Auton Agent Multi-Agent Syst (2009) 19:76–88
DOI 10.1007/s10458-008-9063-8

On BAN logic and hash functions or: how an unjustified inference rule causes problems

Wouter Teepe

Published online: 17 September 2008

Abstract BAN logic, an epistemic logic for analyzing security protocols, contains an unjustifiable inference rule. The inference rule assumes that possession of $H(X)$ (i.e., the cryptographic hash value of X) counts as a proof of possession of X, which is not the case. As a result, BAN logic exhibits a problematic property, which is similar to unsoundness, but not strictly equivalent to it. We will call this property 'unsoundness' (with quotes). The property is demonstrated using a specially crafted protocol, the *two parrots protocol*. The 'unsoundness' is proven using the partial semantics which is given for BAN logic. Because of the questionable character of the semantics of BAN logic, we also provide an alternative proof of 'unsoundness' which we consider more important.

Keywords BAN logic · Soundness · Two parrots protocol · Cryptographic hash function · Security protocol

1 Introduction

Formal analysis of security protocols requires one to reason about the knowledge of the participants (principals) in the protocol. Critical for a security protocol is that it should not only guarantee that certain information is communicated, but also that certain other information is *not* communicated. For example, external observers should typically not be able to infer session keys which are exchanged in a security protocol.

BAN logic [5, 7],[1] introduced by Burrows, Abadi and Needham, is an epistemic logic[2] crafted for analyzing security protocols. It models at an abstract level the knowledge of the principals in a protocol. The principals are supposed to have only polynomially many

[1] See Appendix A for a taxonomy of the papers presenting the BAN logic.
[2] For a thorough treatment of epistemic logic, consult [16, 25].

W. Teepe (✉)
Department of Computer Science, Digital Security group, Radboud University Nijmegen,
Postbus 9010, 6500 GL Nijmegen, The Netherlands
e-mail: w.teepe@cs.ru.nl

computational resources. It was the first logic of its kind, and has had a tremendous influence on protocol analysis: it has helped revealing weaknesses in known protocols, and many logics are based on it. This is not to say that there has been no criticism of BAN logic. For one thing, a *full* semantics is lacking, and many attempts have been made to fix this problem [2, 17, 21, 35, 27, 32, 33, 13]. Moreover, the logic fails to detect some very obvious protocol flaws [26].

Though the semantics of BAN logic is generally considered unclear, it is for our purposes important to note that BAN logic *does* have a partial semantics, which is defined over a part of the formal language of BAN logic.[3]

The general consensus about BAN-descendant logics appears to be that these logics are computationally sound (detected protocol flaws are indeed flaws), but certainly not computationally complete (they may fail to detect certain protocol flaws). Recent work includes attempts to bridge the gap between the formal (i.e., BAN-descendant) approach and the computational approach to security logics [1], and attempts to obtain completeness results for BAN-descendant logics in a kind of Kripke-style semantics [10–12]. In the Multi-Agent Systems world, BAN logic has been widely used (see for example, [3]).

In this paper we show a problem of BAN logic [5, 7] that has, to our knowledge, not yet been identified, despite all research into formal protocol analysis. The problem is this: BAN logic is not 'sound'. That is, false statements can be obtained by the application of existing inference rules from true assumptions. In Sect. 2 we will elaborate on the concept of 'unsoundness'. The problem is caused by one particular questionable inference rule. In Sect. 2 we will also explain the reasoning mistake behind this questionable inference rule. As a result of the reasoning mistake, the inference rule does not have a *computational justification*, which is discussed in Sect. 3. Section 4 shows the protocol we use in our unsoundness proof and Sect. 5 shows all inference rules used in our proof. Section 6 shows the actual proof. In Sect. 7 we will give an alternative proof, but in the questionable semantics of BAN logic; therefore, we regard our proof of Sect. 6 more important. We close with some remarks on the relevance of our results.

2 Cryptographic hash functions and justified beliefs

A cryptographic hash function is a function $H: \{0,1\}^* \rightarrow \{0,1\}^k$ which is computationally feasible to compute, but for which the inverse is computationally infeasible. In particular, computing the inverse of a hash function takes $O(2^k)$ operations. Thus, a cryptographic hash function is *one-way*: it is computationally infeasible to construct a message x such that $H(x)$ yields a given value h [14].[4]

Cryptographic hash functions have a lot of applications, including password protection, manipulation detection and the optimization of digital signature schemes. Unfortunately however, the class of applications is sometimes overestimated. Consider for example the following quote from security expert Schneier [30, p. 31]:

> If you want to verify someone has a particular file (that you also have), but you don't want him to send it to you, then you ask him for the hash value. If he sends you the correct hash value, then it is almost certain that he has that file.

[3] This partial semantics is defined in Sect. 13 of the original BAN papers [5, 7].

[4] For an extensive treatment of cryptographic hash functions, consult [28].

Unfortunately, this claim is false. The problem is that in the above situation sketch, there is no mention that the hash value of the file should be kept totally secret. If there is somebody who is willing to publish the hash value of the file, anybody can 'prove' possession of the file.

The authors of BAN logic [5, 7] made the same reasoning mistake as Bruce Schneier, and incorporated into their logic an inference rule reflecting the above mentioned questionable reasoning.[5] The name of the questionable rule is **H** and the rule will be shown in Sect. 3. As a result of this, BAN logic is not 'sound'. Essential in our proof is the fact that belief in BAN logic is considered to be *justified belief*.

But first, let us recapitulate what soundness is. A proof procedure is sound if it proves only valid formulae. In particular, from justified ('true') formulae it should be impossible to infer an unjustifiable ('false') formula. A proof of soundness generally involves a formal system and a class of models (a *semantics*): a proof of soundness essentially shows that every formula that is *derivable* ($\vdash$) in the formal system is *observable* ($\models$) in all relevant models of the semantics (i.e., $S \vdash X$ implies $s \models X$ for all models s).

A related concept, 'soundness'[6] ($S \vdash P \equiv X$ implies $S \vdash X$) relies on the definition of the modal operator *belief* ($\equiv$) in BAN logic which denotes *true justified belief*. As opposed to beliefs in general, a true justified belief should be true. To see what the authors of BAN logic consider belief, let us look at the following excerpt from [9, page 7]:

> More precisely, define knowledge as truth in all states (as in [18][7]); our notion of belief is a rudimentary approximation to knowledge, and it is simple to see that if all initial beliefs are knowledge then all final beliefs are knowledge and, in particular, they are true.

In this paper, we will first prove 'unsoundness' in Sect. 6 and then unsoundness (without quotes) in Sect. 7. In our 'unsoundness' proof, all initial beliefs are clearly knowledge, though one of the obtained final beliefs is not knowledge, in particular, it is false. Thus, by inferring an unjustified belief in BAN logic from true assumptions, we prove that BAN logic is not 'sound'. In particular, this means that it is impossible to create a semantics in which BAN logic is sound.

We will adhere to the convention of BAN logic to talk about the *beliefs* of the principals as if they are not necessarily also *knowledge*. Which is appropriate, as the initial beliefs may fail to be knowledge. However the claim that true initial beliefs and BAN inferences will result in nothing but true beliefs, turns out to be unwarranted. The culprit is an unwarranted BAN inference rule, which we discuss in the next section.

3 On the computational justification of beliefs

In the analysis of security protocols, if a principal obtains a new belief, there has to be a computational justification for the newly obtained belief. For example, if a principal sees a message cryptographically signed with private key K^{-1}, it is justified to believe that the message originates from the principal owning private key K^{-1}. The computational justification is in this case that it is computationally infeasible for principals other than the one owning

[5] See Appendix A for a detailed discussion of the papers presenting BAN logic, and which papers exactly contain the reasoning mistake.

[6] Note the quotes, which distinguish 'soundness' from soundness.

[7] This is a reference to a preliminary paper. The final paper is [19]—WT.

private key K^{-1} to construct a message signed with this key. This type of justification is *essential* if security is of concern.[8]

With this consideration in mind, it is worth noting the following excerpt from page 266 of the BAN paper [5], (resp. pages 41–42 of [7]):

> Obviously, trust in protocols that use hash functions is not always warranted. If H is an arbitrary function, nothing convinces one that when A has uttered $H(m)$ he must have also uttered m. In fact, A may never have seen m. This may happen, for instance, if the author of m gave $H(m)$ to A, who signed it and sent it. This is similar to the way in which a manager signs a document presented by a subordinate without reading the details of the document. However, the manager expects anyone receiving this signed document to behave as though the manager had full knowledge of the contents. Thus, provided the manager is not careless and the hash function is suitable, signing a hash value should be considered the same as signing the entire message.

This quote contains an assumption which is, in our opinion, unreasonable: the manager expects anyone receiving the signed document to make an unjustified inference, based on the assumption that "the manager is not careless". The unjustified inference is to infer a statement which is not warranted. Of course, any principal including the manager may be free to desire any behavior from other principals. But is it reasonable to expect beliefs to be obtained which are not computationally justified?

It is reasonable to assume that any principal, upon seeing $\{H(N)\}_{K^{-1}}$ will believe that the manager has seen and signed $H(N)$, since it is computationally too difficult for any principal *other than the manager* to construct the signature. However, it is not reasonable to assume that any principal, upon believing that a manager has seen and signed $H(N)$, believes that the manager has seen N, as there is *no computational problem* that would justify such a belief. Anybody may have computed $H(N)$ from N, in particular someone may have told the manager $H(N)$ but not N. Therefore, the expectation of a manager that other principals should act as if the manager knows N, is not warranted.

In fact, the text quoted above is the justification of the inference rule **H** in BAN logic.[9] We believe the identified problematic assumption explains the problems that arise from the inference rule **H**.

The *hashing* inference rule **H** reads, as given on page 266 of [5] (resp. page 42 of [7]):

$$\mathbf{H} \qquad \frac{P \models Q \sim H(X),\ P \lhd X}{P \models Q \sim X}$$

This rule formalizes that if P believes that Q once conveyed $H(X)$, and P receives a message X, then P may infer that Q knows X.

The message X that P receives, need not be sent by Q. Therefore, this rule is problematic, as it essentially infers belief (by P) of 'possession' (by Q) of the message X from P believing that Q once conveyed $H(X)$. This rule leads to the 'unsoundness' of BAN logic. Fortunately, none of the authentication logics that descend from BAN logic, adopts the **H** inference rule.

Because the most commonly used signature schemes use cryptographic hash functions, the **H** inference rule was added to BAN logic to facilitate the analysis of such signature schemes.

[8] Consider the alternative: we do not want principals to believe a message is sent by Santa Claus just because the name 'Santa Claus' is written beneath it; writing the name 'Santa Claus' is an exercise just as easy for Santa Claus himself as it is for anybody else.

[9] The name of this inference rule has been given by the writer of this text.

With this inference rule at hand, we can see how the two parrots protocol demonstrates the 'unsoundness' of BAN logic.

4 The two parrots protocol

To prove the 'unsoundness' of BAN logic, we rely on a protocol. The rather simple *two parrots protocol*, shown in Fig. 1, will demonstrate the 'unsoundness'. Alice (denoted A) chooses a random number N, sends it to Cecil (denoted C), who returns the number. Then Alice sends the cryptographic hash of the number to Bob (denoted B), and Bob signs this hash value and returns it to Alice. As Bob only sees the cryptographic hash value of N, and a cryptographic hash function is one-way, Bob does not learn N itself. Of course, Cecil might privately disclose N to Bob, but this does not happen in the two parrots protocol. Thus, though by private channels Bob might learn N, the protocol certainly does not guarantee this.

After Alice receives Bob's message, she knows that Bob received her message containing the hash value of the random number. However, Alice cannot, as a result of the protocol, conclude that Bob knows N. Neither can Alice conclude that Bob *does not* know N. Unfortunately, according to the analysis of the two parrots protocol in BAN logic, Alice will believe that Bob knows N.

In the two parrots protocol, the message N is transmitted without protection. Thus, one can argue that Bob could learn N by mere eavesdropping. For the sake of simplicity, we use a *very simple protocol* that suffices to demonstrate our observation on BAN logic. Of course, protection of N can be achieved by encryption of the messages between Alice and Cecil. Our proof can be easily extended to obtain the same result for such an altered protocol. Moreover, our proof does not rely on Bob eavesdropping.

Thus, though Bob could learn N through either an assistant (Cecil disclosing N to Bob) or through eavesdropping, the communication in the two parrots protocol simply does not guarantee that Bob knows N, and therefore also does not warrant Alice believing that Bob knows N.

When we want to formally analyze the protocol in BAN logic, we need to *transcribe* it into BAN logic. First, we have the protocol assumptions

$$A \models \overset{K}{\mapsto} B, \quad A \models N, \quad A \models \sharp(N)$$

which state that A knows the public key K of B, A knows N, and A believes N to be *fresh*. A newly generated random number is particularly fresh. Then, we have the protocol itself:

$$\begin{array}{lll} \text{step 1} & S_1: & A \to C : N \\ \text{step 2} & S_2: & C \to A : N \\ \text{step 3} & S_3: & A \to B : H(N) \\ \text{step 4} & S_4: & B \to A : \{H(N)\}_{K^{-1}} \end{array}$$

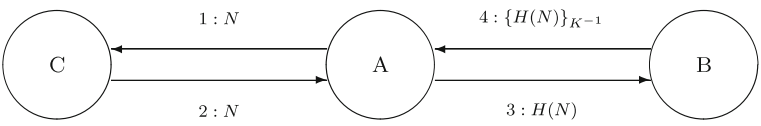

Fig. 1 The two parrots protocol

This protocol description is rather straightforward. In general, the message X cryptographically signed with the private key corresponding to public key K is denoted as $\{X\}_{K^{-1}}$. Thus, any agent that knows K can verify the signature and read X. In particular, in the two parrots protocol Alice can verify that message 4 was signed and sent by Bob.

To quickly see how the two parrots protocol interacts with the **H** inference rule, observe that message 2 ($C \rightarrow A: N$) can be used to obtain the second precondition of **H**, and that message 4 ($B \rightarrow A: \{H(N)\}_{K^{-1}}$) can be used to obtain the first precondition of **H**. Thus, messages 2 and 4 are the essential messages of the protocol. The other messages can be considered mere 'glue'.

What is achieved by a protocol can be stated in *claims*. For the two parrots protocol, the following claim is true:

$$\text{It will not be the case that } B \models N$$

which essentially states that B will not know N. Note that this is true because

1. B only sees $H(N)$,
2. the inverse of $H(\cdot)$ is hard to compute ($H(\cdot)$ is a one-way function), and
3. B has only polynomially many computational resources.

The problem that we identify in BAN logic (see Sect. 6) has the effect that the following statement can also be inferred in BAN logic:

$$A \models B \models N$$

which states that A will believe that B will know N. This belief of A is not computationally justified (see Sect. 3).

5 Used inference rules

The proof of 'unsoundness' in Sect. 6 involves three inference rules of BAN logic.[10] Inference rule **H** has already been given in Sect. 3, the other two rules are:

1. the *message meaning* inference rule number *ii* as given on page 238 of [5] (resp. page 6 of [7]):

$$\textbf{MM} \qquad \frac{P \models \overset{K}{\mapsto} Q, \quad P \lhd \{X\}_{K^{-1}}}{P \models Q \mid\sim X}$$

   This rule formalizes that if P knows Q's public key, and P receives a message X signed with Q's private key, P may infer that Q once sent X.[11]
2. the *nonce-verification* inference rule as given on page 238 of [5] (resp. page 6 of [7]):

$$\textbf{NV} \qquad \frac{P \models \sharp(X), \quad P \models Q \mid\sim X}{P \models Q \models X}$$

   This rule formalizes that if P believes X to be *fresh* (it originates in the current session), and P believes that Q once conveyed X, then P may infer that Q believes X (in the current session).[12]

[10] These names of these inference rules have been given by the writer of this text.

[11] Inference rule **MM** has been questioned by Wedel and Kessler, as it is invalid if interpreted according to their semantics [35]. However, they point out that it is unclear whether BAN logic itself or their semantics of BAN logic is to blame for that.

[12] This rule relies on the assumption that only beliefs are communicated.

6 Proof of 'unsoundness' of BAN logic

In this section, we will present our formal proof. In our proof, we use the term 'unjustified belief'. This might be perceived as unnecessarily harsh or misleading, but we will argue that this is the right formulation, even in lack of a clear semantics of BAN logic as a whole. The central construct of BAN logic, $\models$, is defined as follows on page 236 of [5] (resp. page 4 of [7]):

> $P \models X$: *P believes X*, or *P* would be entitled to believe *X*. In particular, the principal *P* may act as though *X* is true. This construct is central to the logic.

In our proof, we obtain a result of the form $P \models X$, where X is *not warranted*. It *might* be the case that X were true, if some more communication were to occur than prescribed by the two parrots protocol and considered in our proof. Therefore, and in this way, we deem "unjustified belief" the appropriate term for such an X. With this explanation given, let us formulate our main theorem:

Theorem 1 ('Unsoundness' of BAN logic) *Within BAN logic (as defined in [5, 7]) it is possible to derive unjustifiable beliefs. More precisely, a statement of the form* $A \models X$ *can be derived while it is also consistent to assume that B does not know N.*

Proof (*derivability*) Consider the two parrots protocol, whose BAN idealization is given in Sect. 4. It is trivial to verify that A, C and B are capable of sending the messages they ought to send in the two parrots protocol.

As a result of protocol step 2 (S_2), the following statement is inserted:

$$A \lhd N \tag{1}$$

As a result of protocol step 4 (S_4), the following statement is inserted:

$$A \lhd \{H(N)\}_{K^{-1}} \tag{2}$$

Using inference rule **MM**, assumption $A \models \overset{K}{\mapsto} B$ and (2), we can infer:

$$A \models B \mathrel{|\sim} H(N) \tag{3}$$

Using inference rule **H**, (3) and (1), we can infer:

$$A \models B \mathrel{|\sim} N \tag{4}$$

Using inference rule **NV**, assumption $A \models \sharp(N)$ and (4), we can infer:

$$A \models B \models N \tag{5}$$

Statement (5) should definitely not be derivable from the two parrots protocol. With all protocol assumptions satisfied and only valid inferences applied, an unjustifiable belief is established. More precisely, A believes $B \models N$, while it is also consistent to assume that B does not know N, and nobody tells B about N. Therefore, $A \models B \models N$ is unjustified. □

The culprit is the inference rule **H**. This problem cannot be fixed by adding inference rules in such a way that $B \models N$ can be inferred, as this would thwart the definition of a cryptographic hash function: then N would be derivable from $H(N)$. Such a 'fix' would increase the number of computationally unjustified inference rules from (at least) one to two.

Note that one more inference step is needed after application of the **H** rule before a false belief is established. This is because we need to obtain *belief of belief*, which cannot be directly inferred from **H**.[13]

7 The semantic approach

In the original BAN papers [5, 7], a rather limited semantics is given for a part of the formal language of BAN logic. This semantics has been subject to an enormous amount of criticism. For one thing, the semantics is *very* closely tied to the formal language of BAN logic: what is derivable in the logic is by definition observable in the semantics. Arguably, the semantics is *so* closely tied to the formal language that it is of no additional value. Except for it being the subject of criticism, the semantics has hardly ever been used.

In Sect. 6 we have explained why we used the formulation "unjustified belief" in a proof that does not rely on any formal semantics. Therefore, we have consistently used quotes around the term *unsoundness*. In this section we will provide a proof based on a semantics: therefore, we may omit the quotes around unsoundness. However, for this proof we need to disregard all criticisms of the semantics of BAN logic. Therefore, we regard our proof in the previous section as more important. But it is of course up to the reader to choose what he likes best:

1. to agree with our use of "unjustified belief" in the previous section, and with it agree with the semantics-free proof of 'unsoundness' (shown in the previous section), or
2. to accept the semantics of BAN logic, regardless of all its shortcomings, and with it agree to our proof of unsoundness (shown in this section).

Before we show a run of the two parrots protocol in the semantics of BAN logic, it is appropriate to summarize this semantics:

- A *local state* of a principal P at moment i is a tuple $(\mathscr{M}_P, \mathscr{B}_P)$, where $\mathscr{M}_P$ is the set of messages seen ($\lhd$) by P up to moment i, and $\mathscr{B}_P$ is the set of beliefs ($\models$) of P. These sets enjoy closure properties which correspond to the inference rules of the logic. For compactness and ease of reading, we have only included elements in these sets which are relevant for our purposes. Moreover, $i \leq j$ implies $\mathscr{M}_P(i) \subseteq \mathscr{M}_P(j)$ and similar for $\mathscr{B}_P$.
- A *global state* s is a tuple containing the local states of all principals. If s is a global state, then s_P is the local state of P in s and $\mathscr{M}_P(s)$ and $\mathscr{B}_P(s)$ are the corresponding sets of seen messages and beliefs. In our case the principals are A, B and C, and a global state s is the triple (s_A, s_B, s_C).
- A *run* is a finite sequence of global states $s_0, \ldots, s_n$.
- A *protocol run* of a protocol of n steps of the form $(P_i \rightarrow Q_i : X_i)$ is a run of length $n + 1$, where s_0 corresponds to the protocol assumptions and where $X_i \ni \mathscr{M}_{Q_i}(s_i)$ for all i such that $0 < i \leq n$.

To be able to show a run of the two parrots protocol which is convenient to read, we will first name and give all local states. Then, we will give the full protocol run in which the names of these local states are used. For naming the local states, we adhere to the following convention: $s_P^{n,\ldots,n'}$ is the local state of principal P in the global states $n, \ldots, n'$.

[13] Note that in BAN logic, the semantics of *belief* ($\models$) is defined, while the semantics of *once said* ($\mid\!\sim$) is still "largely a mystery" (literal quote from [5, 7, 4]).

The local states of principals A, B and C are as follows:

$$
\begin{array}{lll}
 & \mathscr{M}_A & \mathscr{B}_A \\
s_A^{0,1} & = (\ \emptyset, & \{\overset{K}{\mapsto} B, N, \sharp(N)\} \quad) \\
s_A^{2,3} & = (\ \{N\}, & \{\overset{K}{\mapsto} B, N, \sharp(N)\} \quad) \\
s_A^{4} & = (\ \{N, \{H(N)\}_{K^{-1}}\}, & \{\overset{K}{\mapsto} B, N, \sharp(N), \\
 & & B \mid\sim H(N), B \mid\sim N, B \models N\} \ ) \\
 & \mathscr{M}_B & \mathscr{B}_B \\
s_B^{0,1,2} & = (\ \emptyset, & \emptyset \quad) \\
s_B^{3,4} & = (\ \{H(N)\} & \{H(N), \{H(N)\}_{K^{-1}}\} \quad) \\
 & \mathscr{M}_C & \mathscr{B}_C \\
s_C^{0} & = (\ \emptyset, & \emptyset \quad) \\
s_C^{1,2,3,4} & = (\ \{N\} & \{N\} \quad)
\end{array}
\tag{6}
$$

The following is a *run* of the two parrots protocol:

$$s_0, s_1, s_2, s_3, s_4 \tag{7}$$

where s_i are the global states after the consecutive steps of the protocol:

$$
\begin{array}{lllll}
 & s_A & s_B & s_C & \\
s_0 = (\ & s_A^{0,1} & s_B^{0,1,2} & s_C^{0} & \ ) \\
s_1 = (\ & s_A^{0,1} & s_B^{0,1,2} & s_C^{1,2,3,4} & \ ) \\
s_2 = (\ & s_A^{2,3} & s_B^{0,1,2} & s_C^{1,2,3,4} & \ ) \\
s_3 = (\ & s_A^{2,3} & s_B^{3,4} & s_C^{1,2,3,4} & \ ) \\
s_4 = (\ & s_A^{4} & s_B^{3,4} & s_C^{1,2,3,4} & \ )
\end{array}
\tag{8}
$$

Now that we have specified a protocol run of the two parrots protocol, we can give our alternative proof of unsoundness:

Theorem 2 (Unsoundness of BAN logic) *Within BAN logic (as defined in [5, 7]) it is possible to derive false beliefs from true premises. More precisely, a statement of the form $A \models X$ can be derived while the statement X itself is false.*

Proof (observability) As shown in statement (5) of the *derivability* proof in Sect. 6, we can derive in BAN logic the sentence $A \models B \models N$ in *every* run S_1, S_2, S_3, S_4 of the two parrots protocol. Thus, we have:

$$S_1, S_2, S_3, S_4 \vdash A \models B \models N \tag{9}$$

Global state s_4 corresponds to the semantics after *a particular* protocol run S_1, S_2, S_3, S_4 of the two parrots protocol. That is, a protocol run in which no eavesdropping occurs and

no extra messages are sent. When we take the model as given in equations (6)–(8), we can observe that 'A believes that B knows N': $B \models N \in \mathscr{B}_A(s_4)$, which gives us:

$$s_4 \models A \models B \models N \tag{10}$$

On the other hand, we can also observe in our model that 'B does not know N': $N \notin \mathscr{B}_B(s_4)$, which gives us:

$$s_4 \not\models B \models N \tag{11}$$

Thus, the belief of A as given in (10) is not true in *a particular* protocol run as shown in (11). The false belief of A as given in (10), is nevertheless derivable (9) in *every* protocol run. Thus, it is possible to derive a false belief within BAN logic. □

Let us quote one last excerpt from Sect. 13, on page 269 of [5] (resp. pages 47–48 of [7]):

> Clearly, some beliefs are false. This seems essential to a satisfactory semantics. […] Most beliefs happen to be true in practice, but the semantics does not account for this coincidence. To guarantee that all beliefs are true we would need to guarantee that all initial beliefs are true.

The existence of false beliefs in the semantics as such is not a problem, the problem is that some false beliefs are derivable from true ones.

8 Discussion

The formal approach to protocol analysis essentially started with BAN logic. Many critiques of BAN logic have appeared, mentioning its incompleteness (i.e., inability to detect some obvious problems, cf. [26]) and its poor semantics (among many others, see [2]). Nevertheless, these critiques have not been a reason to abandon the *way of thinking* introduced by BAN logic [20]. The many augmentations to BAN logic (most notably, AT [2], GNY [17], AUTLOG [21, 35], VO [27], SVO [32, 33] and SVD [13]) show the trust in the formal approach which originates from BAN logic. In our opinion, this consensual trust in the *way of thinking* introduced by BAN logic is justified. While obtaining completeness has long been regarded as impossible, the soundness of BAN logic itself has never been seriously doubted. Wedel and Kessler identified rules in BAN, AT and GNY which are invalid in their semantics, but they point out that it is unclear whether the inference rules or their semantics are to blame for that [35]. Various more recent results [1, 10–12, 31] provide directions on how completeness could be obtained for formal protocol analysis.

Our 'unsoundness' result does not at all invalidate the approach that BAN logic employs for protocol analysis. It should merely count as a warning to those who wish to *complete* their logic. All augmentations of BAN logic are incomplete in the sense that they do not accommodate all cryptographic primitives known to date. These logics are essentially 'just big enough' to capture the problems the authors intend to capture. And to be fair, this has been difficult enough already. Just a few BAN-descendant logics accommodate cryptographic hash functions, none of them accommodate fancy primitives like (to name just an example) oblivious transfer [29, 15].

The problem we identified in this paper can be addressed by removing the inference rule **H** from the set of allowed rules in BAN logic. Removing *one* known problem does of course not guarantee the absence of other problems. Therefore, it might still be that even then BAN is

'unsound' or even unsound (without quotes), but at least not as a result of problems identified in this paper. The inference rule **H** is not vital for most applications of BAN logic. In [34] we propose some candidate inference rules which may be added to BAN-like logics in case the logic does need to model hash functions.

The fact that none of the hash-accommodating BAN-descendant logics adopts the **H** inference rule, can probably be explained by the observation that constructing a good logic is already so difficult that none of the authors will have felt the urge to include an inference rule into their logic that was not needed to capture the problem the author intended to capture. Nevertheless, it is remarkable that we are apparently the first to find this result on a paper which has been so extensively studied and which is almost two decades old.

So far, we know of only one publication which relies on the faulty **H** inference rule [3]. In this publication, the SET protocol[14] is analyzed in BAN logic. It remains open whether the authors' assessment of SET holds in a BAN logic with the inference rule **H** omitted.

Acknowledgements I would like to thank the anonymous referees, Rineke Verbrugge and Wiebe van der Hoek for their enthusiasm, insightful comments and very useful suggestions.

Appendix A: a taxonomy of versions of the BAN paper

The seminal paper "A logic of authentication" has a respectable number of versions. Its precursor, "Authentication: a practical study in belief and action" was presented at the second conference on Theoretical Aspects of Reasoning About Knowledge in March 1988 [4, 18 pages]. Then, there is the DEC technical report, which was published in February 1989 and revised in February 1990 [7, 49 pages]. In April 1989, the work was submitted to the Royal Society of London, which published it in December 1989 [5, 39 pages]. Also in December 1989, a revised version of the article was presented on the twelfth ACM Symposium on Operating Systems Principles, which was also published in the ACM SIGOPS Operating Systems Review [6, 13 pages]. This led to a paper in the ACM Transactions on Computer Systems in February 1990 [8, 19 pages]. In May 1994, an appendix to the DEC technical report was published [9, 10 pages].

The most notable distinction between these versions is that in the ACM-published versions and the DEC appendix, the notation of many operators has changed from symbols (e.g. $\models$) to linguistic terms (e.g. **believes**). These versions refer to the DEC technical report for full reference. The DEC technical report and the Royal Society version [7, 5] should be considered the most complete versions, due to their size and the fact that these papers are most often used in self-references of the authors. Martín Abadi considers the Royal Society version the most definite one (on his homepage). These two versions of the article contain a Sect. 12, "On Hashing", which introduces and discusses the inference rule essential in this paper. These two versions also contain a Sect. 13, "Semantics", which defines the partial semantics for BAN logic, used in Sect. 7 of this paper.

[14] SET stands for *Secure Electronic Transactions* [22–24]. The protocol was introduced by VISA and Mastercard for online payments, but it has never been widely adopted or deployed.

References

1. Abadi, M., & Rogaway, P. (2002). Reconciling two views of cryptography (the computational soundness of formal encryption). *Journal of Cryptology, 15*(2), 103–127.
2. Abadi, M., & Tuttle, M. (1991). A semantics for a logic of authentication. In *Proceedings of the Tenth Annual ACM Symposium on Principles of Distributed Computing* (pp. 201–216). Montreal, August 1991.
3. Agray, N., van der Hoek, W.,& de Vink, E. P. (2002). On BAN logics for industrial security protocols. In B. Dunin-Kęplicz & E. Nawarecki (Eds.), *Proceedings of the Second International Workshop of Central and Eastern Europe on Multi-Agent Systems, volume 2296* of *Lecture Notes in Artificial Intelligence* (pp. 29–36). Berlin/Heidelberg, 2002.
4. Burrows, M., Abadi, M., & Needham, R. (1988). Authentication: A practical study in belief and action. In M. Vardi (Ed.), *Proceedings of the Second Conference on Theoretical Aspects of Reasoning About Knowledge* (pp. 325–342).
5. Burrows, M., Abadi, M., & Needham, R. (1989a). A logic of authentication. *Proceedings of the Royal Society of London, Series A, Mathematical and Physical Sciences, 426*(1871), 233–271.
6. Burrows, M., Abadi, M., & Needham, R. (1989b). A logic of authentication. *ACM SIGOPS Operating Systems Review (Proceedings of the 12th ACM Symposium on Operating Systems Principles), 23*(5), 1–13.
7. Burrows, M., Abadi, M., & Needham, R. (1990a). A logic of authentication. *Technical Report 39, Digital Equipment Corporation Systems Research Center*, 28 February 1989. Revised on 22 February 1990.
8. Burrows, M., Abadi, M., & Needham, R. (1990b). A logic of authentication. *ACM Transactions on Computer Systems, 8*(1), 18–36.
9. Burrows, M., Abadi, M., & Needham, R. (1994) . A scope of a logic of authentication. *Appendix to DEC SRC Research Report 39, Digital Equipment Corporation Systems Research Center*, 13 May 1994.
10. Cohen, M., & Dam, M. (2005a). A completeness result for BAN logic. In *Prococeedings of Methods for Modalities 4*, Berlin.
11. Cohen, M., & Dam, M. (2005b). Logical omniscience in the semantics of BAN logics. In A. Sabelfeld (Ed.), *Proceedings of the Foundations of Computer Security '05—FCS '05* (pp. 121–132), Chicago.
12. Cohen, M., & Dam, M. (2007). A complete axiomatization of knowledge and cryptography. Technical Report Trita-CSC-TCS-2007:1, School of Computer Science and Communication, The Royal Institute of Technology (KTH), Stockholm, Sweden, 2007.
13. Dekker, A. H. (2000). C3PO: A tool for automatic sound cryptographic protocol analysis. In *Proceedings of the 13th IEEE Computer Security Foundations Workshop—CFSW-13* (pp. 77–87). IEEE Computer Society Press, 2000.
14. Diffie, W., & Hellman, M. E. (1976). New directions in cryptography. *IEEE Transactions on Information Theory, IT, 22*(6), 644–654.
15. Even, S., Goldreich, O., & Lempel, A. (1985). A randomized protocol for signing contracts. *Communications of the ACM, 28*(6), 637–647.
16. Fagin, R., Halpern, J. Y., Moses, Y., & Vardi, M. Y. (1995) *Reasoning about Knowledge*. Cambridge, MA: MIT Press.
17. Gong, L., Needham, R., & Yahalom, R. (1990). Reasoning about belief in cryptographic protocols. In D. Cooper & T. Lunt (Eds.), *Proceedings 1990 IEEE Symposium on Research in Security and Privacy*, (pp. 234–248). Los Angeles: IEEE Computer Society Press.
18. Halpern, J. Y., & Moses, Y. (1984). Knowledge and common knowledge in a distributed environment. In T. Kameda, J. Misra, J. Peters, & N. Santoro (Eds.), *Proceedings of the Third Annual ACM Symposium on Principles of Distributed Computing* (pp. 50–61). ACM: ACM Press.
19. Halpern, J. Y., & Moses, Y. (1990). Knowledge and common knowledge in a distributed environment. *Journal of the ACM, 37*(3), 549–587.
20. Halpern, J. Y., Pucella, R., & van der Meyden, R. (2003). Revisiting the foundations of authentication logics, Manuscript.
21. Kessler, V., & Wedel, G. (1994). AUTLOG—an advanced logic of authentication. In *Proceedings of the 7th Computer Security Foundations Workshop (CSFW'94)* (pp. 90–99). Los Alamitos: IEEE Computer Society Press.
22. Mastercard & Visa. (1997a). *The SET standard book 1: Business description, version 1.0*. SETCO, May 31 1997.
23. Mastercard & Visa. (1997b). *The SET standard book 2: Programmer's guide, version 1.0*. SETCO, May 31 1997.
24. Mastercard & Visa. (1997c). *The SET standard book 3: Formal protocol definitions, version 1.0*. SETCO, May 31 1997.

25. Meyer, J. -J. Ch., & van der Hoek, W. (1995). *Epistemic logic for AI and Computer Science*. Cambridge University Press.
26. Nessett, D. M. (1990). A critique of the Burrows, Abadi and Needham Logic. *ACM SIGOPS Operating Systems Review, 24*(2), 35–38.
27. van Oorschot, P. C. (1993). Extending cryptographic logics of belief to key agreement protocols (extended abstract). In *Proceedings of the First ACM Conference on Computer and Communications Security* (pp. 232–243). New York: ACM Press, November 1993.
28. Preneel, B. (1993). *Analysis and design of cryptographic hash functions*. PhD thesis, Katholieke Universiteit Leuven, January 1993.
29. Rabin, M. O. (1981). How to exchange secrets by oblivious transfer. *Technical Report TR-81*. Aiken Computation Laboratory, Harvard University.
30. Schneier, B. (1990). *Applied Cryptography*. New York: Wiley.
31. Syverson, P. (2000). Towards a strand semantics for authentication logic. In S. Brookes, A. Jung, M. Mislove, & A. Scedrov (Eds.), *Electronic notes in theoretical computer science, vol. 20*.
32. Syverson, P., & van Oorschot, P. C. (1994). On unifying some cryptographic protocol logics. In *Proceedings of the 1994 IEEE Computer Society Symposium on Research in Security and Privacy* (pp. 14–28). IEEE Computer Society Press, May 1994.
33. Syverson, P., & van Oorschot, P. C. (1996). A unified cryptographic protocol logic. Report 5540-227, Center for High Assurance Computer Systems, Naval Research Laboratory (NRL CHACS).
34. Teepe, W. (2006). *Reconciling information exchange and confidentiality—a formal approach*. PhD thesis, Rijksuniversiteit Groningen.
35. Wedel, G., & Kessler, V. (1996). Formal semantics for authentication logics. In E. Bertino, H. Kurth, G. Martella, & E. Montolivo (Eds.), *Computer Security—ESORICS 96: 4th European Symposium on Research in Computer Security Rome, number 1146* in Lecture Notes in Computer Science (pp. 219–241). Berlin: Springer-Verlag.