Лекция 8. Методы обеспечения качества обслуживания: анализ очередей

Дисциплина: Сетевые технологии

Цель лекции

Изучить основные понятия теории массового обслуживания и рассмотреть применение модели M/M/1 для анализа обработки пакетов в компьютерных сетях.

1. Понятие и назначение анализа очередей

В компьютерных сетях пакеты данных поступают на сетевые устройства и ожидают обработки. Если пакеты поступают быстрее, чем устройство может их передавать, образуется очередь.

Очереди возникают в маршрутизаторах, коммутаторах, шлюзах и других сетевых устройствах. Они позволяют временно хранить пакеты до момента их передачи через выходной интерфейс.

Например, несколько компьютеров одновременно передают данные через один маршрутизатор. Все пакеты поступают на один выходной интерфейс. Если пропускной способности интерфейса недостаточно для немедленной передачи всех пакетов, часть из них помещается в очередь.

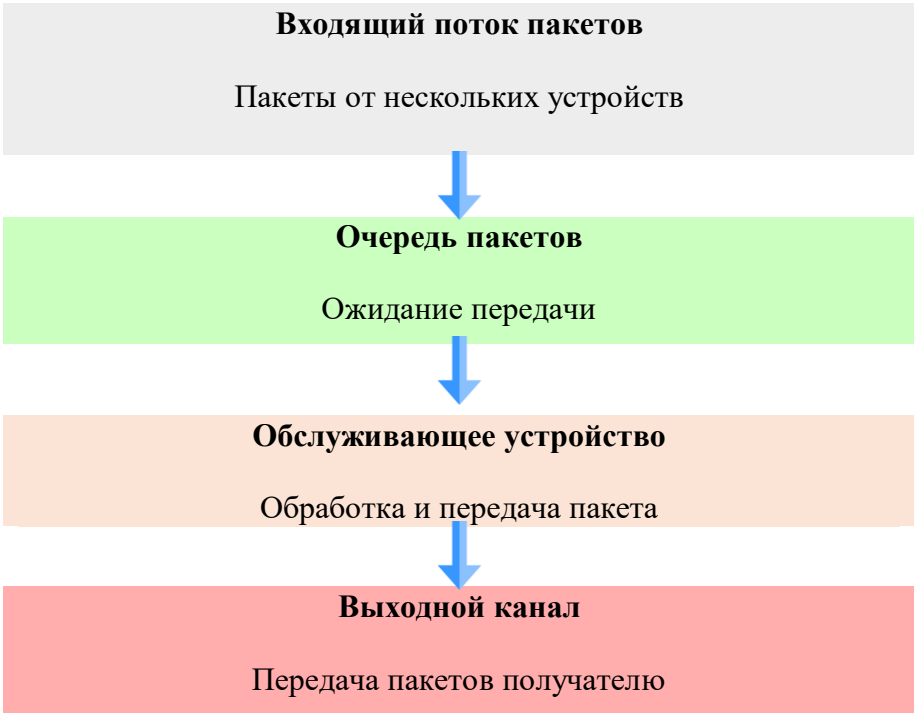


Рисунок 1. Общая схема обработки пакетов с очередью

Анализ очередей позволяет определить:

- среднее количество пакетов в очереди;

- среднее время ожидания пакета;
- среднее время пребывания пакета в системе;
- вероятность того, что устройство занято;
- вероятность перегрузки системы.

Эти показатели используются для оценки производительности сети и качества обслуживания (QoS).

2. Основные элементы системы массового обслуживания

Система массового обслуживания представляет собой модель, в которой поступающие заявки ожидают обработки и получают обслуживание от одного или нескольких устройств.

В компьютерных сетях заявками могут выступать пакеты данных, а обслуживающим устройством может быть маршрутизатор или выходной интерфейс коммутатора.

Основные элементы системы:

1. Источник заявок. Устройства или приложения, которые передают пакеты в сеть.

2. Входящий поток. Последовательность пакетов, поступающих на обработку.

3. Очередь. Место временного хранения пакетов до начала обслуживания.

4. Обслуживающее устройство. Устройство или канал, который обрабатывает и передаёт пакеты.

5. Выходной поток. Пакеты, которые прошли обслуживание и покинули систему.

Для описания таких систем применяются математические модели. Одна из наиболее известных моделей называется M/M/1.

3. Знакомство с моделью M/M/1

Модель M/M/1 относится к теории массового обслуживания. Она описывает систему с одним обслуживающим устройством, случайным поступлением заявок и случайным временем обслуживания.

Обозначение M/M/1 состоит из трёх частей:

Обозначение	Значение
Первая M	Поступление заявок по пуассоновскому потоку
Вторая M	Экспоненциальное распределение времени обслуживания
1	Одно обслуживающее устройство

Буква M происходит от слова Markovian, которое связано с марковскими свойствами модели.

Основные предположения модели

Модель M/M/1 использует следующие условия:

- пакеты поступают независимо друг от друга в соответствии с пуассоновским потоком;
- время обслуживания каждого пакета имеет экспоненциальное распределение;
- система содержит одно обслуживающее устройство;
- пакеты обслуживаются по принципу FIFO, то есть в порядке поступления;
- очередь имеет неограниченную ёмкость в классической модели;
- количество поступающих пакетов и время обслуживания не зависят друг от друга.

В реальных сетях эти условия выполняются не всегда. Поэтому M/M/1 используется прежде всего для упрощённого анализа и оценки поведения очереди.

4. Основные параметры и формулы модели M/M/1

Для расчёта характеристик очереди необходимо определить два основных параметра.

Интенсивность поступления пакетов **λ** показывает среднее количество пакетов, поступающих в систему за единицу времени.

Интенсивность обслуживания **μ** показывает среднее количество пакетов, которые обслуживающее устройство может обработать за единицу времени.

Обычно оба параметра измеряются в пакетах в секунду.

Например, если за одну секунду в маршрутизатор поступает в среднем 80 пакетов, то:

λ =80 пакетов/с

Если маршрутизатор может обслужить 100 пакетов за секунду, то:

μ=100 пакетов/с

4.1. Коэффициент загрузки системы

Коэффициент загрузки показывает, насколько интенсивно используется обслуживающее устройство.

$$\rho = \frac{\lambda}{\mu}$$

где:

- ρ — коэффициент загрузки;
- λ— интенсивность поступления;

- μ— интенсивность обслуживания.

Для устойчивой работы модели M/M/1 необходимо условие:

$$\lambda < \mu, \ \rho<1$$

Если λ приближается к μ, очередь в среднем становится длиннее, а время ожидания увеличивается.

Если λ≥μ, классическая модель с неограниченной очередью не имеет устойчивого стационарного режима.

4.2. Вероятность отсутствия пакетов в системе

Вероятность того, что система пуста, определяется формулой:

$$P_0 = 1 - \rho$$

Например, если коэффициент загрузки равен 0,8, вероятность отсутствия пакетов в системе составляет 0,2, или 20%.

4.3. Среднее количество пакетов в системе

Среднее количество пакетов в системе включает пакеты в очереди и пакет, который в данный момент обслуживается.

$$L = \frac{\rho}{1 - \rho}$$

4.4. Среднее количество пакетов в очереди

Среднее количество пакетов, которые ожидают обслуживания:

$$L = \frac{\rho^2}{1 - \rho}$$

Этот показатель помогает оценить, насколько сильно загружена очередь сетевого устройства.

4.5. Среднее время пребывания пакета в системе

Время пребывания включает ожидание в очереди и обслуживание.

$$W = \frac{\rho}{\mu - \lambda}$$

4.6. Среднее время ожидания в очереди

Среднее время ожидания до начала обслуживания определяется формулой:

$$W_q = \frac{\rho}{\mu - \lambda}$$

Также по закону Литтла выполняются соотношения:

$$L = \lambda W$$

$$L_q = \lambda W_q$$

Все эти формулы применяются при условии устойчивой работы системы, когда λ<μ.

5. Модель M/M/1 как модель обработки пакетов

В компьютерной сети модель M/M/1 можно использовать для описания выходного интерфейса маршрутизатора.

Предположим, что пакеты поступают от нескольких устройств и направляются через один выходной канал. Маршрутизатор обрабатывает пакеты по одному в соответствии с заданной дисциплиной обслуживания.

В такой модели:

- поступление пакетов соответствует входящему трафику;
- очередь соответствует буферу выходного интерфейса;
- обслуживание соответствует передаче пакетов через канал;
- время ожидания отражает задержку, вызванную очередью;
- время пребывания включает ожидание и обслуживание.

Пример модели M/M/1 в сети

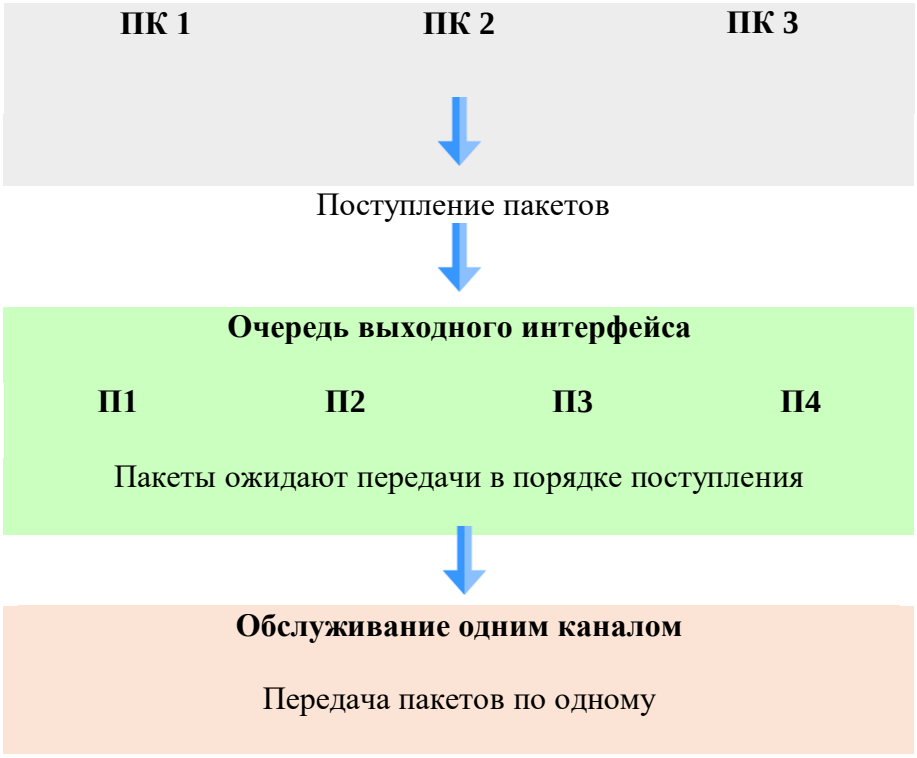


Рисунок 2. Упрощённое представление выходной очереди сетевого устройства

При низкой интенсивности поступления большинство пакетов передаётся без длительного ожидания. При увеличении нагрузки очередь растёт. В результате увеличивается задержка передачи данных.

Следовательно, анализ M/M/1 позволяет оценить, насколько эффективно используется выходной канал и как изменение сетевой нагрузки влияет на задержку пакетов.

6. Пример расчёта характеристик очереди

Рассмотрим выходной интерфейс маршрутизатора со следующими параметрами:

- интенсивность поступления пакетов: \(\lambda=80\) пакетов/с;
- интенсивность обслуживания: \(\mu=100\) пакетов/с.

Шаг 1. Рассчитаем коэффициент загрузки.

$$\rho = \frac{80}{100} = 0{,}8$$

Коэффициент загрузки составляет 80%. Поскольку \(\rho<1\), система устойчива.

Шаг 2. Найдём вероятность отсутствия пакетов.

$$P_0 = 1 - 0{,}8 = 0{,}2$$

Вероятность того, что система пуста, равна 20%.

Шаг 3. Рассчитаем среднее количество пакетов в системе.

$$L = \frac{0{,}8}{1 - 0{,}8}$$

В среднем в системе находятся 4 пакета, включая пакет на обслуживании.

Шаг 4. Рассчитаем среднее количество пакетов в очереди.

$$L = \frac{0{,}8^2}{1 - 0{,}8} = \frac{0{,}64}{0{,}2}\,3{,}2$$

В среднем 3,2 пакета ожидают обслуживания.

Шаг 5. Найдём среднее время пребывания пакета.

$$W = \frac{1}{100 - 80} = 0{,}05\ c$$

$$W = 50\ \text{мс}$$

Среднее время пребывания пакета в системе составляет 50 мс.

Шаг 6. Найдём среднее время ожидания.

$$W_q = \frac{0{,}8}{100 - 80} = 0{,}04\ c$$

W_q=40 мс

Среднее время ожидания в очереди составляет 40 мс. Остальные 10 мс в среднем приходятся на обслуживание пакета.

Результаты расчёта

Показатель	Результат
Интенсивность поступления \(\lambda\)	80 пакетов/с
Интенсивность обслуживания \(\mu\)	100 пакетов/с
Коэффициент загрузки \(\rho\)	0,8
Вероятность отсутствия пакетов \(P_0\)	20%
Среднее количество пакетов в системе \(L\)	4
Среднее количество пакетов в очереди \(L_q\)	3,2
Среднее время пребывания \(W\)	50 мс
Среднее время ожидания \(W_q\)	40 мс

Результаты показывают, что даже при устойчивой работе сети пакеты могут значительное время ожидать передачи. Поэтому одной оценки пропускной способности недостаточно для анализа качества обслуживания.

7. Влияние нагрузки на время ожидания

Рассмотрим, как изменяется время пребывания пакета при увеличении интенсивности поступления. Интенсивность обслуживания остаётся равной 100 пакетам/с.

λ, пакетов/с	Загрузка ρ	Среднее время пребывания W
20	0,2	12,5 мс
50	0,5	20 мс
80	0,8	50 мс
90	0,9	100 мс
95	0,95	200 мс

При увеличении загрузки с 80% до 90% среднее время пребывания пакета возрастает с 50 до 100 мс. При загрузке 95% оно достигает 200 мс.

Таким образом, небольшое увеличение нагрузки при высокой загрузке может вызвать значительный рост задержки. Это важно учитывать при проектировании сетей для IP-телефонии, видеоконференций и других приложений, чувствительных к задержкам.

8. Значение анализа очередей для обеспечения QoS

Анализ очередей позволяет оценить работу сетевых устройств и определить условия, при которых качество обслуживания ухудшается.

Основные задачи анализа:

1. Определение допустимой нагрузки на выходной интерфейс.
2. Оценка среднего времени ожидания пакетов.
3. Выявление риска чрезмерного роста очередей.
4. Выбор пропускной способности канала.
5. Обоснование применения приоритетного обслуживания.
6. Оценка влияния перегрузки на приложения.

Для снижения задержек могут применяться увеличение пропускной способности канала, распределение трафика по нескольким маршрутам, управление очередями и настройка приоритетов.

Следует учитывать, что классическая модель M/M/1 не описывает все особенности реальной сети. Она не учитывает сложные зависимости между пакетами, различия в размерах пакетов и ограниченную ёмкость буфера. Поэтому её результаты следует рассматривать как теоретическую оценку.

Заключение

В этой лекции мы рассмотрели назначение анализа очередей и основные элементы системы массового обслуживания. Особое внимание было уделено модели M/M/1, которая описывает обработку пакетов одним обслуживающим устройством.

Мы изучили формулы для расчёта коэффициента загрузки, среднего количества пакетов в очереди и времени ожидания. На примере маршрутизатора было показано, что при увеличении нагрузки задержка может резко возрасти, даже если пропускная способность канала ещё не превышена.

Модель M/M/1 помогает оценить производительность сетевых устройств и выбрать меры для улучшения качества обслуживания.

Контрольные вопросы

1. Что такое анализ очередей и для чего он применяется в компьютерных сетях?
2. Какие элементы входят в систему массового обслуживания?
3. Что означает обозначение M/M/1?
4. Какие предположения используются в модели M/M/1?
5. Что такое интенсивность поступления и интенсивность обслуживания?
6. Как рассчитывается коэффициент загрузки системы?
7. При каком условии модель M/M/1 работает в устойчивом режиме?
8. Чем отличается среднее время ожидания от среднего времени пребывания пакета?
9. Как увеличение нагрузки влияет на длину очереди и задержку?
10. Какие ограничения имеет модель M/M/1 при анализе реальных компьютерных сетей?