Logical Omniscience in the Semantics of BAN Logic[*]

Mika Cohen Mads Dam

School of Information and Communication Technology, KTH, Stockholm, Sweden
{mikac,mfd}@imit.kth.se

Abstract

BAN logic is an epistemic logic for verification of cryptographic protocols. A number of semantics have been proposed for BAN logic, but none of them capture the intended meaning of the epistemic modality in a satisfactory way. This is due to the so-called *logical omniscience problem*: Agents are ”ideal reasoners” in existing semantics, while agents in BAN logic have only limited cryptographic reasoning powers. Logical omniscience is unavoidable in Kripke semantics, the standard semantical framework in epistemic logic. Our proposal is to generalize the epistemic accessibility relation of Kripke semantics so that it changes not only the current execution point, but also the currently predicated message. When instantiated on message passing systems, the semantics validates BAN logic. It makes agents introspective (”self-aware”) of their own knowledge and of their own actions of sending, receiving and extracting.

Keywords: BAN logic; Epistemic logic; Kripke semantics; Security protocols; Logical omniscience problem

1 Introduction

BAN logic, proposed by Burrows, Abadi and Needham in the late eighties, is an epistemic logic for verification of cryptographic protocols ([4]). From a practical point of view, BAN logic has turned out to be quite successful: It produces short, informative derivations that can reveal subtle protocol errors. However, despite a number of semantics proposed for BAN and BAN-like logic (cf. [1, 5, 8, 10, 11, 12, 14]), the semantics of the epistemic (knowledge) modality in BAN logic remains problematic. This is a serious problem, since it makes it unclear what a proof in BAN logic establishes, and it makes an analysis of BAN logic in semantical terms, for instance using model checking, of limited value.

The basic problem when interpreting BAN's knowledge modality is the well-known *logical omniscience problem*. As an example, under BAN's idealized treatment of cryptography it is reasonable to assume the entailment M *fresh* $\models \{M\}_k$ *fresh*. However, the entailment *a knows M fresh* $\models$ *a knows* $\{M\}_k$ *fresh* should not be validated since in BAN logic agent a knows M is inside $\{M\}_k$ only when a knows k. From the point of view of modal logic, the example shows the failure of the *rule of normality* that allows inference of an entailment a *knows* $F_1 \models a$ *knows* F_2 from the entailment $F_1 \models F_2$. As another example, in the context of the NSSK protocol it is reasonable to assume the entailment s *said* $n, b, k, \{k, a\}_{k_b} \models k_b$ *good for* $b \cdot s$ since the former message is only ever uttered by s when it so happens that k_b is b:s server key (and therefore is good for communication between b and s). Yet, the entailment

$$a \textit{ knows } s \textit{ said } n, b, k, \{k, a\}_{k_b} \models a \textit{ knows } k_b \textit{ good for } b \cdot s \quad (1)$$

[*] Work supported by the Swedish Research Council grants 621-2003-2597 and 622-2003-6108

should not be validated, since in BAN logic agent a can deduce what key $\{k,a\}_{k_b}$ is locked with only if a already knows k_b. In fact, from (1) together with BAN's message meaning rule, we would get the entailment

$$a \textit{ sees } \{\textit{from s} : n, b, k, \{k,a\}_{k_b}\}_{k_a},\ a \textit{ knows } k_a \textit{ good for } a \cdot s \models a \textit{ knows } k_b \textit{ good for } b \cdot c$$

which diverges even more strongly from the intended meaning in BAN logic.

Logical omniscience (the rule of normality) is intimately tied to the use of Kripke semantics. In this type of semantics the modality *a knows* is interpreted through an epistemic accessibility relation $\sim_a$ connecting execution points that are equivalent up to a's restricted power of observation: At execution point s, a *knows* F just in case F holds at every accessible execution point s', $s \sim_a s'$.

Since all Kripke semantics validate the rule of normality, it follows that we need to look to non-Kripkean semantics to avoid validities that are unfaithful to the intended meaning in BAN logic. We suggest a generalization of Kripke semantics that lets the jump from the current execution point to an epistemically accessible execution point affect the predicated messages. The intuition is as follows. Say an agent a views a cipher text M at the current execution point s. As in Kripke semantics we assume that a may be unsure about what execution point she is at, because s and some other execution point s' share the same history up to a's observation powers. In addition, a may be unsure about what the cipher text contains, because a has observed the same properties of M at s as she would have observed of some corresponding cipher text M' at s'. For instance, if a extracts M from the third message a received at s, then a extracts M' from the third message a received at s'; if a cannot decrypt M at s, then a cannot decrypt M' at s', and so on.

To reflect the correspondence between messages at different execution points we relativize accessibility to message renamings. We write $s \sim_a^r s'$ when renaming r carries each message M at s to a corresponding message $r(M)$ at s'. With the relativized accessibility relation, a generalization of Kripke semantics is immediate:

$$s \models a \textit{ knows } F(M) \Leftrightarrow \forall s' : \forall r : s \sim_a^r s' \Rightarrow s' \models F(r(M))\ .$$

For instance, agent a knows that M is fresh, if all corresponding messages at epistemically accessible execution points are fresh.

This semantics avoids logical omniscience, since the predicated message M might change under r as we move from s to an epistemically accessible point s'. There is, however, an interesting weakening of normality which continues to hold, namely the closure of knowledge under validities that only mention keys used by the agent.

$$F_1 \models F_2 \Rightarrow a \textit{ uses Keys}(F_1, F_2),\ a \textit{ knows } F_1 \models a \textit{ knows } F_2$$

where *Keys*(F_1, F_2) contains all message terms that are applied as keys in F_1 and F_2. To illustrate, from the entailment x *fresh* $\models \{x\}_y$ *fresh* we can infer the entailment *a uses* y, *a knows* x *fresh* $\models$ *a knows* $\{x\}_y$ *fresh*. By universal substitution of message terms for variables, we can then conclude the entailment

$$a \textit{ uses } K,\ a \textit{ knows } M \textit{ fresh } \models a \textit{ knows } \{M\}_K \textit{ fresh} \tag{2}$$

for arbitrary (complex) message terms K and M, even when keys other than K are applied in M.

After instantiating the semantics on message passing systems, we show that agents are introspective of their own knowledge, i.e. the modal logic *S5* axioms hold, as is the custom in computer science applications of epistemic logic. Furthermore we show that agents are introspective of their own actions of sending, receiving and extracting (decryption and un-pairing of received messages). For instance, we show introspection of received messages: *a received M* $\models$ *a knows a received M*. While this is immediate from the truth condition for knowledge, it is rather significant. Firstly, it is the central point when validating BAN logic. The unsoundness of BAN logic in related Kripke semantics, such as [1, 12, 14], can ultimately be tied back to the fact that agents are not introspective (in the above sense) of their received messages [1]. As soon as a Kripke semantics hides part of an agents local state to the agent herself, as these semantics do, we lose introspection of received messages. Secondly, introspection of received messages in combination with the above weakening of normality has an interesting implication: knowledge of cryptographic structure may at times transcend the discerning power of the keys used.

We complete the model construction by interpreting the atomic BAN predicates on message passing systems and show soundness of BAN logic. The interpretation we propose involves a fixed point construction to identify keys used with keys known, a construction which may be of independent interest. Finally the paper is closed by a discussion of related and future work, in particular the prospects for using the weakened rule of normality to eliminate BAN's idealization step.

Our semantical investigations so far cover only the symmetric key part of BAN logic. We expect no difficulties in extending the semantics to asymmetric cryptography.

2 BAN Logic

Language Assume a set of agents $a, b, ...$, a set of *message atoms* $k, n, ...$, a set of *message variables* $x, y, z, ...$, and a set of *atomic predicates* p. The set of *message terms* and *statements* are defined by:

Statements $F ::= \; p(M) \mid$ *a knows* F
Message terms $M, M' ::= \; F \mid a \mid k \mid x \mid M, M' \mid \{M\}_{M'} \mid$ *from a : M*

A closed message term, or *message*, is a message term with no variables. A message term is open if it is not closed. Though the BAN language lacks negation, we prove a result (Theorem 9.2) for a language extended with negation ($\neg$) of statements.

Intuitively, atomic statement $p(M)$ expresses the proposition that message M satisfies property p, the operator $\cdot, \cdot$ represents pairing of messages, the operator $\{\cdot\}.$ represents encryption and the operator *from* $\cdot : \cdot$ represents sender field annotation. Message terms include sender field annotations and statements, because as BAN logic is usually applied, it proves properties of so called *idealized* protocols, protocols where messages may include a sender field and messages may contain statements expressing propositions.

The set of atomic predicates includes, at least, the four *atomic BAN predicates*: *a sees, a said, fresh, good for* $a{\cdot}b$ as well as the special atomic predicate *a uses*. Their intended informal meaning is as follows. The predicate *a sees* is true of a message if a can extract the message from something a received. Analogously, *a said* is true of a message if a can extract the message from something a sent. A message *fresh* if it did not circulate until recently. A message satisfies *good for* $a \cdot b$ if every circulated message encrypted with this message as key was said by a or b. Finally, *a uses* a message if a uses that message as a key for decryption and encryption.

Proof rules The rules of BAN logic are summarized in Table 1. We use *Knows* to represent an arbitrary sequence of 0 or more epistemic modalities. Table 1 leaves some conditions implicit: We have omitted symmetric variations and closure under cut and weakening. Note that certain rules assume that agents do not misuse idealizations. For instance, rule *R1*, the *message meaning rule*, assumes that sender fields inside cipher texts are reliable. Also, rule *R7*, the *nonce verification rule*, assumes that agents only say statements known to be true while fresh.

While the original BAN paper ([4]) reads the epistemic modality as ”agent a believes that”, BAN logic is intuitively consistent with a knowledge interpretation. As in [8, 10], we adopt a knowledge interpretation and add the axiom T. The atomic BAN predicate $jurisdiction$ thereby becomes superfluous, and is therefore removed. For a more detailed discussion we refer the reader to [8]. Notice that we generalize the customary modal logic axiom T (*a knows F* $\vdash$ *F*) to arbitrary iterations of epistemic modalities, by adding *Knows* to antecedent and consequent.

[1] Only [1] was intended to validate BAN.

$R1.$ *a sees* $\{$*from b : M*$\}_{M'}$, *a knows* M' *good for* $a \cdot b \vdash a\,knows$ *b said* M
$R2.$ *a knows* M *fresh* $\vdash$ *a knows* M, M' *fresh*
$R3.$ *a knows* M *fresh*, *a knows* M' *good for* $a \cdot b \vdash$ *a knows* $\{M\}_{M'}$ *fresh*
$R4.$ *a sees* $M, M' \vdash$ *a sees* M
$R5.$ *a sees* $\{M\}_{M'}$, *a knows* M' *good for* $a \cdot b \vdash$ *a sees* M
$R6.$ *a knows b said* $M, M' \vdash$ *a knows b said* M
$R7.$ *a knows* $M_1, ..., F, ..., M_n$ *fresh*, *a knows b said* $M_1, ..., F, ..., M_n, \vdash$ *a knows b knows* F
$T.$ *Knows a knows* $F \vdash$ *Knows* F

Table 1: BAN proof rules

3 Semantics for the Non-Epistemic Language Fragment

In computer science, epistemic logics are customarily interpreted on *multi-agent systems* [6], pairs $\mathsf{S} = \langle S, | \rangle$, where S is a non-empty set of execution points and $|$ is a local state projection assigning a local state $s|a$ to each agent a and execution point s. Intuitively, the local state contains all the data currently accessible to that agent. For instance, when modeling a communication protocol, the local state of an agent might be derived from the initial condition plus the sequence of send and receive actions she has performed so far. A *multi-agent model* on S is a triple $\mathsf{M} = \langle S, |, I \rangle$, where I is an interpretation of atomic predicates. That is, to each atomic predicate p and each execution point $s \in S$, the interpretation I assigns the set $I(p, s)$ of messages (closed message terms) that satisfy p at s.

Closed statements are true w.r.t. an execution point s in a model M. The truth condition for atomic closed statements and negation (of closed statements) are as expected: $s \models_{\mathsf{M}} p(M) \Leftrightarrow M \in I(p, s)$ and $s \models_{\mathsf{M}} \neg F \Leftrightarrow s \not\models_{\mathsf{M}} F$. The truth condition for epistemic closed statements is left to section 4. Open statements are true w.r.t. an assignment V of messages to message variables, and an execution point s in a model M. Assignments are lifted to arbitrary message terms in the usual way; write $|M|_V$ for the value of M under V. The truth condition for open statements is: $V, s \models_{\mathsf{M}} F(M) \Leftrightarrow s \models_{\mathsf{M}} F(|M|_V)$.

If Δ is a set of statements, we write $V, s \models_{\mathsf{M}} \Delta$ if $V, s \models_{\mathsf{M}} F$, for all $F \in \Delta$. If C is a class of models: $\Delta \models_{\mathsf{C}} F$, if and only if, for all models M in C, for all execution points s in M and for all assignments V, if $V, s \models_{\mathsf{M}} \Delta$ then $V, s \models_{\mathsf{M}} F$.

4 Semantics for Knowledge

We interpret the epistemic modality through a generalized accessibility relation $\sim_a$ that relates not only execution points, but also messages at one execution point to messages at another. The intuition is that a cipher text M at the current execution point s may correspond, for a, to a different cipher text M' at an epistemically accessible execution point s'. That is, M at s could, for all a knows, be M' at s'. Let r be a *renaming* of messages, i.e. a function in the set of messages, defined for all messages. If r maps every message at s to a corresponding message at s', we say that r is a *counterpart mapping* between s and s' for agent a, and write $s \sim_a^r s'$. Given this ternary accessibility relation $\sim_a$, Kripke semantics can be generalized in an obvious way:

$$s \models_{\mathsf{M}} \textit{a knows F(M)} \Leftrightarrow \forall s' \in S : \forall r : s \sim_a^r s' \Rightarrow s' \models_{\mathsf{M}} F(r(M)) \; .$$

Here, $F(M)$ is any statement in the message term M. We do not assume that message M is somehow accessible to agent a in s, such as once said, or seen, by a. Agents may well know things about messages that are not accessible to them. In fact, this is an essential part of BAN logic (as witnessed by, for instance, axiom *R2*).

Counterpart mappings must be transparent to the set of available keys. A renaming r is *transparent* to a set Π of messages, in symbols $\Pi \rhd r$, if r respects all cryptographic structure accessible when using Π as keys: Π (used

C1. $M' \in \Pi \Rightarrow r(\{M\}_{M'}) = \{r(M)\}_{r(M')}$	C2. $r(M, M') = r(M), r(M')$
C3. r is injective	C4. r is surjective
C5. $r(F(M)) = F(r(M))$	C6. $r(\textit{from a : M}) = \textit{from r(a) : r(M)}$
C7. $r(k) = k$, k is agent name or message atom	

Table 2: Requirements for $\Pi \rhd r$

as keys) cannot distinguish a sequence $M_1, M_2, ...$ from $r(M_1), r(M_2), ...$ Formally, we stipulate that $\Pi \rhd r$, if and only if, each condition in Table 2 above is satisfied. Condition C1 says that encryption structure is plain, or clear, when the appropriate key is available, condition C2 says that pairing structure is always plain, conditions C3 and C4 say that distinct messages appear distinct, condition C5 says that atomic predicates and propositional operators are plain text, condition C6 says that sender field structure is plain, and condition C7, finally, says that agent names and message atoms are plain text.

Lemma 4.1

1. $\Pi \rhd \iota$ *where* ι *is the identity on messages*
2. $\Pi \rhd r,\ \ r(\Pi) \rhd r' \Rightarrow \Pi \rhd (r' \circ r)$
3. $\Pi \rhd r \Rightarrow r(\Pi) \rhd r^{-1}$
4. $\Pi \rhd r,\ \ \Pi \supseteq \Pi' \Rightarrow \Pi' \rhd r$

Proof. (1) and (4) are immediate. We prove (2) here. The proof of (3) is similar. Assume $\Pi \rhd r$ and $r(\Pi) \rhd r'$. Only requirement C1 of Table 2 is non-trivial. Assume $M' \in \Pi$. By the assumptions, $r(\{M\}_{M'}) = \{r(M)\}_{r(M')}$ and $r(M') \in r(\Pi)$. Thus, $r'(\{r(M)\}_{r(M')}) = \{r'(r(M))\}_{r'(r(M'))} = \{(r' \circ r)(M)\}_{(r' \circ r)(M')}$, i.e., $(r' \circ r)(\{M\}_{M'}) = r'(\{r(M)\}_{r(M')}) = \{r'(r(M))\}_{r'(r(M'))} = \{(r' \circ r)(M)\}_{(r' \circ r)(M')}$. □

Counterpart mappings must, furthermore, respect the current local state of the agent; we assume a renaming can be lifted pointwise to a permutation on local states. For r to be a counterpart mapping between s and some other point s', we require that r transforms the local state of the agent at s into her local state at s'.

The idea, then, is to relate the states s and s' under the renaming r for agent a, in symbols $s \sim_a^r s'$, just in case r transforms the local state of a at s into the local state of a at s' and r respects the keys used by the agent at s:

$$s \sim_a^r s' \Leftrightarrow r(s|a) = s'|a \text{ and } I(\textit{a uses}, s) \rhd r \ . \tag{3}$$

Each multi-agent model thus determines a unique ternary epistemic accessibility relation $\sim_a$. In section 9 below we address the apparent asymmetry of (3) and show that under the definitions of *uses* which we consider, whenever $s \sim_a^r s'$ then $s' \sim_a^{r^{-1}} s$.

5 Crypto Normality

The semantics avoids logical omniscience (the rule of normality). To see this, let $S = \{s\}$, $I(p, s) = \{\{M\}_{M'}\}$, $I(\textit{a uses}, s) = \emptyset$ and $s|a = \emptyset$. Then there is a renaming r such that $r(\{M\}_{M'}) \neq \{M\}_{M'}$ and $s \sim_a^r s$. Thus, $\not\models_{\mathsf{M}}$ *a knows* $p(\{M\}_{M'})$. Yet, $\models_{\mathsf{M}} p(\{M\}_{M'})$.

There is, however, an interesting weakening of normality which continues to hold. To formulate this, let *Keys*(M) be the set of message terms applied as keys in M such that *Keys*$(\{M\}_{M'}) = \{M'\} \cup$ *Keys*$(M) \cup$ *Keys*(M'), *Keys*$(M, M') =$ *Keys*$(M) \cup$ *Keys*(M'), *Keys*(*from a : M*) $=$ *Keys*(M), *Keys*$(P(M)) =$ *Keys*(M), *Keys*$(k) = \emptyset$, if k is message atom or agent name, and *Keys*$(x) = \emptyset$, for message variables x. For example, *Keys*$(\{w, \{x, k\}_y\}_z) = \{y, z\}$. Let *Keys*$(\Pi) = \cup_{M \in \Pi}$*Keys*$(M)$, write *a uses* Π for the set $\{$*a uses M* $\mid M \in \Pi\}$, and write *a knows* Δ for the set $\{$*a knows* $F \mid F \in \Delta\}$.

Lemma 5.1 $|\mathrm{Keys}(M)|_V \rhd r \Rightarrow r(|M|_V) = |M|_{r \circ V}$

Proof. By induction over the structure of M. The base step, where M is a variable, an agent name or message atom, is immediate from requirement C7 of Table 2. For the induction step, assume that the property holds for messages M_1 and M_2, i.e. $|$*Keys*$(M_1)|_V \rhd r \Rightarrow |M_1|_{r \circ V} = r(|M_1|_V)$ and $|$*Keys*$(M_2)|_V \rhd r \Rightarrow |M_2|_{r \circ V} = r(|M_2|_V)$. Assume $|$*Keys*$(\{M_1\}_{M_2})|_V \rhd r$. Then, $|$*Keys*$(M_1)|_V \cup |Keys(M_2)|_V \cup \{|M_2|_V\} \rhd r$. By the induction assumption and Lemma 4.1.4, $|M_1|_{r \circ V} = r(|M_1|_V)$ and $|M_2|_{r \circ V} = r(|M_2|_V)$. Then, by requirement C1 of Table 2, $r(|\{M_1\}_{M_2}|_V) = r(\{|M_1|_V\}_{|M_2|_V}) = \{r(|M_1|_V)\}_{r(|M_2|_V)} = \{|M_1|_{r \circ V}\}_{|M_2|_{r \circ V}} = |\{M_1\}_{M_2}|_{r \circ V}$. Showing that pairing and idealization constructions preserve the property is analogous. □

From Lemma 5.1 we get the weak normality rule.

Theorem 5.2 (Crypto Normality) *If* $\Delta \models_{\mathsf{M}} F$ *then a uses Keys*(Δ, F), *a knows* $\Delta \models_{\mathsf{M}}$ *a knows* F.

Crypto normality says that an agents knowledge is closed under logical validities in which all the keys applied are used by the agent. By itself, crypto normality may appear overly restricted, since all keys used in Δ or F must also be used by a. Crypto normality becomes more powerful, however, when combined with the rule of substitution.

Theorem 5.3 (Rule of Substitution) *Let* σ *be any substition of (possibly open) message terms for message variables. If* $\Delta \models_{\mathsf{M}} F$ *then* $\sigma(\Delta) \models_{\mathsf{M}} \sigma(F)$.

In conjunction, the two rules allow interesting inferences to be made, such as (2) in section 1.

6 Message Passing Systems

We instantiate models in message passing systems (cf. [6]), as in the BAN literature. Since the definitions are standard and well-known, we will only briefly hint at them. In a message passing system execution proceeds in rounds. During the first round, initial shared and private possessions are established. From then on, at each round, every agent either sends a message, receives a message or performs some unspecified internal action. By a *message passing model* we mean a multi-agent system $\mathsf{M} = \langle S, |, I\rangle$ based on a message passing system S. We require that the local state $s|a$ of an agent a consists of a first round of initializations followed by a's local history of send and receive actions. As an immediate consequence, agents know which messages they send and receive. Assume predicates *a received* and *a sent*, with *I(a received, s)* = $\{M \mid a$ has received *M* at $s\}$, and *I(a sent, s)* interpreted analogously. The following introspection principle is easily seen to be valid:

Proposition 6.1 (Receive and send introspection) *For message passing models:*

1. *a received* $M \models$ *a knows a received* M
2. *a sent* $M \models$ *a knows a sent* M

To see this, assume $s \sim_a^r s'$. Then, $r(s|a) = s'|a$, i.e. if a received M at s then a received $r(M)$ at s', and correspondingly for messages sent by a. While easily proved, Proposition 6.1 is nonetheless of some consequence. To begin with, the unsoundness of BAN logic in related Kripke semantics, such as [1, 12, 14], ultimately ties back to the failure of Proposition 6.1. When a Kripke semantics hides part of an agents local state from the agent, as these semantics do, we lose receiving and sending introspection: Say a received a cipher text M at s. Then there might be some point s' which is indistinguishable for a from the current point s, but where a received a different cipher text M', not M. Moreover, Proposition 6.1 in combination with crypto normality (Theorem 5.2) has some interesting, and perhaps surprising, implications for knowledge of cryptographic structure. We explore these implications in the section 7.

7 Knowledge of the Unseen

Prima facie it might be thought that an agents knowledge of cryptographic structure depends solely on what keys she uses. However, the mere finding of a cipher text at a certain place might alone indicate something about its contents. For instance, after the second protocol step in the Needham Shröder shared key protocol (NSSK) between principals a and b and with key server s, agent a knows the contents of the ticket she is to forward to b, despite the fact that she cannot decrypt it. The semantics respects such intuitions. To illustrate, assume that message passing model M implements NSSK between a, b and s. We may expect the following: *a received* $\{n, b, k, x\}_{k_a}$, k_a *good for* $a \cdot s \models_{\mathsf{M}} x$ *contains* k, a. (The meaning of *contains* should be clear from the context, while the precise semantics of *good* is not an issue in this example.) By crypto normality (Theorem 5.2) and universal substitution (Theorem 5.3), *a knows a received* $\{n, b, k, \{k, a\}_{k_b}\}_{k_a}$, *a knows* k_a *good for* $a \cdot s$, *a uses* $k_a \models_{\mathsf{M}}$ *a knows* $\{k, a\}_{k_b}$ *contains* k, a. By receiving introspection (Proposition 6.1), *a received* $\{n, b, k, \{k, a\}_{k_b}\}_{k_a}$, *a knows* k_a *good for* $a \cdot s$, *a uses* $k_a \models_{\mathsf{M}}$ *a knows* $\{k, a\}_{k_b}$ *contains* k, a. Thus, if k_a is a's server key and a receives $\{n, b, k, \{k, a\}_{k_b}\}_{k_a}$, then a knows the contents of $\{k, a\}_{k_b}$ even though a is not using k_b as a key.

The reason why the semantics supports deductions such as the above is that the set of counterpart mappings is limited not only by the current keys, but also by the current local state. Say renaming r is transparent to the keys used at the current point s, in symbols $I(\textit{a uses}, s) \rhd r$. This does not guarantee, however, that r is a counterpart mapping from s to any execution point s': There might be no s' in the given system such that $r(s|a) = s'|a$. In this case the agent can rule out r even though r is transparent to her current keys.

8 Interpreting BAN's Atomic Predicates

To complete the semantics for BAN logic, only the atomic predicates remain. This is a subject of subtle and somewhat bewildering variability (cf. [1, 8, 10]). We do not claim our definitions are canonical. Our goal is to show that the renaming semantics can be completed to a meaningful interpretation which validates BAN.

The way the predicates are explained informally in section 2, once the interpretation of *uses* is fixed, the interpretation of *sees*, *said* and *good* follow in a fairly straightforward fashion. Specifically, for *sees* we require that $I(\textit{a sees}, s)$ is the smallest set Π that includes a's initial possessions, the messages a has received at s and such that Π is closed under decryption with keys used ($\{M\}_{M'} \in \Pi$ and $M' \in I(\textit{a uses}, s) \Rightarrow M \in \Pi$) and un-pairing ($M, M' \in \Pi \Rightarrow M \in \Pi$ and $M' \in \Pi$) and sender-field removal (*from b:* $M \in \Pi \Rightarrow M \in \Pi$). The predicate *said* is defined analogously for sent messages, except that a's initial possessions are not included. For *good* we require that $M \in I(\textit{good for } a \cdot b, s)$, if and only if, whenever $\{M'\}_M$ is a sub term of some message in $I(\textit{c received}, s)$, then both M' and $\{M'\}_M$ are in $I(\textit{a said}, s)$ or both M' and $\{M'\}_M$ are in $I(\textit{b said}, s)$, for any agent c and any message M'. We leave the interpretation of the predicate *fresh* open, merely requiring that it is independent of the interpretation of *uses* and that it is closed under the sub-term relation ($M \in I(\textit{fresh}, s) \Rightarrow M, M' \in I(\textit{fresh}, s)$, $M', M \in I(\textit{fresh}, s)$, $\{M\}_{M'} \in I(\textit{fresh}, s)$, and $\{M'\}_M \in I(\textit{fresh}, s)$). One could satisfy these requirements by defining, similarly to [8], a message as fresh if it is not a subterm of any message said by anyone more than d rounds back, for some fixed d. Interpreting the predicate *fresh* is somewhat problematic, but it is peripheral to the issues addressed in this paper. We refer the reader to [8, 10] for more detailed discussions.

We then turn to the predicate *uses*. An immediate observation is that the interpretation of *uses* must validate the entailment

$$\text{a } \textit{knows } M \textit{ good for } a \cdot b \models a \textit{ uses } M \ . \tag{4}$$

This requirement is fundamental, since otherwise rules *R1*, *R3*, and *R5* (Table 1) will not be validated.

A possible approach to the definition of *uses* is to view *uses* and *sees* as synonyms, so that a key is used by an agent just in case it is possessed initially or it is received, or it can be obtained by decryption and un-pairing from used messages. This kind of "operational" view is taken, with variations, in most papers on semantics for BAN like logics. The problem with this definition is that it does not validate (4), unless the class of message passing systems is restricted in some way. For instance a model M may satisfy an entailment such as: a *receives* $k, a, b \models_{\mathsf{M}} k, x$ *good for* $a \cdot b$. Then, by crypto normality (Theorem 5.2) and receive introspection (prop. 6.1.1), a *receives* $k, a, b \models_{\mathsf{M}} a$ *knows* k, x *good for* $a \cdot b$, but it might well be that a has not seen k, x, contradicting (4). This counterexample can be fixed, of course, by disallowing complex terms as keys. But other, similar counterexamples would still require restricting the class of allowed message passing systems. For instance, if we allowed a model specific dependency between properties of different message atoms, say a *receives* $k, a, b \models_{\mathsf{M}} k'$ *good for* $a \cdot b$, then a might be able to conclude that k' is good without actually seeing it, again contradicting (4).

We propose an alternative definition of *uses* which we believe is of independent interest. The idea is to consider a key to be used by an agent just in case the agent knows some property p of that key. Since properties (sees, said, etc.) are defined by means of *uses* itself, a recursive definition is called for. An inductive, rather than a coinductive, definition seems appropriate, since *a uses* should contain the set of keys that a has gathered some positive information about. Adopting this approach we thus define the interpretation function on a message passing system S as a least interpretation function I (in an order extended point wise from set containment) such that $s \models_{\langle S,I \rangle}$ *a uses M*, if and only if, $s \models_{\langle S,I \rangle}$ *knows* $p(M)$ for some atomic BAN predicate p. (We leave the local state projection | implicit.) If we call models that use this definition of the interpretation function *inductive*, we obtain:

Theorem 8.1 *Every message passing system determines a unique inductive model.*

Proof. Assume a message passing system S. The interpretation function in an inductive model on S is, by definition, the least fixed point of the following function f that assigns an interpretation function $f(I)$ to every possible interpretation function I on S. For predicate *uses*, $f(I)(a\ uses, s) = \{M \mid\ \exists$ atomic BAN predicate p : $s \models_{\langle S,I \rangle} a$ *knows* $p(M)\}$ and, for atomic BAN predicates p, $f(I)(p, s)$ is defined with $f(I)(a\ uses, s')$ as the keys used for any agent a at any point s'. From lemma 4.1.4, f is monotone. Therefore, f has a least fixed point. □

Inductive models obviously satisfy the requirement (4) above. In fact, as far as requirement (4) is concerned, we could have defined *uses* in terms of predicates *good* alone, so that $s \models$ *a uses M*, if and only if, $s \models$ *a knows M good for* $a \cdot b$ for some agent b. Perhaps, such a solution would be even more faithful to intuitions in BAN logic, but it would not be quite satisfactory for some protocols (Yahalom is an example) where keys need to be used before they are known to be good. Inductive models offer, in our opinion, an interesting, more extensional, alternative to the more traditional operational models.

9 Introspection Properties

We have already seen (Proposition 6.1) that agents in message passing models are introspective of their received and sent messages. In this section, we observe some further introspection properties in inductive models. We emphasize that these results also hold for models based on an operational interpretation of *uses*.

Lemma 9.1 *For inductive models M:*

1. $s \sim_a^{\iota} s$
2. $s \sim_a^{r} s',\ \ s' \sim_a^{r'} s'' \Rightarrow s \sim_a^{r' \circ r} s''$
3. $s \sim_a^{r} s' \Rightarrow s' \sim_a^{r^{-1}} s$

Proof. (1) Immediate from Lemma 4.1.1. The proof of (2) is similar to (3) and left out. For (3) we first prove, using fixed point induction, that $s \sim_a^r s' \Rightarrow I(a\ \textit{uses}, s') \subseteq r(I(a\ \textit{uses}, s))$ where I is the interpretation function in M. Let I_j be the interpretation function at step j in the fixed point construction of the proof of Theorem 8.1, such that $I_0 = \emptyset$, $I_{j+1} = f(I_j)$, and $I_\delta = \cup_{j<\delta} I_j$, if δ is a limit ordinal. Let M_{I_j} be M with the interpretation I replaced by I_j. We show for all j that

$$I_j(a\ \textit{uses}, s) \rhd r \wedge r(s|a) = s'|a \Rightarrow I_j(a\ \textit{uses}, s') \subseteq r(I_j(a\ \textit{uses}, s)) \tag{5}$$

The property holds for I_0, since $I_0(a\ \textit{uses}, s') = \emptyset$. For successor ordinals, assume (5) holds for j. Assume $I_{j+1}(a\ \textit{uses}, s) \rhd r$ and $r(s|a) = s'|a$. Pick any message M' such that $M' \in I_{j+1}(a\ \textit{uses}, s')$. By C4 in Table 2, $M' = r(M)$ for some message M. Then $s' \models_{\mathsf{M}_{I_{j+1}}} a\ \textit{uses}\ (r(M))$. By the definition of I_{j+1}, there is an atomic predicate p such that $s' \models_{\mathsf{M}_{I_j}} a\ \textit{knows}\ p(r(M))$. Since $I_j \subseteq I_{j+1}$, by Lemma 4.1.4, $I_j(a\ \textit{uses}, s) \rhd r$. By Lemma 4.1.3, $r(I_j(a\ \textit{uses}, s)) \rhd r^{-1}$, so by the induction hypothesis and Lemma 4.1.4, $I_j(a\ \textit{uses}, s') \rhd r^{-1}$. Since $M' = r(M)$, we want to show that $M \in I_{j+1}(a\ \textit{uses}, s)$. By definition of I_{j+1}, it suffices to show that $s \models_{\mathsf{M}_{I_j}} a\ \textit{knows}\ p(M)$. So pick any renaming r' and any execution point $s'' \in S$ such that $I_j(a\ \textit{uses}, s) \rhd r'$ and $r'(s|a) = s''|a$. Since $I_j(a\ \textit{uses}, s) \rhd r$, by the induction hypothesis, and conditions C3 and C4 of Table 2, $r^{-1}(I_j(a\ \textit{uses}, s')) \subseteq I_j(a\ \textit{uses}, s)$. By Lemma 4.1.4, $r^{-1}(I_j(a\ \textit{uses}, s')) \rhd r'$. By Lemma 4.1.2 it follows that $I_j(a\ \textit{uses}, s') \rhd r' \circ r^{-1}$. By the assumptions on r' we get that $r' \circ r^{-1}(s'|a) = r'(r^{-1}(s'|a)) = r'(s|a) = s''|a$. Since we showed $s' \models_{\mathsf{M}_{I_j}} a\ \textit{knows}\ p(r(M))$ we obtain that $s'' \models_{\mathsf{M}_{I_j}} p(r' \circ r^{-1} \circ r(M))$. Since r' and s'' are arbitrary, it follows that $s \models_{\mathsf{M}_{I_j}} a\ \textit{knows}\ p(M)$ which completes the successor part of the induction argument. The limit case is routine.

For the proof of the main statement (3), assume then that $s \sim_a^r s'$, i.e. $I(a\ \textit{uses}, s) \rhd r$ and $r(s|a) = s'|a$. By Lemma 4.1.3, $r(I(a\ \textit{uses}, s)) \rhd r^{-1}$. We also obtain, from the above induction, that $I(a\ \textit{uses}, s') \subseteq r(I(a\ \textit{uses}, s))$. By Lemma 4.1.4, $I(a\ \textit{uses}, s') \rhd r^{-1}$, so $s' \sim_a^{r^{-1}} s$, which completes the proof. □

Using Lemma 9.1 the modal logic S5 properties follow directly.

Theorem 9.2 (Knowledge introspection) *For inductive models:*

1. *a knows* $F \models F$
2. *a knows* $F \models$ *a knows a knows* F
3. $\neg$ *a knows* $F \models$ *a knows* $\neg$ *a knows* F

Validity (1) in Theorem 9.2 is, of course, not an introspection property. Rather, it can be seen as the distinguishing line between knowledge and belief. In fact, (1) holds in all models, not only inductive models. From Theorem 9.2, it follows that agents are also introspective of used and seen messages:

Corollary 9.3 (Use and sees introspection) *For inductive models:*

1. *a uses* $M \models$ *a knows a uses* M
2. *a sees* $M \models$ *a knows a sees* M

Proof. (1) is immediate from Theorem 9.2. (2) follows from crypto normality (Theorem 5.2), rule of substitution (Theorem 5.3), receive introspection (Proposition 6.1.1), and use introspection (1). □

Loosely speaking, sees introspection implies that agents are introspective of extracted messages. Since sees introspection depends on receive introspection (Proposition 6.1) it fails in the related Kripke semantics of [1, 12, 14]. For similar reasons (see section 6), use introspection also fails in these semantics, when cipher texts are allowed as keys.

10 Soundness of BAN logic

As observed in section 2, some BAN rules assume that agents do not misuse idealizations. Accordingly, in our soundness result we restrict attention to *honest* models, models where *from b:* $M \in I($*a said* $, s) \Rightarrow a = b$ and where $M_1, ..., F, ..., M_n$ *fresh* , *a said* $M_1, ..., F, ..., M_n \models$ *a knows* F. Again, we refer the reader to [8] for details.

Soundness for each BAN rule (Table 1) is now a rather immediate application of the following corollary, where *a knows* $\{M_1, ..., M_n\}$ *good* is short for *a knows* M_1 *good for* $a \cdot b_1$, ..., *a knows* M_n *good for* $a \cdot b_n$.

Corollary 10.1 *Let σ be any substitution of message terms for variables. For inductive models M: If $\Delta \models_M F$ then a knows $\sigma($Keys$(\Delta, F))$ good, a knows $\sigma(\Delta) \models_M$ a knows $\sigma(F)$.*

Proof. Immediate from crypto normality (Theorem 5.2), rule of substitution (Theorem 5.3) and requirement (4) in section 8. □

Theorem 10.2 *BAN logic is sound w.r.t. honest inductive models.*

Proof. Rule R4 (Table 1) is immediate. Rule R5 is immediate from requirement (4) in section 8. Each remaining rule is a direct application of Corollary 10.1 on some trivial validity. For instance, rule R3 follows from the fact that y *fresh* $\models \{y\}_z$ *fresh*. Rule R1 needs, in addition, sees introspection (Corollary 9.3.2), while rule *T* needs Theorem 9.2.1. □

11 Related Work

Our use of a ternary accessibility relation is most closely related to possibility relations in counterpart semantics [9]. It is, as far as we know, the first computationally grounded such semantics in epistemic logic.

In the BAN logic literature the semantics most closely related to ours are the Kripke semantics of [1, 12, 14] where the local state of an agent is partly hidden from the agent. In our framework we can recover a binary accessibility relation similar to those used in [1, 12, 14] by letting $s \sim_a s'$ iff $s \sim_a^r s'$ for some renaming r. In fact, our notion of transparent renaming can be seen as related to the message congruences of [1], and to the states of knowledge and belief of [3, 13]. As we have pointed out, however, a Kripke semantics resulting from such a binary accessibility relation $\sim_a$ is both too strong and too weak for BAN: It makes agents logically omniscient, yet fails essential introspection principles [2].

There are, of course, semantics in the literature that do in fact avoid logical omniscience (cf. [6]). But no such semantics has been shown to work for BAN-like logics. Furthermore, these semantics tend to break rather more radically than ours with Kripke semantics. One possible approach is to subdivide knowledge into an implicit and an explicit part. Implicit knowledge would be "ideal" knowledge to which logical omniscience applies, and explicit knowledge would be somehow circumscribed to reflect agents limited reasoning abilities. For instance, [7] specifies adversary capabilities in terms of abstract knowledge extraction algorithms, and [2] uses an awareness predicate to constrain, at each state, the predicates which of which an agent is aware, related to the comprehended messages of [12].

12 Conclusion

We have introduced a semantics that validates BAN logic, yet avoids the rule of normality (logical omniscience). The semantics satisfies crypto normality, a weak version of normality that filters out infeasible cryptographic reasoning powers. The semantics makes agents introspective of their own knowledge and their own actions of

[2] But we acknowledge that only [1] was intended as a semantics for BAN.

sending, receiving and extracting. We have showed how knowledge of cryptographic structure may at times transcend the discriminatory power of the keys used. Finally, we found that knowledge and keys used could be defined as simultaneous fixed points, making the keys used equal to the keys known.

A semantical foundation for BAN logic opens up the possibility of sound model checking of BAN logic specifications. Also, the semantics might be used to improve various elements of the protocol verification process in BAN. The crypto normality rule is a case in point. Using this rule we can sidestep the often criticized ”idealization step” in BAN verifications. To illustrate, say we want to establish the following property of NSSK:

$$a \textit{ knows } k_a \textit{ good for } a \cdot s,\ a \textit{ knows n fresh},\ a \textit{ sees } \{n, b, k, \{k, a\}_{k_b}\}_{k_a} \models a \textit{ knows } k \textit{ good for } a \cdot b \tag{6}$$

As BAN is usually applied, one would instead prove a property of an ”idealization” of the protocol where the message $\{n, b, k, \{k, a\}_{k_b}\}_{k_a}$ has been annotated with sender field and the goodness predicate. As an alternative, we introduce non-epistemic protocol specific validities:

$$k_a \textit{ good for } a \cdot s,\ \textit{n fresh},\ \textit{s said } \{n, b, k, x\}_{k_a} \models k \textit{ good for } a \cdot b \tag{7}$$

$$k_a \textit{ good for } a \cdot s \models \neg\ \textit{a said } \{n, b, k, x\}_{k_a} \tag{8}$$

which arguably express the required properties of the protocol rather more precisely. Starting from a (protocol independent) triviality,

$$\neg\ \textit{a said } \{x\}_y,\ \textit{a sees } \{x\}_y,\ y \textit{ good for } a \cdot s \models \textit{s said } \{x\}_y, \tag{9}$$

we get specification (6) by lifting (7), (8) and (9) to epistemic validities using crypto normality (Corollary 10.1), then applying sees introspection (Corollary 9.3) and knowledge introspection (Theorem 9.2)

We have focused on BAN logic, not in particular deference to BAN, but simply because BAN is the standard logic in its family. A first question to answer is whether our semantics really captures the intended meaning of BAN formulas. A completeness result for a collection of rules which stays acceptably close to BAN's original set-up would help answer this question affirmatively, and we are currently working to address this issue.

It would be of interest also to use our semantics to support epistemic security protocol logics beyond the propositional level. An extension to first-order μ-calculus with rudimentary temporal operators would allow the BAN primitives to be defined, and thus eliminate much of the apparent arbitrariness in the choice of basic vocabulary in the BAN literature. Furthermore, a first-order extension would allow reasoning that exploits partial knowledge of complex data structures; this may be useful in the context of e.g. payment protocols, where different parts of the negotiated data structure remain hidden from different principals.

References

[1] Martín Abadi and Mark Tuttle. A semantics for a logic of authentication. In *Proceedings of the Tenth Annual ACM Symposium on Principles of Distributed Computing*, pages 201–216. ACM Press, August 1991.

[2] Rafael Accorsi, David A. Basin, and Luca Viganò. Towards an awareness-based semantics for security protocol analysis. *Electr. Notes Theor. Comput. Sci.*, 55(1), 2001.

[3] Pierre Bieber. A logic of communication in hostile environments. In *Proceedings of the Computer Security Foundation Workshop III*, pages 14–22. IEEE Computer Society Press, 1990.

[4] Michael Burrows, Martín Abadi, and Roger M. Needham. A logic of authentication. *ACM Trans. Comput. Syst.*, 8(1):18–36, 1990.

[5] Anthony H. Dekker. C3po: A tool for automatic sound cryptographic protocol analysis. In *PCSFW: Proceedings of The 13th Computer Security Foundations Workshop*, pages 77–87. IEEE Computer Society Press, 2000.

[6] Ronald Fagin, Joseph Y. Halpern, Yoram Moses, and Moshe Y. Vardi. *Reasoning About Knowledge*. MIT Press, 1995.

[7] Joseph Y. Halpern and Riccardo Pucella. Modeling adversaries in a logic for security protocol analysis. In *FASec*, pages 115–132, 2002.

[8] Joseph Y. Halpern, Riccardo Pucella, and Ron van der Meyden. Revisiting the foundations of authentication logics. Manuscript, 2003.

[9] David Lewis. Counterpart theory and quantified modal logic. *Journal of Philosophy*, 65:113–126, 1968.

[10] Paul F. Syverson. Towards a strand semantics for authentication logics. In *Electronic Notes in Theoretical Computer Science*, 20,2000.

[11] Paul F. Syverson and Paul C. van Oorschot. On unifying some cryptographic protocol logics. In *Proc. IEEE Symposium on Research in Security and Privacy*, pages 14–28. IEEE CS Press, May 1994.

[12] Paul F. Syverson and Paul C. van Oorschot. A unified cryptographic protocol logic. NRL Publication 5540-227, Naval Research Lab, 1996.

[13] M.-J. Toussaint and P. Wolper. Reasoning about cryptographic protocols. In J. Feigenbaum and M. Merritt, editors, *Distributed Computing and Cryptography*, volume 2 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pages 245–262. American Mathematical Society, 1989.

[14] Gabriele Wedel and Volker Kessler. Formal semantics for authentication logics. In E. Bertino, H. Kurth, and G. Martella, editors, *ESORICS'96*, LNCS 1146, pages 219–241. Springer-Verlag, 1996.