JUNE 20, 2018
VERSION 1.0

Skype Private Conversation

Technical white paper

Microsoft

Contents

Introduction

This document provides a brief technical explanation of Skype Private Conversation.

Skype Private Conversation is an optional mode of messaging when all audio calls, texts, or media files like images, audio, or video are end-to-end encrypted. A Private Conversation supports one device per user, and is only between two users. A user can switch devices during a conversation, but the previous messages won't appear on the new device.

The Private Conversation feature opens a new chat window and encrypts sent and received messages, so no third party, including Microsoft, can see what's being shared. This means that some features available in Skype, like Translator, Cortana, Add-ins, and Bots, will not be available.

Skype Private Conversation end-to-end encryption uses Signal Protocol by Signal.

Terminology

A **device** is a computer or mobile phone running the Skype application with one user logged in. A device is identified by a globally unique **deviceId**.

All the following key-pairs are generated using Curve25519. The private key component never leaves the device and is protected in device-local storage. In this document, public keys are uppercase, while private keys are lowercase.

Identity Key (**IK**, **ik**) is a long-term key-pair associated with the device. It is generated after a user logs in. The Identity Key is unique per device.

Signed Pre-Key (**SPK**, **spk**) is a medium-term key-pair that is rotated periodically. It is signed by **ik**.

Pre-Key (**PK**, **pk**) is a one-time key-pair used to set up conversations between devices. It is generated on a per-needed basis.

Signal Protocol Library is the underlying encryption provider used by devices.

Client setup

After the user has logged in, an Identity Key and Signed Pre-Key are generated. The Signed Pre-Key is then periodically renewed every two weeks.

Conversation setup

A Private Conversation is established between a pair of devices designated by users participating in the conversation. Users can switch devices during a Private Conversation, but users can't have the same session open simultaneously across multiple devices. For example, when users A and B start a Private Conversation, they exchange a handshake, which sets up their devices and transfers the keys to set up the end-to-end encryption.

The handshake has three steps:

1. Invite

User A sends a message from one device to invite user B to a Private Conversation. The invite provides the following information to B: **`IK`**(**A**), **`SPK`**(**A**), **`PK`**(**A**), **`deviceId`**(**A**). B receives the invite on all their devices.

2. Accept

User B accepts the invite and their device sends an accept message to A containing the following information: **`IK`**(**B**), **`SPK`**(**B**), **`PK`**(**B**), **`deviceId`**(**B**).

3. Confirm

This is an automatic response from A notifying all devices of both A and B of negotiated endpoints.

The steps may be executed asynchronously, with the server keeping the data for the other party to retrieve. The lifetime of a handshake is limited to seven days and/or validity of **`SPK`**s, whichever expires first.

The Signal Protocol Library uses the X3DH algorithm [1] to establish the end-to-end encrypted session between two Skype users. After the session is established, users A and B can view the fingerprint of the session and confirm the session is secure. Switching to another device uses the same handshake algorithm and is conversation based. When a user switches the conversation to another device, the previous messages are not moved to the new device.

Users can leave a Private Conversation at any time, including during the handshake setup. Users can resend the invitation, and a new Pre-Key will be provided each time. Until a handshake is complete, neither A nor B can send messages to the Private Conversation.

What is encrypted

A Private Conversation uses the same infrastructure as any other Skype conversation: Messaging and calling use Skype token authentication. During a Private Conversation, all content, including audio calls, text messages, and media files (images, audio, video), are end-to-end encrypted. Standard information such as delivery time, read receipts, and alike, are included to ensure the message was delivered.

Text messages

Before sending a message, both sender and receiver must complete a handshake and setup Signal Protocol session. Any party can then send the first message to the conversation.

- Sender generates a per-message random symmetric encryption key **Ke**, initialization vector **IV**, and authentication key **Ka**.
- Sender encrypts plaintext of the message to be sent using AES-256 in CBC mode with key **Ke** and **IV** to obtain ciphertext **CT**.
- Sender computes authentication tag **T** using HMAC-SHA256 over **IV** and **CT** with key **Ka**.
- Sender computes SHA-256 hash over **IV**, **CT** and **T** to obtain hash value **H**.
- Sender then encrypts tuple **Ke**, **Ka**, **H**, **IV**, **T** to the intended recipient device using the Signal Protocol session established with the device (see [2]) to obtain a dictionary **D** = {<**device**>: <**encrypted data**>}.
- Sender then sends **CT** and dictionary **D** to the intended recipient.

The server sends the messages to all recipient devices. Devices not listed in the dictionary will drop the messages and not decrypt the contents. A dictionary is used to allow the extension to multiple devices in the future.

The client application requests all pending messages from the server and only shows messages in the chat window that are meant for the device and can be decrypted. Encrypted messages are stored up to 30 days to support offline messaging.

Media and file sharing

Media sharing works in a similar fashion as message sending.

- Sender generates per-file random encryption key **Ke**, initialization vector **IV**, and authentication key **Ka**.
- Sender encrypts, on the device, file/media to be shared using AES-256 in CBC mode with key **Ke** and **IV** to obtain ciphertext **CT**.
- Sender computes HMAC-SHA256 over **IV** and **CT** with key **Ka** to obtain authentication tag **T**.
- Sender computes SHA-256 hash over **IV**, **CT**, and **T** to obtain hash value **H**.
- Sender uploads **IV**, **CT**, **T** to the file storage and obtains URL of the file on the server.
- Sender sends tuple **Ke**, **Ka**, **H**, **URL** to the recipient(s) over a pre-established Private Conversation.

URL is server generated and does not contain original file name and other potentially sensitive information. URL is sent in the message metadata in plaintext form, as we need the server to manage the lifetime of the content in accordance with respective regulations. Content uploaded to file storage is encrypted by sender’s device and can only be decrypted by the intended recipient; Skype and Microsoft can’t decrypt this content.

Media files are stored for up to 30 days.

Calls

End-to-end encrypted calls can only be initiated from an existing Private Conversation.

- Caller and callee establish a Private Conversation session.
- Caller generates a cryptographically random encryption key and sends it during the Private Conversation (that is, encrypted by Signal Protocol, see [2]).
- Caller encrypts the SRTP using the generated encryption key, which can only be decrypted by the callee.
- After the call is setup, the media packets are encrypted using the SRTP keys.

Conclusion

Skype Private Conversation protects end-to-end content between the sender and recipient using the Signal Protocol. This prevents Skype, Microsoft, and third parties from decrypting user content. This means Skype users have privacy, but it also means features such as Translator, Cortana, etc. can't be used with messages. Users can verify the security of their conversations by checking the key setup between two users. At any time, users can exit a Private Conversation mode in Skype messenger.

References

[1] - The X3DH Key Agreement Protocol ➔

[2] - The Double Ratchet Algorithm ➔