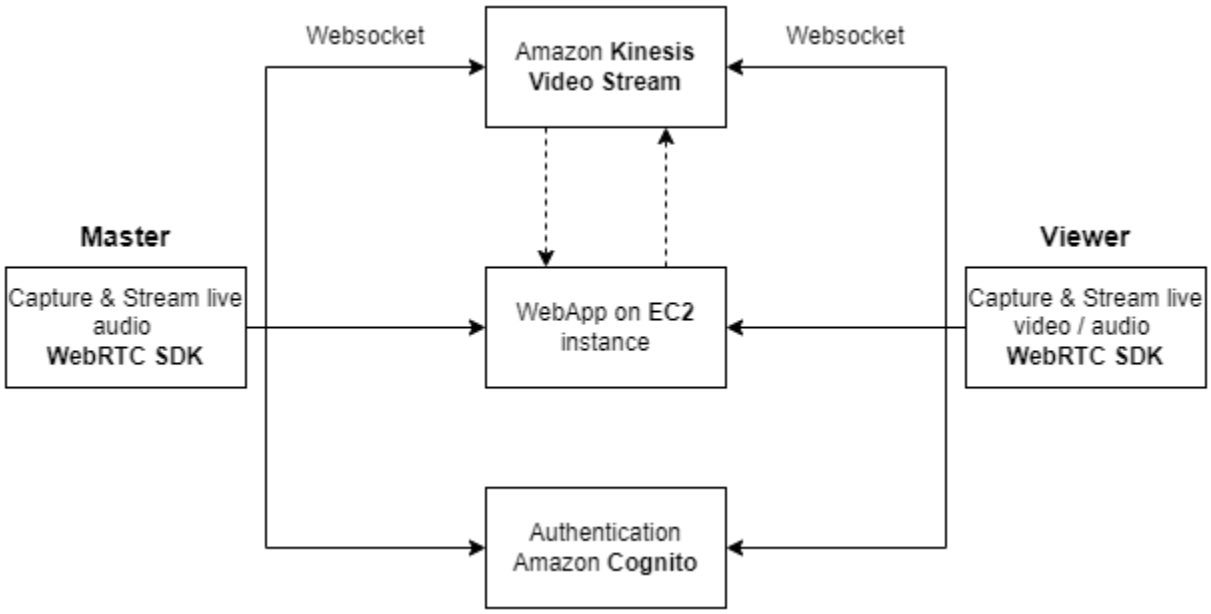
CSE 523-524 Project
VR Avatar Talking Face for Remote Video App
(Vedant Sawant)

Overall Architecture:

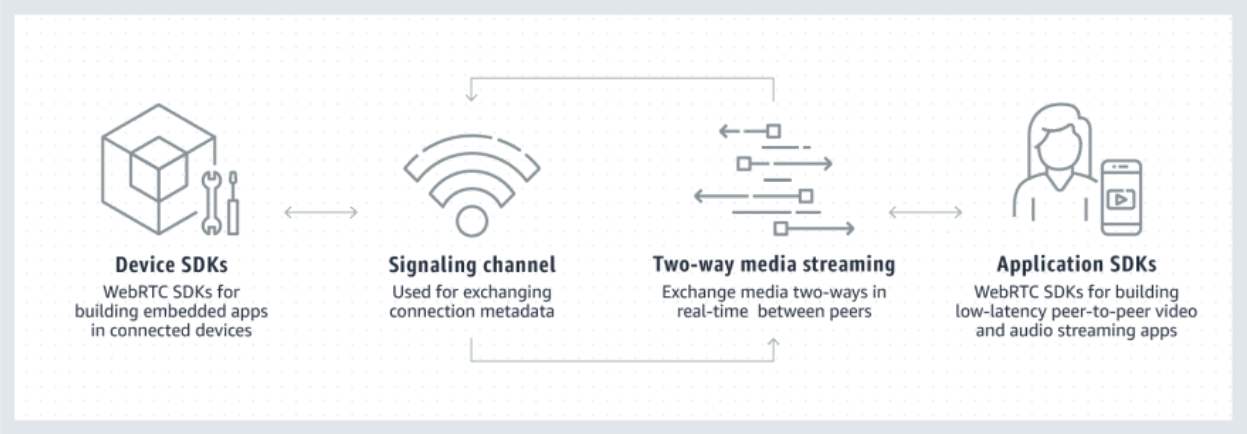


Amazon Kinesis Video Streams with WebRTC uses a signaling channel to set up, control, and terminate a peer-to-peer connection by exchanging signaling messages. Peers include a master that initiates the connection and viewers that can only exchange media with the master.

The process is as follows:

- The Master creates a meeting and chooses audio/video.
- When the video button is clicked, Master sends an SDP offer containing session information and ICE candidates to Viewer through a WebSocket.
- The Viewer generates an SDP Answer with information about the session and sends it back to the Master.
- They perform connectivity checks using STUN requests and mark valid ICE candidates.
- After negotiating, one valid pair is chosen and media starts flowing between the applications.

WebRTC Architecture:



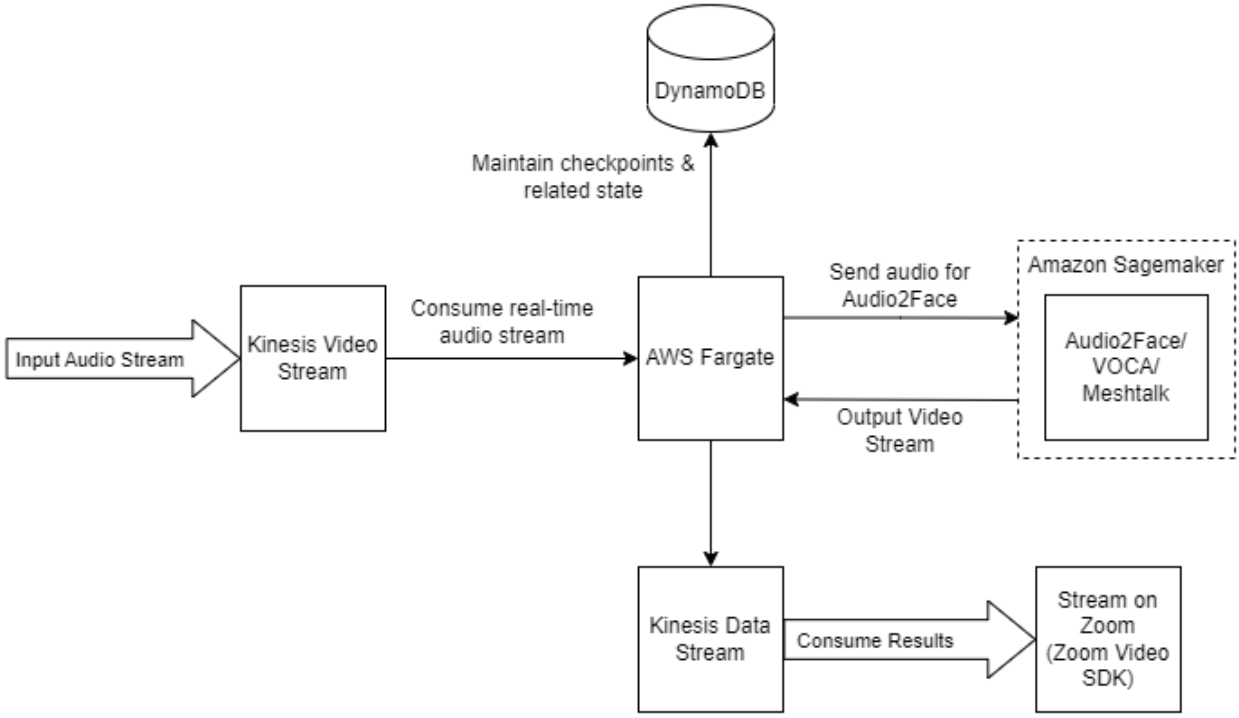
(Image: Amazon)

WebRTC uses peering techniques for real-time data exchange between connected peers and provides low-latency media streaming.

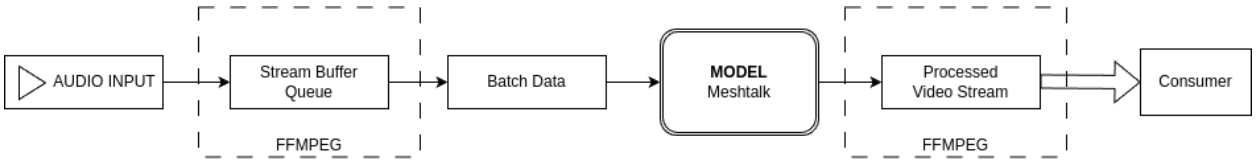
WebRTC used the following protocols:

- STUN is used to discover public addresses
- TURN servers are used to bypass restrictions
- SDP is used to describe the multimedia content of the connection.
- ICE allows web browsers to connect with peers and ICE candidates are used for communication by the sending peer

Media Flow & Processing:



The Kinesis Video Streams client library enables the processing of media across distributed workers, manages invocation of SageMaker endpoints, and publishing of inference results into Kinesis data streams.
The library determines which streams to process, connects to them, and refreshes them periodically. It pulls media fragments from the streams, extracts, and decodes, and invokes the SageMaker endpoint. The results are published into a Kinesis data stream for consumption.
The AWS CloudFormation template can be used to automate the deployment of all relevant AWS infrastructure, including an ECS cluster using Fargate, a DynamoDB table for maintaining checkpoints, and CloudWatch resources for monitoring. The library will be used with a SageMaker endpoint that accepts audio data.
Once the output video stream is received from the ML model, it can be encoded to a format that can be streamed to Zoom using API provided by Zoom.


AUDIO INPUT
Stream Buffer Queue
FFMPEG
Batch Data
MODEL
Meshtalk
Processed Video Stream
FFMPEG
Consumer