

# QoE метрика в видеоплатформе Яндекса

Коровин Василий  
Yandex Infrastructure



# Василий Коровин

Аналитик-разработчик  
Player Development Department



# Infrastructure

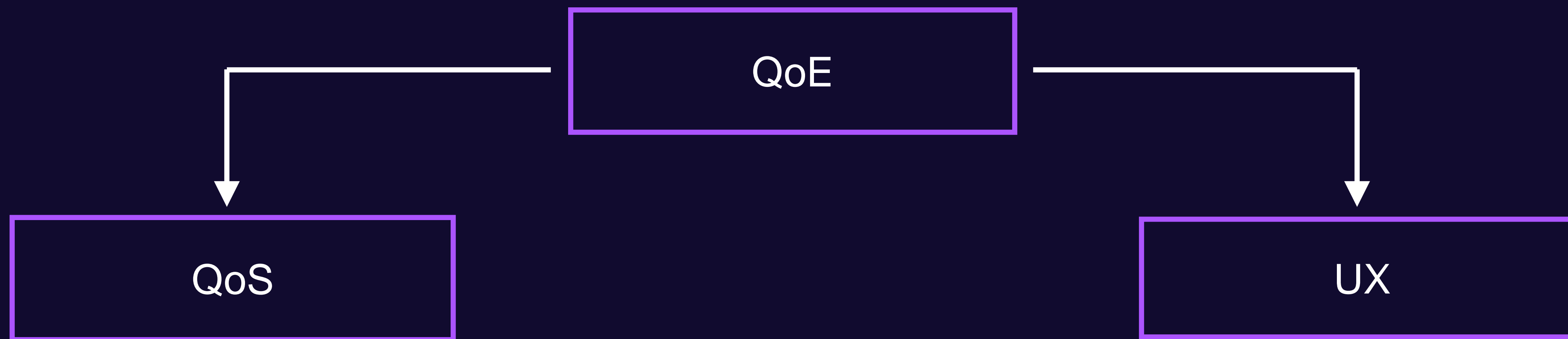
Мы создаём и предоставляем внутреннюю инфраструктуру Яндекса — фундамент из продуктов и технологий, на базе которого тысячи инженеров разрабатывают, деплоят и эксплуатируют основные сервисы Яндекса: Поиск, Такси, Маркет, Алису, Кинопоиск и многие другие.

# Что это за аббревиатура — QoE?

QoE — Quality of Experience — метрика качества восприятия



# Чуть-чуть теории



## Quality of Service

Объективные  
параметры, которые  
мы можем измерять:

- Логи

## User experience

Субъективное  
восприятие сервиса:

- Эмоции
- Опросы

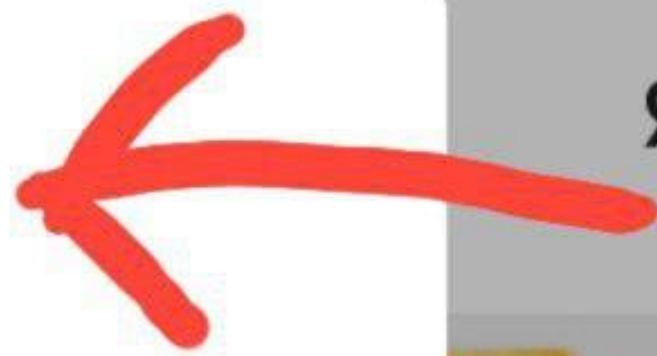
# Мы тут все собрались оценить счастье пользователей. Но чем и как?

MOS (Mean Opinion Score) —  
среднее арифметическое  
по метрикам, по которым  
мы оцениваем качество  
работы системы





Василий  
★ 5.00



Яндекс Go

Моё местоположение  
Льва Толстого 16

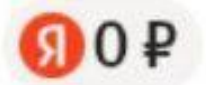
Помощь рядом  
Округление



Способы оплаты  
Подтвердите карту



Карта Пэй



Плюс



Бизнес-аккаунт



МультиТранспорт

Стать водителем

Семейный аккаунт  
Поездки по семейной карте для близких

окааты

Безопасность



Проблемы: большая смещённость, большая зависимость от неподконтрольных факторов

Многие наши правила и процедуры мешают хорошо выполнять работу.

Совершенно не согласен

Не согласен

Скорее не согласен

Скорее согласен

Согласен

Совершенно согласен

1. Почти во всем моя жизнь соответствует моему идеалу.

Совершенно не согласен

Не согласен

Немного не согласен

Неопределенно (и согласен, и не согласен)

Немного согласен

Согласен

Совершенно согласен

Я считаю, что получаю неплохую оплату за свой труд.

Совершенно не согласен

Не согласен

Скорее не согласен

Скорее согласен

Согласен

Совершенно согласен

# Примеры из жизни

## G-модель в сетевых играх

$$\text{MOS}(x) = \max(4.33 - 3.08^{-9}x^3 + 1.18^{-5}x^2 - 1.15^{-2}x, 1),$$

где  $x = \min(\text{Ping} + 0.686\text{Jitter}, 650)$

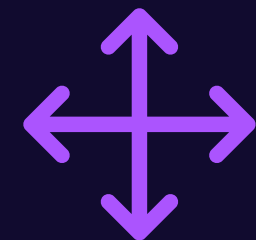


# Пример оценки качества видео для компании Huawei U-vMOS

5-балльная шкала оценки контента зависит от следующего:



Контент — включает разрешение, цвета



Загрузка — параметры сети, скорость загрузки

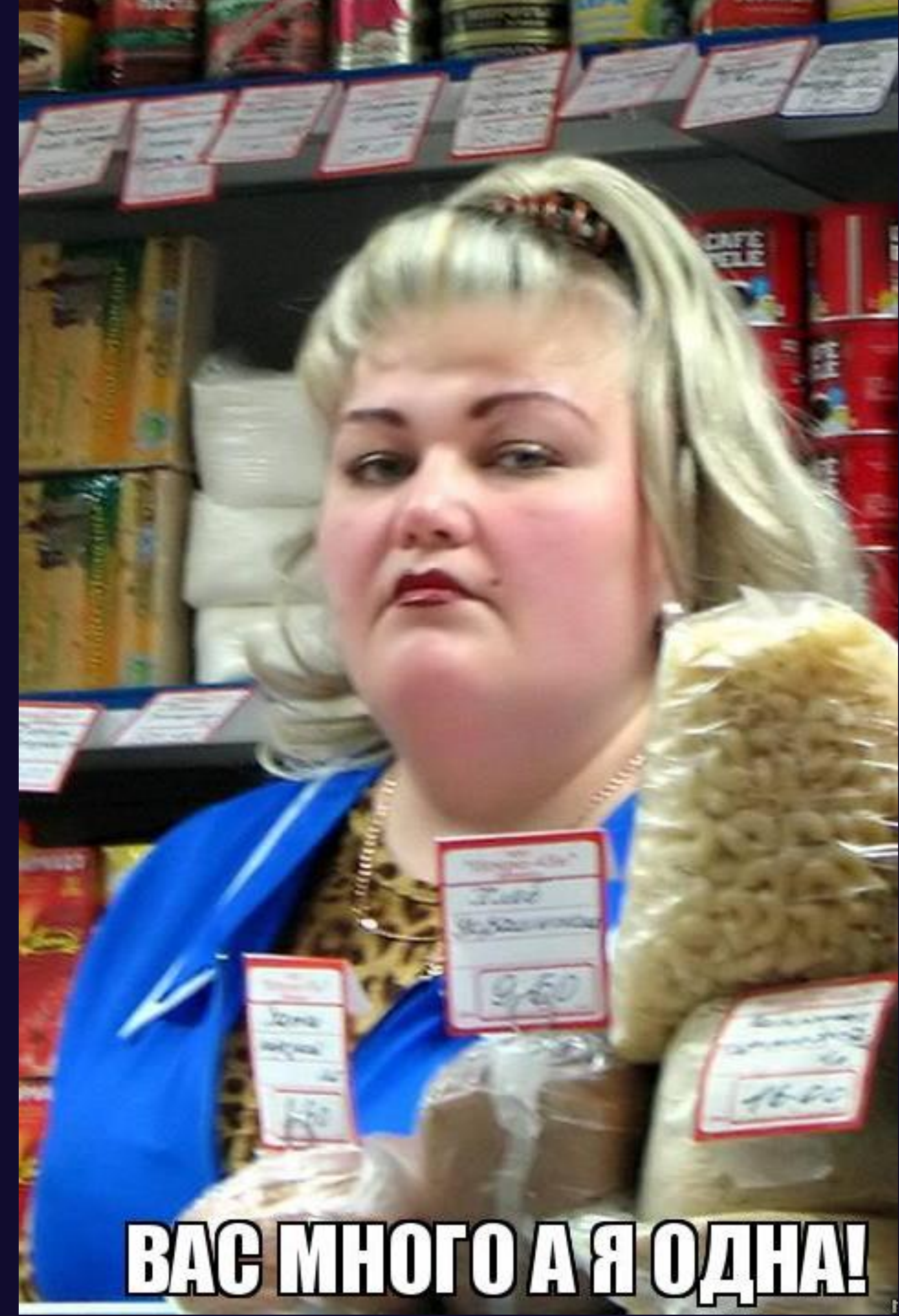


Воспроизведение — оценка качества вещания

$$U\text{-vMOS} = (sQuality - 1) \left( 1 + \frac{\alpha(sInteraction - 1) + \beta(sView - 1)}{4(\alpha + \beta)} \right)$$

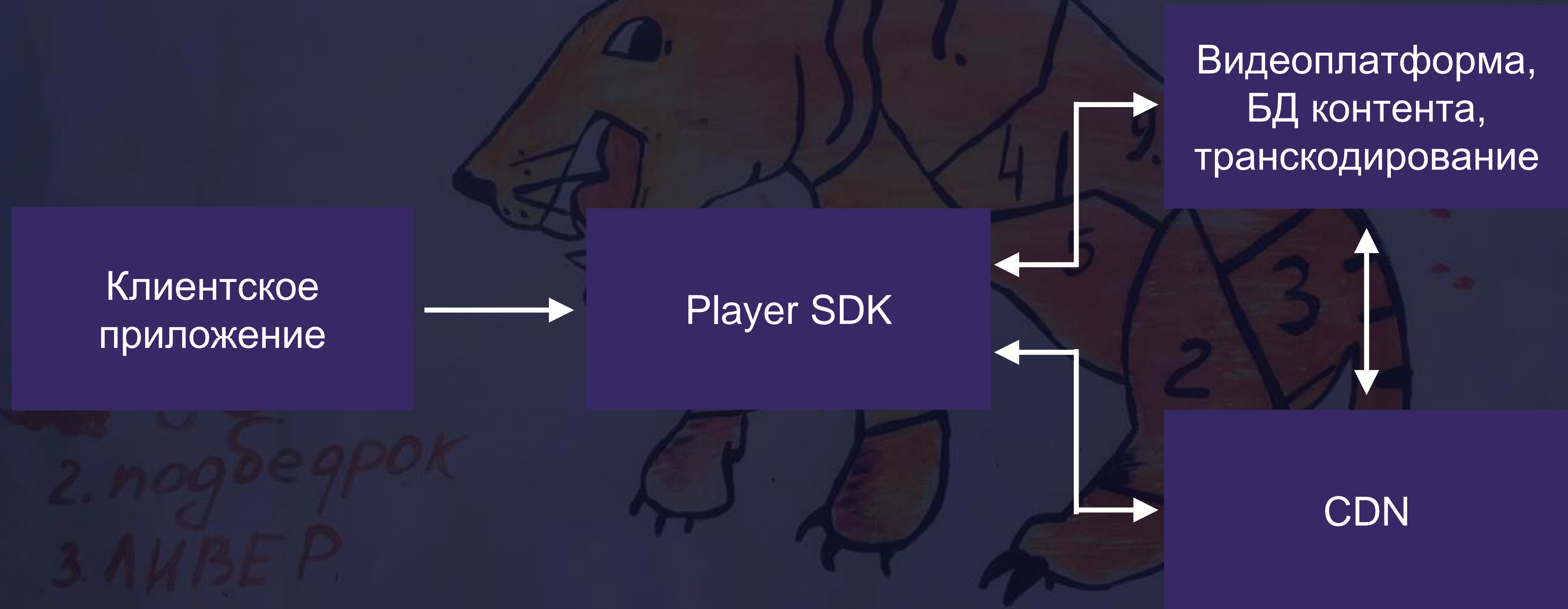
# А зачем нам QoE-метрика, что мы от неё хотим?

- 1 Информация о текущем состоянии сервиса
- 2 О причине проблем на сервисе
- 3 О том, как влияют наши изменения на сервис



**ВАС МНОГО А Я ОДНА!**

# Как устроена видеоплатформа



# Что же было раньше? Появление плеера

На что смотрим: статусы ответов CDN

Метрика: доля успешных запросов

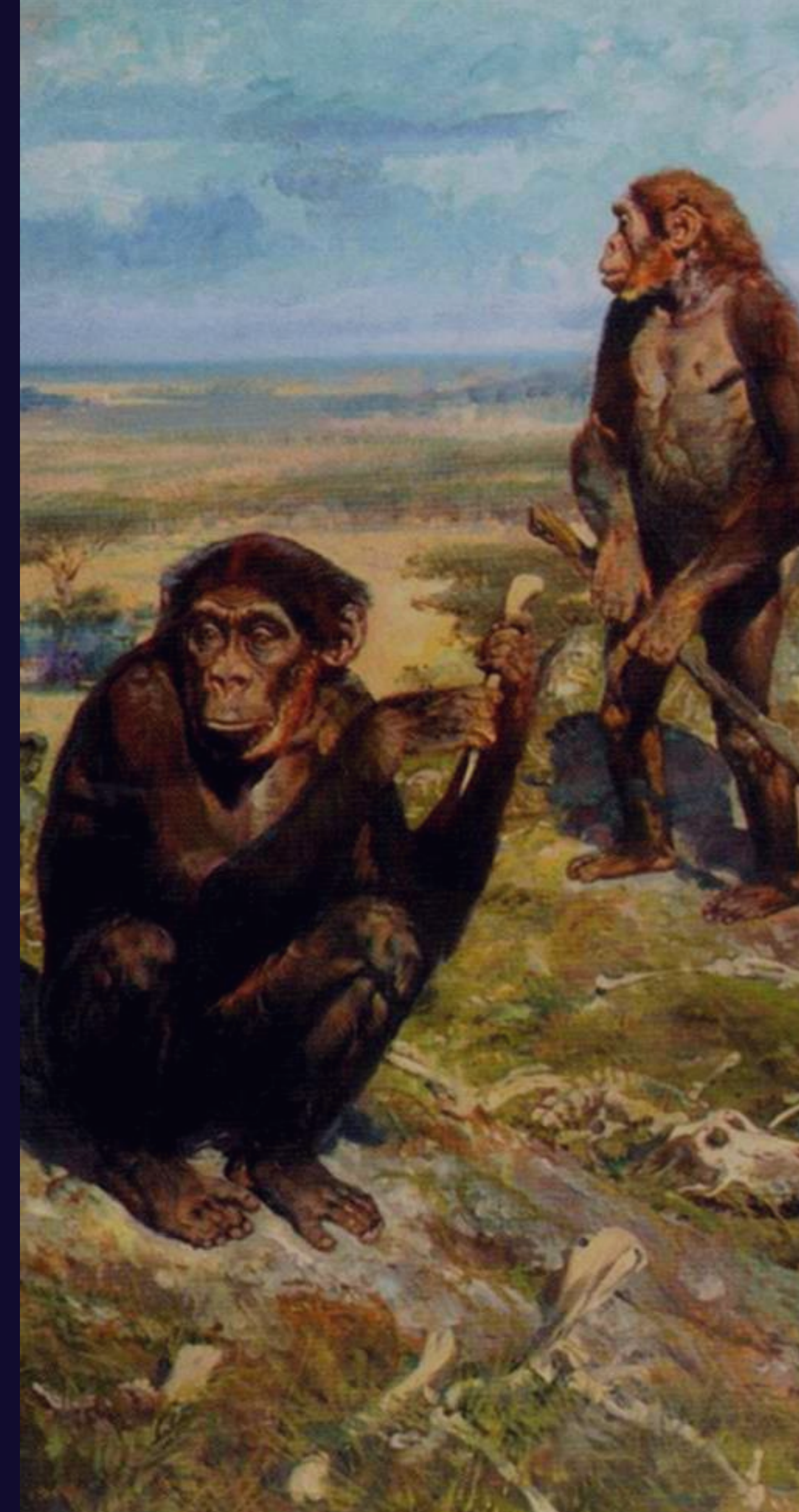
## Плюсы:

видим состояние CDN

## Минусы:

- не видим всего остального
- не понимаем, что у клиента
- не понимаем, почему так
- и много-много чего ещё

Это лишь состояние CDN



# Первые логи

## Логирование ошибок

Метрика: количество ошибок

### Плюсы:

уже видим причины проблемы

### Минусы:

- по сути, это работает / не работает плеер
- нет возможности фильтрации
- нет никаких «промежуточных» состояний



# Начинаем собирать heartbeat'ы...

Метрика: количество heartbeat'ов

## Плюсы:

это может служить сигналом  
о негативных или позитивных  
изменениях, но...

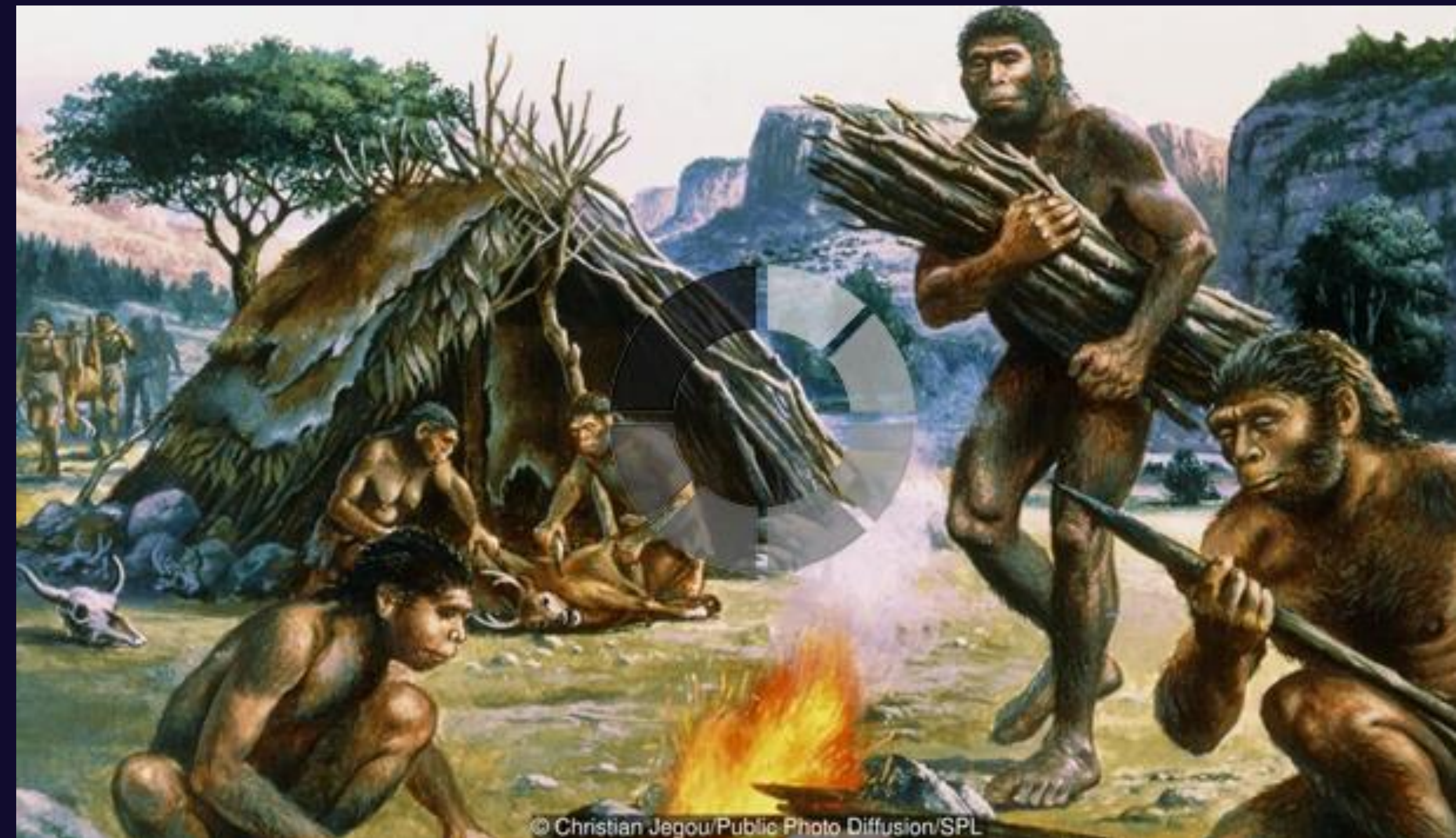
## Минусы:

непонятны причины  
этих изменений



# И что такое столлды?

- Таким словом мы называем задержки в видео (калька с английского)
- Метрика: отношение времени столлды ко всему времени работы плеера



# 1 Логи принимают современный вид



Сессия видеосмотра — комбинация VSID (идентификатор созданного) и videoContentID (идентификатора контента)

# 2 Появляются понятия сессии просмотра и событий



События — Start, SetSource, FatalError, PlayerAlive и т. п.



# Unhappy end

Посессионная метрика, основанная на двух последних событиях

## Плюсы:

- неплохо определяем плохие сессии
- понятность

## Минусы:

- надо постоянно следить за актуальностью событий
- есть проблемы в хороших сессиях



# Наша текущая цель — получить метрику, которая будет:

1

Простая

2

Понятная

3

Чувствительная  
к изменениям

4

Коррелирующая  
с нашими внутренними  
представлениями  
о прекрасном

# Мы решили провести брейнштурм, чтобы понять, от чего зависит негативный опыт пользователя

Качество картинки

Качество звука

Качество сети

Фатальные ошибки плеера

Отказы

Задержки в процессе

Задержка старта

# В чём заключается наша идея

Инициализация

Прерывания

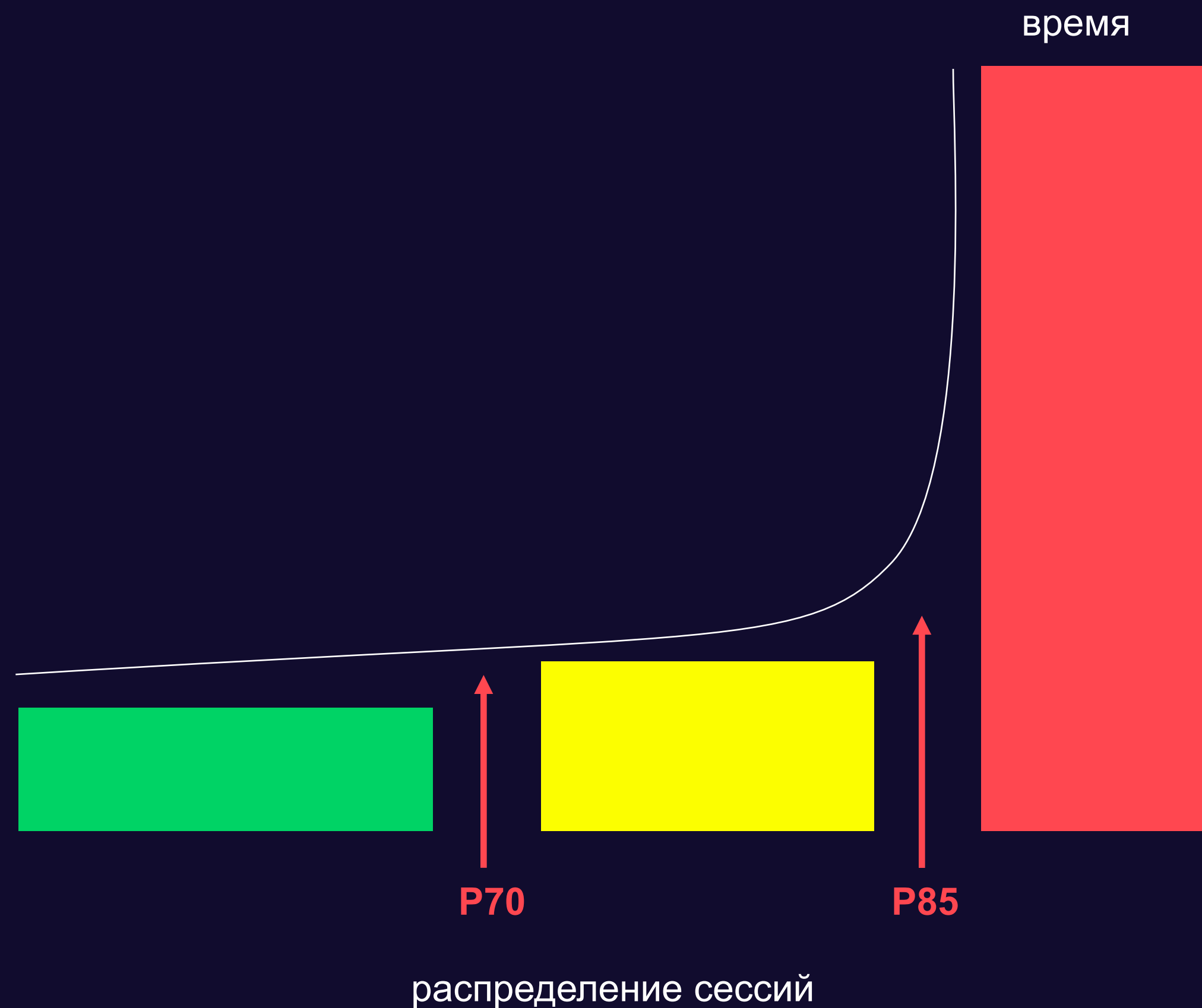
Отказ

Фатальные ошибки

Качество видео

Для каждого выбранного нами параметра мы решили построить критерии хорошей (1) / средней (2) / плохой сессии (3)

# Инициализация



Мы исходим из нескольких соображений:

1. Все три категории должны быть
2. Адекватность с точки зрения жизни
3. Отзывчивость на наши изменения

Нами выбраны границы **P85** и **P70**

# Буферизация прерывания

- Чем дольше столлды, тем хуже
- Чем больше столлдов, тем хуже



- Количество столлдов
- Длина максимального столлда

Эти две мысли мы преобразуем в два считаемых параметра

Также строим распределение перцентилей и принимаем границы как P85 и P70

# Фатальные ошибки

**Мы считаем, что фатальные ошибки - это плохо  
нет/есть**

## Отказы

**Отказы — это сессии, которые имеют столлды  
(трафик был, но при этом время просмотра = 0)**



# Как можно оценить качество видео?

VMAF, SSIM — довольно дорого и требовательно к ресурсам

**Нашу оценку качества видео**

**мы взяли у компании Мух.**

**Это оценка показываемой**

**нами картинки от 100 до 0.**

Для того чтобы разобраться, надо сначала установить, что такое *upscale*



$$upscale = \max\left(\frac{Csize}{Quality} - 1, 0\right)$$

# Но как быть, если качества нет?



$$upscale = \max\left(\frac{\min(Csize, MaxQuality)}{Quality} - 1, 0\right)$$

# Ещё вопрос пользовательского качества



$$upscale = \max\left(\frac{\min(Csize, MaxQuality, UserCap)}{Quality} - 1, 0\right)$$

**1** Считая *upscale* для каждого момента времени сессии, мы можем посчитать **КОЭФФИЦИЕНТ КАЧЕСТВА ВИДЕО**

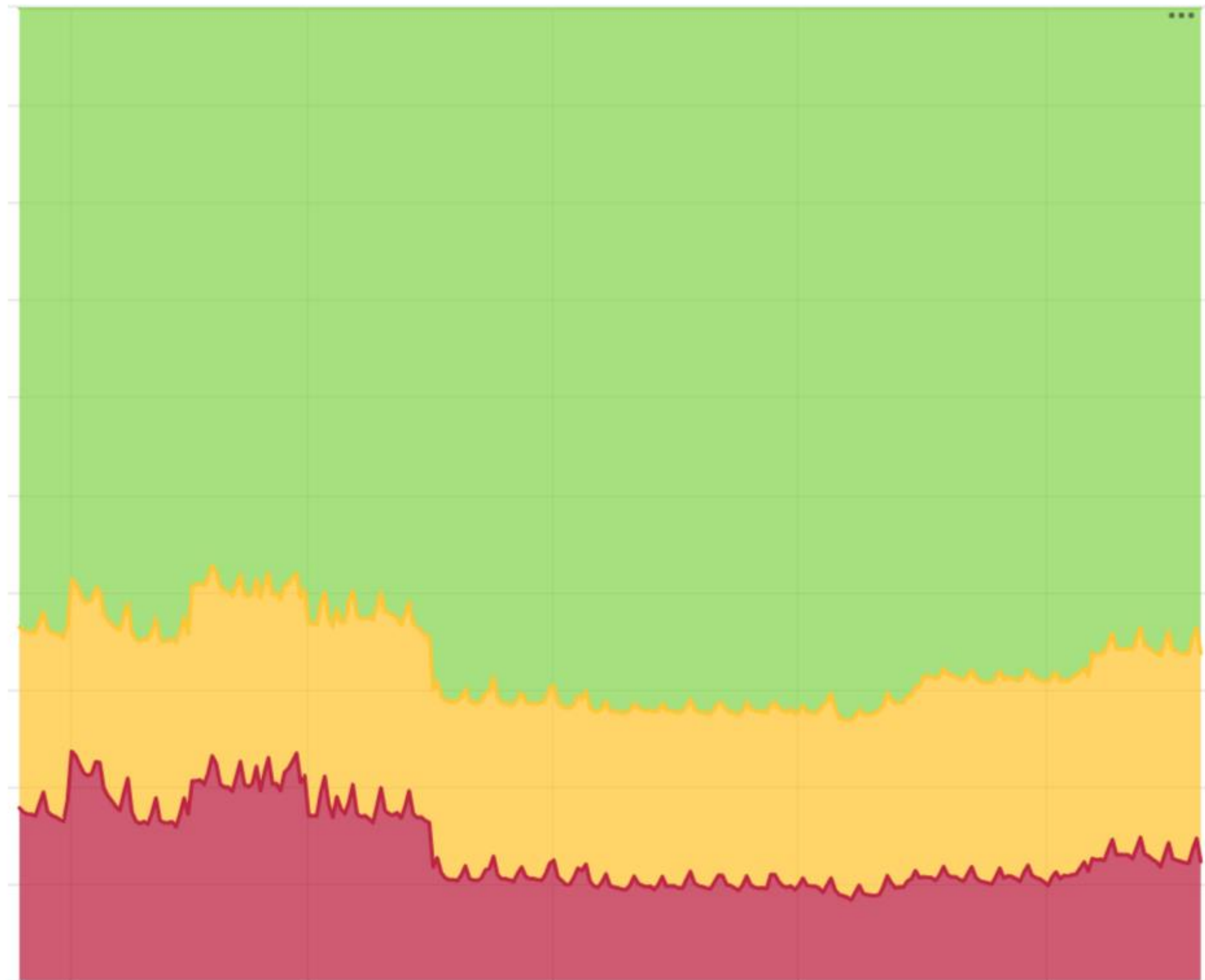
$$VQS = 100e^{-0.33(0.15 \max(\textit{upscale}) + 0.85\text{avg}(\textit{upscale}))}$$

**2** Границы для *VQS* мы принимаем как 95 (это соответствует постоянному *upscale* 15%) и 87.5 (постоянный *upscale* примерно 50%)

Итого мы посчитали для каждой сессии 6 параметров,  
а суммарную оценку сессии мы делаем как

$\text{MAX}(P1, P2, \dots P6)$

и строим ежедневный график распределения  
красных/жёлтых/зелёных сессий



Доля зеленых сессий



68

abc\_service\_name Нет выб...

device Нет выбранных зна...

from\_block Нет выбранны...

service Нет выбранных зн...

labels\_from Нет выбранны...

day 19.12.2022 - 09.10.2023



# Плюсы такого подхода



Простота



Гибкость

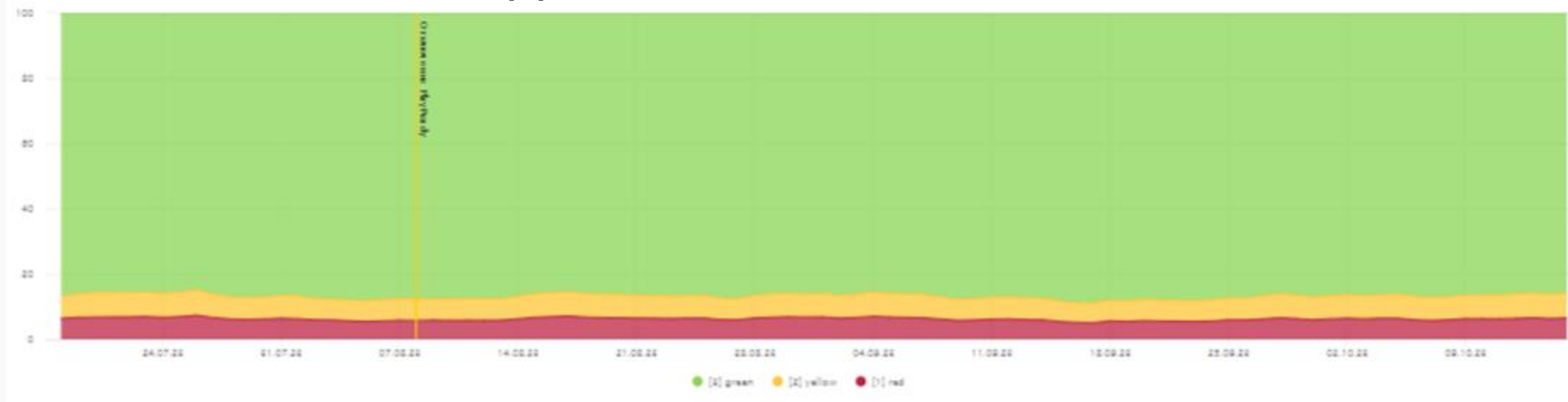


Лёгкость в деле  
выявления проблемы

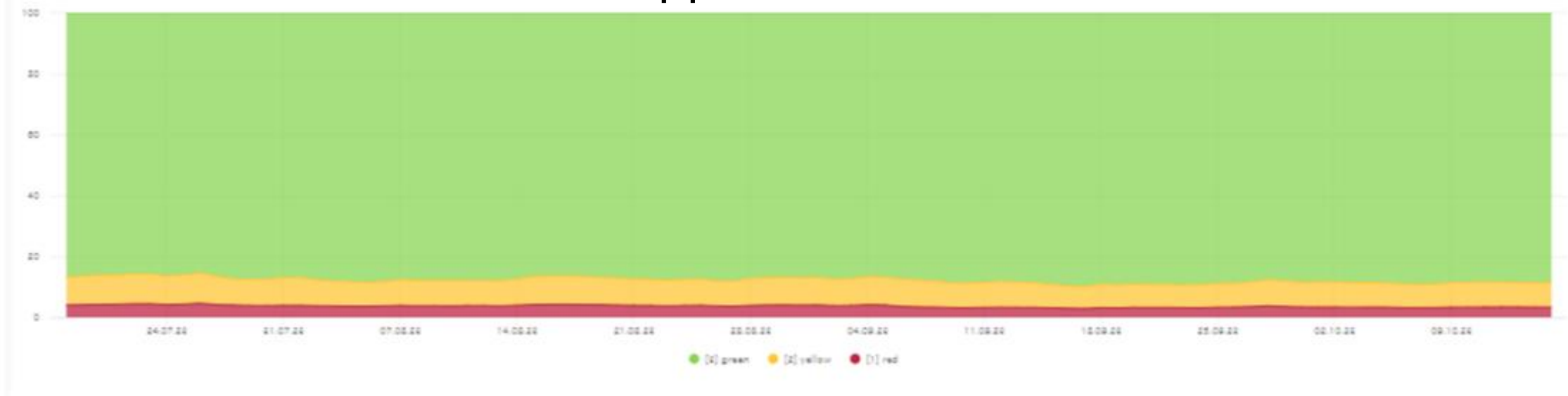


Наглядность

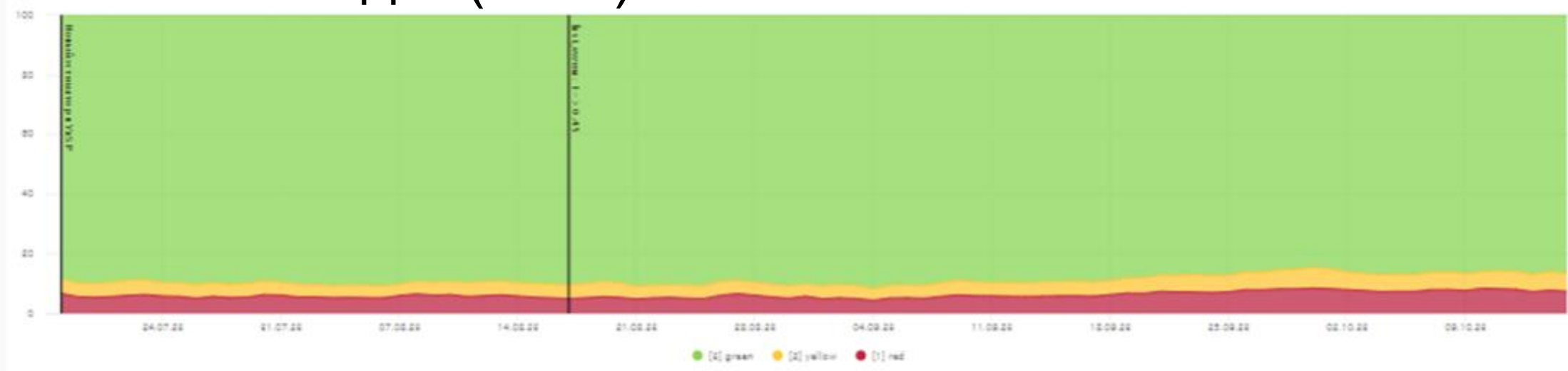
### Количество столлдов



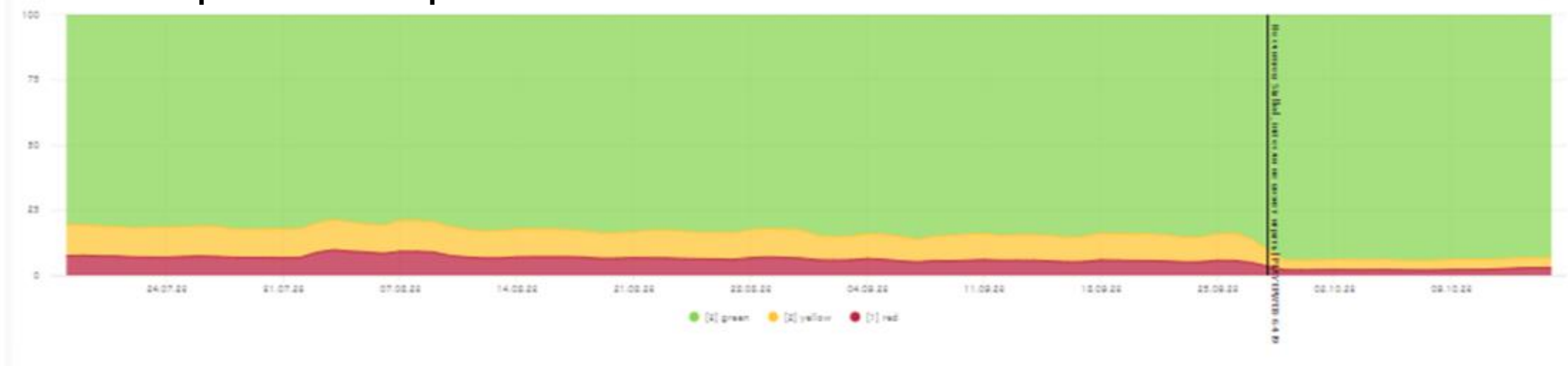
### Максимальный столлд



### Качество видео(VQS)



### Инициализация



### Отказы



### Фатальные ошибки

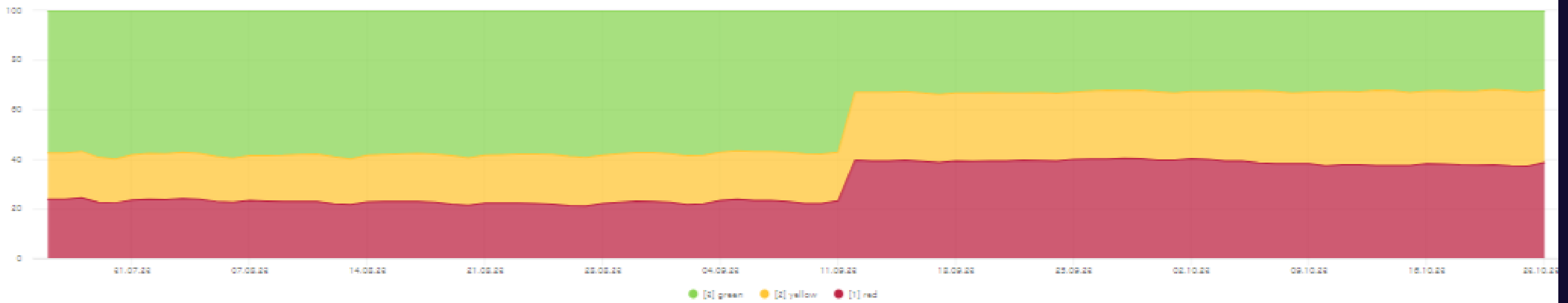


**Небольшой пример**

**ИЗ ЖИЗНИ**

Все дашборды

Общее состояние

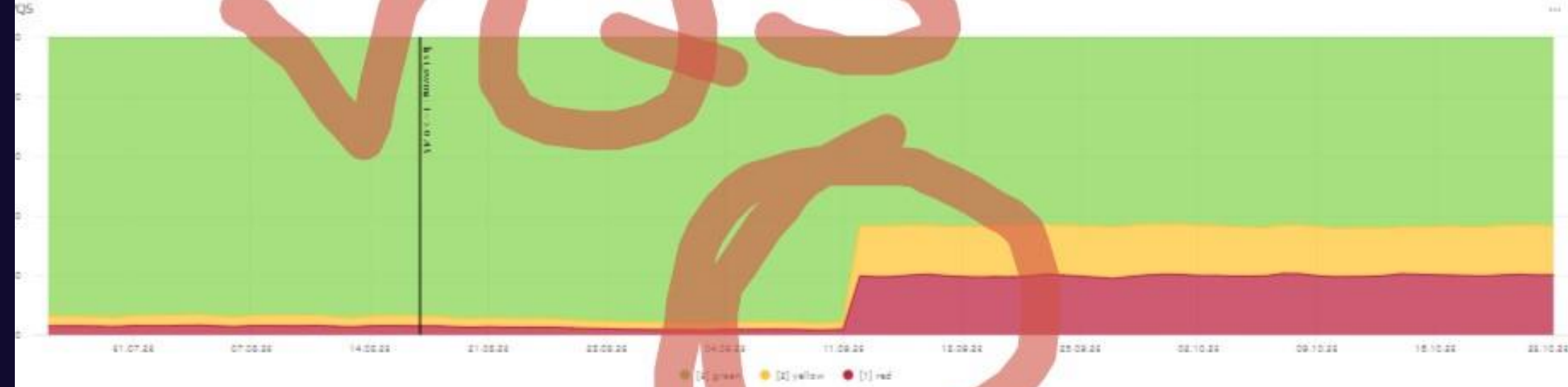




Количество столбов неинициализации, короткая видео: желтые - более 3, красные - более 3  
 длинные видео: желтые - более 4, красные - более 14



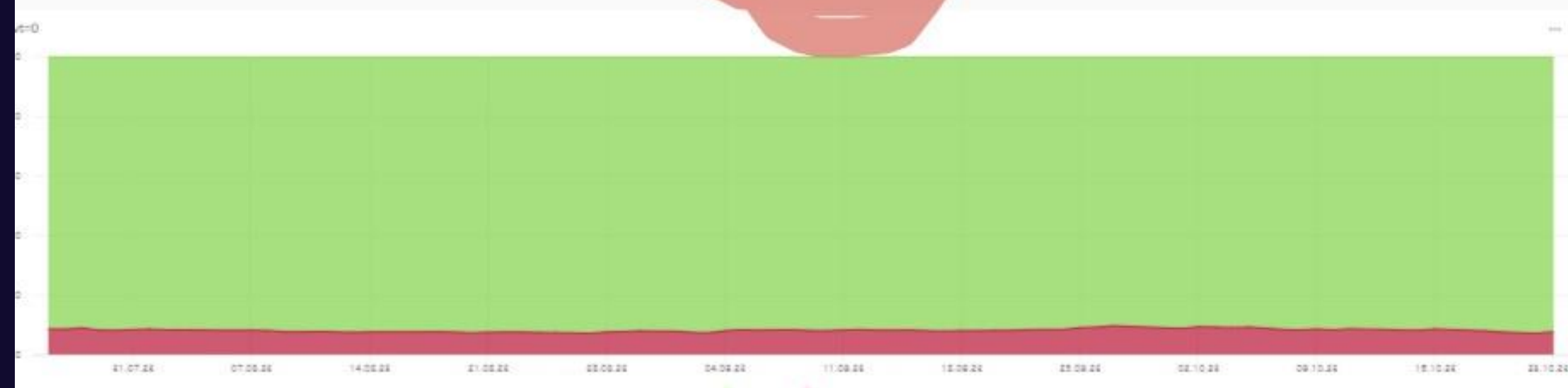
Длина максимального столба неинициализации:  
 короткая видео: желтые - более 0.45с; красные более 3.19с  
 длинные видео: желтые - более 3.27с; красные более 21.63с



Q5 - qds менее 95 (аналогично постоянному адеквату в 1.15 раз), красные - qds менее 87.5 (аналогично постоянному адеквату в 1.4 раза)



Время инициализации:  
 короткая видео: желтые - более 2.73с, красные - более 5.30с  
 длинные видео: желтые - более 5.35с; красные - более 12.86с



vos

# Дальнейшие шаги

- 1 Звук
- 2 Качество сети
- 3 Пользовательский опыт
- 4 Потерянные кадры

**Спасибо  
за внимание!**

**Коровин Василий**  
Yandex Infrastructure

# Подписывайтесь на наши канал и чат



Канал



Чат

