ProVerif Analysis of the ZRTP Protocol

Riccardo Bresciani and Andrew Butterfield
Foundations and Methods Group, Trinity College Dublin
Lero – the Irish Software Engineering Research Centre
bresciar@scss.tcd.ie, Andrew.Butterfield@scss.tcd.ie

Abstract

When some agents want to communicate through a media stream (for example voice or video), the Real Time Protocol *(RTP) is used. This protocol does not provide encryption, so it is necessary to use* Secure RTP *(SRTP) to secure the communication. In order for this to work, the agents need to agree on key material and ZRTP provides them with a procedure to perform this task: it is a key agreement protocol, which relies on a Diffie-Hellman exchange to generate SRTP session parameters, providing confidentiality and protecting against* Man-in-the-Middle *attacks even without a public key infrastructure or endpoint certificates. This is an analysis of the ZRTP protocol performed with ProVerif, which tests security properties; in order to perform the analysis, the protocol has been modeled in the applied π-calculus.*

1. Introduction

In recent years research has strongly focused on proofs of security. The verification step to ensure that a computer program or a protocol has certain requested properties is a crucial one, and this task should ideally be done by formal reasoning, rather than by tests and simulations, as the latter approach is not as exhaustive as the formal one.

There are two possible approaches to protocol verification: the formal model and the computational model. In the first model, we are in a highly idealized setting, therefore this can be effectively implemented in fully-automated protocol verifiers. The second approach borrows ideas from complexity theory and requires much more human intervention in proofs, and it is only recently being automated. [5]

These verification techniques allow us to uncover design faults that may remain hidden for years. There are a lot of examples that can be recalled on this topic, for example a successful application of verification in the formal model can be found in [9, 10]: the popular Needham-Schroeder protocol dates back to 1978, but it was just in 1995 that Gavin Lowe found that an attack on the protocol was possible and proposed a modification. To achieve this goal Gavin Lowe used the tool FDR, which is a model checker for CSP.

Besides generic model checkers such as FDR, there are tools which have been conceived with communication protocols in mind. In the present paper we use Bruno Blanchet's ProVerif [1, 3]: if the original Needham-Schroeder protocol is analysed with this tool, this same security flaw can be uncovered and a trace of the attack given.

The purpose of the present paper is to present the results of a proof of security for the ZRTP protocol in the formal model.

This protocol has been submitted as a RFC to the IETF by Philip Zimmermann, Alan Johnston and Jon Callas [12] and has been *tested* in its reference *implementation* — the open source software *Zfone* — by Detica Forensics, which undertook a basic *black block* analysis [8].

Our goal is to have the protocol *verified* at the higher level of its *specification*.

1.1. The ZRTP Protocol

ZRTP is run to provide key agreement and parameter negotiation to establish an SRTP session — SRTP (*Secure RTP*) is a secure profile of RTP (*Real-time Transport Protocol*), that deals with security and privacy issues that are not built-in to RTP. Agents wishing to communicate by means of SRTP must agree on session keys and parameters in order to establish a secure session: these negotiations may be effectively made through ZRTP.

ZRTP bases the key agreement procedure on a Diffie-Hellman exchange and on cached secrets established in preceding sessions (if any): this creates a new shared secret, from which all key material can be derived by means of one-way functions.

In case no valid secret is found in the cache (or in cacheless implementations), the protocol is vulnerable to a *Man-in-the-Middle* attack. To ensure that this attack has not been performed, ZRTP provides a method to detect it: the agents have a *short authentication string* (*SAS*), which is a one-

Copyright © 2010, Infonomics Society

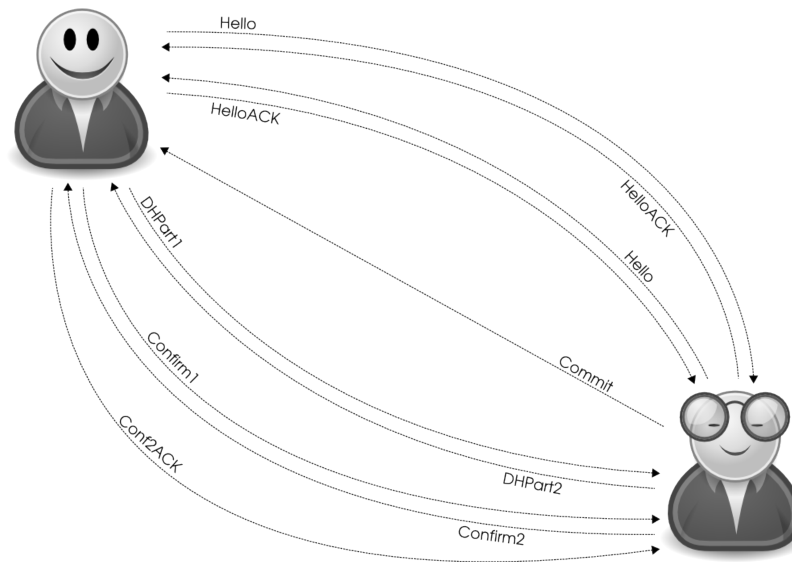

Figure 1. Key agreement call flow.

way function of the ZRTP session key and that can be either verbally compared or validated by an optional signature. If the short authentication strings do not match, an attack may be taking place.

Keying material is destroyed at the end of each session, thus ZRTP offers perfect forward secrecy.

Some key continuity is kept by means of cached secrets: after each successful key agreement the cache is updated with a secret, which is a one-way function of the new secret generated through the key agreement procedure.

One key feature of ZRTP is that it does not rely on SIP signaling for key management, on any server or on any kind of public key infrastructure: only the endpoints have to interact for the key agreement to be performed.

ZRTP provides also protection against *Denial-of-Service* attacks, as it offers a way to detect and reject false ZRTP messages.

2. Protocol Description

In this section we will provide a brief description of the protocol, in order to show the way it works and how it enables two agents to agree on the key material and parameters needed for an SRTP session.

ZRTP has three possible working modes:

- the *Diffie-Hellman mode* is based on a Diffie-Hellman exchange: all SRTP keys are computed from the secret value computed by each party;
- the *Multistream mode* is usable only if there is already an active SRTP session between the endpoints: new SRTP keys for a new stream can be derived from a the preceding Diffie-Hellman exchange, avoiding the expensive computations of a new one;
- the *Preshared mode* does not rely on a Diffie-Hellman exchange, but on previously cached secrets only. This is secure as long as the secret cache is not corrupted. Indeed, even in this case, it mantains the perfect forward secrecy of the protocol, as keying material is deleted as soon as each session is terminated.

The present paper addresses the Diffie-Hellman mode only, as it is the setting where an attack can be performed: if no attack has succeeded in this session, the agents share reliable secrets, therefore all subsequent sessions in Multistream or Preshared mode that rely on them are safe, under the assumption that integrity of the secret cache is preserved.

During the run of the protocol two agents exchange messages: the *initiator* and the *responder*.

The key agreement and negotiation algorithm can be divided into 4 steps (see Figure 1):

- discovery — the agents exchange information about their identity and the supported session parameters;
- hash committment — the *initiator* starts the key agreement procedure;
- Diffie-Hellman exchange and key derivation — public keys are exchanged;
- confirmation — the endpoints acknowledge the successful key agreement.

These steps are presented in more detail in the remainder of this section.

Copyright © 2010, Infonomics Society

2.1. Discovery

During the discovery phase the initiator and the responder exchange their ZRTP identifiers (`ZIDX` is a unique 96 bit random identifier, generated only once when the client is set up for the first time). Additionally, they gather information about each other's capabilities, in terms of supported ZRTP versions, hash functions, ciphers, authorization tag lengths, key agreement types, and *SAS* algorithms.

The messages exchanged during this phase are called `Hello` messages. An acknowledgement is sent upon receipt of each of these messages: this is the `HelloACK` message — the initiator may skip sending his acknowledgement and reply immediately with a `Commit` message, explained in the following subsection.

2.2. Hash commitment

The next step towards the negotiation of the key material is made with the hash commitment: besides containing all the session parameters that will have to be used and defining the roles in the protocol (it is symmetrical in the discovery phase, and the agent sending this message is the one who is willing to act as the initiator), it contains a value that commits the initiator not to change his Diffie-Hellman key pair.

In fact the initiator creates his key pair $(\mathtt{svI}, \mathtt{pvI})$ prior to sending the hash commitment:

$$\mathtt{pvI} = g^{\mathtt{svI}} \bmod p$$

But the initiator cannot reveal this immediately, as doing so could enable the other party (or an attacker) to choose maliciously his keys depending on the initiator's choice. The solution is to send a hash of the key, concatenated to a hash of the message sent by the responder during the discovery phase:

$$\mathtt{hvI} = \mathcal{H}(\text{Initiator's } \mathtt{DHPart2} | \text{Responder's } \mathtt{Hello})$$

The message hash protects the protocol against *bid-down attacks*, that aim at making the agents rely on weaker algorithms, as an attacker may alter the information on supported session parameters. Finally the hash commitment prevents the agents from being able to influence deterministically the *SAS*: it is a function of the exchanged messages, so it can be influenced by an opportunistic choice of the agents' key — this cannot be done as the responder chooses his keys before knowing the initiator keys, and the initiator chooses his key before sending the hash commitment, that binds him to that choice.

2.3. Diffie-Hellman exchange and key derivation

After the hash commitment the agents can perform the Diffie-Hellman exchange, which will enable the agents to derive a new shared secret `s0`: the messages exchanged in this phase are `DHPart1` from the responder and `DHPart2` from the initiator. Different elements contribute to this secret and are contained in these messages — explicitly or implicitly, by means of a *Hash-based Message Authentication Code* (HMAC).

The first of these elements is the Diffie-Hellman result `dhr`, that the agents compute by modular exponentiation of the other party's public key to the power of their own private key — thus the value they have computed is known only to the two of them:

$$\begin{aligned} \mathtt{dhr} &= \mathtt{pvI}^{\mathtt{svR}} \bmod p \\ &= (g^{\mathtt{svI}} \bmod p)^{\mathtt{svR}} \bmod p \\ &= g^{\mathtt{svR}\cdot\mathtt{svI}} \bmod p \\ &= (g^{\mathtt{svR}} \bmod p)^{\mathtt{svI}} \bmod p \\ &= \mathtt{pvR}^{\mathtt{svI}} \bmod p \end{aligned}$$

Along with the public keys, the agents send also a HMAC — $\mathcal{H}_M(k, s)$, where k is the key and s is the string on which the key is applied — for each secret `secretX` that they already share (a maximum of two retained secrets in the cache and two other optional secrets that depend on the environment where ZRTP is running): this allows them to distinguish matching secrets from non-matching ones (and discard them):

$$\begin{aligned} \mathtt{secretX\text{-}idR} &= \mathcal{H}_M(\mathtt{secretX}, \mathtt{"Responder"}) \\ \mathtt{secretX\text{-}idI} &= \mathcal{H}_M(\mathtt{secretX}, \mathtt{"Initiator"}) \end{aligned}$$

The matching shared secrets will be concatenated with the hash of all the exchanged messages — the hash of this concatenation[1] will be the new shared secret between the agents:

$$\begin{aligned} \mathtt{mh} =& \mathcal{H}(\text{Responder's } \mathtt{Hello} | \mathtt{Commit} | \\ & \qquad | \mathtt{DHPart1} | \mathtt{DHPart2}) \\ \mathtt{s0} =& \mathcal{H}(\mathtt{dhr} | \mathtt{mh} | \mathtt{s1} | \mathtt{s2} | \mathtt{s3}) \end{aligned}$$

All the key material needed to establish a SRTP session will be derived from this shared secret by means of a variation of the HMAC function, keyed with different strings depending on the particular key to be generated. Also the *SAS* is derived in this way. Once these operations have been completed, the endpoints exchange the confirmation messages to acknowledge that the key agreement procedure has been successful. They contain an encrypted block, which is encrypted using the newly generated keys: verifying the secrecy of this block will be the way to ensure that the protocol is safe.

[1] Actually some other constant parameters are also concatened in this hash, following the requirements listed in NIST SP800-56A: these extra parameters are omitted for the sake of simplicity. See [12] for more details.

Copyright © 2010, Infonomics Society

2.4. Confirmation

The endpoints can now use s0 to generate a ZRTP session key and SRTP master keys and salts — separate in each direction for each media stream — using the key derivation function, which is an HMAC function taking the length (where the obtained values should be truncated) and the KDFContext (defined as the concatenation of ZIDI, ZIDR and mh) as extra parameters: this provides keys of the length required by the chosen SRTP algorithm.

For the sake of simplicity these parameters will always be omitted and we note this function simply as $\mathcal{K}$: the purpose of the present work is a formal proof of security, where computational aspects are not taken into account, so the functions $\mathcal{K}$ and $\mathcal{H}_M$ are substantially the same thing, as the length parameter is irrelevant from a non-computational point of view and the value of KDFContext is publicly known.

Each session has a unique identifier, computed using the key derivation function, where the usual argument KDFContext is replaced by the concatenation of ZIDI and ZIDR — for this reason we note this function as $\mathcal{K}' \neq \mathcal{K}$ in the expression of ZRTPSess:

$$\texttt{ZRTPSess} = \mathcal{K}'(\texttt{s0}, \texttt{"ZRTP Session Key"})$$

Next they compute their HMAC keys (hmackey) and the new retained secret rs0:

$$\texttt{hmackeyR} = \mathcal{K}(\texttt{s0}, \texttt{"Responder HMAC key"})$$
$$\texttt{hmackeyI} = \mathcal{K}(\texttt{s0}, \texttt{"Initiator HMAC key"})$$
$$\texttt{rs0} = \mathcal{H}_M(\texttt{s0}, \texttt{"retained secret"})$$

After this, the agents generate the ZRTP keys, which will be destroyed only at the end of the call signaling session: this will allow ZRTP Preshared mode to generate new SRTP key-salt pairs for new concurrent media streams between the same endpoints, within the limit of the call signaling session (*i.e.* in case of separate calls, each call has its own ZRTP keys).

$$\texttt{ZRTP-keyR} = \mathcal{K}(\texttt{s0}, \texttt{"Responder ZRTP Key"})$$
$$\texttt{ZRTP-keyI} = \mathcal{K}(\texttt{s0}, \texttt{"Initiator ZRTP Key"})$$

After this s0 is deleted: in the protocol draft the authors put great emphasis on the importance of deleting s0 as soon as it is no longer needed, as this prevents the possibility of recreating the keys — this is important in case something may go wrong in that session, for example an attacker somehow gaining access to that value. All other key material will be deleted as soon as it is no longer used, in any case no later than the end of the session.

The agents compute the *SAS* value to be able to make sure that no *Man-in-the-Middle* attack has taken place:

$$\texttt{sashash} = \mathcal{K}(\texttt{ZRTPSess}, \texttt{"SAS"})$$
$$\texttt{sasvalue} = \text{[Rightmost 32 bits of]}\ \texttt{sashash}$$

Finally the agents can send the confirmation messages Confirm1 and Confirm2, which are exchanged essentially for three reasons:

1. they confirm that the whole key agreement procedure was successful and encryption is working, and they enable automatic detection of *Man-in-the-Middle attacks*;
2. they allow the CFB-encrypted transmission of the *SAS* Verified flag (V), so that no passive observer can learn whether the agents have the good habit of verifying the *SAS*;
3. they allow the CFB-encrypted transmission of the hash image H0 (see subsection 2.5);
4. they may contain an optional signature for the validation of *SAS*.

The confirmation messages contain a ciphered part composed by the cache expiration interval for rs0, an optional signature and an 8 bit unsigned integer, which contains — among other flags — the *SAS Verified flag*.

The encrypted part of a confirmation message is ciphered via the CFB algorithm, using the computed ZRTP key: its initialization vector is sent in the message, along with an HMAC (hmac) covering the encrypted part:

$$\texttt{hmac} = \mathcal{H}_M(\texttt{hmackeyX}, \text{Encrypted part of } \texttt{ConfirmX})$$

The responder sends the acknowledgment message Conf2ACK upon receipt of the confirmation message from the initiator.

After the confirmation procedure, both parties discard the rs2 secret and replace it by the rs1, rs1 by rs0.

It must be noted that if one endpoint fails to update the secret cache, there still could be a secret to rely on for a subsequent key agreement: the non-updated rs1 will match the updated rs2, if the rs1s were matching in the current key agreement.

2.5. Protection of the message exchange

The message exchange is protected with a chain of HMACs that cover each message: each HMAC is keyed with a value that is transmitted in clear only in the following message (thus it can be verified only in that moment). Moreover the values used as keys for the HMACs are generated from an 8-word nonce by subsequent hashing: for this reason they are referred to as *hash images*. The coherency of a hash image can be verified upon receipt of the following one. The original 8-word nonce is transmitted in the encrypted part of the confirmation messages.

Copyright © 2010, Infonomics Society

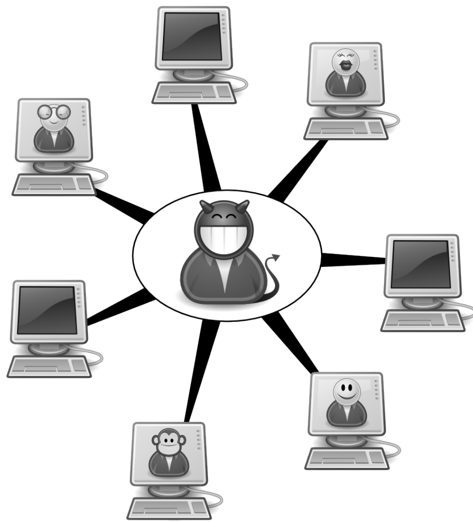

Figure 2. The Dolev-Yao model.

3. Analysis

In the proposed analysis the protocol is modeled in the Applied π-calculus [2]: the agents taking part in the protocol are expressed as two concurrent processes. The agents interact in a scenario, which is normally referred to as *Dolev-Yao model*, described in the following subsection.

3.1. The Dolev-Yao model

The Dolev-Yao model [7], schematised in Figure 2, assumes that:

- the net is under the intruder's control: messages can be intercepted and altered. New messages can be injected to the net;
- the cryptographic primitives are perfect;
- the protocol admits any number or participants and any number of parallel sessions;
- the protocol messages can be of any size.

The above formal model can be effectively captured by automatic protocol verifiers and it is much easier to be implemented than computational models: most automatic proofs on protocols have been done in this model.

In this model we can reason about an idealized version of the protocol, so we can abstract from the implementation issues: for example a flaw in an implementation of a protocol due to overflow will not be detected in the formal model, but a flaw due to misconception of the protocol will be found by a protocol verifier.

3.2. Applied π-calculus and ProVerif

In order to reason about cryptographic protocols, Martín Abadi and Cédric Fournet have built the applied π-calculus [2] on top of Milner's π-calculus [11]: the main thing is that names are replaced by terms (the atomic values of the π-calculus are not enough to deal efficiently with the complexity of a cryptographic protocol). By using equational theories, it is possible to take full advantage of this calculus to test security properties of communication protocols, as they allow us to account for equational properties of the functions used in the protocols.

The syntax of the applied π-calculus is shown in Figure 3.

Once the protocol has been modeled in the Applied π-calculus (see Figures 4 and 5), the analysis can be performed with the tool ProVerif [1, 3] and will provide a formal proof of security for the model.

ProVerif translates the Applied π-calculus process that describes the protocol into a set of Horn clauses, that account for the initial knowledge of the attacker, for the inference rules he can apply to broaden his knowledge pool and for the messages that can be sent over the communication

TERMS M, N :==

Variables x,y,z

Names a,b,c,k,s

Constructor $f(M_1, \ldots, M_n)$

PROCESSES P, Q :==

Output $\overline{M}\langle N \rangle.P$
The term N is output to channel M, then P

Input $M(x).P$
A value is input on channel M and bound to x, then P

Destructor $let\ x = g(M_1, \ldots, M_n)\ in\ P\ else\ Q$

Conditional $if\ M = N\ then\ P\ else\ Q$

Nil process 0

Parallel $P|Q$

Replication $!P$

Restriction $(\nu a).P$
Declare a as a new locally scoped name, then P

Figure 3. The syntax of the applied π-calculus.

Copyright © 2010, Infonomics Society

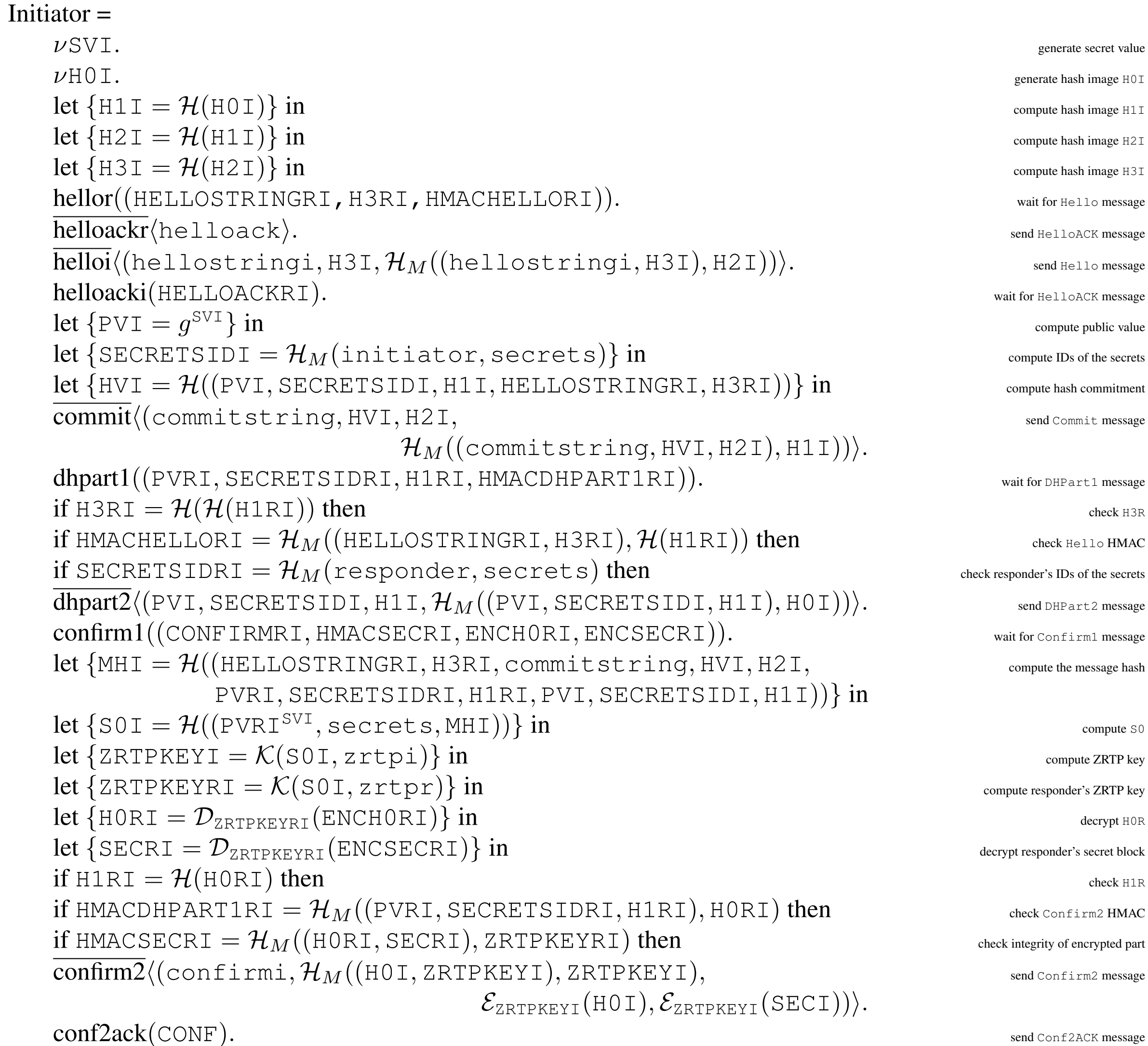

Figure 4. The "Initiator" process

channels.

These clauses are built on two basic predicates, that state either that the attacker knows a certain term or that a certain message can be sent.

In general ProVerif can be used to verify trace and equivalence properties of protocols. Among the trace properties we are interested in testing secrecy properties, *i.e.* whether a Dolev-Yao attacker is able to derive a term from the messages exchanged among the agents.

The resolution process consists of applying a resolution algorithm to this set of rules and deriving new clauses, which must not allow an attacker to compromise the protocol — for example we cannot derive a clause stating that a secret term is sent unencrypted on a public channel.

In the case of the present paper, the goal is to prove that secrecy of data encrypted under the key, on which the agents have agreed by means of the protocol, is preserved and data is not disclosed to an attacker.

3.3. Protocol model and results

The protocol has been modeled in the following way:

- there is no mismatch in the secrets: the key agreement procedure can rely on this for key generation. This is ideally the typical run of the protocol, when *SAS* has been verified in the very first session between the agents and the secret cache has been correctly updated in each subsequent session;
- publicly known dummy constants have been used for

Copyright © 2010, Infonomics Society

Figure 5. The "Responder" process

what does not concern security;

- no negotiations are done during the discovery phase, thus the hash function is predefined and publicly known, as well as encryption algorithms, ZRTP version and so on.

We challenge the adversary to derive the terms that are sent encrypted under the negotiated key in the confirmation messages: if there is no way that an adversary can derive them by applying the rules, then the protocol is safe, as this means that the key agreement procedure has not been compromised and thus the key negotiated between the endpoints is a safe one.

Among the functions that will be used in the protocol there are one-way functions, such as the hashing function $\mathcal{H}$, the function to compute HMACs $\mathcal{H}_M$ and the key derivation function $\mathcal{K}$: for these functions only a constructor is declared. The lack of the appropriate destructor makes it impossible to recover the argument passed to any of these functions.

The encryption functions are different, as a destructor is declared: when a message is encrypted under the key k via the function $\mathcal{E}$, it can be recovered by using the function $\mathcal{D}$, provided that the correct key k is passed to this function.

Finally the model is equipped with an equational theory that accounts for the commutative property of the exponen-

Copyright © 2010, Infonomics Society

tial:

$$(g^x)^y = (g^y)^x$$

The messages are distinguished one from the other by having a different channel for each message type: channels are declared as free names and they belong to the initial knowledge of the attacker, *i.e.* any data flowing through these channels is knowable by the attacker. Since the beginning the attacker also knows any constant used in the protocol, such as the base of the exponentials or the constant strings. The terms to be sent under encryption are declared as *private* free names: this means that they do not belong to the initial knowledge pool of the attacker, *i.e.* there will be no Horn clause stating that the attacker knows those free names. ProVerif shows the protocol to be secure in a Dolev-Yao network, as the attacker cannot derive these terms: if the key agreement procedure can be performed, then we have the formal proof that an attacker cannot have compromised it and have broken into the session.

4. Conclusions

In the present paper the protocol run has been modeled as two concurrent processes that interact by exchanging messages, synchronizing on every message exchange.

The model does not bother with all the negotiation procedure of the discovery phase, as this is unessential to prove the security of the protocol: according to the Dolev-Yao model, the cryptographic functions are idealized, so every algorithm is just as strong as any other; moreover the chosen algorithms are publicly known, as they are sent in clear in the `Commit` message.

The analysis performed on the protocol has formally proven that ZRTP is a safe key agreement protocol: two endpoints that use it to agree on a key can be sure that their communications are secured against any attack. For this to happen it is crucial that there are some pre-shared secrets: if this is not the case, ProVerif shows that a *Man-in-the-Middle attack* is possible. This is the reason why one needs to use *SAS* to ensure that this attack has not been performed on the first session between the two agents: in this session a reliable shared secret will be created, and therefore all the subsequent sessions will be secured.

It must be noted that this is true under the assumption that *SAS* provides an effective way to detect the presence of an attacker. [6]

More in general, the present paper highlights the benefits of using the applied π-calculus and ProVerif to reason about cryptographic protocols: the model of the protocol accounts for all the peculiarities of a typical run of the ZRTP protocol and therefore provides a good support for reasoning about ZRTP, in view of future modifications and improvements.

5. Acknowledgements

The present work has emanated from research conducted with the financial support of *Science Foundation Ireland*, grant 08-RFP-CMS1277.

The authors wish to thank *Bruno Blanchet*, *Steve Kremer* and *Phil Zimmermann* for their helpful comments and suggestions on earlier work that has lead to this paper.

6. References

[1] M. Abadi and B. Blanchet. Analyzing Security Protocols with Secrecy Types and Logic Programs. *Journal of the ACM*, 52(1):102–146, Jan. 2005.

[2] M. Abadi and C. Fournet. Mobile values, new names, and secure communication. In *POPL '01: Proceedings of the 28th ACM SIGPLAN-SIGACT symposium on Principles of programming languages*, pages 104–115, New York, NY, USA, 2001. ACM.

[3] B. Blanchet. An efficient cryptographic protocol verifier based on prolog rules. In *14th IEEE Computer Security Foundations Workshop*, pages 86–100, 2001.

[4] B. Blanchet. An Efficient Cryptographic Protocol Verifier Based on Prolog Rules. In *14th IEEE Computer Security Foundations Workshop (CSFW-14)*, pages 82–96, Cape Breton, Nova Scotia, Canada, 2001. IEEE Computer Society.

[5] B. Blanchet. *Vérification automatique de protocoles cryptographiques : modèle formel et modèle calculatoire.* Mémoire d'habilitation à diriger des recherches, Université Paris-Dauphine, 2008.

[6] R. Bresciani. The ZRTP protocol — Analysis on the Diffie-Hellman mode. (TCD-CS-2009-13), June 2009.

[7] D. Dolev and A. C. Yao. On the security of public-key protocols. *IEEE Transaction on Information Theory*, 2(29):198–208, March 1983.

[8] I. Livingstone and A. Clark. Zfone – forensic analysis, April 2008.

[9] G. Lowe. An attack on the needham-schroeder public-key authentication protocol. *Inf. Process. Lett.*, 56(3):131–133, 1995.

[10] G. Lowe. Breaking and fixing the needham-schroeder public-key protocol using fdr. In *TACAs '96: Proceedings of the Second International Workshop on Tools and Algorithms for Construction and Analysis of Systems*, pages 147–166, London, UK, 1996. Springer-Verlag.

[11] R. Milner. *Communicating and Mobile Systems: the Pi-Calculus.* Cambridge University Press, June 1999.

[12] P. Zimmermann, A. Johnston, and J. Callas. ZRTP: Media Path Key Agreement for Secure RTP. January 2010.

Copyright © 2010, Infonomics Society