

Verifying a Hotel Key Card System

Tobias Nipkow

Institut für Informatik, TU München

Abstract. Two models of an electronic hotel key card system are contrasted: a state based and a trace based one. Both are defined, verified, and proved equivalent in the theorem prover Isabelle/HOL. It is shown that if a guest follows a certain safety policy regarding her key cards, she can be sure that nobody but her can enter her room.

1 Introduction

This paper presents two models for a hotel key card system and the verification of their safety (in Isabelle/HOL [5]). The models are based on Section 6.2, *Hotel Room Locking*, and Appendix E in the book by Daniel Jackson [2]. Jackson employs his Alloy system to check that there are no small counterexamples to safety. We confirm his conjecture of safety by a formal proof.

Most hotels operate a digital key card system. Upon check-in, you receive a card with your own key on it (typically a pseudorandom number). The lock for each room reads your card and opens the door if the key is correct. The system is decentralized, i.e. each lock is a stand-alone, battery-powered device without connection to the computer at reception or to any other device. So how does the lock know that your key is correct? There are a number of similar systems and we discuss the one described in Appendix E of [2]. Here each card carries two keys: the old key of the previous occupant of the room (key_1), and your own new key (key_2). The lock always holds one key, its “current” key. When you enter your room for the first time, the lock notices that its current key is key_1 on your card and recodes itself, i.e. it replaces its own current key with key_2 on your card. When you enter the next time, the lock finds that its current key is equal to your key_2 and opens the door without recoding itself. Your card is never modified by the lock. Eventually, a new guest with a new key enters the room, recodes the lock, and you cannot enter anymore.

After a short introduction of the notation we discuss two very different specifications, a state based and a trace based one, and prove their safety and their equivalence. The complete formalization is available online in the *Archive of Formal Proofs* at afp.sf.net.

1.1 Notation

HOL conforms largely to everyday mathematical notation. This section introduces further non-standard notation and in particular a few basic data types with their primitive operations.

Types The type of truth values is called *bool*. The space of total functions is denoted by \Rightarrow . Type variables start with a quote, as in *'a*, *'b* etc. The notation $t::\tau$ means that term t has type τ .

Functions can be updated at x with new value y , written $f(x := y)$. The range of a function is *range* f , *inj* f means f is injective.

Pairs come with the two projection functions $fst :: 'a \times 'b \Rightarrow 'a$ and $snd :: 'a \times 'b \Rightarrow 'b$.

Sets have type *'a set*.

Lists (type *'a list*) come with the empty list $[]$, the infix constructor \cdot , the infix $@$ that appends two lists, and the conversion function *set* from lists to sets. Variable names ending in “s” usually stand for lists.

Records are constructed like this $\langle f_1 = v_1, \dots \rangle$ and updated like this $r(f_i := v_i, \dots)$, where the f_i are the field names, the v_i the values and r is a record.

Datatype *option* is defined like this

$$\text{datatype } 'a \text{ option} = \text{None} \mid \text{Some } 'a$$

and adjoins a new element *None* to a type *'a*. For succinctness we write $[a]$ instead of *Some a*.

Note that $\llbracket A_1; \dots; A_n \rrbracket \Longrightarrow A$ abbreviates $A_1 \Longrightarrow \dots \Longrightarrow A_n \Longrightarrow A$, which is the same as “If A_1 and \dots and A_n then A ”.

2 A State Based Model

The model is based on three opaque types *guest*, *key* and *room*. Type *card* is just an abbreviation for $\text{key} \times \text{key}$.

The state of the system is modelled as a record which combines the information about the front desk, the rooms and the guests.

```
record state =
  owns :: room  $\Rightarrow$  guest option
  currk :: room  $\Rightarrow$  key
  issued :: key set
  cards :: guest  $\Rightarrow$  card set
  roomk :: room  $\Rightarrow$  key
  isin :: room  $\Rightarrow$  guest set
  safe :: room  $\Rightarrow$  bool
```

Reception records who *owns* a room (if anybody, hence *guest option*), the current key *currk* that has been issued for a room, and which keys have been *issued* so far. Each guest has a set of *cards*. Each room has a key *roomk* recorded in the lock and a set *isin* of occupants. The auxiliary variable *safe* is explained further below; we ignore it for now.

In specification languages like Z, VDM and B we would now define a number of operations on this state space. Since they are the only permissible operations on the state, this defines a set of *reachable* states. In a purely logical environment

like Isabelle/HOL this set can be defined directly by an inductive definition. Each clause of the definition corresponds to a transition/operation/event. This is the standard approach to modelling state machines in theorem provers.

The set of reachable states of the system (called *reach*) is defined by four transitions: initialization, checking in, entering a room, and leaving a room:

init:

$\text{inj } \text{initk} \implies$
 $\langle \text{owns} = (\lambda r. \text{None}), \text{currk} = \text{initk}, \text{issued} = \text{range } \text{initk},$
 $\text{cards} = (\lambda g. \{\}), \text{roomk} = \text{initk}, \text{isin} = (\lambda r. \{\}),$
 $\text{safe} = (\lambda r. \text{True}) \rangle \in \text{reach}$

check-in:

$\llbracket s \in \text{reach}; k \notin \text{issued } s \rrbracket \implies$
 $s \langle \text{currk} := (\text{currk } s)(r := k), \text{issued} := \text{issued } s \cup \{k\},$
 $\text{cards} := (\text{cards } s)(g := \text{cards } s \cup \{(\text{currk } s \ r, k)\}),$
 $\text{owns} := (\text{owns } s)(r := \text{Some } g),$
 $\text{safe} := (\text{safe } s)(r := \text{False}) \rangle \in \text{reach}$

enter-room:

$\llbracket s \in \text{reach}; (k, k') \in \text{cards } s \ g; \text{roomk } s \ r \in \{k, k'\} \rrbracket \implies$
 $s \langle \text{isin} := (\text{isin } s)(r := \text{isin } s \ r \cup \{g\}),$
 $\text{roomk} := (\text{roomk } s)(r := k'),$
 $\text{safe} := (\text{safe } s)(r := \text{owns } s \ r = \lfloor g \rfloor \wedge \text{isin } s \ r = \{ \} \wedge k' = \text{currk } s \ r$
 $\quad \vee \text{safe } s \ r) \rangle \in \text{reach}$

exit-room:

$\llbracket s \in \text{reach}; g \in \text{isin } s \ r \rrbracket \implies$
 $s \langle \text{isin} := (\text{isin } s)(r := \text{isin } s \ r - \{g\}) \rangle \in \text{reach}$

There is no check-out event because it is implicit in the next check-in for that room: this covers the cases where a guest leaves without checking out (in which case the room should not be blocked forever) or where the hotel decides to rent out a room prematurely, probably by accident. Neither do guests have to return their cards at any point because they may loose cards or may pretended to have lost them. We will now explain the events.

init Initialization requires that every room has a different key, i.e. that *currk* is injective. Nobody owns a room, the keys of all rooms are recorded as issued, nobody has a card, and all rooms are empty.

enter-room A guest may enter if either of the two keys on his card equal the room key. Then *g* is added to the occupants of *r* and the room key is set to the second key on the card. Normally this has no effect because the second key is already the room key. But when entering for the first time, the first key on the card equals the room key and then the lock is actually recoded.

exit-room removes an occupant from the occupants of a room.

check-in for room *r* and guest *g* issues the card (*currk* *s* *r*, *k*) to *g*, where *k* is new, makes *g* the owner of the room, and sets *currk* *s* *r* to the new key *k*.

The reader can easily check that our specification allows the intended distributed implementation: entering only reads and writes the key in that lock, and check-in only reads and writes the information at reception.

In contrast to Jackson we require that initially distinct rooms have distinct keys. This protects the hotel from its guests: otherwise a guest may be able to enter rooms he does not own, potentially stealing objects from those rooms. Of course he can also steal objects from his own room, but in that case it is easier to hold him responsible. In general, the hotel may just want to minimize the opportunity for theft.

The main difference to Jackson’s model is that his can talk about transitions between states rather than merely about reachable states. This means that he can specify that unauthorized entry into a room should not occur. Because our specification does not formalize the transition relation itself, we need to include the *isin* component in order to express the same requirement. In the end, we would like to establish that the system is *safe*: only the owner of a room can be in a room:

$$\llbracket s \in \text{reach}; g \in \text{isin } s \ r \rrbracket \implies \text{owns } s \ r = \lfloor g \rfloor$$

Unfortunately, this is just not true. It does not take a PhD in computer science to come up with the following scenario: because guests can retain their cards, there is nothing to stop a guest from reentering his old room after he has checked out (in our model: after the next guest has checked in), but before the next guest has entered his room. Hence the best we can do is to prove a conditional safety property: under certain conditions, the above safety property holds. The question is: which conditions? It is clear that the room must be empty when its owner enters it, or all bets are off. But is that sufficient? Unfortunately not. Jackson’s Alloy tool took 2 seconds [2, p. 303] to find the following “guest-in-the-middle” attack:

1. Guest 1 checks in and obtains a card (k_1, k_2) for room 1 (whose key in the lock is k_1). Guest 1 does not enter room 1.
2. Guest 2 checks in, obtains a card (k_2, k_3) for room 1, but does not enter room 1 either.
3. Guest 1 checks in again, obtains a card (k_3, k_4) , goes to room 1, opens it with his old card (k_1, k_2) , finds the room empty, and feels safe ...

After Guest 1 has left his room, Guest 2 enters and makes off with the luggage.

Jackson now assumes that guests return their cards upon check-out, which can be modelled as follows: upon check-in, the new card is not added to the guest’s set of cards but it replaces his previous set of cards, i.e. guests return old cards the next time they check in. Under this assumption, Alloy finds no more counterexamples to safety — at least not up to 6 cards and keys and 3 guests and rooms. This is not a proof but a strong indication that the given assumptions suffice for safety. We prove that this is indeed the case.

It should be noted that the system also suffers from a liveness problem: if a guest never enters the room he checked in to, that room is forever blocked. In practice this is dealt with by a master key. We ignore liveness.

2.1 Formalizing Safety

It should be clear that one cannot force guests to always return their cards (or, equivalently, never to use an old card). We can only prove that if they do, their room is safe. However, we do not follow Jackson's approach of globally assuming everybody returns their old cards upon check-in. Instead we would like to take a local approach where it is up to each guest whether he follows this safety policy. We allow guests to keep their cards but make safety dependent on how they use them. This generality requires a finer grained model: we need to record if a guest has entered his room in a safe manner, i.e. if it was empty and if he used the latest key for the room, the one stored at reception. The auxiliary variable *safe* records for each room if this was the case at some point between his last check-in and now. The main theorem will be that if a room is safe in this manner, then only the owner can be in the room. Now we explain how *safe* is modified with each event:

init sets *safe* to *True* for every room.

check-in for room *r* resets *safe s r* because it is not safe for the new owner yet.

enter-room for room *r* sets *safe s r* if the owner entered an empty room using the latest card issued for that room by reception, or if the room was already safe.

exit-room does not modify *safe*.

The reader should convince his or herself that *safe* corresponds to the informal safety policy set out above. Note that a guest may find his room non-empty the first time he enters, and *safe* will not be set, but he may come back later, find the room empty, and then *safe* will be set. Furthermore, it is important that *enter-room* cannot reset *safe* due to the disjunct $\vee \text{safe } s \ r$. Hence *check-in* is the only event that can reset *safe*. That is, a room stays safe until the next *check-in*. Additionally *safe* is initially *True*, which is fine because initially injectivity of *initk* prohibits illegal entries by non-owners.

Note that because none of the other state components depend on *safe*, it is truly auxiliary: it can be deleted from the system and the same set of reachable states is obtained, modulo the absence of *safe*.

We have formalized a very general safety policy of always using the latest card. A special case of this policy is the one called *NoIntervening* by Jackson [2, p. 200]: every *check-in* must immediately be followed by the corresponding *enter-room*.

2.2 Verifying Safety

All of our lemmas are invariants of *reach*. The complete list, culminating in the main theorem, is this:

- Lemma 1.**
1. $s \in \text{reach} \implies \text{currk } s \ r \in \text{issued } s$
 2. $\llbracket s \in \text{reach}; (k, k') \in \text{cards } s \ g \rrbracket \implies k \in \text{issued } s$
 3. $\llbracket s \in \text{reach}; (k, k') \in \text{cards } s \ g \rrbracket \implies k' \in \text{issued } s$

4. $s \in \text{reach} \implies \text{roomk } s \ k \in \text{issued } s$
5. $s \in \text{reach} \implies \forall r \ r'. (\text{currk } s \ r = \text{currk } s \ r') = (r = r')$
6. $s \in \text{reach} \implies (\text{currk } s \ r, k') \notin \text{cards } s \ g$
7. $\llbracket s \in \text{reach}; (k_1, k) \in \text{cards } s \ g_1; (k_2, k) \in \text{cards } s \ g_2 \rrbracket \implies g_1 = g_2$
8. $\llbracket s \in \text{reach}; \text{safe } s \ r \rrbracket \implies \text{roomk } s \ r = \text{currk } s \ r$
9. $\llbracket s \in \text{reach}; \text{safe } s \ r; (k', \text{roomk } s \ r) \in \text{cards } s \ g \rrbracket \implies \text{owns } s \ r = \lfloor g \rfloor$

Theorem 1. If $s \in \text{reach}$ and $\text{safe } s \ r$ and $g \in \text{isin } s \ r$ then $\text{owns } s \ r = \lfloor g \rfloor$.

The lemmas and the theorem are proved in this order, each one is marked as a simplification rule, and each proof is a one-liner: induction on $s \in \text{reach}$ followed by *auto*.

Although, or maybe even because these proofs work so smoothly one may like to understand why. Hence we examine the proof of Theorem 1 in more detail. The only interesting case is *enter-room*. We assume that guest g_1 enters room r_1 with card (k_1, k_2) and call the new state t . We assume $\text{safe } t \ r$ and $g \in \text{isin } t \ r$ and prove $\text{owns } t \ r = \lfloor g \rfloor$ by case distinction. If $r_1 \neq r$, the claim follows directly from the induction hypothesis using $\text{safe } s \ r$ and $g \in \text{isin } t \ r$ because $\text{owns } t \ r = \text{owns } s \ r$ and $\text{safe } t \ r = \text{safe } s \ r$. If $r_1 = r$ then $g \in \text{isin } t \ r$ is equivalent with $g \in \text{isin } s \ r \vee g = g_1$. If $g \in \text{isin } s \ r$ then $\text{safe } s \ r$ follows from $\text{safe } t \ r$ by definition of *enter-room* because $g \in \text{isin } s \ r$ implies $\text{isin } s \ r \neq \emptyset$. Hence the induction hypothesis implies the claim. If $g = g_1$ we make another case distinction. If $k_2 = \text{roomk } s \ r$, the claim follows immediately from Lemma 1.9 above: only the owner of a room can possess a card where the second key is the room key. If $k_1 = \text{roomk } s \ r$ then, by definition of *enter-room*, $\text{safe } t \ r$ implies $\text{owns } s \ r = \lfloor g \rfloor \vee \text{safe } s \ r$. In the first case the claim is immediate. If $\text{safe } s \ r$ then $\text{roomk } s \ r = \text{currk } s \ r$ (by Lemma 1.8) and thus $(\text{currk } s \ r, k_2) \in \text{cards } s \ g$ by assumption $(k_1, k_2) \in \text{cards } s \ g_1$, thus contradicting Lemma 1.6.

This detailed proof shows that a number of case distinctions are required. Luckily, they all suggest themselves to Isabelle via the definition of function update ($:=$) or via disjunctions that arise automatically.

2.3 An Extension

To test the flexibility of our model we extended it with the possibility for obtaining a new card, e.g. when one has lost one's card. Now reception needs to remember not just the current but also the previous key for each room, i.e. a new field $\text{prevk} :: \text{room} \Rightarrow \text{key}$ is added to *state*. It is initialized with the same value as *currk*: though strictly speaking it could be arbitrary, this permits the convenient invariant $\text{prevk } s \ r \in \text{issued } s$. Upon check-in we set *prevk* to $(\text{prevk } s)(r := \text{currk } s \ r)$. Event *new-card* is simple enough:

$$\begin{aligned} &\llbracket s \in \text{reach}; \text{owns } s \ r = \lfloor g \rfloor \rrbracket \\ &\implies s(\text{cards} := (\text{cards } s)(g := \text{cards } s \ g \cup \{(\text{prevk } s \ r, \text{currk } s \ r)\})) \in \text{reach} \end{aligned}$$

The verification is not seriously affected. Some additional invariants are required

$$\begin{aligned}
s \in \text{reach} &\implies \text{prevk } s \ r \in \text{issued } s \\
\llbracket s \in \text{reach}; \text{owns } s \ r' = \lfloor g \rfloor \rrbracket &\implies \text{currk } s \ r \neq \text{prevk } s \ r' \\
\llbracket s \in \text{reach}; \text{owns } s \ r = \lfloor g \rfloor; g \neq g' \rrbracket &\implies (k, \text{currk } s \ r) \notin \text{cards } s \ g'
\end{aligned}$$

but the proofs are still of the same trivial induct-auto format.

Adding a further event for loosing a card has no impact at all on the proofs.

3 A Trace Based Model

The only clumsy aspect of the state based model is *safe*: we use a state component to record if the sequence of events that lead to a state satisfies some property. That is, we simulate a condition on traces via the state. Unsurprisingly, it is not trivial to convince oneself that *safe* really has the informal meaning set out at the beginning of subsection 2.1. Hence we now describe an alternative, purely trace based model, similar to Paulson's inductive protocol model [6]. The events are:

datatype *event* =

Check-in guest room card | *Enter guest room card* | *Exit guest room*

Instead of a state, we have a trace, i.e. list of events, and extract the state from the trace:

$\text{initk} :: \text{room} \Rightarrow \text{key}$
 $\text{owns} :: \text{event list} \Rightarrow \text{room} \Rightarrow \text{guest option}$
 $\text{currk} :: \text{event list} \Rightarrow \text{room} \Rightarrow \text{key}$
 $\text{issued} :: \text{event list} \Rightarrow \text{key set}$
 $\text{cards} :: \text{event list} \Rightarrow \text{guest} \Rightarrow \text{card set}$
 $\text{roomk} :: \text{event list} \Rightarrow \text{room} \Rightarrow \text{key}$
 $\text{isin} :: \text{event list} \Rightarrow \text{room} \Rightarrow \text{guest set}$
 $\text{hotel} :: \text{event list} \Rightarrow \text{bool}$

Except for *initk*, which is completely unspecified, all these functions are defined by primitive recursion over traces:

$\text{owns } [] \ r = \text{None}$
 $\text{owns } (e \cdot s) \ r =$
 $(\text{case } e \text{ of } \text{Check-in } g \ r' \ c \Rightarrow \text{if } r' = r \text{ then } \lfloor g \rfloor \text{ else } \text{owns } s \ r$
 $\quad | - \Rightarrow \text{owns } s \ r)$
 $\text{currk } [] \ r = \text{initk } r$
 $\text{currk } (e \cdot s) \ r =$
 $(\text{let } k = \text{currk } s \ r$
 $\quad \text{in case } e \text{ of } \text{Check-in } g \ r' \ c \Rightarrow \text{if } r' = r \text{ then } \text{snd } c \text{ else } k \mid - \Rightarrow k)$
 $\text{issued } [] = \text{range initk}$
 $\text{issued } (e \cdot s) = \text{issued } s \cup (\text{case } e \text{ of } \text{Check-in } g \ r \ c \Rightarrow \{\text{snd } c\} \mid - \Rightarrow \emptyset)$
 $\text{cards } [] \ g = \emptyset$
 $\text{cards } (e \cdot s) \ g =$
 $(\text{let } C = \text{cards } s \ g$
 $\quad \text{in case } e \text{ of } \text{Check-in } g' \ r \ c \Rightarrow \text{if } g' = g \text{ then } \{c\} \cup C \text{ else } C \mid - \Rightarrow C)$

$roomk \ [] \ r = initk \ r$
 $roomk \ (e \cdot s) \ r =$
 $(let \ k = roomk \ s \ r$
 $in \ case \ e \ of \ Enter \ g \ r' \ (x, y) \Rightarrow if \ r' = r \ then \ y \ else \ k \ | \ - \Rightarrow k)$
 $isin \ [] \ r = \emptyset$
 $isin \ (e \cdot s) \ r =$
 $(let \ G = isin \ s \ r$
 $in \ case \ e \ of \ Check-in \ g \ r \ c \Rightarrow G$
 $\quad | \ Enter \ g \ r' \ c \Rightarrow if \ r' = r \ then \ \{g\} \cup G \ else \ G$
 $\quad | \ Exit \ g \ r' \Rightarrow if \ r' = r \ then \ G - \{g\} \ else \ G)$

However, not every trace is possible. Function *hotel* tells us which traces correspond to real hotels:

$hotel \ [] = True$
 $hotel \ (e \cdot s) =$
 $(hotel \ s \wedge$
 $\quad (case \ e \ of \ Check-in \ g \ r \ (k, k') \Rightarrow k = currk \ s \ r \wedge k' \notin issued \ s$
 $\quad | \ Enter \ g \ r \ (k, k') \Rightarrow (k, k') \in cards \ s \ g \wedge roomk \ s \ r \in \{k, k'\}$
 $\quad | \ Exit \ g \ r \Rightarrow g \in isin \ s \ r))$

Alternatively we could have followed Paulson [6] in defining *hotel* as an inductive set of traces. The difference is only slight.

3.1 Formalizing Safety

The principal advantage of the trace model is the intuitive specification of safety. Using the auxiliary predicate *no-Check-in*

$no-Check-in \ s \ r \equiv \neg(\exists g \ c. \ Check-in \ g \ r \ c \in set \ s)$

we define a trace to be *safe₀* for a room if the card obtained at the last *Check-in* was later actually used to *Enter* the room:

$safe_0 \ s \ r \equiv \exists s_1 \ s_2 \ s_3 \ g \ c.$
 $s = s_3 \ @ \ [Enter \ g \ r \ c] \ @ \ s_2 \ @ \ [Check-in \ g \ r \ c] \ @ \ s_1 \wedge no-Check-in \ (s_3 \ @ \ s_2) \ r$

A trace is *safe* if additionally the room was empty when it was entered:

$safe \ s \ r \equiv \exists s_1 \ s_2 \ s_3 \ g \ c.$
 $s = s_3 \ @ \ [Enter \ g \ r \ c] \ @ \ s_2 \ @ \ [Check-in \ g \ r \ c] \ @ \ s_1 \wedge$
 $no-Check-in \ (s_3 \ @ \ s_2) \ r \wedge isin \ (s_2 \ @ \ [Check-in \ g \ r \ c] \ @ \ s_1) \ r = \{\}$

The two notions of safety are distinguished because, except for the main theorem, *safe₀* suffices.

The alert reader may already have wondered why, in contrast to the state based model, we do not require *initk* to be injective. If *initk* is not injective, e.g. $initk \ r_1 = initk \ r_2$ and $r_1 \neq r_2$, then $[Enter \ g \ r_2 \ (initk \ r_1, k), Check-in \ g \ r_1 \ (initk \ r_1, k)]$ is a legal trace and guest *g* ends up in a room he is not the owner of. However, this is not a safe trace for room r_2 according to our definition. This reflects that hotel rooms are not safe until the first time their owner has entered them. We no longer protect the hotel from its guests.

3.2 Verifying Safety

Lemma 1 largely carries over after replacing $s \in reach$ by *hotel* s and *safe* by *safe*₀. Only properties 5 and 6 no longer hold because we no longer assume that *roomk* is initially injective. They are replaced by two somewhat similar properties:

Lemma 2.

1. $\llbracket hotel (s_2 @ Check-in\ g\ r\ (k, k') \cdot s_1);$
 $k' = currk (s_2 @ Check-in\ g\ r\ (k, k') \cdot s_1)\ r \rrbracket$
 $\implies r' = r$
2. $\llbracket hotel (s_2 @ [Check-in\ g\ r\ (k, k')] @ s_1); no-Check-in\ s_2\ r \rrbracket$
 $\implies (k', k') \notin cards (s_2 @ Check-in\ g\ r\ (k, k') \cdot s_1)\ g'$

Both are proved by induction on s_2 . In addition we need some easy structural properties:

- Lemma 3.**
1. $issued (s @ s') = issued\ s \cup issued\ s'$
 2. $no-Check-in\ s_2\ r \implies owns (s_2 @ s_1)\ r = owns\ s_1\ r$
 3. $no-Check-in\ s_2\ r \implies currk (s_2 @ s_1)\ r = currk\ s_1\ r$

The main theorem again correspond closely to its state based counterpart:

Theorem 2. If *hotel* s and *safe* $s\ r$ and $g \in isin\ s\ r$ then $owns\ s\ r = \lfloor g \rfloor$.

Let us examine the proof of this theorem to show how it differs from the state based version. For the core of the proof let $s = s_3 @ [Enter\ g'\ r\ (k, k')] @ s_2 @ [Check-in\ g'\ r\ (k, k')] @ s_1$ and assume $isin\ s_2 @ [Check-in\ g'\ r\ (k, k')] @ s_1\ r = \emptyset$ (0). By induction on s_3 we prove

$$\llbracket hotel\ s; no-Check-in (s_3 @ s_2)\ r; g \in isin\ s\ r \rrbracket \implies g' = g$$

The actual theorem follows by definition of *safe*. The base case of the induction follows from (0). For the induction step let $t = (e \cdot s_3) @ [Enter\ g'\ r\ (k, k')] @ s_2 @ [Check-in\ g'\ r\ (k, k')] @ s_1$. We assume *hotel* t , *no-Check-in* $((e \cdot s_3) @ s_2)\ r$, and $g \in isin\ s\ r$, and show $g' = g$. The proof is by case distinction on the event e . The cases *Check-in* and *Exit* follow directly from the induction hypothesis because the set of occupants of r can only decrease. Now we focus on the case $e = Enter\ g''\ r'\ c$. If $r' \neq r$ the set of occupants of r remains unchanged and the claim follow directly from the induction hypothesis. If $g'' \neq g$ then g must already have been in r before the *Enter* event and the claim again follows directly from the induction hypothesis. Now assume $r' = r$ and $g'' = g$. From *hotel* t we obtain *hotel* s (1) and $c \in cards\ s\ g$ (2), and from *no-Check-in* $(s_3 @ s_2)\ r$ and (0) we obtain *safe* $s\ r$ (3). Let $c = (k_1, k_2)$. From Lemma 1.8 and Lemma 3.3 we obtain *roomk* $s\ r = currk\ s\ r = k'$. Hence $k_1 \neq roomk\ s\ r$ by Lemma 2.2 using (1), (2) and *no-Check-in* $(s_3 @ s_2)\ r$. Hence $k_2 = roomk\ s\ r$ by *hotel* t . With Lemma 1.9 and (1–3) we obtain $owns\ t\ r = \lfloor g \rfloor$. At the same time we have $owns\ t\ r = \lfloor g' \rfloor$ because *hotel* t and *no-Check-in* $((e \cdot s_3) @ s_2)\ r$: nobody has checked in to room r after g' . Thus the claim $g' = g$ follows.

The details of this proof differ from those of Theorem 1 but the structure is very similar.

theorem *safe*: **assumes** *hotel s* and *safe s r* and $g \in \text{isin } s \ r$
shows $\text{owns } s \ r = \lfloor g \rfloor$

proof –

```

{ fix  $s_1 \ s_2 \ s_3 \ g' \ k \ k'$ 
  let  $?s = s_3 @ [\text{Enter } g' \ r \ (k, k')] @ s_2 @ [\text{Check-in } g' \ r \ (k, k')] @ s_1$ 
  assume  $0: \text{isin } (s_2 @ [\text{Check-in } g' \ r \ (k, k')] @ s_1) \ r = \{\}$ 
  have  $\llbracket \text{hotel } ?s; \text{no-Check-in } (s_3 @ s_2) \ r; \ g \in \text{isin } ?s \ r \rrbracket \implies g' = g$ 
  proof(induct  $s_3$ )
    case Nil thus ?case using 0 by simp
  next
    case (Cons  $e \ s_3'$ )
    let  $?b = [\text{Enter } g' \ r \ (k, k')] @ s_2 @ [\text{Check-in } g' \ r \ (k, k')] @ s_1$ 
    let  $?s = s_3' @ ?b$  and  $?t = (e \cdot s_3') @ ?b$ 
    show ?case
    proof(cases  $e$ )
      case (Enter  $g'' \ r' \ c$ )[simp]
      show  $g' = g$ 
      proof cases
        assume [simp]:  $r' = r$ 
        show  $g' = g$ 
        proof cases
          assume [simp]:  $g'' = g$ 
          have 1: hotel  $?s$  and 2:  $c \in \text{cards } ?s \ g$  using  $\langle \text{hotel } ?t \rangle$  by auto
          have 3: safe  $?s \ r$  using  $\langle \text{no-Check-in } ((e \cdot s_3') @ s_2) \ r \rangle 0$ 
            by(simp add:safe-def) blast
          obtain  $k_1 \ k_2$  where [simp]:  $c = (k_1, k_2)$  by force
          have  $\text{roomk } ?s \ r = k'$ 
            using safe-roomk-currk[OF 1 safe-safe[OF 3]]
             $\langle \text{no-Check-in } ((e \cdot s_3') @ s_2) \ r \rangle$  by auto
          hence  $k_1 \neq \text{roomk } ?s \ r$ 
            using no-checkin-no-newkey[where  $s_2 = s_3' @ [\text{Enter } g' \ r \ (k, k')] @ s_2$ ]
            1 2  $\langle \text{no-Check-in } ((e \cdot s_3') @ s_2) \ r \rangle$  by auto
          hence  $k_2 = \text{roomk } ?s \ r$  using  $\langle \text{hotel } ?t \rangle$  by auto
          with only-owner-enter-normal[OF 1 safe-safe[OF 3]] 2
          have  $\text{owns } ?t \ r = \lfloor g \rfloor$  by auto
          moreover have  $\text{owns } ?t \ r = \lfloor g' \rfloor$ 
            using  $\langle \text{hotel } ?t \rangle \langle \text{no-Check-in } ((e \cdot s_3') @ s_2) \ r \rangle$  by simp
          ultimately show  $g' = g$  by simp
        end
      end
    end
  next
    assume  $g'' \neq g$  thus  $g' = g$  using Cons by auto
  qed
next
  assume  $r' \neq r$  thus  $g' = g$  using Cons by auto
qed
qed (insert Cons, auto)
qed
} with prems show  $\text{owns } s \ r = \lfloor g \rfloor$  by(auto simp:safe-def)
qed

```

Fig. 1. Isar proof of Theorem 2

3.3 Eliminating *isin*

In the state based approach we needed *isin* to express our safety guarantees. In the presence of traces, we can do away with it and talk about *Enter* events instead. We show that if somebody enters a safe room, he is the owner:

Theorem 3. If *hotel* (*Enter* $g\ r\ c\cdot s$) and $\text{safe}_0\ s\ r$ then $\text{owns}\ s\ r = \lfloor g \rfloor$.

From $\text{safe}_0\ s\ r$ it follows that s must be of the form $s_2 @ [\text{Check-in } g_0\ r\ c'] @ s_1$ such that *no-Check-in* $s_2\ r$. Let $c = (x, y)$ and $c' = (k, k')$. By Lemma 1.8 we have *roomk* $s\ r = \text{currk}\ s\ r = k'$. From *hotel* (*Enter* $g\ r\ c\cdot s$) it follows that $(x, y) \in \text{cards}\ s\ g$ and $k' \in \{x, y\}$. By Lemma 2.2 $x = k'$ would contradict $(x, y) \in \text{cards}\ s\ g$. Hence $y = k'$. With Lemma 1.9 we obtain $\text{owns}\ s\ r = \lfloor g \rfloor$.

Having dispensed with *isin* we could also eliminate *Exit* to arrive at a model closer to the ones in [2].

Finally one may quibble that all the safety theorems proved so far assume safety of the room at that point in time when somebody enters it. That is, the owner of the room must be sure that once a room is safe, it stays safe, in order to profit from those safety theorems. Of course, this is the case as long as nobody else checks in to that room:

Lemma 4. If $\text{safe}_0\ s\ r$ and *no-Check-in* $s'\ r$ then $\text{safe}_0\ (s' @ s)\ r$.

It follows easily that Theorem 3 also extends until check-in:

Corollary 1. If *hotel* (*Enter* $g\ r\ c\cdot s' @ s$) and $\text{safe}_0\ s\ r$ and *no-Check-in* $s'\ r$ then $\text{owns}\ s\ r = \lfloor g \rfloor$.

3.4 Completeness of *safe*

Having proved correctness of *safe*, i.e. that safe behaviour protects against intruders, one may wonder if *safe* is complete, i.e. if it covers all safe behaviour, or if it is too restrictive. It turns out that *safe* is incomplete for two different reasons. The trivial one is that in case *initk* is injective, every room is protected against intruders right from the start. That is, $[\text{Check-in } g\ r\ c]$ will only allow g to enter r until somebody else checks in to r . The second, more subtle incompleteness is that even if there are previous owners of a room, it may be safe to enter a room with an old card c : one merely needs to make sure that no other guest checked in after the check-in where one obtained c . However, formalizing this is not only messy, it is also somewhat pointless: this liberalization is not something a guest can take advantage of because there is no (direct) way he can find out which of his cards meets this criterion. But without this knowledge, the only safe thing to do is to make sure he has used his latest card. This incompleteness applies to the state based model as well.

4 Equivalence

Although the state based and the trace based model look similar enough, the nagging feeling remains that they could be subtly different. Hence I wanted to show the equivalence formally. This was very fortunate, because it revealed

some unintended discrepancies (no longer present). Although I had proved both systems safe, it turned out that the state based version of safety was more restrictive than the trace based one. In the state based version of *safe* the room had to be empty the first time the owner enters with the latest card, whereas in the trace based version any time the owner enters with the latest card can make a room safe. Such errors in an automaton checking a trace property are very common and show the superiority of the trace based formalism.

When comparing the two models we have to take two slight differences into account:

- The initial setting of the room keys *initk* in the trace based model is an arbitrary but fixed value. In the state based model any injective initial value is fine.
- As a consequence (see the end of Section 3.1) *state.safe* is initially true whereas *Trace.safe* is initially false.

Since many names occur in both models they are disambiguated by the prefixes *state* and *Trace*.

In the one direction I have shown that any hotel trace starting with an injective *initk* gives rise to a reachable state when the components of that state are computed by the trace functions:

$$\begin{aligned}
 & \llbracket inj\ initk; hotel\ t \rrbracket \\
 \implies & \langle state.owns = Trace.owns\ t, currk = Trace.currk\ t, \\
 & \quad issued = Trace.issued\ t, cards = Trace.cards\ t, roomk = Trace.roomk\ t, \\
 & \quad isin = Trace.isin\ t, \\
 & \quad safe = \lambda r. Trace.safe\ t\ r \vee Trace.owns\ t\ r = None \rangle \\
 & \in reach
 \end{aligned}$$

Conversely, for any reachable state there is a hotel trace leading to it:

$$\begin{aligned}
 s \in reach & \implies \\
 \exists t\ ik. & \\
 initk = ik & \longrightarrow \\
 hotel\ t \wedge & \\
 state.cards\ s = Trace.cards\ t \wedge & \\
 state.isin\ s = Trace.isin\ t \wedge & \\
 state.roomk\ s = Trace.roomk\ t \wedge & \\
 state.owns\ s = Trace.owns\ t \wedge & \\
 state.currk\ s = Trace.currk\ t \wedge & \\
 state.issued\ s = Trace.issued\ t \wedge & \\
 state.safe\ s = (\lambda r. Trace.safe\ t\ r \vee Trace.owns\ t\ r = None) &
 \end{aligned}$$

The precondition *initk* = *ik* just says that we can find some interpretation for *initk* that works, namely the one that was chosen as the initial setting for the keys in *s*.

The proofs are almost automatic, except for the *safe* component. In essence, we have to show that the procedural *state.safe* implements the declarative *Trace.safe*. The proof was complicated by the fact that initially it was not true and I had to debug *Trace.safe* by proof. Unfortunately Isabelle's current

counterexample finders [1,7] did not seem to work here due to search space reasons. Once the bugs were ironed out, the following key lemma, together with some smaller lemmas, automated the correspondence proof for *safe*:

$$\begin{aligned} & \text{hotel } (\text{Enter } g \ r \ (k, k') \cdot t) \implies \\ & \text{Trace.safe } (\text{Enter } g \ r \ (k, k') \cdot t) \ r = \\ & (\text{Trace.owns } t \ r = \lfloor g \rfloor \wedge \text{Trace.isin } t \ r = \emptyset \wedge k' = \text{Trace.currk } t \ r \vee \\ & \text{Trace.safe } t \ r) \end{aligned}$$

In addition we used many lemmas from the trace model, including Theorem 2.

5 Conclusion

We have seen two different specification styles in this case study. The state based one is conceptually simpler, but may require auxiliary state components which express properties of the trace that lead to that state. And it may not be obvious if the definition of the state component correctly captures the desired property of the trace. A trace based specification expresses those properties directly. The proofs in the state based version are all automatic whereas in the trace based setting 4 proofs (out of 15) require special care, thus more than doubling the overall proof size. It would be interesting to test Isabelle's emerging link with automatic first-order provers [3] on the trace based proofs.

There are two different proof styles in Isabelle: unstructured apply-scripts [5] and structured Isar proofs [8,4]. Figure 1 shows an example of the latter. Even if the reader is unfamiliar with Isar, it is easy to see that this proof is very close to the version given in the text. Although apply-scripts are notoriously obscure, and even the author may not have an intuitive grasp of the structure of the proof, in our kind of application they also have advantages. In the apply-style, Isabelle's proof methods prove as much as possible automatically and leave the remaining cases to the user. This leads to much shorter (but more brittle) proofs: The (admittedly detailed) proof in Figure 1 was obtained from an apply-script of less than half the size.

The models given in this paper are very natural but by no means the only possible ones. Jackson himself uses an alternative trace based one which replaces the list data structure by an explicit notion of time. It would be interesting to see further treatments of this problem in other formalisms, for example temporal logics.

Acknowledgments. Daniel Jackson got me started on this case study, Stefano Berardi streamlined my proofs, and Larry Paulson commented on the paper at short notice.

References

1. Stefan Berghofer and Tobias Nipkow. Random testing in Isabelle/HOL. In J. Cuellar and Z. Liu, editors, *Software Engineering and Formal Methods (SEFM 2004)*, pages 230–239. IEEE Computer Society, 2004.
2. Daniel Jackson. *Software Abstractions. Logic, Language, and Analysis*. MIT Press, 2006.

3. Jia Meng, Claire Quigley, and Lawrence C. Paulson. Automation for interactive proof: First prototype. *Information and Computation*. In press.
4. Tobias Nipkow. Structured Proofs in Isar/HOL. In H. Geuvers and F. Wiedijk, editors, *Types for Proofs and Programs (TYPES 2002)*, volume 2646, pages 259–278, 2003.
5. Tobias Nipkow, Lawrence Paulson, and Markus Wenzel. *Isabelle/HOL — A Proof Assistant for Higher-Order Logic*, volume 2283. 2002. <http://www.in.tum.de/~nipkow/LNCS2283/>.
6. Lawrence C. Paulson. The inductive approach to verifying cryptographic protocols. *J. Computer Security*, 6:85–128, 1998.
7. Tjark Weber. Bounded model generation for Isabelle/HOL. In W. Ahrendt, P. Baumgartner, H. de Nivelle, S. Ranise, and C. Tinelli, editors, *Selected Papers from the Workshops on Disproving and the Second International Workshop on Pragmatics of Decision Procedures (PDPAR 2004)*, volume 125(3) of *Electronic Notes in Theoretical Computer Science*, pages 103–116, 2005.
8. Markus Wenzel. *Isabelle/Isar — A Versatile Environment for Human-Readable Formal Proof Documents*. PhD thesis, Institut für Informatik, Technische Universität München, 2002. <http://tumb1.biblio.tu-muenchen.de/publ/diss/in/2002/wenzel.html>.