Honest Ideals on Strand Spaces*

F. Javier THAYER Fábrega Jonathan C. HERZOG
Joshua D. GUTTMAN
The MITRE Corporation
{jt, jherzog, guttman}@mitre.org

Abstract

In security protocol analysis, it is important to learn general principles that limit the abilities of an attacker, and that can be applied repeatedly to a variety of protocols. We introduce the notion of an ideal*—a set of messages closed under encryption and invariant under composition with arbitrary messages—to express such principles.*

In conjunction with the strand space formalism, we use the concept of ideals *to prove bounds on a penetrator's capabilities, independent of the security protocol being analyzed. From this we prove a number of correctness properties of the Otway Rees protocol, using these results to explain the limitations of the protocol.*

1 Introduction and Review

A security protocol is a sequence of messages between two or more parties in which encryption is used to provide authentication or to distribute cryptographic keys for new sessions. In this paper we extend the ideas of [7], in which we introduced the concept of a *strand space* and used it to formulate and prove correctness properties for the Needham-Schroeder-Lowe protocol.

In this paper, we will develop more of the algebra of messages. We will be more explicit about the structure that we need to assume on the set of messages, under the operations of encryption and message concatenation (Section 2). In [7], to simplify the exposition, we assumed that these formed a free algebra.

We will also introduce additional algebraically natural sets of messages—we call them *ideals*—that make it easier to state and prove general facts about the powers of the penetrator (Section 3). An *ideal* is a set of messages closed under encryption and invariant under composition with arbitrary messages.

These general theorems about the powers of the penetrator are independent of the protocols to be analyzed, so that they can be re-used effectively for many protocols. A typical specimen asserts that if a legitimate protocol entity never utters any message in an ideal I, then a penetrator can never utter any message in I either (Section 4). We call these kinds of theorems "bounds on the penetrator."

We have applied these methods to analyze the Otway-Rees protocol and the Yahalom protocol [6]. In this paper (Section 5), we will use Otway-Rees to illustrate the utility of the penetrator bounds. Our results explain in a very clear way exactly what the protocol establishes, and what its fundamental limitations are.

In order to make the paper self-contained, we review some of the terminology of our earlier paper [7] in the remainder of this introduction.

1.1 Strands

Throughout the paper, A will denote the set of messages that can be exchanged between principals in a protocol.[1] We will refer to the elements of A as *terms*. In a protocol, principals can either send or receive terms. We will represent sending a term as the occurrence of that term with positive sign, and receiving a term as its occurrence with a negative sign.

Definition 1.1 *A* signed term *is a pair* $\langle \sigma, a \rangle$ *with* $a \in \mathsf{A}$ *and* σ *one of the symbols* $+, -$. *We will write a signed term as* $+t$ *or* $-t$. $(\pm\mathsf{A})^*$ *is the set of finite sequences of signed terms. We will denote a typical element of* $(\pm\mathsf{A})^*$ *by* $\langle\, \langle \sigma_1, a_1 \rangle, \ldots, \langle \sigma_n, a_n \rangle \,\rangle$.

Definition 1.2 *A* strand space *is a set* Σ *with a* trace *mapping* $tr : \Sigma \to (\pm\mathsf{A})^*$.

*This work was supported by the National Security Agency through US Army CECOM contract DAAB 07-96-C-E601. Appears in *Proceedings, 1998 Computer Security Foundations Workshop*, June 1998, Rockport, MA.

[1] In this paper, we will use a $\mathsf{sansserif}$ style for sets like A and its important subsets, and for the basic operators on A. In [7], **bold face** was used for these as well as for other items.

In particular applications of the theory, the mapping tr may fail to be injective because we may need to distinguish between various instances of the same trace. For instance, to model authentication properties of certain protocols it may be necessary to distinguish identical traces originating from different principals, or to model simple replay attacks we may need to distinguish identical traces originating from the same principal.

Fix a strand space Σ.

1. A *node* is a pair $\langle s, i\rangle$, with $s \in \Sigma$ and i an integer satisying $1 \leq i \leq \mathrm{length}(\mathrm{tr}(s))$. The set of nodes is denoted by $\mathcal{N}$. We will say the node $\langle s, i\rangle$ belongs to the strand s. Clearly, every node belongs to a unique strand.

2. If $n = \langle s, i\rangle \in \mathcal{N}$ then $\mathrm{term}(n)$ is $(\mathrm{tr}(s))_i$, i.e. the ith signed term in the trace of s. Similarly, $\mathrm{uns\_term}(n)$ is $((\mathrm{tr}(s))_i)_2$, i.e. the unsigned part of the ith signed term in the trace of s.

3. If $n_1, n_2 \in \mathcal{N}$, $n_1 \to n_2$ means $\mathrm{term}(n_1) = +a$ and $\mathrm{term}(n_2) = -a$. It means that node n_1 sends the message a, which may be received by n_2, creating a causal link between their strands.

4. If $n_1, n_2 \in \mathcal{N}$, then $n_1 \Rightarrow n_2$ means n_1, n_2 occur on the same strand s with $n_1 = \langle s, i\rangle$ and $n_2 = \langle s, i+1\rangle$. It expresses the causal dependence of a later action on its predecessor.

$\mathcal{N}$ becomes an ordered graph with both sets of edges $n_1 \to n_2$ and $n_1 \Rightarrow n_2$.

1.2 Bundles

A *bundle* in a strand space is a finite subgraph of the node graph $\mathcal{N}$, for which we can regard the edges as expressing the causal dependencies of the nodes.

Definition 1.3 *Let $\mathcal{C}$ be a set of edges, and let $\mathcal{N}_{\mathcal{C}}$ be the set of nodes incident with any edge in $\mathcal{C}$. $\mathcal{C}$ is a bundle if:*

1. *$\mathcal{C}$ is finite.*

2. *If $n_1 \in \mathcal{N}_{\mathcal{C}}$ and $\mathrm{term}(n_1)$ is negative, then there is a unique n_2 such that $n_2 \to n_1 \in \mathcal{C}$.*

3. *If $n_1 \in \mathcal{N}_{\mathcal{C}}$ and $n_2 \Rightarrow n_1$ then $n_2 \Rightarrow n_1 \in \mathcal{C}$.*

4. *$\mathcal{C}$ is acyclic.*

We will speak of a node as being in the bundle $\mathcal{C}$ if in fact it is in $\mathcal{N}_{\mathcal{C}}$.

A well-formed bundle is illustrated in Figure 1, although this bundle does not exemplify a useful protocol.

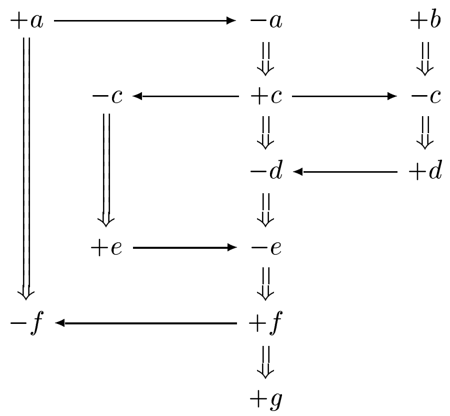


Figure 1. A Bundle

Definition 1.4 *If $\mathcal{C}$ is a bundle and $s \in \Sigma$, then the $\mathcal{C}$* height *of s, denoted $\mathit{height}_{\mathcal{C}}(s)$, is the largest $i \leq \mathit{length}(\mathit{tr}(s))$ such that $s \in \Sigma$ and $\langle s, i\rangle \in \mathcal{C}$.*

$\mathcal{C}$ contains *s if $\mathit{height}_{\mathcal{C}}(s) = \mathit{length}(\mathit{tr}(s))$.*

Clearly $\langle s, j\rangle \in \mathcal{C}$ for all $j \leq \mathrm{height}_{\mathcal{C}}(s)$. The model intentionally allows strands representing legitimate protocol agents to have less than full height.

Definition 1.5 *If s is a strand and $\mathcal{C}$ a bundle, the $\mathcal{C}$-trace of s is the restriction of $\mathit{tr}(s)$ to the integer interval $\{1, \ldots, \mathit{height}_{\mathcal{C}}(s)\}$. A partial $\mathcal{C}$-trace of s is the restriction of $\mathit{tr}(s)$ to any interval $\{1, \ldots, k\}$ for $k \leq \mathit{height}_{\mathcal{C}}(s)$.*

Definition 1.6 *Suppose that $\mathcal{S}$ is a set of edges, i.e. a subset of the union of $\to$ and $\Rightarrow$, and let $\mathcal{N}_{\mathcal{S}}$ be the set of nodes incident with any edge in $\mathcal{S}$.*

Then $\prec_{\mathcal{S}}$ is the transitive closure of $\mathcal{S}$, and $\preceq_{\mathcal{S}}$ is the reflexive, transitive closure of $\mathcal{S}$.

Each relation is a subset of $\mathcal{N}_{\mathcal{S}} \times \mathcal{N}_{\mathcal{S}}$. Moreover, $n \prec_{\mathcal{S}} n'$ means that there is a sequence of one or more edges (of either kind) belonging to $\mathcal{S}$ leading from n to n'. Similarly, $n \preceq_{\mathcal{S}} n'$ means that there is a sequence of zero or more edges belonging to $\mathcal{S}$ leading from n to n'.

Lemma 1.7 *Suppose $\mathcal{C}$ is a bundle. Then $\preceq_{\mathcal{C}}$ is a partial order, i.e. a reflexive, antisymmetric, transitive relation. Every non-empty subset of the nodes in $\mathcal{C}$ has $\preceq_{\mathcal{C}}$-minimal members.*

When a bundle $\mathcal{C}$ is understood, we will simply write $\preceq$.

1.3 Messages

In the remainder of this paper, we will specialize the set of messages A and assume it has additional structure intended to model message construction and message encryption. We specialize A by introducing:

- A set $\mathsf{T} \subset \mathsf{A}$ of texts (representing the atomic messages), and a disjoint set $\mathsf{K} \subset \mathsf{A}$ of cryptographic keys.
- A unary operator $\mathsf{inv} : \mathsf{K} \to \mathsf{K}$. We assume that inv maps each member of a key pair for an asymmetric cryptosystem to the other, and that it maps a symmetric key to itself.
- Two binary operators

$$\mathsf{encr} : \mathsf{K} \times \mathsf{A} \to \mathsf{A}$$
$$\mathsf{join} : \mathsf{A} \times \mathsf{A} \to \mathsf{A}$$

To follow accepted notation, we will write: $\mathsf{inv}(K) = K^{-1}$, $\mathsf{encr}(K, m) = \{m\}_K$ and $\mathsf{join}(a, b) = a\,b$. To minimize the use of parentheses in our notation, we will implicitly associate terms on the right. Thus $a\,b\,c$ is an abbreviation of $a\,(b\,c)$. Note that nothing is stated about the kind of encryption used here.

We will refer to the range of encr, namely the ciphertexts of the form $\{h\}_K$, as E. We will refer to the set of terms of the form $a\,b$, as C. A term is *simple* if it is an element of $\mathsf{K} \cup \mathsf{E} \cup \mathsf{T}$. Note that the range of encryption is included in the simple terms.

1.4 Infiltration

A *penetrator set* consists of a set of keys $\mathsf{K}_\mathcal{P}$. It consists of all keys initially known to the penetrator. Typically it would contain all public keys, all private keys of penetrators and all symmetric keys K_{px}, K_{xp} initially shared between the penetrator and a principal that plays by the protocol rules. It may also contain "lost keys" that are known to the penetrator, either because of the carelessness of a nonmalicious principal, or else because the pentrator has succeeded in some cryptanalysis.

The actions available to the penetrator are encoded in a set of *penetrator traces* that summarize his ability to discard messages, generate well known messages, piece messages together, and apply cryptographic operations using keys that become available to him.

Definition 1.8 *A* penetrator trace *is one of the following:*

M. *Text message:* $\langle +t \rangle$ *where* $t \in \mathsf{T}$

F. *Flushing:* $\langle -g \rangle$

T. *Tee:* $\langle -g,\ +g,\ +g \rangle$

C. *Concatenation:* $\langle -g,\ -h,\ +g\,h \rangle$

S. *Separation into components:* $\langle -g\,h,\ +g,\ +h \rangle$

K. *Key:* $\langle +K \rangle$ *where* $K \in \mathsf{K}_\mathcal{P}$.

E. *Encryption:* $\langle -K,\ -h,\ +\{h\}_K \rangle$.

D. *Decryption:* $\langle -K^{-1},\ -\{h\}_K,\ +h \rangle$.

It is also possible to extend the set of penetrator traces given here if it is desired to model some special ability of the penetrator, such as the ability to cryptanalyze some kinds of encrypted messages, without any essential change to our overall framework.

Definition 1.9 *An* infiltrated strand space *is a pair* $(\Sigma, \mathcal{P})$ *with* Σ *a strand space and* $\mathcal{P} \subseteq \Sigma$ *such that tr*(p) *is a penetrator trace for* $p \in \mathcal{P}$.

A strand $s \in \Sigma$ *is a* penetrator strand *if it belongs to* $\mathcal{P}$*, and a node is a* penetrator node *if the strand it lies on is a penetrator strand. Otherwise we will call it a* non-penetrator *or* regular *strand or node.*

A node n *is a* **M**, **F**, *etc. node if* n *lies on a penetrator strand with a trace of kind* **M**, **F**, *etc.*

We would not expect an infiltrated strand space to realize all of the penetrator traces of type **M**. In that case, the space could not model unguessable nonces. The more useful spaces Σ lack **M**-strands for many text values, which the legitimate participants can use as fresh nonces.

2 Unique Readability

When reasoning about terms, it is important to know whether they are ambiguous, in the sense that there are different ways to "read" or "parse" them. We can draw conclusions more effectively if portions (at least) of the term can be read in only one way. In [7], we took the short way with this issue, assuming that the algebra of messages is free. In this section, we develop a more flexible algebraic framework that allows (for instance) message concatenation to be associative. The conclusions of [7] remain true in this more realistic context.

Axiom 2.1 handles the case of a term that can be regarded as a ciphertext, and asserts that it can be regarded as a ciphertext in only one way. Axiom 2.2 deals with the decomposition of composite terms, and their relation to other terms.

Axiom 2.1 *If* $\{t\}_K = \{t'\}_{K'}$ *then* $K = K'$ *and* $t = t'$.

We will refer to this assumption as the *free encryption assumption*; other authors such as Paulson [5] and Marrero et al [4] make similar assumptions.

There exist interpretations of the theory we are presenting in which this axiom is satisfied, for instance, the set of formal expressions built from K and T using the operations join and encr. However, in the most common application of the theory—namely cryptography—Axiom 2.1 is false, because there are many relations in the algebra of real messages. For instance a cardinality argument immediately establishes that there must be many distinct pairs of an input block of plaintext and a DES key (for instance) that yield the same block of ciphertext. Nevertheless, a good cryptosystem makes it hard to find pairs that will collide in this way. Moreover, there should be very few different *meaningful* texts for which there exist keys that will cause them to collide, for most notions of "meaningful." Axiom 2.1 idealizes the situation by assuming that there are none. This matches our goal, namely to determine whether protocols have weaknesses independent of the choice of cryptosystem. An initial step in giving a complete answer to this question is to consider whether there would still be weaknesses, even if the cryptosystem is ideal.

Axiom 2.2 *No simple element is in* C*, the range of* join*. If* $p\,a = q\,b$ *with* p, q *simple, then* $p = q$ *and* $a = b$*.*

We turn now to the question of decomposing a composite term into a succession of simple components. An *exhausted* term is one for which this is no longer possible.

Definition 2.3 $a \in \mathsf{A}$ *is* exhausted *iff* a *cannot be expressed in not the form* $p\,h$ *for* p *simple.*

Clearly any simple term is exhausted since by Axiom 2.2, it cannot be written in the form $a\,b$.

Proposition 2.4 *For any term* a *either*

1. *a can be expressed as $p_1\ \cdots\ p_k\,h$ where where each p_i is simple and h is exhausted. If such a representation exists it is unique.*
2. *There is a unique pair of infinite sequences $\{p_i\}_{1\leq i}$, $\{h_i\}_{1\leq i}$ where each p_i is simple such that $a = p_1\ \cdots\ p_k\,h_k$.*

PROOF. Suppose

$$p_1\ \cdots\ p_k\,h = q_1\ \cdots\ q_n\,g$$

where p_i, q_j are simple and $k < n$. Applying Axiom 2.2 repeatedly,

$$h = q_{k+1}\ \cdots\ q_n\,g$$

contradicting the assumption h is exhausted. Similarly, we can exclude $n < k$. Thus $k = n$ and it follows immediately from Axiom 2.2 that $p_i = q_i$ and $h = g$. If there is no representation of a in the form stated, then for any any k there is a unique representation of a in the form $p_1\ \cdots\ p_k\,h$ where h is not exhausted. ■

Definition 2.5 *Let* $a \in \mathsf{A}$ *and* $1 \leq k$*:*

1. *a has width k iff $a = p_1\ \cdots p_{k-1}\,h$ where each p_i is simple and h exhausted.*
2. *a has width $+\infty$ if a does not have a representation $p_1\ \cdots p_{k-1}\,h$ where each p_i is simple and h is exhausted.*

Lemma 2.6 *Any* $a \in \mathsf{A}$ *has width* k *for exactly one* $1 \leq k \leq \infty$*.*

PROOF. Follows immediately from Proposition 2.4. ■

This lemma may be used in various forms to show that sets of terms of certain forms are disjoint from each other. For instance, the result of concatenating an atomic text and a key never collides with the result of concatenating two texts before any member of A, which we need in treating the Otway-Rees protocol (Section 5).

Proposition 2.7 *The set of terms of the form* $h\,K$ *is disjoint from the set of terms of the form* $h\,h'\,a$ *for all* $h, h' \in \mathsf{T}$*,* $K \in \mathsf{K}$ *and* $a \in \mathsf{A}$*.*

Attacks that might exist if there are terms that may be "read" as having more than one form are referred to as *type flaw attacks* [2]. Some type flaw attacks seem implausible, in the sense that most implementations would not be vulnerable to them, while others are more troublesome. We will not consider type flaws further in the current paper, although there are various possible approaches to extending strand spaces to model them.

3 Ideals

We introduce the concept of *ideal* for two purposes:

1. To make it easier to formulate general facts about the penetrator's capabilities.
2. As a technical device for stating assumptions and proving facts about the subterm relationship. In our previous paper, [7] we made the simplifying assumption that the message algebra was free and so no additional assumptions were necessary to guarantee results such as Corollary 3.14 below.

Definition 3.1 *If* $\mathsf{k} \subseteq \mathsf{K}$*, a* k*-ideal of* A *is a subset* I *of* A *such that for all* $h \in I$*,* $g \in \mathsf{A}$ *and* $K \in \mathsf{k}$

1. $h\,g, g\,h \in I$.
2. $\{h\}_K \in I$.

The smallest k*-ideal containing* h *is denoted* $I_{\mathsf{k}}[h]$*.*

We now define a subterm relation $\sqsubset$ that uses the structure of message composition and encryption specific to Section 1.3.

Definition 3.2 h *is a* subterm *of* g*, written* $h \sqsubset g$ *is defined as* $g \in I_{\mathsf{K}}[h]$*.*

This definition gives a more restricted notion of subterm then one might have expected. In particular, $K \not\sqsubset (\{h\}_K)$ unless K already happened to be a subterm of h. Restricting subterms in this way reflects an assumption about the penetrator's capabilities, to wit, that keys can be obtained from cyphertext only if they are embedded in the text that was encrypted. This might not always be the case—for instance, if a dictionary attack is possible—but it is the assumption we will make in this paper. Future work within the same framework could certainly relax the assumption.

Proposition 3.3 *$\sqsubset$ is a transitive, reflexive relation. Moreover, if $h, g \in \mathsf{A}$ and $K \in \mathsf{K}$, then*

1. *$h \sqsubset h\,g$ and $g \sqsubset h\,g$.*
2. *$h \sqsubset \{h\}_K$.*

PROOF. Clearly $h \in I_{\mathsf{K}}[h]$, so $h \sqsubset h$. If $g \sqsubset g'$, then $g' \in I_{\mathsf{K}}[g]$. If in addition $h \sqsubset g$, then $g \in I_{\mathsf{K}}[h]$, so by the definitions $I_{\mathsf{K}}[g] \subseteq I_{\mathsf{K}}[h]$. Therefore $g' \in I_{\mathsf{K}}[h]$.

If $h, g \in \mathsf{A}$ and $K \in \mathsf{K}$, then clearly $h\,g, g\,h, \{h\}_K \in I_{\mathsf{K}}[h]$. ■

Axiom 3.4 *If t is a simple term and $g\,h \in I_{\emptyset[t]}$ then either $g \in I_\emptyset[t]$ or $h \in I_\emptyset[t]$.*

Axiom 3.5 *If $K \in \mathsf{K}$, $t \in \mathsf{T}$, $e \in \mathsf{E}$ and $c \in \mathsf{C}$.*

1. *$e \not\sqsubset K$.*
2. *$e \not\sqsubset t$.*
3. *$K \not\sqsubset t$.*
4. *$c \not\sqsubset K$.*
5. *$c \not\sqsubset t$.*

It also follows that $t \not\sqsubset K$, although this fact is not needed here.

Lemma 3.6 *The sets K, T, E and C are pairwise disjoint.*

PROOF. Since $t \sqsubset t$ the result follows from immediately from Axioms 3.5 and 2.2.

Definition 3.7 *Suppose $\mathsf{k} \subseteq \mathsf{K}$. $s \in \mathsf{A}$ is a k-subterm of $t \in \mathsf{A}$, written $s \sqsubset_{\mathsf{k}} t$ iff $t \in I_{\mathsf{k}}[s]$.*

If $s \sqsubset_\emptyset t$, then we use the expression *s is a visible subterm of t*.

Proposition 3.8 *$\sqsubset_{\mathsf{k}}$ is a transitive, reflexive relation. Moreover, $h \sqsubset_{\mathsf{k}} g$ implies $h \sqsubset g$.*

PROOF. To prove $\sqsubset_{\mathsf{k}}$ is a transitive, reflexive relation, see the proof of Proposition 3.3. If $h \sqsubset_{\mathsf{k}} g$ then $g \in I_{\mathsf{k}}[h] \subseteq I_{\mathsf{K}}[h]$ so $h \sqsubset g$ as asserted. ■

Definition 3.9 *If $S \subseteq \mathsf{A}$, $I_{\mathsf{k}}[S]$ is the smallest k-ideal containing S.*

The ideal structure is very simple:

Proposition 3.10 *If $S \subseteq \mathsf{A}$, $I_{\mathsf{k}}[S] = \bigcup_{x \in S} I_{\mathsf{k}}[x]$.*

PROOF. The property of being a k-ideal is equivalent to closure under the mappings $x \mapsto x\,a$, $x \mapsto a\,x$ and $x \mapsto \{x\}_k$ for $k \in \mathsf{k}$. Thus the union of k-ideals is a k-ideal. Thus $\bigcup_{x \in S} I_{\mathsf{k}}[x]$ is a k-ideal which contains S. Clearly $\bigcup_{x \in S} I_{\mathsf{k}}[x] \subseteq I_{\mathsf{k}}[S]$. ■

Lemma 3.11 *Let $S_0 = S$, $S_{i+1} = \{\{g\}_K : g \in I_\emptyset[S_i], K \in \mathsf{k}\}$. Then $I_{\mathsf{k}}[S] = \bigcup_i I_\emptyset[S_i]$.*

PROOF. By induction, $S_i \subseteq I_{\mathsf{k}}[S]$, so $\bigcup_i I_\emptyset[S_i] \subseteq I_{\mathsf{k}}[S]$. In the other direction, $\bigcup_i I_\emptyset[S_i]$ is clearly a k-ideal which contains S. ■

Proposition 3.12 *Suppose $S \subseteq \mathsf{A}$, and every $s \in S$ is simple. If $g\,h \in I_{\mathsf{k}}[S]$ then either $g \in I_{\mathsf{k}}[S]$ or $h \in I_{\mathsf{k}}[S]$.*

PROOF. In virtue of the previous lemma, $g\,h \in I_\emptyset[S_i]$ for some i. By Proposition 3.10, $g\,h \in I_\emptyset[x]$ for some $x \in S_i$. This x is simple, as either $i = 0$, in which case $S_i = S$, or else $i = j + 1$, in which case each $x \in S_i$ is of the form $\{h\}_K$, and hence simple. Thus by Axiom 3.4, either $g \in I_\emptyset[x]$ or $h \in I_\emptyset[x]$.

Proposition 3.13 *Suppose $K \in \mathsf{K}$; $S \subseteq \mathsf{A}$; and for every $s \in S$, s is simple and is not of the form $\{g\}_K$. If $\{h\}_K \in I_{\mathsf{k}}[S]$, then $h \in I_{\mathsf{k}}[S]$.*

PROOF. Assume $K \in \mathsf{K}$, $\{h\}_K \in I_{\mathsf{k}}[S]$ and $h \notin I_{\mathsf{k}}[S]$. Let I' be the set difference $I_{\mathsf{k}}[S] \setminus \{\{h\}_K\}$. Clearly $S \subseteq I'$, since S does not contain anything encrypted with outermost key K. Moreover I' is a k-ideal: Since $I_{\mathsf{k}}[S]$ is already an ideal and $\{h\}_K$ is not of the form $a\,b$, I' clearly satisfies the join closure condition for ideals. If $\{h\}_K = \{h_1\}_{K'}$ for $h_1 \in I'$, then by Axiom 2.1 (free encryption), $h = h_1 \in I' \subseteq I_{\mathsf{k}}[S]$ a contradiction. Thus I' is an ideal which contains S. This contradicts the definition of $I_{\mathsf{k}}[S]$ as the smallest ideal which contains S. ■

In Proposition 3.13, S may contain a term $\{g\}_{K'}$ where $K' \neq K$ and g in turn contains subterms encrypted in K.

Corollary 3.14 *Suppose $K \neq K'$ and $\{h'\}_{K'} \sqsubset \{h\}_K$. Then $\{h'\}_{K'} \sqsubset h$.*

PROOF. The assumption means $\{h\}_K \in I_{\mathsf{K}}[\{h'\}_{K'}]$, which by the Proposition implies $h \in I_{\mathsf{K}}[\{h'\}_{K'}]$.

Proposition 3.15 *Suppose $K \in \mathsf{K}$; $S \subseteq \mathsf{A}$; and every $s \in S$ is simple and is not of the form $\{g\}_K$. If $\{h\}_K \in I_{\mathsf{k}}[S]$ for $K \in \mathsf{K}$, then $K \in \mathsf{k}$.*

$$\begin{array}{ccc} \notin I & \notin I & \in I \\ \pm\bullet \cdots\cdots\Rightarrow & \pm\bullet \Longrightarrow & +\bullet \end{array}$$

Figure 2. Entry Point for I

The proof is similar to the proof of Proposition 3.13.

PROOF. Assume $K \in \mathsf{K}$, $\{h\}_K \in I_{\mathsf{k}}[S]$ and $K \notin \mathsf{k}$. As in the preceding proposition, let $I' = I_{\mathsf{k}}[S] \setminus \{\{h\}_K\}$. For the same reason as before, $S \subseteq I'$ and I' satisfies the join closure condition for ideals. Moreover, by free encryption, $\{h\}_K$ is not of the form $\{h'\}_{K'}$ for any $K' \in \mathsf{k}$. Thus I' is an ideal which contains S. This contradicts the definition of $I_{\mathsf{k}}[S]$. ■

4 Origination and Honesty

Definition 4.1 *Suppose Σ is a strand space, $\mathcal{N}$ the set of nodes of Σ. An unsigned term t* originates *on $n \in \mathcal{N}$ iff: term(n) is positive; $t \sqsubset$ term(n); and whenever n' precedes n on the same strand, $t \not\sqsubset$ term(n'). An unsigned term t is* uniquely originating *iff t originates on a unique $n \in \mathcal{N}$.*

Definition 4.2 *A node m is an* entry point *for $I \subseteq \mathsf{A}$ if and only if term(m) is positive, term$(m) \in I$ and for all nodes m' which precede m on the same strand, term$(m') \notin I$.*

We sometimes write $m' \Rightarrow^+ m$ to mean that m' precedes m on the same strand.

Proposition 4.3 *Suppose $\mathcal{C}$ is a bundle over A. If m is minimal in $\{m \in \mathcal{C} : \text{term}(m) \in I\}$, then m is an entry point for I.*

PROOF. If term$(m) = -h$, then by Definition 1.3 Clause 2, there is a node $m' \in \mathcal{C}$ with term$(m') = +h$, violating minimality. If $m' \Rightarrow^+ m$ and term$(m') \in I$, then using Definition 1.3 Clause 3 repeatedly, $m' \in \mathcal{C}$, again contradicting minimality. ■

Definition 4.4 *A set $I \subseteq A$ is* honest *relative to a bundle $\mathcal{C}$ if and only if whenever a penetrator node p is an entry point for I, p is an* **M** *node or a* **K** *node.*

Thus, I is honest relative to $\mathcal{C}$ if the penetrator can achieve entry into I only by a lucky guess: either he utters the right nonce or other text in a lucky **M** node, or he utters the right key in a lucky **K** node. He does not deduce it via his abilities to decrypt and encrypt, or to concatenate and separate.

Our main theorem interrelates the structure of ideals with the possible cases for a penetrator strand.

Theorem 4.5 *Suppose $\mathcal{C}$ is a bundle over A; $S \subseteq \mathsf{T} \cup \mathsf{K}$; $\mathsf{k} \subseteq \mathsf{K}$; and $\mathsf{K} \subseteq S \cup \mathsf{k}^{-1}$. Then $I_{\mathsf{k}}[S]$ is honest.*

PROOF. Let $I = I_{\mathsf{k}}[S]$. Because $I \cap \mathsf{K} = S \cap \mathsf{K}$, we may infer $\mathsf{K} \setminus I = \mathsf{K} \setminus S \subseteq \mathsf{k}^{-1}$. Also, since $S \subseteq \mathsf{T} \cup \mathsf{K}$, the set S contains nothing encrypted and no concatenations, so Propositions 3.12 and 3.13 can be applied.

Suppose m is a penetrator node and an entry point for I. We now consider the various kinds of strands on which a penetrator node can occur. By the definition of entry point, m cannot be on a strand of kind **F** or kind **T**. Consider now the remaining cases:

C. m is on a strand with trace $\langle -g, -h, +h\,g\rangle$. Since $h\,g \in I$, by Proposition 3.12, one of g, h must be in I, contradicting the definition of entry point.

S. m is on a strand with trace $\langle -h\,g, +h, +g\rangle$. Since term$(m)$ must be positive, m is either the second or third node of the strand, so either $h \in I$ or $g \in I$. By the ideal property, $h\,g \in I$, contradicting the definition of entry point.

D. m belongs to a strand with trace $\langle -K_0^{-1}, -\{h\}_{K_0}, +h\rangle$. By the assumption that m is an entry point for I, $K_0^{-1} \notin I$. Hence, $K_0^{-1} \notin S$. However, $\mathsf{K} \subseteq S \cup \mathsf{k}^{-1}$. Therefore $K_0^{-1} \in \mathsf{k}^{-1}$, so $K_0 \in \mathsf{k}$. By the k-ideal property of I, $\{h\}_{K_0} \in I$, contradicting the definition of entry point.

E. m belongs to a strand with trace $\langle -K', -h, +\{h\}_{K'}\rangle$. By assumption $\{h\}_{K'} \in I$. By Proposition 3.13, $h \in I$, contradicting the definition of entry point.

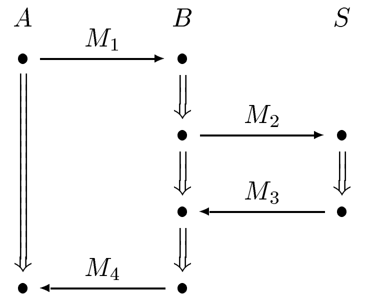
The only remaining possibilities are that m is on a strand of kind **M** or of kind **K** as asserted. ■

In our analysis of Otway-Rees, we use two corollaries of this main result. The first allows us to conclude (in some situations) that if a key is transmitted that is not originally known to the penetrator, then a regular (i.e. non-penetrator) node has provided the entry point.

Corollary 4.6 *Suppose $\mathcal{C}$ is a bundle, $\mathsf{K} = S \cup \mathsf{k}^{-1}$ and $S \cap \mathsf{K}_{\mathcal{P}} = \emptyset$. If term$(m) \in I_{\mathsf{k}}[S]$ for some $m \in \mathcal{C}$, then for some regular node $n \in \mathcal{C}$, n is an entry point for $I_{\mathsf{k}}[S]$.*

PROOF. Suppose m is minimal in $\{n \in \mathcal{C} : \text{term}(n) \in I_{\mathsf{k}}[S]\}$. By Proposition 4.3, m is an entry point for $I_{\mathsf{k}}[S]$. Since by assumption m is not regular (and so must be a penetrator node), Theorem 4.5 implies m is either a penetrator node of kind **M** or of kind **K**.

However, since $\mathsf{K} = S \cup \mathsf{k}^{-1}$, $S \subseteq \mathsf{K}$. Hence $I_{\mathsf{k}}[S] \cap \mathsf{T} = \emptyset$, so m is not of kind **M**. Because $S \cap \mathsf{K}_{\mathcal{P}} = \emptyset$, m is not of kind **K**. ■



Where

1. $M_1 = M\,A\,B\,\{N_a\,M\,A\,B\}_{K_{AS}}$.
2. $M_2 = M\,A\,B\,\{N_a\,M\,A\,B\}_{K_{AS}}\,\{N_b\,M\,A\,B\}_{K_{BS}}$.
3. $M_3 = M\,\{N_a\,K_{AB}\}_{K_{AS}}\,\{N_b\,K_{AB}\}_{K_{BS}}$.
4. $M_4 = M\,\{N_a\,K_{AB}\}_{K_{AS}}$.

Figure 3. Message Exchange in Otway-Rees

The second corollary gives a condition under which encryption guarantees a non-penetrator origin.

Corollary 4.7 *Suppose $\mathcal{C}$ is a bundle; $\mathsf{K} = S \cup \mathsf{k}^{-1}$; $S \cap \mathsf{K}_{\mathcal{P}} = \emptyset$; and no regular node $\in \mathcal{C}$ is an entry point for $I_{\mathsf{k}}[S]$. Then any term of the form $\{g\}_K$ for $K \in S$ does not originate on a penetrator strand.*

PROOF. By Corollary 4.6, for every node $m \in \mathcal{C}$, term$(m) \notin I = I_{\mathsf{k}}[S]$. Suppose $t_1 = \{g\}_K$ for $K \in S$ originates on a penetrator strand m. By inspection, m cannot occur on a penetrator strand of kind **F**, **T**, **K**, **M**, **C** or **S**. Consider the remaining cases:

E. m occurs on a strand with trace $\langle -K_0, -h, +\{h\}_{K_0}\rangle$. Now $K_0 \notin I$ and so $K_0 \neq K$. Since $\{g\}_K \sqsubset \{h\}_{K_0}$, Corollary 3.14 implies $\{g\}_K \sqsubset h$, contradicting the definition of entry point.

D. m belongs to a strand with trace $\langle -{K_0}^{-1}, -\{h\}_{K_0}, +h\rangle$. If $\{g\}_K \sqsubset h$, then $\{g\}_K \sqsubset \{h\}_{K_0}$, contradicting the definition of entry point. ■

5 Otway-Rees: The Protocol

This protocol has three roles: initiator, responder, and server. The goal of the protocol is to mutually authenticate initiator and responder and to distribute a session key generated by a server. See Figure 3.

To provide a mathematical model of this protocol, we further refine the assumptions on the algebra A.

- A set $\mathsf{T}_{\text{name}} \subseteq \mathsf{T}$ of names.
- A mapping $K : \mathsf{T}_{\text{name}} \to \mathsf{K}$. This is intended to denote the mapping which associates to each principal the key it shares with the server. In the literature on this protocol this mapping is usually written using subscripts $K(A) = K_{AS}$. We assume the mapping $A \mapsto K_{AS}$ is injective. We also assume $K_{AS} = K_{AS}^{-1}$, i.e. that the protocol is using symmetric cryptography.

We will adopt some conventions on variables for the remainder of this section:

- Variables A, B range over T_{name};
- Variables K, K' range over K;
- Variables N, M (or the same letters decorated with subscripts) range over $\mathsf{T} \setminus \mathsf{T}_{\text{name}}$, i.e. those texts that are not names.

Other letters such as G and H range over all of A. We would emphasize that N_a is just a variable, having no reliable connection to A, whereas K_{AS} is the result of applying the function K to the argument A. Thus, the latter reliably refers to the long term key shared between A and S.

- Init$[A, B, N, M, K]$ is the set of strands $s \in \Sigma$ whose trace is

  $$\langle +\, M\,A\,B\,\{N\,M\,A\,B\}_{K_{AS}}, -\,M\,\{N\,K\}_{K_{AS}}\rangle$$

  Σ_{init} is the union of the range of Init.

- Resp$[A, B, N, M, K, H, H']$ is defined when $N \not\sqsubset H$; its value then is the set of strands in Σ whose trace is

  $$\begin{aligned}\langle &-\; M\,A\,B\,H,\\ &+\; M\,A\,B\,H\,\{N\,M\,A\,B\}_{K_{BS}},\\ &-\; M\,H'\,\{N\,K\}_{K_{BS}},\\ &+\; M\,H'\rangle\end{aligned}$$

  Σ_{resp} is the union of the range of Resp.

- Serv$[A, B, N_a, N_b, M, K]$ is defined if $K \notin \mathsf{K}_{\mathcal{P}}$, $K \notin \{K_{AS} : A \in \mathsf{T}_{\text{name}}\}$ and $K = K^{-1}$; its value then is the set of strands in Σ whose trace is:

  $$\begin{aligned}\langle &-\; M\,A\,B\,\{N_a\,M\,A\,B\}_{K_{AS}}\,\{N_b\,M\,A\,B\}_{K_{BS}},\\ &+\; M\,\{N_a\,K\}_{K_{AS}}\,\{N_b\,K\}_{K_{BS}}\rangle\end{aligned}$$

  Σ_{serv} is the union of the range of Serv.

The condition $N \not\sqsubset H$ in the definition of a responder strand ensures that the nonce N must originate on the strand Resp$[A, B, N, M, K, H, H']$ itself. A protocol participant cannot inspect the contents of H to enforce this condition, since under normal operation of the protocol, H is cyphertext inaccessible to the participant. Rather, we assume that this condition is enforced by a probabilistic mechanism.

Lemma 5.1 *If $f(\vec{v}) \cap f(\vec{v}') \neq \emptyset$, then $\vec{v} = \vec{v}'$, when f is one of the mappings* Serv, Init, Resp.

Lemma 5.2 *The sets $\Sigma_{serv}, \Sigma_{init}, \Sigma_{resp}$ are pairwise disjoint.*

PROOF. It suffices to prove the sets of traces are disjoint. Originator traces begin with a positive term. The second term of of a responder trace has width at least 4, whereas for a server trace the width is exactly 3.

Definition 5.3 *An* Otway-Rees strand space *is an infiltrated strand space Σ such that $\Sigma = \Sigma_{serv} \cup \Sigma_{init} \cup \Sigma_{resp} \cup \mathcal{P}$.*

This union is disjoint, by Lemma 5.2 and the observation that $\mathcal{P}$ contains no strands of the same form as $\Sigma_{\mathrm{serv}} \cup \Sigma_{\mathrm{init}} \cup \Sigma_{\mathrm{resp}}$.

Fix an Otway-Rees strand space Σ over A.

We sometimes find it convenient to use the $*$ to indicate union over some indices. Thus for instance Resp$[A, B, N_b, M, K, *, *] =$

$$\bigcup_{H, H'} \mathrm{Resp}[A, B, N_b, M, K, H, H']$$

6 Otway-Rees: Secrecy

We first prove that session keys distributed by the server cannot be disclosed unless the penetrator possesses one of the long-term keys used in the run. We show that a session key can never occur in a form in which it is not encrypted by the participants' long-term keys.

Theorem 6.1 *Suppose $\mathcal{C}$ is a bundle in Σ; $A, B \in \mathsf{T}_{name}$; K is uniquely originating; $K_{AS}, K_{BS} \notin \mathsf{K}_{\mathcal{P}}$; and $s_{serv} \in$* Serv$[A, B, N_a, N_b, M, K]$ *has $\mathcal{C}$-height 2.*

Let $S = \{K_{AS}, K_{BS}, K\}$ and $\mathsf{k} = \mathsf{K} \setminus S$. For every node $m \in \mathcal{C}$, term$(m) \notin I_{\mathsf{k}}[K]$.

PROOF. By Proposition 3.10, it suffices to prove the stronger statement that for every node m, term$(m) \notin I_{\mathsf{k}}[S]$. Since $S \cap \mathsf{K}_{\mathcal{P}} = \emptyset$, $\mathsf{k} = \mathsf{k}^{-1}$ and $\mathsf{K} = \mathsf{k} \cup S$, by Corollary 4.6 it suffices to show that no regular node m is an entry point for $I_{\mathsf{k}}[S]$.

We will argue by contradiction and assume m is a regular node which is an entry point for $I_{\mathsf{k}}[S]$. Since m is an entry point for $I_{\mathsf{k}}[S]$, by the definitions, it follows that term(m) is an element of $I_{\mathsf{k}}[S]$. By 3.10, this implies that one of the keys K, K_{AS}, K_{BS} is a subterm of term(m). Now no regular node contains any key of the form K_{XS} as a subterm. In fact the only keys which occur as subterms of term(m) for m regular, are the session keys emanating from a server. But by assumption the set of such keys is disjoint from the set of keys of the form K_{XS}. It thus follows K must be a subterm of term(m).

If m is a positive regular node on a strand s, then $K \sqsubset$ term(m) implies either:

1. $s \in \Sigma_{\mathrm{serv}}$ and $m = \langle s, 2 \rangle$, in which case K is the session key of s; or
2. $s \in$ Resp$[*, *, *, *, *, H, *]$, $m = \langle s, 2 \rangle$, and $K \sqsubset H$.

In case 2, m is not an entry point for $I_{\mathsf{k}}[S]$, because $H \sqsubset \langle s, 1 \rangle$, which is a preceding negative node.

So consider case 1. By the unique origination of K, $s = s_{\mathrm{serv}}$, so term$(m) = M\ \{N_a\, K\}_{K_{AS}}\ \{N_b\, K\}_{K_{BS}}$. By Proposition 3.12, either

1. $M \in I_{\mathsf{k}}[S]$, or
2. $\{N_a\, K\}_{K_{AS}} \in I_{\mathsf{k}}[S]$, or
3. $\{N_b\, K\}_{K_{BS}} \in I_{\mathsf{k}}[S]$.

But the first is impossible by Axiom 3.5; the second and third are impossible by Proposition 3.15. ■

7 Otway-Rees: Authentication

In this section we will prove the authentication guarantees that Otway-Rees provides to its initiator and responder. It is also possible to prove that the protocol provides authentication guarantees to the server [6], but we will not do so here. We first "import" the consequence of Corollary 4.6 that we will need to prove the authentication goals.

Proposition 7.1 *Consider a bundle $\mathcal{C}$ in Σ. Suppose $X \in \mathsf{T}_{name}$ is such that $K_{XS} \notin \mathsf{K}_{\mathcal{P}}$. Then no term of the form $\{g\}_{K_{XS}}$ for $X \in \mathsf{T}_{name}$ can originate on a penetrator node in $\mathcal{C}$.*

PROOF. Let $S = \{K_{XS}\}$ and $\mathsf{k} = \mathsf{K}$. To apply Corollary 4.7, we must check that no regular node is an entry point for $I_{\mathsf{K}}[S]$, or equivalently, that K_{XS} does not originate on any regular node.

A key K originates on a regular node only if it is a session key K originating on a server strand $s \in$ Serv$[*, *, *, *, K, *, *]$. However, by the definition of Σ_{serv}, the session key K is never a long term key K_{XS}.

Hence, we may apply Corollary 4.7 to $I_{\mathsf{K}}[S]$, so any term $\{g\}_{K_{XS}}$ can only originate on a regular node. ■

Proposition 7.2 *If $\{H\}_{K_{XS}}$ originates on a regular strand s, then:*

1. *If $s \in \Sigma_{serv}$, then $H = N\,K$.*
2. *If $s \in \Sigma_{init}$, then $H = N\,M\,X\,C$ for $X, C \in \mathsf{T}_{name}$.*
3. *If $s \in \Sigma_{resp}$, then $H = N\,M\,C\,X$ for $X, C \in \mathsf{T}_{name}$.*

PROOF. By the definition of originating (Definition 4.1), if the term $\{H\}_{K_{XS}}$ originates on m, then m is positive.

If $s \in \Sigma_{\text{init}}$ then $m = \langle s, 1\rangle$. Thus term(m) is of the form $M\,A\,B\,\{N\,M\,A\,B\}_{K_{AS}}$. The only encrypted subterm of this term, $\{N\,M\,A\,B\}_{K_{AS}}$, is of form 2.

If $s \in \Sigma_{\text{resp}}$, then the positive nodes of s are $\langle s, 2\rangle$ and $\langle s, 4\rangle$. The encrypted subterms of $\langle s, 2\rangle$ have plaintext of forms 2 and 3 respectively, while the encrypted subterm of $\langle s, 4\rangle$ has form 1.

A similar argument holds if $s \in \Sigma_{\text{serv}}$. ■

Corollary 7.3 *Suppose s is a regular strand of Σ.*

1. *If $\{N\,K\}_{K_{XS}}$ originates on s, then either*
   - $s \in$ Serv$[A, X, N, N', M, K]$
   - $s \in$ Serv$[X, B, N', N, M, K]$

   for some A, B, N', M. In either case the term originates on the node $\langle s, 2\rangle$ and K originates on s.
2. *If $\{N\,M\,A\,B\}_{K_{AS}}$ originates on s, with $A \neq B$ then*
   - $s \in$ Init$[A, B, N, M, K]$

   for some K. The term originates on the node $\langle s, 1\rangle$ and N originates on s.
3. *If $\{N\,M\,A\,B\}_{K_{BS}}$ originates on s, with $A \neq B$ then*
   - $s \in$ Resp$[A, B, N, M, K, H, H']$

   for some K, H, H'. The term originates on the node $\langle s, 2\rangle$ and N originates on s.

PROOF. Since s is regular, $s \in \Sigma_{\text{serv}} \cup \Sigma_{\text{init}} \cup \Sigma_{\text{resp}}$. Apply Propositions 7.2 and 2.7. ■

7.1 Initiator's Guarantee

The following theorem asserts that if a bundle contains a strand $s \in \Sigma_{\text{init}}$, then under reasonable assumptions, there are regular strands $s_{\text{resp}} \in \Sigma_{\text{resp}}$ and $s_{\text{serv}} \in \Sigma_{\text{serv}}$ which agree on the initiator, responder, and M values.

Theorem 7.4 *Suppose $\mathcal{C}$ is a bundle in Σ; $A \neq B$; N_a is uniquely originating in $\mathcal{C}$; and $K_{AS}, K_{BS} \notin \mathsf{K}_{\mathcal{P}}$.*

If $s \in$ Init$[A, B, N_a, M, K]$ has $\mathcal{C}$-height 2, then there are regular strands

- *$s_{resp} \in$ Resp$[A, B, N_b, M, *, *, *]$ of $\mathcal{C}$-height at least 2.*
- *$s_{serv} \in$ Serv$[A, B, N_a, N_b, M, K]$ of $\mathcal{C}$-height 2.*

PROOF. The assumption of the theorem means

$$\begin{aligned}\langle +\ & M\,A\,B\,\{N_a\,M\,A\,B\}_{K_{AS}},\\ -\ & M\,\{N_a\,K\}_{K_{AS}}\rangle\end{aligned}$$

is the $\mathcal{C}$-trace of a strand s.

Since $K_{AS} \notin \mathsf{K}_{\mathcal{P}}$, by Proposition 7.1, $\{N_a\,K\}_{K_{AS}}$ originates on a regular node in $\mathcal{C}$. By Corollary 7.3, this node belongs to a strand s_{serv} which satisfies one of the conditions:

1. $s_{\text{serv}} \in$ Serv$[A, X, N_a, N, M_1, K]$
2. $s_{\text{serv}} \in$ Serv$[X, A, N, N_a, M_1, K]$

where $X \in \mathsf{T}_{\text{name}}$, and $N, M_1 \in \mathsf{T}$. Since $\langle s_{\text{serv}}, 2\rangle \in \mathcal{C}$, s_{serv} has $\mathcal{C}$-height 2.

If condition 1 holds, $\{N_a\,M_1\,A\,X\}_{K_{AS}} \sqsubset$ term$(\langle s_{\text{serv}}, 1\rangle)$. By Proposition 7.1, $\{N_a\,M_1\,A\,X\}_{K_{AS}}$ originates on a regular strand s_1, and by Corollary 7.3, N_a originates on the same strand s_1. By the unique origination of N_a, $s_1 = s$. Thus $M_1 = M$ and $X = B$, and $s_{\text{serv}} \in$ Serv$[A, B, N_a, N, M, K]$.

By Proposition 7.1, $\{N\,M\,A\,B\}_{K_{BS}}$ originates on a regular node in $\mathcal{C}$. By Corollary 7.3, this node is the second on a strand $s_{\text{resp}} \in$ Resp$[A, B, N, M, *, *, *]$. Since $\langle s_{\text{resp}}, 2\rangle \in \mathcal{C}$, it follows s_{resp} has $\mathcal{C}$-height at least 2.

Suppose that condition 2 holds instead. Then $\{N_a\,M_1\,X\,A\}_{K_{AS}}$ is a subterm of term$(\langle s_{\text{serv}}, 1\rangle)$. By Proposition 7.1, $\{N_a\,M_1\,X\,A\}_{K_{AS}}$ originates on a regular strand s_1, and by Corollary 7.3, N_a originates on the same strand s_1. By the unique origination of N_a, $s_1 = s$. Hence by Corollary 7.3, $\{N_a\,M_1\,X\,A\}_{K_{AS}} = \{N_a\,M_1\,A\,B\}_{K_{AS}}$, so $A = B$, contradicting an assumption. ■

Remarks. Even though the intention of the protocol design is to have B receive $H = \{N_a\,M\,A\,B\}_{K_{AS}}$ from A there is no way to prevent a penetrator from replacing $\{N_a\,M\,A\,B\}_{K_{AS}}$ with garbage. Moreover a penetrator can prevent the output of the server from reaching B. Thus, we cannot show that B has $\mathcal{C}$-height > 2.

7.2 Responder's Guarantee

The responder can rest assured that if a bundle contains a strand $s \in \Sigma_{\text{resp}}$, then under familiar assumptions there are regular strands $s_{\text{init}} \in \Sigma_{\text{init}}$ and $s_{\text{serv}} \in \Sigma_{\text{serv}}$ which agree on the initiator, responder, and M values. Its proof is very similar to the proof of Theorem 7.4.

Theorem 7.5 *Suppose $\mathcal{C}$ is a bundle in Σ; $A \neq B$; N_b is uniquely originating in $\mathcal{C}$; and $K_{AS}, K_{BS} \notin \mathsf{K}_{\mathcal{P}}$.*

If $s \in$ Resp$[A, B, N_b, M, K, H, H']$ has $\mathcal{C}$-height at least 3, then there are regular strands

- $s_{init} \in$ Init$[A, B, *, M, *]$ *of $\mathcal{C}$-height at least 1.*
- $s_{serv} \in$ Serv$[A, B, *, N_b, M, K]$ *of $\mathcal{C}$-height 2.*

PROOF. The assumption of the proposition means the $\mathcal{C}$-trace of s contains at least:

$$\begin{aligned}\langle &- M\,A\,B\,H,\\ &+ M\,A\,B\,H\,\{N_b\,M\,A\,B\}_{K_{BS}},\\ &- M\,H'\,\{N_b\,K\}_{K_{BS}}\rangle\end{aligned}$$

Since $K_{BS} \notin \mathsf{K}_{\mathcal{P}}$, by Proposition 7.1, $\{N_b\,K\}_{K_{BS}}$ originates on a regular node in $\mathcal{C}$. By Corollary 7.3, this node belongs to a strand s_{serv} which satisfies one of the following two conditions:

1. $s_{\text{serv}} \in$ Serv$[B, X, N_b, N, M_1, K]$
2. $s_{\text{serv}} \in$ Serv$[X, B, N, N_b, M_1, K]$

where $X \in \mathsf{T}_{\text{name}}$, and $N, M_1 \in \mathsf{T}$. Since $\langle s_{\text{serv}}, 2\rangle \in \mathcal{C}$, s_{serv} has $\mathcal{C}$-height 2.

If condition 1 holds, then $\{N_b\,M_1\,B\,X\}_{K_{BS}} \sqsubset \langle s_{\text{serv}}, 1\rangle$. By Proposition 7.1, $\{N_b\,M_1\,B\,X\}_{K_{BS}}$ originates on a regular strand s_1, and by Corollary 7.3, N_b originates on the strand s_1. By the unique origination of N_b, $s = s_1$. Hence $\{N_b\,M_1\,B\,X\}_{K_{BS}} = \{N_b\,M_1\,A\,B\}_{K_{BS}}$, so that $B = A$, contradicting an assumption.

Suppose that condition 2 holds instead. Again, $\{N_b\,M_1\,X\,B\}_{K_{BS}} \sqsubset \langle s_{\text{serv}}, 1\rangle$. By Proposition 7.1, $\{N_b\,M_1\,X\,B\}_{K_{BS}}$ originates on a regular strand s_1, and by Corollary 7.3, N_b originates on the strand s_1. By the unique origination of N_b, $s_1 = s$. Thus, $M_1 = M$ and $X = A$, and $s_{\text{serv}} \in$ Serv$[A, B, N, N_b, M, K]$.

By Proposition 7.1, $\{N\,M\,A\,B\}_{K_{AS}}$ originates on a regular node in $\mathcal{C}$. By Corollary 7.3, this node belongs to a strand $s_{\text{init}} \in$ Init$[A, B, N, M, *]$. s_{init} has $\mathcal{C}$-height at least 1. ■

Remarks. As in the previous theorem there are some penetrator behaviors that cannot be prevented. For instance the penetrator could take the encrypted session key that B is supposed to pass on to A and throw it away. Hence, we can not show that the initiator's strand has $\mathcal{C}$-height > 1.

More significantly, the above argument makes vividly clear why the BAN modification to Otway-Rees [1, Section 4] might fail, as was shown by Mao and Boyd [3]. In that modification the nonce N_b is outside the encryption. Though it is still true, when condition 2 holds, that the term $\{M_1\,X\,B\}_{K_{BS}}$ originates on a regular strand s_1, this term does not contain N_b. Hence, s_1 may not be an origination point for N_b, and we can no longer conclude that $s_1 = s$.

Indeed, the BAN modification also requires a weakening of Theorem 7.4, as we can no longer infer that the responder and the server strands will agree on the responder's nonce N_b.

7.3 The Missing Guarantee

The authentication theorems do not establish something that we had expected they would, namely that if a bundle $\mathcal{C}$ contains complete initiator and responder strands, then they agree on the session key distributed.

That is, one cannot strengthen Theorem 7.4 by replacing the asterisk by K to obtain $s_{\text{resp}} \in$ Resp$[A, B, N_b, M, K, *, *]$. Nor can one strengthen Theorem 7.5 by replacing an asterisk by K to obtain $s_{\text{init}} \in$ Init$[A, B, *, M, K]$. The reason is that there is a counterexample, a bundle $\mathcal{C}$ (illustrated in Figure 4) in which each player has a complete strand in $\mathcal{C}$, and they agree on A, B, and M, but they do not agree on K.

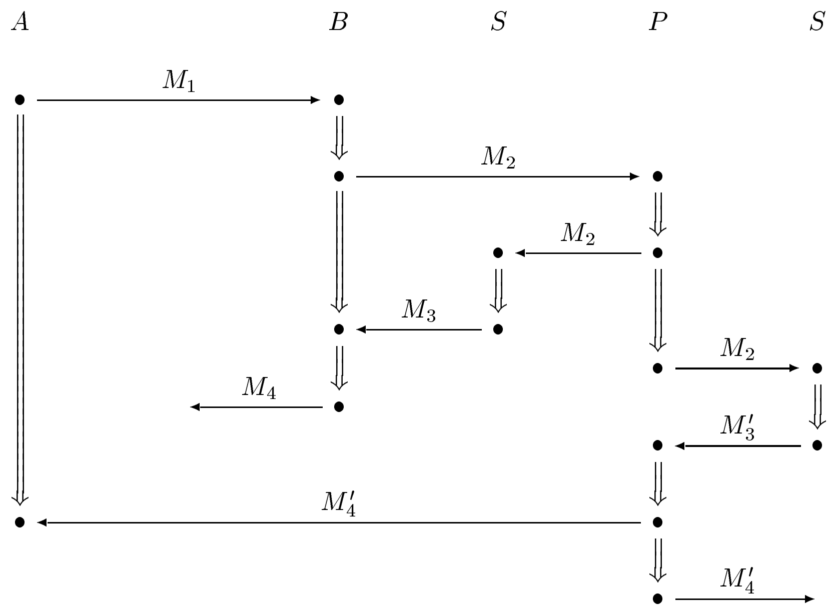
Although this protocol has been studied very carefully in the past (e.g. [1, 3, 5]), this weakness appears not to be explicit in the literature. For instance, the BAN authors [1, Section 4] suggest the contrary, that the two participants at the end each believe of a (single) key K_{AB} that it is a good shared key for A and B. The authors comment that neither principal can know whether the key is known to the other, but this is presumably because neither principal knows whether the other has completed his strand. Paulson [5], despite his very detailed argument, does not comment on this point.

Presumably this protocol weakness is not serious, as no shared keys are disclosed. However, it serves to illustrate the subtleties that remain poorly understood even in very familiar protocols.

8 Discussion

This paper, an extension of [7], has served two purposes.

First, we have developed new algebraic machinery—the notion of ideal—to supplement the strand space idea, and to prove general, re-usable bounds on the penetrator (Sections 3–4). Our methods exploit two partial orderings, namely the subterm relation $\sqsubset$ between terms and the $\preceq$ relation between nodes. Inductive characteristics of these orderings are formulated via the notion of an ideal in the case of $\sqsubset$, and via a least element principle in the case of $\preceq$. In addition, the strand space machinery, together with our treatment of unique origination, provides great power for localizing the crucial steps in potential attacks. One knows on which strand a particular event must occur, and the form of the term at the relevant node. This gives finer grained



Where

1. $M_1 = M\ A\ B\ \{N_a\ M\ A\ B\}_{K_{AS}}$.
2. $M_2 = M\ A\ B\ \{N_a\ M\ A\ B\}_{K_{AS}}\ \{N_b\ M\ A\ B\}_{K_{BS}}$.
3. $M_3 = M\ \{N_a\ K_{AB}\}_{K_{AS}}\ \{N_b\ K_{AB}\}_{K_{BS}}$.
4. $M_3' = M\ \{N_a\ K'_{AB}\}_{K_{AS}}\ \{N_b\ K'_{AB}\}_{K_{BS}}$.
5. $M_4 = M\ \{N_a\ K_{AB}\}_{K_{AS}}$.
6. $M_4' = M\ \{N_a\ K'_{AB}\}_{K_{AS}}$.

Figure 4. An Otway-Rees Weakness: Mismatched Keys

control over the analysis than other methods seem to us to provide.

Second, we have used our methods to provide simple and revealing proofs about a particular protocol (Sections 5–7). These proofs show that even in the case of a very well-studied protocol, there remain fine points that have not been understood.

The specific algebraic properties we have considered are still elementary. They are applied under assumptions (such as "free encryption") that are still restrictive. However, it is likely that the approach can be used in the case of message algebras with less restrictive assumptions.

Acknowledgements. We are grateful to Sylvan Pinsky and his colleagues at NSA for support, encouragement, and discussions. Shim Berkovits, Marion Michaud, and John Vasak patiently helped us improve the presentation. Peter Ryan taught us a great deal about the field, and provided the initial stimulus for doing the work.

We are also grateful to the anonymous referees for their meticulous and insightful comments.

References

[1] Michael Burrows, Martín Abadi, and Roger Needham. A logic of authentication. *Proceedings of the Royal Society*, Series A, 426(1871):233–271, December 1989. Also appeared as SRC Research Report 39 and, in a shortened form, in ACM Transactions on Computer Systems 8, 1 (February 1990), 18-36.

[2] Ulf Carlsen. Cryptographic protocol flaws. In *Proceedings 7th IEEE Computer Security Foundations Workshop*, pages 192–200. IEEE Computer Society, 1994.

[3] Wenbo Mao and Colin Boyd. Towards the formal analysis of security protocols. In *Proceedings of the Computer Security Foundations Workshop VI*, pages 147–158. IEEE Computer Society Press, 1993.

[4] Will Marrero, Edmund Clarke, and Somesh Jha. A model checker for authentication protocols. In Cathy Meadows and Hilary Orman, editors, *Proceedings of the DIMACS Workshop on Design and Verification of Security Protocols*. DIMACS, Rutgers University, September 1997.

[5] Lawrence C. Paulson. Proving properties of security protocols by induction. In *10th IEEE Computer Security Foundations Workshop*, pages 70–83. IEEE Computer Society Press, 1997.

[6] F. Javier Thayer Fábrega, Jonathan C. Herzog, and Joshua D. Guttman. Strand spaces. Technical report, The MITRE Corporation, November 1997.

[7] F. Javier Thayer Fábrega, Jonathan C. Herzog, and Joshua D. Guttman. Strand spaces: Why is a security protocol correct? In *1998 IEEE Symposium on Security and Privacy*, 1998.