
Яндекс

Яндекс

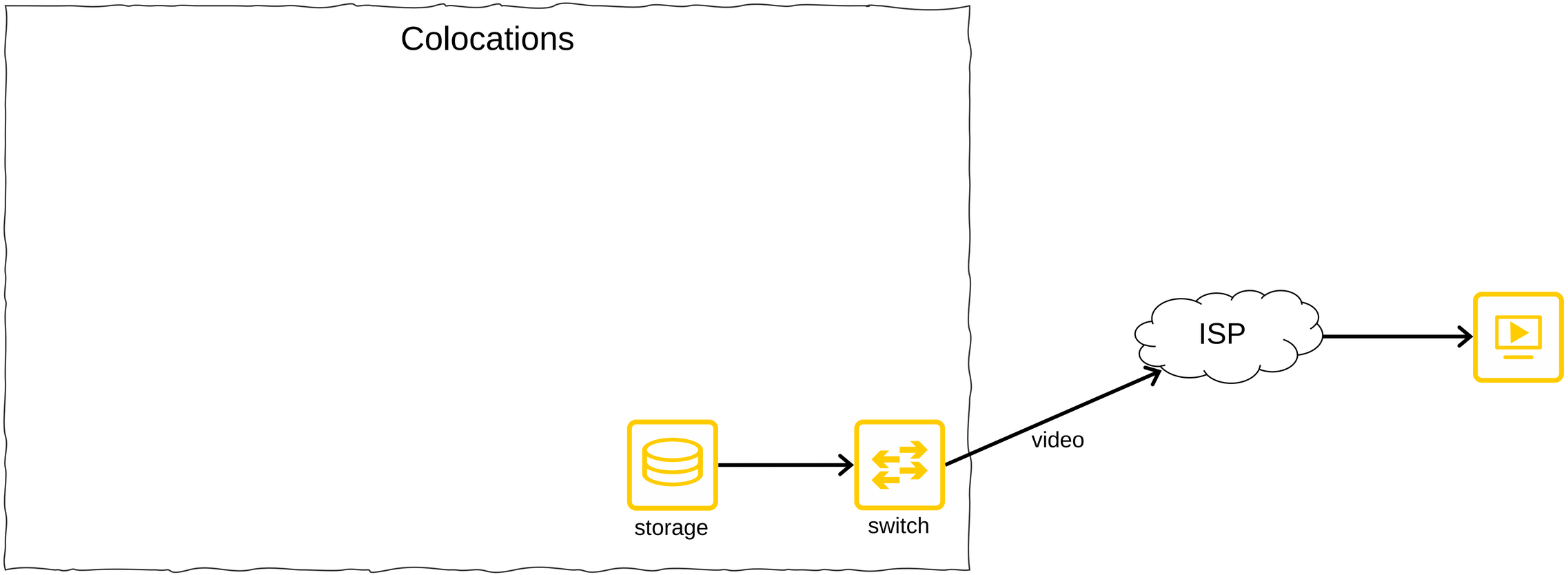
Дрессированный медийный трафик Яндекса

Алексей Щуров, Платформа видеотрансляций

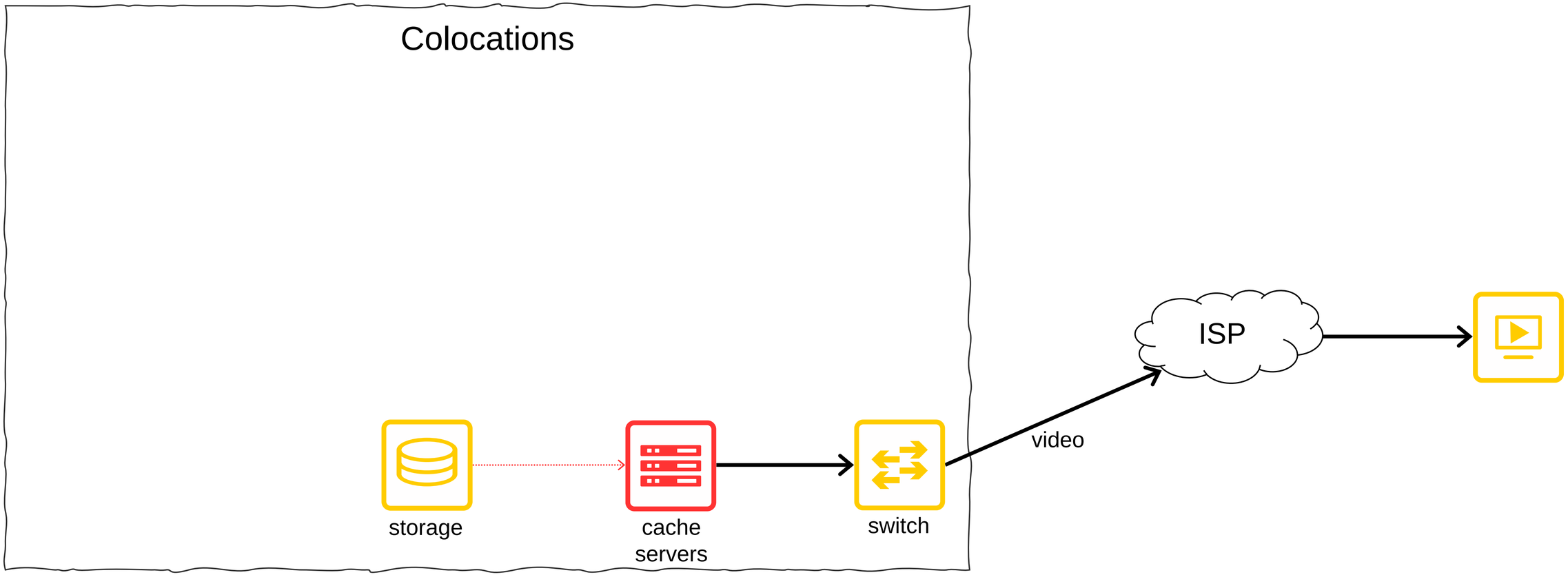
① Что хотим получить?

- Отказоустойчивый и масштабируемый сервис
- Максимальную эффективность утилизации ресурсов

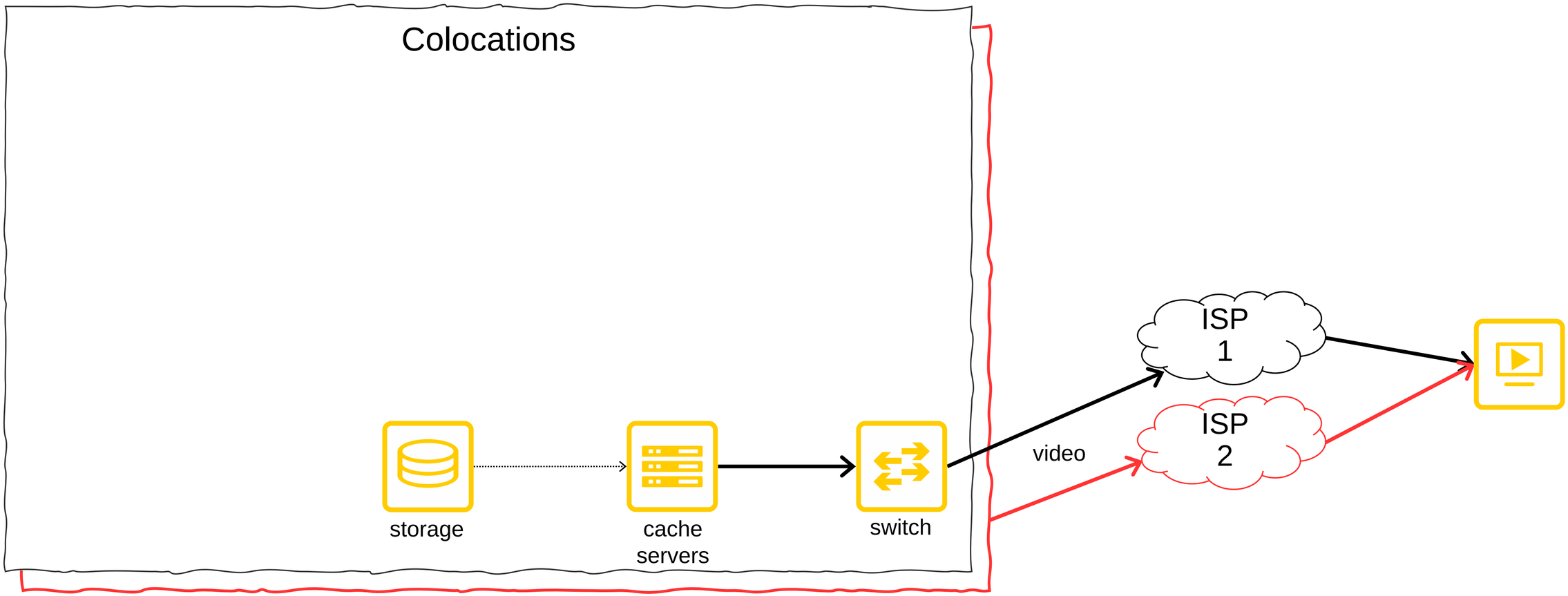
① Минимально возможный сервис



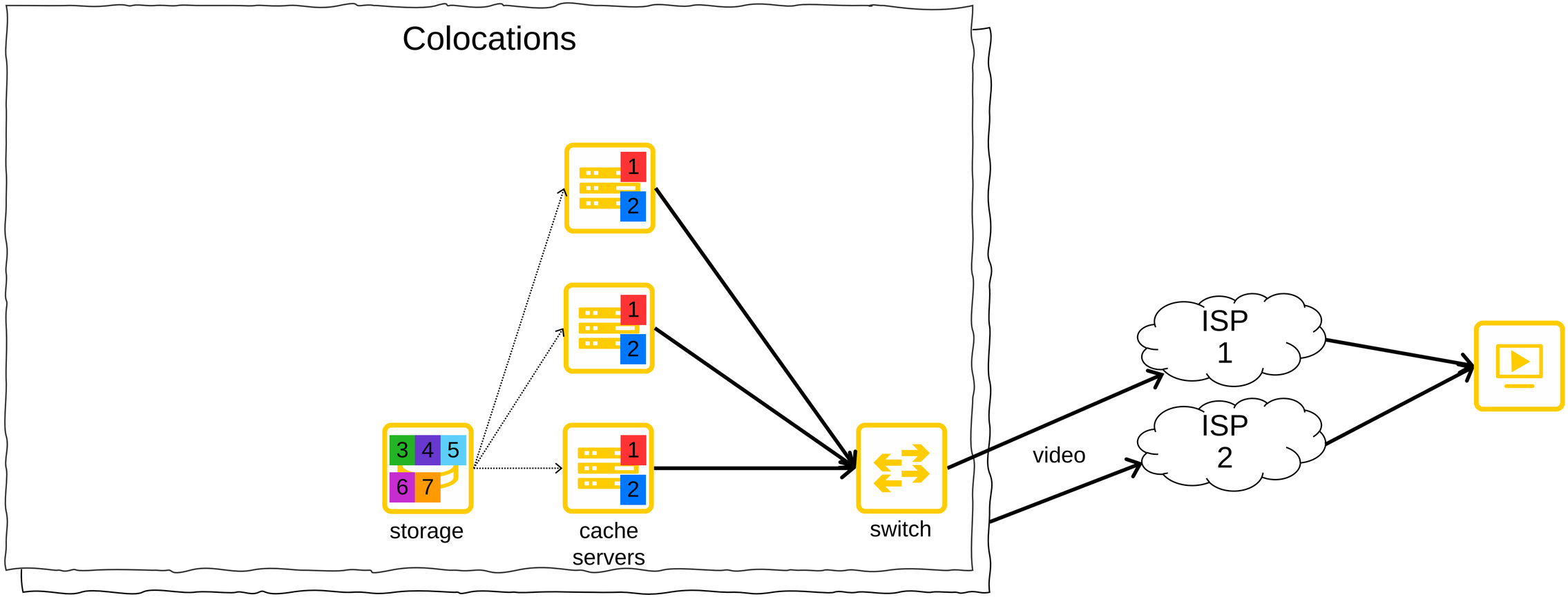
① Больше трафика → Кэш-серверы



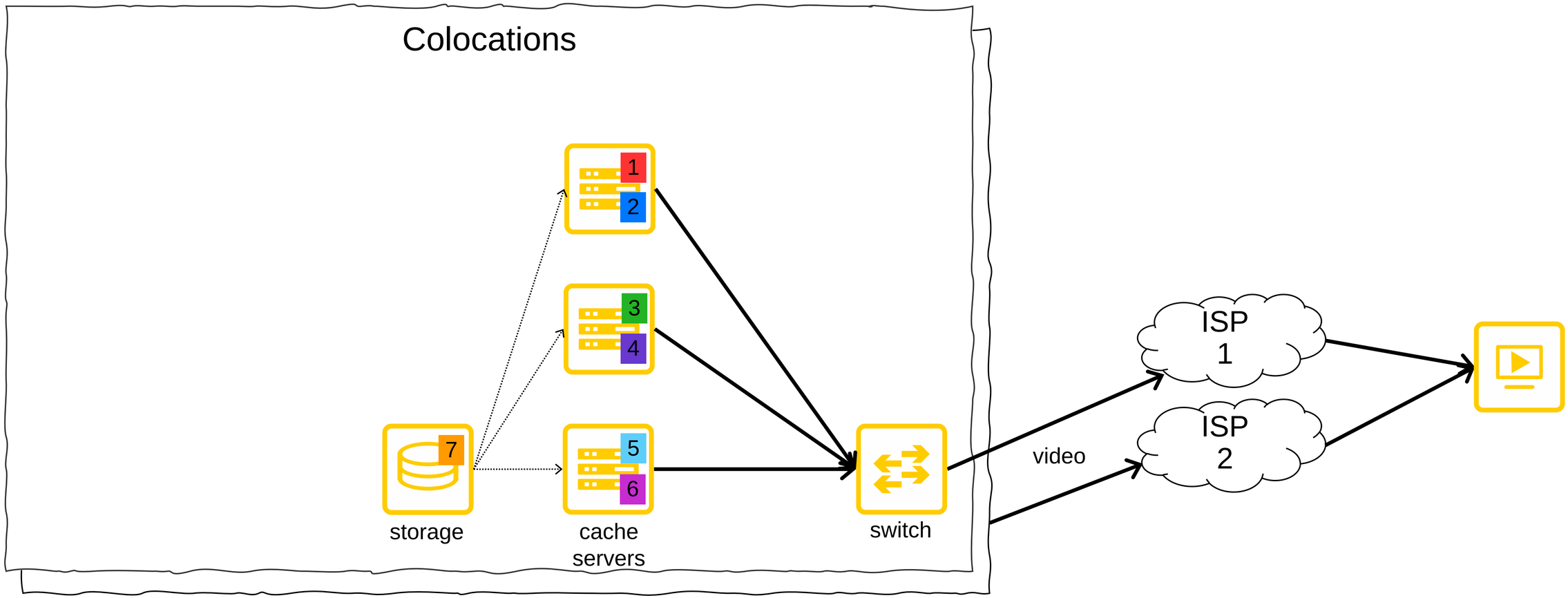
① Резервирование



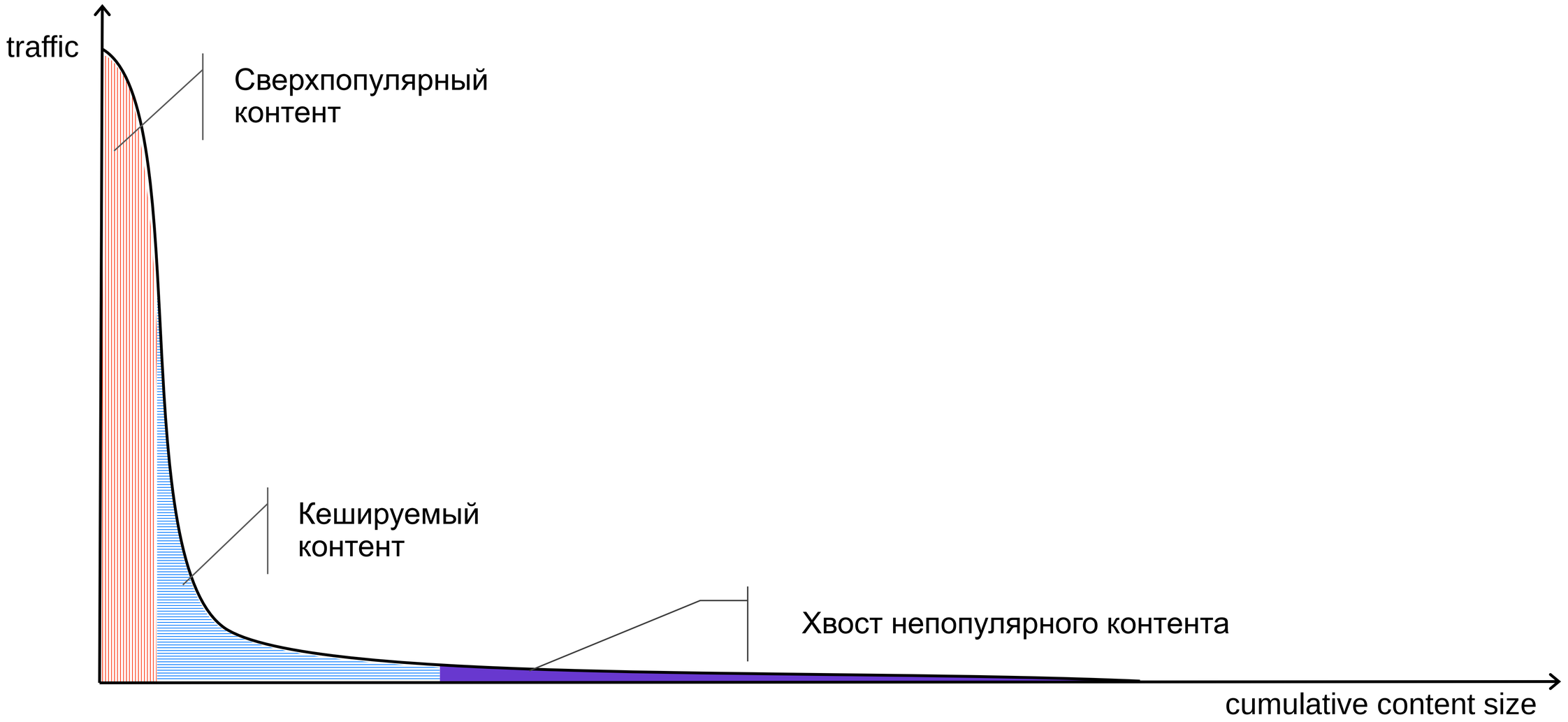
① Много контента



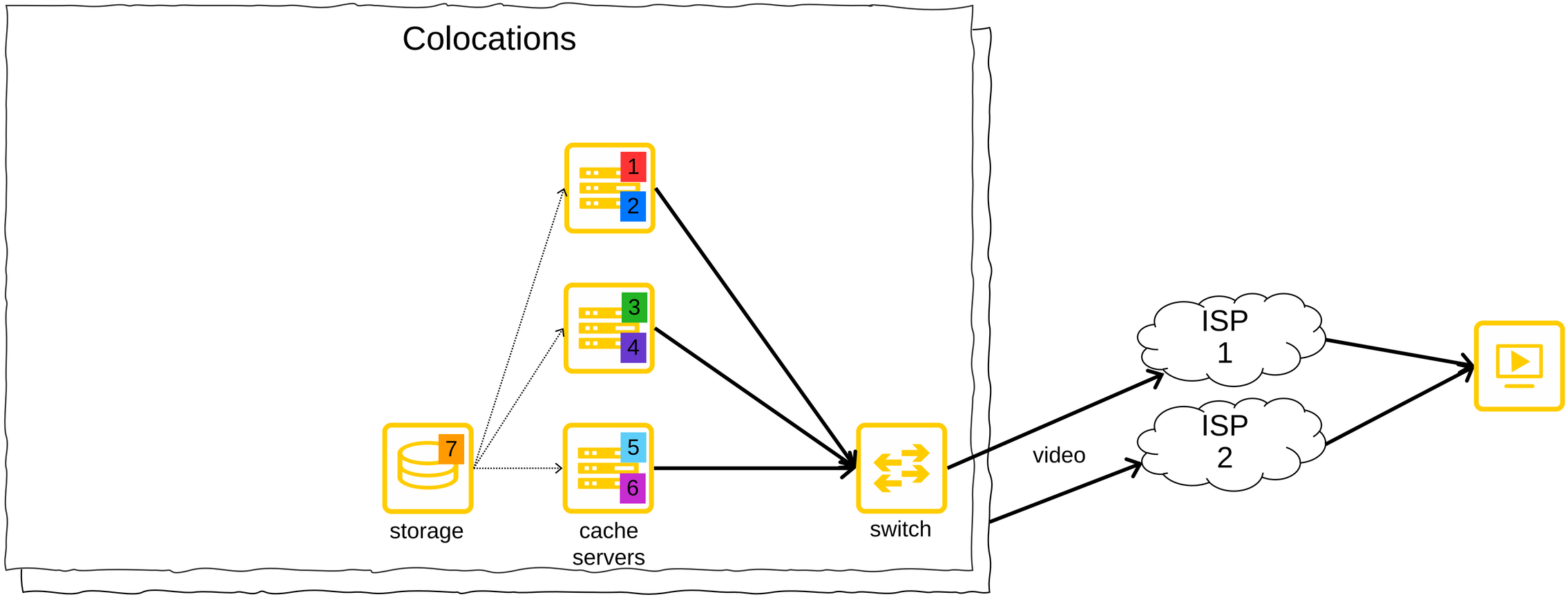
① Шардинг


Colocations
1
2
3
4
5
6
7
storage
cache
servers
switch
video
ISP
1
ISP
2

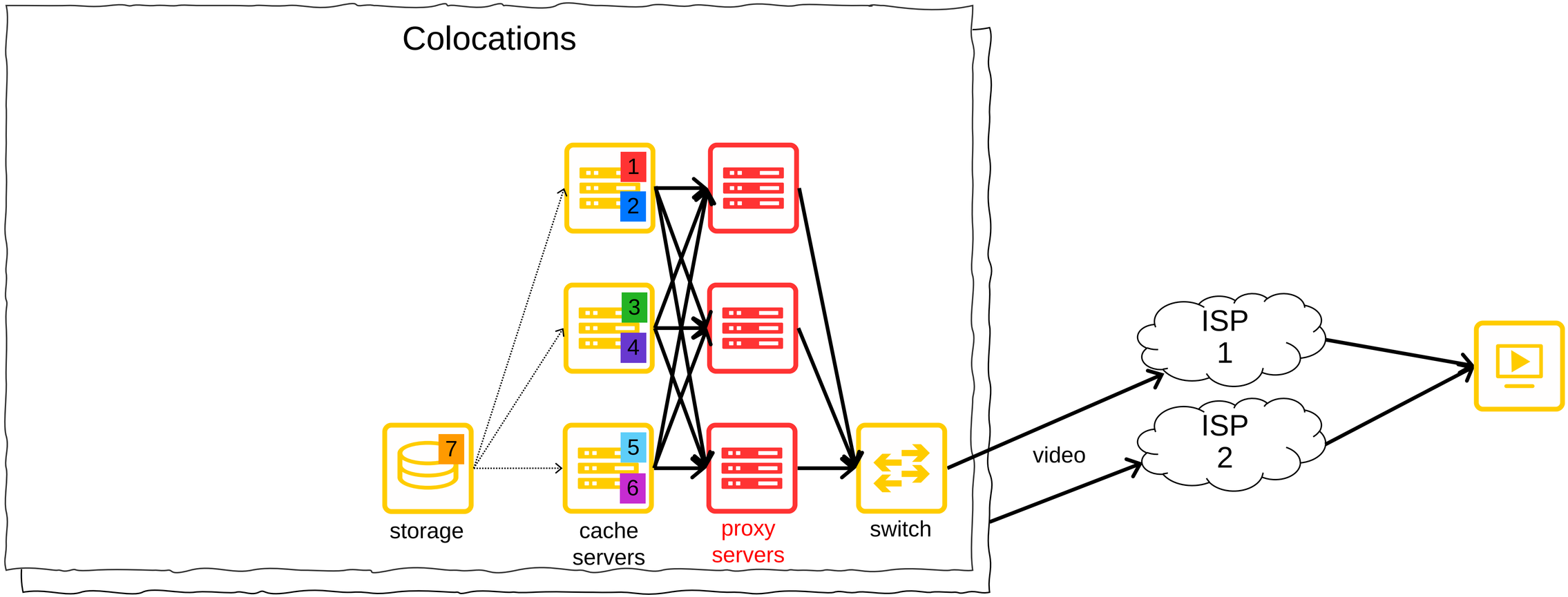
① Популярность контента



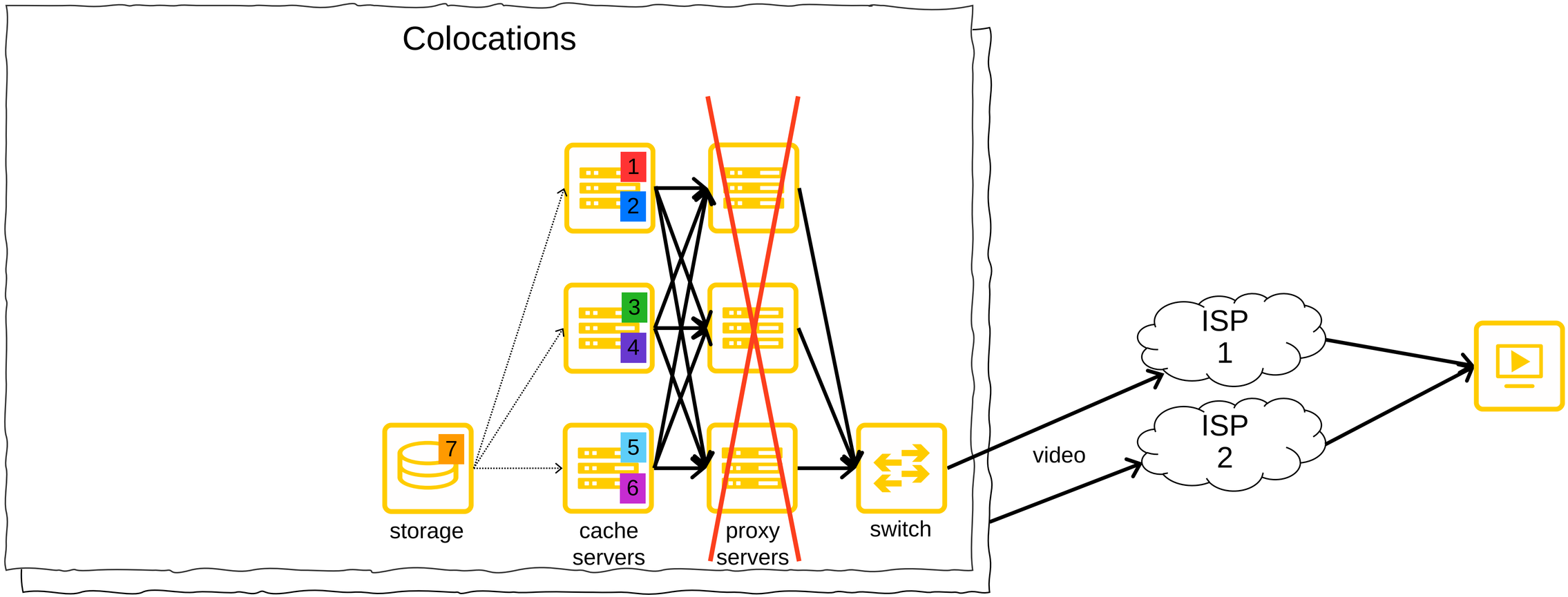
① Шардинг


Colocations
1
2
3
4
5
6
7
storage
cache
servers
switch
video
ISP
1
ISP
2

① Шардинг


Colocations
1
2
3
4
5
6
7
storage
cache servers
proxy servers
switch
video
ISP 1
ISP 2

① Шардинг


Colocations
1
2
3
4
5
6
7
storage
cache servers
proxy servers
switch
video
ISP 1
ISP 2

① Плейлисты

DASH:

<MPD ...>

<BaseURL>https://**isp1-edge2-msk1.video.tld**/content/98765/dash/</BaseURL>

<BaseURL>https://**isp5-edge6-msk2.video.tld**/content/98765/dash/</BaseURL>

...

HLS:

#EXTM3U

#EXT-X-VERSION:3

#RESOLUTION=**1280x720**

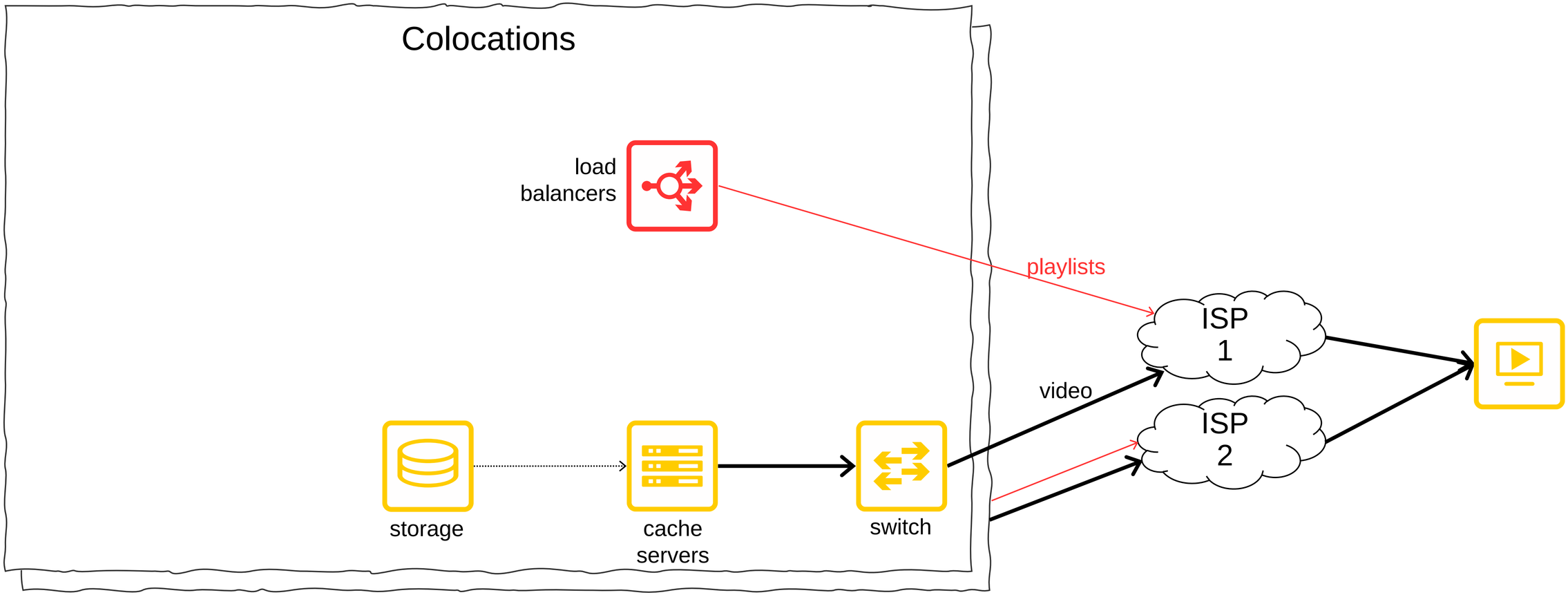
https://**isp1-edge2-msk1.video.tld**/content/98765/hls/index-v11-a4.m3u8

#RESOLUTION=**1280x720**

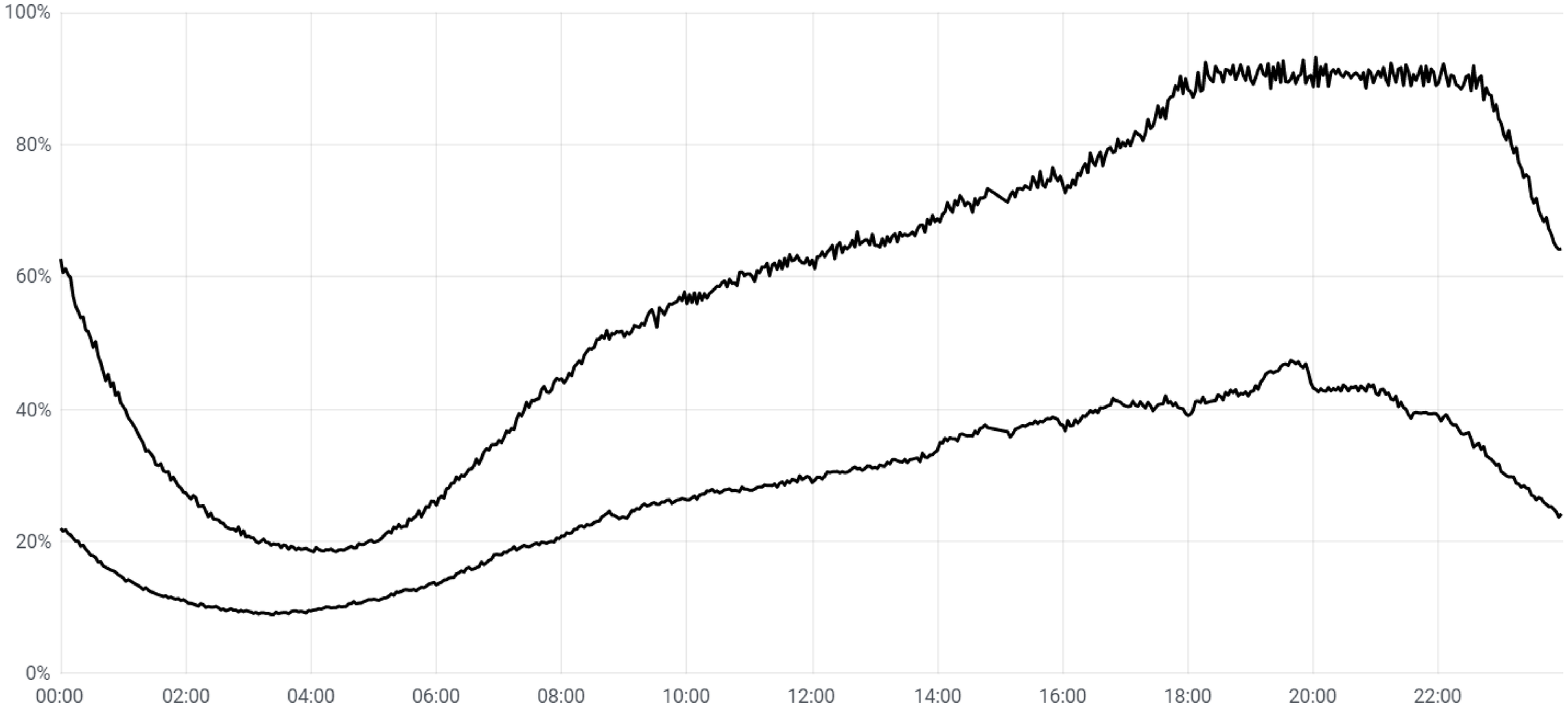
https://**isp5-edge6-msk2.video.tld**/content/98765/hls/index-v11-a4.m3u8

…

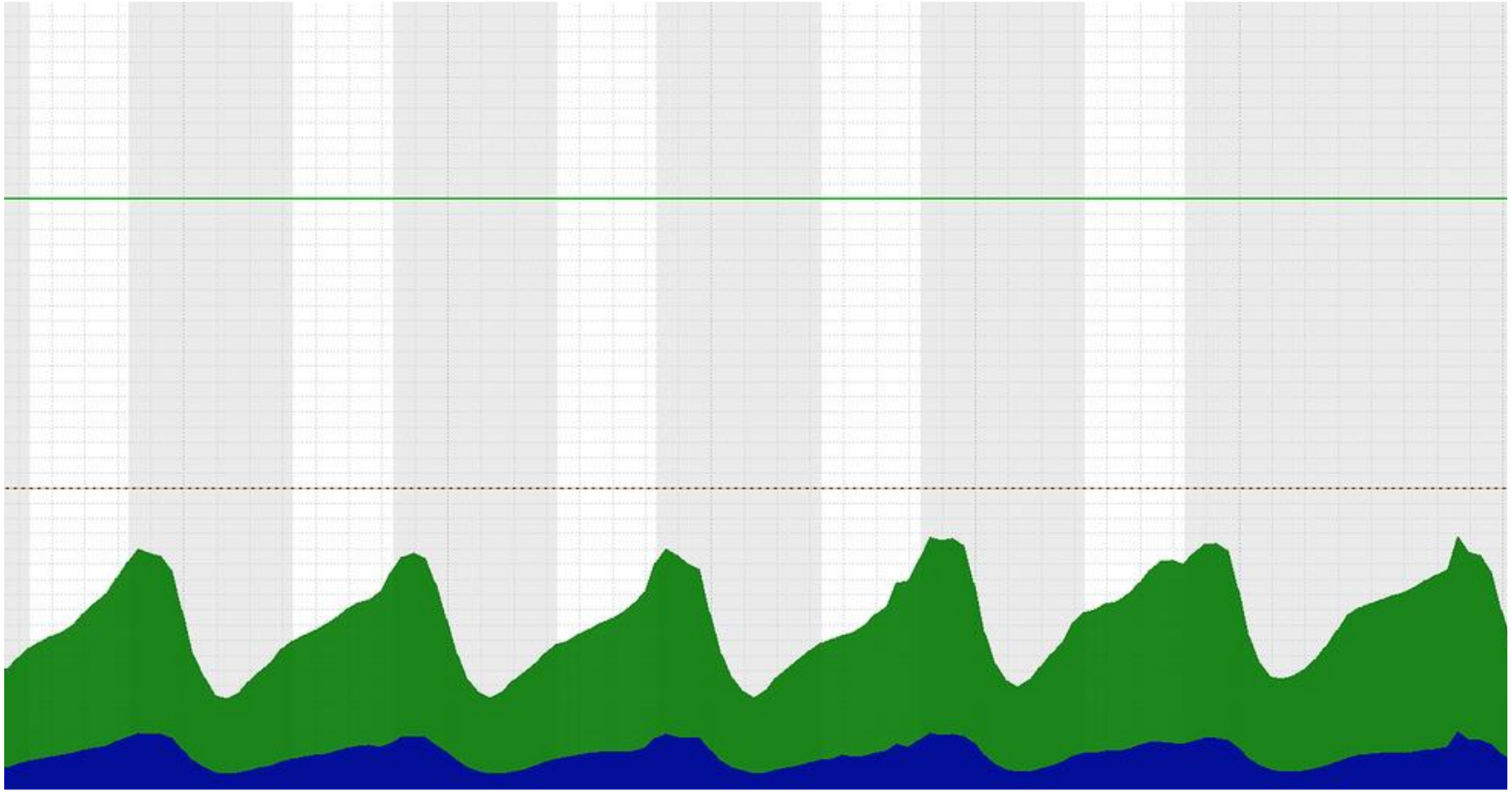
① Балансировщик



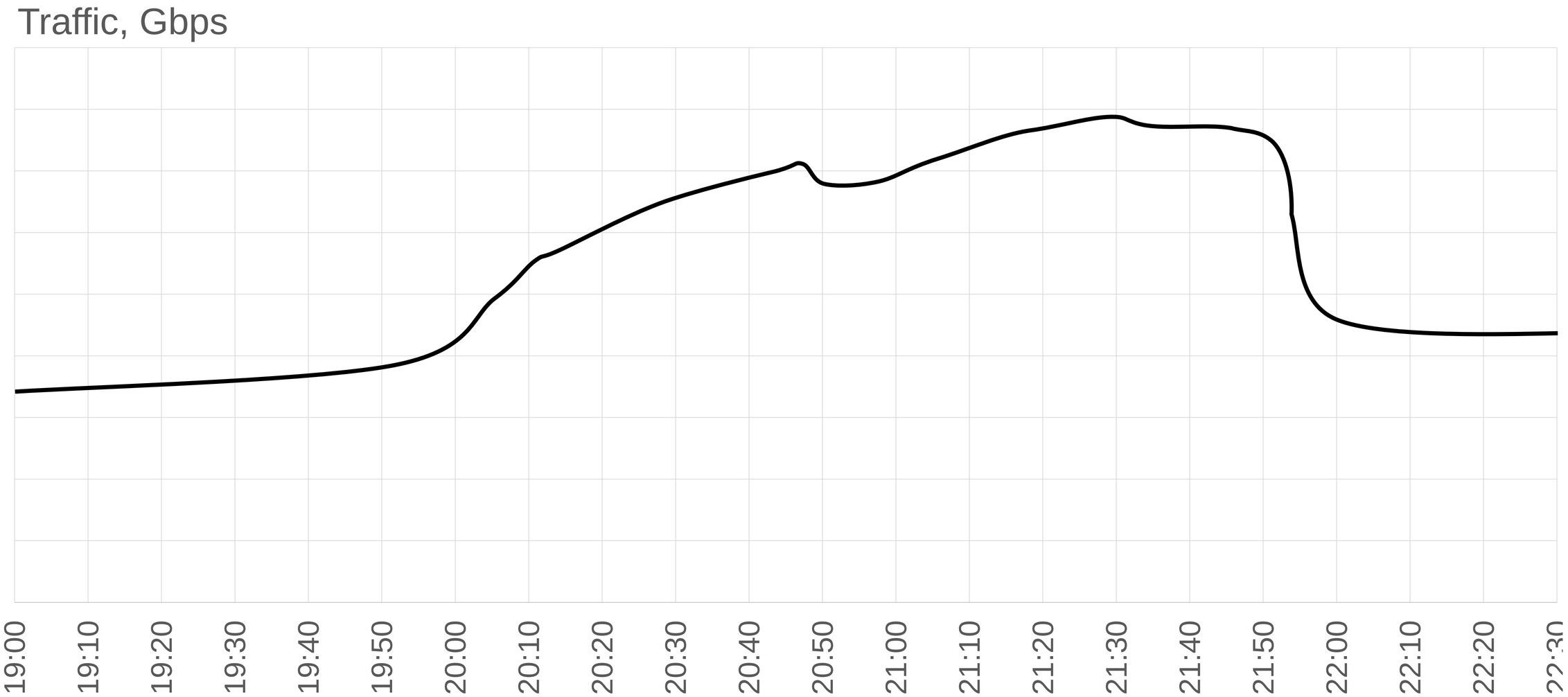
① Неравномерность загрузки



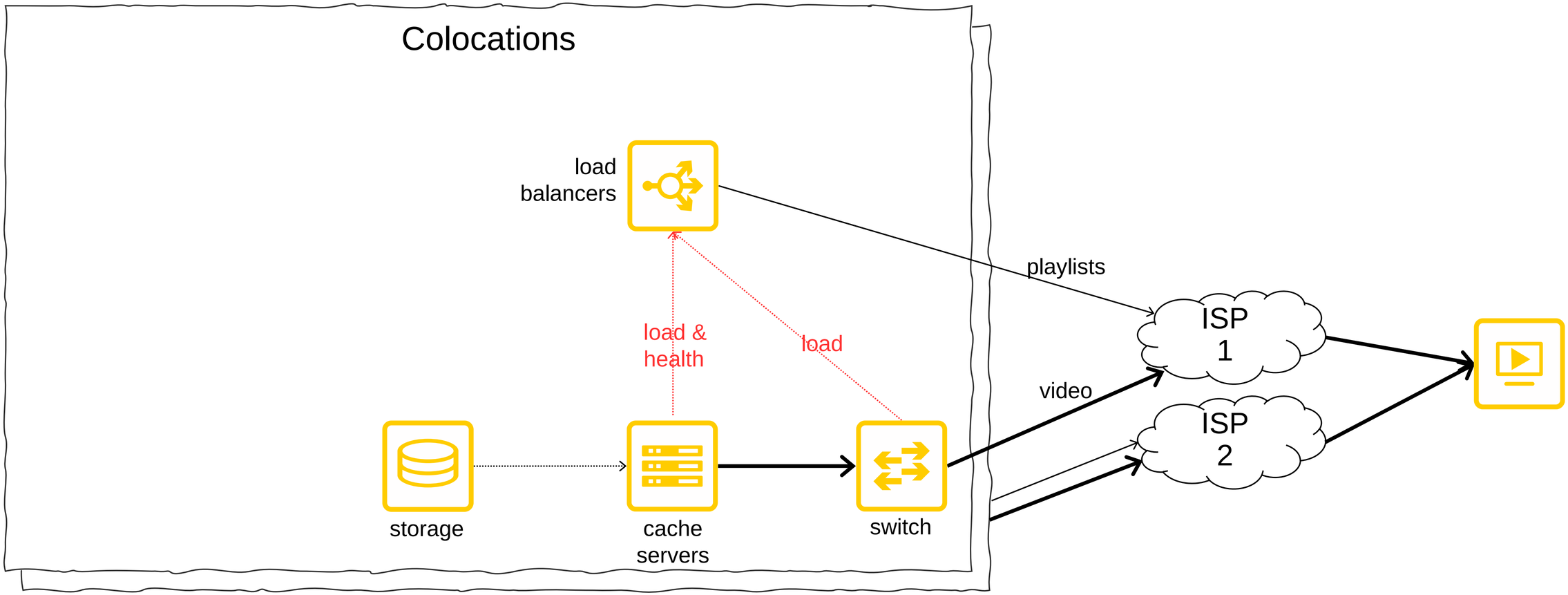
① Регулярный и пиковый трафик



① Что это за событие?



① Сбор метрик



① Что получилось?

Реализация 1

- Отказоустойчивость и масштабируемость
- Эффективность утилизации серверов
- Эффективность утилизации каналов связи

① Что получилось?

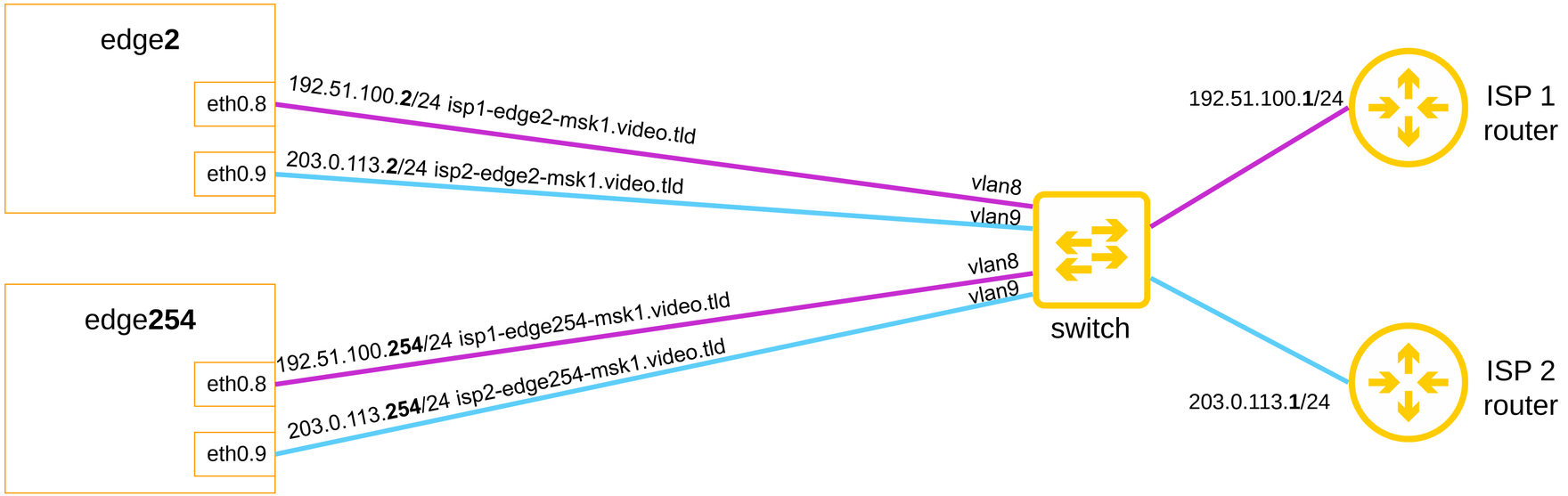
Реализация 1

- Отказоустойчивость и масштабируемость
- Эффективность утилизации серверов
- Эффективность утилизации каналов связи
- Дорогие транзитные каналы в интернет

② Что хотим улучшить?

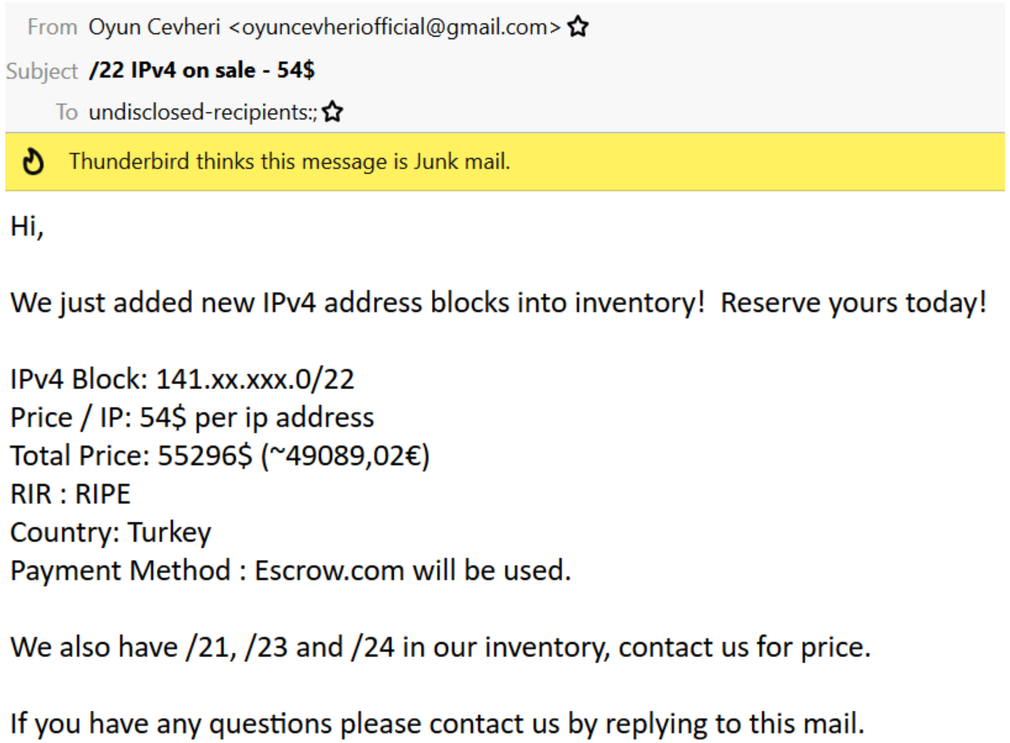
- Снизить стоимость доставки
- Повысить скорость загрузки контента

② Source routing



② А если подключить 100 операторов?

- Количество IPv4 адресов =
  кол-во операторов * кол-во серверов
- Отсутствие возможности подключить IX



From Oyun Cevheri <oyuncevheriofficial@gmail.com>
Subject **/22 IPv4 on sale - 54$**
To undisclosed-recipients:;

Thunderbird thinks this message is Junk mail.

Hi,

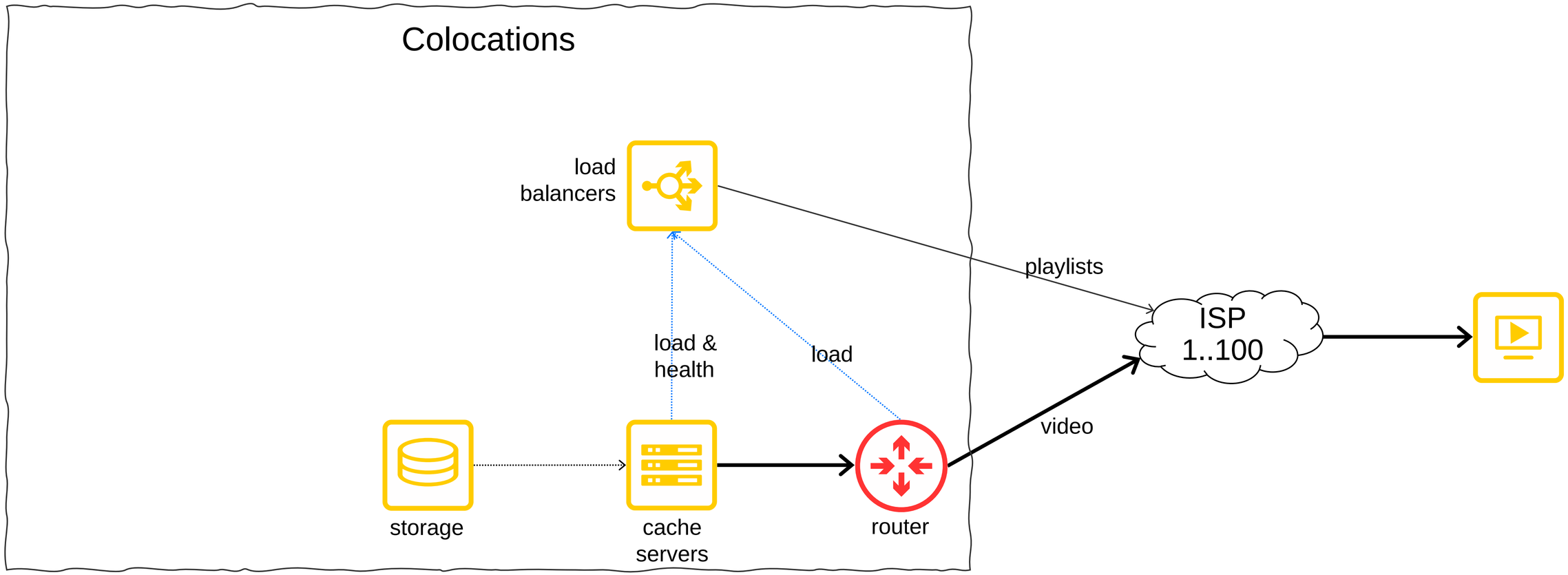
We just added new IPv4 address blocks into inventory! Reserve yours today!

IPv4 Block: 141.xx.xxx.0/22
Price / IP: 54$ per ip address
Total Price: 55296$ (~49089,02€)
RIR : RIPE
Country: Turkey
Payment Method : Escrow.com will be used.

We also have /21, /23 and /24 in our inventory, contact us for price.

If you have any questions please contact us by replying to this mail.

② Маршрутизатор



② Traffic Engineering в BGP

Управление внешним трафиком:

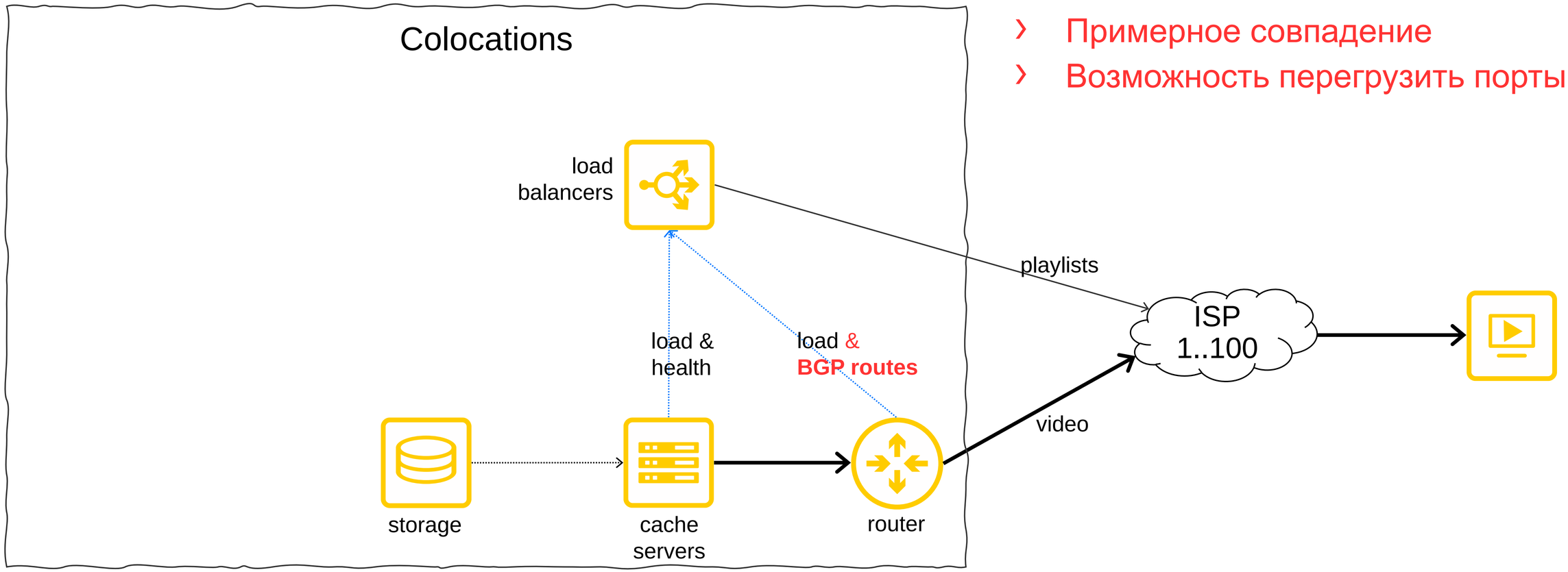
- Local preference
- Multi exit discriminator (MED)
- Weight / Preference / Administrative Distance
- Multipath

② Traffic Engineering в BGP

Управление внешним трафиком:

- Local preference
- Multi exit discriminator (MED)
- Weight / Preference / AD
- Multipath

② Предугадывание поведения маршрутизатора



- Примерное совпадение
- Возможность перегрузить порты

② Метрики выбора маршрутов

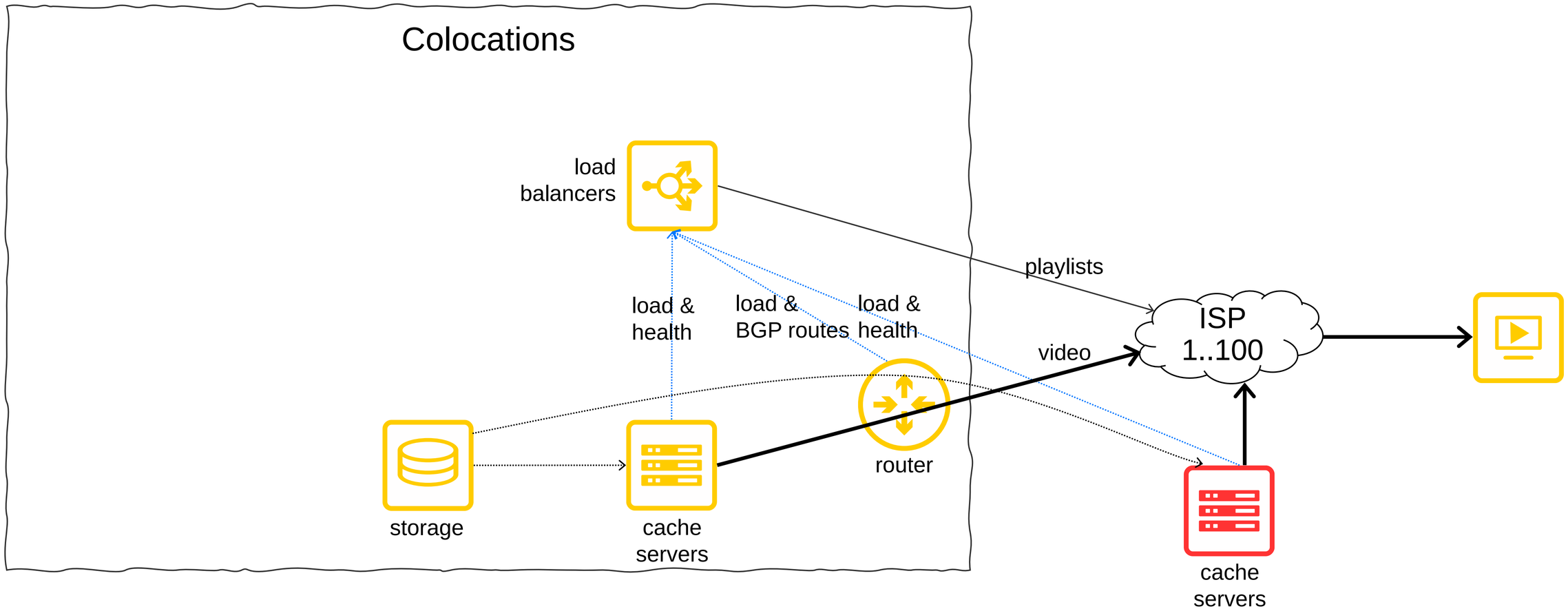
Стандартное поведение bgp:

- as-path, включая prepend
- localpref
- origin code
- med
- community

На самом деле необходимо учитывать:

- Загруженность сетевых интерфейсов
- Скорость загрузки контента, включая перегрузки на пути к конечному клиенту
- Эксплуатационные расходы, включая ожидания операторов связи

② Кэши у операторов



② Что получилось?

Реализация 1

- Отказоустойчивость и масштабируемость
- Эффективность утилизации серверов
- ~~Эффективность утилизации каналов связи~~
- ~~Дорогие транзитные каналы в интернет~~

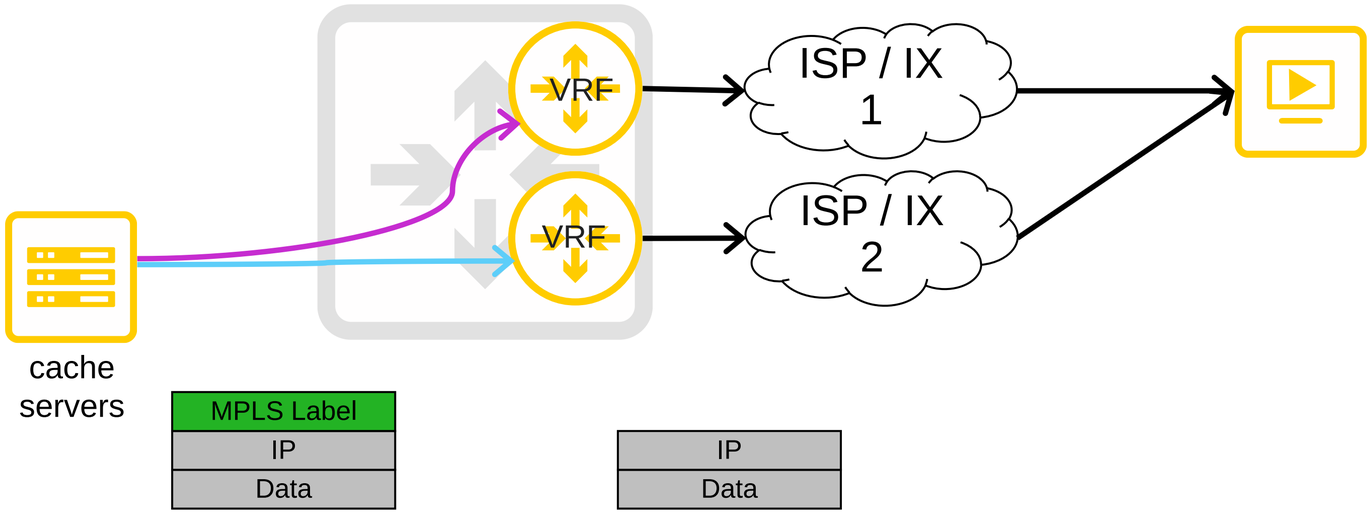
Реализация 2

- Отказоустойчивость и масштабируемость
- Эффективность утилизации серверов
- **Сниженная эффективность утилизации каналов связи**
- **Выбор маршрутов до пользователей**

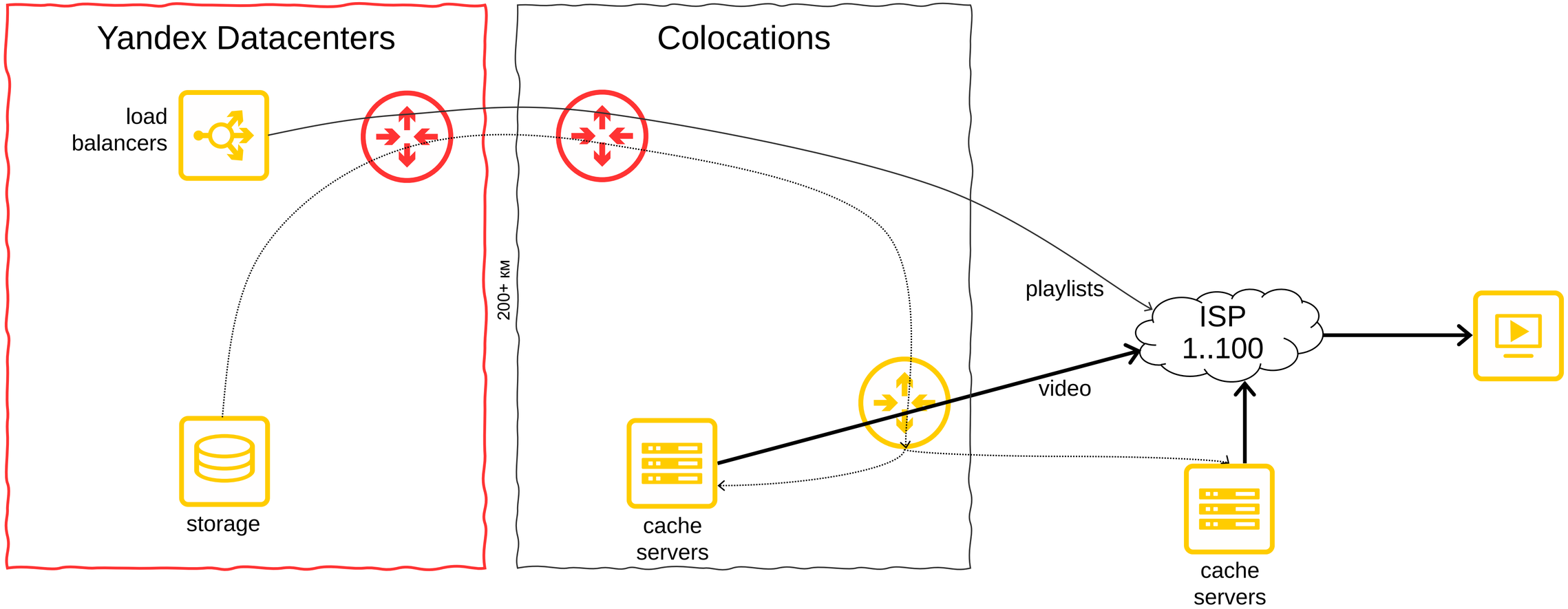
③ Что хотим улучшить?

- Вернуть управляемость трафика
- Оптимизировать доставку непопулярного контента

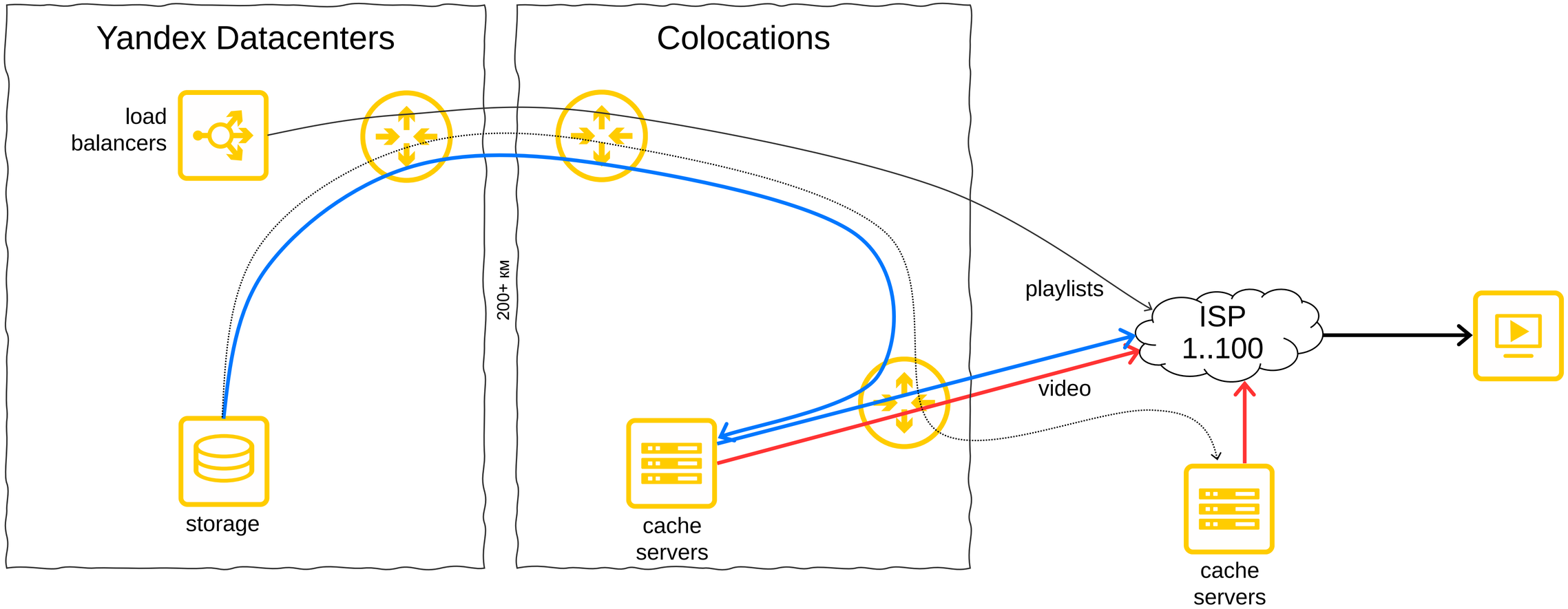
③ Контролируемый выбор стыков (MPLS)



③ Далёкие Датацентры



③ Проксирование непопулярного контента



③ Что мы рассматривали

- inner+outer MPLS Label

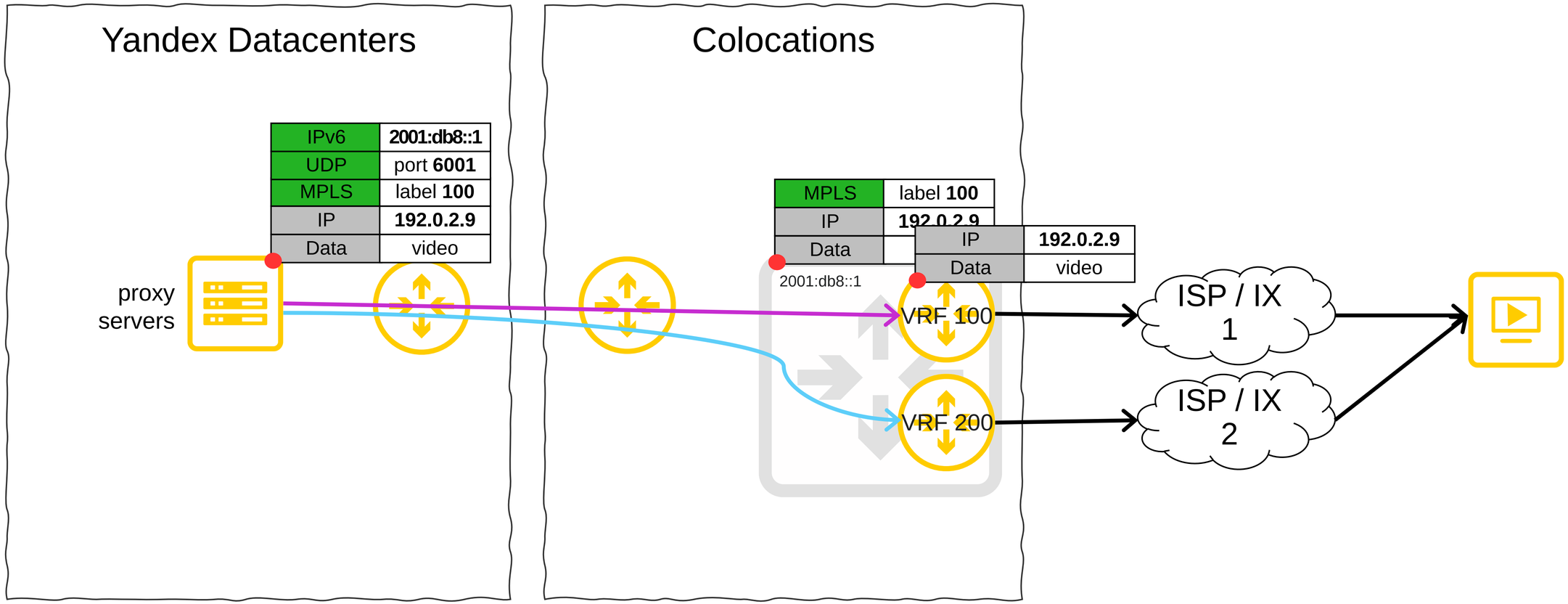
③ Что мы рассматривали

- ~~inner+outer MPLS Label~~
- Google Espresso: MPLS over GRE

③ Что мы рассматривали

- ~~inner+outer MPLS Label~~
- ~~Google Espresso: MPLS over GRE~~
- MPLS over UDP

③ Ещё больше абстракций (MPLS+UDP)



③ Маршрутизация на хостах

- Синхронизация глобального контекста
- Миллионы маршрутов в каждом сервере
- Двойная маршрутизация

③ Глобальная маршрутизация

DASH:

<MPD ...>

<BaseURL>https://isp1-edge2-msk1.video.tld/content/98765/dash/**link_id=1234**/</BaseURL>

<BaseURL>https://isp5-edge6-msk2.video.tld/content/98765/dash/**link_id=9000**/</BaseURL>

...

HLS:

#EXTM3U

#EXT-X-VERSION:3

#RESOLUTION=**1280x720**

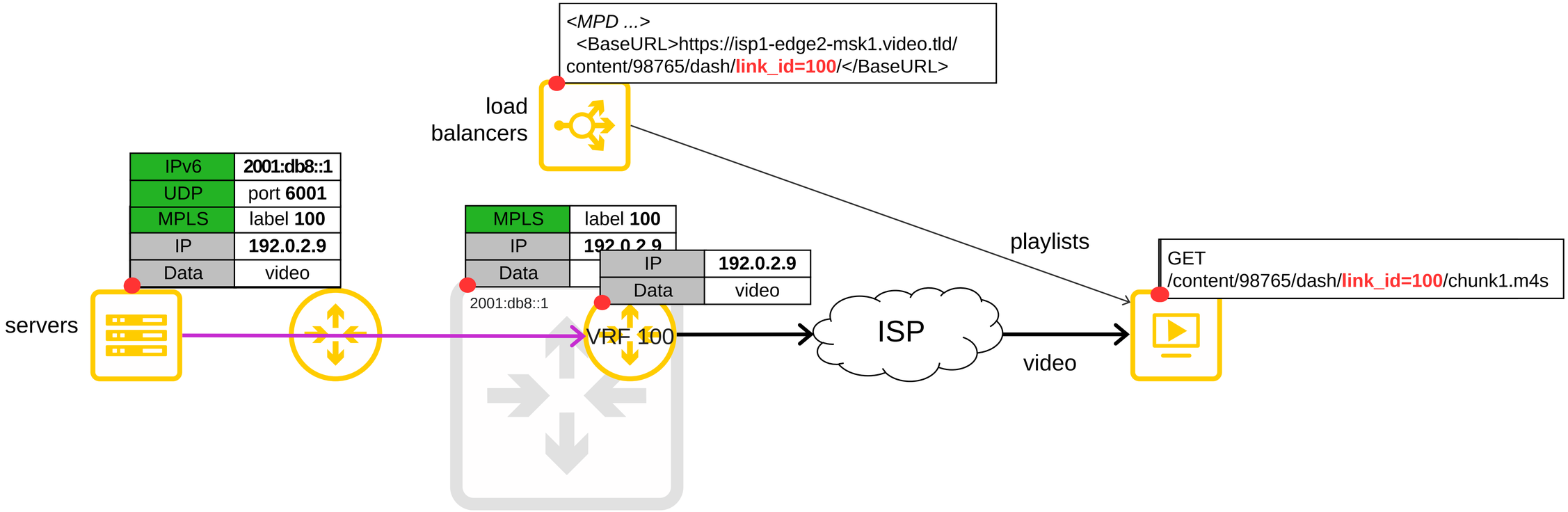
https://isp1-edge2-msk1.video.tld/content/98765/hls/**link_id=1234**/index-v11-a4.m3u8

#RESOLUTION=**1280x720**

https://isp5-edge6-msk2.video.tld/content/98765/hls/**link_id=9000**/index-v11-a4.m3u8

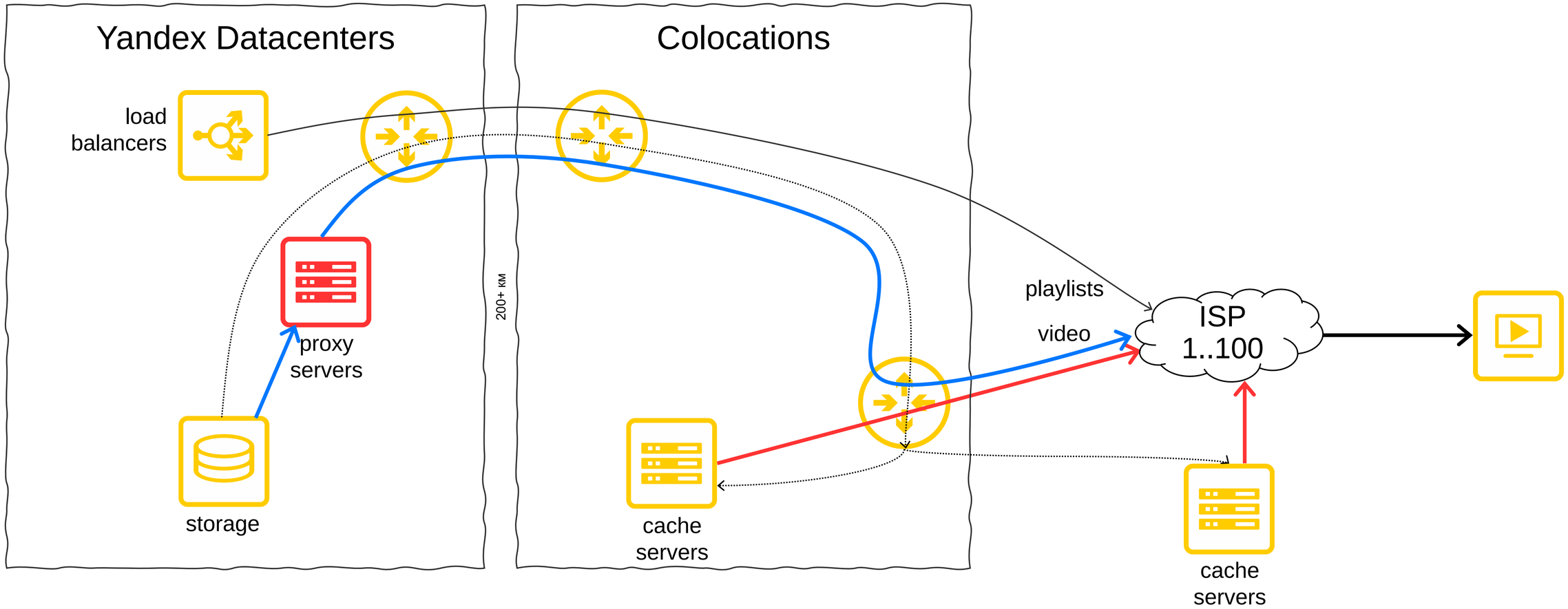
…

③ Круговорот пакетов в природе




```
# ip rule add from all fwmark 0x64 lookup 999100
# ip -4 route add table 999100 default encap mpls 999100 dev strm-b0635 scope link mtu 1450 advmss 1410
# ip link add name strm-b0635 numtxqueues 8 numrxqueues 8 type ip6tnl remote 2001:db8::1 \
> local ${LOCAL_ADDR} encap fou encap-sport auto encap-dport 6001 encaplimit none
```

③ Проксирование непопулярного контента



③ Что получилось?

Реализация 2

- Отказоустойчивость и масштабируемость
- Эффективность утилизации серверов
- ~~Сниженная эффективность утилизации каналов связи~~
- Выбор маршрутов до пользователей

Реализация 3

- Отказоустойчивость и масштабируемость
- Эффективность утилизации серверов
- **Эффективность утилизации каналов связи**
- Выбор маршрутов до пользователей
- **Экономия места в Colocations**
- **Ещё большая масштабируемость сети (L3 CLOS)**

Выводы

Traffic Engineering про:

1. Трафик и
2. Ресурсы

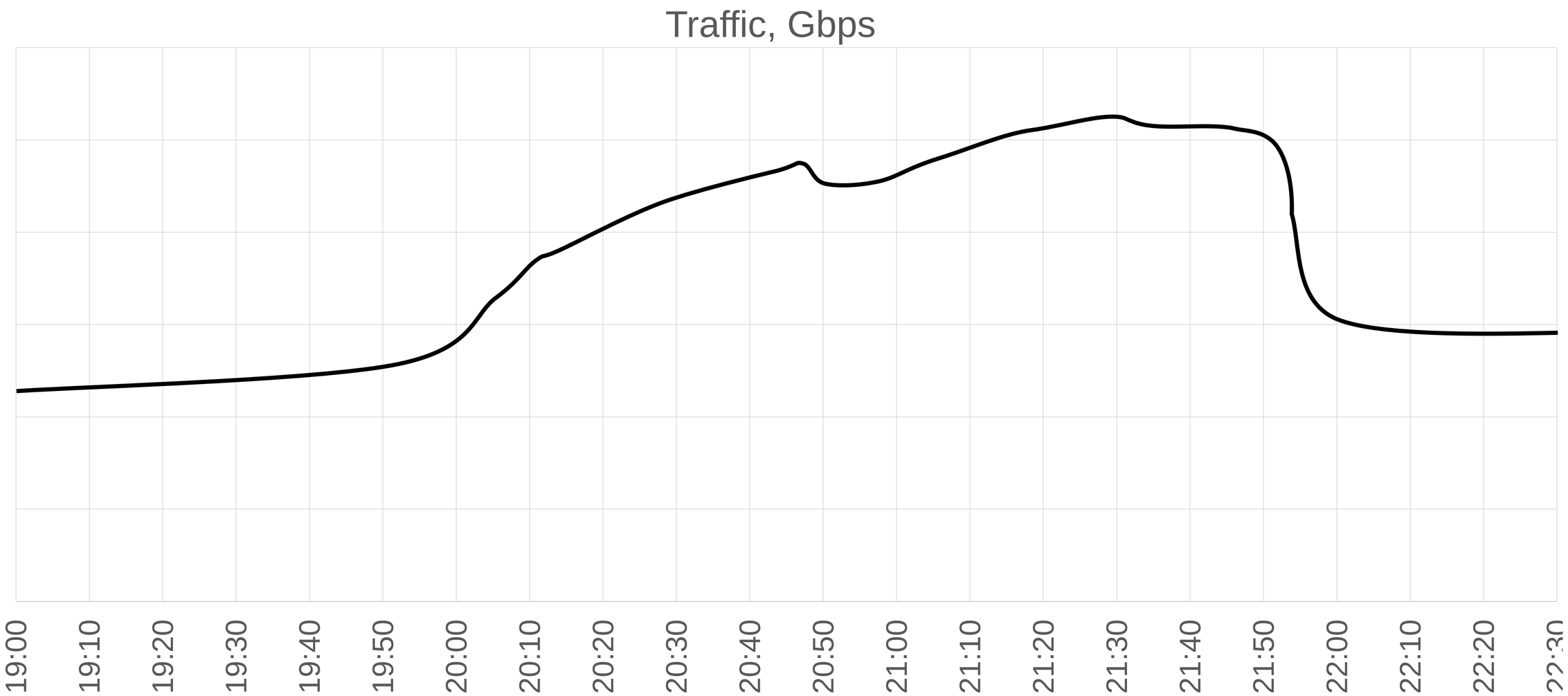
- Может быть разный, все три реализации рабочие.
- Инкапсуляция поможет пробраться через любой лес сетевого оборудования.

В следующей серии: SRv6

Пример конфигурации для сервера:
ip rule add from all fwmark **0x64** lookup **999100**
ip route add table **999100** default encap seg6 mode encap segs **2a02:db8:1::4:100** dev eth0

Пример конфигурации для Juniper:
set routing-instances **Peer-link-999100** protocols bgp source-packet-routing srv6 locator lab1 end-dt6-sid **2a02:db8:1::6:100**
set routing-instances **Peer-link-999100** protocols bgp source-packet-routing srv6 locator lab1 end-dt4-sid **2a02:db8:1::4:100**
set routing-options source-packet-routing srv6 locator lab1 **2a02:db8:1::/64**

~~Это просто какое-то промо~~ ☺



Яндекс

Алексей Щуров

Платформа видеотрансляций

aschurov@yandex-team.ru

+79160624477

https://www.linkedin.com/in/aschurov/