



Открытый генератор трафика T-RAF и исследование производительности сети

Орлова Маргарита (НИУ «МЭИ»)





Работодатели
и партнеры



Основные направления научных исследований

Высокоточные вычисления,
в том числе, на архитектуре
МП Эльбрус

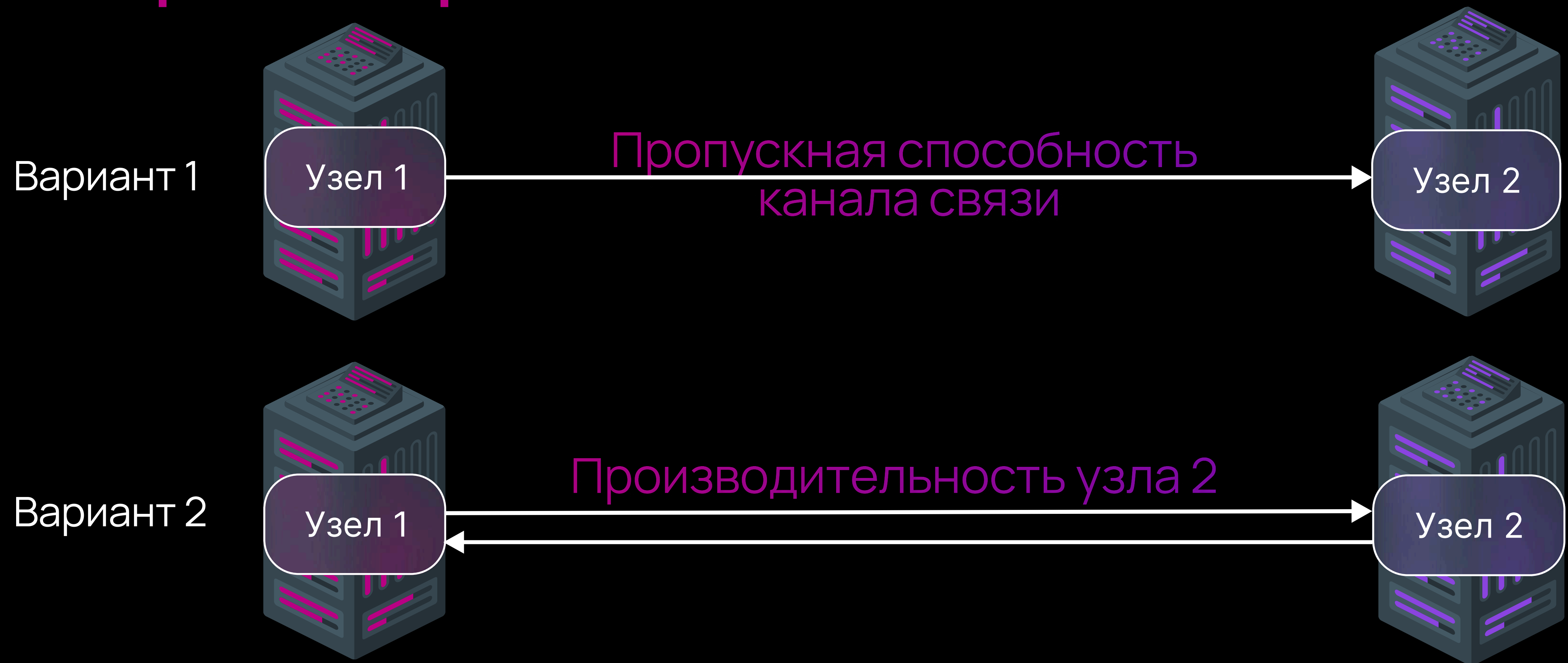
Исследование производительности
вычислительных комплексов, систем
и сетей

Цифровая обработка
многомерных сигналов

Интеллектуальные системы
синтеза СВЧ устройств

Моделирование
распределенных систем

Исследование характеристик сети



Характеристики сети

Пропускная способность [бит/с]

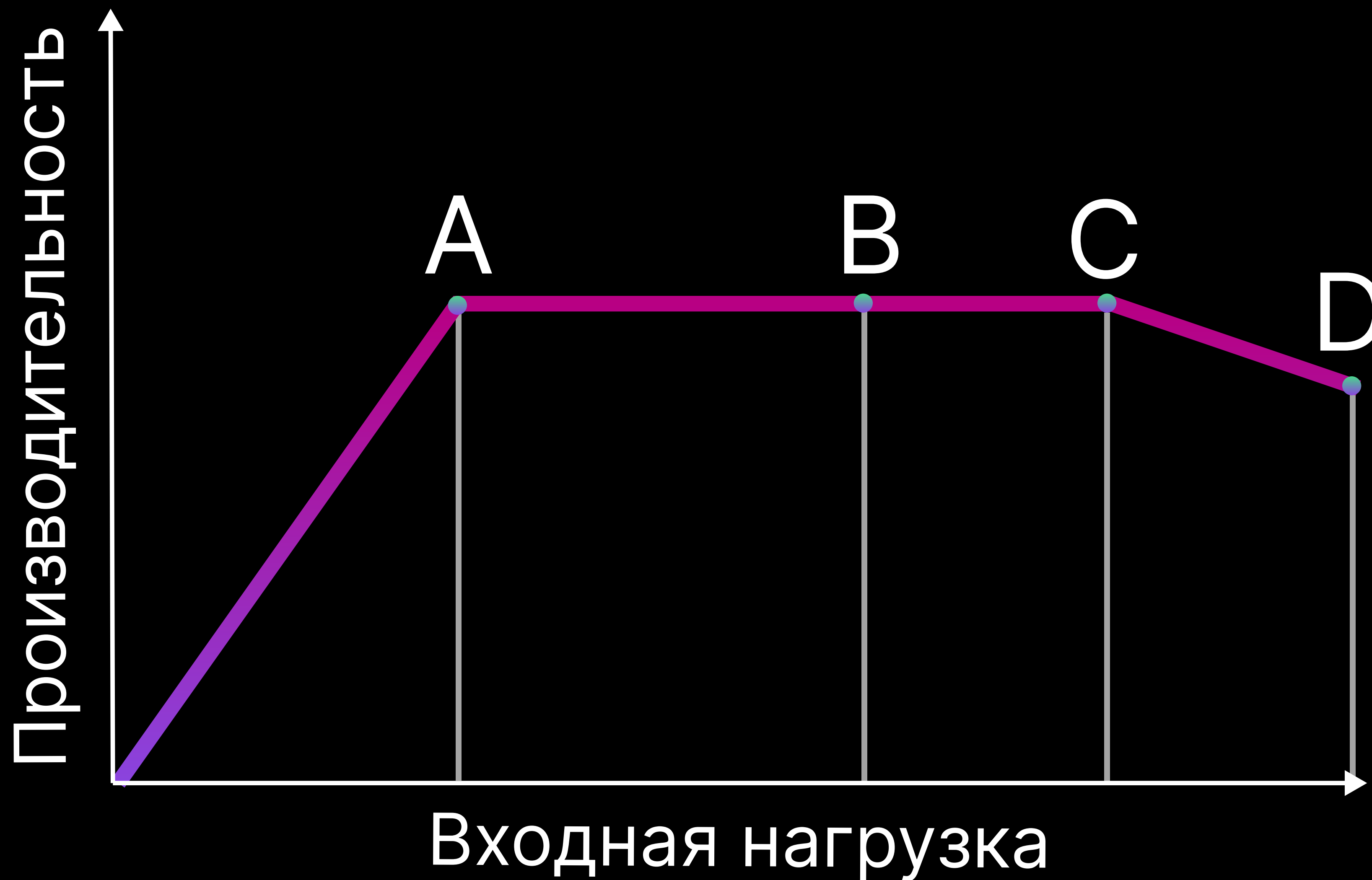
Потери [%]

Сквозная задержка [с]

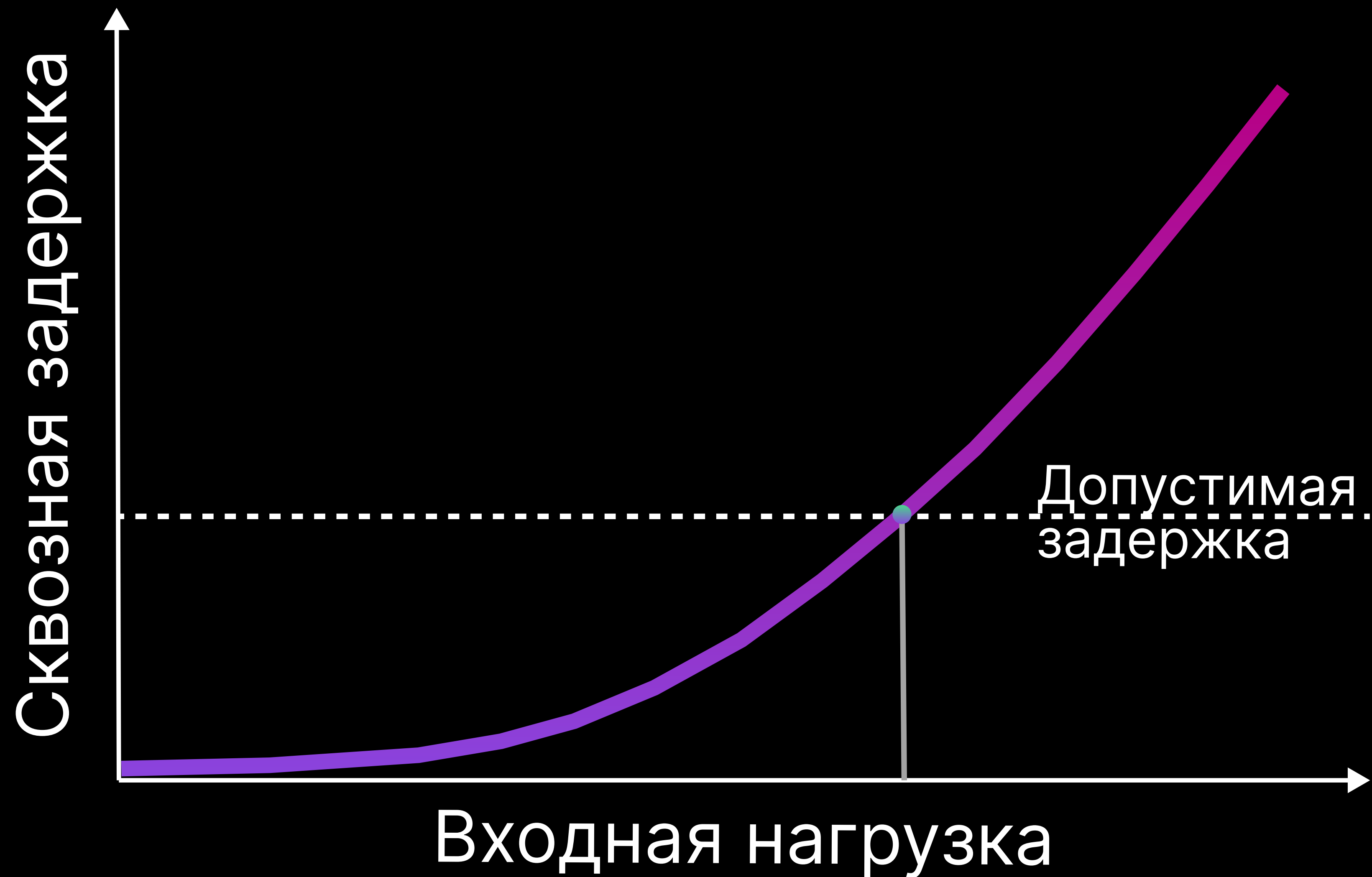
Производительность [пакет/с]

Джиттер (флуктуации задержки) [с]

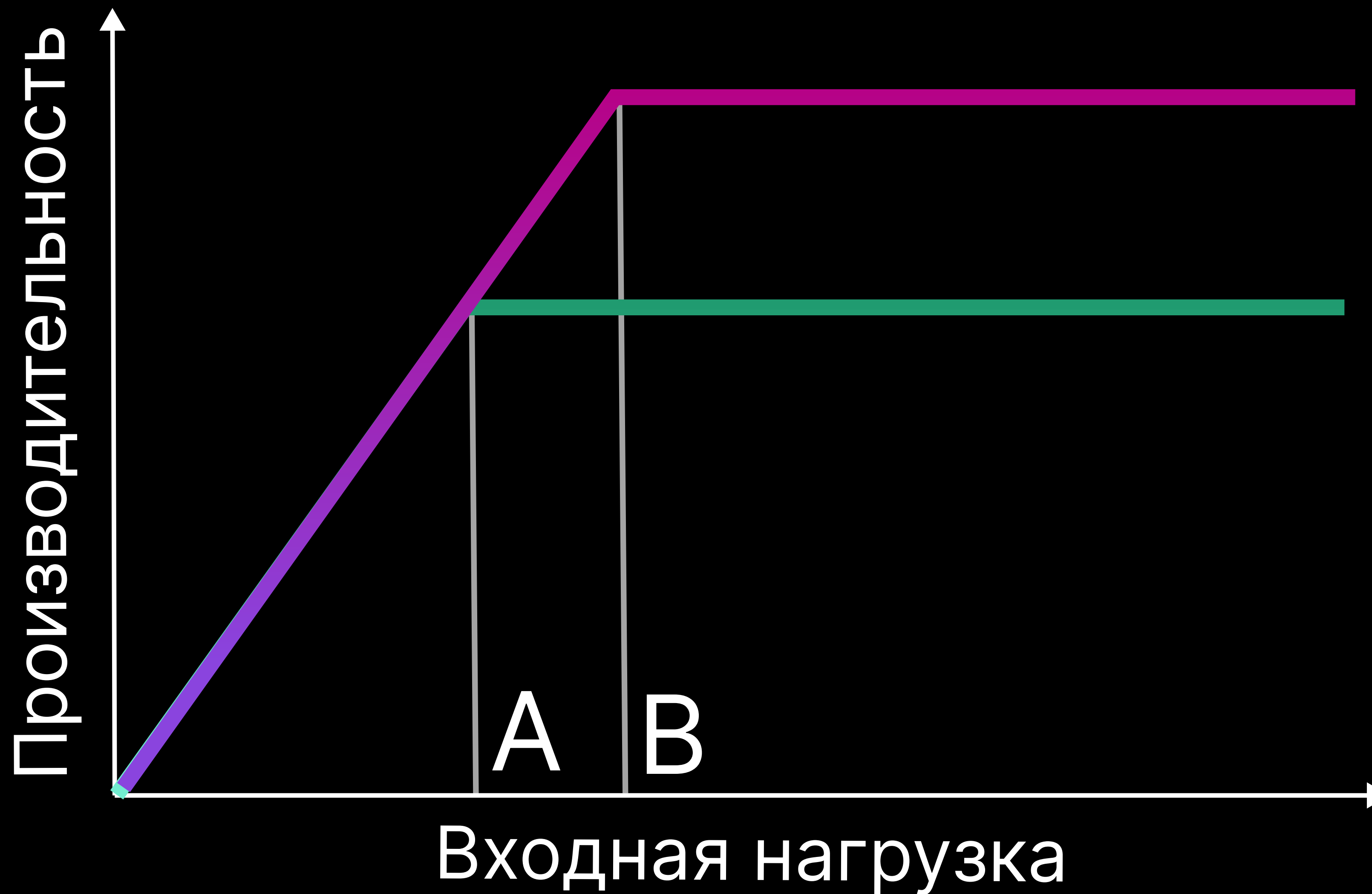
Производительность



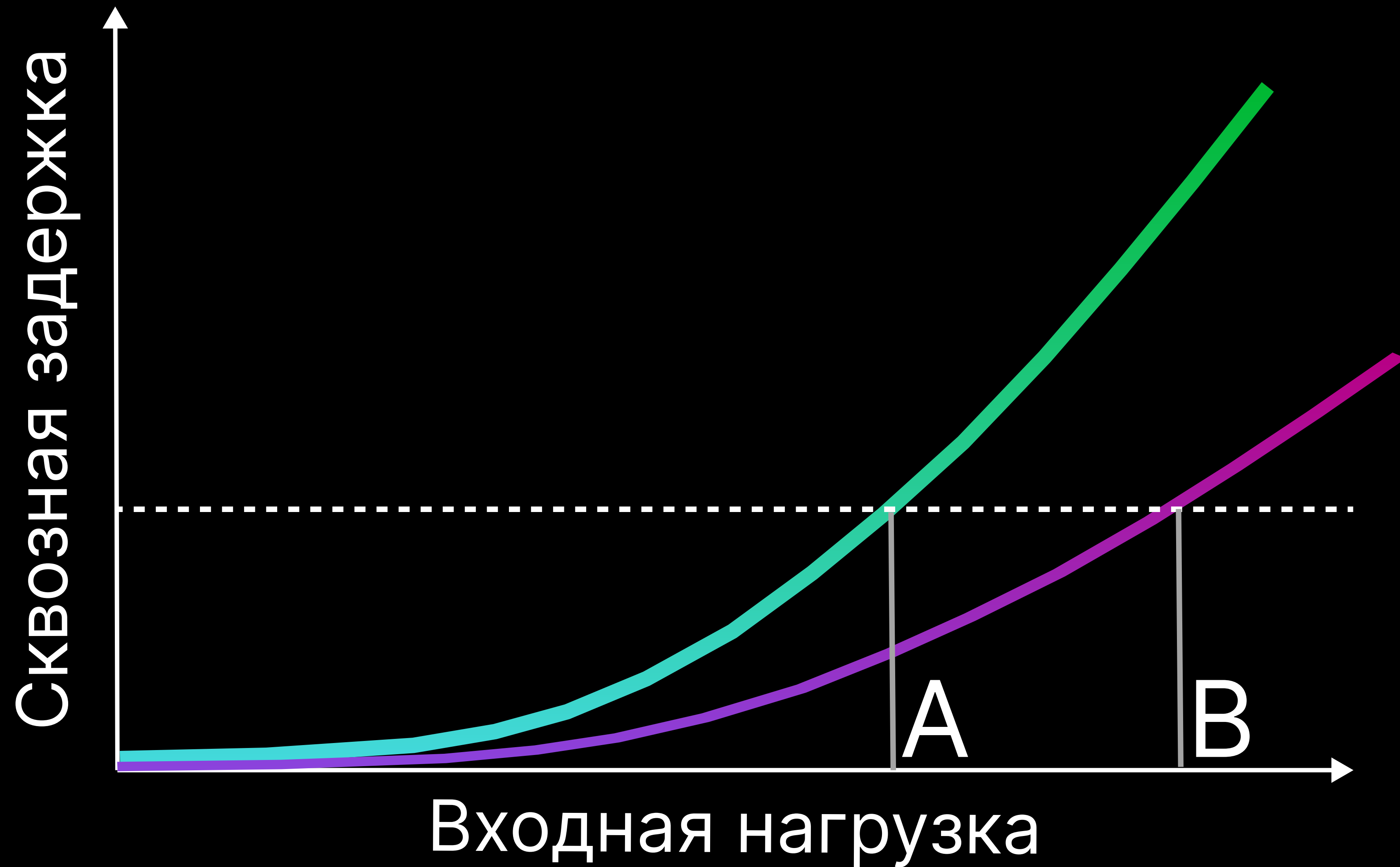
Сквозная задержка



Производительность



Сквозная задержка



Сравнение генераторов трафика

Единица передачи (кадр/пакет/сегмент/запрос к БД)

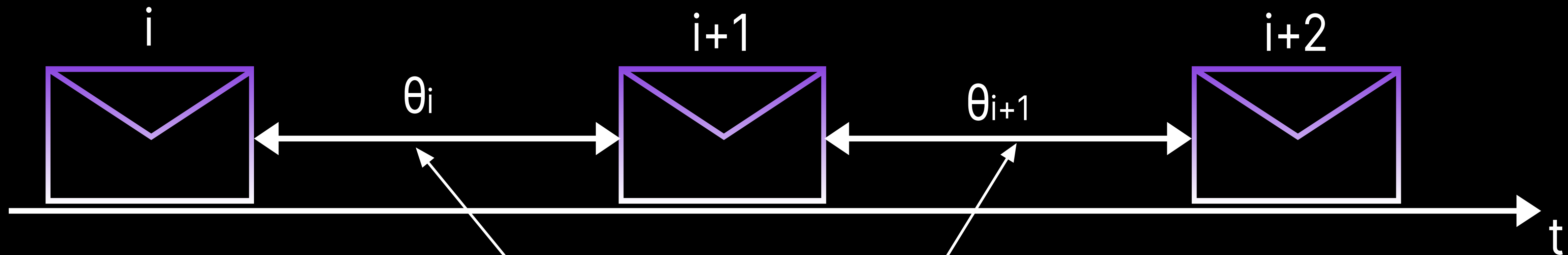
Распределенный эксперимент

Интервал отправки пакета

Сквозная задержка

Производительность

Интервал отправки пакетов



механизм учёта и контроля

Модель трафика

D (Deterministic) — детерминированное распределение,
 $\theta = \text{const}$ (предсказуемое поведение, нет гибкости)

M (Markovian) — интервал отправки пакетов распределен по экспоненциальному закону, вероятность отправки пакета
 $P(T > \theta) = e^{-\lambda\theta}$ (простота анализа и моделирования)

G (General) — общее распределение
(сложность анализа и предсказания поведения системы)

Популярные генераторы трафика

iperf/iperf3

EtherateMT

TRex

pktgen

Сравнение генераторов трафика

	iperf/iperf3	TRex	EtherateMT	pktgen
Единица передачи	TCP/UDP (L4)	L2-L7	кадр (L2)	кадр (L2)
Сквозная задержка	Нет	с помощью PTP	с помощью PTP	Нет
Распределенный эксперимент	вручную	вручную	вручную	вручную
Производительность	средняя	высокая	высокая	высокая
Интервал отправки пакета	G	G, вручную*	D	G

Синхронизация времени

PTP, IEEE 1588

Аппаратная метка времени (точность ~ 10-100 ns)
Программная метка времени (точность ~ 10 us)

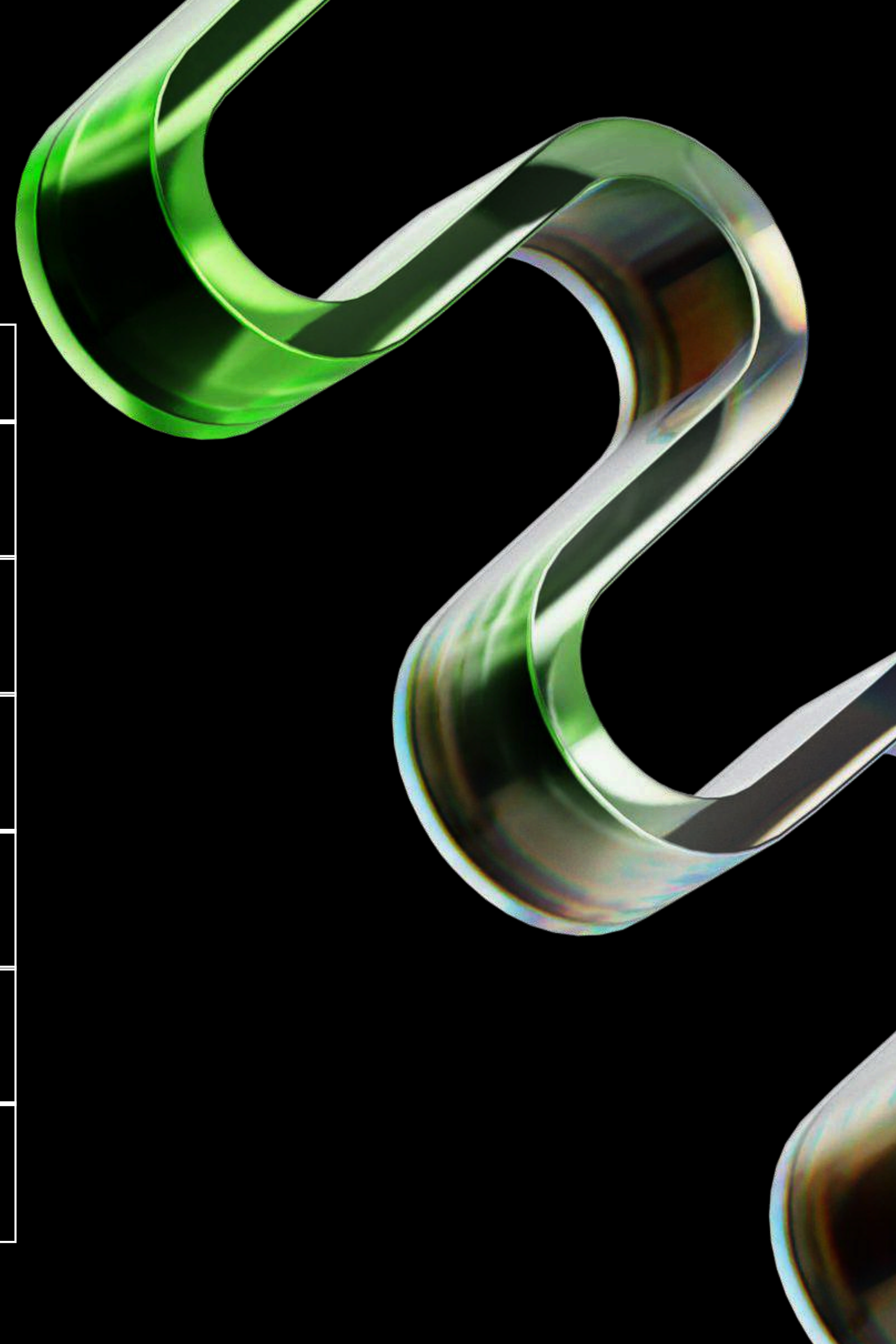
GPS

Приемник с поддержкой Pulse Per Second (PPS) (точность ~ 10-100 ns)

Специализированное оборудование

Масштаб

Операция	Типичная задержка
Простые арифметические и логические операции, процессор 3.5 GHz	~ 300 ps
Доступ к L1-кэшу	~ 0,5 – 2 ns
Ethernet 100 Gbps пакет 64Б	5,12 ns
1 Gbps пакет 64Б	512 ns
Syscall	~ 0,3 – 2 us
Обработка пакета в ядре Linux	~ 30 – 400 us



Генератор T-RAF

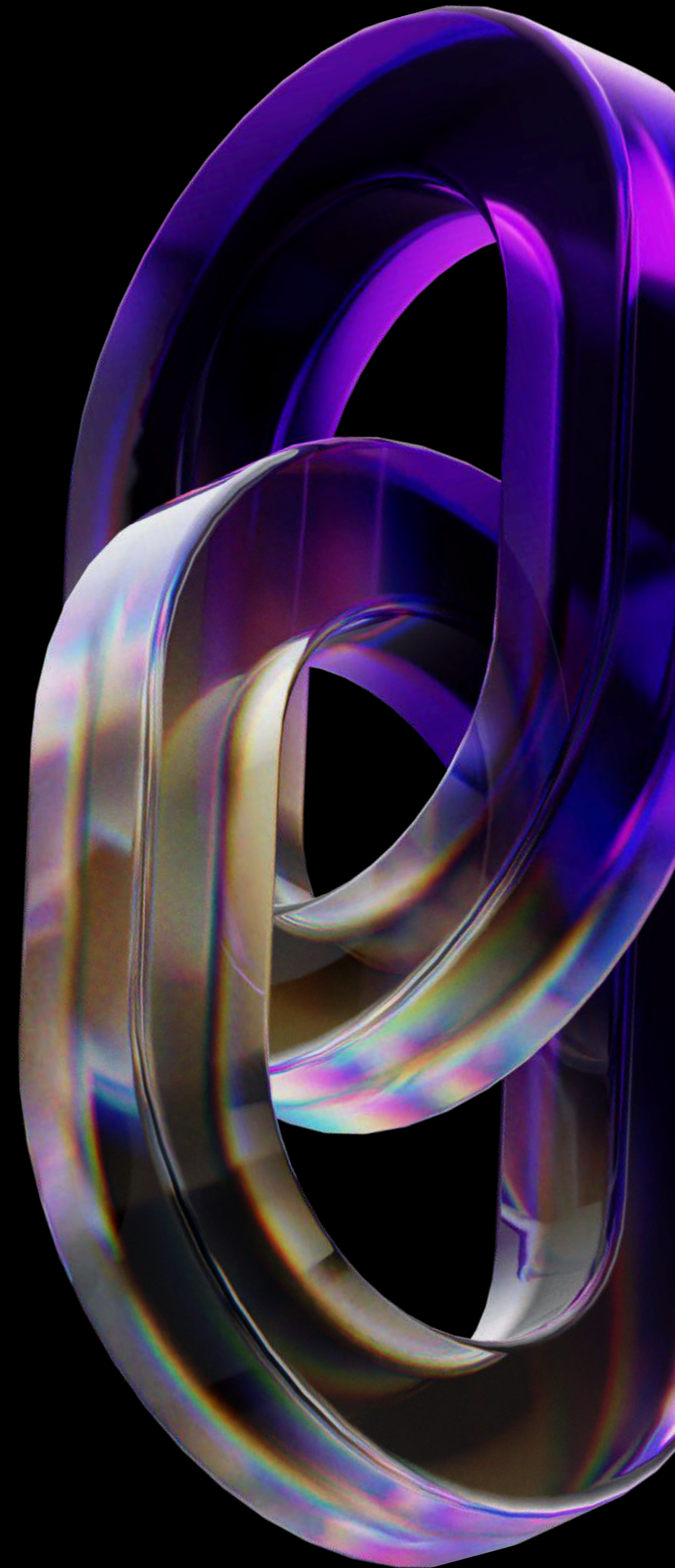
Единица передачи: UDP датаграмма/пакет

Измеряет сквозную задержку

Позволяет настроить распределенный эксперимент

Высокая производительность

Интервал отправки пакета: M и D



```
git clone https://gitflic.ru/project/mpei-vmss/t-raf.git  
cd t-raf  
mkdir build && cd build && cmake ..  
make
```

Сборка 
программы


```
git clone https://github.com/ntop/PF_RING.git  
cd PF_RING/kernel  
make  
sudo make install  
cd PF_RING/kernel  
sudo insmod ./pf_ring.ko
```

PF_RING
МОДУЛЬ



Характеристика	UDP сокет	RAW сокет	PF_RING сокет
Уровень	Transport (L4)	Network (L3)	Link (L2)
Доступ к полям пакета	Только payload	IP-пакет	Ethernet-кадр
Нужен root-доступ?	Нет (обычно)	Да	Да

Сокеты
t-raf



```
experiment:
```

```
# Продолжительность эксперимента в секундах
```

```
time: 10
```

```
# Путь к папке для сохранения результатов
```

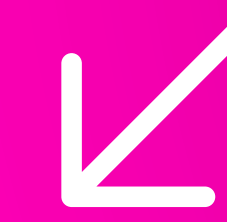
```
# Каждый узел создает файлы
```

```
# client создаёт файл <output_directory>/<имя_роли>-sent.csv
```

```
# server создаёт файл <output_directory>/<имя_роли>-<порт>-rcv.csv
```

```
output_directory: './results/'
```

Настройка ↗
эксперимента



Настройка эксперимента

```
# Описание сети
network:
# Параметры узла с ролью "server1"
server1:
# IP адрес
ip: "127.0.0.2"
# Имя интерфейса
int: 'lo2'
traffic:
# Может отправлять (client)/получать трафик (server), ниже пример для
client1
server:
# ТОЛЬКО если ports = [0], Если True, сокет использует pfring, если False -
RAW СОКЕТЫ
use_pfring: 'True'
enabled: 'True'
# Выключить запись в файл (для снижения задержки обработки пакета)
disable_output: 'False'
# 0 - для pfring или RAW СОКЕТОВ, 1 - 65535 - UDP сокет
ports: [50001]
# Размер принимаемых пакетов
length: 64
```

```
client1:
# IP адрес
ip: "127.0.0.3"
# Имя интерфейса
int: 'lo3'
traffic:
  client:
#    Тип отправляемого трафика:
#    'm' или 'exponential' – экспоненциальное распределение (по умолчанию,
случайное),
#    'd' or 'deterministic' – детерминированное распределение (через равные
промежутки),
#    'max' – минимальная задержка между пакетами, укажите количество
отправляемых пакетов "npackets"
#    'n' or 'none' – только синхронизация времени
distribution: 'm'
# Интенсивность отправки пакетов
intensity: 80000
# Размер IP пакета (64 – 65575)
length: 64
# Если distribution: 'max'
# npackets: > 0 – отправляет пакеты до достижения количества npackets
# npackets: 0 – отправляет пакеты
npackets: 2000000
```

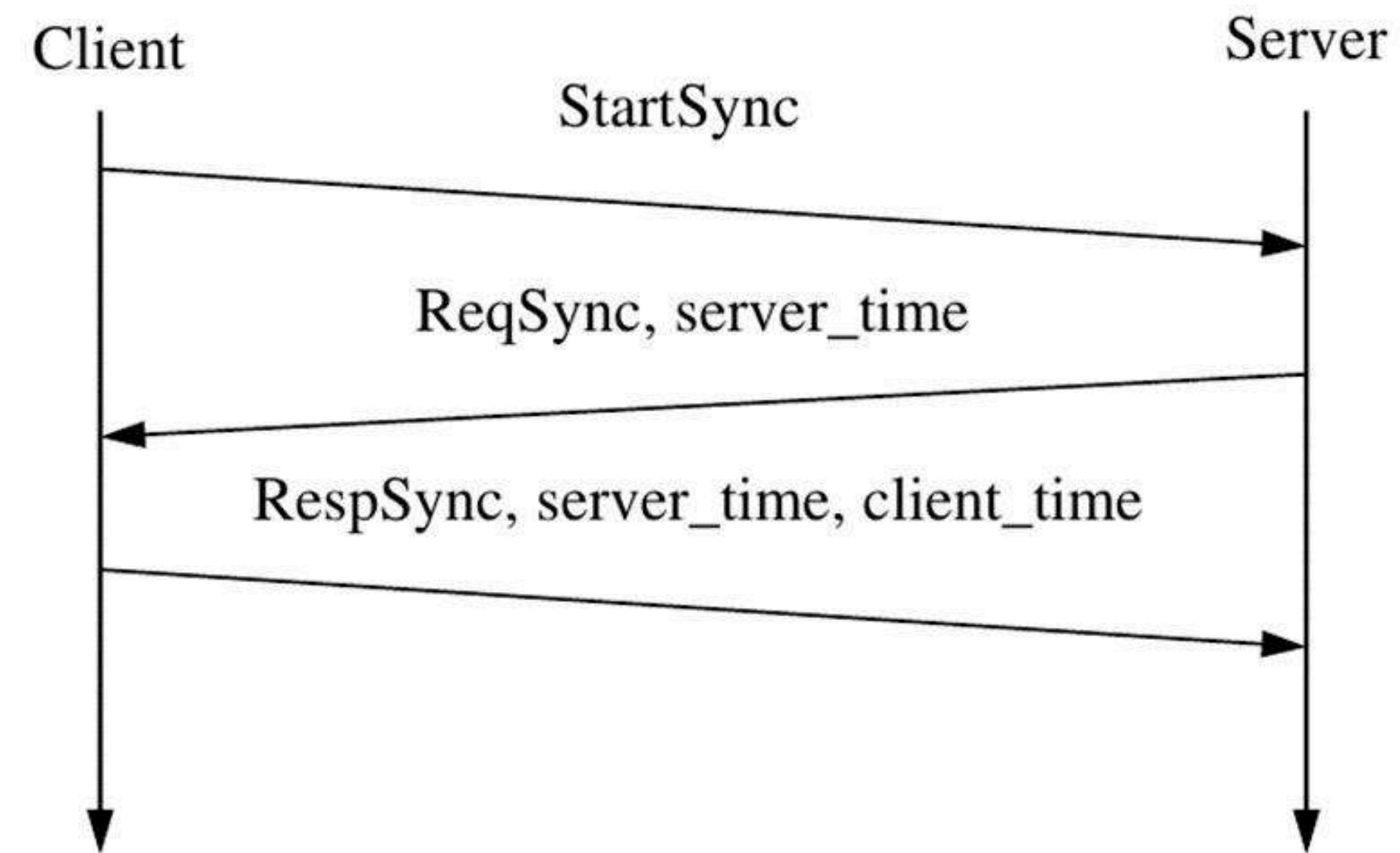
↙ Настройка эксперимента

Настройка эксперимента

```
# IP адрес получателя пакетов
# Альтернативный способ указания нескольких IP адрес для одного direction_mac
# direction: ['127.0.0.2', '127.0.0.3', '127.0.0.4', '127.0.0.5']
# length of packet sequence sent continuously to one of direction IPs, default
value is 1, parameter has no meaning if only one direction IP is specified
# sequence_length: 1
# direction: '127.0.0.2'
# MAC адрес получателя пакетов
# direction_mac: '00:00:00:00:00:00'
# port to send data, 0 – если используется pfring или RAW сокеты
# port: 50001
# Если port:0,
# use_pfring: 'True' – pfring сокеты
# use_pfring: 'False' – RAW сокеты
# use_pfring: 'True'
# single_thread: 'True' – генерация пакетов выполняется в главном потоке
# single_thread: 'False' – генерация пакетов выполняется в отдельном потоке
# single_thread: 'True'
server:
  enabled: 'False'
```


Скопируйте YAML файл на узлы,
участвующие в эксперименте
Запустите генерацию командой:
t-raf < путь к файлу YAML > < Название роли
устройства в конфиге >

Запуск ↗
генератора



Режим
синхронизации
времени



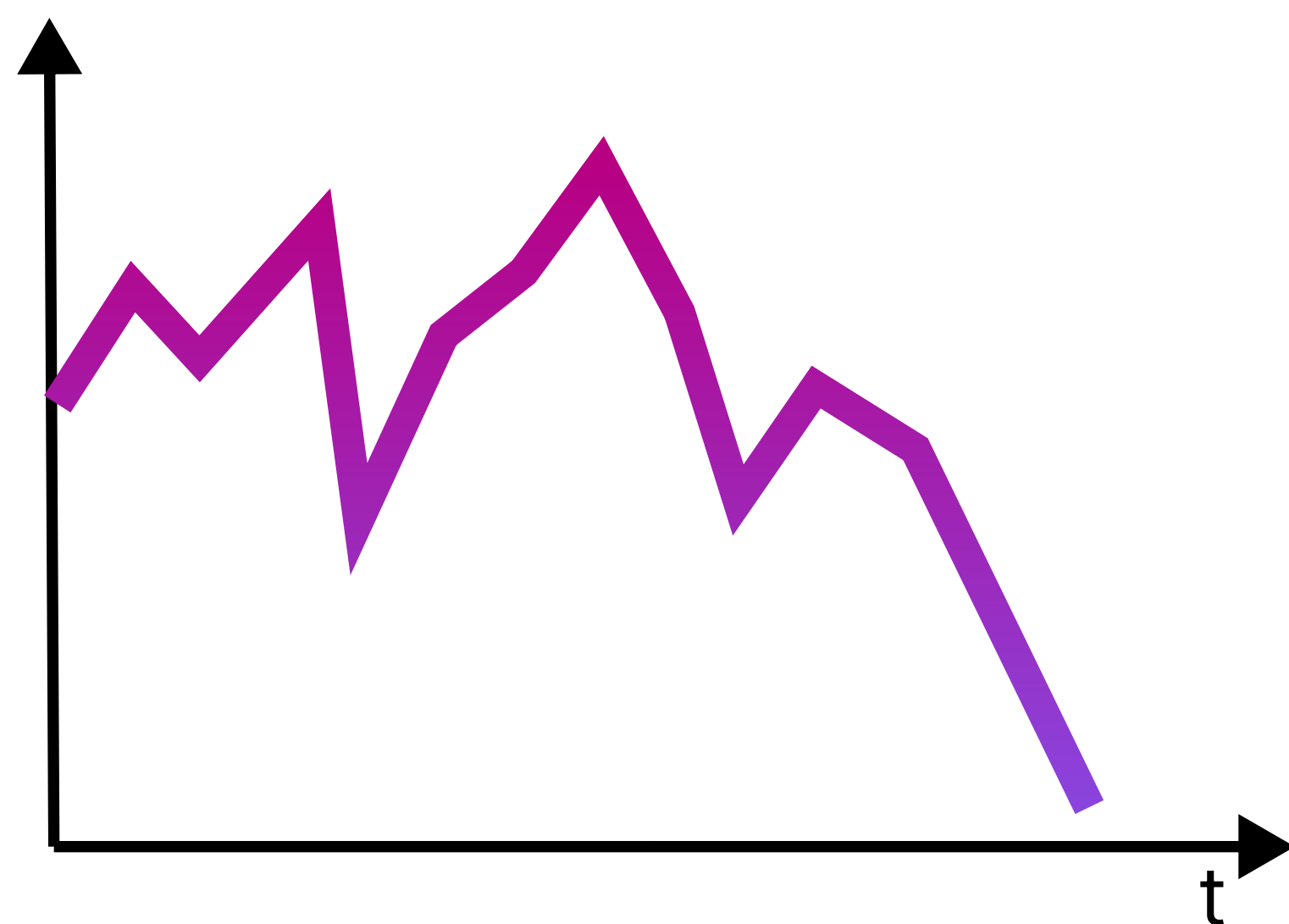
Результаты экспериментов

<имя_роли>-sent.csv

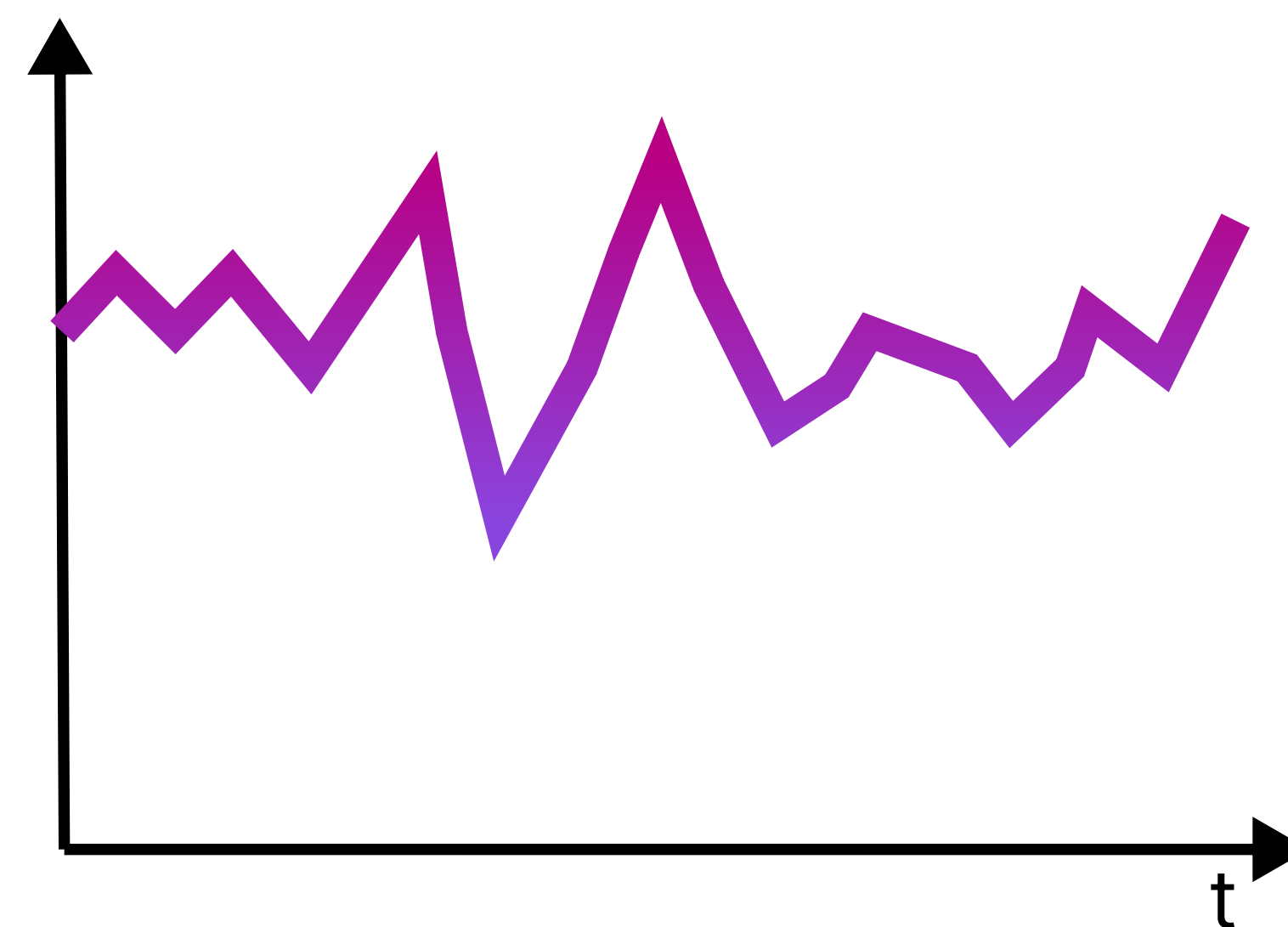
<имя_роли>-<порт>-rcv.csv

<метка времени нс>,
 θ_i ,
<номер пакета>

<метка времени нс>,
<номер пакета>,
<сквозная задержка>



Низкая производительность



Высокая производительность

Неприятные
аномалии ↗

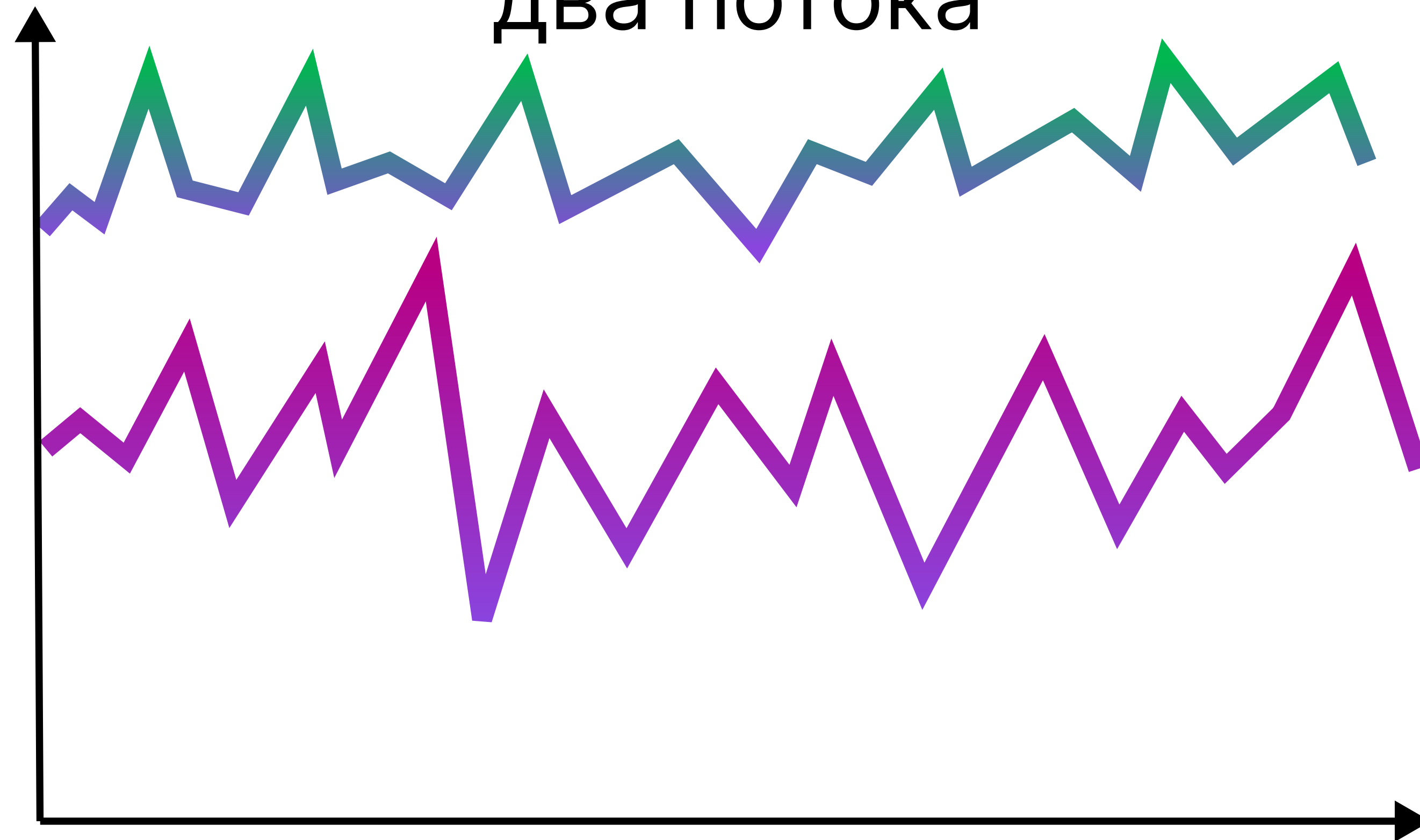
Продолжительность
эксперимента | Выборка

№ пакета	Задержка обработки, нс	Интервал поступления пакета, с
0	112759991	-
1	466255	0,157
2	113779030	2,694
3	103435049	3,063
4	682389	0,576
5	325330	1,058
6	354514	2,17
7	320391	0,987
8	337182	0,564
9	287388	0,072
10	440576	0,237

Неприятные
↙ аномалии

Кеширование

два потока



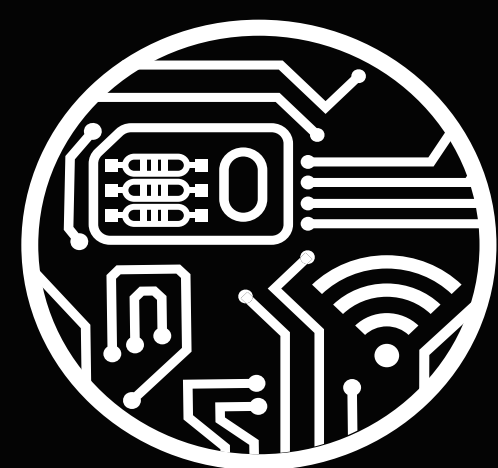
Неприятные
аномалии

Хеширование



Неприятные
аномалии

Потери



ВМСС

КАФЕДРА ВЫЧИСЛИТЕЛЬНЫХ МАШИН,
СИСТЕМ И СЕТЕЙ

Спасибо за внимание!



Контакты: OrlovaMA@mpei.ru