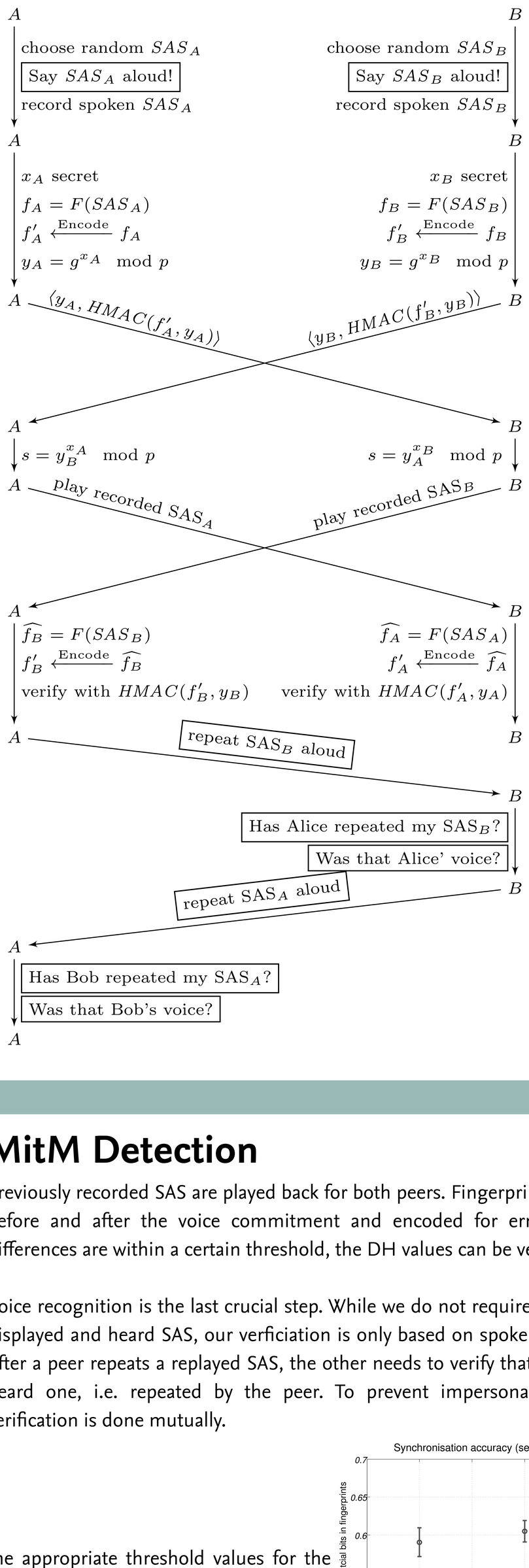
Technische Universität Braunschweig

Institut für Betriebssysteme und Rechnerverbund

Handsfree ZRTP - A Novel Key Agreement for RTP, Protected By Voice Commitments

Dominik Schürmann and Stephan Sigg
Technische Universität Braunschweig | Institut für Betriebssysteme und Rechnerverbund
< schuerm | sigg >@ibr.cs.tu-bs.de



Motivation

Several mobile applications were released that provide end-to-end secure voice calls without prior manual key exchange or additional identification infrastructure. Prominent examples are

- Silent Phone by Silent Circle
- Redphone by Open WhisperSystems
- CSipSimple (open source SIP application)

All those implementations use the Diffie-Hellman based ZRTP key exchange to protect against Man-in-the-Middle (MitM) attackers. ZRTP's 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 this concept of voice recognition by utilising audio fingerprinting to replace the manual comparison of SAS.

ZRTP

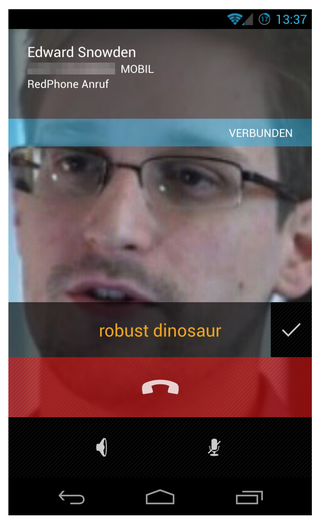
Fundamentals:

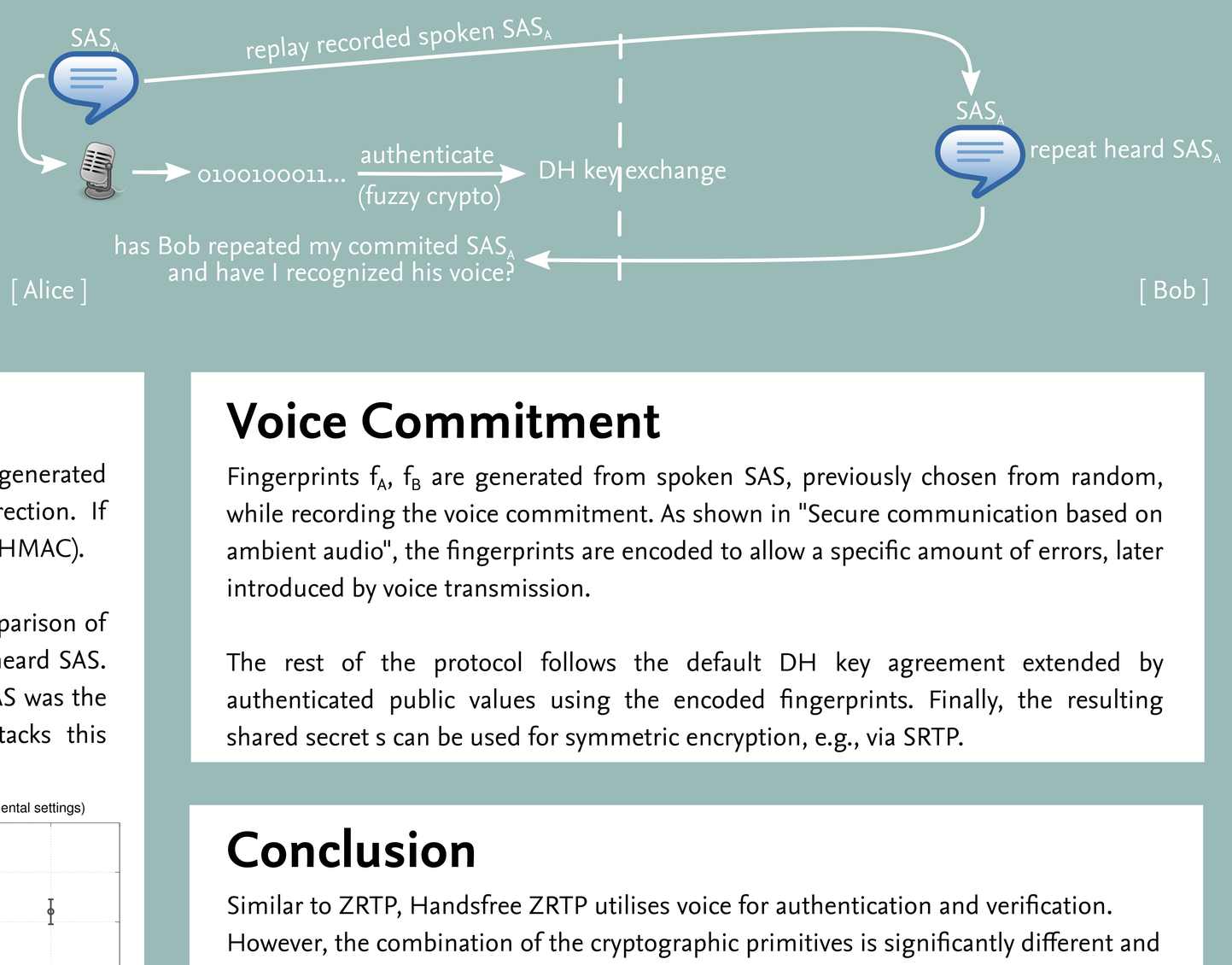
- Based on well-known Diffie-Hellman (DH) key exchange
- Hashes of public DH values are calculated for authentication.
- Allows Short Authentication Strings (SAS) instead of long hashes, a Hash Commitment is executed before exchanging public DH values.
- Actual authentication is done by recognizing the peer's voice, while comparing these SAS

→ MitM protection by voice recognition and comparison of displayed SAS

Disadvantages:

- Cumbersome SAS comparison step
- Will an uneducated user cancel established calls when hearing a different SAS?
- Unusable on devices without displays, e.g., hands-free equipment



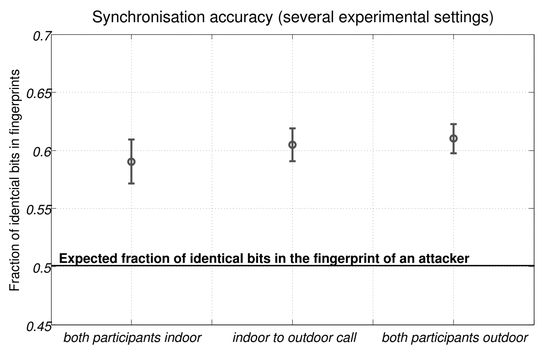


MitM Detection

Previously recorded SAS are played back for both peers. Fingerprints are generated before and after the voice commitment and encoded for error correction. If differences are within a certain threshold, the DH values can be verified (HMAC).

Voice recognition is the last crucial step. While we do not require a comparison of displayed and heard SAS, our verficiation is only based on spoken and heard SAS. After a peer repeats a replayed SAS, the other needs to verify that her SAS was the heard one, i.e. repeated by the peer. To prevent impersonation attacks this verification is done mutually.

The appropriate threshold values for the error correcting codes were defined by conducting real-world experiments with two Android Smartphones.



Voice Commitment

Fingerprints f_A, f_B are generated from spoken SAS, previously chosen from random, while recording the voice commitment. As shown in "Secure communication based on ambient audio", the fingerprints are encoded to allow a specific amount of errors, later introduced by voice transmission.

The rest of the protocol follows the default DH key agreement extended by authenticated public values using the encoded fingerprints. Finally, the resulting shared secret s can be used for symmetric encryption, e.g., via SRTP.

Conclusion

Similar to ZRTP, Handsfree ZRTP utilises 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. This work is a first sketch of a new way to protect against MitM attackers using audio fingerprinting and fuzzy cryptography.