

# Балансировка сетевого трафика

к.ф.-м.н., м.н.с. Степанов Евгений Павлович

# Программа курса

## Подходы:

- 1. Управление перегрузкой**
  - Современные протоколы управления перегрузкой TCP
- 2. Демультимплексирование/мультиплексирование**
  - Многопоточные транспортные протоколы
  - Маршрутизация на уровне интернет провайдеров
  - Network Coding
- 3. Сегментация**
  - TCP Proxy
- 4. Балансировка**
  - Балансировка нагрузки и управление трафиком
- 5. Преобразование сообщений**
  - FEC
  - Сжатие

## Модели оценки качества сервиса:

- Активные и пассивные измерения
- NS3: моделирование поведения сети с высокой точностью
- Сетевое исчисление: математический подход к качеству сервиса

## Примеры:

- Управление сетевыми ресурсами в Центрах Обработки Данных
- Обеспечение качества сервиса в сетях доставки контента
- Пропускная способность по требованию

# Балансировка нагрузки

*в контексте компьютерных сетей*

- Метод распределения трафика между несколькими хостами, линиями связи и коммутирующими устройствами
- Цели балансировки
  - Повышение степени утилизации ресурсов
  - Повышение пропускной способности
  - Уменьшение времени отклика системы
  - Борьба с отказами и перегрузками

# Балансировка нагрузки

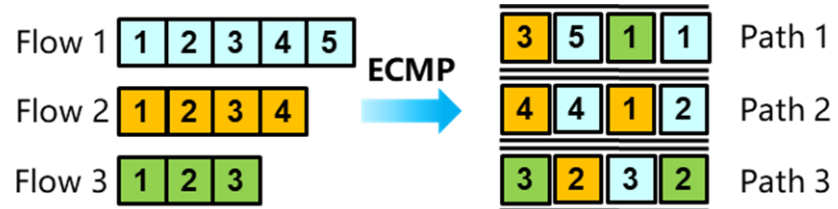
## *классификация*

- По устройствам и по времени
  - Burstable – способ оплаты трафика, когда расчёты исходят из фактически потреблённой полосы в рамках выделенной
- Статическая и динамическая
  - Учёт нагрузки vs простота реализации
- Централизованная и распределённая
  - Выделенный контроль за балансировкой

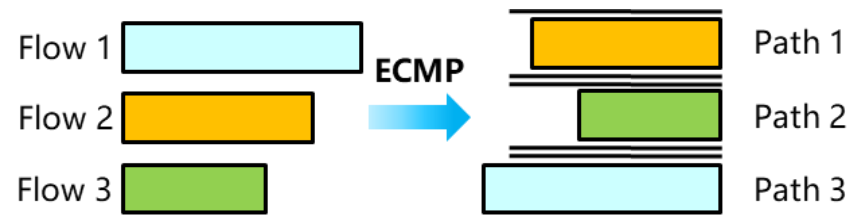
# Балансировка нагрузки

## классификация

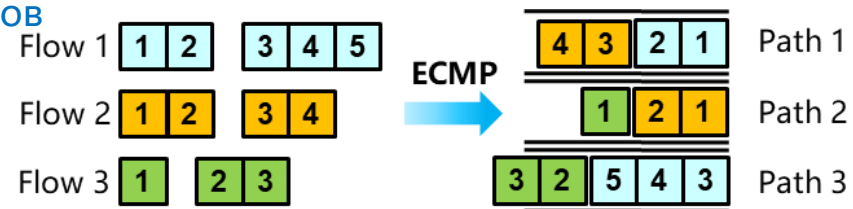
### ➤ Балансировка пакетов



### ➤ Балансировка потоков



### ➤ Балансировка всплесков



### Недостатки:

Неупорядоченный приход пакетов приводит к снижению скорости

Не учитывается размер потоков

Не учитывается размер всплесков

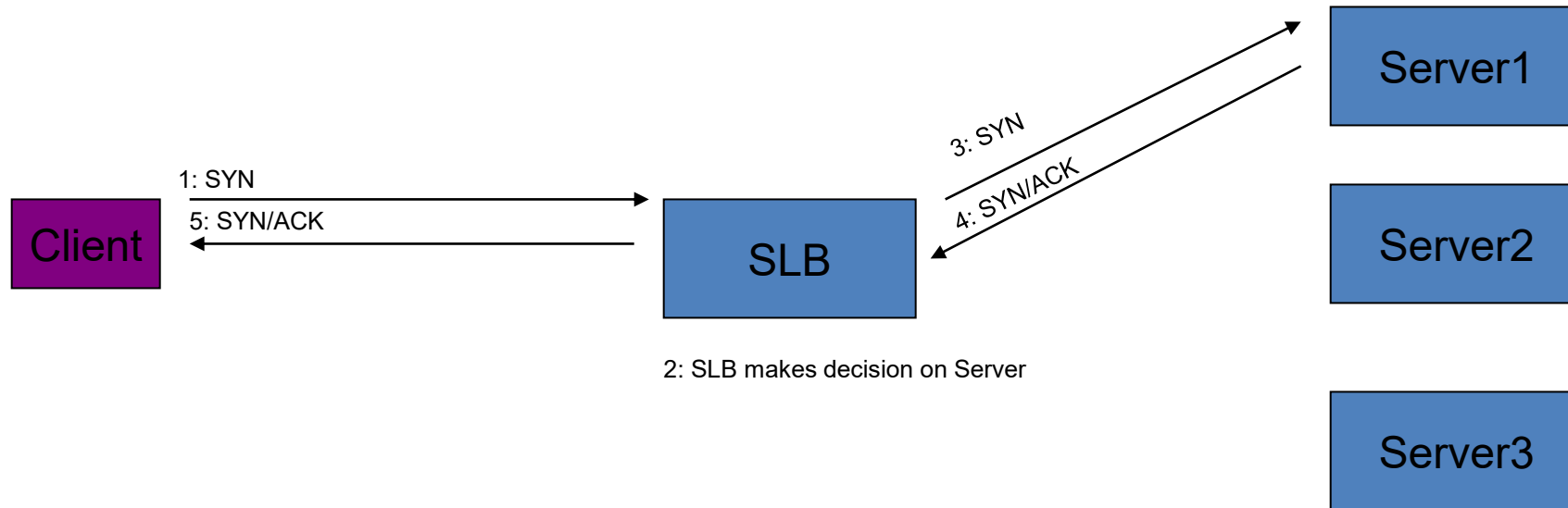
➤ Sinha S., Kandula S., Katabi D. Harnessing TCP's burstiness with flowlet switching //Proc. 3rd ACM Workshop on Hot Topics in Networks (Hotnets-III). – 2004.

# Балансировка трафика по нескольким маршрутам

Недостатки передачи данных по одному пути:

- Ограничение пропускной способности сети
  - Маршрутизация по одному пути не позволяет достигнуть оптимума функции утилизации
  - Производительность в угоду простоте
- Подверженность колебаниям
  - Ранний адаптивный протокол маршрутизации, который учитывал задержку на линиях связи оказался нестабильным

# Server Load Balancing (SLB)



# Стратегии балансировки

Статические:

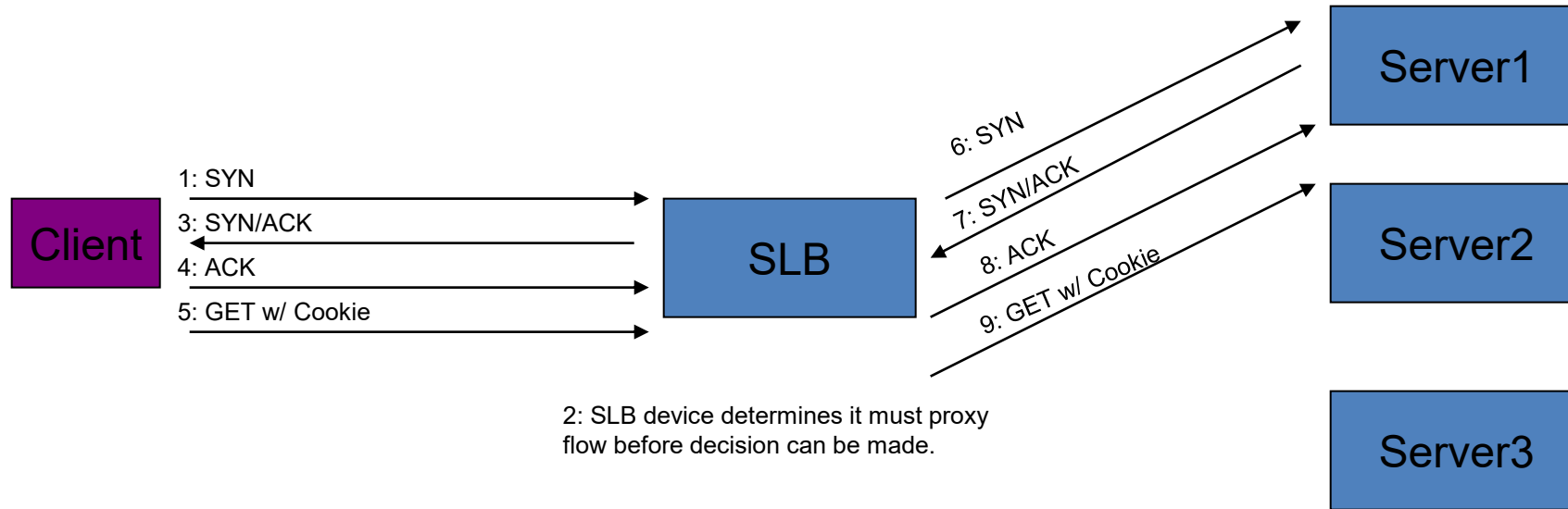
- Round Robin
- Hash-based
  - Разные методы расчёта

Динамические:

- Least Connections
  - Выбирается интерфейс, обслуживающий наименьшее количество соединений
- Lowest Response Time
  - Выбирается интерфейс с наименьшим временем завершения запроса

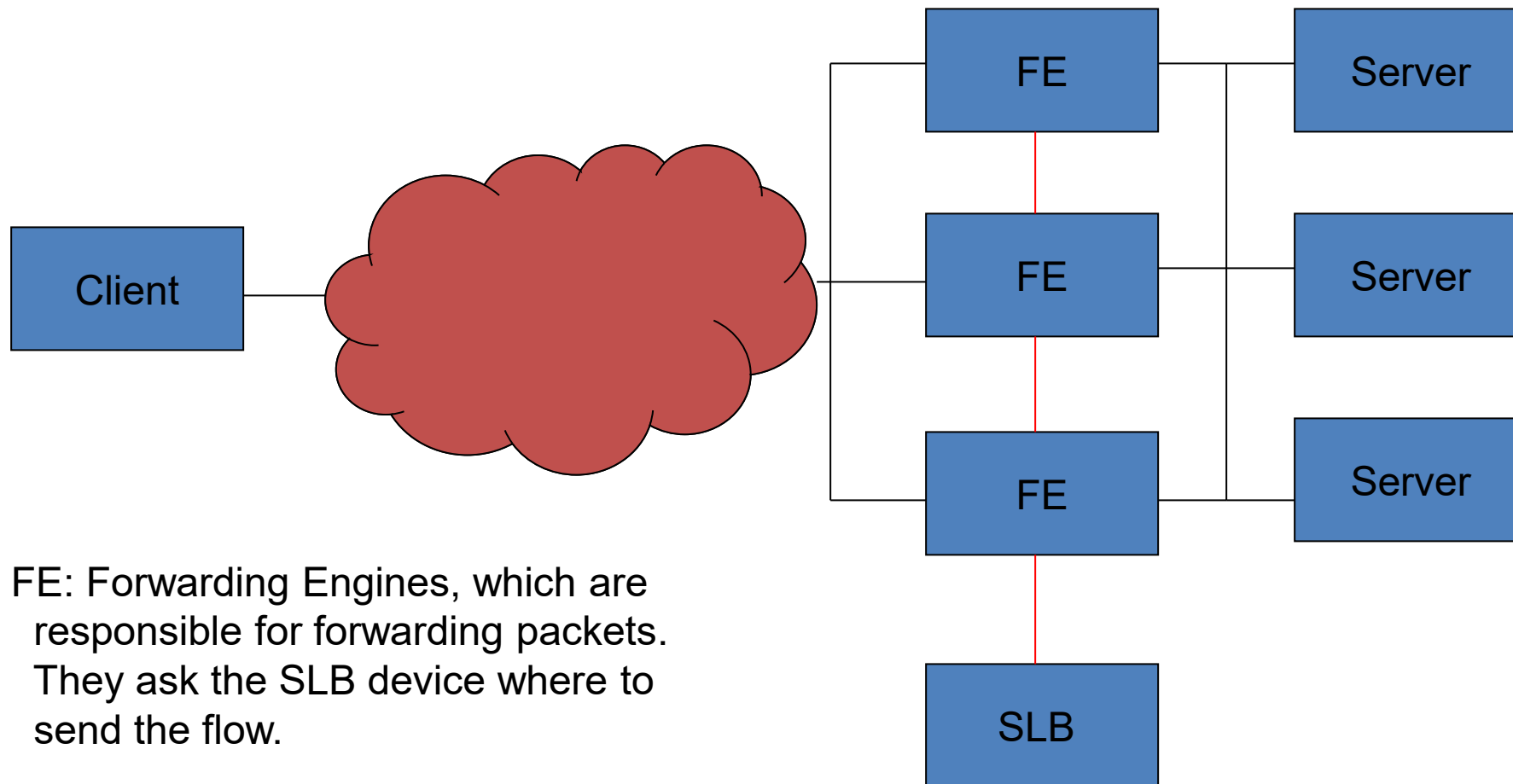


# Балансировка на 5+

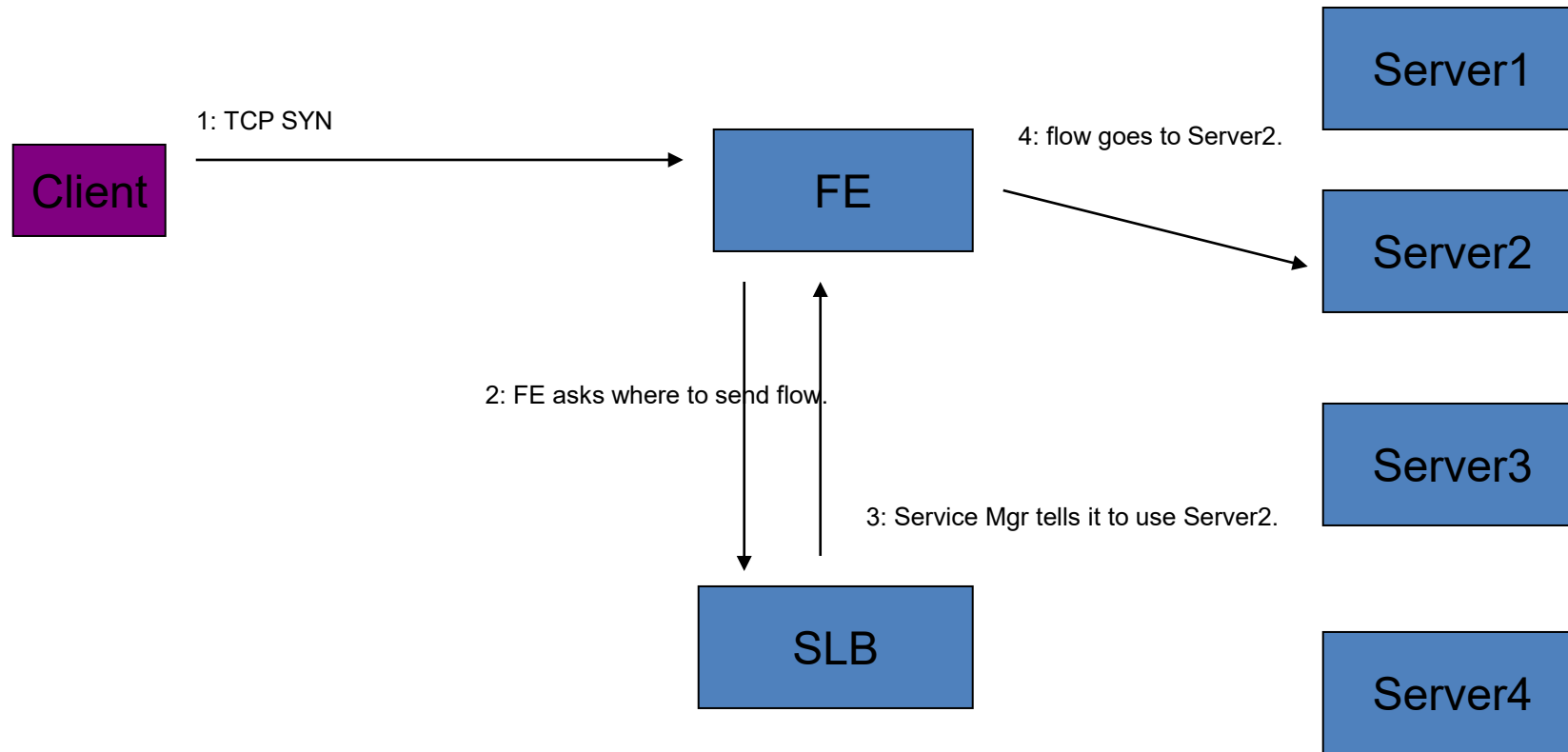


При работе на L5 балансировщик не может определить, куда перенаправлять поток данных заранее => Он должен взять часть работы сервера на себя

# Распределённый балансировщик

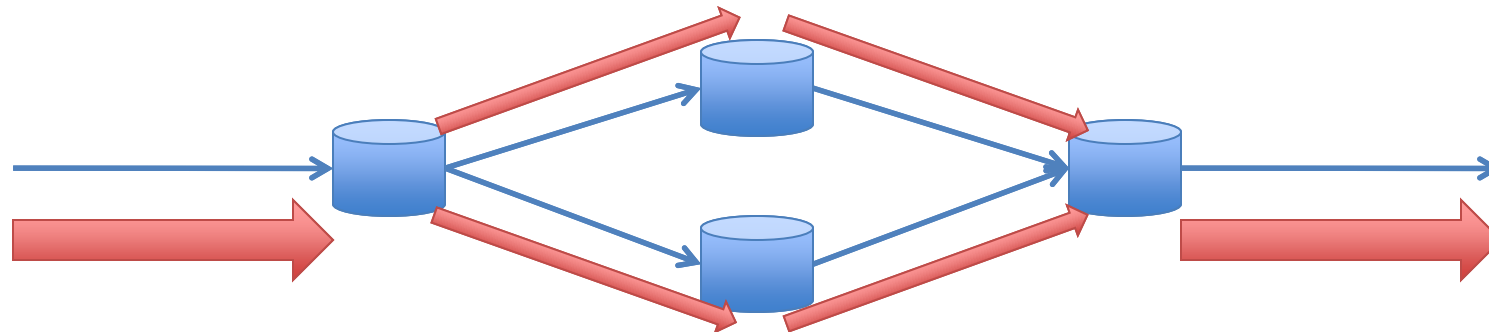


# Распределённый балансировщик



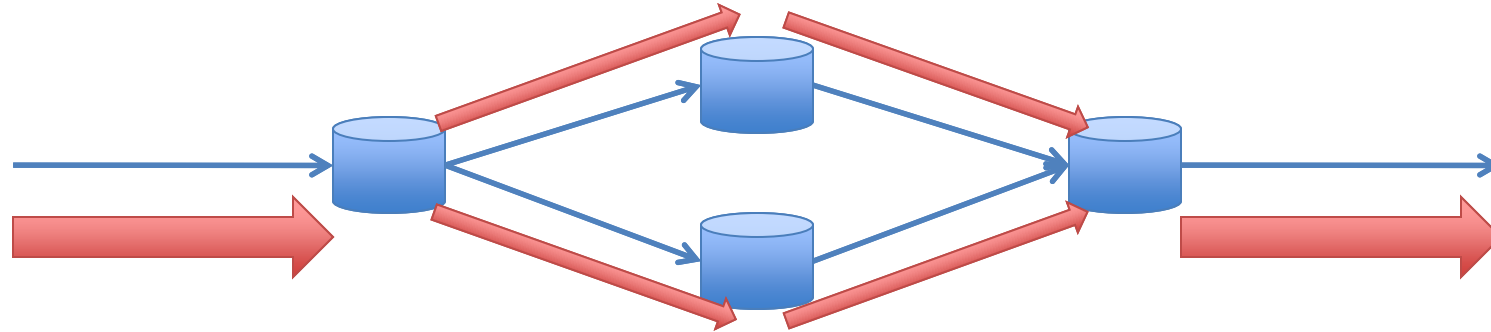
Subsequent packets flow directly from Client to Server2 thru the FE.  
The FE must notify the SLB device when the flow ends.

# Equal-cost multipath routing (ECMP)



- ECMP
  - Распределяет нагрузку между множеством путей с одинаковыми метриками маршрутизации
- Round-Robin использовать нельзя:
  - Переупорядочивание пакетов (TCP такого не любит)
  - Разные пути могут иметь разные RTT, PLR...
  - На разных путях может быть разный MTU

# Equal-cost multipath routing (ECMP)



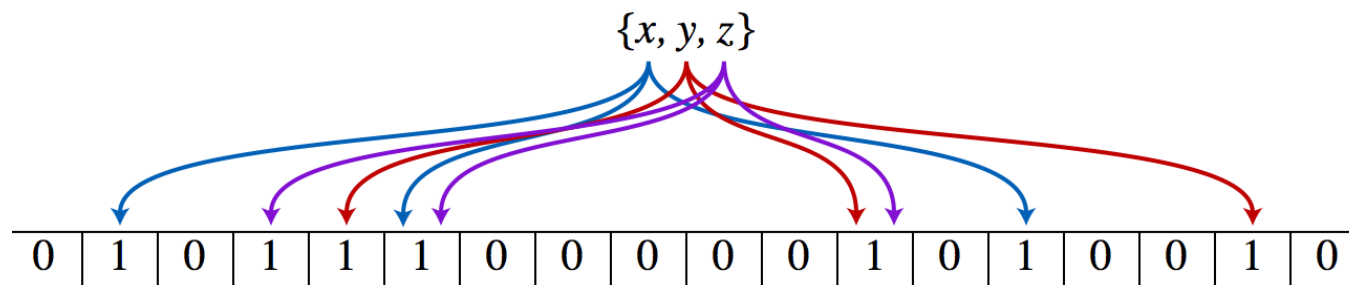
- Для балансировки используется хэширование
  - количество бакетов = количеству выходов
  - Хэш на основе адресов пакетов сохраняет привязку потоков по выходам

# Basic Hash Techniques

- Simple approach for uniform data
  - If data distributed uniformly over  $N$ , for  $N \gg n$
  - Hash  $fn = \langle data \rangle \bmod n$
  - Fails goal of uniformity if data not uniform
- Non-uniform data, variable-length strings
  - Typically split strings into blocks
  - Perform rolling computation over blocks
    - CRC32 checksum
    - Cryptographic hash functions (SHA-1 has 64 byte blocks)

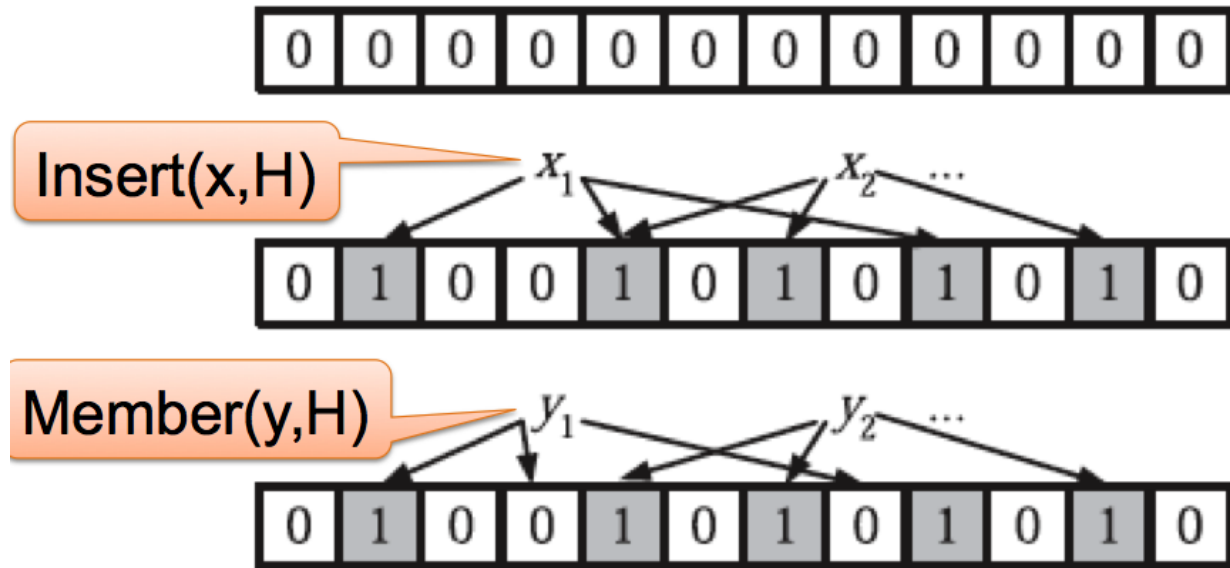
# Bloom Filters

- Data structure for probabilistic membership testing
  - Small amount of space, constant time operations
  - False positives possible, no false negatives
  - Useful in per-flow network statistics, sharing information between cooperative caches, etc.
- Basic idea using hash fn's and bit array
  - Use  $k$  independent hash functions to map item to array
  - If all array elements are 1, it's present. Otherwise, not



# Bloom Filters

## Example Bloom Filter $k=3$

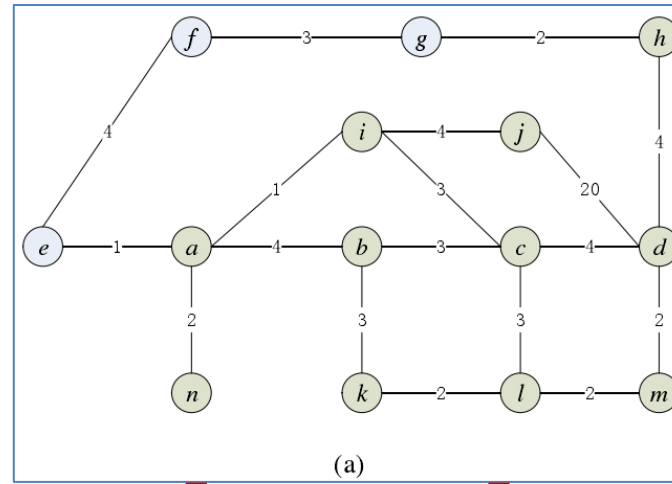


$y_1$  = is not in  $H$  (why ?);  $y_2$  may be in  $H$  (why ?)

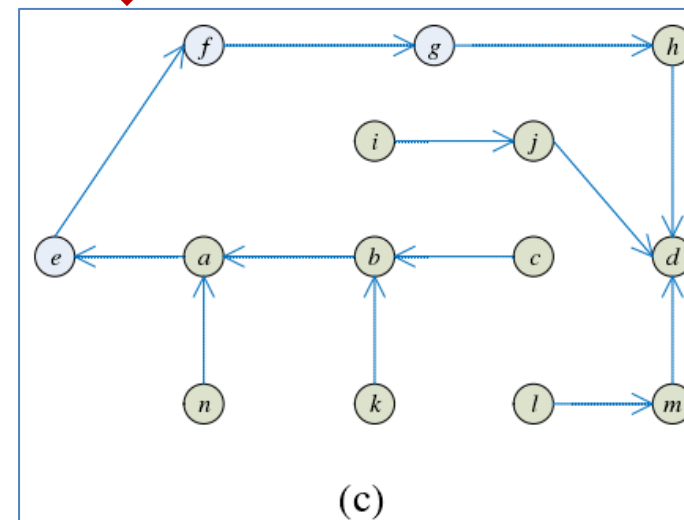
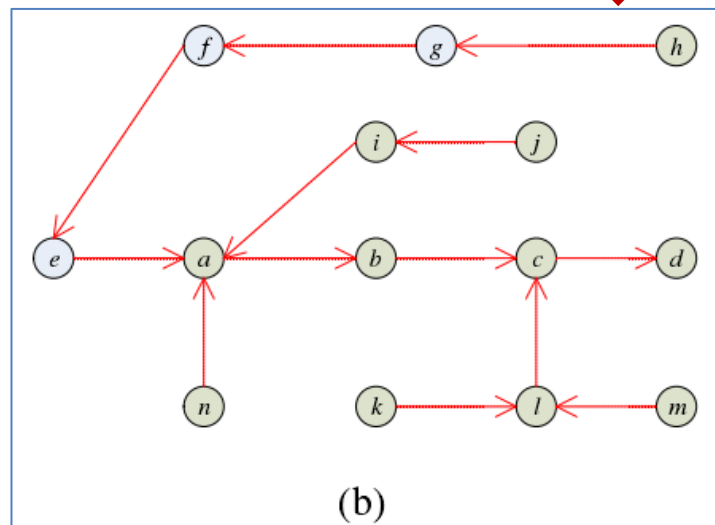


# Maximally Redundant Trees

- 2 trees  $O(n)$
- 3 trees  $O(n^2)$
- 4 trees  $O(n^3)$

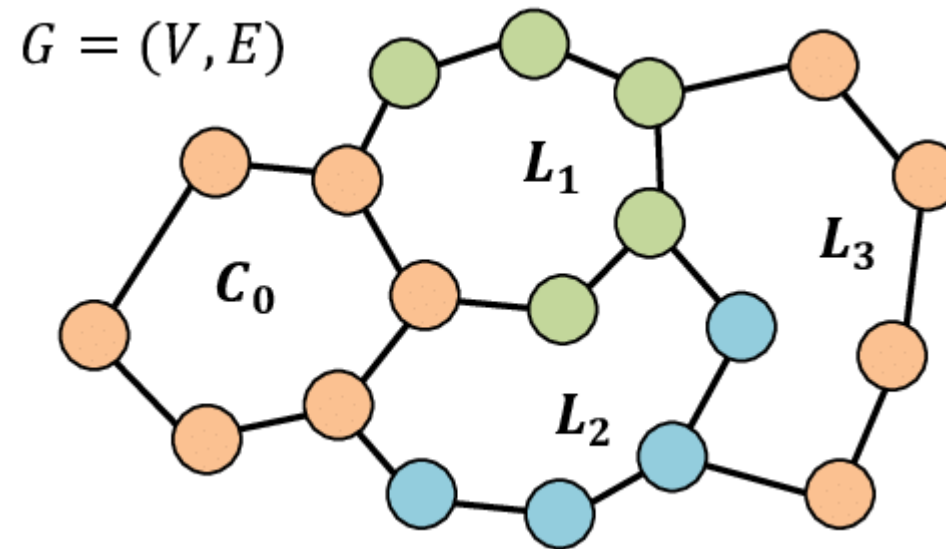


To destination  $d$ , each node to  $d$  has two paths.

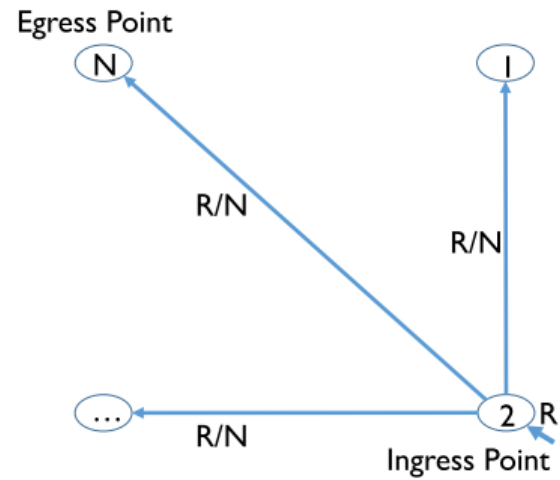


# Ear decomposition

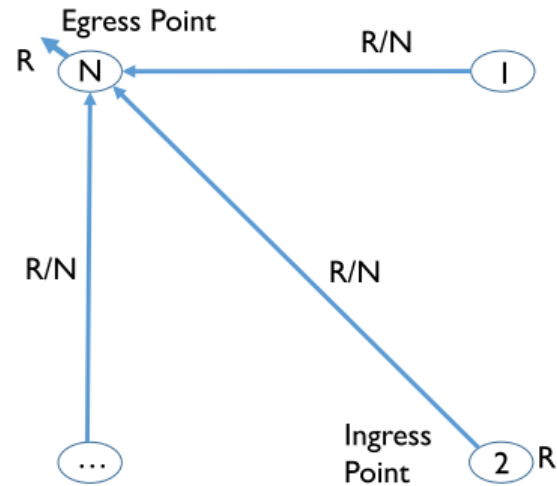
An **ear decomposition** of an undirected graph  $G$  is a partition of its set of edges into a sequence of ears, such that the one or two endpoints of each ear belong to earlier ears in the sequence and such that the internal vertices of each ear do not belong to any earlier ear.



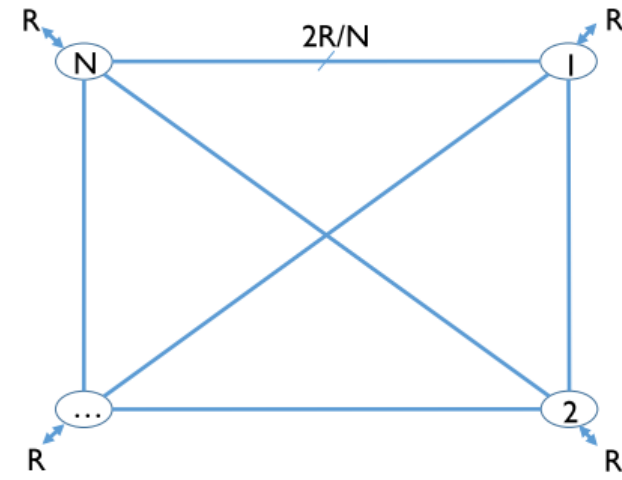
# Valiant Load Balancing



(a) VLB Stage 1

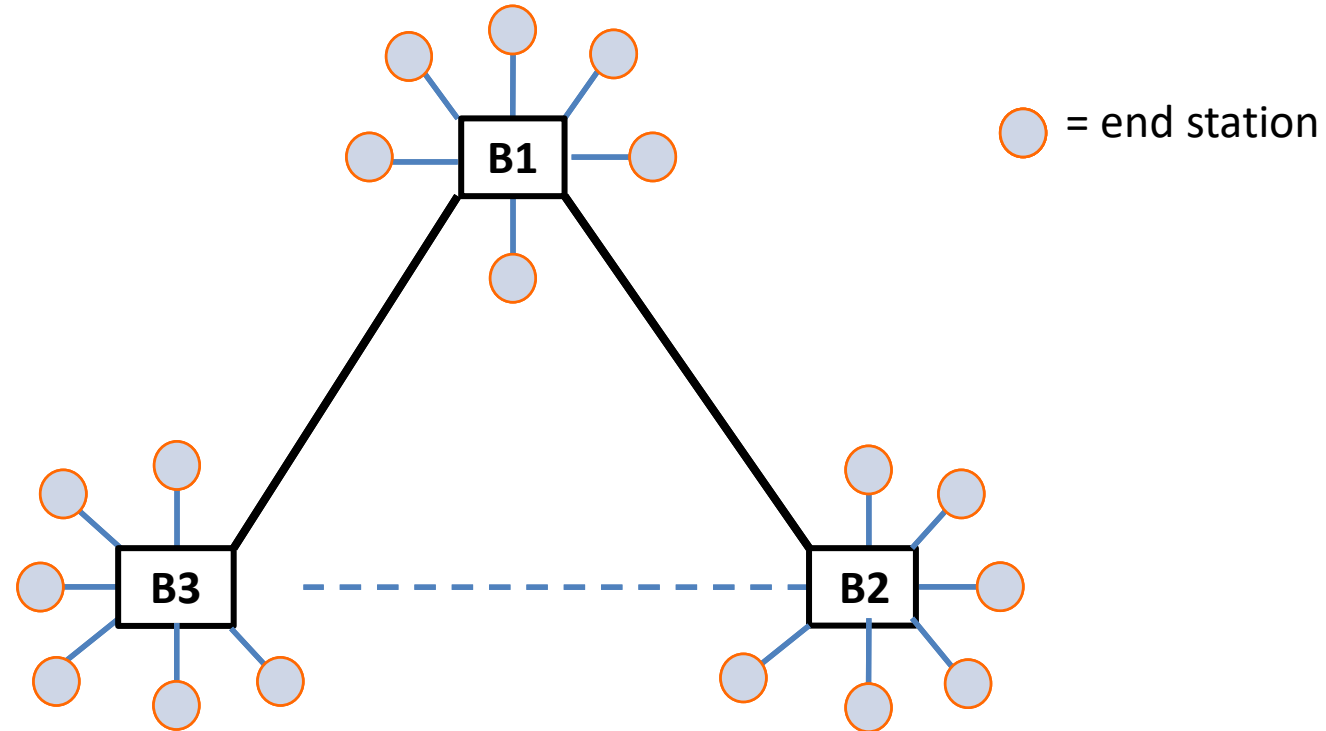


(b) VLB Stage 2



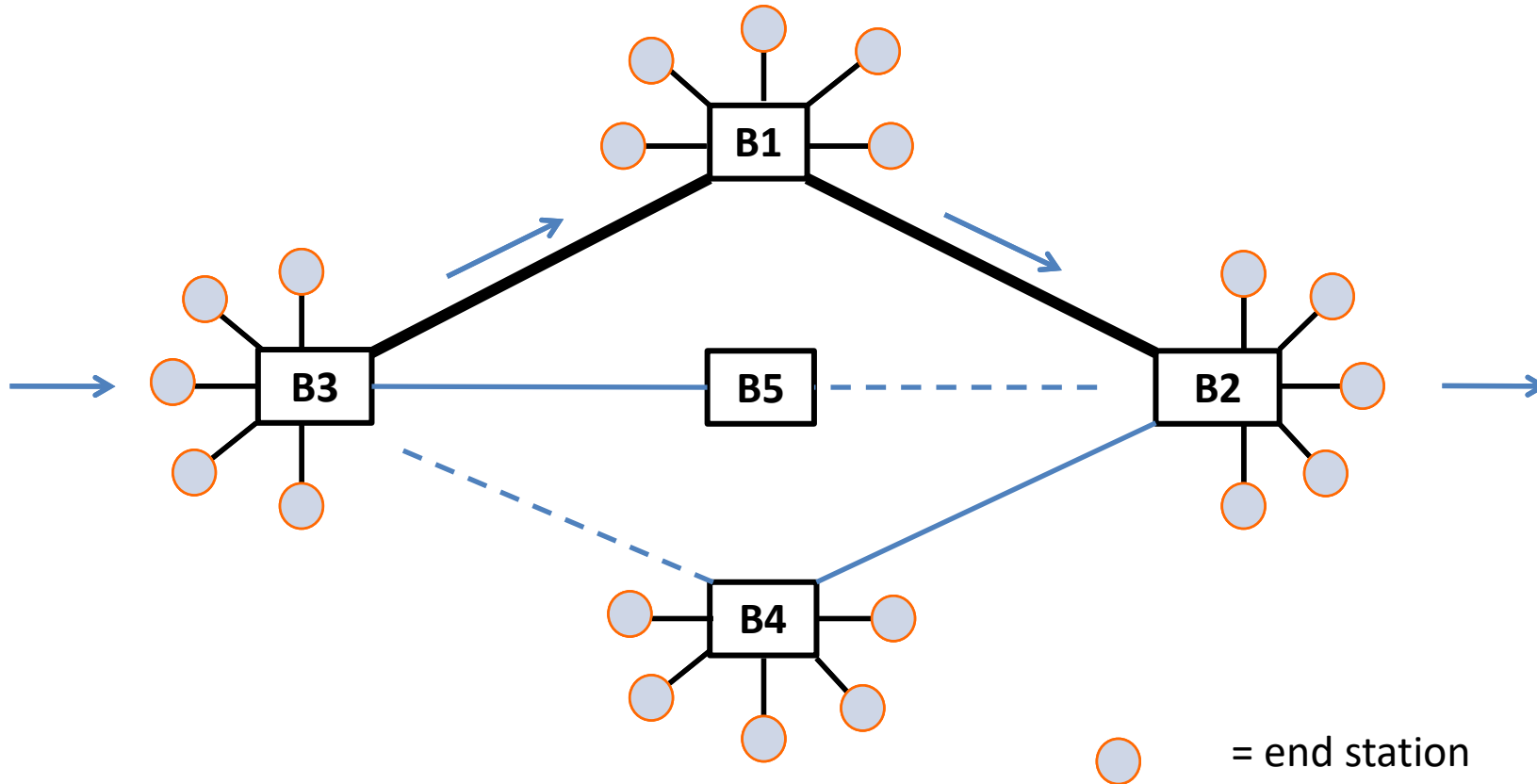
(c) VLB Stage 1+2

# Unicast Least Cost Paths



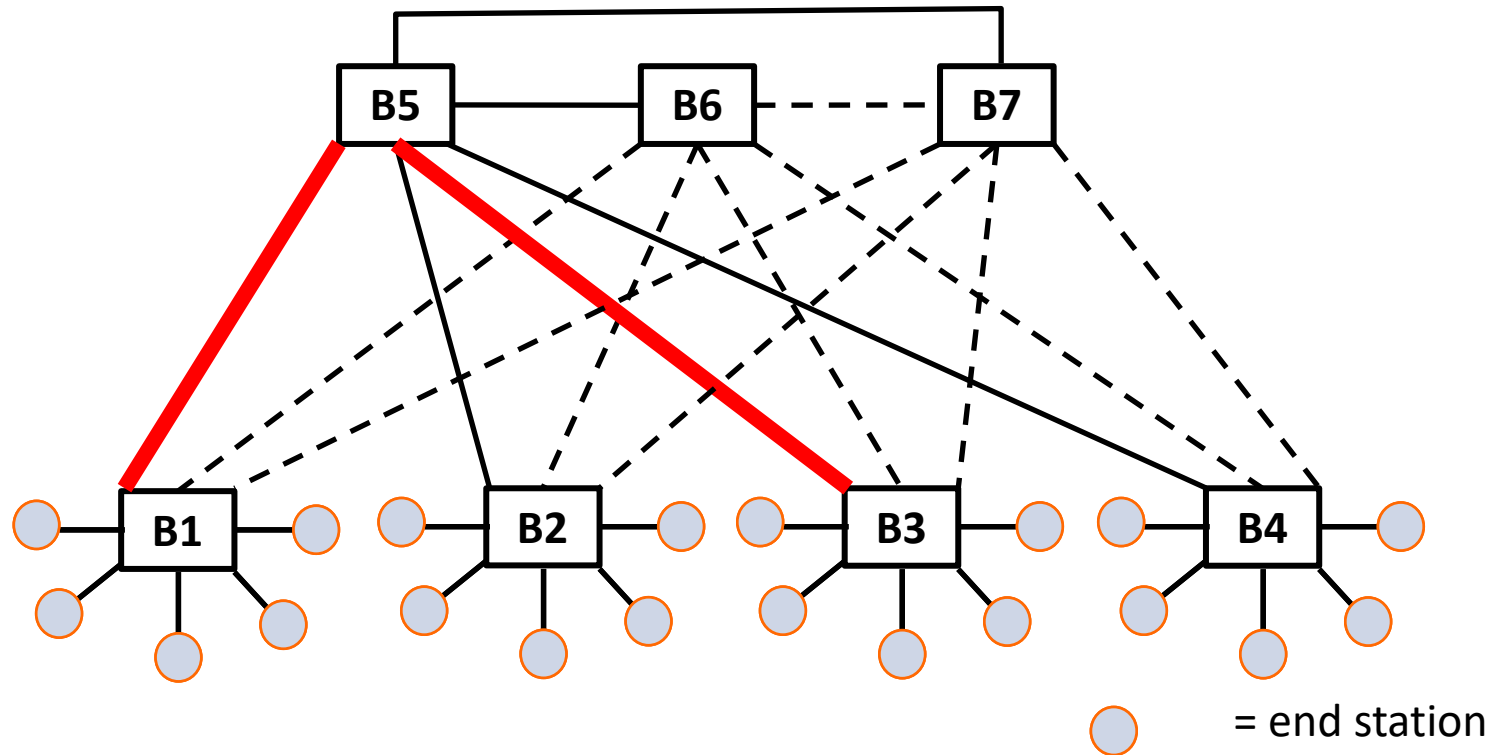
Spanning tree eliminates loops  
by disabling ports

# Unicast Multi-Pathing



Bridges limit traffic to one path

# Multi-Pathing (Unicast)



Bridges limit traffic to one path

# Протокол TRILL



# Протокол TRILL

- Основан на простой идее:
  - Добавить к передаваемым пакетам новый L2 заголовок с hop-counter'ом
  - Маршрутизировать пакеты с помощью IS-IS
  - Удалить дополнительные заголовки перед тем, как доставить пакет
- Преимущества:
  - Передача по лучшему пути
  - Встроенная поддержка Equal Cost Multi Path
  - Поддержка передачи на несколько назначений



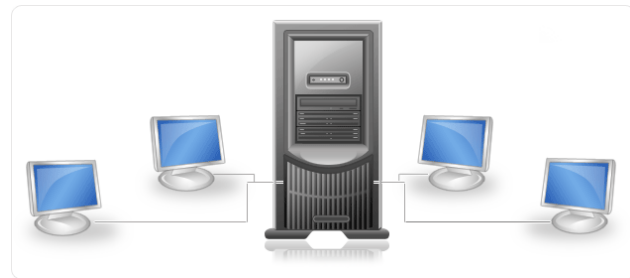
# TRILL has a poem

I hope that we shall one day see  
    A graph more lovely than a tree.  
A graph to boost efficiency  
    While still configuration-free.  
A network where RBridges can  
    Route packets to their target LAN.  
The paths they find, to our elation,  
    Are least cost paths to destination!  
With packet hop counts we now see,  
    The network need not be loop-free!  
RBridges work transparently,  
    Without a common spanning tree.

- By Ray Perlner (Radia Perlman's son)

# Меняется парадигма организации вычислений

- На смену клиент-сервисной архитектуре пришли Cloud Computing и концепция Software as a Service
- При высокой концентрации вычислительных ресурсов сеть становится узким местом



# При увеличении числа пользователей пропускная способность разделяется

Решение – увеличение  
количества передающих  
станций

**wimax**



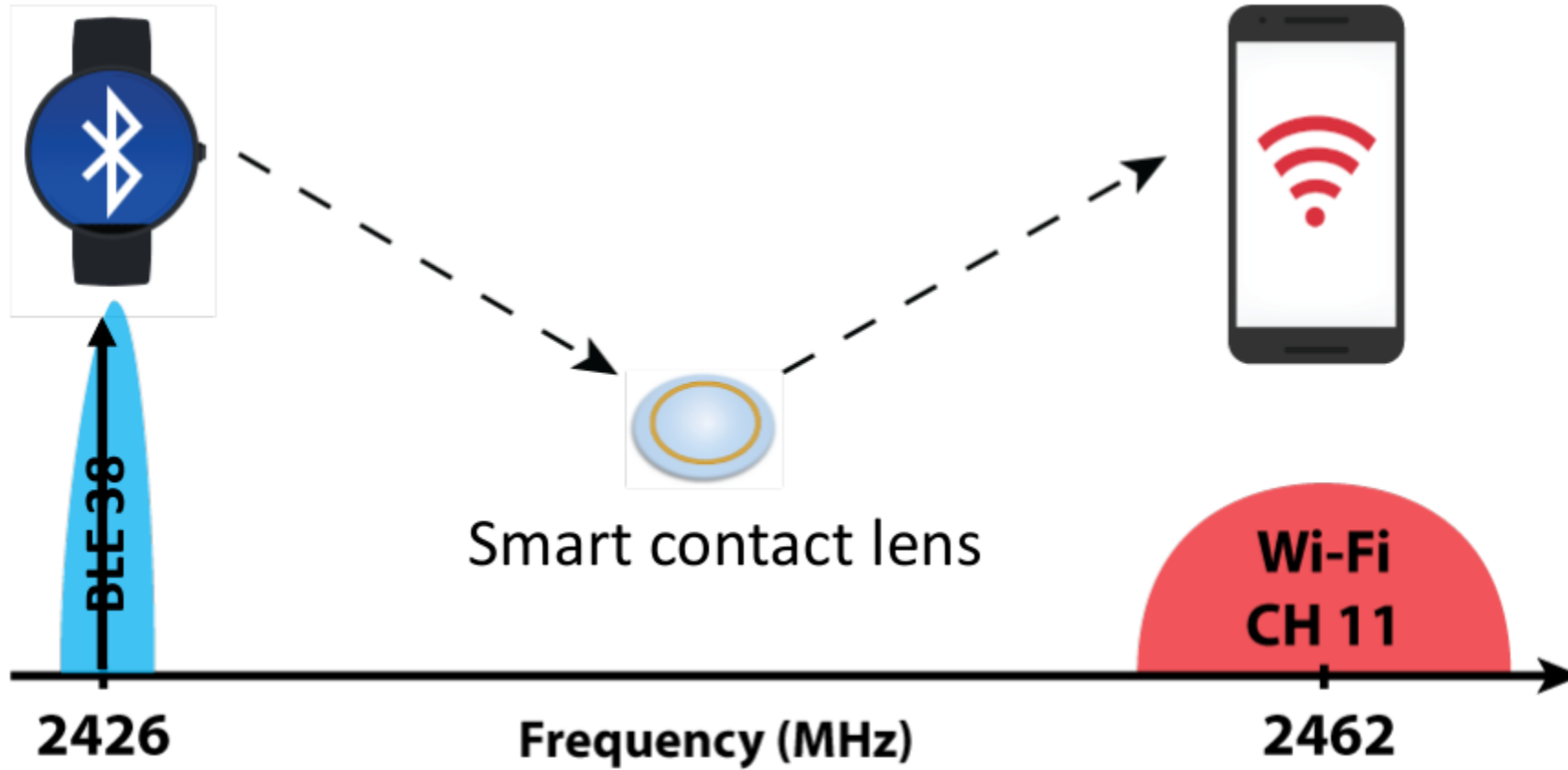
**Слишком дорого!**

**WiFi**

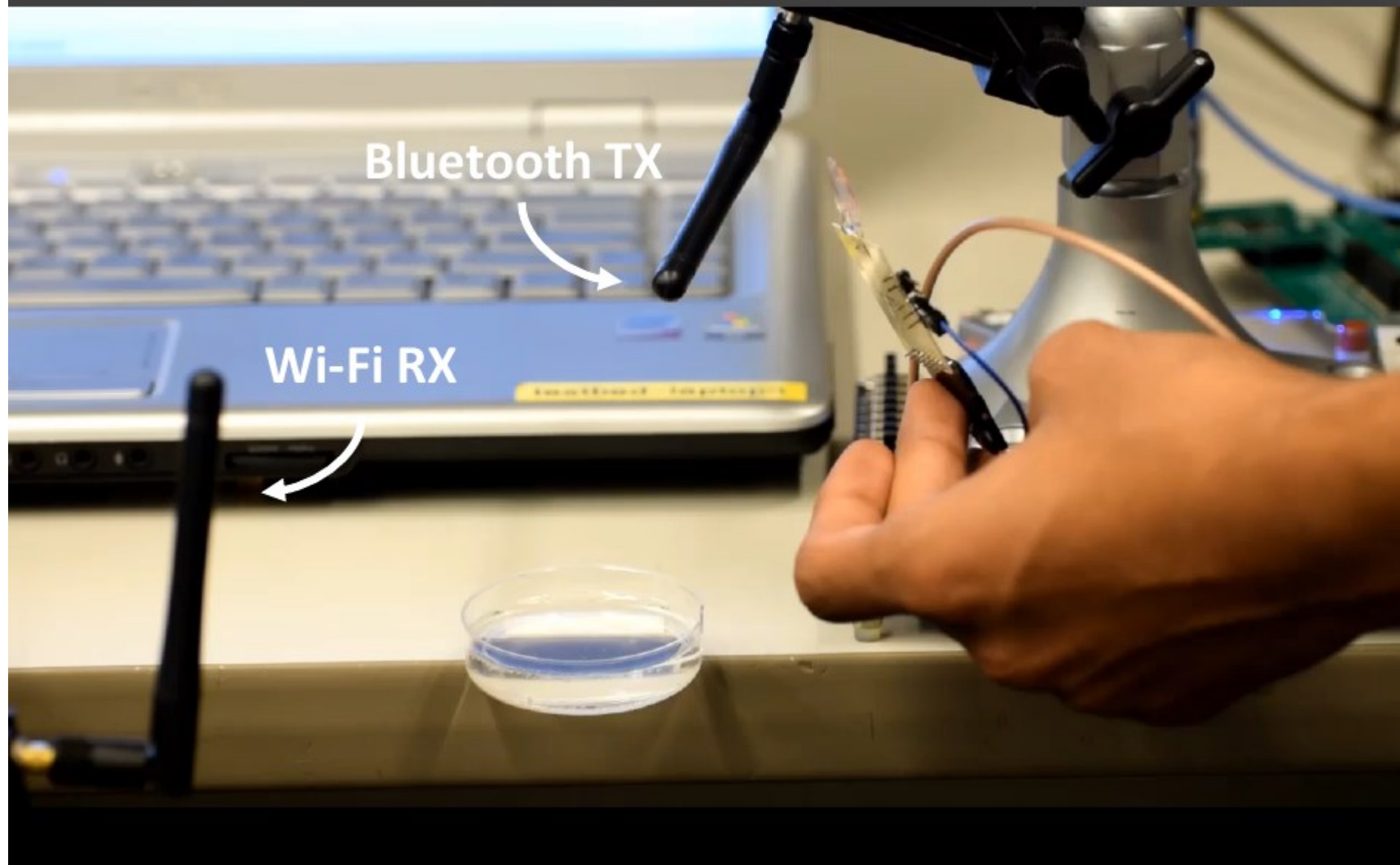


[Почему нельзя мультиплексировать  
передачу данных через разные каналы?!]

# Interscatter Communication



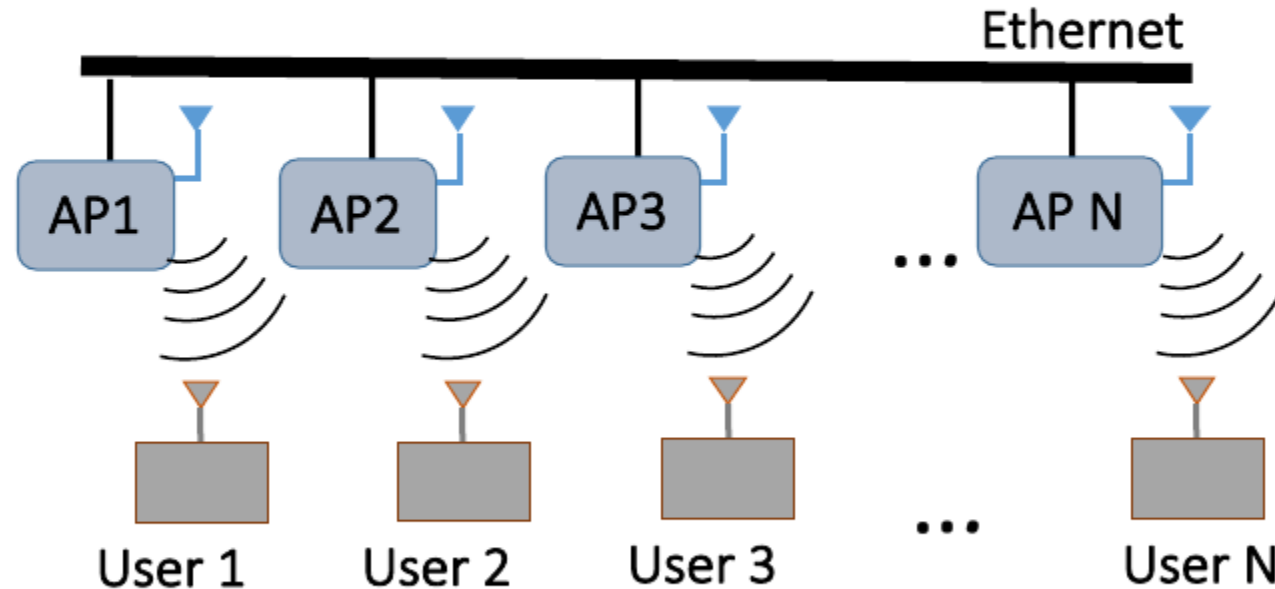
# Contact Lens Prototype



# Real-time Distributed MIMO Systems

## Distributed MIMO is the Holy Grail

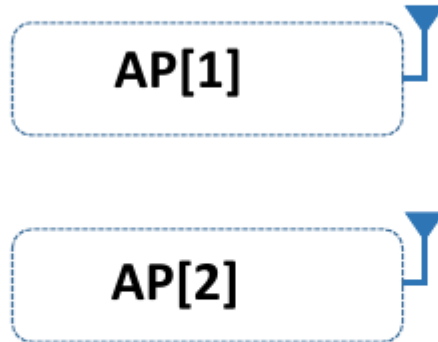
Distributed protocol for APs to act as a huge MIMO transmitter with sum of antennas



**N APs → N times higher throughput**

# Real-time Distributed MIMO Systems

## What happens with Distributed MIMO?

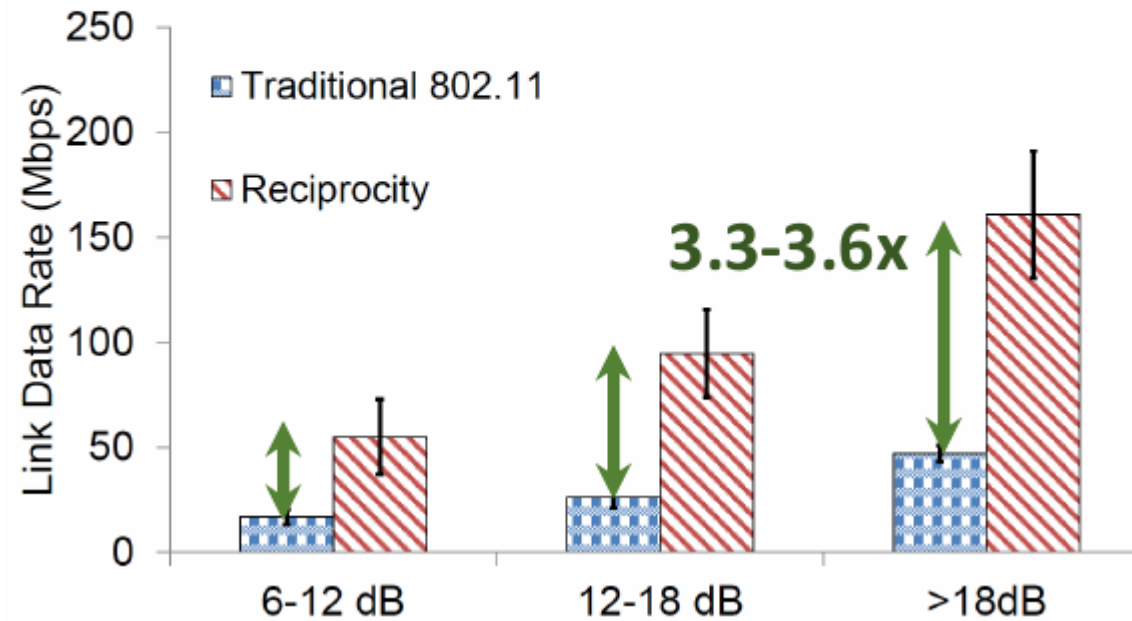


Separate devices → Different Crystals

→ RF chains have oscillator offset relative to each other

# Real-time Distributed MIMO Systems

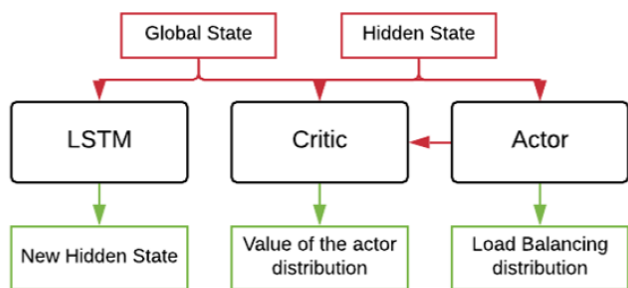
## MegaMIMO 2.0 vs. Traditional 802.11



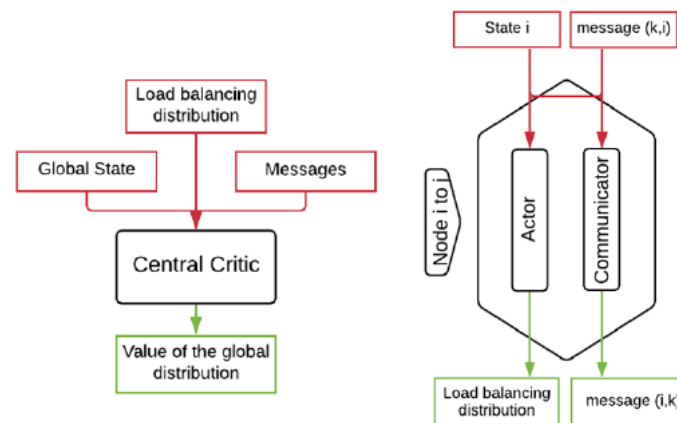


# Методы машинного обучения в задаче балансировки трафика

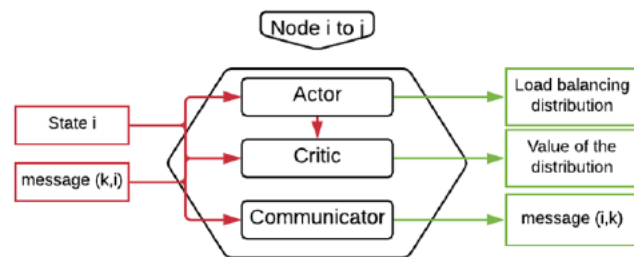
- Наиболее распространенная модель – Actor-Critic
- Работа балансировщика определяется состоянием, действием и наградой (методы поиска оптимальной политики)



Централизованный подход



Полуцентрализованный подход



Распределенный подход

# На каком уровне проводить балансировку нагрузки?

- На уровне каналов
  - TRILL, Wireless MIMO
- На уровне маршрутизаторов
  - ECMP, VLB, etc.
- На уровне транспорта
  - SCTP
  - MPTCP
- На уровне приложений
  - BitTorrent

# Программа курса

## Подходы:

- 1. Управление перегрузкой**
  - Современные протоколы управления перегрузкой TCP
- 2. Демультимплексирование/мультиплексирование**
  - Многопоточные транспортные протоколы
  - Маршрутизация на уровне интернет провайдеров
  - Network Coding
- 3. Сегментация**
  - TCP Proxy
- 4. Балансировка**
  - Балансировка нагрузки и управление трафиком
- 5. Преобразование сообщений**
  - FEC
  - Сжатие

## Модели оценки качества сервиса:

- Активные и пассивные измерения
- NS3: моделирование поведения сети с высокой точностью
- Сетевое исчисление: математический подход к качеству сервиса

## Примеры:

- Управление сетевыми ресурсами в Центрах Обработки Данных
- Обеспечение качества сервиса в сетях доставки контента
- Пропускная способность по требованию