Основы построения телекоммуникационных систем и сетей

Лекция №14
«Методы теории очередей»

профессор Соколов Н.А.

Общие положения

В начале XX века стали активно развиваться телефонные сети. Возникли новые задачи планирования этих сетей. Одна из первых задач заключалась в расчете емкости пучка каналов при заданной вероятности потерь. А.К. Эрланг вывел формулу, позволившую решить эту задачу. Данная формула была приведена в девятой лекции. Считается, что именно с работ Эрланга началось развитие теории телетрафика. Единице трафика в 1946 году решением МСЭ было присвоено название "Эрланг". Первые системы коммутации работали по алгоритму с потерями. Это означает, что при отсутствии свободного обслуживающего прибора заявка теряется. Использование программного управления позволило ввести дисциплину обслуживания с ожиданием. Это увеличило эффективность обслуживания заявок. Широкое применение данного алгоритма обслуживания привело к тому, что вместо словосочетания "Теория телетрафика" стало чаще использоваться название "Теория очередей". В настоящее время "Теория очередей" широко используется для исследования телекоммуникационных сетей, транспортных систем, сферы торговли.

Классификация (1)

В 1961 году Кендалл (D.G. Kendall) ввел следующее обозначение для систем массового обслуживания: "$A/B/n$". Символ "A" определяет процесс поступления заявок, обозначение "B" указывает на распределение времени занятия, а величина n равна числу обслуживающих приборов. Для более полного описания систем телетрафика позже было введено расширенное обозначение Кендалла:

$A/B/n/K/S/X$

где:

- K – количество мест для ожидания в очереди,
- S – число обслуживаемых абонентов,
- X – дисциплина обработки заявок.

В первой позиции классификации чаще всего стоит символ M. Это означает, что входящий поток является пуассоновским. Для более сложных моделей используются символы GI (произвольный рекуррентный закон) и G (произвольный закон с возможной корреляцией).

Во второй позиции обычно используется один из следующих символов: M (экспоненциальное распределение), D (постоянное время обслуживания), E_k (распределение Эрланга $k-го$ порядка) G (произвольный закон распределения). Реже встречаются другие символы.

Классификация (2)

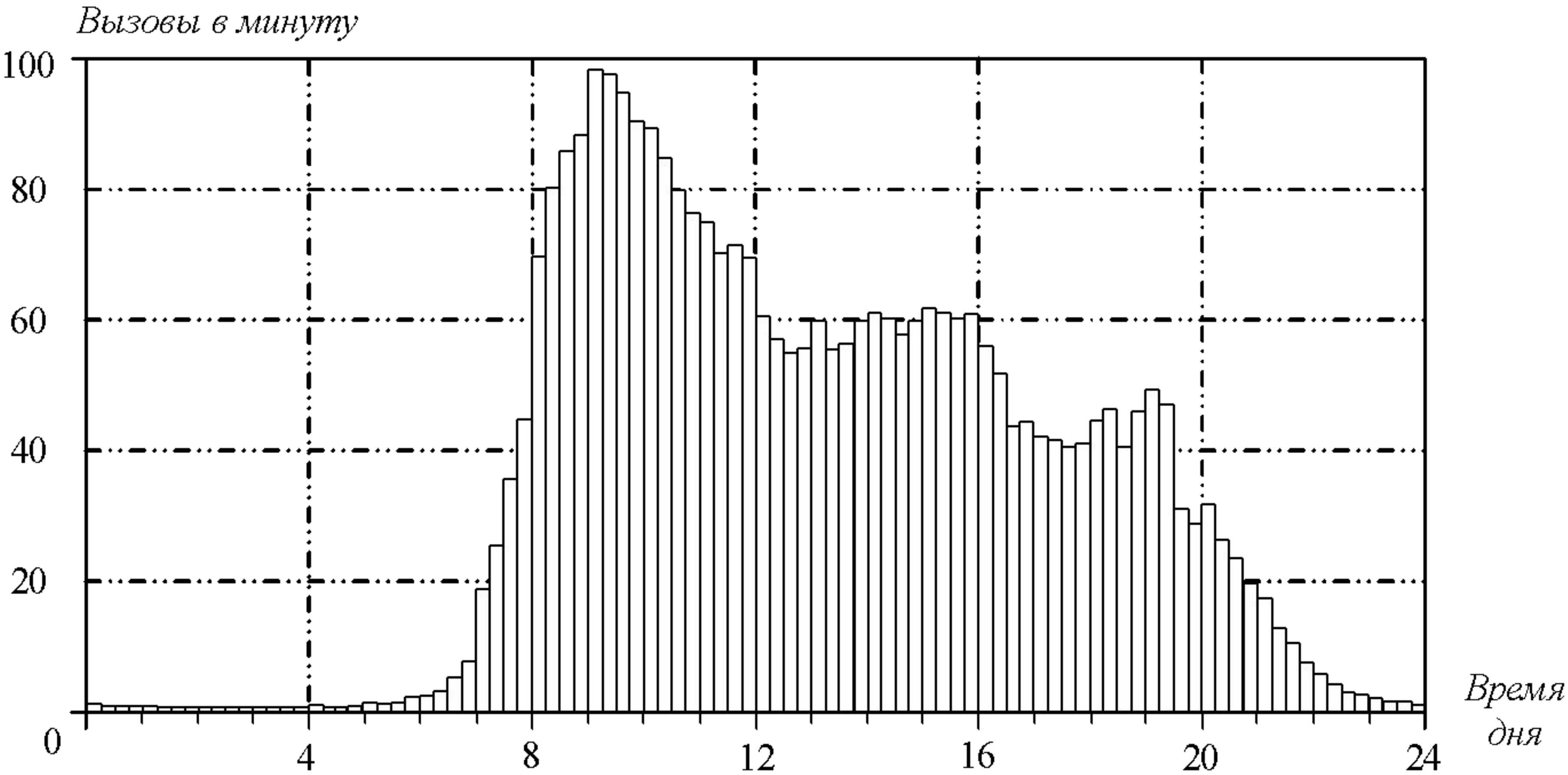
Классификация алгоритмов обслуживания была приведена в девятой лекции. Обычно используются такие дисциплины:

- FCFS: first come – first served (первым пришел – первым обслужен),
- LCFS: last come – first served (последним пришел – первым обслужен),
- SIRO: service in random order (обслуживание в случайной порядке),
- SJF: shortest job first (обслуживание, начиная с "коротких" заявок).

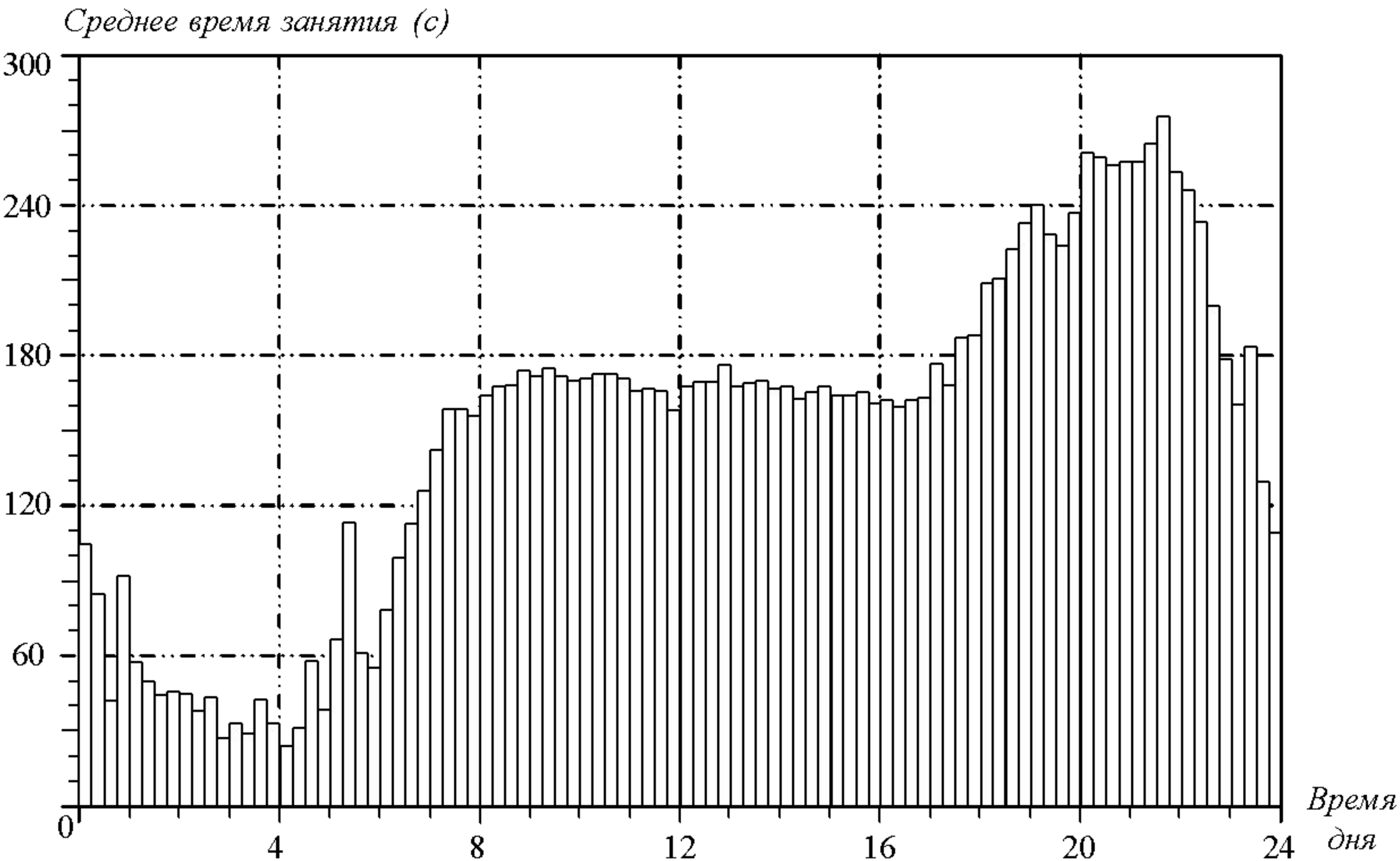
В некоторых случаях вводятся приоритеты. Существует принципиальное различие между двумя видами приоритетов: без прерывания и с прерыванием обслуживаемого требования (non-preemptive and preemptive). Для первой дисциплины поступившая заявка ждет окончания обслуживания менее приоритетного требования. Она будет обслужена сразу после окончания обработки менее приоритетной заявки. При использовании второй дисциплины обслуживание заявки с низшим уровнем приоритета прерывается. Обычно выделяют три способа обслуживания прерванной заявки:

- дообслуживание с момента приостановки менее приоритетного требования,
- обслуживание прерванной заявки заново,
- потеря прерванной заявки.

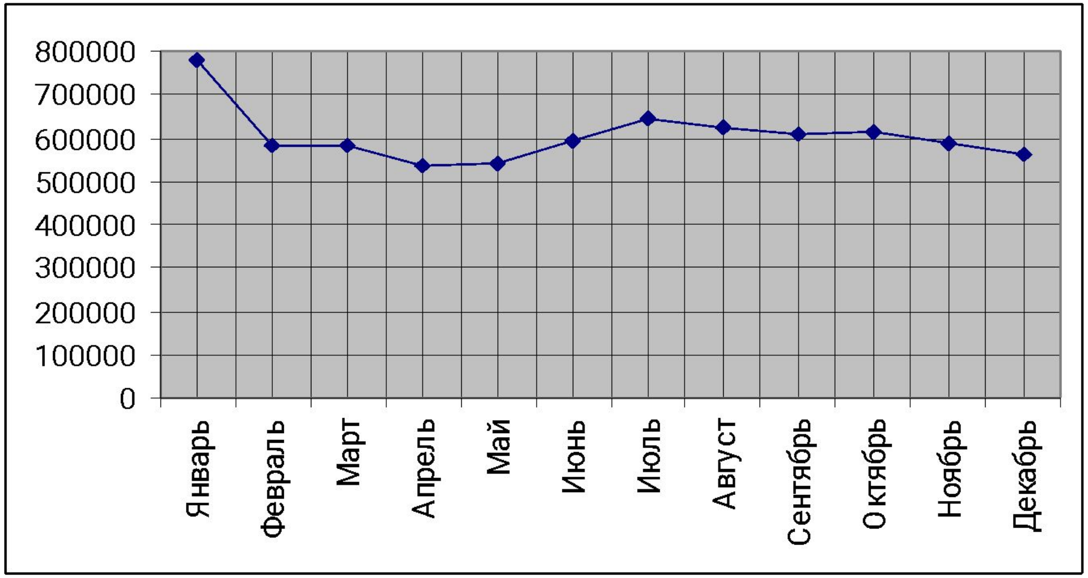
Входящий поток заявок



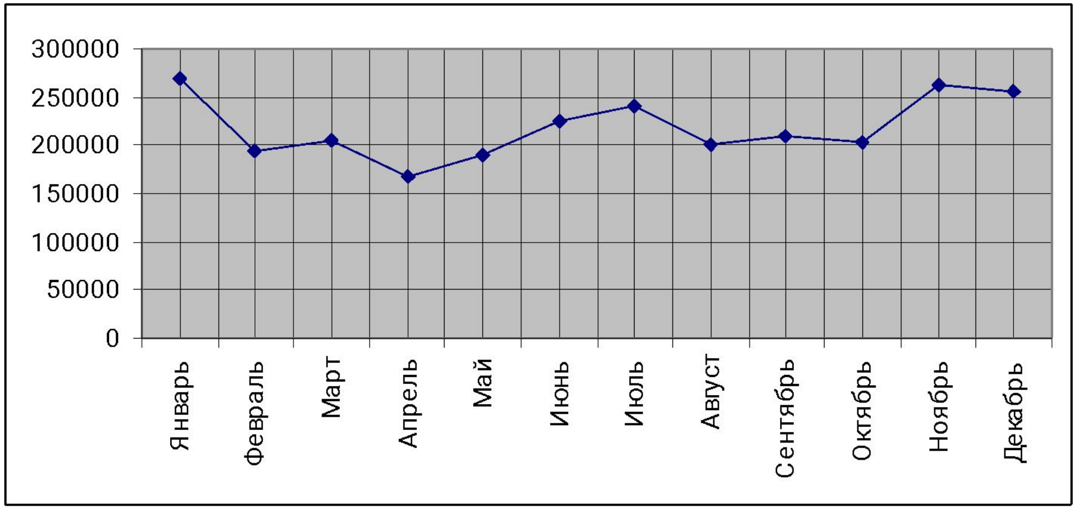
Время обслуживания заявок



Количество обращений в службу «09»

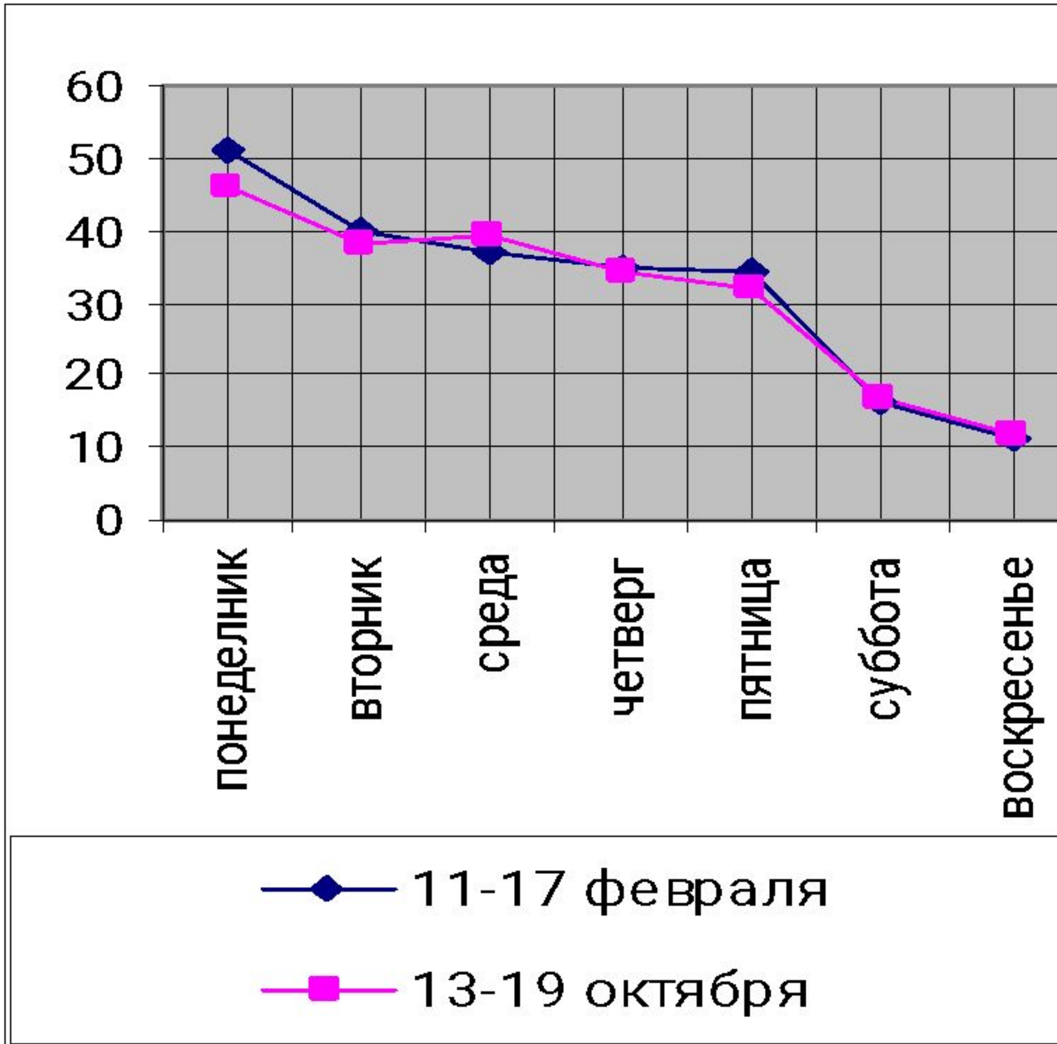


Город «А»



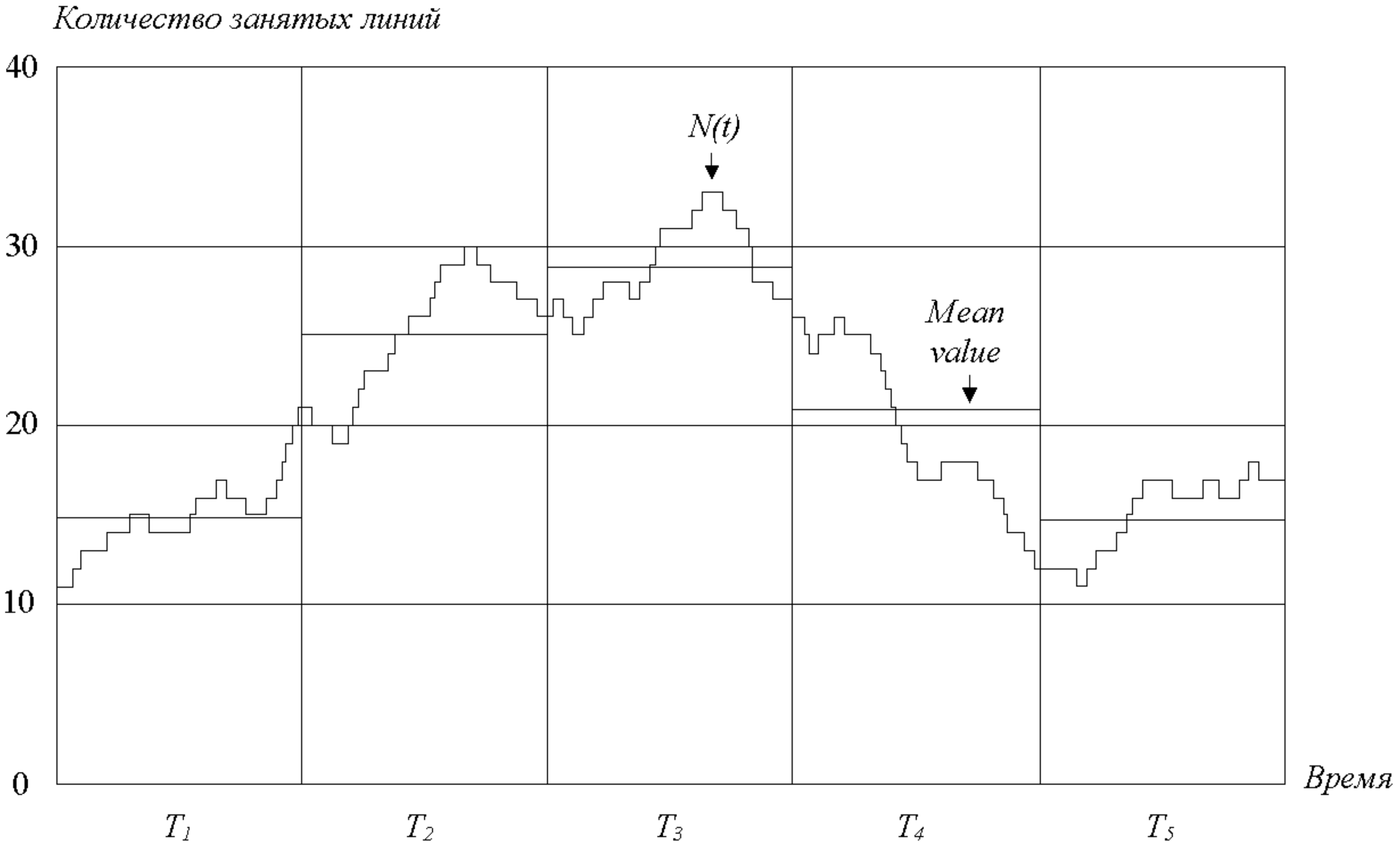
Город «В»

Количество обращений за неделю

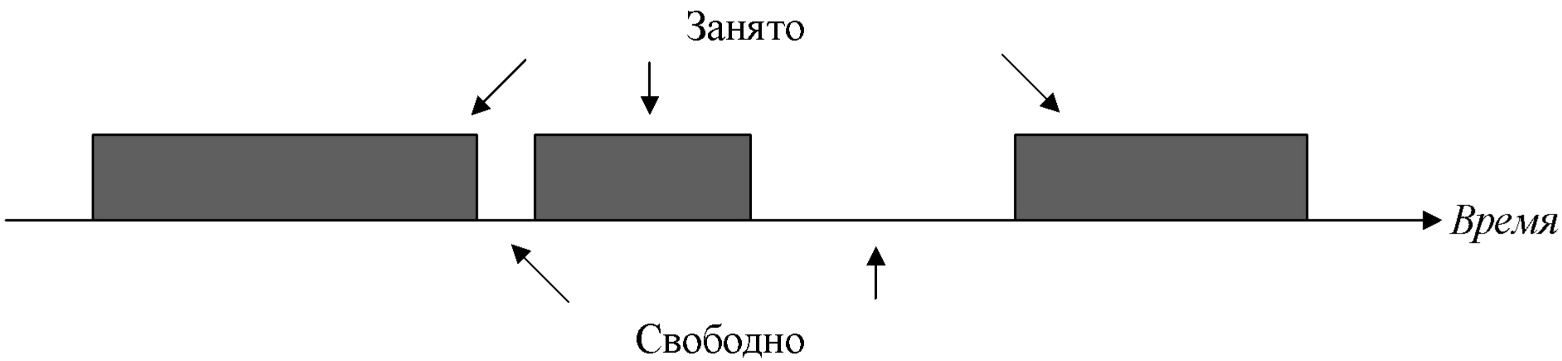


Трафик справочной службы «09»

Занятие и освобождение линий



Два состояния линии



Для планирования телекоммуникационной сети очень важными характеристиками систем с очередями являются:

- среднее время задержки,
- квантиль функции распределения времени задержки.

Именно эти два показателя нормируются в рекомендациях ITU-T и в стандартах ETSI. Для решения иных задач, не входящих в процесс планирования сети, представляют интерес и другие характеристики queueing system.

Примеры вычислений (1)

Среднее значение длительности задержки определяется такой суммой:

$$S^{(1)} = W^{(1)} + B^{(1)}.$$

Первое слагаемое – среднее значение длительности ожидания начала обслуживания заявок. Второе слагаемое – среднее значение длительности обслуживания заявок. Очевидно, что основные сложности связаны с расчетом величины $W^{(1)}$, то есть первого слагаемого.

Преобразование Лапласа-Стилтьеса функции распределения длительности задержки заявок вычисляется следующим образом:

$$S^{*}(s) = W^{*}(s)B^{*}(s).$$

В этой формуле основные сложности связаны с первым сомножителем, который представляет собой преобразование Лапласа-Стилтьеса функции распределения длительности ожидания начала обслуживания заявок.

Примеры вычислений (2)

Для вычисления $W^{(1)}$ в системах $M/G/1$ была получена формула Поллячека-Хинчина:

$$W^{(1)} = \frac{\lambda B^{(2)}}{2(1-\rho)}$$

где:

- λ – интенсивность входящего трафика,
- $B^{(2)}$ – второй момент времени занятия,
- ρ – нагрузка системы.

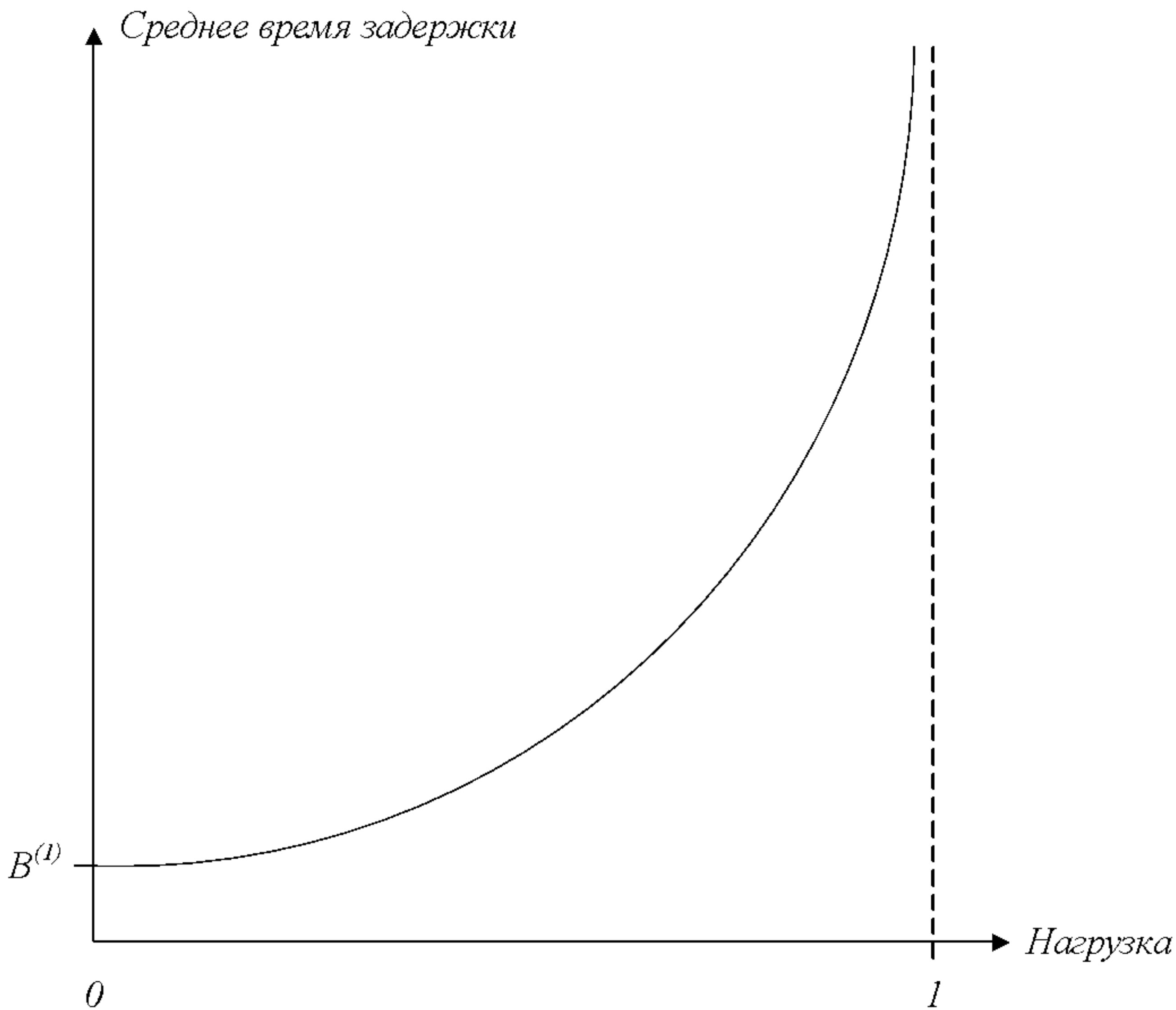
Величина $B^{(2)}$ может быть вычислена по первому моменту $B^{(1)}$ и соответствующему коэффициенту вариации C_B:

$$B^{(2)} = B^{(1)}\left(1 + C_B^2\right).$$

Нагрузка системы равна $\lambda B^{(1)}$. Иногда вместо величины $B^{(1)}$ используется интенсивность обслуживания μ. Она связана с $B^{(1)}$ простым соотношением: $B^{(1)} = \mu^{-1}$. Учитывая формулу (14.12), получаем:

$$W^{(1)} = \frac{\rho\left(1 + C_B^2\right)}{2(1-\rho)} B^{(1)}.$$

Задержка как функция нагрузки



Оценка квантиля (1)

Оценка квантиля подразумевает получение выражения для расчета функции распределения $S(t)$. Для системы $M/M/1$ искомое выражение имеет такой вид:

$$S(t) = 1 - e^{-(1-\rho)\mu t}.$$

Для модели $M/D/1$ функция распределения длительности ожидания была получена Кроммелином (C.D. Crommelin). Обычно, результаты расчета функции $W(t)$ представляются в графической форме. Они известны как кривые Кроммелина. Формула для вычисления $W(t)$ при $B^{(1)} = \tau$ представима в следующей форме:

$$W(t) = (1-\lambda\tau)\sum_{k=0}^{\left[\frac{t}{\tau}\right]} \frac{[\lambda(k\tau - t)]^k}{k!} e^{\lambda(t-k\tau)}.$$

Квадратные скобки над знаком суммы указывают на тот факт, что определяется целая часть от результата деления t на τ. Функция $S(t)$ для модели $M/D/1$ определяется очевидным соотношением:

$$S(t) = \begin{cases} 0, & npu\ t < \tau \\ W(t-t), & npu\ t \geq \tau \end{cases}.$$

Оценка квантиля (2)

Для системы $M/G/1$ было получено уравнение Поллячека-Хинчина, позволяющее определить преобразование Лапласа-Стилтьеса для функций $W(t)$ и $S(t)$:

$$W^*(s) = \frac{s(1-\rho)}{s-\lambda+\lambda B^*(s)}, \quad S^*(s) = \frac{s(1-\rho)}{s-\lambda+\lambda B^*(s)} B^*(s).$$

Эти выражения позволяют найти любые моменты длительности ожидания и задержки заявок. Получить соответствующие функции распределения для большинства моделей очень сложно.

Практический интерес связан с определением характеристик систем с очередями, состоящих из нескольких фаз обслуживания. В силу аддитивности математического ожидания среднее значение суммарной задержки $S_T^{(1)}$ заявок, проходящих через N фаз обслуживания, определяется следующим образом:

$$S_T^{(1)} = \sum_{j=1}^{N} S_j^{(1)}.$$

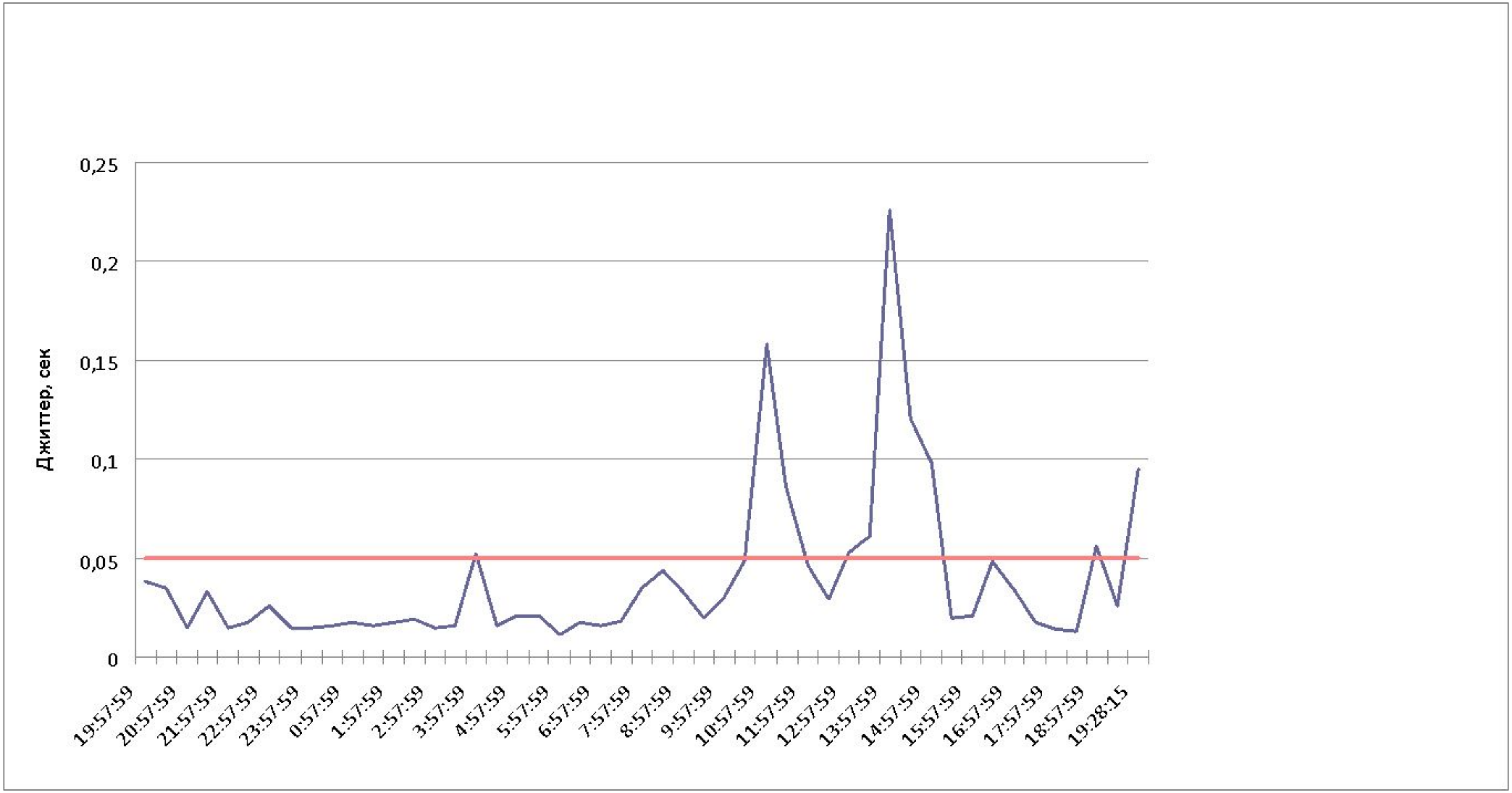
Оценка квантиля (3)

Величина $S_j^{(1)}$ – среднее значение времени задержки заявок на $j-ой$ фазе обслуживания. Предполагается, что времена задержки заявок на всех фазах являются взаимно независимыми случайными величинами. Для этого же предположения суммарная функция распределения времени задержки заявок $S_T(t)$ определяется через преобразования Лапласа-Стилтьеса $S_T^*(s)$:

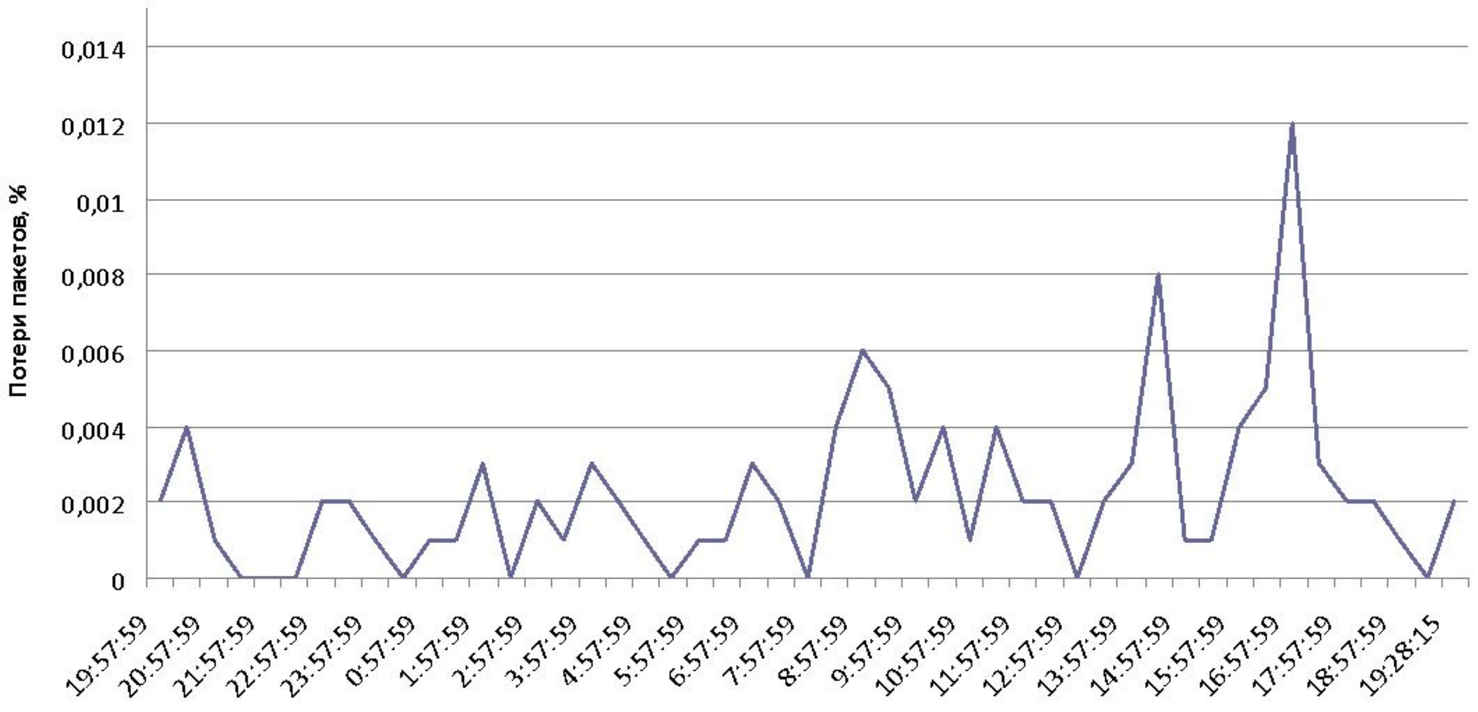
$$S_T^*(s) = \prod_{J=1}^{N} S_j^*(s)\,.$$

В этой формуле $S_j^*(s)$ – преобразование Лапласа-Стилтьеса функции распределения времени задержки заявок на $j-ой$ фазе обслуживания.

Измеренные значения QoS (1)



Измеренные значения QoS (2)



Вопросы?