

Модели оценки качества сервиса

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

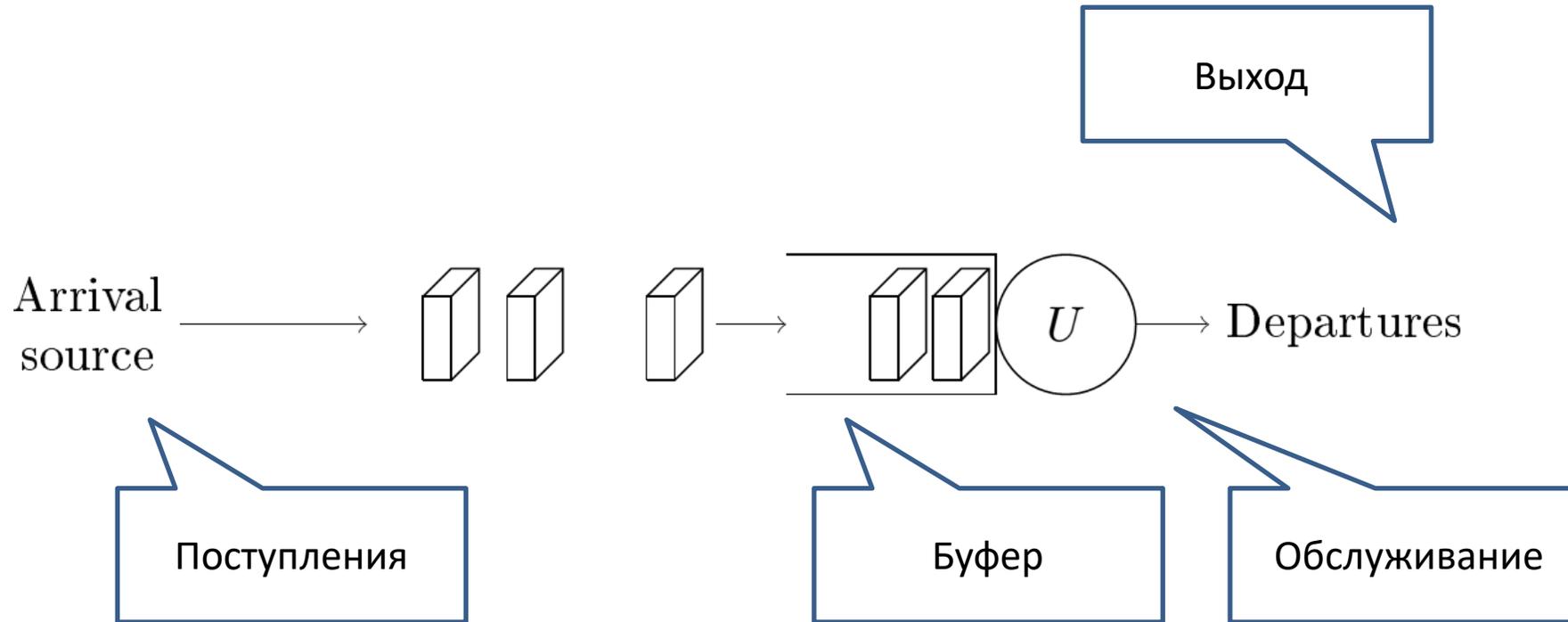
Примеры:

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

Очереди



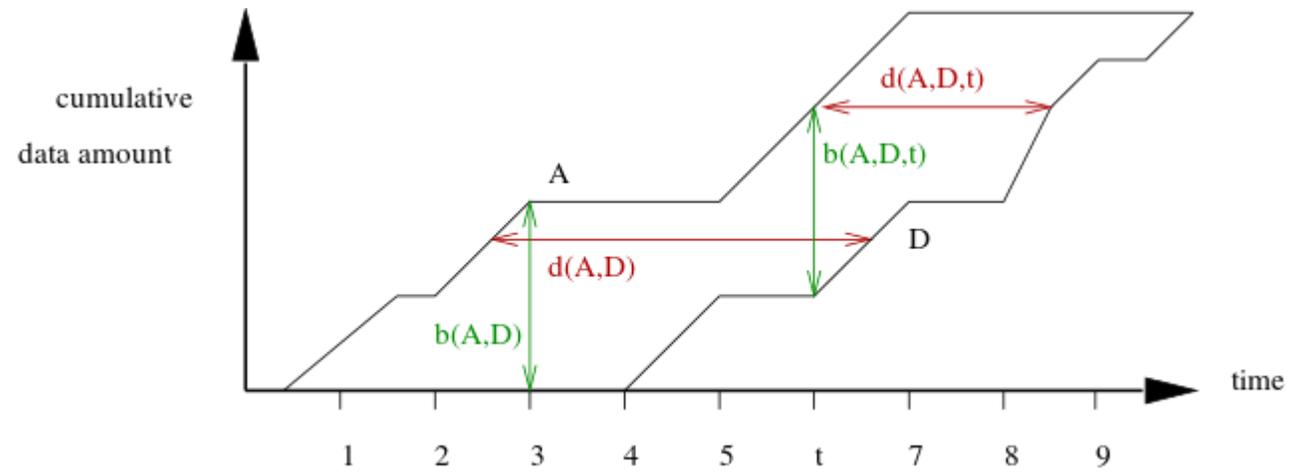
Очереди



Анализ очередей

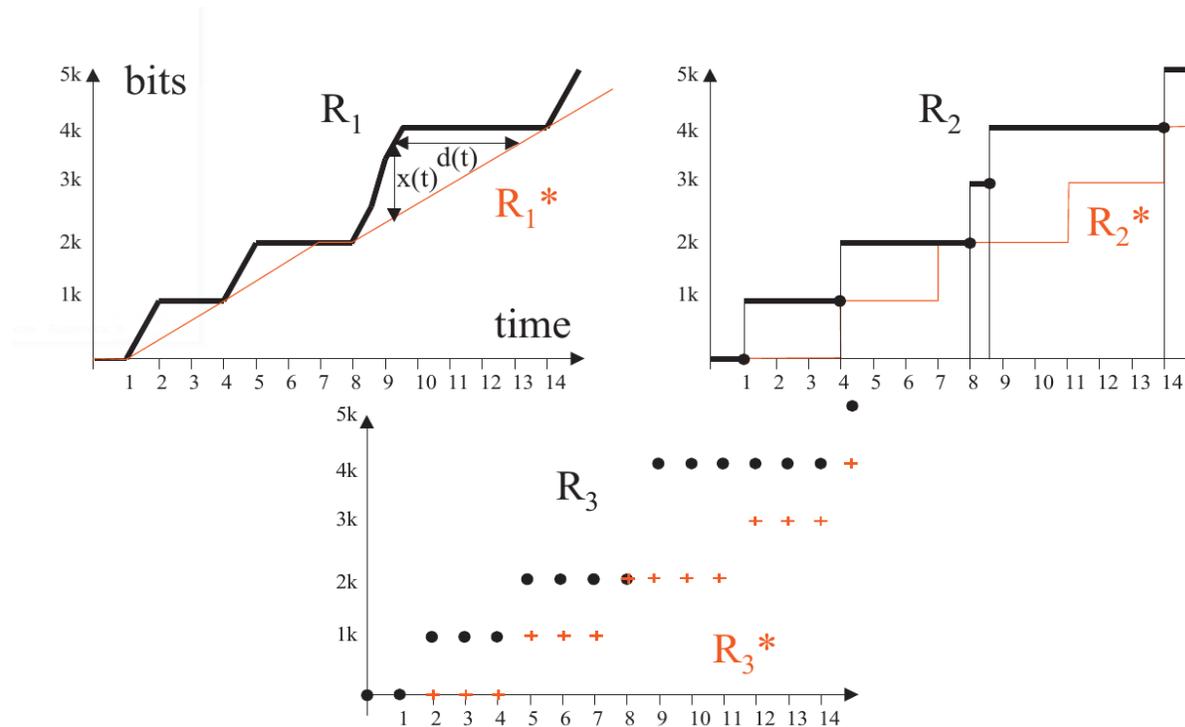
- Erlang – теория очередей, 1909
 - The Theory of Probabilities and Telephone Conversations *Nyt Tidsskrift for Matematik*, Vol. 20, No. B. (1909), pp. 33-39 by Agner K. Erlang
- Cruz – Сетевое исчисление, 1991
 - [1] *R. L. Cruz. A calculus for network delay, Part I: Network elements in isolation. IEEE Transactions on Information Theory, 37(1):114-131, January 1991.*
 - [2] *R. L. Cruz. A calculus for network delay, Part II: Network analysis. IEEE Transactions on Information Theory, 37(1):132-141, January 1991.*

Детерминированное сетевое исчисление



Модели потока данных

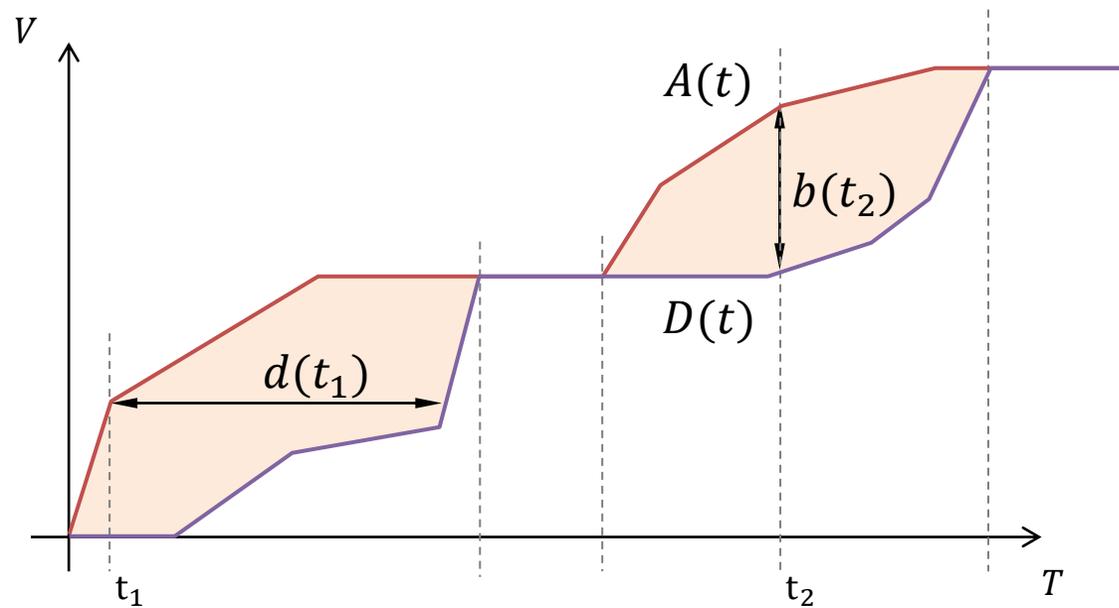
1. С дискретным временем (например, $t = 0, 1, 2, 3, \dots$).
2. С непрерывным временем, модель жидкости (R – непрерывная)
3. С непрерывным временем, общая модель (R – непрерывная справа или слева)



Основные определения

- **Функция прибытия** $A(t)$ описывает зависимость суммарного количества данных, поступивших на обработчик от времени
- **Функция отправки** $D(t)$ – зависимость количества переданных данных потока от времени
- Каждый обработчик может быть описан перечислением пар вида $\langle A(t), D(t) \rangle$
- **Отставание (backlog)** $b(t)$ – выражает количество данных, находящихся *внутри* обработчика
- **Период отставания** – промежуток в течение которого функция отставания строго положительна
- **Задержка (delay)** $d(t)$ – время прохождения через обработчик той порции данных, которая поступила на него в момент времени t

Функции поступления и отправки



Кривые нагрузки и сервиса

- Вид функций поступления и отправки в большинстве случаев неизвестен
 - Необходима аппроксимация
- **Кривая нагрузки** $\alpha(t)$ -- накопительная функция, для которой выполнено условие

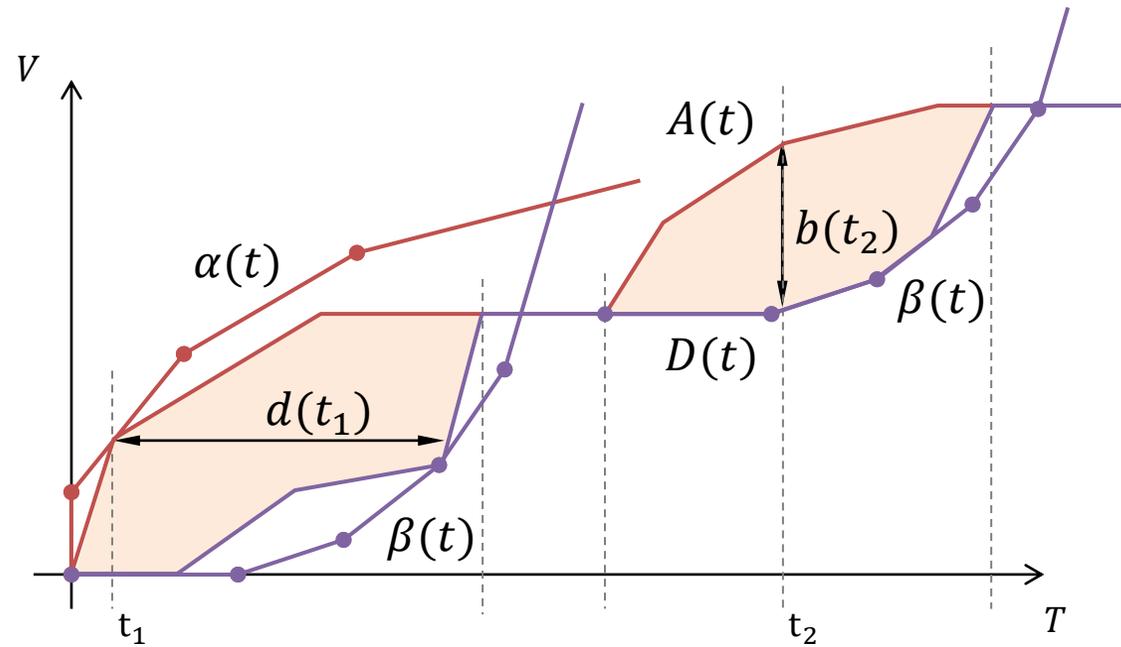
$$\forall t, \tau: A(t + \tau) - A(t) \leq \alpha(\tau)$$

- за время τ поступает не больше $\alpha(\tau)$ данных
- **Кривая сервиса** $\beta(t)$ – накопительная функция, для которой внутри каждого периода оставание выполняется условие

$$\forall t, \tau: D(t + \tau) - D(t) \geq \beta(\tau)$$

- за время τ обслуживается не меньше $\beta(\tau)$ данных

Кривые нагрузки и сервиса



Min-plus алгебра

- Обычная алгебра $R, +, *$
 - Поле
 - $a * (b + c) = a * b + a * c$
- Min-plus алгебра $R \cup \{+\infty\}, \wedge (\text{inf, min}), +$
 - Коммутативное идемпотентное полукольцо (диоид)
 - $a + (b \wedge c) = (a + b) \wedge (a + c)$

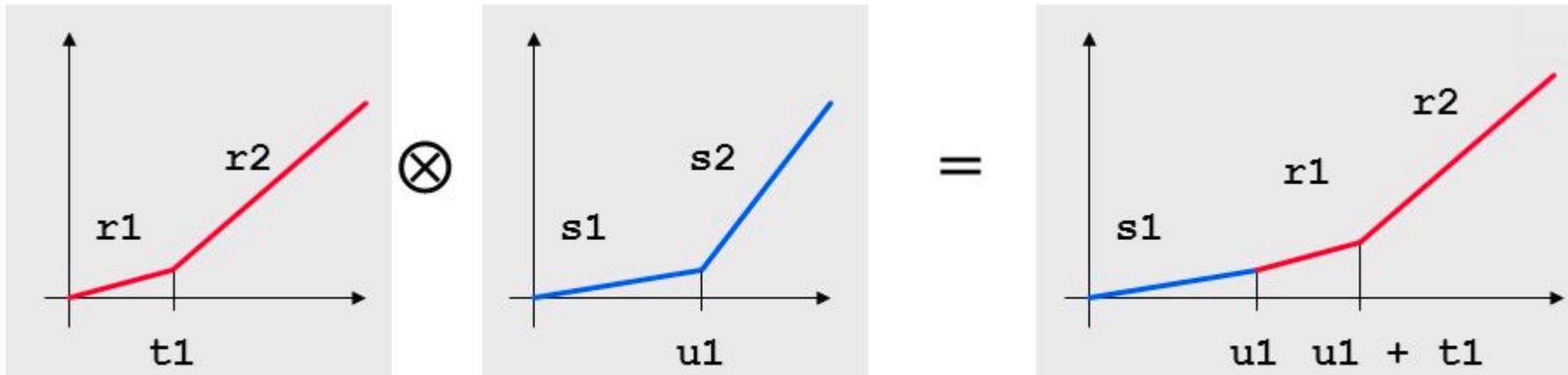
Min-plus свертка

- Обычная свертка

$$- (f \otimes g)(t) = \int_0^t f(t-s)g(s)ds$$

- Min-plus свертка

$$- (f \otimes g)(t) = \inf_{0 \leq s \leq t} \{f(t-s) + g(s)\}$$

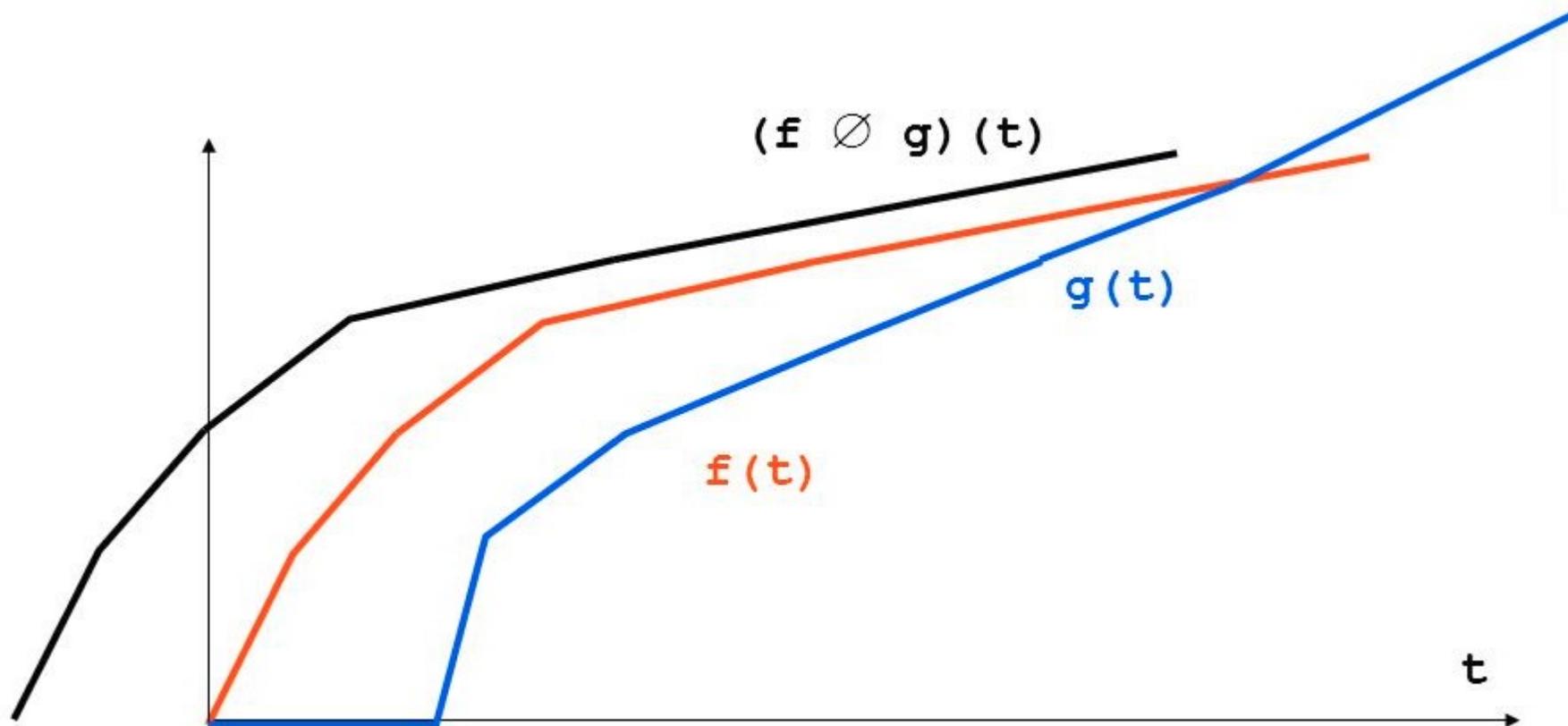


Min-plus свертка

- *Выпуклые функции*: если f и g выпуклые функции, тогда $f \otimes g$ тоже выпуклая. В частности, если f и g – выпуклые и кусочно-линейные функции, то $f \otimes g$ может быть получена соединением концов различных линейных кусков функций f и g в порядке увеличения наклона.

Обратная min-plus свертка

- $(f \oslash g)(t) = \sup_{u \geq 0} \{f(t+u) - g(u)\}$.



Полученные оценки

Оценка отставания:

$$b(t) \leq v(\alpha, \beta) = (\alpha \oslash \beta)(0)$$

Оценка задержки (FIFO):

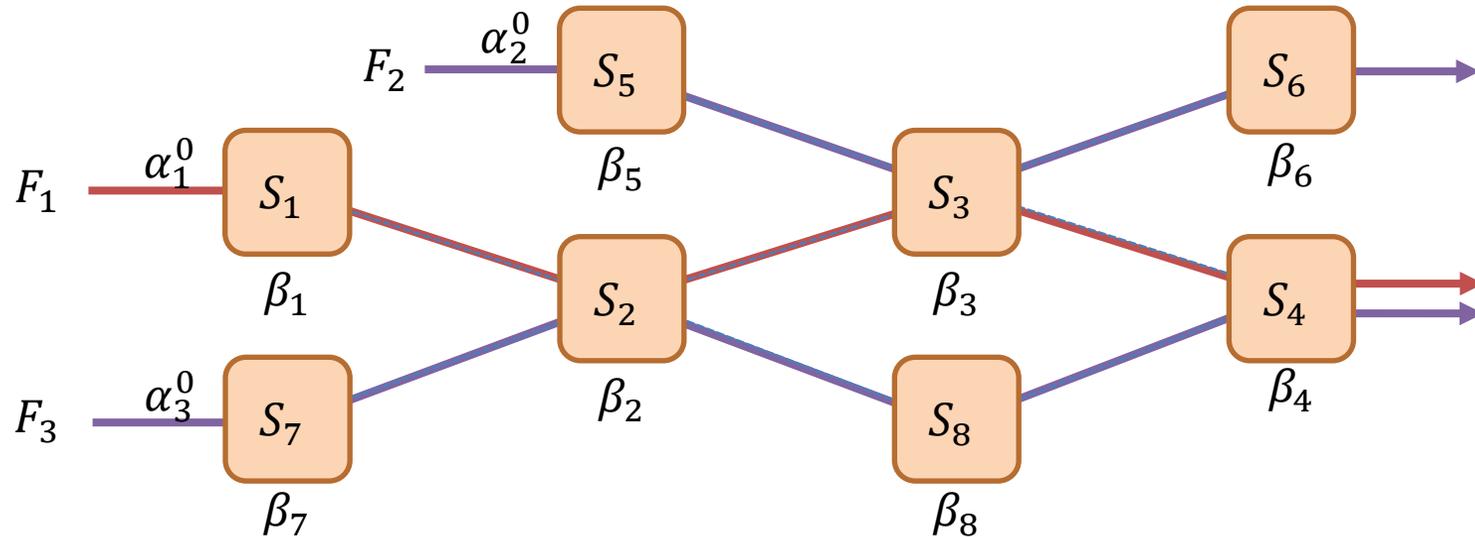
$$d(t) \leq h(\alpha, \beta) = \sup_t \{ \inf \{ \tau \geq 0 \mid \alpha(t) \leq \beta(t + \tau) \} \}$$

Композиция обработчиков:

Если обработчики S_1 и S_2 с кривыми сервиса β_1 и β_2 образуют **тандем**, то их систему описывает кривая сервиса $\beta = \beta_1 \otimes \beta_2$

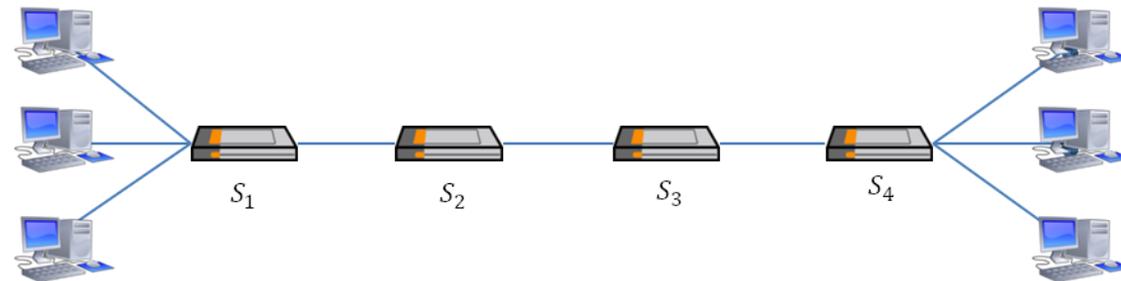
Применение Network Calculus

- Separated Flow Analysis
- Линейное программирование

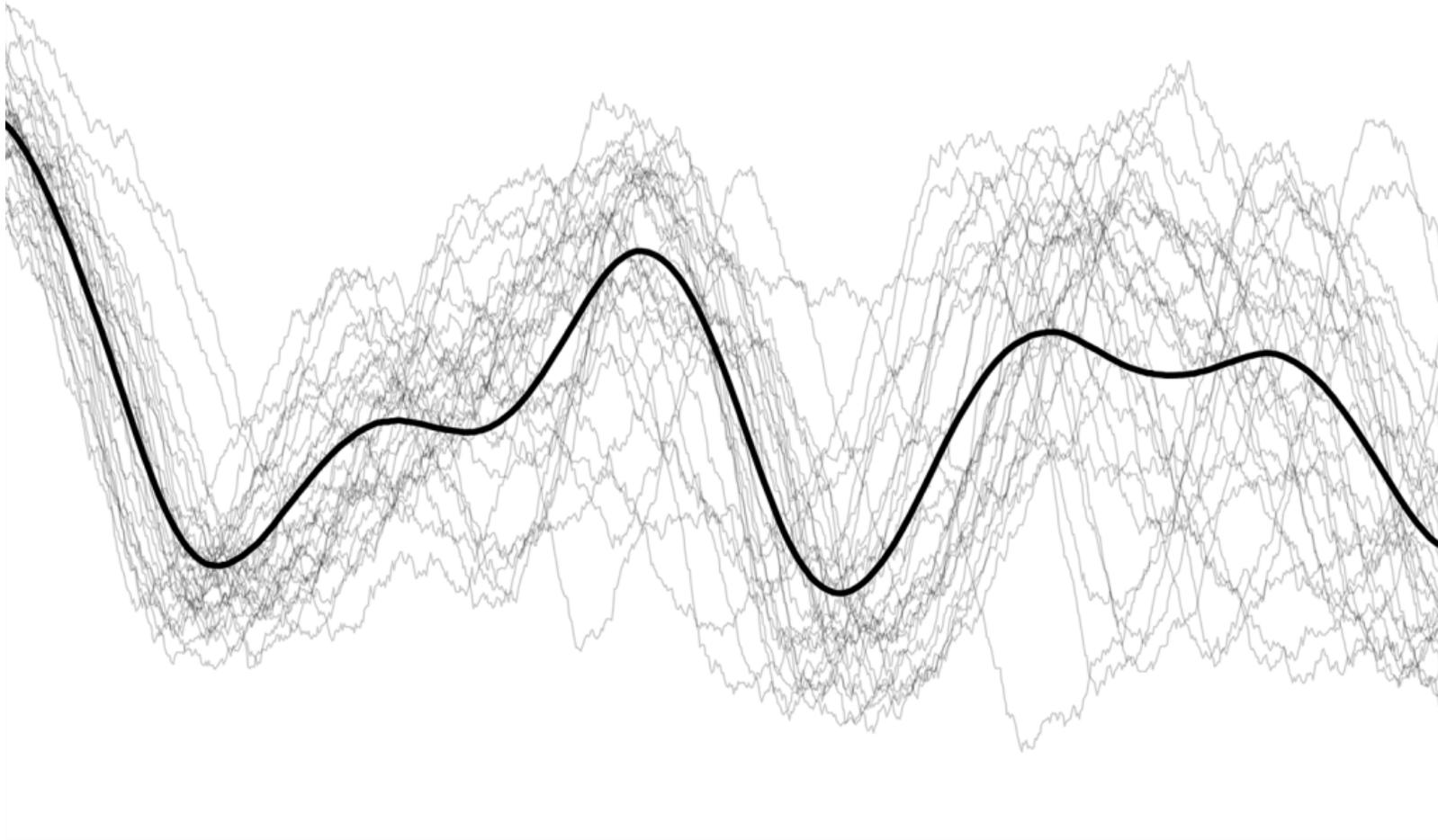


Применение Network Calculus

Длина	Аудио			Видео			Данные		
	NS3	SF	LP	NS3	SF	LP	NS3	SF	LP
2	159	12348	12348	244	11881	11881	244	9204	9204
3	281	16053	12519	459	23681	12045	366	18167	9334
4	403	23123	12689	488	35640	12209	515	27231	9457
5	525	33558	12860	692	47877	12373	639	36486	9584
6	647	47359	13031	733	60510	12537	733	46022	9711
7	769	64527	13201	891	73662	12701	962	55934	9838
8	891	85063	13372	998	87459	12866	977	66316	9964
9	1013	108972	13543	1099	102031	13030	1099	77270	10091



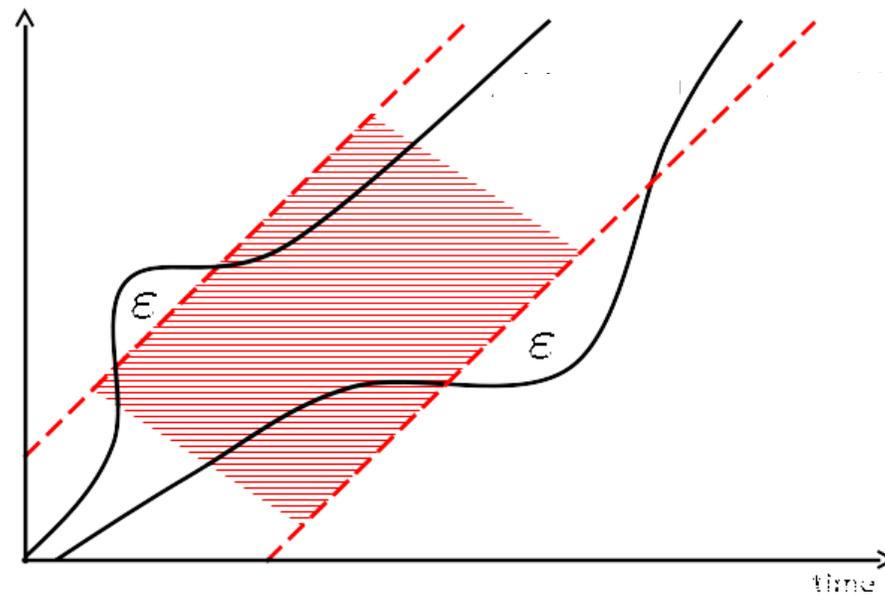
Стохастическое сетевое исчисление



Общая идея

Позволить нарушения кривых нагрузки и сервиса с некоторой вероятностью

-> теряем точность результата



Основные принципы

- Задать распределение для поступающего трафика и обслуживания обработчиков
- Получить афинную оценку
- Связать афинную оценку с погрешностью оценки
- Сохраняем подходы из детерминированного сетевого исчисления

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