Poster: Handsfree ZRTP - A Novel Key Agreement for RTP, Protected by Voice Commitments

Dominik Schürmann
TU Braunschweig
d.schuermann@tu-braunschweig.de

Stephan Sigg
TU Braunschweig
stephan.sigg@udo.edu

1. INTRODUCTION

Recently, several mobile applications were released that claim to provide secure Voice-over-IP communications. Most of these, e.g., Redphone by Open WhisperSystems[1] or Silent Phone by Silent Circle[2], are utilizing ZRTP [4] to establish session keys for end-to-end security. ZRTP was designed to achieve key exchange without trusted third parties or certificate infrastructure, while providing a way to protect against Man-in-the-Middle (MitM) attacks. The basic idea is that the caller and callee can verify that no MitM attacker is present by recognizing the voice of the peer, while comparing Short Authentication Strings (SAS).

We rethink ZRTP's concept of voice recognition by utilising audio fingerprinting to replace the manual comparison of SAS. This enables the use of devices without displays and hands-free equipment. It provides end-to-end secure communications in cars, while the driver focuses on the street.

We discuss shortcomings of ZRTP, present our novel authentication protocol and discuss results from a case study on utilising audio fingerprints to establish a common secret via a remote connection. With this poster, we aim at gathering feedback and discussing attack scenarios, before implementing a prototype.

ZRTP extends Diffie-Hellman (DH) key agreement by protection against MitM. It forces the callee to release a 256 bit hash h_B of the DH public part y_B in advance of exchanging the public parts y_A, y_B themselves. Due to this hash commitment, an attacker can only guess once with a chance of 1:65536 by using 16 bit SAS [4]. The actual protection against MitM attacks is done by reading the SAS (generated from a combination of h_B, y_A, and y_B) aloud, while the peer has to compare the heard SAS with a displayed one and simultaneously recognise the voice of the caller. For ease of use and pronunciation SAS are words in modern ZRTP implementations. An analogous verification can protect against sophisticated replay attacks, where an attacker impersonates one peer and calls in its name. In this case, the attacker has to use actual recorded speech to conduct a conversation, which is hard to perform unnoticed. Mutual authentication protects against impersonation as it also forces the attacker to read the SAS (or a part of it) aloud, which is tied to this specific key exchange by hash commitment [2].

1.1 Shortcomings of ZRTP

In ZRTP, manual voice recognition is the central innovation to protect against MitM attacks. The comparison of spoken SAS with displayed ones is needed to provide protection against replay attacks of SAS. Therefore, ZRTP inherently requires a display on both communicating devices. In addition, we argue that from a usability perspective, the comparison step distracts users from the crucial step of voice authentication. To improve this process, we propose to utilise audio fingerprints of spoken and received SAS to provide authentication. Audio fingerprints provide sufficient entropy and can be used in conjunction with fuzzy cryptography to establish shared secrets [3].

[1] `http://www.whispersystems.org`
[2] `https://silentcircle.com/web/security/`

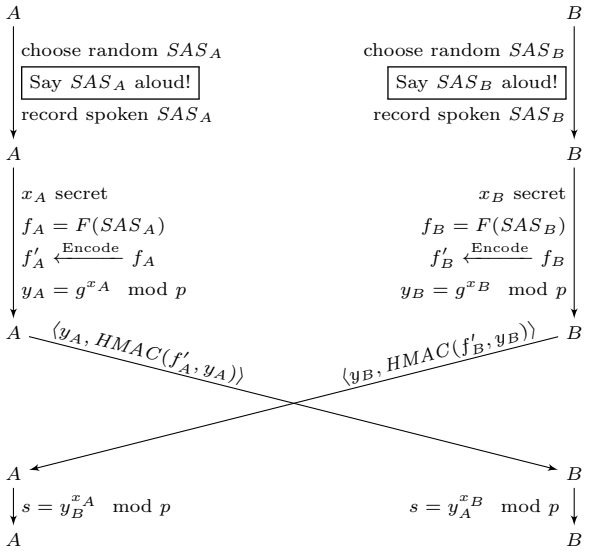

Figure 1: HMAC authenticated Diffie-Hellman with Voice Commitment

2. PROTOCOL DESCRIPTION

Our protocol consists of two parts: A voice commitment and the protection against MitM attacks.

2.1 Voice Commitment

Prior to the DH key exchange, both peers choose a random SAS and speak it out aloud, while it is recorded by the application. In Figure 1, audio fingerprints f_A and f_B are then created from this recording and encoded by an appropriate error correcting code to f'_A and f'_B (cf. [3]). The rest of the protocol follows the well studied DH key agreement, while the public parts are authenticated with $HMAC$ using f'_A and f'_B as secrets. The shared secret s is finally used to

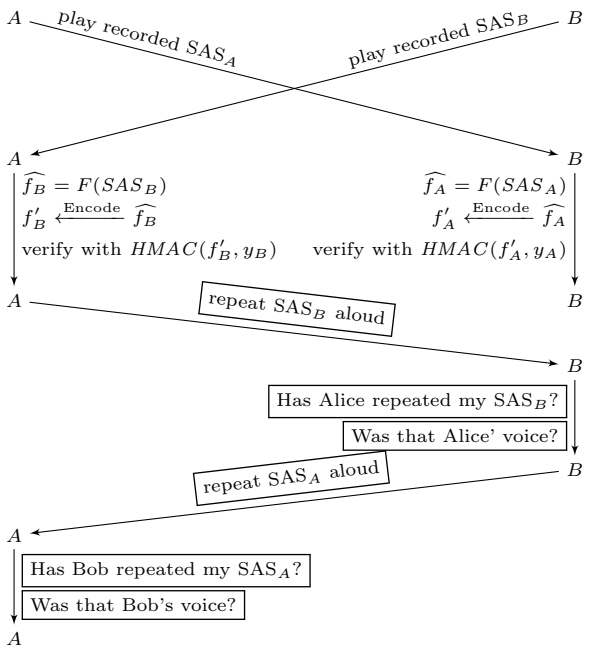

Figure 2: Protection against MitM attacks

encrypt the following conversation utilising SRTP [1].

2.2 Protection against MitM attacks

As depicted in Figure 2, the previously recorded SAS_A and SAS_B are played back to be heard by both peers. The audio fingerprint algorithm F produces fingerprints $\widehat{f_A}, \widehat{f_B}$ that are similar but, due to voice encoding and transmission, not identical to the values f_A, f_B, previously calculated in the voice commitment. If the differences are within a certain threshold [3], error correcting codes can be utilised to encode $\widehat{f_A}, \widehat{f_B}$ to identical secrets f'_A, f'_B. These are used as secrets for $HMAC$ to verify the authenticated DH public parts.

The crucial voice verification is done by repeating the received SAS_B by Alice. Bob needs to check whether Alice repeated his previously committed SAS_B and needs to recognise Alice' voice. Like in ZRTP an additional protection against sophisticated replay attacks is added by analogous repeating and verifying, started by Bob.

Our proposed protocol requires both partners only to speak out an SAS, validate the communication partner via her voice and repeat the SAS obtained. The protocol is feasible for devices without a display since the SAS validity is checked unobtrusively by audio fingerprints.

3. REMOTE FINGERPRINTS

The proposed protocol requires audio fingerprints of sufficient similarity from recorded SAS before and after the transmission over a remote connection. Since in contemporary communication systems, audio is encoded by various codecs, thereby altering e.g. frequency or noise level, the fingerprint generated at two sides of a remote connection might differ significantly. We conducted a case study in which fingerprints are calculated from speech audio on both ends of a mobile communication to study their similarity. In three settings with different environmental conditions we created

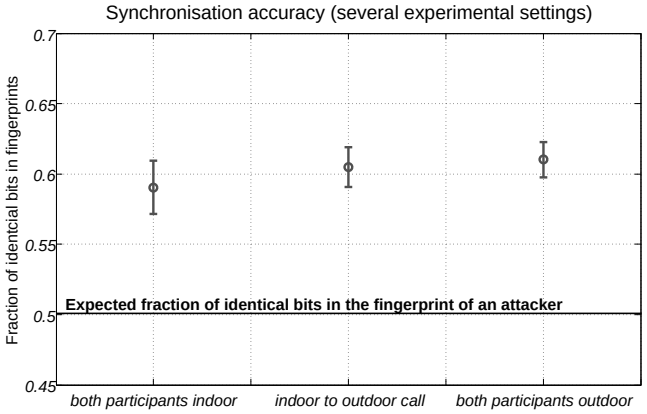

Figure 3: Average similarity and standard deviation experienced for fingerprints in three scenarios

fingerprints from audio that was captured during an ongoing phone call. The recorded audio was simply taken from the microphone of the phone and covers local and remote audio. For the case studies, a Nexus 4 and an LG Optimus Speed (Optimus 2x) have been utilised in the German D2 cellular network. We distinguish between situations in which both phones are indoors, both phones are outdoors or one is indoors and the other outdoors. In each of these situations, the experiment has been repeated nine times. The audio fingerprints have been conducted according to the parameters detailed in [3]. Figure 3 summarises the results achieved. Observe that there is sufficient similarity in fingerprints so that indeed identical keys may be generated by utilising fuzzy cryptography.

4. CONCLUSION AND FUTURE WORK

Similar to ZRTP, our protocol utilises recorded voice for authentication and verification. However, the combination of the cryptographic primitives is significantly different and the utilisation of audio fingerprinting leads to a more convenient, less manual and simple protocol. We investigated the feasibility of using audio fingerprints via a remote connection in a case study. This work is a first sketch of a new way to protect against MitM attackers using audio fingerprinting and fuzzy cryptography. We are currently evaluating the cryptographic correctness and building a prototype to conduct a usability analysis.

5. REFERENCES

[1] M. Baugher, D. McGrew, M. Naslund, E. Carrara, and K. Norrman. The Secure Real-time Transport Protocol (SRTP). RFC 3711 (Proposed Standard), Mar. 2004. Updated by RFC 5506.

[2] M. Petraschek, T. Hoeher, O. Jung, H. Hlavacs, and W. Gansterer. Security and usability aspects of Man-in-the-Middle attacks on ZRTP. *Journal of Universal Computer Science*, 14(5):673–692, 2008.

[3] D. Schürmann and S. Sigg. Secure Communication Based on Ambient Audio. *IEEE Transactions on Mobile Computing (TMC)*, 12(2):358–370, 2013.

[4] P. Zimmermann, A. Johnston, and J. Callas. ZRTP: Media Path Key Agreement for Unicast Secure RTP. RFC 6189 (Informational), Apr. 2011.