**Agenda item:** 9.8

**Source:** Qualcomm Inc.

**Title: [FS\_AVATAR] User-Avatar authentication in AR calls**

**Document for** Discussion andAgreement

# Introduction

In this contribution, we discuss the issue of impersonation using 3D avatars and propose a solution to minimize this risk.

# Security challenges

The integration of 3D avatars in augmented reality (AR) calls has become increasingly popular. These 3D avatars offer users a way to represent themselves virtually in a more personalized and interactive manner, mimicking the user's facial expressions and body gestures.

The use of 3D avatars in AR calls provides a dynamic and immersive experience, allowing users to engage with others in a virtual environment that can be more engaging and expressive than traditional video calls. These avatars can be customized to resemble the user closely and can be dressed according to the user’s desire.

While this technology presents numerous opportunities for enabling a more immersive user experience, it also brings with it significant risks, particularly around user privacy and security questions. Unfortunately, the very features that make 3D avatars appealing also open the door to potential misuse, particularly in the form of impersonation. When a user creates a 3D avatar, especially one that closely resembles their real-life appearance, this model can potentially be misappropriated by malicious actors.

One of the most significant risks occurs when the 3D avatar base model is exchanged with other participants in the call for animation. Once the assets are downloaded, a malicious user may offer it in the future as their own base avatar model to other participants.

Impersonation can be particularly dangerous in scenarios where 3D avatars are used for authentication or verification purposes. For instance, some platforms may use the uniqueness of a user's avatar as part of a security protocol. If an attacker successfully replicates or steals the avatar, they could bypass these security measures, leading to unauthorized access to accounts or personal information.

# Potential Authentication Solution

In order to minimize the risk of impersonation by other participants of an AR call that are granted access to the user’s base avatar model, the following measures should be taken:

* DRM protection of the base avatar model and encryption of key components of the base avatar model, such as the blendshapes, the mesh components, key textures, and/or the skin binding information,
* Encrypted storage of user-authenticating information in the base avatar model,
* During an AR call, receivers continuously authenticate the user by matching user-authenticating information from the base avatar model to the same information extracted from live user streams.

NOTE: the security risks of storing user authenticating features as part of the base avatar model are FFS.

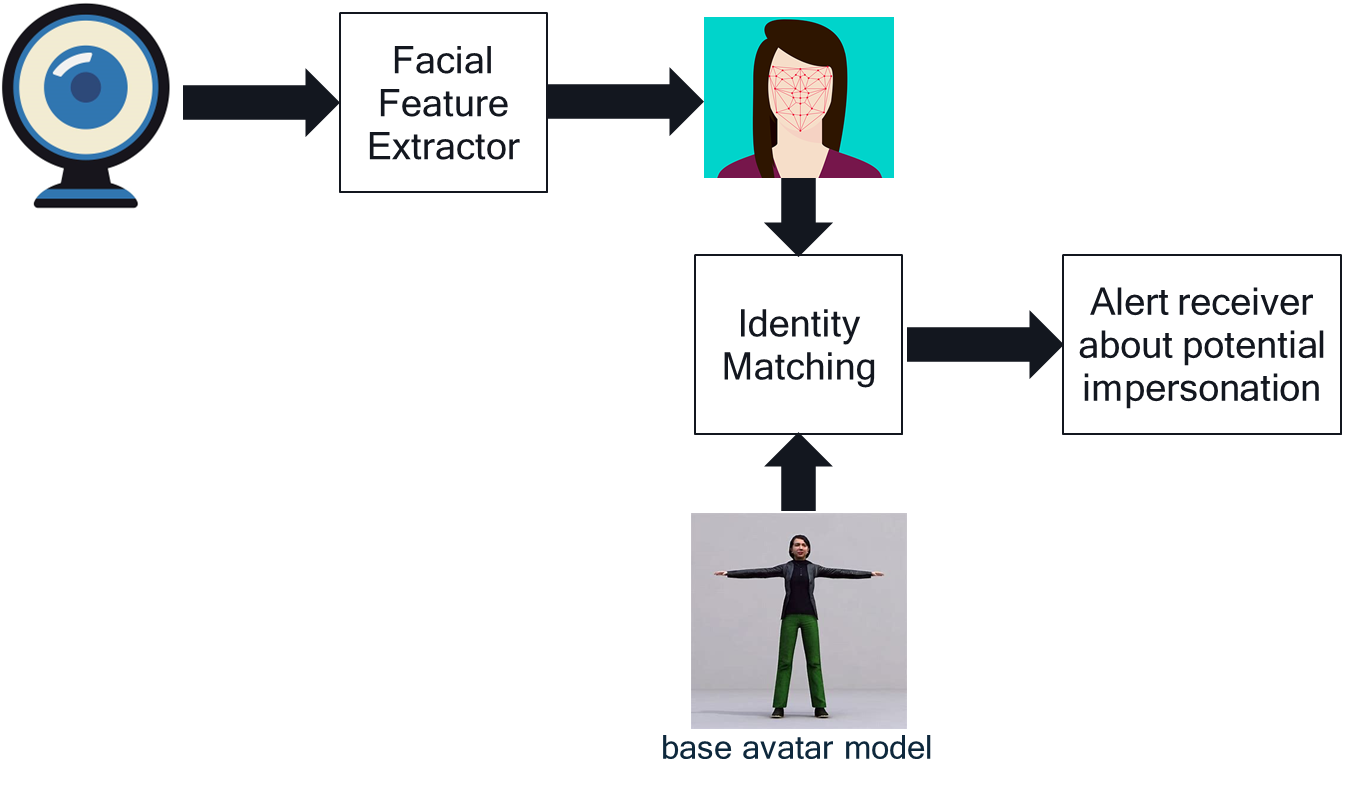
One such user-authenticating information are the facial features, which can be stored as part of the user’s base avatar model and encrypted using their private key to prevent tampering or replacement.

During an AR call the receivers access these features by decrypting them using the public security key of the sender. They also periodically run the same facial feature extraction algorithm on key frames from the user’s video stream and compare them to the features in the base avatar model. A strong or prolonged deviation is a strong indicator of an impersonation attempt. The same approach can be used with user’s voice by deploying a voice feature extractor.

Facial features are key points or regions on the face that capture significant information about an individual's appearance, often used in facial recognition, emotion detection, and identity verification systems. These features typically include landmarks such as the eyes, nose, mouth, and the contours of the face, which provide critical data about the geometry and structure of the face. The extraction of these features involves identifying specific points on the face, known as facial landmarks, which can number anywhere from a few dozen to several hundred depending on the complexity of the model. These landmarks represent the spatial coordinates of important facial regions, allowing algorithms to map and analyze the face's shape and expressions. Beyond landmarks, other facial features can include texture patterns, skin tone, and even fine details like wrinkles or freckles, which contribute to the uniqueness of a person's face.

The extraction of facial features is typically performed using algorithms like Active Shape Models (ASM) [1], Active Appearance Models (AAM) [2], or more recently, deep learning-based methods like Convolutional Neural Networks (CNNs) [3]. Traditional methods, such as ASM and AAM, rely on statistical models to learn the shape and appearance of facial features from a training set and then fit these models to new images by adjusting the parameters to match the observed data. CNNs, on the other hand, are trained on large datasets of facial images and automatically learn to identify and extract relevant features by processing the image through multiple layers of convolutions and pooling. These networks can capture both local and global features, allowing them to detect subtle variations in facial expressions or slight changes in orientation. Modern approaches often combine landmark detection with feature extraction in a unified framework, resulting in robust and accurate models that can be used in real-time applications such as facial recognition systems or augmented reality.

The following diagram depicts the procedure:



# Proposal

We propose to adopt the content of sections 2 and 3 to the PD and recommend the development of a solution for the prevention of impersonation. We also propose to document the requirement on the Avatar Representation Format to support user authentication and prevention of impersonation.

# References

[1] Wan et al., An accurate active shape model for facial feature extraction, Pattern Recognition Letters, Volume 26, Issue 15, November 2005

[2] Edwards et al., Face recognition using active appearance models, Lecture Notes in Computer Science, volume 1407

[3] Wang et al., Deep Face Recognition, A Survey, Computer Vision and Pattern Recognition, August 2020