Загальні принципи управління мережними ресурсами в ТКС

**1. Якість обслуговування як мета мережного управління**

**1.1 Загальні поняття та визначення щодо якості обслуговування**

Основною метою мережного управління, в кінцевому результаті, є забезпечення заданних показників якості обслуговування (Quality of service, QoS). Підтримка якості обслуговування в сучасних, перш за все, мультисервісних ТКС є досить трудомістким завданням і вимагає узгодженого розв’язання цілого комплексу задач управління мережними ресурсами. Однією з найважливіших складових системи мережного управління є підсистема управління мережними ресурсами на рівнях транспортної мережі та мереж доступу до неї. Засоби управління мережними ресурсами на цих рівнях мають брати активну участь у процесі забезпечення заданих показників якості обслуговування та підвищення продуктивності ТКС в цілому на підставі:

* ведення та постійного оновлення єдиної бази даних щодо стану ТКС – її топології, завантаженості вузлів, трактів передачі та ін.;
* забезпечення високого рівня відмовостійкості мережі;
* реорганізації доступу до використання та збалансованого завантаження доступних мережних ресурсів;
* автоматизованого контролю параметрів трафіка користувачів у відповідності до укладеної умови щодо якості обслуговування (Service Level Agreement, SLA);
* раціональної організації та адаптивної зміни стратегій маршрутизації трафіка;
* реконфігурації режимів роботи мережного обладнання, в тому числі настроювання механізмів пріоритетної обробки пакетів на всіх або частині мережних вузлів.

Відповідно до рекомендацій ITU-Т Е.800 якість обслуговування (QoS) – це певна інтегральна оцінка, яка визначає ступінь задоволеності користувача наданою йому послугою зв'язку. Це визначення уточнене в рекомендації Е.860: "Якість обслуговування – ступінь відповідності обслуговування, надаваного користувачеві постачальником, угоді між ними". Це надає ще більшу важливість угодам (трафік-контрактам) SLA між користувачами й постачальниками послуг.

Фахівцями компанії Cіsco уведене таке визначення терміну "якість обслуговування" – "Здатність мережі забезпечити необхідний сервіс заданому трафіку в певних технологічних рамках (Frame Relay, ATM, Ethernet й 802.1 мережі, SONET і ІP мережі)". Відповідно до змісту RFC 2475 під сервісом варто розуміти набір характеристик передачі пакетів в одному напрямку одним або декількома мережними маршрутами.

Типове рішення завдань щодо QoS охоплює такі області:

- класифікацію аплікацій (applications) із призначенням пріоритетів (маркуванням) і диференціюванням трафіка;

- профілювання мережного трафіка;

- обмеження (за необхідністю) інтенсивності трафіка, який надходить від користувачів;

- управління чергами зі встановленням черговості обробки пакетів на мережних вузлах;

- маршрутизація мережного трафіка.

Вирішення задач щодо забезпечення QoS у рамках наведених областей, визначаючи свій набір управляючих механізмів і протоколів, має відбуватися взаємопов’язано та узгоджено. Наприклад, класифікація трафіка багато в чому визначає особливості вирішення інших завдань, таких як маршрутизація, розподіл канальних та буферних ресурсів і т.ін. Стандартизація технічних вимог щодо якості послуг у мережах зв'язку здійснюється на глобальному рівні Міжнародним союзом електрозв'язку (ІTU), на міжнародному регіональному рівні – Європейським інститутом телекомунікаційних стандартів (ETSІ), Асоціацією телекомунікаційної промисловості (TIA), Американським національним інститутом стандартів (ANSІ) та ін. Всі ці організації взаємодіють одна з одною при розробці стандартів, особливо коли мова йде про стандарти глобального характеру.

## 1.2 Класифікація показників якості обслуговування в ТКС

Точне визначення й параметри якості обслуговування переважно визначаються типом аплікації. Сервіс (якість обслуговування) описується рядом параметрів. Так, наприклад, для передачі голосового трафіка, найважливішими параметрами QoS є середня затримка та її варіація (джитер) на певному інтервалі часу, у той час як втрата деякої частини пакетів припустима. Параметри якості обслуговування можна розбити на три групи:

* параметри пропускної здатності (мінімальна, середня й максимальна швидкість передачі пакетів);
* параметри затримок передачі пакетів (середні й максимальні величини затримок і джитера);
* параметри надійності передачі (рівень втрат і викривлення пакетів).

Вимірювання зазначених параметрів здійснюється на певному інтервалі часу. Чим менше цей часовий інтервал, тим більше жорсткі вимоги ставляться до мережі, а отже, до всіх її елементів, оскільки забезпечення QoS "від краю до краю" вимагає взаємодії всіх вузлів на шляху пакетів трафіка й визначається надійністю, функціональністю й продуктивністю "слабкої ланки".

Для опису номінальної пропускної здатності середовища передачі інформації, протоколу або з'єднання ще досить часто використовується термін смуга пропускання (bandwidth).

Затримка при передачі пакета (packet delay) або латентність (latency) є сумарною величиною, що об’єднує в собі такі різновиди затримок: затримка серіалізації, затримка поширення, час чекання в черзі, затримка комутації (час передачі й обробки пакета), затримка, зв'язана з формуванням трафіка, мережна затримка. Описання і умови виникнення кожного з перерахованих типів затримки наведено в табл.

В умовах виникнення в мережі областей перевантажень, затримки, які пов'язані з очікуванням у черзі, починають переважати над іншими видами затримок і призводять до виникнення тремтіння (packet jitter) – різниці у величині сумарної затримки при передачі різних пакетів того ж самого потоку. Джитер дуже важливий, оскільки саме він визначає максимальну затримку при прийомі пакетів у кінцевому пункті призначення.

Рівень втрат пакетів (packet loss) визначається як частка пакетів, відкинутих мережею під час передачі протягом визначеного інтервалу часу. Основними причинами втрат пакетів є перевантаження мережі й ушкодження пакетів під час передачі лінією зв'язку. Найчастіше відкидання пакетів відбувається з першої причини – у місцях перевантаження, де кількість пакетів, що надходять, набагато перевищує верхню межу розміру вихідної черги. Крім того, відкидання пакетів може викликатися недостатнім розміром вхідного буфера.

Таблиця 1 – Складові затримки під час передачі пакетів

|  |  |  |
| --- | --- | --- |
| Різновид затримки | Визначення | Місце виникнення |
| Затримка серіалізації (serialization delay). Величина постійна для даного каналу передачі даних | Час, необхідний для передачі всіх біт пакета у фізичне середовище. Є функцією розміру пакета і швидкості каналу (ширини смуги пропускання) | На виході будь-якого фізичного інтерфейсу. Звичайно дуже мала для каналу ТЗ і вище |
| Затримка поширення (propagation delay). Величина постійна для даного фізичного каналу | Час, необхідний для досягнення переданому бітові інформації приймаючого пристрою на іншому кінці каналу. Залежить від середовища поширення (точніше від швидкості світла в даному фізичному середовищі) і від відстані | У будь-якому фізичному каналі. Звичайно дуже мала в каналах мереж рівня LAN і на коротких ділянках WAN мереж |
| Затримка в черзі (queuing delay). Величина непостійна | Час, витрачений пакетом на перебуванні в черзі в очікуванні подальшої передачі (вихідна черга) або в очікуванні можливості перетнути комутаційне поле (вхідна черга) (в очікуванні комутації) | Можлива на будь-якому вихідному інтерфейсі. Вхідні черги мало специфічні для маршрутизаторів, однак поширені в комутаторах мереж рівня LAN |
| Час пересилання або обробки (forwarding or processing delay). Величина непостійна | Час, необхідний для прийняття вхідного пакета і його обробки, доти поки пакет не буде поставлений у чергу для подальшої передачі | У будь-якому комутаторному обладнанні, включаючи маршрутизатори, комутатори LAN, Frame Relay, ATM |
| Затримка, пов'язана з формуванням трафіку (shaping delay). Величина непостійна | За умови здійснення формування трафіку, це час, на який пакети, що підлягають передачі, затримуються щоб уникнути втрат пакетів у середовищі Frame Relay або ATM | Скрізь, де включений механізм формування трафіку (shaping), найчастіше в маршрутизаторах, що передають пакети в мережі Frame Relay або ATM |
| Мережна затримка (network delay). Величина непостійна | Затримка, внесена компонентами абонентської мережі | У межах мережі провайдера послуг |

# 1.3 Угода про рівень обслуговування (SLA)

Додаткову складність при забезпеченні QoS викликає той факт, що користувачі оцінюють якість надаваних послуг суб'єктивно (наприклад, при перегляді відеоматеріалів користувача цікавить якість прийнятого зображення). Отже, вони оцінюють якість послуг "у цілому", а не у вигляді сукупності параметрів, значення яких найчастіше не розуміють. У зв'язку із цим оператори зв'язку мають потребу в універсальному способі домовлятися з користувачем про якість надаваних послуг, тобто методі, який би надав для оператора якість послуг з погляду користувача. Таким методом стала "угода про рівень обслуговування", що є важливим елементом, який регламентує роботу всіх механізмів QoS.

Відповідно до рекомендації ITU-Т Е.860, угода про рівень послуг – це формальна угода між двома й більше об'єктами права, яка була досягнута після узгодження характеристик послуги, відповідальності й пріоритету кожної зі сторін. Постачальник послуг має гарантувати, що користувач одержить, щонайменше, обговорений рівень QoS. SLA є контрактом на обслуговування між провайдером і користувачем, що гарантує певні характеристики якості наданих послуг. У договорі SLA визначаються:

* плата за обслуговування залежно від обраного рівня сервісу;
* параметри QoS для даного рівня (середня та (або) максимальна затримка та її варіації, пропускна здатність, максимальний час відновлення мережі після аварій і т.ін. );
* методи вимірювання вищевказаних параметрів;
* штрафні санкції за незабезпечення необхідного рівня QoS;
* будь-які інші додаткові статті за взаємною згодою.

У цілому концепція укладання угоди по QoS у службі ІP є ієрархічною моделлю (рис. 2).



Рисунок 2 – Концепція укладання договору щодо якості обслуговування

Договір SLA у моделі займає найвищий рівень абстракції в специфікації послуги. Параметризація трафіка також виділяється в самостійну підмножину SLS (Servіce Level Specіfіcatіon). При статичному SLS користувач може передавати дані в будь-який час. При динамічному SLS користувач повинен використати сигнальний протокол для запиту необхідних ресурсів мережі й обробки запитів SLS. Блок PHB (Per-Hop Behavіor) – є комбінацією функцій покрокової маршрутизації, класифікації, обробки черг і методів відкидання пакетів на кожному кроці. Блок PDB (Per-Domaіn Behavіor) реалізує концепцію покрокової маршрутизації на рівні домену.

Кожен тип PDB має кількісні і якісні атрибути, які можуть використовуватися для оцінки параметрів обслуговування пакетів заданого типу PDB. Атрибути обчислюються й виміряються практичним шляхом й описують такі параметри як швидкість примусового відкидання пакетів, пропускну здатність, верхнє й нижнє значення параметра затримки за деякий проміжок часу. Атрибути бувають довгострокові й короткострокові. Довгострокові характеризують агрегований трафік за досить великий проміжок часу, короткострокові описують допустимі флуктуації параметрів щодо середніх значень.

**2. Загальна характеристика існуючих мережних ресурсів**

телекомунікаційний якість обслуговування затримка

До мережних ресурсів належать як апаратні ресурси, наприклад пропускні здатності трактів передачі й ліній доступу, буферний простір й процесорний час обчислювальних засобів мережних вузлів, так й інформаційні ресурси – зміст баз даних про стан ТКС, трафік користувачів й ін.

**2.1 Ресурси мережних вузлів**

Сучасні мережні вузли – маршрутизатори забезпечують швидкість обробки пакетів до десятків терабіт за секунду (Тбіт/с). Високопродуктивні маршрутизатори, як правило, є гігабітними (Гбіт/с) з можливістю масштабування до терабітного діапазону. На сьогоднішній день терабітні швидкості забезпечують, наприклад, маршрутизатор TSR-40 від компанії Avici – 5.6 Тбіт/с. Одне шасі маршрутизатора TeraPlex 20 виробництва компанії Pluris забезпечує пропускну здатність в 150 Гбіт/с з масштабуванням до сумарної величини 44 Тбіт/с, а Pluris 2000 масштабується до 149 Гбіт/с у випадку одного пристрою й до 19.2 Тбіт/с при об'єднанні декількох пристроїв.

Платформа маршрутизації Routing Core Platform 7770 компанії Alcatel підтримує швидкість до 640 Гбіт/с на шасі, причому вісім шасі можна об'єднати в єдиний логічний пристрій, у результаті загальна продуктивність складе 20 Тбіт/с. Модель 8812 компанії Procket досягає, як стверджують виробники, продуктивності 960 Гбіт/с і, за попередньою оцінкою, 2 млрд пакетів за секунду. Це досягається за рахунок застосування придбаних або власних спеціалізованих інтегральних схем ASIC (Application-Specific Integrated Circuits), а також розроблених нею мікросхем надвеликого ступеня інтеграції VLSI (Very Large-Scale Integration). Це повністю програмувальні мережні процесори. За рахунок застосування спеціальних мікросхем ASIC маршрутизатор у змозі обробляти на кожній лінійній карті мільйони нових потоків у секунду. Компанія Chiaro Networks випускає маршрутизатори Enstara високої готовності, в основу якого покладена оптична комутуюча матриця – оптичний фазований масив (Optical Phased Array). При цьому згідно з даними виробника, час перемикання складає декілька наносекунд, масштабування здійснюється без втрат і забезпечується необхідна відмовостійкість на рівні 99,999% .

Крім того, компанія Cisco, наприклад, офіційно представила новий маршрутизатор CRS-1 (Carrier Routing System-1), робота над яким велася протягом чотирьох років (рис.3). Новинка орієнтована, у першу чергу, на великих операторів зв'язку, комп'ютерні системи, які щодня обробляють терабайти даних. Пристрій у мінімальній конфігурації має пропускну здатність в 1,2 Тбіт/с, однак за рахунок масштабування даний показник можна довести до 92 Тбіт/с. На такій швидкості весь вміст Бібліотеки Конгресу США можна передати всього за 4,6 секунди. В основі маршрутизатора лежать спеціалізовані процесори для обробки пакетів Sіlіcon Packet Processor (SPP), а також фірмове програмне забезпечення ІOS XR Software, створене на основі коду операційної системи Іnternetwork. На розробку CRS-1 компанія Cіsco витратила біля півмільярда доларів США. Подібних маршрутизаторів щороку продається на суму від 450 тисяч до 2 мільйонів доларів залежно від конфігурації.

**2.2 Ресурси пропускної здатності трактів передачі**

Сьогодні доступні технології WDM й DWDM, які забезпечують в трактах передачі швидкість на одну оптичну носійну 10 Гбіт/с. При використанні частотного плану DWDM стає можливою передача мультиплексного потоку на швидкості 400 Гбіт/с. Нові успіхи в науці й техніці мають сприяти досягненню теоретичного максимуму пропускної здатності 100 Tбіт/с, поширенню технології всехвильового оптичного волокна, застосування якого відкриває дотепер не використовувані властивості через присутність іонів води 1400 нм; нових технологій модуляції й техніки солітонної передачі (метод передачі шляхом спеціальних оптичних імпульсів). В сучасних комерційних лініях оптоволоконного зв’язку досягають пропускної здатності понад 1 Tбіт/с; у лабораторних умовах уже продемонстрована можливість передачі інформації зі швидкістю понад 10 Tбіт/с. Крім того, використання розподілених комбінованих підсилювачів й ербієвих підсилювачів дозволяє збільшити дальність зв'язку без переприйому до 200 км.

**3. Базова архітектура управління мережними ресурсами в ТКС**

Забезпечення наскрізного QoS "з кінця в кінець" (end-to-end) у рамках гетерогенної ТКС припускає використання таких засобів управління (рис.4):

1. Засоби управління, які реалізовані в окремих елементах мережі, наприклад в її вузлах, та забезпечують обробку пакетів відповідно до заданого рівня QoS:

* управління пріоритетами (класифікація й маркування пакетів на границі мережі);
* управління інтенсивністю вхідного та вихідного (профілювання) трафіка;
* управління пропускною здатністю (розподіл канальних ресурсів);
* управління перевантаженнями (запобігання перевантаження й превентивне відкидання пакетів).

1. Протоколи QoS-сигналізації та маршрутизації, які здійснюють координацію роботи мережних елементів.
2. Централізовані функції політики, управління й обліку QoS, які дозволяють адміністраторам мережі цілеспрямовано впливати на мережні елементи для поділу ресурсів мережі між різними видами трафіку відповідно до заданого рівня QoS.



Рисунок 4 – Базова архітектура управління мережними ресурсами

Централізовані функції політики, управління й обліку QoS не є необхідним засобом архітектури управління, але дуже бажані у територіально-розподілених мережах. З огляду на те, що кожен користувач і кожна аплікація прагнуть одержати обслуговування з максимально високим рівнем якості (наприклад, пропускної здатності), то мають існувати засоби, за допомогою яких адміністратор міг би виконувати функцію арбітра й задавати раціональний рівень якості обслуговування для груп або окремих користувачів і аплікацій. Функції політики дозволяють адміністраторові створювати правила, за якими мережні пристрої можуть формально, на підставі набору ознак, розпізнавати окремі типи трафіку й застосовувати до них певні функції управління.

Що стосується єдності політики, то її правила й засоби можуть конфігуруватися окремо в кожному мережному пристрої, або ж бути централізованими. Централізація засобів політик управління та забезпечення QoS припускає єдність правил політики, справедливих для всіх пристроїв мережі, і використання для їхнього зберігання на сервері політик. У цьому випадку адміністратор конфігурує правила політики в одній точці, потім за допомогою спеціального протоколу ці правила поширюються всіма мережними пристроямми, що підтримує якість обслуговування, а мережні пристрої застосовують політику для формування (профілювання) трафіку й управління чергами відповідно до встановлених QoS-вимог.

Служби QoS, в яких працюють централізовані системи підтримки політики, називаються службами QoS, заснованими на політиці (policy-based QoS). Правила політики корисні не тільки для управління QoS, але і для координації мережних пристроїв при виконанні інших функцій, наприклад, функцій захисту трафіка. Тому централізована система політики мережі звичайно базується на загальній довідковій службі мережі (Directory Service), що традиційно зберігає всі облікові дані про користувачів (ім'я – пароль).

Крім перерахованого, слід окремо виділити засоби підвищення ефективності використання канальних ресурсів (Link Efficiency Mechanisms) – фрагментація (дефрагментація) пакетів й їхнє стиснення, використання яких у мережі сприяє підвищенню якості надаваних мережею послуг. Обов'язковим елементом QoS-мережі є наявність механізмів управління доступом (Call Admission Control, CAC), що покликані захистити мережу від перевантажень шляхом запобігання надходженню в мережу трафіка в обсязі більшому, ніж мережа здатна передати. Відмовляючи в обслуговуванні новим потокам, які запитують певний сервіс, механізм САС не тільки захищає трафік даних від відкидання, але і забезпечує стабільну якість передачі вже прийнятих до обслуговування потоків.

Серед названих функцій QoS класифікація, маркування пакетів і управління інтенсивністю реалізуються в приграничних вузлах мережі. Управління інтенсивністю (формування трафіка) реалізується з метою зміни параметрів вхідного трафіка відповідно до заданого профілю, зазначених у SLA.

Відзначимо, що укладання між користувачем і провайдером (оператором зв'язку) угоди про рівень обслуговування (SLA), у якій чітко обумовлюються вимоги щодо розглянутих параметрів мережного з'єднання – смузі пропускання, затримки, джитера і припустимого рівня втрат, а також деяких інших параметрів потоку, є необхідною умовою забезпечення мережею належної якості обслуговування. SLA є свого роду трафік-контрактом.

Основними функціями формування трафіка є його вирівнювання (traffic shaping) і обмеження (traffic policing). Вирівнювання трафіка дозволяє усунути сплески і тим самим зменшити імовірність втрати пакетів даних. Обмеження трафіка полягає у відкиданні пакетів, що не задовольняють заданим параметрам, і здійснюється, наприклад, за допомогою механізму узгодження швидкості доступу (Committed Access Rate, CAR). В основу задач вирівнювання й обмеження трафіка покладені алгоритми дозування трафіка, які мають назви «кошик маркерів» (token bucket) і «діряве відро» (leaky bucket), а також їхні різновиди.

Управління смугою пропускання може здійснюватися шляхом нормування швидкості TCP (TCP rate shaping), що зводиться до перехоплення і маніпулювання розмірами TCP-вікна, або за допомогою механізмів організації й обслуговування черг. Механізм обслуговування черг шляхом регулювання порядку обслуговування пакетів певного потоку (класу) трафіка дозволяє варіювати частоту їхньої обробки і тим самим виділяти визначену смугу пропускання даному потокові (класові). Для забезпечення QoS механізм обслуговування черг повинен мати можливість диференціювання різних потоків і визначення рівня їхніх вимог до обробки. Прикладами механізмів, здатних забезпечити необхідну смугу пропускання в мережах IP, є зважений механізм рівномірного обслуговування черг WFQ (Weighted Fair Queuing), зважений механізм рівномірного обслуговування черг на основі класу CBWFQ (Class-Based Weighted Fair Queuing) і на основі потоку (Flow-Based Distributed Weighted Fair Queuing).

Необхідною умовою для забезпечення мережею гарантованого рівня обслуговування є відсутність у ній перевантажень, тобто станів, при яких мережа не здатна забезпечити узгоджені параметри існуючих з'єднань. Запобігання перевантажень і їхнє усунення в мережах АТМ відповідно до рекомендації ITU–T I.371 і специфікацією форуму ATM „Traffic Management Specification” визначені як основні задачі управління трафіком (Traffic Management), компонентами якого є управління доступом до мережі, управління пріоритетами, формування трафіка, розподіл ресурсів, відкидання пакетів.

Алгоритм відкидання пакетів визначає спосіб регулювання довжини черги у випадку виникнення перевантаження або при наближенні до цього стану. Перший випадок відповідає механізмові обслуговування черг FIFO, що передбачає відкидання усіх вхідних пакетів при досягненні чергою свого максимального значення. В другому випадку задіються активні механізми управління чергами, що дозволяють запобігти перевантаженню шляхом превентивного відкидання пакетів і тим самим попередити джерело про можливе перевантаження. Прикладами активних механізмів управління чергами є алгоритм довільного раннього виявлення RED (Random Early Detection), зважений алгоритм довільного раннього виявлення WRED (Weighted RED). Запобігання перевантаженню в мережах IP можливе за допомогою алгоритму явного повідомлення про перевантаження ECN (Explicit Congestion Notification), а також шляхом управління розмірами TCP-вікна.

Необхідною умовою гарантованого забезпечення QoS є наявність протоколу, що дозволяє аплікаціям проводити сигналізацію своїх QoS-вимог. Протоколи сигналізації QoS дозволяють механізмам QoS окремих вузлів обмінюватися службовою інформацією, що сприяє координації зусиль щодо забезпечення параметрів якості обслуговування уздовж всього шляху проходження потоку, тобто "з кінця в кінець". Яскравим прикладом протоколів даної групи є протокол RSVP (ReSerVation Protocol), призначений для резервування мережних ресурсів для кожного потоку на всіх маршрутизаторах, через які здійснюється доставка інформації, відповідно до запитуваного рівня QoS. RSVP сигналізує про запити резервування ресурсів доступним шляхом в мережі. При цьому велику роль у забезпеченні QoS відіграють протоколи маршрутизації, що визначають шлях, уздовж якого проводитиметься резервування. У світлі вищевикладеного перехід від традиційної маршрутизації найкоротшим шляхом до маршрутизації, що враховує при виборі маршруту QoS-вимоги потоку і наявність мережних ресурсів уздовж усього маршруту доведення (маршрутизація з підтримкою QoS), дозволить значно підвищити можливості мережі в плані гарантованого забезпечення QoS.

Ініціювати роботу протоколу сигналізації може як кінцевий вузол (як в RSVP), так і проміжний пристрій. Наприклад, прикордонний маршрутизатор сервіс-провайдера здатен виконати класифікацію трафіка і зарезервувати в мережі провайдера даному потокові деяку пропускну здатність. У цьому випадку координація мережних пристроїв відбуватиметься не на всьому шляху проходження трафіка, а тільки в межах мережі даного провайдера, що, звичайно, знижує якість обслуговування трафіка.

Один із примітивних засобів сигналізації є маркування пакета позначкою, яка несе інформацію про необхідний для пакета рівень якості обслуговування. Найчастіше з цією метою використовується поле пріоритету (у пакеті IPv4 перші три біти поля Type оf Service, TоS). Маючи таку позначку, пакет переносить уздовж шляху проходження свої вимоги до якості обслуговування, щоправда, у досить узагальненій формі – у зв’язку з тим, що поле пріоритету має усього кілька можливих значень, якість обслуговування надаватиметься диференційовано за декількома агрегованими потоками мережі.

Як саме перераховані механізми управління впливають на параметри з'єднання, показано в табл. 2 – 4.

Таблиця 2 – Засоби QoS , що впливають на смугу пропускання

|  |  |
| --- | --- |
| Засіб QoS | Механізм впливу |
| Стиснення | Стиснення корисної інформації або заголовка, що зменшує загальну кількість біт, які очікують на передачу |
| CAC | Зменшення загального вхідного навантаження в мережі за допомогою відкидання нових запитів |
| Організація черг | Може використовуватися з метою резервування мінімальної кількості смуги пропускання для окремих типів трафіка |

Таблиця 3 – Засоби QoS , що впливають на втрати

|  |  |
| --- | --- |
| Засіб QoS | Механізм впливу |
| Організація черг | Підтримка довших черг збільшує затримку, але запобігає втратам |
| Випадкове відкидання пакетів | Механізм випадкового відкидання пакетів при наближенні розміру черги до точки відмови сповільнює деякі з'єднання TCP. Це зменшує загальне завантаження, скорочує переповнення черги, погіршуючи якість обслуговування (збільшується час відповіді) тільки одного користувача |

Таблиця 4 – Засоби QoS , що впливають на затримку і джитер

|  |  |
| --- | --- |
| Засіб QoS | Механізм впливу |
| Організація черг | Дозволяє розмістити пакети так, щоб чутливі до затримки пакети, наприклад, мовні, були оброблені швидше в порівнянні з менш чутливими, наприклад, пакетами даних |
| Фрагментація пакетів і чергування | Фрагментації великих пакетів даних та їхнє чергування з невеликими за розмірами, але чуттєвими до затримки мовними пакетами, дозволяє уникнути для останніх очікування в черзі, зменшуючи тим самим сумарну затримку |
| Стиснення | Стиснення корисної інформації або заголовка зменшує загальну кількість біт, що підлягають передачі каналом. Вимагаючи тепер меншої кількості смуги пропускання, черги «стискаються», що у свою чергу зменшує затримку. Крім цього, фрагментація пакетів сприяє зменшенню затримки стиснення, тому що обробляється менша кількість біт. Процес стиснення додає деяку затримку обробки |
| Вирівнювання трафіка | Затримка штучно збільшується, тому що при вирівнюванні трафіка використовується буферизація, яка здійснюється з метою зменшення втрат усередині мереж |

Для організації ефективного управління мережею і забезпечення гарантованого QoS необхідними умовами є:

- реалізація QoS «зверху вниз»;

- реалізація QoS «від краю до краю».

Перша умова вимагає узгоджену взаємодію засобів різних рівнів еталонної семирівневої моделі, тому що задача забезпечення гарантованого QoS стосується всіх аспектів функціонування ТКС, не будучи при цьому задачею якого-небудь одного рівня. Тому цілісне розв’язання задачі забезпечення гарантованого QoS можливе тільки шляхом взаємоузгодженого розв’язання підзадач окремих рівнів – від фізичного до прикладного. Використання в рамках однієї мережі різних телекомунікаційних технологій і протоколів, наскрізна взаємодія яких починається, як правило, з мережного рівня, дозволяє зробити висновок про те, що підтримка наскрізного гарантованого QoS у таких мережах найефективніше забезпечується методами мережного рівня, тому в даній дисципліні основну увагу буде зосереджено на забезпеченні QoS засобами саме мережного рівня.

Друга умова вимагає наявності механізмів QoS у мережах різних рівнів ієрархії (оператора зв'язку і підприємства, LAN і WAN), а також їхню координацію.

**4. Рівні якості обслуговування і відповідні архітектурні моделі**

QoS можна розглядати як міру якості передачі і доступності сервісу в мережі, а згідно з цим можна виділити три рівні якості обслуговування: від відсутності всіляких гарантій щодо якості обслуговування до їхнього твердого забезпечення за всіма параметрами передачі. Виділяють такі рівні якості обслуговування:

- негарантована доставка даних (best-effort service);

- сервіс з перевагою (soft QoS) або диференційоване обслуговування (Differentiated Service, DiffServ);

- гарантоване обслуговування (hard QoS).

Негарантована доставка даних (best-effort service) є доставкою за можливістю, яка передбачає тільки забезпечення зв’язності вузлів мережі, і не гарантує забезпечення будь-яких показників якості обслуговування. Негарантована доставка пакетів є на сьогоднішній день єдиною послугою, яка підтримується в Internet. Незважаючи на деяке зниження продуктивності, для більшості аплікацій, що орієнтовані на передачу інформації (наприклад, аплікацій, які забезпечують взаємодію за FTP), ця послуга є цілком достатньою.

Забезпечення QoS у випадку обслуговування з перевагою (DiffServ) досягається шляхом диференціювання вхідного трафіка на підставі його вимог до якості обслуговування і відповідного присвоєння пріоритетів. Ця схема забезпечує обслуговування одних класів трафіка на більш високому рівні в порівнянні з іншими, однак саме по собі диференційоване обслуговування не забезпечує гарантій щодо якості наданих послуг («м'який» QoS). Диференційоване обслуговування зручне для застосовування в мережах з інтенсивним трафіком аплікацій. У цьому випадку важливо забезпечити відокремлення адміністративного трафіка мережі від всього іншого трафіка і призначити йому пріоритет, що дозволяє в будь-який момент часу бути упевненим у зв’язності вузлів мережі.

Гарантоване обслуговування передбачає забезпечення твердих гарантій щодо виконання вимог потоків трафіка. Це досягається шляхом попереднього резервування мережних ресурсів уздовж шляхів доведення відповідно до запитуваного рівня обслуговування в термінах розглянутих вище параметрів передачі. Резервування мережних ресурсів дозволяє зменшити величину затримки передачі пакета за рахунок зменшення затримки комутації і звести до мінімуму рівень втрат пакетів.

Відповідно до описаних рівнів QoS були розроблені архітектурні моделі QoS, що базуються вже на конкретних механізмах і протоколах. Існує дві основні архітектурні моделі: інтегрованих послуг (Integrated Service, IntServ) і диференційованих послуг (Differentiated Service, DiffServ), де перша реалізує рівень гарантованого обслуговування (hard QoS), друга – однойменний рівень диференційованого обслуговування (soft QoS).

Архітектура IntServ з'явилася першою в 1994 році (RFC 1633) і передбачала надання рівня «hard» QoS шляхом резервування ресурсів для кожного потоку індивідуально (рис. 5). Незважаючи на очікувану високу якість обслуговування архітектура не одержала поширення на практиці, що пов’язано з її низкою масштабованістю. У відповідь на недоліки IntServ з'явилася архітектура DiffServ (RFC 2474, RFC 2475, 1998 рік), що передбачає агреговане обслуговування потоків (потоки з однаковими вимогами до сервісу поєднуються в один клас, що обслуговується як єдине ціле). Хоча модель DiffServ не здатна за своєю суттю забезпечити твердих гарантій щодо якості передачі, вона одержала поширення на практиці. Однак цей недолік усе-таки сприяв пошуку нових рішень в області QoS. Це рішення знайдене в спільному використанні IntServ і DiffServ у рамках технології MPLS (MPLS DiffServ Aware Traffic Engineering), де, як у DiffServ, передбачається агрегація потоків у класи, для яких, як у IntServ, резервується смуга і до яких застосовуються функції управління доступом і маршрутизація з урахуванням обмежень, як у Traffic Engineering.