

Две части

RTP-метка	Считает тики оцифровки медиа от случайного старта. Не пакеты, не стенное время.
RTCP SR	Сопоставляет одно RTP-показание со стенным временем NTP в тот же момент.

Частоты RTP-медиа часов

Аудио Opus	48 000 Hz — 960 тиков на 20-мс пакет (RFC 7587).
Аудио G.711	8 000 Hz — 160 тиков на 20-мс пакет (RFC 3551).
Видео	90 000 Hz — 3 000 тиков на кадр при 30 fps (RFC 3551).

Медиа-часы → стенные часы

Формат NTP	64-бит фикс. точка: секунды от 0h UTC 1.01.1900; 32 целых + 32 доли.
Пара из SR	RTP-метка R0 отображается в NTP-время T0 (тот же момент).
Перевод	$\text{стен}(R) = T0 + (R - R0) \div \text{частота часов}$. Делать по каждому потоку.
Lip-sync	Оба потока на одних стенных часах; рендер совпадающих аудио + кадра вместе.

Когда синхронизация ломается

Дрейф часов	разница кварцев	Lip-sync уходит ровно; хуже за долгий звонок.
Сетевой джиттер	неровное прибытие	Синхронизация дёргается, но не в одну сторону.
Нет / неверный SR	нет якоря	Потоки в норме порознь, навсегда не в фазе.
Задержка защёлк.	лимит RTCP	Первые ~1-5 с не в фазе, затем защёлк. (RFC 6051).

Запомнить

- ☐ Аудио и видео — отдельные RTP-потоки с отдельными часами со случайным стартом.
- ☐ RTP-метка считает тики оцифровки медиа, а не пакеты и не стенное время.
- ☐ Аудио тактируется на 48 000 Hz (Opus) или 8 000 Hz; видео — на 90 000 Hz.
- ☐ RTCP-отчёт отправителя сопоставляет одно RTP-показание со стенным временем NTP.
- ☐ Одна пара на поток ставит оба на одну шкалу — это и есть lip-sync.
- ☐ Отчёты должны идти повторно; каждый свежий заново привязывает против дрейфа.
- ☐ RFC 6051 и WebRTC abs-capture-time несут соответствие in-band для быстрой синхронизации.
- ☐ Мидлбокс, регенерирующий SR по своим часам, тихо ломает синхронизацию.