Brief history of media processing with GStreamer at Twilio

GStreamer conference

Oct 23 2025

Jeff Foster, Andrey Kovalenko

WHO WE ARE AND WHERE WE ARE GOING

Twilio in brief
Leader & pioneer in CPaaS solutions

01 Founded in 2008 in San Francisco

02 2024 turnover : $ 4.46 bn

03 APIs : Messaging, Voice, Video, SMS, Email

04 Global presence - 5,500 employees

https://twilio-engineering.atlassian.net/browse/VMEDIA-4530

Where Twilio came from, and where we are going

2008
Twilio founded
(Programmable Voice, Phone numbers)

2010
SMS launch
(Programmable Messaging)

2015
Programmable Video launch

2016
IPO,
Kurento acquisition

2017
Video Group Rooms Launch

2018
Sendgrid Email Launch

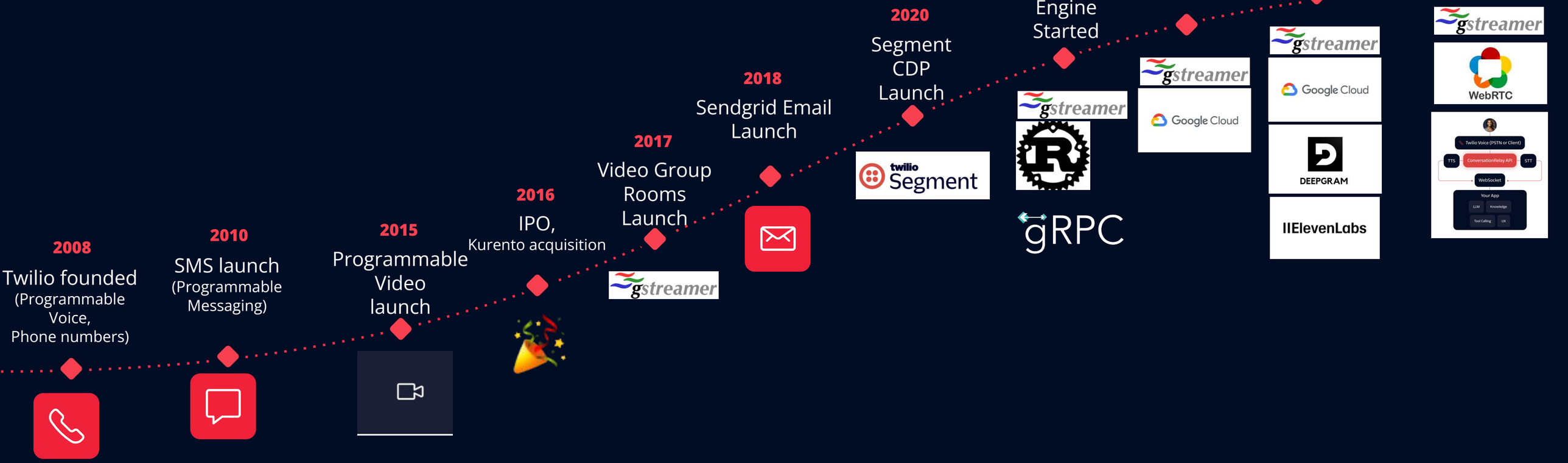
2020
Segment CDP Launch

2022
Twilio Media Engine Started

2023
tme-nlu launch
(Google Dialogflow integration)

2024
tme-streams, tme-processor tme-recorder launch,
Real-time transcriptions GA
Google Speech v2, Deepgram, 11labs

2025
tme-bridge started
(hybrid multi-participant SFU+MCU),
ConversationRelay GA,
Video Transcriptions



Twilio Media Engine

Started in 2022 as part of media stack modernization initiative
Twilio Media Engine is written in Rust and uses GStreamer for all real-time media processing tasks

TME facts

Why GStreamer

- Flexibility in building complex media pipelines
- Huge library of existing media elements
- Active and passionate open-source community
- Rust bindings

Why Rust

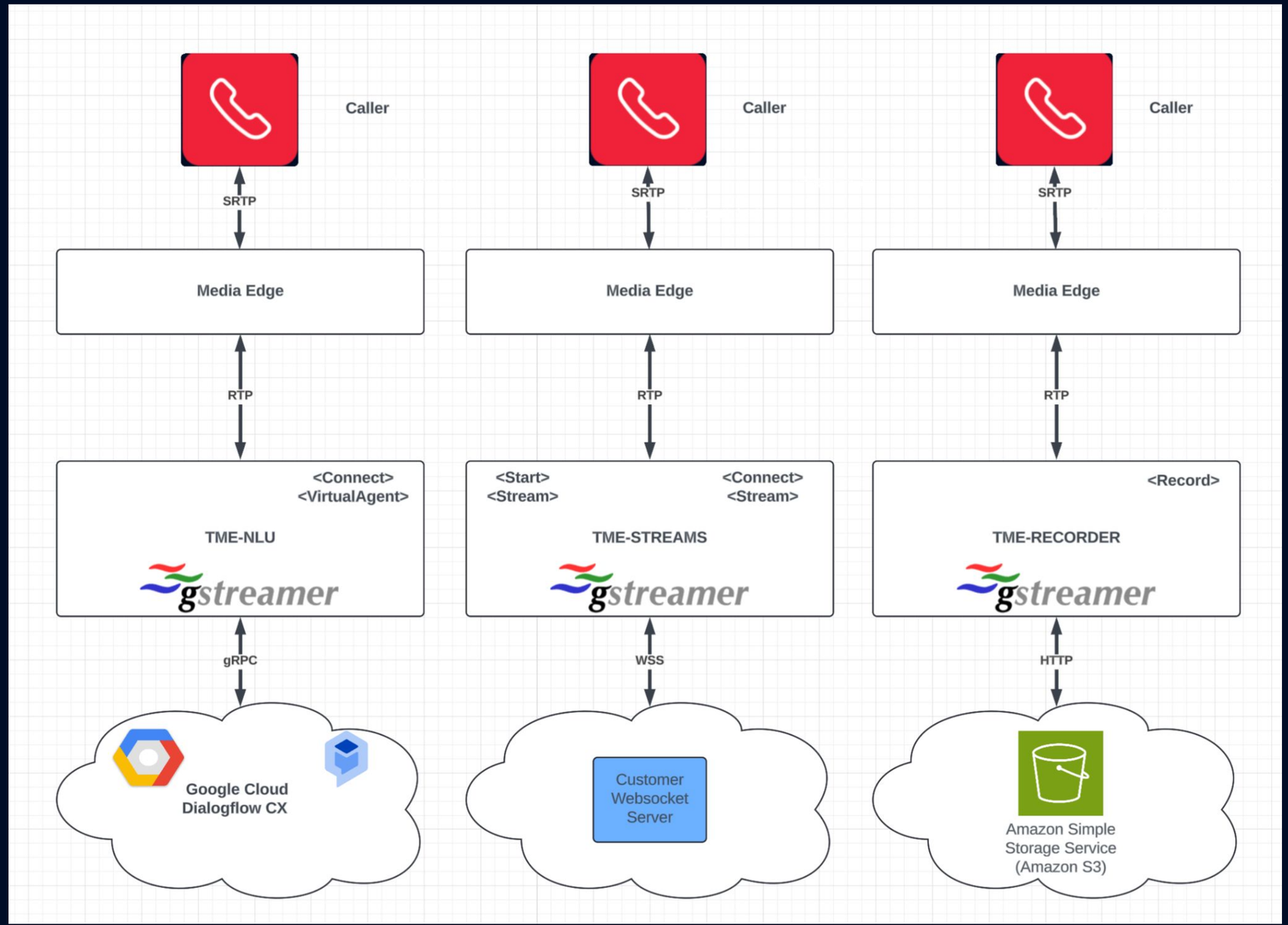
- Memory safety
- Excellent async io runtime
- Great tooling and built-in best practices in the compiler

Why gRPC

- Based on protobuf - flexible and strongly typed, easy to expand API in a backward-compatible manner
- Easy to generate client stubs in any language
- Built-in security and encryption

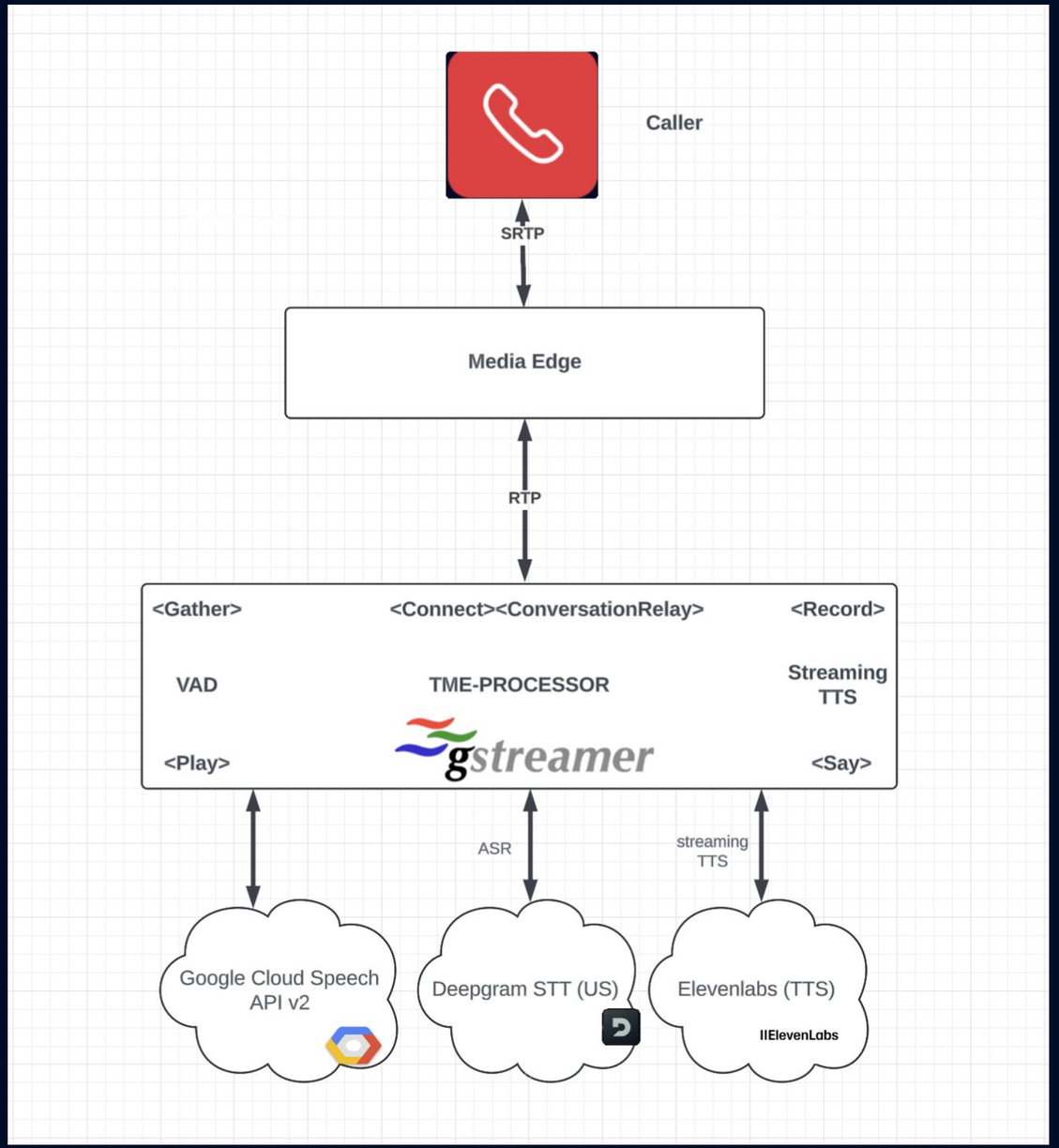
Twilio Media Engine - production use cases

Today GStreamer-based Twilio Media Engine performs a variety of media processing tasks including speech recognition, voice activity detection, recording, DTMF recognition, noise reduction, streaming TTS, audio playback, media streams over websocket, and integration with virtual agent providers and customer-provided LLMs



Twilio Media Engine - production use cases (cont’d)

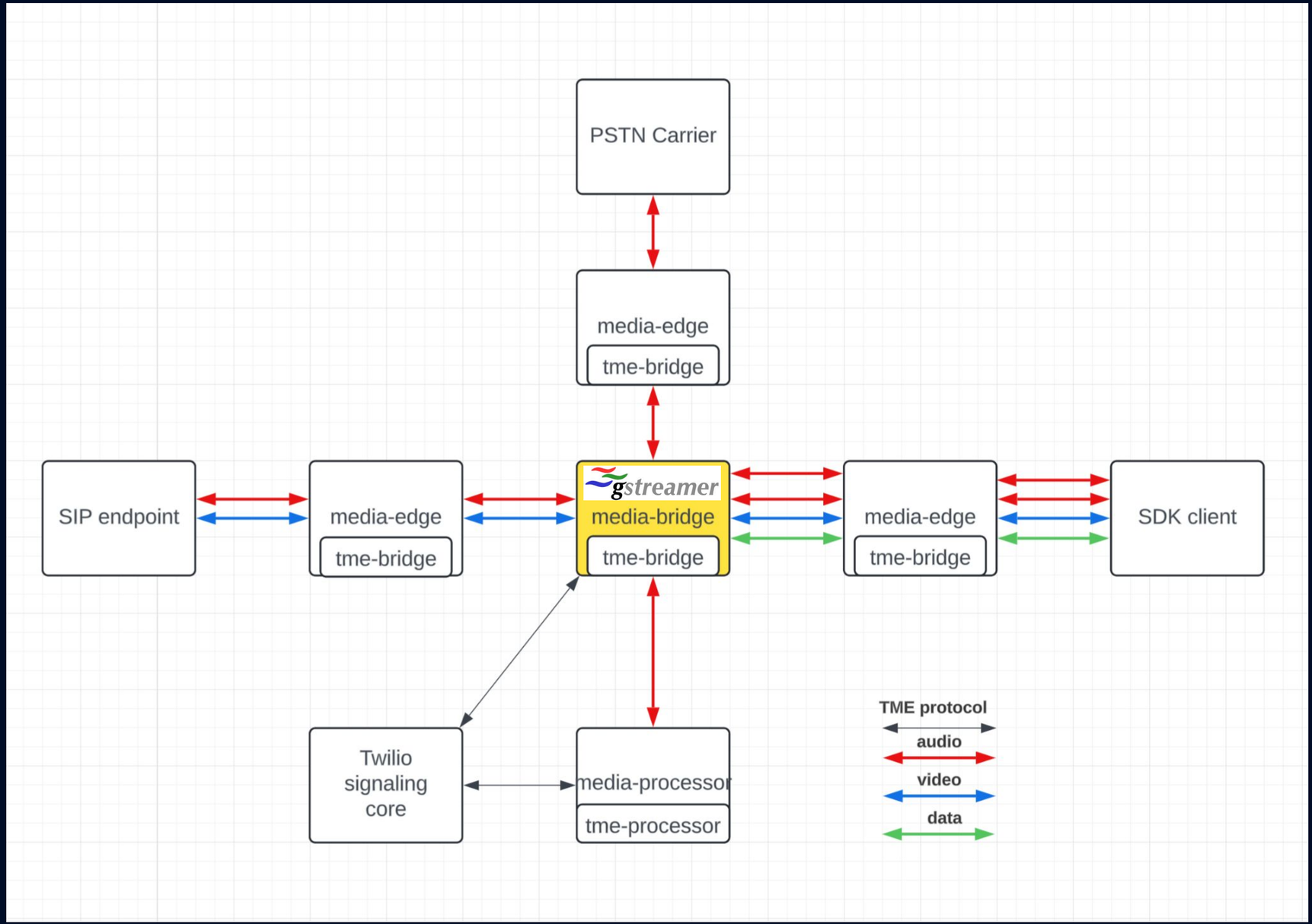
TME Processor acts as a Swiss-army knife of internal media processing allowing media function composition and bi-directional streaming API events to and from multiple Twilio integration partners, including Google, Deepgram and Elevenlabs



Twilio Media Engine - work-in-progress

Twilio Media Engine Bridge hybrid MCU / SFU allows bridging of modern and legacy RTP and WebRTC endpoints such as PSTN carriers, SIP PBXs, and SDK clients, efficiently relaying or mixing media tracks based on the endpoints' capabilities

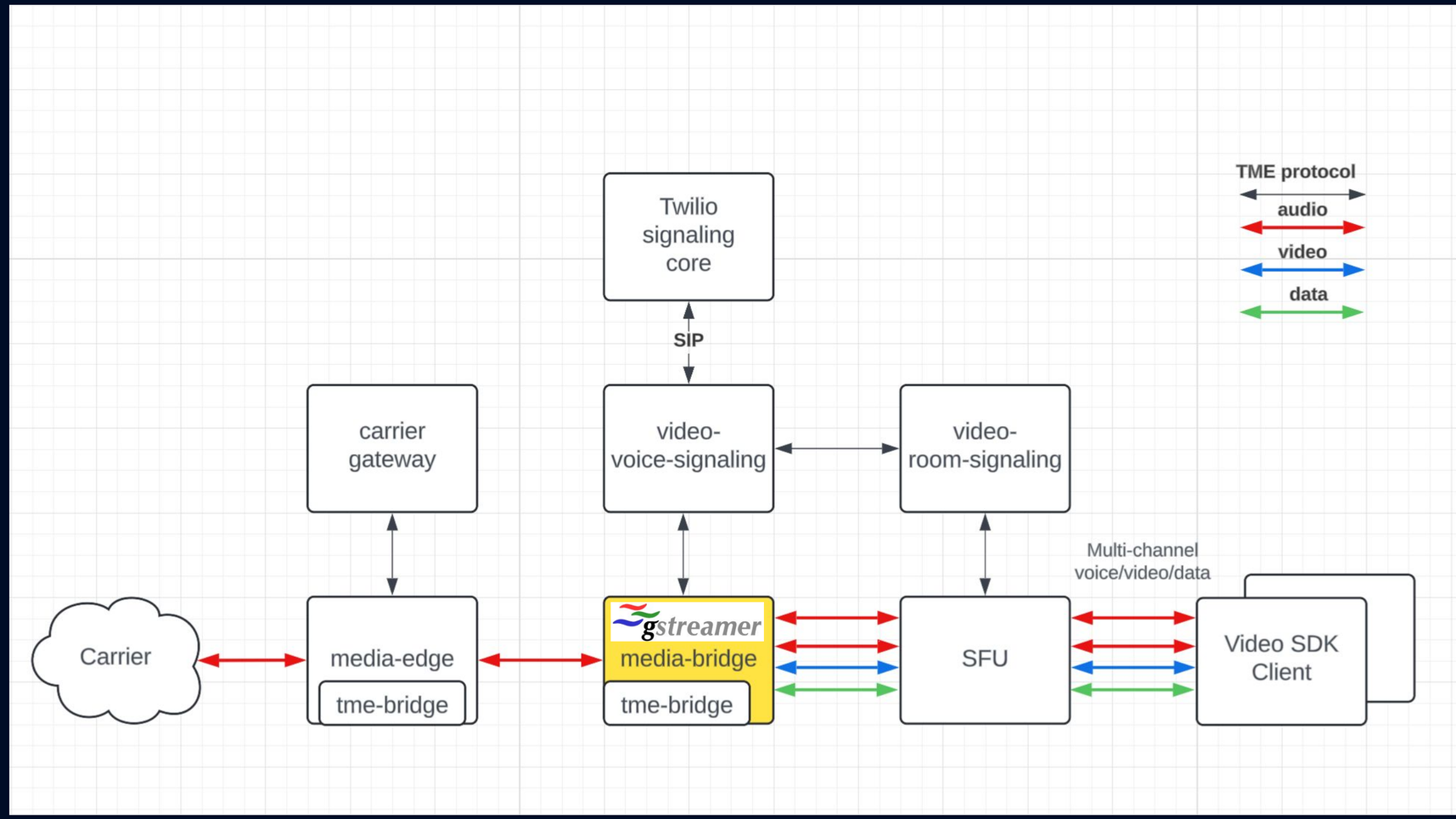
TME Bridge Use Case 1 - internal multi-party mixing / forwarding bridge



Twilio Media Engine - work-in-progress (cont’d)

TME Bridge’s flexible architecture allows it to subscribe voice and data channel tracks and efficiently relay or mix them to legacy PSTN endpoints

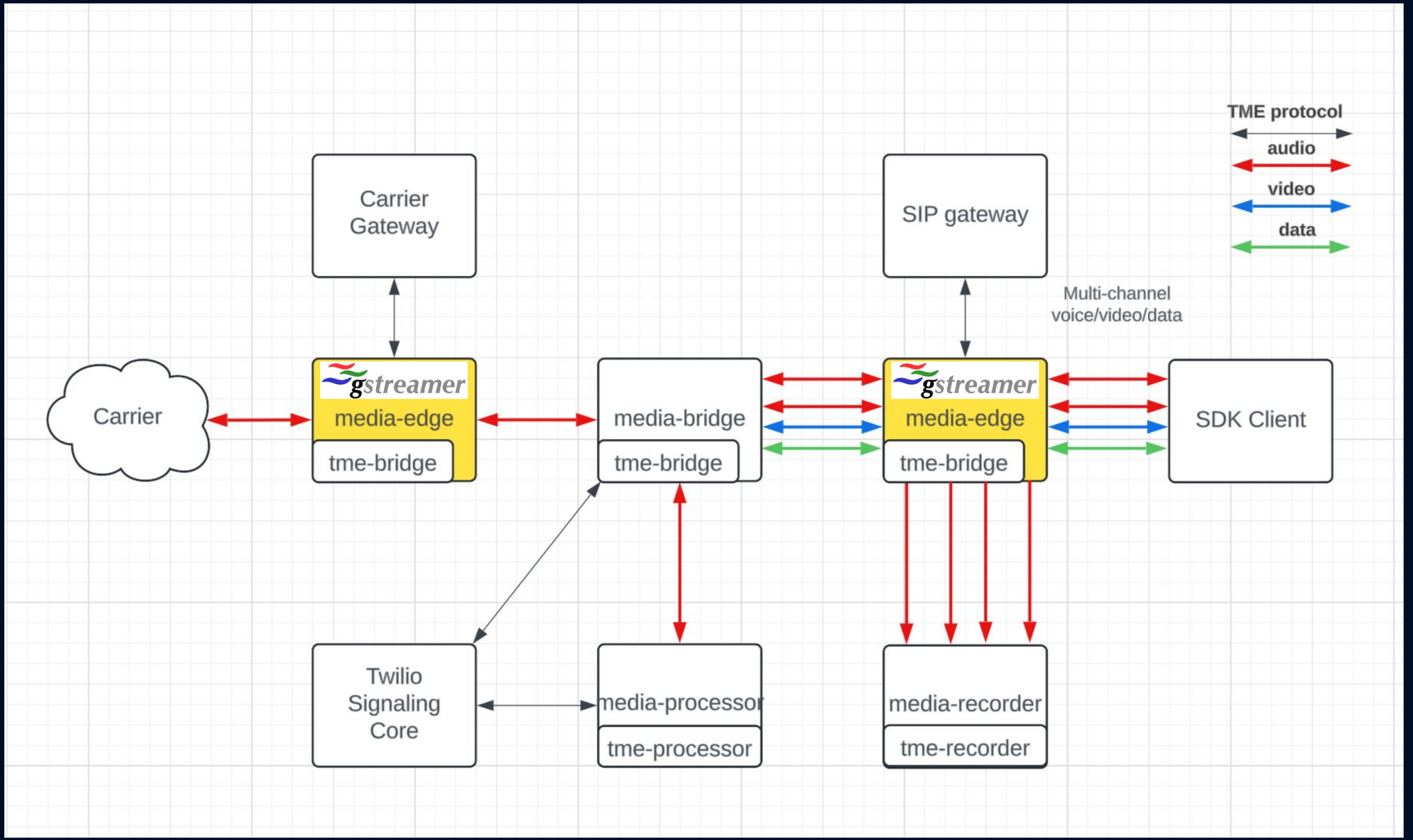
TME Bridge Use Case 2 - bridging voice and video infrastructures



Twilio Media Engine - work-in-progress (cont'd)

TME Bridge's collection of (S)RTP and WebRTC participant types allows it to act a media edge (2-party bridge) between WebRTC capable and RTP endpoints (e.g PSTN carriers)

TME Bridge Use Case 3 - new media edge



EVOLUTION OF MEDIA PROCESSING AT TWILIO

https://twilio-engineering.atlassian.net/browse/VMEDIA-4530

Evolution of media processing at Twilio

Programmable Voice

Multi-party conferencing

Twilio client

iOS + Android

WebRTC

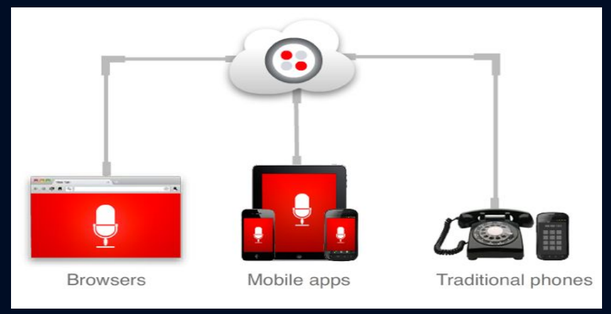
Twilio SIP

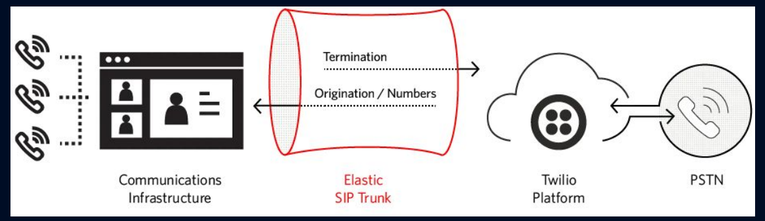
Elastic SIP trunking

Secure SIP trunking

Programmable Video

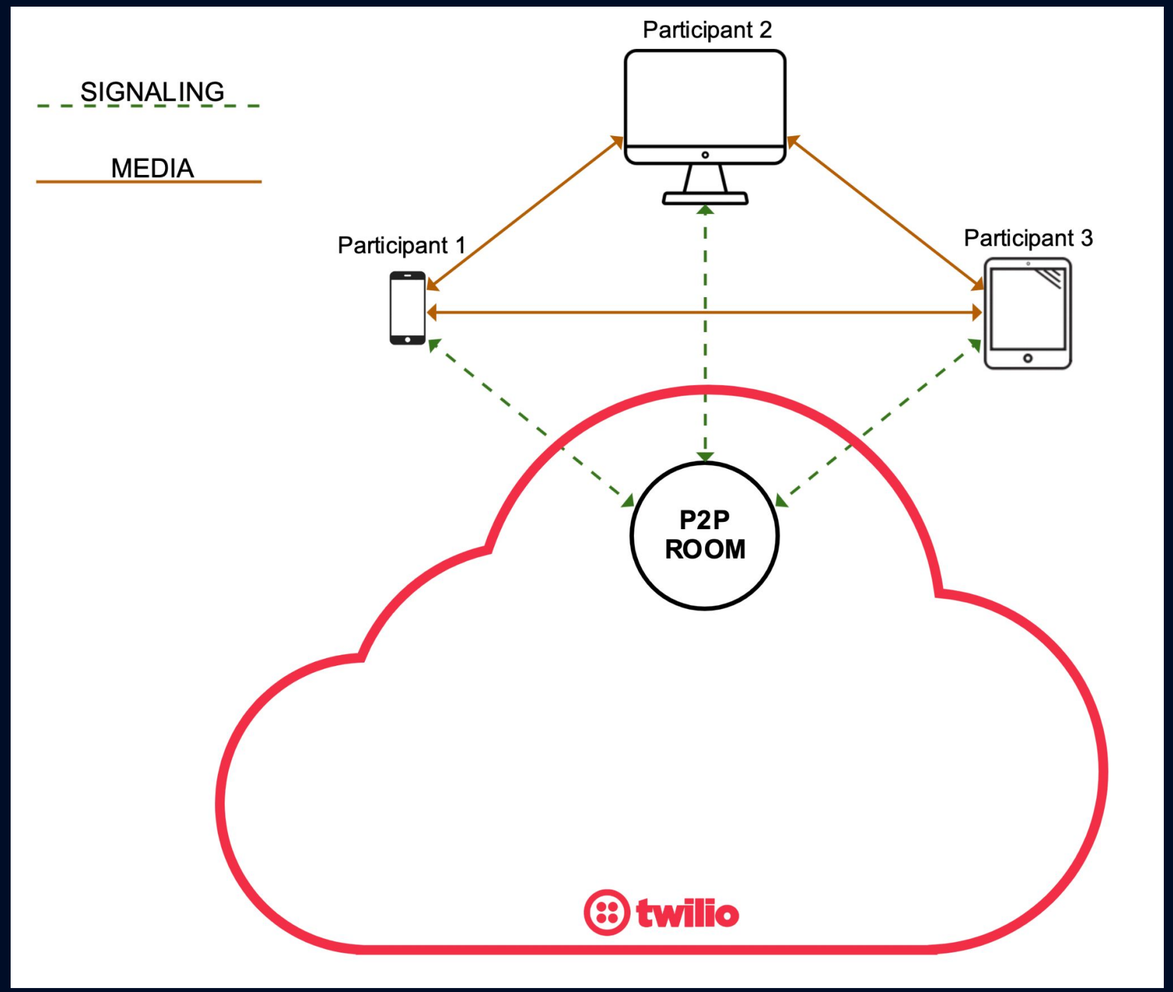
Kurento acquisition



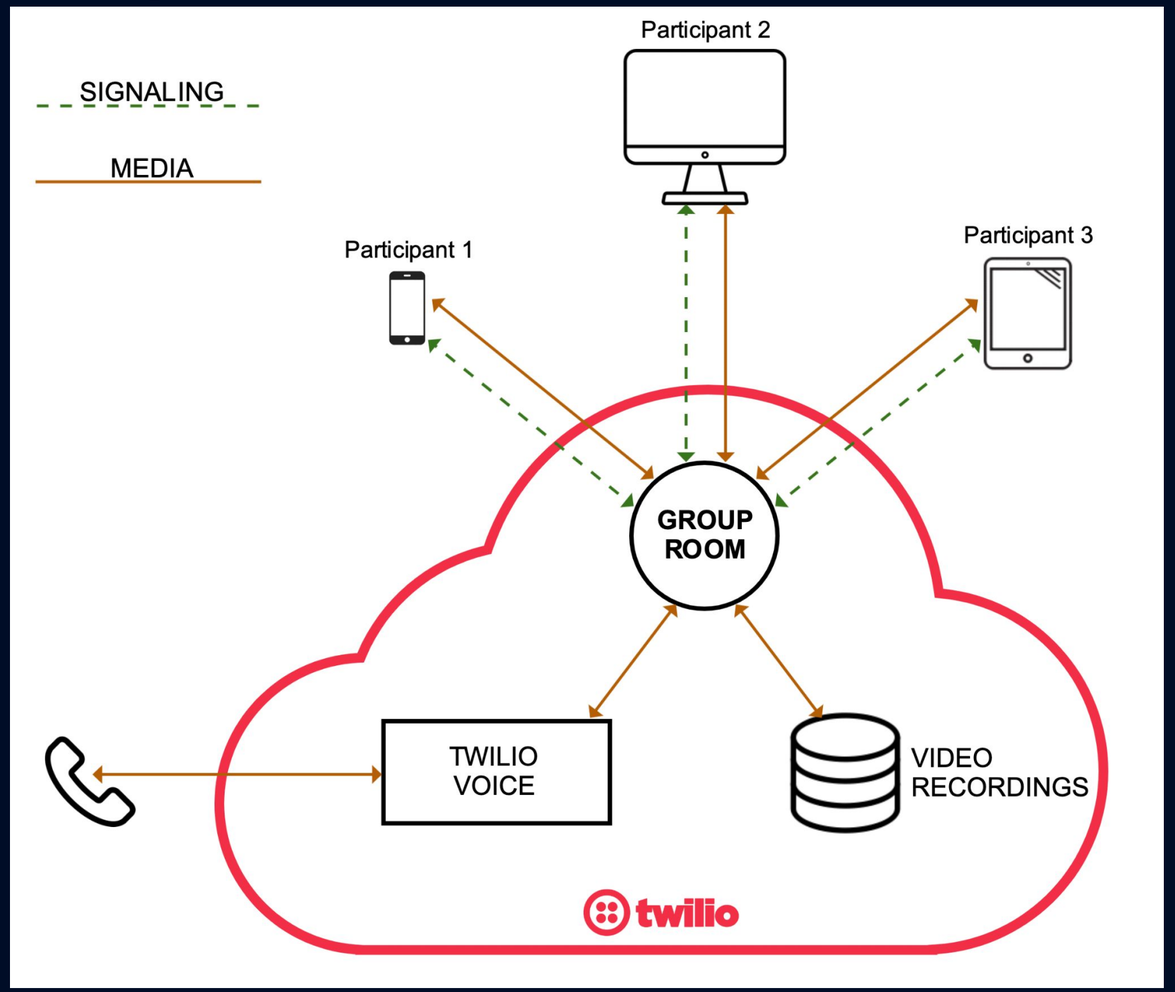


Video Rooms

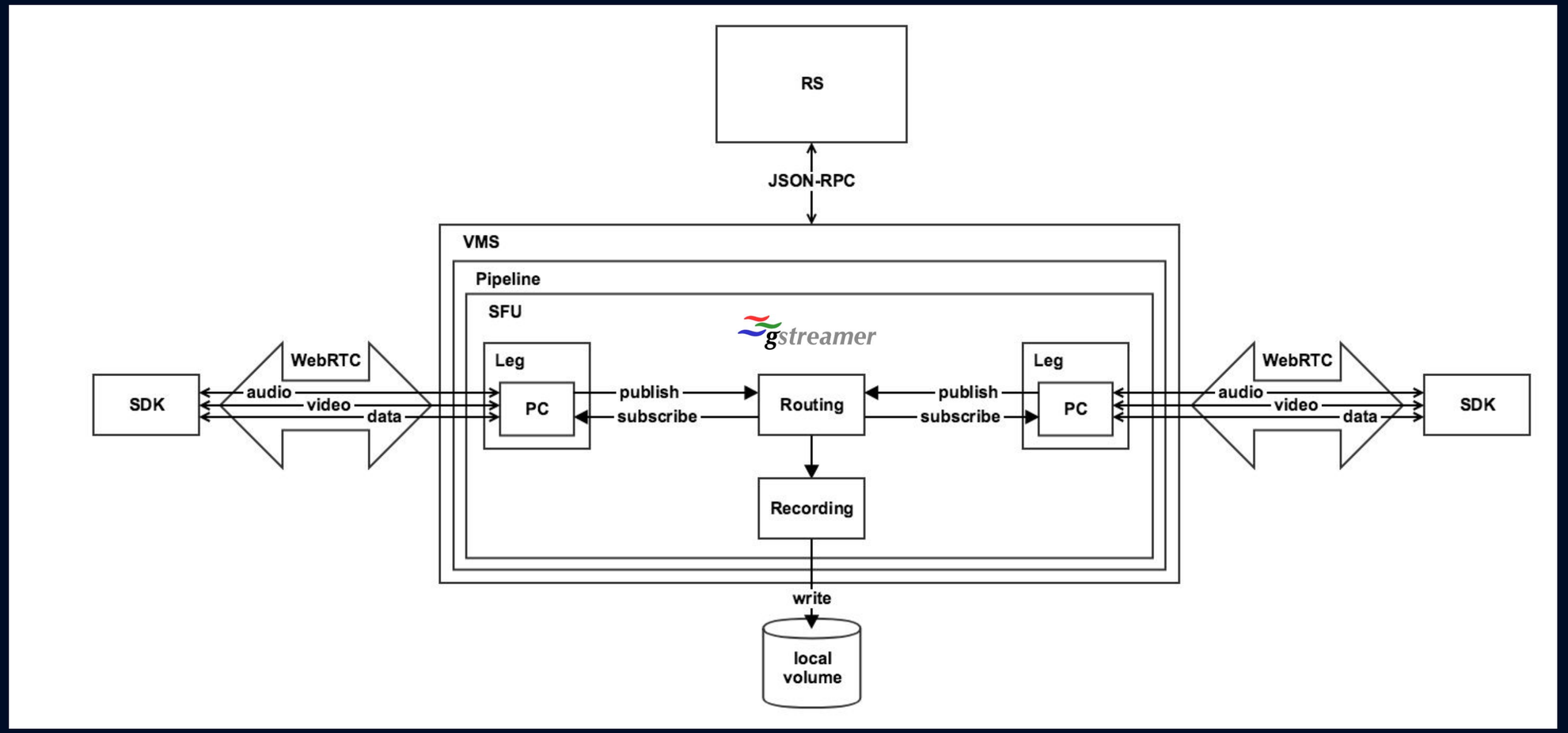
P2P rooms



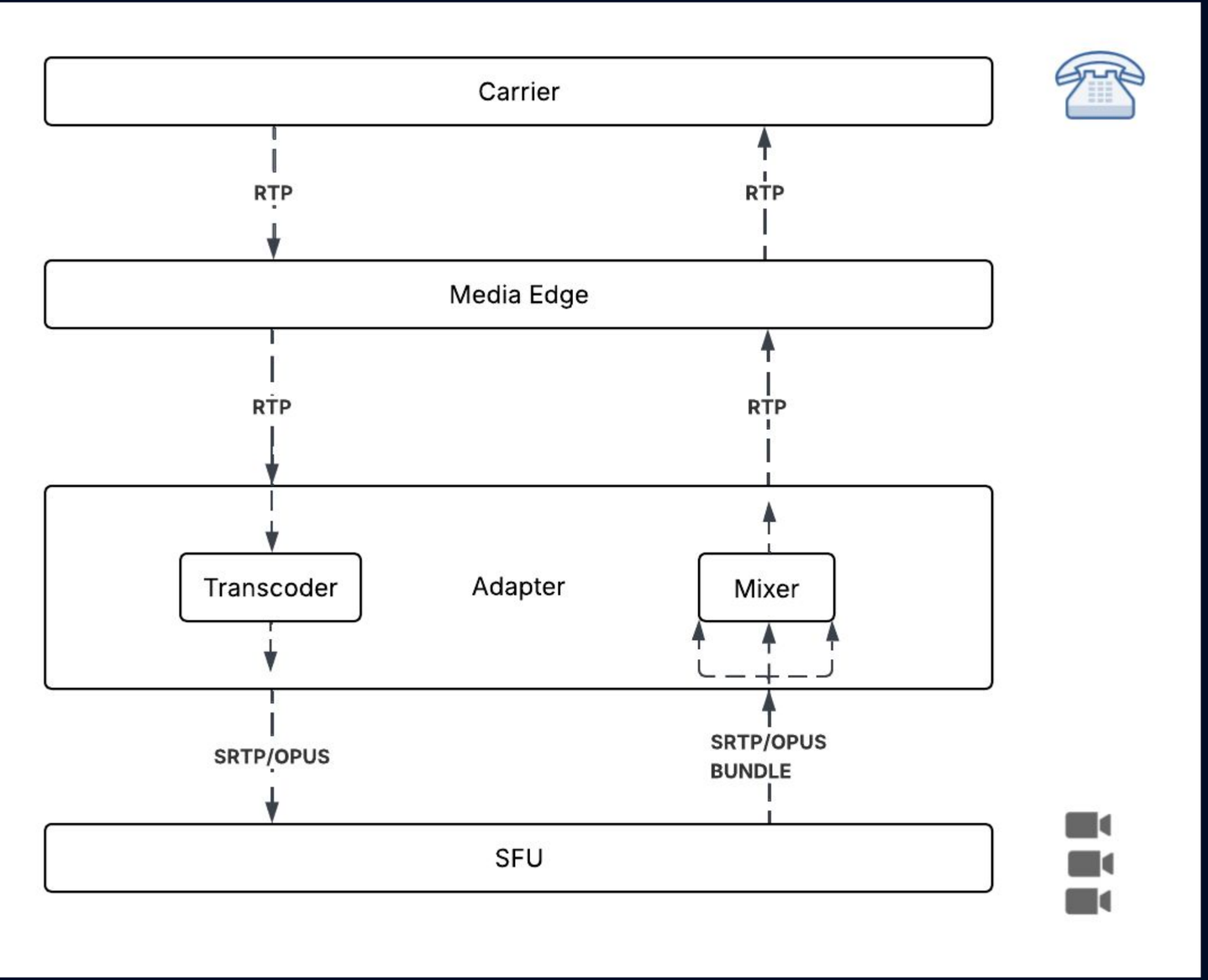
Group rooms



VMS SFU


RS
JSON-RPC
VMS
Pipeline
SFU
gstreamer
Leg
PC
publish
subscribe
Routing
Recording
write
local volume
WebRTC
audio
video
data
SDK

Connecting Programmable Voice and Video



https://twilio-engineering.atlassian.net/browse/VMEDIA-4530

Voice Media Framework

Enhanced AMD

<Gather> speech recognition

Voice Participants in Video Rooms

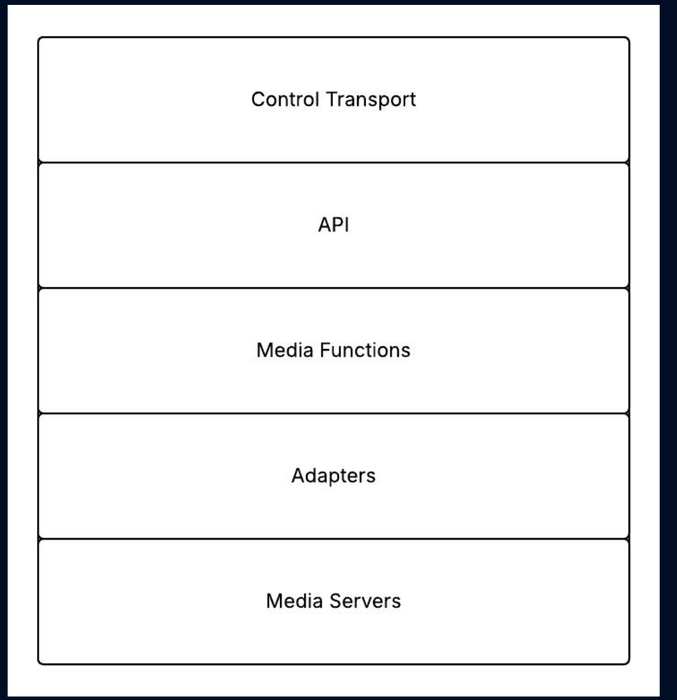
Media Streams

SIPREC SRC

Async AMD

Dialogflow CX

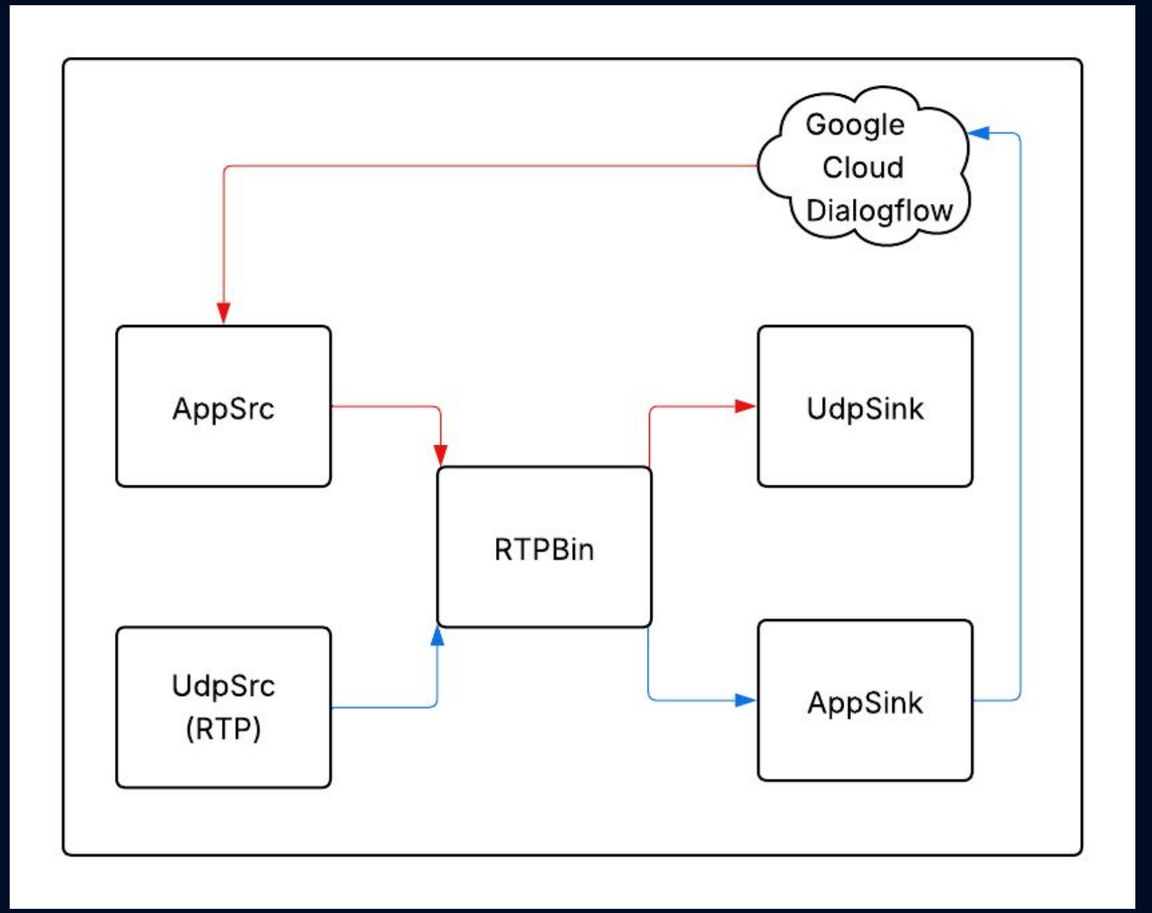
gstreamer



TME + GSTREAMER DEEP DIVE

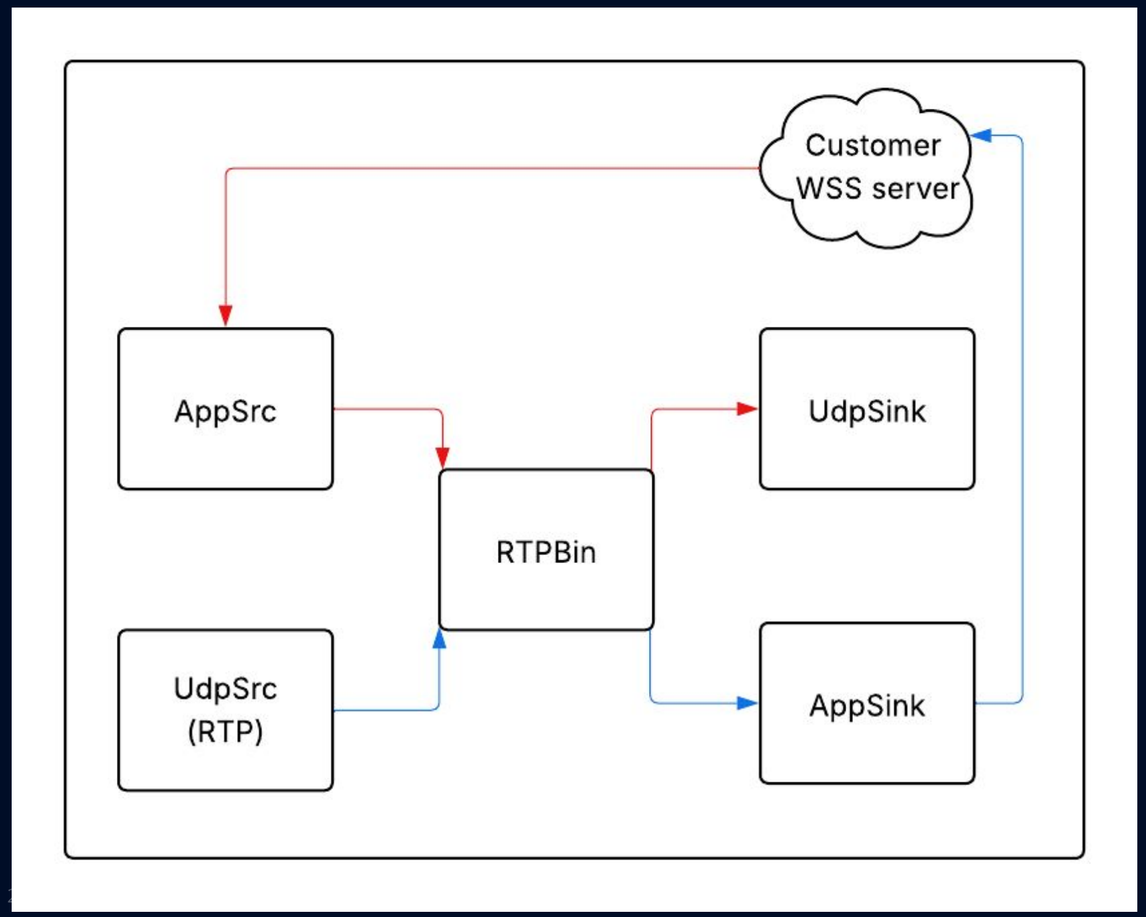
TME NLU

Dialogflow CX integration

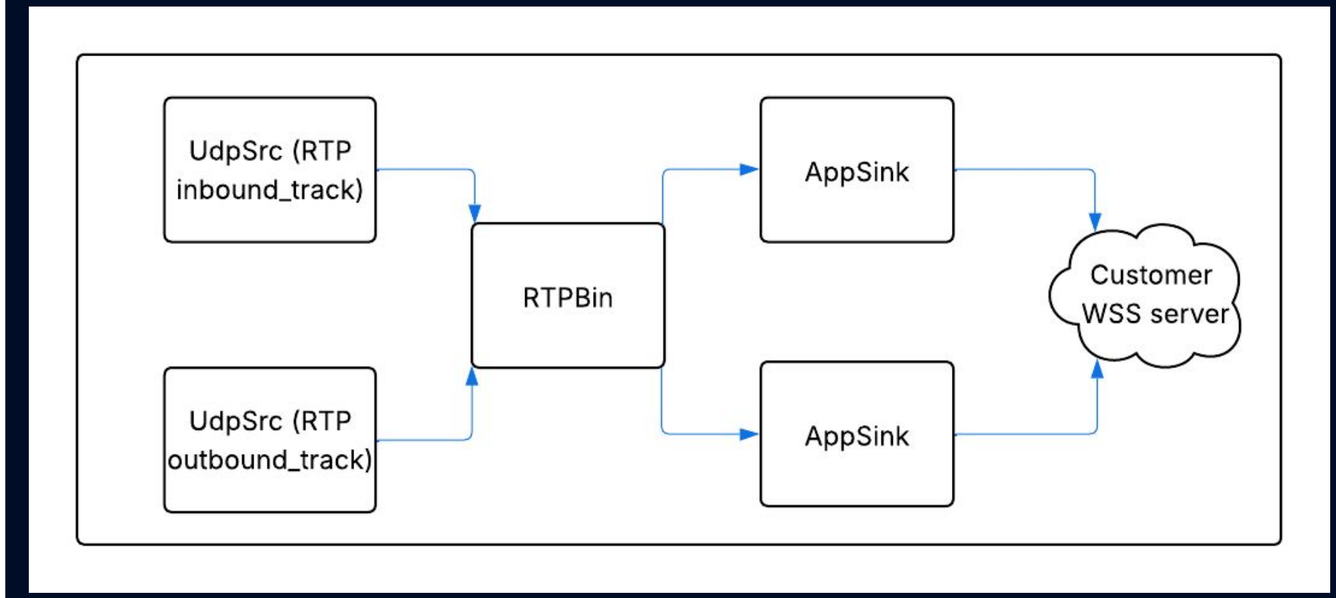


TME STREAMS

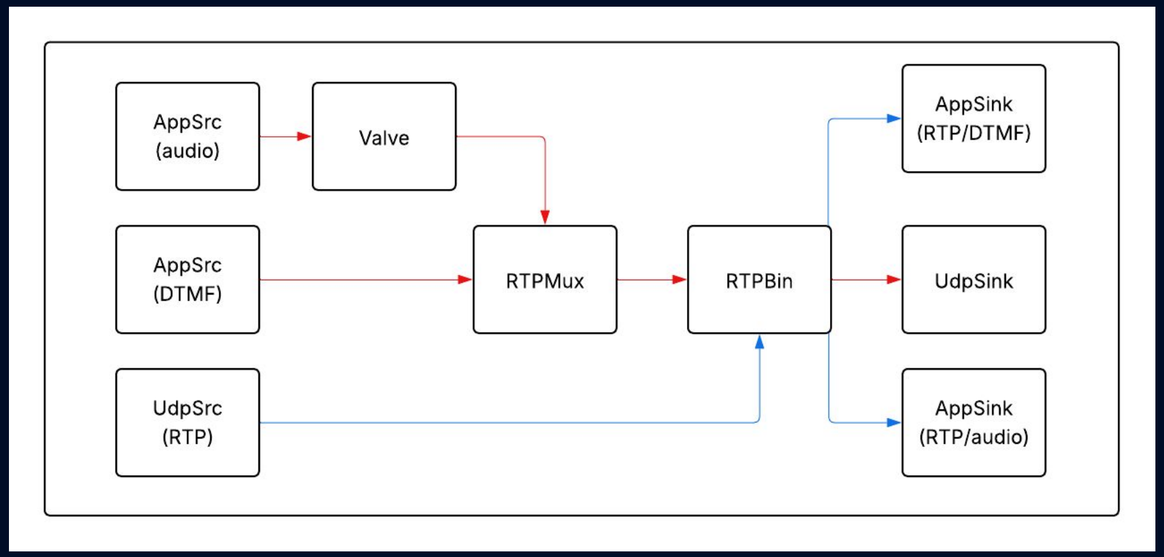
Bidirectional <Connect><Stream>



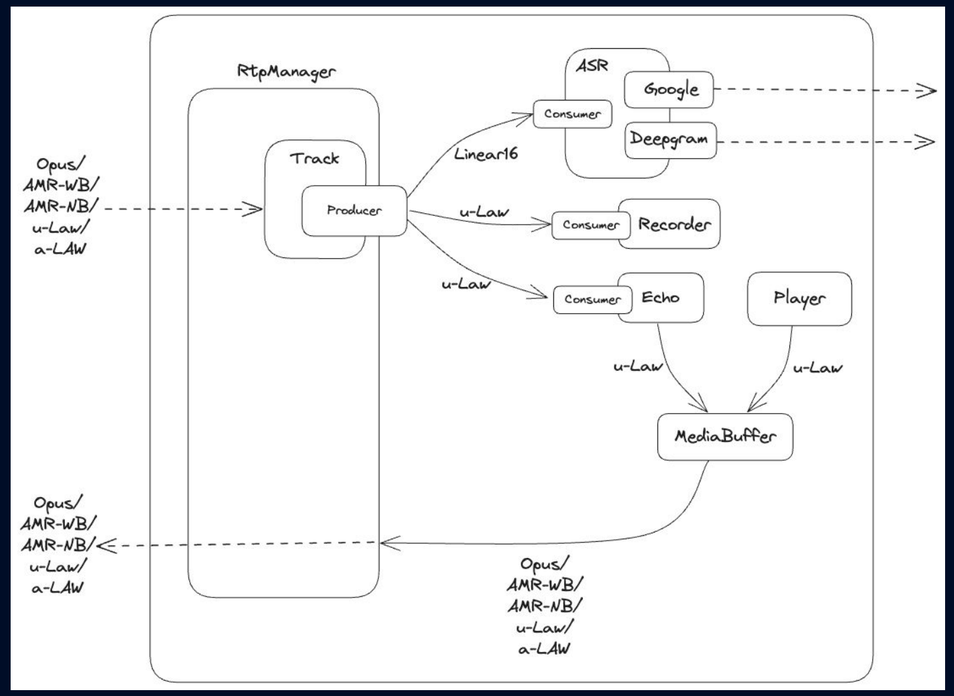
Unidirectional <Start><Stream>

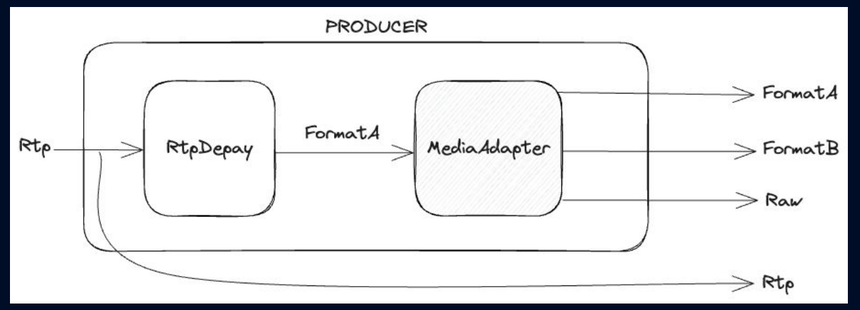


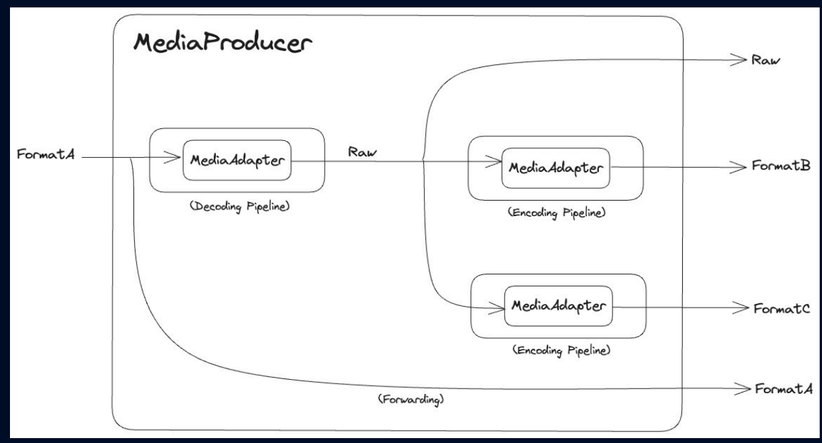
TME PROCESSOR



Composing pipelines







GStreamer learnings

Benefits

- Flexibility
- Safety
- Tooling
- Ecosystem

Challenges

- RTP bin
- Overhead of threads
- Audio + DTMF muxing
- SSRC changes
- WebRTC bin

Future

- tme-bridge in production
- Evolution of GStreamer in rust
- Threadshare elements

Q&A

THANK YOU