

Модуль 1_Моделювання та проектування високошвидкісних мереж, впровадження мережевих рішень конвергентної мережевої інфраструктури

Тема 2. Моделювання та проектування високошвидкісних мереж

Лекція 4. Модель якості роботи мережі

1. Складові якості роботи мережі. Матрична модель якості мереж

2. Первинні та вторинні параметри якості

3. Еталонні з'єднання та можливі процеси подій у мережах

ЛІТЕРАТУРА

1. Стеклов В.К., Беркман Л.Н. Проектування телекомунікаційних мереж. Київ, "Техніка", 2003 – 923 с.
2. Оліфер В.Г. Комп'ютерні мережі. Принципи, технології, протоколи. Посібник для Вузів. 2-е вид. –СПБ. Пітер. 2005. -с.176-201.

Quality of Service - QoS

- Висока вартість каналів передачі даних і складність підвищення швидкості передачі даних за рахунок прокладки додаткових волоконно-оптичних ліній обумовлює **надзвичайно економне відношення до пропускної спроможності каналу в мережах**
- Для нормальної роботи додатків в таких умовах потрібне застосування **методів забезпечення якості обслуговування** Quality of Service
- Іншим могутнім стимулом втілювання механізмів QoS є прагнення до передачі по пакетним мережам всіх типів інформацій, в тому числі і мультимедійної - голосу, зображення, відео
- Трафік таких застосувань відрізняється **чутливістю до затримки** - одного з параметрів, який регулюється за допомогою механізмів QoS
- Тому в більшості технологій, спеціально розроблених для глобальних мереж передачі даних - Frame Relay, ATM, - механізми **QoS є вбудованими**

Альтернатива



- Просте підвищення пропускної спроможності мережі не є гарантією того, що обладнання яке працюють в мережі, отримає те обслуговування, яке потрібне.
- **Екстенсивний шлях розвитку**, коли канали і комунікаційне устаткування мережі замінюється іншим, на порядок продуктивнішим - дає швидкий бажаний результат
 - додатком дістається необхідна для їх якісної роботи частка пропускної спроможності;
 - затримки із-за очікування пакетів в чергах (невід'ємна особливість мереж цього типу) стають невідчутними.
 - *Проте з перебігом часу кількість користувачів мережі зростає, змінюються саме застосування – як правило, трафіки стають більш ресурсоємними, і мережа починає відчувати перенавантаження.*

Типи QoS

- Типи QOS розрізняються по ступеню «строгості», тобто по тому, наскільки твердо сервіс QOS може гарантувати забезпечення певних значень характеристик QoS-пропускної спроможності, затримок, варіацій затримок, рівня втрат пакетів і т.п.
- Можна виділити три типи служб QoS
- **1 Сервіс з максимальними зусиллями**, який також можна назвати відсутністю QoS
 - Забезпечує взаємодію кінцевих вузлів без яких би не було б гарантій.
 - Типові представники таких послуг - класичні мережі Ethernet або IP, які не роблять ніяких відмінностей між пакетами окремих користувачів і додатків і обслуговують ці пакети на основі принципу FIFO (першим прийшов - першим обслужений).
- **2 Сервіс з перевагою (званий також «м'яким» сервісом QOS)**
 - деякі типи трафіку обслуговуються краще, ніж інші.
 - Мається на увазі швидша обробка, в середньому більше пропускної спроможності і менше втрат даних.
 - Це статистична перевага, а не чисельно вираженні гарантії. Точні значення параметрів QoS, які отримають додаток в результаті роботи служби цього типу, невідомі і залежать від характеристик пропонованого мережі трафіку.
 - **Наприклад**, якщо високо пріоритетний трафік пропонує мережі в даний момент часу низьку інтенсивність своїх пакетів, то низько пріоритетний трафік може в цей час отримувати вельми якісне обслуговування - значну пропускну спроможність і низькі затримки. Проте при зміні ситуації, коли високо пріоритетний трафік починає посилати в мережу свої пакети з високою інтенсивністю, низько пріоритетний трафік може взагалі якийсь час не обслуговуватися.

Продовження

- **3. Гарантований сервіс («жорстким», або «істинним», сервіс QoS)** дає статистичні численні гарантії різним потокам трафіку.
- Зазвичай такий вид QoS заснований на **попередньому резервуванні** мережених ресурсів для кожного з потоків, що отримав гарантії обслуговування.
- Трафік, якому виділили ресурси, гарантовано має при проходженні через мережу ті параметри пропускної спроможності або затримок, які визначені для нього в числовому вигляді (звичайно, якщо джерела цього трафіку не порушують обумовлених для них умов і генерують не більше пакетів, чим передбачалося).
- Служби такого типу здатні, наприклад, гарантувати додатку обумовлену пропускну спроможність, не зменшується ні за яких обставин, якою б переобтяженою мережа не ставала.
- Слід зазначити, що гарантії носять статистичний характер, тобто можна гарантувати деяке числові значення якого-небудь параметра тільки з деякою вірогідністю, хай дуже високою, не рівною 1.
- **Наприклад**, з вірогідністю 0,999 можна стверджувати, що затримка пакету не перевищить 100 мс, отже, один пакет з 1000 може затриматися в мережі і на більший час. Другою складовою такого режиму роботи служби QoS є вхідний контроль потоків, яким дані гарантії. Дійсно, дотримувати дані гарантії можна тільки в тому випадку, якщо інтенсивності вхідних в мережу потоків не перевищують граничних нижче зазначених значень. Інакше потік споживатиме більше ресурсів, чим йому виділялося, а значить, іншим потокам дістанеться менше, ніж було заплановано, і гарантії для них дотримані не будуть.

Згода про рівень обслуговування

- *Природною основою нормальної співпраці постачальника і споживача є договір, який в даному випадку називається угодою про рівень обслуговування (Service Level Agreement, SLA).*
- У цьому договорі постачальник послуг і його клієнт визначають наступні позиції:
 1. Параметри якості обслуговування трафіку, які цікавлять споживача і які згоден підтримувати постачальник (за показаннями середня пропускна спроможність, максимальні затримки і варіації затримок, максимальна інтенсивність втрат даних, коефіцієнт готовності сервісу, максимальний час відновлення сервісу після відмови.
 2. Методи вимірювання параметрів якості обслуговування.
 3. Визначення плати за обслуговування.

Система оплати може бути достатньо складною, особливо якщо угода передбачає декілька вирівняний якості обслуговування, які оплачуються за різними тарифами.
 4. Санкції за порушення зобов'язань постачальника послуг із-за забезпечення належної якості обслуговування, а також за відхилення параметрів трафіка користувача від обумовлених значень. Ці санкції можуть виражатися у вигляді штрафів або в іншій формі, наприклад у формі надання сервісу протягом деякого часу безкоштовно або за зниженим тарифом.
 5. Як і любий договір, угода SLA по взаємної згоди постачальника і клієнта може включати велику кількість різних додаткових статей. Наприклад, статтю, що обумовлює умови переходу до більш якісного обслуговування або обслуговуванню з різним рівнем якості в залежності від дня тижня або часу доби.

- 6. Угоду може включати - правила кондиціонування трафіку користувача, тобто обробки трафіку, який виходить за обговорені межі, наприклад трафіку з більшою середньою інтенсивністю на значному проміжку часу.
- Також правила кондиціонування можуть визначати умови відкидання або маркування пакетів-порушників (помічені пакети відкидатимуться мережею не завжди, а тільки у тому випадку, коли мережні пристрої випробовують перевантаження).
- Багато постачальників послуг пропонують своїм клієнтам типові контракти SLA.
- В них не тільки визначений перелік характеристик якості обслуговування, але навіть і їх конкретні числові значення, наприклад: «затримка пакетів, усереднена цінна за місяць, не перевищуватиме 100 мс при передачі між будь-якими двома Вузлами мережі». Типові контракти полегшують життя постачальникам послуг, оскільки для їх реалізації можна обійтися без засобів гарантованої підтримки якості обслуговування. Потрібно тільки підтримувати приблизно постійний рівень запасу пропускної спроможності і пропонувати в SLA ті значення параметрів QoS, які демонструє працююча мережа.
- Згода SLA полягає не тільки між постачальником послугами публічних мереж і корпоративними клієнтами. Достатньо популярним останнім часом стало укладення подібних контрактів між інформаційним відділом підприємства - постачальником транспортних послуг, і споживачами – функціональними відділами підприємства.

Вимоги до якості обслуговування додатків різних типів.

- Сучасна тенденція конвергенції мереж різних типів, привела до необхідності перенесення мережею всіх видів трафіку, а не тільки традиційного для комп'ютерних мереж трафіку додатків доступу до файлів і електронної пошти.
- До теперішнього часу виконана велика робота по класифікації трафіку додатків. Як основні критерії класифікації були прийняті три характеристики трафіку:
 - **1 відносна передбаченість швидкості передачі даних;**
 - **2 чутливість трафіку до затримок пакетів;**
 - **3 чутливість трафіку до втрат і спотворень пакетів.**

Передбаченість швидкості передачі даних

- Відносно передбаченості швидкості передачі даних трафік додатків ділиться на два великі класи:
- 1 потоковий трафік (stream);

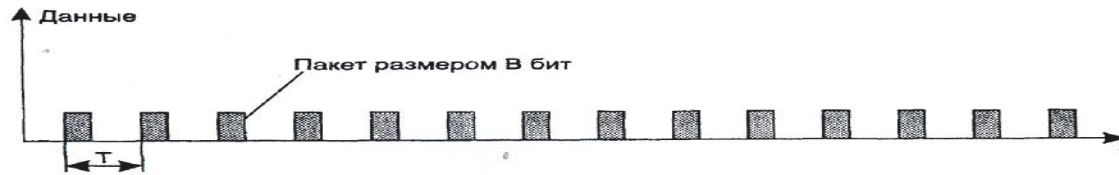
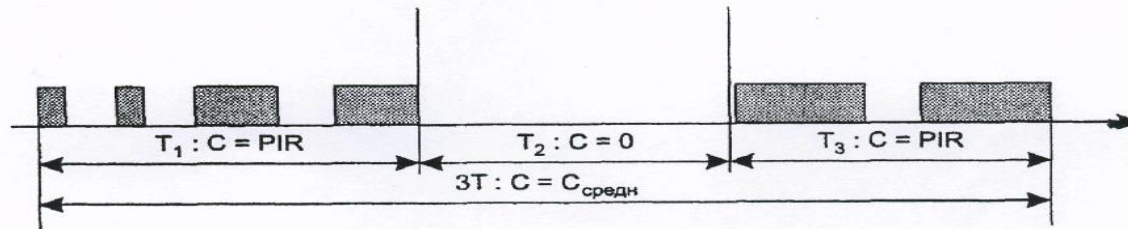


Рис. 8.1. Поточковый трафик

- 2 пульсующий трафік (burst).



- Пульсующий трафік

Потоковий трафік

- Постійна швидкість потокового трафіку (CBR) може бути обчислена шляхом усереднювання на одному періоді: - CBR - B/T бит/с.
- У загальному випадку постійна швидкість потокового трафіку менше номінальної максимальної бітової швидкості протоколу, за допомогою якого передаються дані, оскільки між пакетами існують паузи. Так максимальна швидкість передачі даних за допомогою протоколу Ethernet – 10 Base-T складає 9,76 Мбит/с (випадок кадрів максимальної довжини), що менше за номінальну швидкість цього протоколу, рівну 10 Мбит/с
- Додатки з пульсуючим трафіком відрізняються високим ступенем непередбаченості, коли періоди мовчання змінюються пульсацією, протягом якої пакети «щільно» слідують один за одним. В результаті трафік характеризується змінною бітовою швидкістю (Variable Bit Rate, VBR), що ілюструє мал. 8.2. Так, при роботі додатків файлового сервісу інтенсивність трафіка, що генерується додатком, може падати до нуля, коли файли не передаються, і підвищуватися до максимально доступною, обмеженою тільки можливостями мережі, коли файловий сервер передає файл.

Пульсуючий трафік

- На малюнку показано три періоди вимірювань T_1 , T_2 і T_3 . Для спрощення розрахунків прийняте, що пікові швидкості на першому і третьому періодах рівні між собою і рівні PIR , а всі три періоди мають однакову тривалість T . Враховуючи це, можна обчислити величину пульсації, яка рівна кількості бітів, переданих на періоді пульсації:
$$B = PIR \times T.$$
- Таким чином, величина пульсації для періодів T_1 і T_3 рівна B , а на періоді T_2 - нулю.
- Для приведеного прикладу можна підрахувати коефіцієнт пульсації. (Нагадаємо, що він рівний відношенню пікової швидкості на якому-небудь невеликому періоді часу до середньої швидкості трафіку, зміряної на тривалому періоді часу.) Оскільки пікова швидкість на періоді T_1 (або T_3) рівна B/T , а середня швидкість на сумарному періоді $T_1 + T_2 + T_3$ рівна $2B/3T$, то коефіцієнт пульсації рівний $3/2$.
- Практично будь-який трафік, навіть трафік поточкових застосувань, має не нульовий коефіцієнт пульсації. Просто значення коефіцієнтів пульсації у поточкового і пульсуючого трафіків істотно розрізняються. У додатків з пульсуючим трафіком він зазвичай знаходиться в межах від 2:1 до 100:1, а у поточкових застосувань близький до 1:1. У локальних мережах коефіцієнт пульсації зазвичай вище, ніж в глобальних, оскільки на магістралях глобальних мереж трафік є сумою трафіків багатьох джерел, що по закону великих чисел приводить до згладжування результуючого трафіку.

Чутливість трафіку до затримок пакетів

Ще один критерій класифікації додатків за типом трафіку - *чутливості до затримок пакетів і їх варіацій*.

основні типи додатків в порядку підвищення чутливості до затримок пакетів

- **1 Асинхронні** Практично немає обмежень на час затримки (**еластичний трафік**). Приклад такого застосування - електронна пошта
- **2 Інтерактивні** Затримки можуть бути відмічені користувачами, але вони не позначаються негативно на функціональності додатків. Приклад - текстовий редактор, що працює з віддаленим файлом
- **3 Ізохронні** Є поріг чутливості до варіацій затримок, при перевищенні якого різко знижується функціональність застосування. Приклад - передача голосу, коли при перевищенні порогу варіації затримок в 100-150 мс різко знижується якість відтвореного голосу, а при ≥ 450 втрачається контакт
- **4 Надчутливі до затримок** Затримка доставки даних зводить функціональність додатку до нуля. Приклад - додатки, які управляють технічним об'єктом в реальному часі. При запізненні управляючого сигналу на об'єкті може відбутися аварія.
- *інтерактивність додатку підвищує його чутливість до затримок. Наприклад, інтерактивний телефонний або телевізійна розмова не терпить затримок, що добре помітно при трансляції розмови через супутник. Тривалі паузи в розмові вводять співбесідників в оману, вони втрачають терпіння і починають чергову фразу одночасно.*
- В обладнанні, що застосовується сьогодні, розподіл додатків за цією ознакою на два класи:
- **- асинхронні** До асинхронних відносять ті застосування, які нечутливі до затримок передачі даних в дуже широкому діапазоні, аж до декількох секунд
- **- синхронні** - решта всіх застосувань, на функціональність яких затримки впливають істотно

Чутливість трафіку до втрат і спотворень пакетів

чутливість до втрат пакетів ділять на дві групи.

■ 1 Додатки, чутливі до втрати даних.

Всі додатки, що передають алфавітно-цифрові дані (до яких відносяться текстові документи, коди програм, числові масиви і т. і.), володіють високою чутливістю до втрати окремих фрагментів даних. Такі втрати часто ведуть до повного знецінення успішно прийнятої інформації. Наприклад, відсутність хоч би одного байта в коді програми робить її абсолютно непридатною. Всі традиційні мережеві застосування (файловий сервіс, сервіс баз даних, електронна пошта і т. д.) відносяться до цього типу додатків.

■ 2 Додатки, стійкі до втрати даних. До цього типу відносяться багато застосувань, що передають трафік з інформацією про інерційні фізичні процеси. Стійкість до втрат пояснюється тим, що невелика кількість відсутніх даних можна визначити на основі прийнятих. Так, при втраті одного пакету, що несе декілька послідовних вимірів голосу, відсутні виміри при відтворенні голосу можуть бути замінені апроксимацією на основі сусідніх значень. До такого типу відноситься велика частина додатків, що працюють з мультимедійним трафіком (аудіо і відео додатки). Проте стійкість до втрат має свої межі, тому відсоток втрачених пакетів не може бути великим (*наприклад, не більше 1 %*),

■ *Можна відзначити, що не будь-який мультимедійний трафік такий стійкий до втрат даних, наприклад, скомпресований голос і відео зображення дуже чутливі до втрат, тому відносяться до першого типу додатків.*

Класи додатків

- Між значеннями трьох характеристик якості обслуговування (відносна передбаченість швидкості передачі даних; чутливість трафіку до затримок пакетів; чутливість трафіку до втрат і спотворенням пакетів) немає строгого взаємозв'язку.
- Наприклад, наступне поєднання характеристик додатку «породжуваний трафік - рівномірний потік, додаток ізохронний, стійкий до втрат» відповідає таким популярним застосуванням, як IP-телефонія, підтримка відео конференцій, аудіо віщання через Інтернет. Існують і такі поєднання характеристик, для яких важко привести приклад додатку, наприклад: «породжуваною трафік - рівномірний потік, додаток асинхронний, чуттєве до втрат».
- Стійких поєднань характеристик, що описують певний клас додатків, існує не так вже багато. Так, при стандартизації технології ATM, яка спочатку розроблялася для підтримки різних типів трафіку, було визначено 4 класи додатків: А, В, С і D. Для кожного класу рекомендується використовувати власний набір характеристик QoS, Крім того, для всіх застосувань, не включених ні в один з цих класів, був визначений клас X, в якому поєднання характеристик додатку може бути довільним.
- Класифікація ATM є на сьогодні найбільш детальною і загальною, вона не вимагає від нас знання технологій, використовуваних для передачі цих типів трафіка, тому приведемо її тут (табл. 1).

Клас трафіку	Характеристики
А	Постійна бітова швидкість, чутливість до затримок, передача зі встановленням з'єднання (наприклад, голосовий трафік, трафік телевізійного зображення). Параметри QoS: пікова швидкість передачі даних, затримка, джітер
В	Змінна бітова швидкість, чутливість до затримок, передача зі встановленням з'єднання (наприклад, компресований голос, компресування відео зображення). Параметри QoS: пікова швидкість передачі даних, пульсація, середня швидкість передачі даних, затримка, джітер
С	Змінна бітова швидкість, еластичність, передача зі встановленням з'єднання (наприклад, трафік комп'ютерних мереж, в яких кінцеві вузли працюють по протоколах зі встановленням з'єднань, - frame relay, X.25, TCP). Параметри QoS: пікова швидкість передачі даних, пульсація, середня швидкість передачі даних
D	Змінна бітова швидкість, еластичність, передача без встановлення з'єднання (наприклад, трафік комп'ютерних мереж, в яких кінцеві вузли працюють по протоколах без встановлення з'єднань, - IP/UDP, Ethernet). Параметри QoS не визначені
X	Тип трафіку і його параметри визначаються користувачем

Параметри якості обслуговування

- Трьом критеріям класифікації додатків (передбаченість швидкості передачі даних, чутливість до затримок і чутливість до втрат) відповідають три групи параметрів
- **1 Параметри пропускної спроможності.** До таких параметрів відносяться середня, максимальна (пікова) і мінімальна швидкості передачі даних.
- **2 Параметри затримок.** Використовується середня і максимальна величини затримок, а також середнє і максимальне значення варіацій затримок, тобто відхилень між пакетних інтервалів в трафіку
- **3 Параметри надійності передачі.** Використовується відсоток втрачених пакетів.
- При визначенні всіх цих параметрів важливо, на якому періоді і вимірюється даний параметр. Чим менше цей період, тим більше жорсткими є відповідно вимоги якості обслуговування і тим важче для мережі їх витримати. Тому постачальники послуг віддають перевагу середньомісячній характеристиці., тоді як постачальники послуг мереж frame relay і ATM, що мають в своєму розпорядженні могутні засоби QoS, здатні гарантувати параметри, усереднені на періоді в декілька секунд.
- Для додатків з пульсуючим трафіком якість обслуговування краще всього характеризується середньою швидкістю і максимальною швидкістю, яка потрібна в період пульсації.
- Параметри якості обслуговування зазвичай обмовляються в угоді про рівень обслуговування SLA між користувачем мережі і постачальником послуг. Після укладення угоди користувач і постачальник послуг повинні належним чином набудувати свої програмні і апаратні засоби, щоб вони відпрацьовували обумовлені параметри.

Служба QoS Модель служби QoS

- Мережа - це розподілене середовище, що складається з великої кількості пристроїв для підтримки різних технологій і протоколів. Тому складно дотримувати єдині вимоги по якісному обслуговуванню різних видів трафіку на всьому протязі складеного шляху від одного кінцевого вузла до іншого, тобто **«з кінця в кінець» (end-to-end)**. Завдання ж просування пакетів в такій мережі із заданими параметрами якості обслуговування істотно складніше ніж перша.
- **Для рішення поставлених задач в мережі необхідна служба QoS.** Ця служба має розподілений характер так як її елементи повинні бути присутніми у всіх мережних пристроях, що просувають пакети: комутаторах, маршрутизаторів, серверах доступу. З іншого боку, роботу окремих мережних пристроїв та забезпеченню підтримки QoS потрібно скоординувати, щоб якість обслуговування була однорідною уздовж всього шляху, по якому слідують пакети. Тому служба QoS повинна включати також **елементи централізованого управління**, за допомогою яких адміністратор мережі може конфігурувати механізми QoS в окремих пристроях мережі.
- Базова архітектура служби QoS включає елементи трьох основних типів
- **1 засоби QoS вузла**, що виконують обробку трафіку, що поступає у вузол, відповідно до вимог якості обслуговування;
- **2 протоколи QoS-сигналізації** для координації роботи мережних елементів по підтримці якості обслуговування «з кінця в кінець»;
- **3 централізовані функції політики**, управління і обліку QoS, дозволяють адміністраторам мережі цілеспрямовано впливати на мережеві елементи для розділення ресурсів мережі між різними видами трафіку

Засоби QoS вузла

- Засоби QoS вузла являються основним виконавчим механізмом служби QoS, оскільки саме вони безпосередньо впливають на процес просування пакетів між вхідними і вихідними інтерфейсами комутаторів і маршрутизаторів і, отже, визначають внесок даного пристрою в характеристики якості обслуговування мережі. Засоби QoS вузла можуть включати механізми двох типів:
 - **1 механізми обслуговування черг;**
 - **2 механізми кондиціонування трафіку;**
 - **3 протокола резервування ресурсів.**

механізми обслуговування черг

- Механізми обслуговування черг є необхідним елементом будь-якого пристрою, що працює за принципом комутації пакетів.
- Вони можуть підтримувати різні алгоритми обробки пакетів, які потрапили в чергу, від найпростіших типу FIFO (першим прийшов - першим обслужений) до дуже складних, підтримуючих обробку декількох класів потоків, наприклад алгоритмів пріоритетного, або зваженого, обслуговування.
- За умовчанням в мережних пристроях діє алгоритм черги FIFO, але він достатній тільки для реалізації обслуговування з максимальними зусиллями, а *для підтримки «дійсних» сервісів QoS потрібні складніші механізми.*
- *Черги виникають в ті періоди часу, коли швидкість надходження трафіку стає більше швидкості його пересування.*
- Механізми обслуговування черг *розраховані на роботу якраз в періоди перевантаження* і потрібні для того, щоб потоки як умого менше страждали від існування таких періодів.
~~Затримки від очікування в чергах повинні укладатися в параметри потоку.~~

механізми кондиціонування трафіку

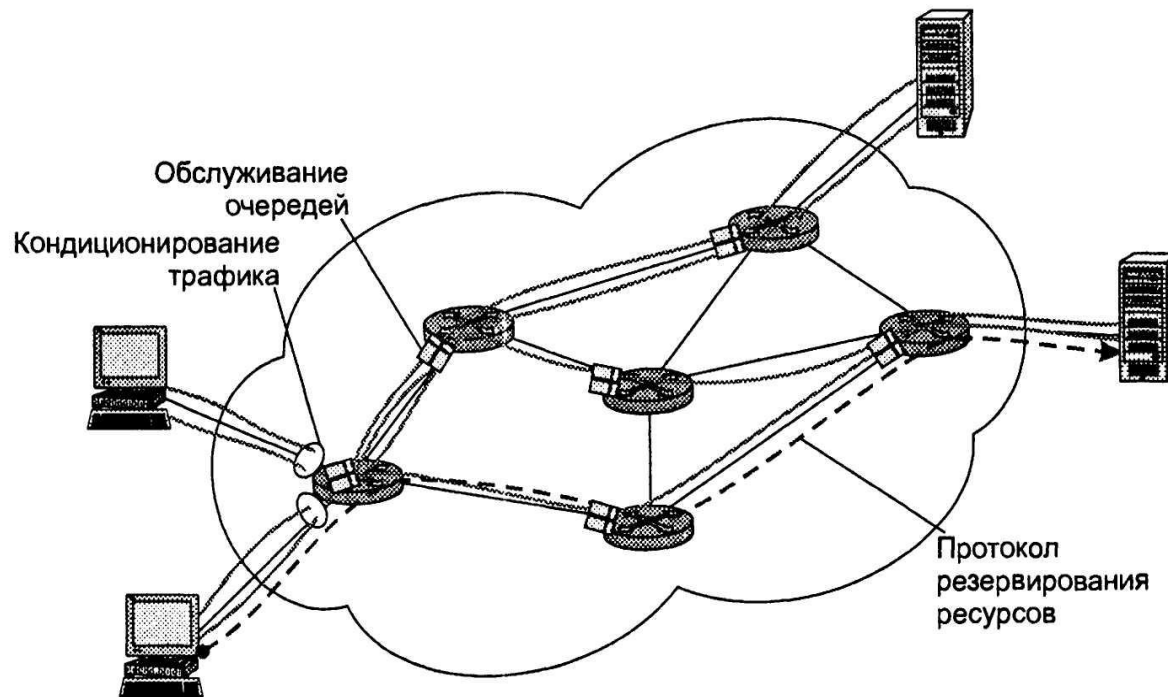
- Механізми другого типу (кондиціонування трафіку) можуть реалізовуватися при створенні таких вимог, коли швидкість просування трафіку потоку узгоджується. із швидкістю надходження цього трафіку у вузол мережі.
- Механізми вирішують задачу створення умов якісного обслуговування трафіку способом - *за рахунок зменшення швидкості надходження потоку :в даний вузол настільки, щоб вона завжди залишалася менше, ніж швидкість просування цього потоку.*

Механізм кондиціонування трафіку

включає виконання функцій

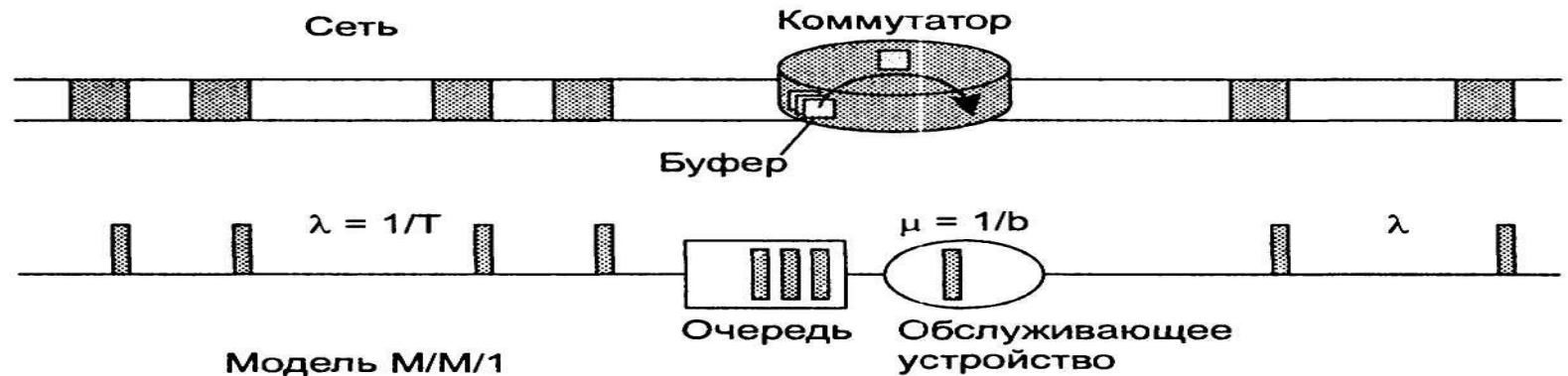
- **1 Класифікація трафіку.** - функція виділяє із загальної послідовності пакетів, пакети одного потоку, що має загальні вимоги до якості обслуговування. Класифікація може виконуватися на основі різних ознак пакету - адрес джерела і призначення, ідентифікаторів додатків, значення пріоритету пакета, значення мітки потоку.
- **2 Профілізація трафіка** на основі правил політики. Для кожного вхідного потоку є відповідний йому набір параметрів QoS - профіль трафіка. У разі порушення параметрів профілю (наприклад, перевищення пульсації або середньої швидкості) відбувається відкидання або маркіровка пакетів. Відкидання пакетів знижує інтенсивність потоку і приводить його параметри у відповідність з профілем. Маркіровка пакетів без відкидання потрібна для того, щоб пакети були обслужені з якістю обслуговування, відмінним від указанного в профілі. Для перевірки відповідності вхідного трафіку заданому профілю механізм кондиціонування виконує вимірювання параметрів потоку.
- **3 Формування трафіку (*shaping*).** функція призначена для додання трафіку тимчасової «форми». За допомогою функції прагнуть згладити пульсації трафіку. Згладжування пульсацій зменшує черги в мережних пристроях, які оброблятимуть трафік далі по потоку.

Протокол резервирования ресурсов



Архитектура системы обеспечения качества обслуживания, основанная на резервировании ресурсов

нужен для автоматизации процедуры резервирования на всем пути следования некоторого потока, то есть «из конца в конец». Протокол резервирования является аналогом протоколов установления соединения в сетях с коммутацией каналов, поэтому он иногда называется **сигнальным**



Основними елементами моделі є:

- 1 вхідний потік абстрактних заявок на обслуговування;
- 2 буфер;
- 3 обслуговуючий пристрій;
- 4 вихідний потік заявок.

Управління чергами в елементах

- Основу засобів підтримки QoS в мережних елементах складають **черги і алгоритми обробки цих черг**. Ці механізми використовуються в будь-якому мережному пристрої - в маршрутизаторі, в комутаторі локальної або глобальної мережі, в кінцевому вузлі виключення складають тільки повторювачі, які пакетів не розрізняють, працюють на рівні потоків бітів.
- потрібен для **обробки періодів тимчасових перевантажень**, коли мережний пристрій не може передавати пакети на вихідний інтерфейс в тому темпі у якому вони поступають для виконання такого просування. Якщо причиною перевантаження є процесорний блок мережного пристрою, то для тимчасового зберігання необроблених пакетів використовується вхідна черга, тобто черга, пов'язана з вхідним інтерфейсом.

- Головним по ступеню впливу на виникнення черг чинником є **коефіцієнт навантаження пристрою (utilization)** - відношення середньої інтенсивності вхідного трафіку пристрою до середньої інтенсивності просування пакетів на вихідний інтерфейс.
- Якщо коефіцієнт навантаження *більше одиниці*, значить, інтенсивність вхідного трафіку постійно вище, ніж інтенсивність просування пакетів на вихідний інтерфейс. Тому черга в пристрої існує завжди, її швидкість намагалась йти до нескінченності, коли б не кінцевий розмір буфера, відведеного під зберігання пакетів, що стоять в черзі. Але і у тому випадку, коли коефіцієнт навантаження менше одиниці, черга теж може існувати, більш того, мати достатньо значну середню довжину.
- *Варіація інтервалів* надходження пакетів є другим важливим чинником, що впливає на поведінку черг. При пульсуючому характері багатьох типів трафіку комп'ютерних мереж, коли коефіцієнт пульсацій рівний 100:1 або більш, черги можуть бути значними. Якщо ж ця варіація відсутня, тобто пакети прибувають строго через певні проміжки часу, як це відбувається у трафіку типу рівномірного потоку, то черга при коефіцієнті завантаження, меншому 1, не виникає.
- Вплив пульсацій трафіку на появу затримок обслуговування добре відомий користувачам сегментів Ethernet, що розділяються. Навіть при значеннях коефіцієнта навантаження сегменту 0,5 затримки доступу до мережі бувають значними, що примушує використовувати ці мережі з коефіцієнтом навантаження сегменту не більше 0,3.

Служба QoS використовує для підтримки гарантованого рівня

QoS складну модель

Це робиться за допомогою наступних методів:

- 1 за рахунок попереднього резервування смуги пропускання для трафіку з відомими параметрами (наприклад, значеннями середньої інтенсивності і коефіцієнту пульсації);
- 2 примусової профілізації вхідного трафіку, що дозволяє підтримувати коефіцієнт навантаження пристрою на потрібному рівні;
- 3 використання складних алгоритмів управління чергами.

Найчастіше в маршрутизаторах і комутаторах застосовуються наступні алгоритми обробки черг:

- 4 традиційний алгоритм FIFO;
- 5 пріоритетне обслуговування (Priority Queing), яке також називають «подавляючим»;
- 6 зважене обслуговування (Weighted Queing, WQ).

Кожен алгоритм розроблявся для вирішення певних завдань і специфічним чином впливає на якість обслуговування різних типів трафіка в мережі. Можливо і комбіноване застосування цих алгоритмів.

Традиційний алгоритм FIFO

- Принцип традиційного алгоритму FIFO полягає в тому, що у разі перевантаження пакети поміщаються в чергу, а при припиненні перевантаження передаються на вихід в тому порядку, в якому поступили, тобто «першим прийшов - першим пішов» (First In - First Out, FIFO).
- У всіх пристроях з комутацією пакетів - це *алгоритм обробки черг за умовчанням*.
- *Перевагою* його є простота реалізації і відсутність потреби в конфігурації.
- *Корінний недолік* - неможливість диференційованої обробки пакетів різних потоків. Всі пакети стоять в загальній черзі на рівних підставах - і пакети чутливого до затримок голосового трафіка, і пакети нечутливого до затримок, але дуже інтенсивного трафіку резервного копіювання, тривалі пульсації якого можуть надовго затримати голосовий пакет.

Висновок: *черги FIFO необхідні для нормальної роботи мережних пристроїв, але вони недостатні для підтримки диференційної якості обслуговування.*

Пріоритетне обслуговування

- Механізм пріоритетного обслуговування заснований на розділенні мережного трафіку на невелику кількість класів і подальшого призначення кожному класу деякої числової ознаки - пріоритету.
- Пріоритети можуть призначатися не тільки комутатором або маршрутизатором, але і додатком на вузлі-відправнику.



ПОЯСНЕННЯ

- У мережному пристрої, що підтримує пріоритетне обслуговування, є декілька черг (буферів), по одній для кожного пріоритетного класу.
- Пакет, що поступив в період перевантажень, поміщається в чергу, відповідну його пріоритетному класу. До тих пір, поки з пріоритетнішої черги не будуть вибрані всі наявні в ній пакети, пристрій не переходить до обробки наступною, менше пріоритетної черги. Тому пакети з низьким пріоритетом обробляються тільки тоді, коли порожні всі вищестоящі черги: з високим, середнім і нормальним пріоритетами.

Зважені черги

- Алгоритм зважених черг розроблений для того, щоб можна було надати всім класам трафіку певний мінімум пропускної спроможності або гарантувати деякі вимоги до затримок. Під вагою даного класу розуміється відсоток такою, що надається класу трафіку пропускної спроможності від повної пропускної спроможності вихідного інтерфейсу.

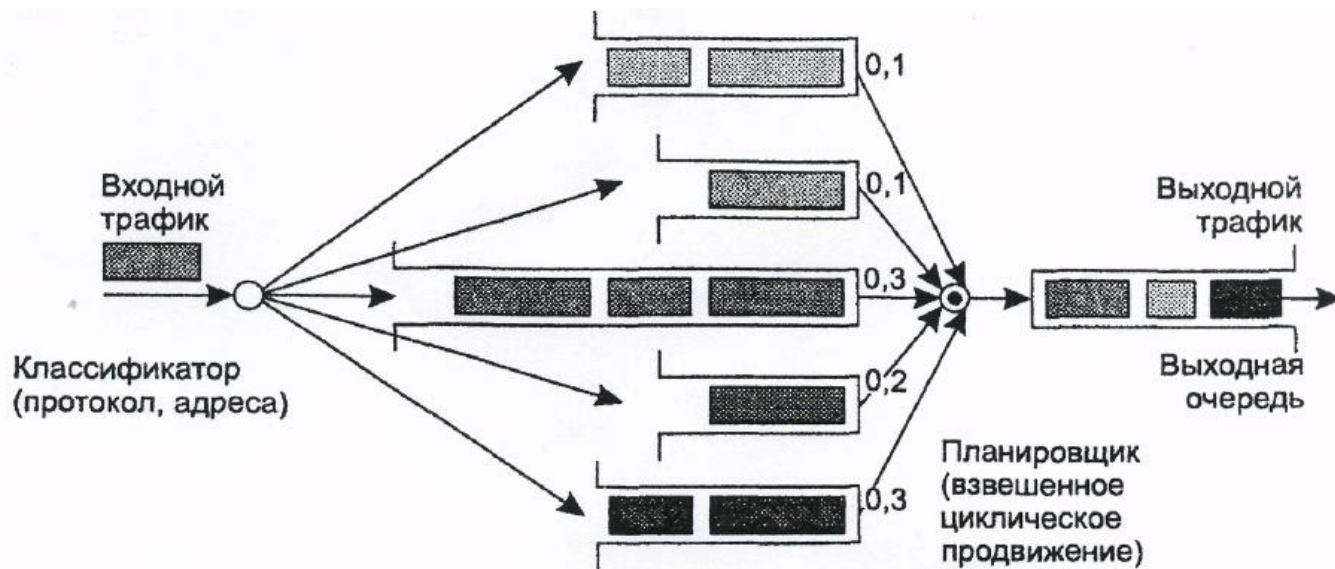


Рис. 8.10. Взвешенные очереди

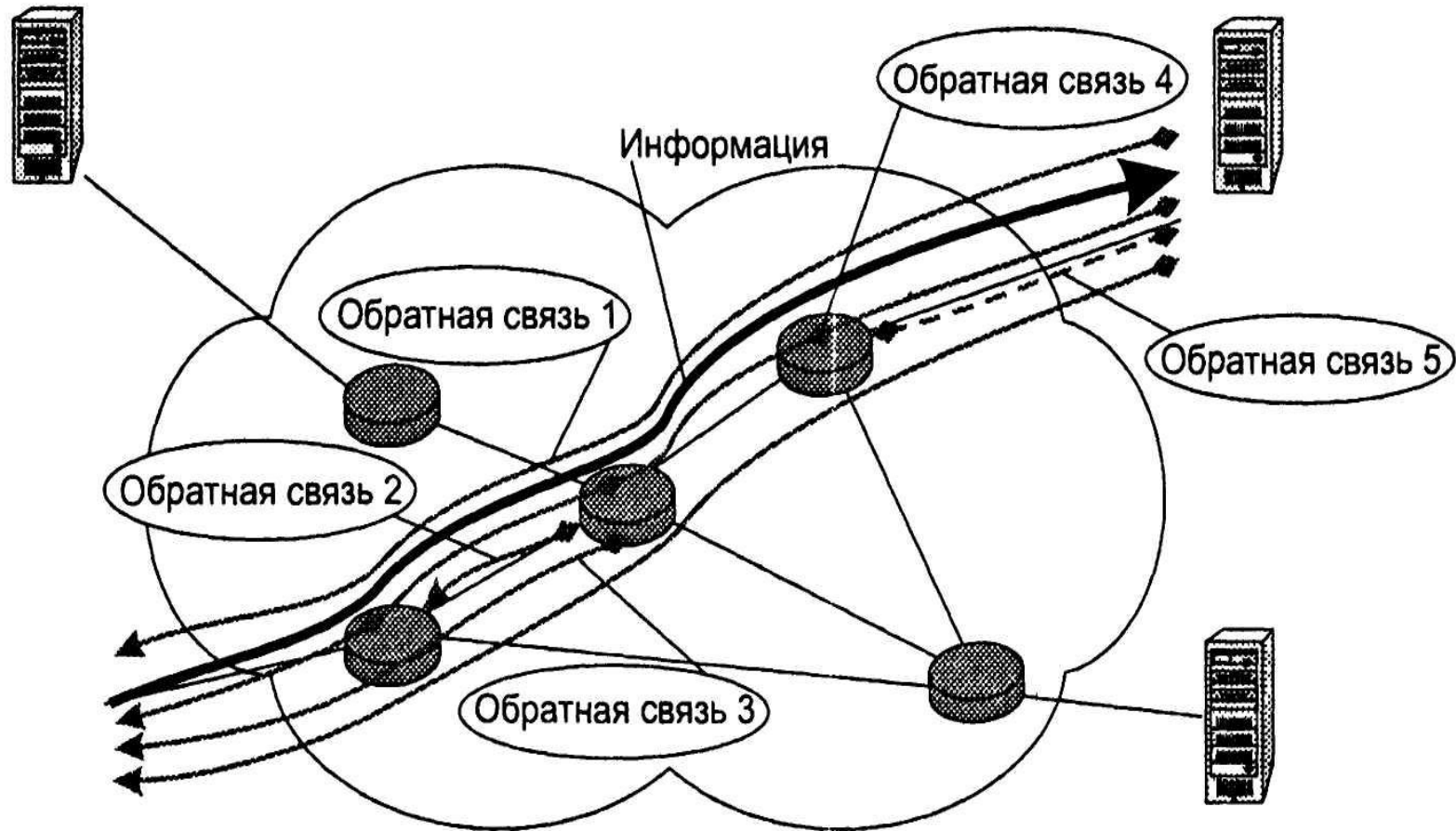
Комбіновані алгоритми обслуговування черг

- Кожен з описаних підходів має свої достоїнства і недоліки. Пріоритетне обслуговування, забезпечуючи мінімальний рівень затримок для черги найвищого пріоритету, не дає ніяких гарантій відносно середньої пропускної здатності для трафіку черг нижчих пріоритетів. Зважене обслуговування забезпечує заданий розподіл середньої пропускної спроможності, але не враховує вимог до затримок.
- Існують комбіновані алгоритми обслуговування черг. У найбільш популярному алгоритмі подібного роду підтримується одна пріоритетна черга при обслуговуванні решти черг відповідно до зваженого алгоритмом. Пріоритетна черга використовується для чутливого до затримкам трафіку, а решта - для еластичного трафіку декількох класів. Кожен клас еластичного трафіку отримує деякий мінімум пропускної спроможності при перевантаженнях. Цей мінімум обчислюється як відсоток від пропускної здатності, що залишилася від пріоритетного трафіку. Потрібно обмежити пріоритетний трафік, щоб він не поглинав всю пропускну здатність ресурсу. Це робиться засобами профілізації трафікаі.
- **Висновки:** *Алгоритми управління чергами не запобігають перевантаженням, а лише деяким «справедливим» чином в умовах дефіциту перерозподіляють ресурси між різними потоками або класами трафіку. Алгоритми управління чергами відносяться до механізмів контролю перевантажень, які починають працювати, коли мережа вже переобтяжена.*

клас засобів, які носять назву механізмів запобігання перенавантаження

- запобігти перевантаженню мережі можна в тому випадку, коли сумарна інтенсивність всіх потоків, що передаються кожним інтерфейсом кожного комутатора мережі, менше пропускної спроможності цього інтерфейсу. Добитися цього можна двома способами - збільшуючи пропускну здатність інтерфейсу або зменшуючи інтенсивності потоків.
- **Перший варіант** відноситься до засобів проектування і планування мережі.
- **Другий варіант** - зменшення інтенсивності потоків - можна реалізувати також двома принципово різними способами.
- Перший спосіб заснований на використанні **механізму зворотного зв'язку**, за допомогою якого переобтяжений вузол мережі, реагуючи на перевантаження, просить попередні вузли, розташовані уздовж маршруту проходження потоку (або потоків, належних одному класу), тимчасово знизити швидкість трафіку. Після того, як перенавантаження в даному вузлі зникне, він посилає інше повідомлення, що дозволяє підвищити швидкість передачі даних.
- Інший спосіб базується на **передчасному резервуванні пропускної здатності** для потоків, що протікають через мережу. Для цього йому необхідна попередня інформація про інтенсивності потоків. Принципи резервування ресурсів ми розглянемо пізніше.

зворотній зв'язок



Учасники зворотного зв'язку

Учасники зворотного зв'язку

- Зворотний зв'язок 1 - організовано між двома кінцевими вузлами мережі
- Зворотний зв'язок 2 - організовано між двома сусідніми комутаторами
- Зворотний зв'язок 3 - організовано між проміжним комутатором вузлом-джерелом
- Зворотний зв'язок 4 - як і у разі зворотного зв'язку 1, повідомлення про перенавантаження породжується вузлом призначення і передається вузлу-джерелу
- Зворотний зв'язок 5 - передача повідомлення про перевантаження вузлу призначення, який перетворить його в повідомлення зворотного зв'язку і відправляє в потрібному напрямі, тобто у напрямі джерела

Інформація зворотного зв'язку

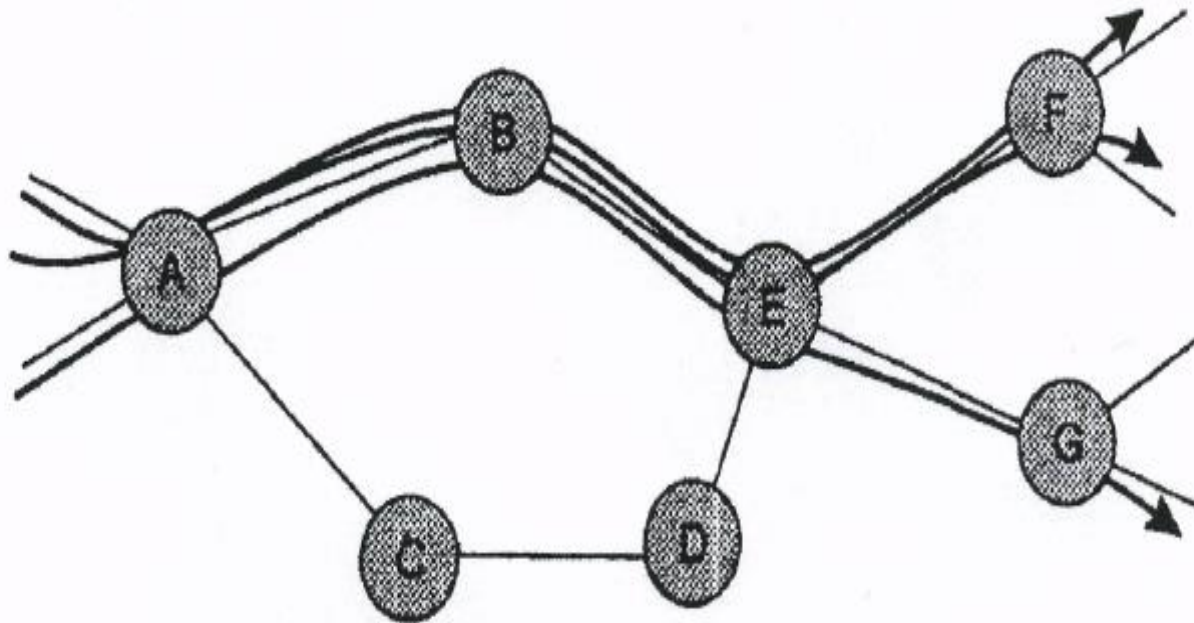
У вживаних сьогодні методах зворотного зв'язку використовуються наступні основні типи повідомлень:

- 1 ознака перевантаження;
- 2 максимальна швидкість передачі;
- 3 максимальний об'єм даних (кредит);
- 4 непрямі ознаки.

Інжиніринг трафіка

- Завдання вибору маршрутів для потоків (або класів) трафіку з урахуванням дотримання вимог QoS вирішують методи інжинірингу трафіку (Traffic Engineering, TI).
- За допомогою цих методів прагнуть добитися ще однієї мети - по можливості максимально і збалансовано завантажити всі ресурси мережі, щоб мережа при заданому рівні якості обслуговування володіла як можна більш високою сумарною продуктивністю.
- Традиційні методи маршрутизації розглядають якнайкращий вибраний маршрут як єдиного можливого, навіть якщо існують інші, хоч і декілька гірші маршрути.

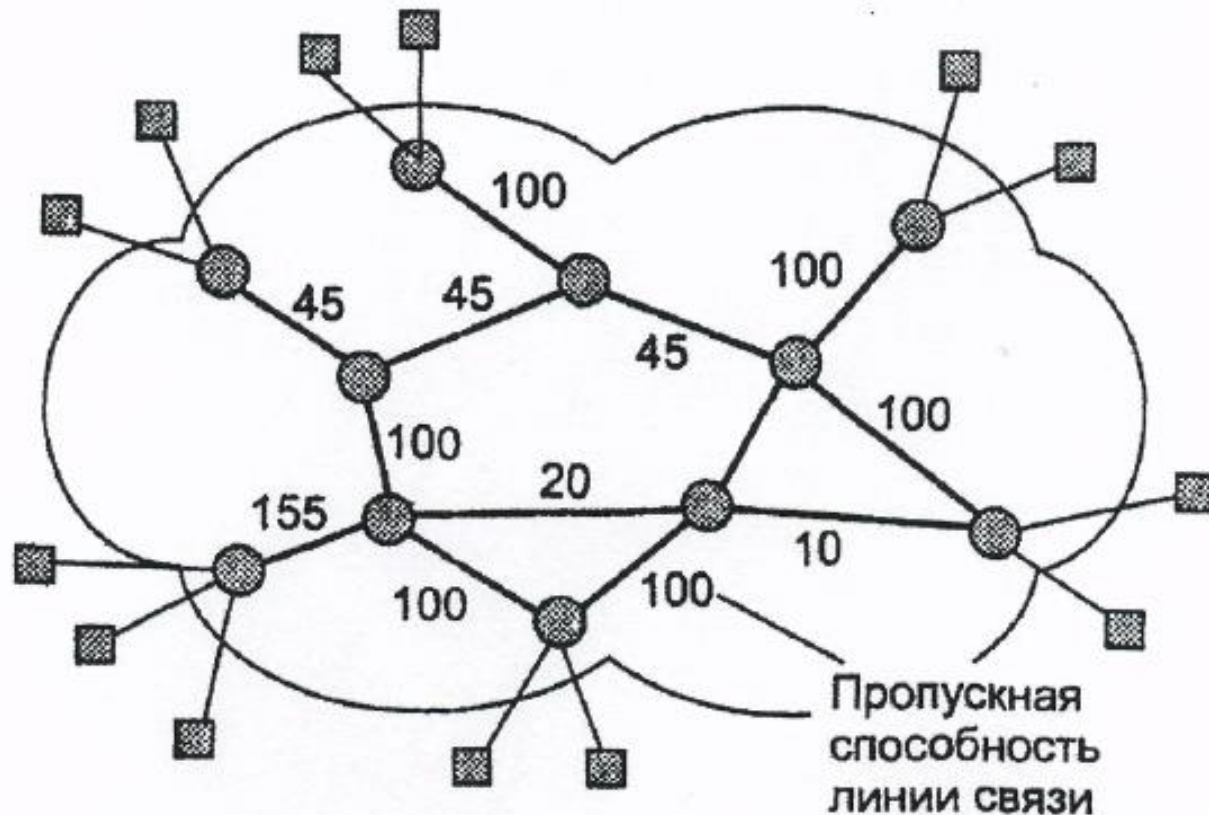
Класичним прикладом неефективності
такого підходу є - мережа з топологією
«риба»



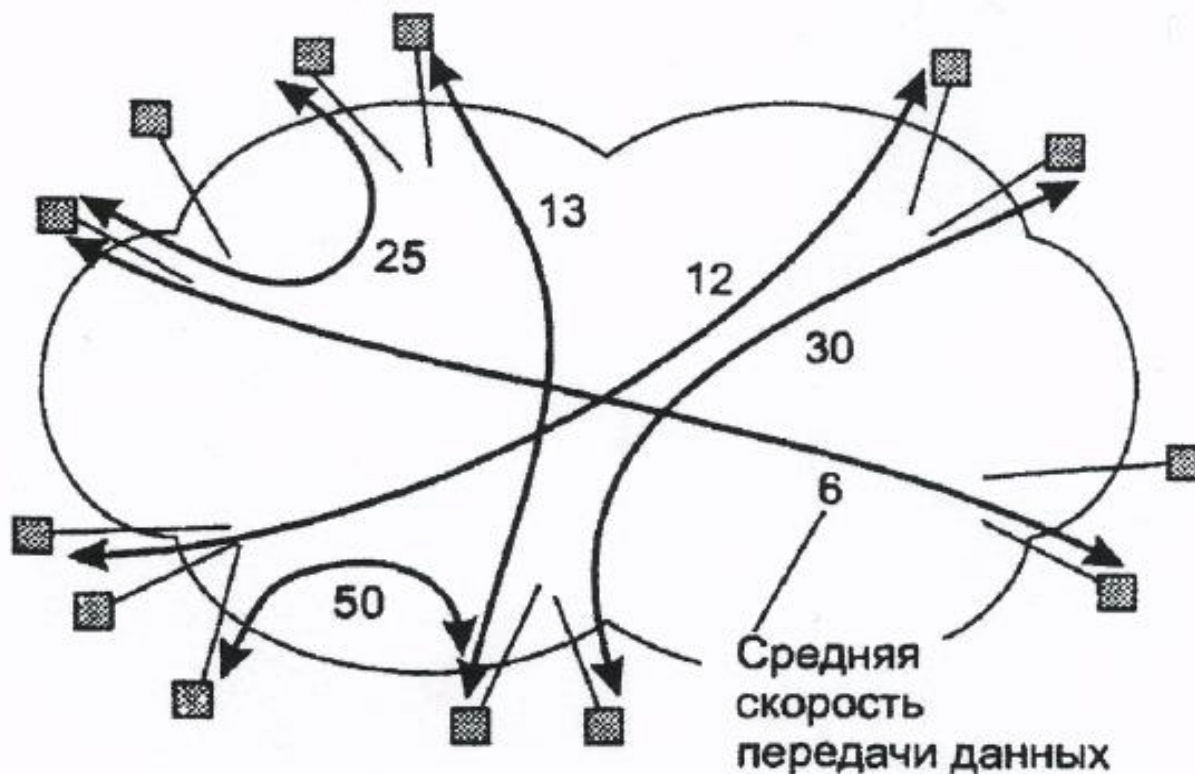
Методи інжинірингу трафіку

Початковими даними для методів інжинірингу трафіку є:

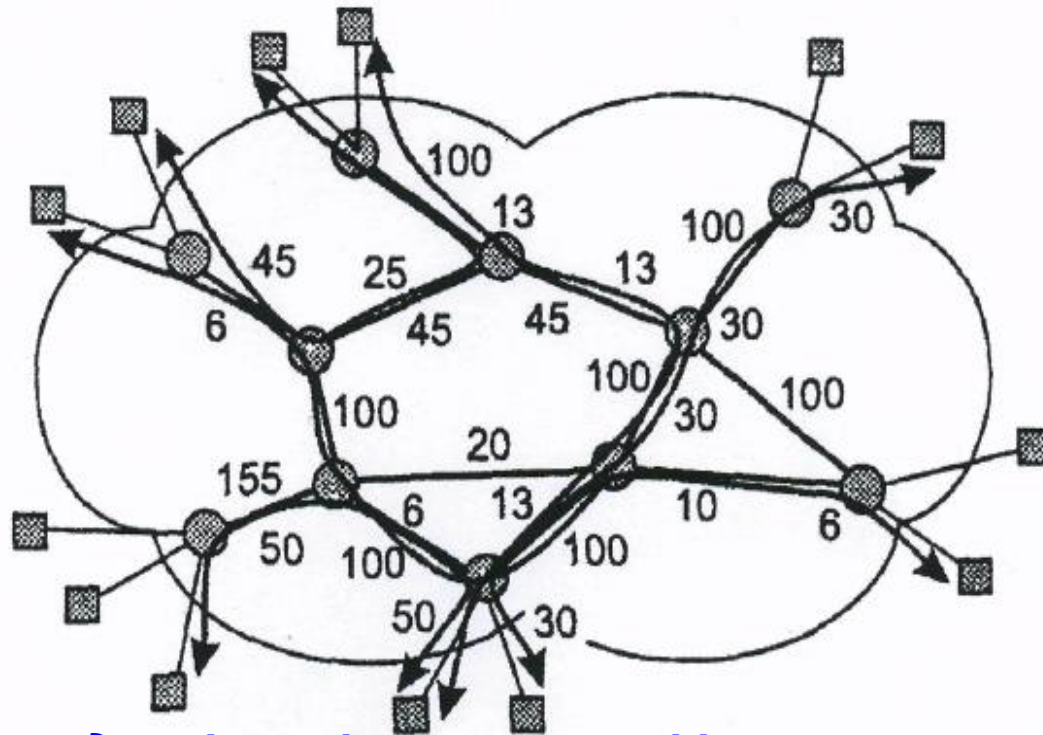
- 1 характеристики мережі, що передає, - її топологія, а також продуктивність складових її комутаторів і ліній зв'язку



2 зведення про запропоноване навантаження мережі, тобто про потоки трафіку, які мережа повинна передати між своїми прикордонними комутаторами



Рішення - розподіл навантаження по мережі - вибір шляхів передачі трафіку



Висновок Методи інжинірингу трафіку полягають у виборі раціональних маршрутів проходження потоків через мережу. Вибір маршрутів забезпечує максимізацію завантаження ресурсів мережі при одночасному дотриманні необхідних гарантій відносно параметрів якості обслуговування трафіку.

Рішення задачі ТІ можна шукати по-різному

- ***По-перше***, можна шукати його передчасно, у фоновому режимі. Для цього потрібно знати початкові дані: топологію і продуктивність мережі, а також вхідні і вихідні точки потоків трафіку і середню швидкість передачі даних в них. Після цього завдання раціонального розподілу шляхів проходження трафіку при фіксованих точках входу і виходу, а також заданому рівні максимального значення коефіцієнта використання ресурсу можна передати деякій програмі, яка за допомогою направленої перебору варіантів знайде рішення. Результатом роботи програми будуть точні маршрути для кожного потоку з вказівкою всіх проміжних комутаторів.
- ***По-друге***, можна вирішувати задачу в оперативному режимі, доручивши її самим комутаторам мережі. Для цього використовуються модифікації стандартних протоколів маршрутизації. Модифікація протоколів маршрутизації полягає в тому, що вони повідомляють один одному не тільки топологічну інформацію, але і текуче значення вільної пропускної спроможності у кожного ресурсу.
- Після того, як рішення знайдене, потрібно його реалізувати, тобто утілити в таблицях маршрутизації. На цьому етапі може виникнути проблема - в тому випадку, якщо ми хочемо прокласти ці маршрути в дейтаграммной мережі. Річ у тому, що таблиці маршрутизації цих мереж враховують тільки адреси призначення пакетів. Комутатори і маршрутизатори таких мереж (наприклад, IP-мереж) не працюють з потоками, для них потік в явному вигляді не існує, кожен пакет при його просуванні є незалежною одиницею комутації.