Video Streaming (RTSP, UDP, SFU) for Low Latency

Purpose

This document explains **why and how low-latency video streaming** is designed for robotics (robot dog / rover / drone), and clarifies the roles of **RTSP**, **UDP**, and **SFU** in your current architecture.

The target reader is an engineer working on **real-time control, perception, and autonomy**, not media playback.

---

1. Design Goal: What "Low Latency" Means in Robotics

In robotics, video is a **sensor**, not entertainment.

Priority order:

```
Latency  >  Jitter  >  Frame Freshness  >  Image Quality  >  Reliability
```

Key implications: - Dropped frames are acceptable - Waiting for retransmission is not - Showing the *latest* frame matters more than showing *every* frame

This immediately disqualifies most TCP-based streaming methods.

---

2. Why TCP-Based Video Is a Problem

TCP behavior

- Guarantees delivery
- Retransmits lost packets
- Enforces in-order delivery

Result for video

- One lost packet can stall the stream
- Latency grows unpredictably
- Control feedback becomes unsafe

Examples (NOT suitable for robotics)

- MJPEG over HTTP
- HLS / DASH

- RTMP

These are acceptable for monitoring, but not for control or autonomy.

---

3. UDP: The Foundation of Real-Time Video

Why UDP is required

UDP provides: - No retransmission delay - No head-of-line blocking - Predictable timing

Behavior: - Packet lost → frame degraded or dropped - Stream continues immediately

This matches robotics needs perfectly.

---

4. RTP: Real-Time Media Transport

Real-time video over UDP almost always uses **RTP (Real-time Transport Protocol)**.

RTP provides: - Sequence numbers - Timestamps - Jitter handling

Important clarification:

> RTP is the **media transport**. Other protocols (RTSP, WebRTC) *control* RTP.

---

5. RTSP Explained (What It Is and What It Is Not)

What RTSP is

RTSP (**Real-Time Streaming Protocol**) is a **control protocol**.

It is used to: - SETUP a stream - PLAY / PAUSE - TEARDOWN

What RTSP is NOT

- RTSP does **not** carry video data itself

Actual data path

```
RTSP (control)
→ RTP (media)
→ UDP (transport)
```

Latency impact

RTSP itself does **not** add significant latency. Latency depends on: - Whether RTP uses UDP (good) - Or RTP is tunneled over TCP (bad)

Your configuration correctly prefers **RTP over UDP**.

---

6. SFU (Selective Forwarding Unit)

Definition

An **SFU** is a real-time media relay that: - Receives RTP streams - Forwards them to multiple subscribers - Does NOT decode or re-encode video

What SFU does

- One stream in → many streams out
- Per-subscriber forwarding
- Minimal added latency (typically <10–15 ms on LAN)

What SFU does NOT do

- No video mixing
- No transcoding
- No AI processing

---

7. Why SFU Is Critical for Multi-Client Robotics

Without SFU:

```
Pi → Mac
Pi → iPhone
Pi → Tablet
```

Problems: - Multiple encodes - High uplink bandwidth - CPU pressure on Pi

With SFU:

```
Pi → SFU → N clients
```

Advantages: - Pi uploads once - Any number of viewers - Consistent latency

This matches your current architecture.

---

8. Where WebRTC Fits

WebRTC is NOT a replacement for RTP

WebRTC is best understood as:

> RTP/UDP + NAT traversal + congestion control + browser compatibility

Under the hood

- Still RTP
- Still UDP
- Often still SFU

Why WebRTC matters

- Native browser support (Safari / iOS)
- Secure transport (SRTP)
- Adaptive bitrate

Relationship to RTSP

- RTSP is common in engineering tools
- WebRTC is required for browser-based clients

Your system already prepares for this via WHEP endpoints.

---

9. Latency Comparison (Typical Values)

Method	Transport	Typical Latency
MJPEG / HTTP	TCP	300–1500 ms
HLS	TCP	5–30 s
RTSP (RTP/UDP)	UDP	80–150 ms
SFU + RTP/UDP	UDP	50–120 ms
WebRTC + SFU	UDP	30–100 ms

Your current **SFU + RTSP (RTP/UDP)** setup is already in the correct latency class for robotics.

---

10. Correct Mental Model (Important)

```
Camera
 → H.264 Encoder
 → RTP
 → UDP
 → SFU
 → Clients (Mac / iPhone / Browser)
```

Control and telemetry remain on a **separate, deterministic channel**.

---

11. Summary

- Low-latency robotics video must be **UDP-based**
- RTP is the actual media carrier
- RTSP controls RTP sessions
- SFU enables multi-client viewing with minimal latency
- WebRTC adds browser and NAT support on top of the same principles

Your current architecture correctly follows these rules and is suitable for: - Teleoperation - Object detection - Autonomous navigation

---

12. When to Evolve Further

Consider moving fully to WebRTC when: - iPhone/Safari becomes a primary viewer - You need WAN / NAT traversal - You want tighter congestion control

RTSP + SFU remains perfectly valid on a controlled LAN.

End of document