used once, then you cannot link two uses of the same tag.

The important point in this paper is that strong unlinkability in fact fails. Indeed for our minimal authentication example we can prove the following inequality, where $\sim$ is a suitable notion of bisimilarity.

$$\text{Linkable System} \not\sim \text{Linkable Spec}$$

The use of bisimilarity grants the attacker more power than trace equivalence, essentially allowing the attacker to resolve certain choices (in this case, to which reader the challenge is sent). We will explain such attacks in the remaining sections of this paper.

### 3.3 Fixing protocols to achieve strong unlinkability

Beyond finding new attacks and shorter attacks, bisimilarity can also be used to provide proofs of strong unlinkability when they exist. For many calculi, it is established that
Bisimilarity is asymptotically more efficient to check than trace equivalence, particularly in the limits\(^\text{[23]}\) (for infinitely many sessions). Indeed, with expertise, finding a bisimulation in the limit is relatively easy here.

Consider the variant of our running example given in Fig. 5. Notice that the error message is encrypted, along with a nonce to ensure that the ciphertext is different on each execution. Note, we assume that a fresh ciphertext \{\(r, \text{error}\)\}_k and a nonce \(m\) are indistinguishable. This prevents an attacker intercepting communications from distinguishing between a correct response and an error message.

To verify the fixed minimal authentication protocol, we consider the following specification, in which the else branch has been modified:

\[
\text{Unlinkable} \triangleq \nu c. (\text{!} \text{Reader} | \text{!} vk. \text{Unlinkable})
\]

The fixed system and specification are thereby stated as follows.

\[
\begin{align*}
\text{Unlinkable System} & \triangleq \nu c. (\text{!} \text{Reader} | \text{!} vk. \text{Unlinkable}) \\
\text{Unlinkable Spec} & \triangleq \nu c. (\text{!} \text{Reader} | \text{!} vk. \text{Unlinkable})
\end{align*}
\]

Indeed, we can prove \(\text{Unlinkable System} \sim \text{Unlinkable Spec}\) holds, where \(\sim\) is a suitable notion of bisimilarity. This establishes that strong unlinkability holds for our fixed basic authentication protocol. The same fix can be applied to BAC, which is a fix within the scope of the ICAO 9303 standard\(^\text{[1]}\), since the standard does not exclude encrypting the error message in the BAC protocol.

### 4 Background on bisimilarity for the applied \(\pi\)-calculus.

We briefly recall a concise formulation of (strong) early bisimilarity for the applied \(\pi\)-calculus. Our presentation makes use of extended processes (in normal form), and a pure labelled transition system which simplifies the analysis of bisimilarity. Note the presentation we adopt here makes it relatively easy to quickly discover our attack.

Extended processes in normal form \(\nu x. (\sigma | P)\) are subject to the restriction that the variables in \(\text{dom}(\sigma)\) are fresh for \(x\), \(\text{fv}(P)\) and \(\text{fv}(\nu y. \sigma)\), for all variables \(y\) (i.e., \(\sigma\) is idempotent, and substitutions are fully applied to \(P)\). We follow the convention that operational rules are defined directly on extended processes in normal forms. Note adopting normal forms removes the need for several additional conditions that must be imposed in other formulations of bisimilarity for the applied \(\pi\)-calculus\(^\text{[3]}\).

We require a standard notion of static equivalence, which checks two processes are indistinguishable in terms of the messages output so far.
Definition 1 (static equivalence). Extended processes in normal form \(\text{vx.}(\sigma \mid P)\) and \(\text{vy.}(\theta \mid Q)\) are statically equivalent whenever, for all pairs of messages \(M\) and \(N\) such that \((\text{fv}(M) \cup \text{fv}(N)) \cap (x \cup y) = \emptyset\), we have \(M\sigma =_E N\sigma\) if and only if \(M\theta =_E N\theta\).

We require the following definitions for composing extended processes in parallel and with substitutions, defined whenever \(z \notin \text{fv}(B) \cup \text{fv}(\rho)\) and \(\text{dom}(\sigma) \cap \text{dom}(\theta) = \emptyset\).

\[
\sigma \mid \theta \mid Q \triangleq \sigma \cdot \theta \mid Q \\
\rho \triangleright z.A \triangleq \nu z.(\rho \mid A) \\
B \triangleright z.A \triangleq \nu z.(B \mid A) \\
\nu z.A \triangleright B \triangleq \nu z.(A \mid B)
\]

The above definitions are employed in our definition of (early) labelled transitions (Fig. 6), which are defined directly on extended processes in normal form. Labels on transitions are either: \(\tau\) — an internal communication; \(\overline{M}(z)\) — an output on channel \(M\) binding the output message to variable \(z\); or \(MN\) — an input on channel \(M\) receiving message \(N\). (Notice if-then-else makes no additional \(\tau\)-transitions in this presentation.)

Fig. 6. An early labelled transition system, plus symmetric rules for parallel composition and choice. The equational theory over message terms can be applied at any point. The set of free variables and \(\alpha\)-conversion are as standard, where \(\text{vx.}P\) and \(M(x).P\) bind \(x\) in \(P\). Define the bound names such that \(\text{bn}(\pi) = \{x\}\) only if \(\pi = \overline{M}(x)\) and \(\text{bn}(\pi) = \emptyset\) otherwise. Define the names such that \(n(MN) = \text{fv}(M) \cup \text{fv}(N)\), \(n(M(x)) = \text{fv}(M) \cup \{x\}\) and \(n(\tau) = \emptyset\).

The early labelled transition system and static equivalence together can be used to define the following (strong) version of early bisimilarity.

Definition 2 (early bisimilarity). A symmetric relation between extended processes \(R\) is an early bisimulation only if, whenever \(A \triangleright R B\) the following hold:

- \(A\) and \(B\) are statically equivalent.
- If \(A \triangleright A'\) there exists \(B'\) such that \(B \triangleright B'\) and \(A' \triangleright R B'\).
Processes \( P \) and \( Q \) are early bisimilar, written \( P \sim Q \), whenever there exists an early bisimulation \( R \) such that \( P \mathcal{R} Q \).

Notice initially we consider here a strong notion of bisimilarity, where the number of internal communications can be counted. This initially simplifies the analysis. To be precise, the strong semantics preserves a notion called image finiteness, which is lost in the weak setting and imposes additional technical challenges. However, later we show attacks discovered lift to the weak setting (by including more observables).

## 5 Finding attacks on privacy using bisimilarity

In order to refer to intermediate states, we can break down the sub-states of \( \text{Reader} \) and \( \text{Linkable} \), from Sec. 3 as follows.

\[
\begin{align*}
W_i & \triangleq a(x),vm.a\{m,x\}_{k_b} & U(n,y)_i & \triangleq \text{if } \text{snd}(\text{dec}(y,k_i)) = n \\
\text{Linkable} & \triangleq \tau(k_i),vn.T(n)_i & \text{then } \text{a}(\text{fst}(\text{dec}(y,k_i))) \\
T(n)_i & \triangleq \text{a}(n).a(y).U(n,y)_i & \text{else } \text{a}(\text{error})
\end{align*}
\]

The real system, which allows multiple instances of the same tag, can perform the following two \( \tau \) actions followed by an output action \( \text{a}(u) \). The idealised specification on the right below follows with the same actions as best it can. Note we abbreviate multiple transitions by writing sequences of actions on the label.

\[
\text{Linkable}_\text{System} \xrightarrow{\tau \bar{\pi}(u)} \text{Broken}_\text{System}' \quad \text{Linkable}_\text{Spec} \xrightarrow{\tau \bar{\pi}(u)} \text{Broken}_\text{Spec}'
\]

The states reached above are of the following form.

\[
\begin{align*}
\text{Broken}_\text{System}' & \triangleq vC, k_1, n_1, n_2, (\{n_i/\nu\} | W_1 | W_1 | \text{!!Reader} | a(y).U(n_1,y)_1 | T(n_2)_1 | \text{!!Linkable}_1 | \text{!!k!!.Linkable}) \\
\text{Broken}_\text{Spec}' & \triangleq vC, k_1, k_2, n_1, n_2, (\{n_i/\nu\} | W_1 | W_2 | \text{!!Reader} | a(y).U(n_1,y)_1 | T(n_2)_2 | \text{!!k!!.Linkable})
\end{align*}
\]

At this point, we can swap the system for the specification (exploiting the symmetry of a bisimulation), and \( \text{Broken}_\text{Spec}' \) performs the sequence of actions below.

\[
\text{Broken}_\text{Spec}', \text{a}.u.\bar{\pi}(v) \xrightarrow{a.v.\bar{\pi}(v)} vC, k_1, k_2, n_1, n_2, m.(\{n_i,m,m_1,m_2,\text{error}/\text{drop},w\} | W_1 | 0 | \text{!!Reader} | 0 | T(n_2)_2 | \text{!!k!!.Linkable})
\]

If the system and specification were equivalent (which they are not), then the system should be able to perform the same actions to reach a state where the system appears to be identical to the idealised specification, from the perspective of the attacker. The longest \( \text{Broken}_\text{System}' \) can keep up this bisimulation game is as follows.

\[
\text{Broken}_\text{System}', \text{a}.u.\bar{\pi}(v) \xrightarrow{a.v.\bar{\pi}(v)} vC, k_1, n_1, n_2, m.(\{n_i,m,m_1,m_2,\text{error}/\text{drop},w\} | W_1 | 0 | \text{!!Reader} | 0 | T(n_2)_1 | \text{!!Linkable}_1 | \text{!!k!!.Linkable})
\]

The important step above is the first input action \( a.u \). This transition affects subprocess \( W_2 \) in \( \text{Broken}_\text{Spec}' \), which evolves to \( \bar{\text{a}}(\{m,n_1\}_{k_2}) \), i.e., a reader ready to respond to challenge \( n_1 \) by using key \( k_2 \). In contrast, \( \text{Broken}_\text{System}' \) can only reach a
state with sub-process $\overline{a}(\{m, n_{1}\}_{k_{1}})$ which is ready to respond to the same challenge $n_{1}$ but using key $k_{1}$ (note there are two equivalent ways the system can act at this point, since two readers with key $k_{1}$ are active — both options lead to the same outcome).

Both the system and specification then proceed with the actions $\overline{a}(v)$, $a\nu$, then $\overline{a}(w)$, corresponding to intercepting the response of the reader $v$, relaying $v$ to the tag, and obtaining the output $w$ of the tag.

After the four transition steps described above, the real system satisfies the equation $w \neq \text{error}$. In contrast, for the idealised specification we have that $w = \text{error}$ holds. Performing this test represents an attacker intercepting the third output on channel $a$, named $w$ above, and checking whether or not it is an error message. If the system does not produce an error following this strategy, then unlinkability is violated. This way, we can link the two sessions, since we have proof that they must involve the same tag.

Notice test $w = \text{error}$ confirms static equivalence is violated.

### 6 Lifting our attack to the setting of labelled bisimilarity for the ICAO 9303 standard BAC protocol

Previously, in Sections 3 and 5 we emphasised that we discussed a slightly simpler protocol than BAC which exhibits the same problems with strong unlinkability. We also used a stronger notion of bisimilarity, allowing the attack to be discovered more easily. These decisions were made in order to present details of the attack more clearly.

We simplify the presentation of our attack by making the following methodological point. When we discover an attack under stronger assumptions, we can lift the attack to a setting with weaker assumptions, and then check the attack is still valid. In this section, we follow exactly this methodology — we describe how the attack lifts to the setting of BAC under a weak notion of bisimilarity, exactly as assumed in the original paper symbolically analysing BAC [4] (which, recall, made the opposite claims without providing proofs).

In order to conduct our analysis, we require a constructor representing message authentication codes. We extend the message language with function $\text{mac}(M, N)$, with no new equations in the message theory. For readability, we employ the abbreviation $\text{let } x = M \text{ in } P \doteq P[\{M/x\}]$ in the following specifications of the UK e-passport and (generic) e-passport reader.

```plaintext
MainUK \doteq \overline{c}(ke, km).d(x).\{x = \text{get\_challenge}\}mnt.z(nt).d(y).
   if snd(y) = \text{mac}(fst(y), km) then
      if nt = fst(snd(dec(fst(y), ke))) then
         vkt.let m = \{(nt, \{fst(dec(fst(y), ke)), k\})\}_{k}, in
         \overline{c}(m, \text{mac}(m, km))
      else \overline{c}(error)
   else \overline{c}(error)

Reader \doteq c_{k}(x_{k}).\overline{c}(\text{get\_challenge}).d(nt).vnrvkr.
   let m = \{(nr, \{nt, kr\})\}_{\text{fst}(x_{k})} \text{ in } \overline{c}(m, \text{mac}(m, \text{snd}(x_{k})))
```


Similarly to our minimal authentication example, we can express the system and idealised specification, respectively, as follows.

\[
\text{SystemUK} \equiv \nu c. (\!k! \text{Reader} | \!k! \nu . \text{MainUK})
\]

\[
\text{SystemUK}' \equiv \nu c. (\!k! \nu . \text{MainUK})
\]

We also employ labelled bisimilarity [3] which makes use of weak transitions, A \xrightarrow{\pi} B which allow zero or more \tau transitions to occur before and after the transition \pi, or zero transitions if \pi = \tau. Notice B \xrightarrow{\pi} B' is the only difference compared to Def. 2.

**Definition 3 (labelled bisimilarity).** A symmetric relation between extended processes \(R\) is a labelled bisimulation only if, whenever \(A \xrightarrow{\pi} B\) the following hold:

- A and B are statically equivalent.
- If \(A \xrightarrow{\pi} A'\) there exists \(B'\) such that \(B \xrightarrow{\pi} B'\) and \(A' \xrightarrow{\pi} B'\).

Labelled bisimilarity \(\approx_{l}\) is the greatest labelled bisimulation.

Now, by following a similar strategy described in the previous section, we can prove that strong unlinkability fails, expressed as follows.

\[
\text{SystemUK} \not\approx_{l} \text{SystemUK}'
\]

A little more work is required, compared to the previous section, since we must count the number of get\_challenge messages sent and received rather than number of \tau transitions. However, we can go through essentially the same symbolic reasoning to discover a similar attack to the previous section. Rather than repeating the same analysis but on a larger specification, we instead present a shorter way to describe such attacks and informally describe how it can be exploited in a practical fashion.

### 6.1 Describing the attack using a modal logic formula

We can concisely describe attacks on privacy using modal logic formulae. Attacks on labelled bisimilarity can be described using the modal logic classical \(\mathcal{FM}(\mathcal{F})\) is for free inputs, \(M\) is for match [24]). A syntax for classical \(\mathcal{FM}\) is presented below.

\[
\begin{align*}
\phi & ::= M = N \quad \text{equality} \quad M \not\equiv N \equiv \neg(M = N) \\
| \phi \land \phi & \quad \text{conjunction} \quad [\pi]\phi \equiv \neg(\pi) \neg \phi \\
| \langle \pi \rangle \phi & \quad \text{diamond} \quad [\lnot \pi]\phi \equiv \neg(\pi) \neg \phi \\
| \neg \phi & \quad \text{negation} \quad \phi \land \psi \equiv \neg(\neg \phi \land \neg \psi)
\end{align*}
\]

The semantics of classical \(\mathcal{FM}\) is presented below.

\[
\begin{align*}
\forall x.(\sigma | P) & \models M = N \iff M\sigma =_E N\sigma \quad \text{and} \quad x \cap (\text{fv}(M) \cup \text{fv}(N)) = \emptyset \\
A & \models \langle \pi \rangle \phi \quad \text{iff} \text{ there exists } B \text{ such that } A \xrightarrow{\pi} B \text{ and } B \models \phi. \\
A & \models \phi_1 \land \phi_2 \quad \text{iff} A \models \phi_1 \text{ and } A \models \phi_2. \\
A & \models \neg \phi \quad \text{iff} A \models \phi \text{ does not hold.}
\end{align*}
\]

Using classical \(\mathcal{FM}\), we can define a witness that two processes are not labelled bisimilar, as expressed using this soundness and completeness theorem. Note this is a more standard classical version of a theorem in related work [21].
Theorem 1 (soundness and completeness). \( P \approx_l Q \), whenever, for all formula \( \phi \), we have \( P \models \phi \) if and only if \( Q \models \phi \).

From the contrapositive of the above theorem, whenever \( P \not\approx_l Q \), there exists a formula \( \phi \) such that \( P \models \phi \) holds, but \( Q \not\models \phi \).

In the case of the failure of strong unlinkability of the UK BAC protocol, we have the following classical \( \mathcal{FM} \) formula, say \( \psi \).

\[
\langle d \text{ get challenge} \rangle \langle \overline{c}(x) \rangle \langle \overline{c}(y) \rangle \langle \overline{c}(z) \rangle \\
\quad x = \text{get challenge} \wedge y = \text{get challenge} \wedge z \neq \text{get challenge} \wedge \\
\quad [d z] (\langle \overline{c}(u) \rangle \langle d u \rangle \langle \overline{c}(v) \rangle (u \neq \text{get challenge} \wedge v \neq \text{get challenge} \wedge v \neq \text{error}) \\
\quad \vee [\overline{c}(w)] (w = \text{get challenge}) )
\]

For this formula we can verify \( \text{SystemUK} \models \psi \) holds. Clearly, interpreting such a witness for non-bisimilarity requires considerable expertise. The first part of the formula, until input \([d z]\), starts an e-passport session and two reader sessions, and then sends the challenge, named \( z \) in the formula, from the e-passport. The later branches of the formula check whether or not the reader sessions are with the same e-passport or not. The critical step is \([d z]\), which ranges over all ways in which the challenge \( z \) can be fed back into the system as an input. In the bisimulation game, this corresponds to a swapping of perspective, where the idealised specification leads, rather than the system (as illustrated in the attack on the minimal authentication protocol in Sec. 5). In practical terms, this means that the attacker takes control over where the input \( d z \) is performed.

Now consider \( \text{SystemUK}' \). We show that \( \text{SystemUK}' \not\models \psi \). Notice that the branch \([\overline{c}(w)] (w = \text{get challenge}) \) covers the possibility that the input is fed in when a \text{get challenge} message is expected, leaving no possible output actions other than those starting a fresh session. Notice also the possibility of an error occurring too early (\( u = \text{error} \)) is also accommodated. Importantly, regardless of how \( \text{SystemUK}' \) plays the first four actions, in the state reached, there exists an input \( d u \) which fails to match any of the eventualities described by the formula.

Note there are many such distinguishing formulae, each describing subtly different attacks on strong unlinkability. We select this one, as it formally justifies the practical description of the attack in the next section.

6.2 Practical steps to implement a discovered attack

Here, we give an example of a practical attack that might be carried out in the real world, based on the attack on strong unlinkability given in the previous section. We assume the presence of a Dolev-Yao [17] adversary, who can block or redirect messages. Importantly, we assume that the adversary cannot interfere with the credentials on the e-passport, for example by snooping on an OCR session.

The aim of our attack will be to identify the e-passport who has most recently interacted with a specific reader device (which need not be under adversary control). For example, in an airport, the attacker may wish to identify people who have travelled through the “priority” lane, as they are more likely to be airline staff or other people of interest. The attack proceeds as follows:
Fig. 7. Attack on UK e-passport: implementation involving fake reader and fake e-passport, informally. The critical moment is choosing where to feed $nt$. Assume $Msg = \langle nr, (nt, kr) \rangle_{ke}$, $Msg' = \langle nt, (nr, kt) \rangle_{ke}$, $R = \langle Msg, \text{mac}(Msg, km) \rangle$ and $C = \langle Msg', \text{mac}(Msg', km) \rangle$.

1. An honest agent has their OCR details read by the targeted reader device.
2. The adversary blocks any RF communication between the (now-scanned) e-passport and the reader. The agent presumes that the machine is faulty and moves on.
3. The adversary brings a custom reader device close to an agent. This custom reader initiates the BAC protocol with the agent’s e-passport.
   - The fake reader does not make use of, or attempt to read, any OCR data. It acts as if this phase has already been completed.
4. The fake reader relays messages from the e-passport to the reader suspended in (2), for example by using a RF retransmitter located close to the reader.
   - The suspended reader still has OCR data stored from the earlier step.
5. If the e-passport that the adversary is communicating with is indeed the e-passport that was scanned by the reader (as is depicted in Figure 7), then the protocol will complete successfully, and the adversary will see an encrypted data packet.
   - If the e-passport does not match the previously scanned one, the adversary will see a constant error message.

The adversary never learns the keys of any e-passport in this case, but they do not need to - they need only distinguish whether or not the final message is a constant term or an encrypted packet.

In Fig. 7 we highlight three key timing constraints on this attack. The hard constraint, labelled $t_2$, is the maximum time a genuine e-pasport reader waits between issuing a request and receiving a response from an e-passport. We conducted experiments on open source e-pasport readers and found that $t_2$ is bounded above by approximately 1.1 seconds. To perform this experiment we implemented a fake e-passport, to interact with an open source e-pasport reader.

https://github.com/tananaev/passport-reader
The constraint $t_3$ represents how long an e-passport is willing to wait before receiving the next command after sending a challenge. It has no technical upper bound, as a tag remains active (and awaiting commands) for as long as it is powered. The flow of messages in Fig. 7 shows it is possible to arrange $t_3$, such that it is bounded above by a few seconds. Therefore, if the e-passport itself implements a timeout (which typically they do not) it would be easy to stay within that timeout bound.

A key practical concern in step (2) is the duration for which an e-passport reader will hold on to OCR details, indicated as $t_1$ in Fig. 7. This is dependent on the specific firmware implementation of the reader (the OCR reader and RF session with BAC combined). Certainly for open source readers for smart phones, this is not an obstacle. To avoid this attack, airport e-passport readers should require that an upper bound is enforced on $t_1$. It is also unknown if a reader discards stored OCR data after it believes it has finished executing the BAC protocol. This should be enforced, to ensure that the attack cannot be repeated (i.e. we can attempt to link only one passport with the last OCR scan).

An important point is that, if we interpret Fig. 7 simply as a trace of inputs and outputs then it is not an attack. To see why, observe that even if the suspended reader has different keys as expected in the idealised specification, then another (currently unused) reader can be employed to produce the same sequence of actions. The use of bisimilarity is essential.

7 Conclusions

Our primary contribution is to clear up confusion regarding the unlinkability of the BAC, when implemented with a single plaintext error message, as in the UK e-passport. We clarify that, contrary to claims previously made [4], there is a real attack on strong unlinkability. The attack can be discovered quickly from the strategy that causes the search for bisimulation to fail. While the attack is not as easy as the known attack on the French e-passport, the attack is practical, as clarified in Sec. 6.2.

Our secondary contribution is to clarify how different modelling assumptions may lead to different conclusions about the unlinkability of BAC. Our conclusion from this survey is that in a model faithful to the problem, there is no attack on the unlinkability of BAC that can be described as a simple trace. Our attacks on strong unlinkability make non-trivial use of bisimilarity; in practical terms this corresponds to the attacker making a decision about which reader receives a challenge. Thus we have:

- An attack, Sec. 6 correcting the original claim about strong unlinkability expressed in terms of bisimilarity [4].
- No attack expressible as a trace, a claim supported by our DEEPSEC code in Sec. 2 and by adapting [19].

Note that in both cases, we make the assumption that the initial configuration of the system is not fixed, as discussed in Sec. 2. Also, in both cases, internal communications, modelled by $\tau$-transitions, are assumed to be unobservable and a get challenge message is observable.
We also make significant methodological contributions. We discovered this attack by employing a state-of-the-art approach to bisimilarity checking, never before applied to a problem of this complexity. Our attack was discovered systematically, by the following methodology.

1. We search for a proof using a finer notion of bisimilarity called open bisimilarity \([25,7,26,21]\), which lazily explores the state space.
2. When a distinguishing strategy is discovered using open bisimilarity, we determine whether it is an attack by constructing a distinguishing formula in an intermediate modal logic called intuitionistic \(\mathcal{IFM}\) \([21,22]\).
3. Given our formula, we check whether the formula is still distinguishing under classical assumptions. This confirms there is also an attack on early bisimilarity.
4. We check the attack is also valid in the setting of labelled bisimilarity \([3]\) (for which \(\tau\)-transitions are silent), by checking where a lack of image finiteness allows additional processes to be created that may have an impact on the analysis.

While the above methodology discovers and confirms our attack systematically, undoubtedly employing the above methodology required mastery of state-of-the-art work on bisimilarity. Thus future work includes improving tool support.

Further perspectives on BAC and unlinkability. We note that the impact of our work extends beyond the BAC protocol. Attacks on strong unlinkability we discover can be adapted to a wide range of authentication protocols. We propose a general form for an authentication protocol that may fail strong unlinkability.

- The same keys are used between the e-document and multiple readers.
- A failed authentication session behaves observably differently from a successful authentication session.

Note observable differences between successful and failed sessions may be due to an error message, as in the French and UK implementations of the BAC protocol; but may also be due to the presence or absence of a valid message expected during authentication. Therefore our attack adapts also to a variant of BAC that signals a failed authentication session without any error message.

The latter point may be trickier to mitigate in practice. It may be possible to observe the presence or absence of a message exchanged after authentication is complete. In the ICAO 9303 standard, this phase is called the secure messaging phase. Such a practical extension of our attack is a concern perpendicular to the study of the BAC protocol in this work.

Another modelling dimension is the question of whether attacks such as those highlighted in this paper are down to inadequate definitions of unlinkability. A way to avoid our attacks by modifying the definition of unlinkability is to sequentialise entire sessions, such that exactly one reader starts and one passport starts, and both must have used up all their actions before proceeding with any action in a new session. This essentially models the situation where a round trip between an e-passport and remote reader becomes infeasible (e.g., due to stricter timeouts). The current work however focuses on clarifying established definitions of unlinkability.
Acknowledgements. We thank the following people for their time and knowledge during the investigation of these results: Vincent Cheval, Ugo Chirico, Stéphanie Delaune, Lucca Hirschi, and Steve Kremer.

References

1. Machine readable travel documents, part 11: Security mechanisms for MRTDs. Tech. Rep. Doc 9303. Seventh Edition, International Civil Aviation Organization (ICAO) (2015), https://www.icao.int/publications/Documents/9303_p11_cons_en.pdf
2. ISO 15408-2: Common criteria for information technology security evaluation. part 2: Security functional requirements. Tech. Rep. CCMB-2017-04-002, ISO/IEC standard (2017), https://www.commoncriteriaportal.org/files/ccfiles/CCPART2V3.1R5.pdf
3. Abadi, M., Blanchet, B., Fournet, C.: The applied pi calculus: Mobile values, new names, and secure communication. J. ACM 65(1), 1:1–1:41 (2017)
4. Arapinis, M., Chothia, T., Ritter, E., Ryan, M.: Analysing unlinkability and anonymity using the applied pi calculus. In: Computer Security Foundations Symposium (CSF), 2010 23rd IEEE, pp. 107–121. IEEE (2010)
5. Avoine, G., Beaujeant, A., Hernandez-Castro, J., Demay, L., Teuwen, P.: A survey of security and privacy issues in epassport protocols. ACM Comput. Surv. 48(3), 47:1–47:37 (2016)
6. Bender, J., Fischlin, M., Kügler, D.: Security analysis of the PACE key-agreement protocol. In: Samarati, P., Yung, M., Martinelli, F., Ardagna, C.A. (eds.) Information Security. pp. 33–48. Springer (2009)
7. Briais, S., Nestmann, U.: Open bisimulation, revisited. Theoretical Computer Science 386(3), 236–271 (2007)
8. Brusío, M., Chatzikokolakis, K., Etalle, S., den Hartog, J.: Linking unlinkability. In: Palamidessi, C., Ryan, M.D. (eds.) Trustworthy Global Computing. pp. 129–144. Springer Berlin Heidelberg, Berlin, Heidelberg (2013)
9. Cheval, V.: Automatic verification of cryptographic protocols: privacy-type properties. PhD thesis, Laboratoire Spécification et Vérification, ENS Cachan (2012)
10. Cheval, V.: APTE: an algorithm for proving trace equivalence. In: International Conference on Tools and Algorithms for the Construction and Analysis of Systems. pp. 387–392. Springer (2014)
11. Cheval, V., Comon-Lundh, H., Delaune, S.: A procedure for deciding symbolic equivalence between sets of constraint systems. Information and Computation 255(2 Part 1), 94 – 125 (2017)
12. Cheval, V., Kremer, S., Rakotonirina, I.: DEEPSEC: Deciding equivalence properties in security protocols theory and practice. In: 2018 IEEE Symposium on Security and Privacy (S&P), pp. 529–546 (2018)
13. Chothia, T.: Analysing the mute anonymous file-sharing system using the pi-calculus. In: Najm, E., Pradat-Peyre, J.F., Donzéau-Gouge, V.V. (eds.) Formal Techniques for Networked and Distributed Systems - FORTE 2006. pp. 115–130. Springer Berlin Heidelberg, Berlin, Heidelberg (2006)
14. Chothia, T., Smirnov, V.: A traceability attack against e-passports. In: Sion, R. (ed.) Financial Cryptography and Data Security, 14th International Conference, FC 2010, Tenerife, Canary Islands, Spain, January 25-28, 2010, Revised Selected Papers. Lecture Notes in Computer Science, vol. 6052, pp. 20–34. Springer (2010)
15. Cortier, V., Rusinowitch, M., Zalisnescu, E.: Relating two standard notions of secrecy. Logical Methods in Computer Science 3(3) (2007)
16. Cremers, C.: The Scyther tool: Verification, falsification, and analysis of security protocols. In: International Conference on Computer Aided Verification. pp. 414–418. Springer (2008)
17. Dolev, D., Yao, A.: On the security of public-key protocols. IEEE Transactions on Information Theory 2(29) (1983)
18. Feldhofer, M., Dominikus, S., Wolkstorfer, J.: Strong authentication for RFID systems using the AES algorithm. In: Joye, M., Quisquater, J.J. (eds.) Cryptographic Hardware and Embedded Systems - CHES 2004. Lecture Notes in Computer Science, vol. 3156, pp. 357–370. Springer (2004)
19. Hirschi, L., Baelde, D., Delaune, S.: A method for verifying privacy-type properties: the unbounded case. In: Security and Privacy (S&P), 2016 IEEE Symposium on. pp. 564–581. IEEE (2016)
20. Hirschi, L., Baelde, D., Delaune, S.: A method for unbounded verification of privacy-type properties. Journal of Computer Security 27(3), 277–342 (2019)
21. Horne, R.: A bisimilarity congruence for the applied π-calculus sufficiently coarse to verify privacy properties (arXiv:1811.02536) (2018), https://arxiv.org/abs/1811.02536
22. Horne, R., Ahn, K.Y., Lin, S.W., Tiu, A.: Quasi-open bisimilarity with mismatch is intuitionistic. In: Dawar, A., Grädel, E. (eds.) In Proceedings of 33rd Annual ACM/IEEE Symposium on Logic in Computer Science, Oxford, United Kingdom, July 9-12, 2018. pp. 26–35 (2018)
23. Kanellakis, P.C., Smolka, S.A.: CCS expressions, finite state processes, and three problems of equivalence. Inf. Comput. 86(1), 43–68 (1990)
24. Milner, R., Parrow, J., Walker, D.: Modal logics for mobile processes. Theoretical Computer Science 114(1), 149–171 (1993)
25. Sangiorgi, D.: A theory of bisimulation for the π-calculus. Acta Informatica 33(1), 69–97 (1996)
26. Tiu, A., Dawson, J.: Automating open bisimulation checking for the spi calculus. In: 2010 23rd IEEE Computer Security Foundations Symposium. pp. 307–321. IEEE (2010)