

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ АВІАЦІЙНИЙ УНІВЕРСИТЕТ
ФАКУЛЬТЕТ АЕРОНАВІГАЦІЇ,
ЕЛЕКТРОНІКИ ТА ТЕЛЕКОМУНІКАЦІЙ
КАФЕДРА ТЕЛЕКОМУНІКАЦІЙНИХ ТА РАДІОЕЛЕКТРОННИХ СИСТЕМ**

ДОПУСТИТИ ДО ЗАХИСТУ
Завідувач кафедри

Роман ОДАРЧЕНКО
“ ” 2022 р.

**КВАЛІФІКАЦІЙНА РОБОТА
(ПОЯСНЮВАЛЬНА ЗАПИСКА)**

ВИПУСКНИКА ОСВІТНЬОГО СТУПЕНЯ МАГІСТР

Тема: «Методи контролю навантажень SIP-серверів у мережах NGN».

Виконавець: _____ Іван МАКАРОВ
(підпис)

Керівник: _____ Ірина ТЕРЕНТЬЄВА
(підпис)

Консультанти з окремих розділів пояснювальної записки:

Консультант розділу «Охорона праці» _____ Батир ХАЛМУРАДОВ
(підпис)

Консультант розділу «Охорона навколишнього середовища»
_____ Євгеній БОВСУНОВСЬКИЙ
(підпис)

Нормоконтролер: _____ Денис БАХТІЯРОВ
(підпис)

Київ 2022

НАЦІОНАЛЬНИЙ АВІАЦІЙНИЙ УНІВЕРСИТЕТ

Факультет аеронавігації, електроніки та телекомунікацій

Кафедра телекомунікаційних та радіоелектронних систем

Спеціальність 172 «Телекомунікації та радіотехніка»

Освітньо-професійна програма 172 «Телекомунікаційні системи та мережі»

ЗАТВЕРДЖУЮ

Завідувач кафедри

Роман ОДАРЧЕНКО

“ _____ ” _____ 2022 р.

ЗАВДАННЯ на виконання дипломної роботи

Макарова Івана Сергійовича

(прізвище, ім'я, по батькові випускника в родовому відмінку)

1. Тема дипломної роботи(проекту): «Методи контролю навантажень SIP-серверів у мережах NGN».

затверджена наказом ректора від «07» вересня 2022 р. №1321/ст

2. Термін виконання роботи: з 05.09.2022 р. по 30.11.2022 р.

3. Вихідні дані до роботи: NGN, Протокол SIP, VoIP, Round Robin, IP.

4. Зміст пояснювальної записки:

- Розділ 1 «Мережі NGN. Протокол SIP.»

- Розділ 2 «Технології в NGN»

- Розділ 3 «SIP-сервер. Навантаження на SIP-сервер. Контроль навантажень на SIP-сервери.»

- Розділ 4 «Дослідження методу балансування навантаження за кількістю транзакцій на сервері»

- Розділ 5 «Охорона праці»

- Розділ 6 «Охорона навколишнього середовища»

5. Перелік обов'язкового графічного (ілюстративного) матеріалу: 10 рисунків.

6. Календарний план-графік:

№ пор.	Завдання	Термін виконання	Відмітка про виконання
1	Розробити деталізований зміст розділів диплому	05.09.2022- 06.09.2022	Виконано
2	Вступ	07.09.2022- 10.09.2022	Виконано
3	«Мережі NGN. Протокол SIP.»	12.09.2022- 05.10.2022	Виконано
4	«Технології в NGN»	06.10.2022- 15.10.2022	Виконано
5	«SIP-сервер. Навантаження на SIP-сервер. Контроль навантажень на SIP-сервери.»	17.10.2022- 25.10.2022	Виконано
6	«Дослідження методу балансування навантаження за кількістю транзакцій на сервері»	25.10.2022- 05.11.2022	Виконано
7	Охорона праці	05.11.2022- 14.11.2022	Виконано
8	«Охорона навколишнього середовища»	14.11.2022- 19.11.2022	Виконано
9	Усунення недоліків дипломної роботи	21.11.2022- 30.11.2022	Виконано

7. Консультанти з окремих розділів

		Дата, підпис
--	--	--------------

Розділ	Консультант (посада, П.І.Б.)	Завдання видав	Завдання прийняв
Охорона праці	к.м.н., проф. Батир ХАЛМУРАДОВ		
Охорона навколишнього середовища	к.т.н., доц. Євгеній БОВСУНОВСЬКИЙ		

8. Дата видачі завдання: “22” серпня 2022 р.

Керівник кваліфікаційної роботи _____
(підпис керівника)

Ірина ТЕРЕНТЬЄВА
(П.І.Б.)

Завдання прийняв до виконання _____
(підпис випускника)

Іван МАКАРОВ
(П.І.Б.)

РЕФЕРАТ

Кваліфікаційна робота «Методи контролю навантажень SIP-серверів у мережах NGN» містить 80 сторінок, 10 рисунків, 13 використаних джерел.

SIP, SIP-СЕРВЕР, NGN, КОНТРОЛЬ НАВАНТАЖЕННЯ

Об'єкт дослідження – Методи контролю навантажень SIP-серверів.

Предмет дослідження – SIP-сервер.

Мета кваліфікаційної роботи – Дослідити методи контролю навантажень SIP-серверів.

Метод дослідження – порівняльний метод, статистичний аналіз, структурний аналіз, прийоми аналізу наукової літератури.

ЗМІСТ

КВАЛІФІКАЦІЙНА РОБОТА.....	1
ЗАВДАННЯ.....	2
РЕФЕРАТ.....	5
ПЕРЕЛІК СКОРОЧЕНЬ, УМОВНИХ ПОЗНАЧЕНЬ.....	9
ВСТУП.....	10
РОЗДІЛ 1	12
МЕРЕЖІ NGN. SIP-СЕРВЕРИ.	12
ПРОТОКОЛ SIP.	12
1.1. Мережі NGN	12
1.2. Архітектура NGN	14
1.3. Рішення в мережах NGN.....	16
1.4. Протокол SIP	19
ВИСНОВКИ ДО РОЗДІЛУ 1.....	24
РОЗДІЛ 2	25
ТЕХНОЛОГІЇ NGN	25
2.1. Next Generation Network.....	25
2.2. Протоколи мереж NGN	29
2.2.1. Базові протоколи стеку TCP/IP	30
2.2.2. Протокол IP.....	30
2.2.3. Протокол ICMP.....	31
2.2.4. Протокол TCP	31
2.2.5. UDP	32
2.2.6. Сигнальні протоколи SIP та H.323	32
2.2.7. Протокол MGCP	34
ВИСНОВКИ ДО РОЗДІЛУ 2.....	34
РОЗДІЛ 3	35
SIP-СЕРВЕР. НАВАНТАЖЕННЯ НА SIP-СЕРВЕР.....	35

3.1. SIP-сервер	35
3.1.1. Принцип роботи SIP-серверу та його переваги	38
3.2. Перевантаження SIP-серверу.....	40
ВИСНОВКИ ДО РОЗДІЛУ 3.....	44
РОЗДІЛ 4	45
КОНТРОЛЬ НАВАНТАЖЕНЬ НА SIP-СЕРВЕРИ. ДОСЛІДЖЕННЯ МЕТОДУ БАЛАНСУВАННЯ НАВАНТАЖЕННЯ ЗА КІЛЬКІСТЮ ТРАНЗАКЦІЙ НА СЕРВЕРІ	45
4.1. Аналіз схем балансування навантаження у кластері SIP серверів	45
4.1.1. За значенням ідентифікатора.....	47
4.1.2. Round Robin.....	47
4.1.3. Weighted Round Robin	48
4.1.4. Destination Hash Scheduling та Source Hash Scheduling.....	48
4.1.5. Sticky Sessions.....	48
4.2. Розробка методу балансування навантаження з урахуванням транзакцій.....	49
4.3. Дослідження ефективності методу.....	52
ВИСНОВКИ ДО РОЗДІЛУ 4.....	53
РОЗДІЛ 5	55
ОХОРОНА ПРАЦІ.....	55
5.1. Опис робочого місця та умови праці співробітника ІТ-відділу.....	55
5.2. Перелік шкідливих та небезпечних чинників, що діють на суб'єкта. 56	
5.2.1. Мікроклімат (підвищена або занижена вологість, швидкість руху повітря)	57
5.2.2. Підвищене значення електричного струму, що може проходити через людину і становити небезпеку для нього	57
5.2.3. Електромагнітне випромінювання	58
5.2.4. Розрахунок швидкості руху повітря в приміщенні ІТ-відділу, враховуючи основні санітарно-гігієнічні параметри.....	59
5.3. Пожежна безпека.....	60
5.3.1. Правила евакуації з приміщень.	61
ВИСНОВОК ДО РОЗДІЛУ 5.....	62

РОЗДІЛ 6	64
ОХОРОНА НАВКОЛИШНЬОГО СЕРЕДОВИЩА	64
6.1. Завдання системи екологічного управління	64
6.2. Система заходів захисту здоров'я користувачів корпоративною системою Інтернет мережі.	68
ВИСНОВКИ ДО РОЗДІЛУ 6.....	74
ВИСНОВКИ	75
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	79

ПЕРЕЛІК СКОРОЧЕНЬ, УМОВНИХ ПОЗНАЧЕНЬ

АТС - Автоматична телефонна станція

VoIP - Voice over IP

IP - Internet Protocol

SIP - Session Initiation Protocol

QoS - Quality of Service

NGN - New Generation Network

MSAN - Multi-Service access node

ATM - Asynchronous Transfer Mode

RTP - Real-time Transport Protocol

CoS - Class of Service

SDP - Session Description Protocol

FIFO - First In, First Out

LIFO - Last In, First Out

RR - Round Robin

TCP - Transmission Control Protocol

UDP - User Datagram Protocol

ICMP - Internet Control Message Protocol

MGCP - Media Gateway Control Protocol

ВСТУП

В даний час телекомунікаційні технології знаходяться на такому високому рівні розвитку, що впроваджуються в будь-які пристрої, починаючи з найпростіших комп'ютерних модемів, що забезпечують доступ в Інтернет, і закінчуючи системами дистанційного керування автомобілем або побутовою технікою, реалізованими за допомогою мобільних телефонів (смартфони). На фоні бурхливого розвитку телекомунікацій змінюються основи будь-якого виду зв'язку - змінюються телефонні мережі, способи доступу до них, методи і системи проектування АТС, які забезпечують доступ абонентів до послуг зв'язку. Особливо цікаво розглянути можливості проектування і створення АТС, що використовуються повсюдно в самих різних компаніях - починаючи від міні АТС, до послуг яких вдаються невеликі компанії, стартапи, закінчуючи потужними комунікаційними структурами, які допомагали національним і міжнародним організаціям, офіси яких розташовуються на великій відстані один від одного.

Постійне вдосконалення і поширення в усіх областях мережі Інтернет призвело до зародження нового виду зв'язку, який ґрунтується на використанні протоколів Інтернет і IP - серверів, здатних перетворювати в цифрову форму звичайний голосовий сигнал. Цей новий вид зв'язку назвали IP або VoIP телефонією. даний вид технології дозволив перетворити аудіо- та відео-переговори, у зручний, якісний, універсальний і що також важливо дешевий засіб спілкування, який став доступним практично кожному. Тож, пропоную детальніше розглянути, як це працює зсередини.

Мета кваліфікаційної роботи полягає в дослідженні методів збільшення ефективності та пропускну здатності SIP-серверів у мережах нового покоління.

Для досягнення мети даного дослідження були поставлені такі **завдання:**

- Дослідити розвиток мереж нового покоління та IP-телефонії в них
- Проаналізувати можливі методи контролю навантажень на SIP-сервери
- Здійснити покращення показів ефективності та пропускну здатності SIP-серверу удосконаленням одного з методів контролю навантаження на сервер.

Об'єктом дослідження є методи забезпечення контролю навантажень на SIP-сервер.

Предметом дослідження є SIP-сервер у мережі NGN.

Методи досліджень: порівняльний метод, статистичний аналіз, структурний аналіз, прийоми аналізу наукової літератури, порівняльний метод.

Практичне значення отриманих результатів матеріали дипломної роботи рекомендується використовувати при вивченні протоколів, що використовуються в IP-телефонії. Вони можуть використовуватися студентами для аудиторної та самостійної роботи.

Структура роботи визначена метою і завданнями дослідження. Дипломна робота складається зі вступу, шести розділів, висновків та списку використаних джерел. У вступі визначено мету і завдання дослідження, які використовуються методи, описана актуальність, об'єкт і предмет, новизна, практична значущість роботи.

РОЗДІЛ 1

МЕРЕЖІ NGN. SIP-СЕРВЕРИ.

ПРОТОКОЛ SIP.

1.1. Мережі NGN

NGN — це MSAN, в яку можна інтегрувати сервіси передавання голосу, даних та мультимедіа. Вона базована на IP-мережі. В цій мережі передбачено розділену комутацію на сигнальну інформацію(для комутації абонентів та використання сервісів) та дані користувача(інформація спрямована клієнту, голос, відео, тощо). В цих двох складових інформації шляхи проходження не зобов'язані збігатись.[1]

До характеристик такої мережі можна віднести:

- передачу інформації з пакетною комутацією, спосіб передачі даних, у якому передане повідомлення розбивається кілька спеціально оформлених порцій - пакетів, кожен із яких передається незалежно. При цьому пакети можуть переміщатися від передавача до одержувача різними шляхами. На відміну від каналної комутації у мережу працює за принципом «збережи та передай». Перш ніж передати далі, вузол мережі зберігає прийнятий пакет. При такому способі передачі є кілька переваг, оскільки він дає можливість краще використовувати вільні ресурси, розділяючи їх між різними одночасними сеансами передачі. При пакетній комутації відсутня затримка пов'язана із створенням сеансу зв'язку; [1]

- поділ функцій управління між пропускнуою спроможністю каналу-носія, викликом/сеансом, а також додатком/послугами;

- розв'язка між наданням послуг та транспортуванням та надання відкритих інтерфейсів;

- підтримка широкого спектру послуг, додатків та механізмів на основі уніфікованих блоків обслуговування (включаючи послуги у реальному

масштабі часу, у потоковому режимі, в автономному режимі та мультимедійні послуги);

- можливості широкосмужової передачі даних із наскрізною функцією QoS (якості обслуговування), або ж Broadband – це технологія передачі даних по мережі (можна використовувати як мідний, так і оптико-волоконний кабель), за якої дані передаються у вигляді модульованих радіочастотних сигналів. Також існує широкосмужова передача в стільникових мережах, тобто mobile broadband, що дозволяє передавати бездротовими каналами як цифрові дані, так і голос; [1]

- взаємодія з існуючими мережами за допомогою відкритих інтерфейсів;

- необмежений доступ користувачів до різних постачальників послуг;

- різноманітність схем ідентифікації;

- зближення послуг між фіксованим(провідним, стаціонарним) та рухомим(безпровідним, мобільним) зв'язком;

- незалежність пов'язаних з обслуговуванням функцій від використовуваних технологій транспортування;

- підтримка різних технологій "останньої милі", це канал, що з'єднує кінцеве (клієнтське) обладнання із вузлом доступу провайдера (оператора зв'язку). Наприклад, при наданні послуги підключення до мережі Інтернет — це ділянка від порту комутатора провайдера на вузлі зв'язку до порту маршрутизатора клієнта в його офісі. Для послуг комутованого (dial-up, дайлапного) підключення остання миля — це ділянка між модемом користувача та модемом (модемним пулом) провайдера. В останню милю зазвичай не включається розведення проводів усередині будівлі. До технологій останньої милі зазвичай відносять xDSL, FTTx, Wi-Fi, WiMax, DOCSIS, зв'язок PLC;

- виконання всіх регламентуючих вимог, наприклад, для аварійного зв'язку, захисту інформації, конфіденційності, законного перехоплення тощо.

[2]

А щодо задач NGN, згідно з міжнародними рекомендаціями такі мережі повинні виконувати такі функції як:

- сприяти чесній конкуренції;
- заохочувати приватні інвестиції;
- визначати принципи архітектури та можливості для приведення у відповідність до різних регламентуючих вимог;
- забезпечувати відкритий доступ до мереж;
- забезпечувати універсальне надання послуг та доступ до них;
- сприяти забезпеченню рівних можливостей для населення;
- сприяти різноманітності змісту, включаючи культурне та мовне розмаїття.[3]

В такій транспортній платформі, як NGN поєднано різноманітні види послуг і її використання доволі сильно розширює спектр послуг, які можливо надати абоненту. У разі проектування мережі наступного покоління слід прийняти до уваги один з важливіших факторів — забезпечення гідної якості обслуговування. З початком використання MSAN здебільшого став _підхід до завдання рівня обслуговування на підставі вимог самих клієнтів до якості послуг. Тобто створення покращеної системи керування якістю наданих послуг в NGN є більш ніж актуальним.[3]

1.2. Архітектура NGN

Однією з основних відмінностей концепції NGN від мережевих інфраструктур, що реалізуються до цього, є перехід до принципово іншої функціональної моделі. У класичній ТФОП основними функціональними елементами були вузли доступу та вузли комутації різного рівня. При цьому обладнання вузла комутації вирішувало одночасно кілька завдань: комутацію потоків користувальницької інформації, обробка виклику та надання послуг. Реалізація інтерфейсів між цими функціями була внутрішньою справою виробника системи комутації і підлягала регламентації. Розвиток класичної

ТФОП, пов'язане, перш за все, з появою технології ISDN, дозволило дещо розділити функції обробки сигналізації та комутації потоків інформації користувача. Як результат, з'явилися нові функціональні елементи, такі як виділені пункти транзиту сигналізації (STP), а топологія сигнальної мережі почала відрізнятися від топології мережі комутації. З іншого боку, сигнальна мережа вирішувала питання транзиту сигнальних повідомлень, але завдання обробки інформації рівня ISUP, отже, управління комутацією, у разі вирішувалася у точці збігу топологій, т. е. в системах комутації.[14]

Концепція NGN, в першу чергу, характеризується чітким поділом трьох рівнів з'єднання відповідно до їх функціональних завдань: для комутації та передачі мовної інформації використовується транспортний функціональний рівень, для передачі інформації сигналізації – рівень сигналізації, а надання послуг, відмінних від базових, здійснюється з боку рівня послуг. У цьому між рівнями визначено інтерфейси, які є об'єктом стандартизації. Здобувши подібну незалежність один від одного, рівні надалі можуть розвиватися самостійно. Більше того, з погляду адміністративного поділу мережі може порушуватися питання про те, щоб послуги різних рівнів надавалися різними операторами.[14]

Іншою особливістю інфраструктури NGN є використання універсальних технологій транспортної мережі, що базуються на технології пакетної комутації. У класичних мережах надання послуг ТФОП базувалося на технології комутації каналів, а надання послуг доступу до мереж передачі даних і даних передбачало або формування нової транспортної структури, або неефективне використання існуючого транспорту мережі з комутацією каналів. Тоді як у мережах NGN пакетні технології, визначені передачі даних, використовуються для надання всіх видів послуг.

Архітектуру мережі зв'язку, побудованої відповідно до концепції NGN, показано на рисунку 1.1.

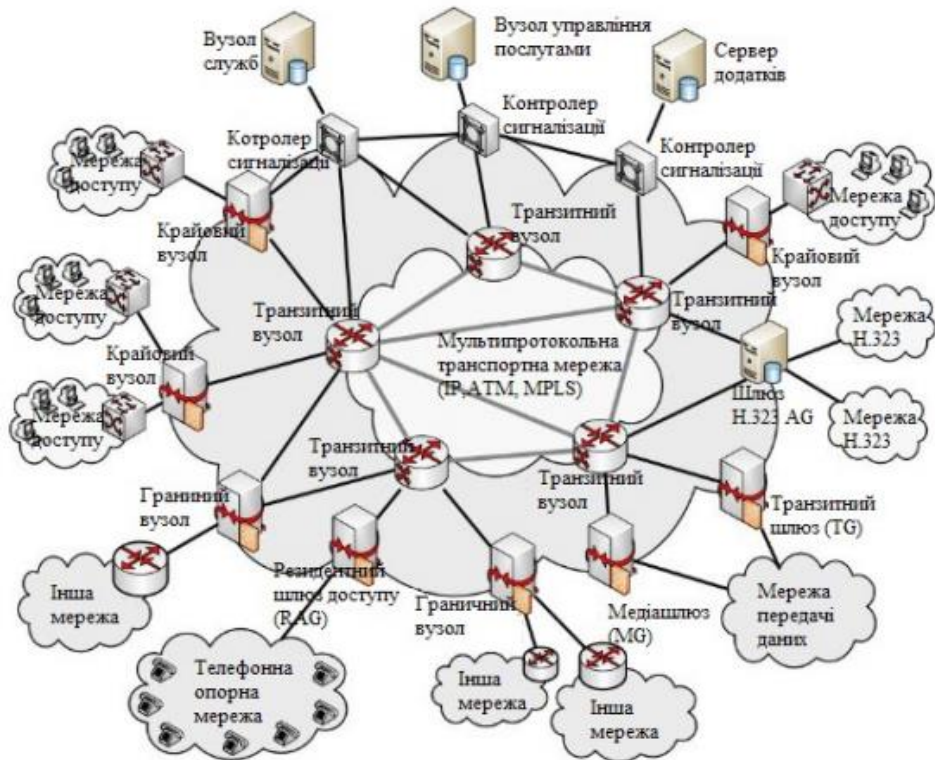


Рис. 1.1. Архітектура мультисервісних мереж

1.3. Рішення в мережах NGN

В транспортній мережі для виконання функцій комутації та передавання використовують транзитні вузли, шлюзи для підключення традиційних мереж зв'язку, кінцеві вузли для забезпечення доступу користувачів до мультисервісної мережі та контролери сигналізації, що здійснюють функції управління викликами, з'єднаннями та обробки інформації сигналізації. Використання загальних контролерів дає право розглядати їх як цільну систему комутації. Призначенням транспортного рівня є надання послуг перенесення, прозора передача інформації користувача та комутація. Транспортний рівень базується на основі пакетних технологій передачі інформації, використовуючи технології ATM, які включають вбудовані засоби забезпечення якості обслуговування та можуть використовуватися при створенні NGN практично без змін і IP, що потребують реалізації в них додаткової функції забезпечення якості обслуговування.[3] Користувачам

надається загальна інтегрована платформа передачі даних з високою надійністю, забезпечення QoS , високою продуктивністю. Від технологій багато в чому залежить рівень роботи всієї NGN і кількість сервісів, що надаються. Саме універсальна транспортна мережа є базою NGN, яка здійснює функції транспортного рівня і рівня управління комутацією.

Завданням рівня управління комутацією і передачею інформації є маршрутизація викликів, обробка інформації сигналізації та управління потоками. На цьому рівні використовується Softswitch (надання голосових і мультимедійних послуг) або технологія програмної комутації. Виконується забезпечення керуванням з'єднання та первинного управління викликами. До того ж гнучкий комутатор (SoftSwitch) відповідає за управління викликами, обробку протоколів, послуг мобільного зв'язку, розподілом ресурсів, а також за інтерфейси програмування додатків та управлінням доступу до медіашоюзів. В NGN підтримується більше число інтерфейсів та вища пропускна здатність, ніж в традиційних мережах. Все це приводить до необхідності перегляду принципів і підходів до мережевого управління для NGN. Система управління опирається на використанні об'єктно-орієнтованої розподіленої структури. Гнучка комутація, яка використовується для управління установлених з'єднань є основною технологією та цьому рівні, а метою є перенесення інформації в режимі реального часу. Однією з головних особливостей систем управління NGN є відкрита модульна архітектура, що дозволяє розробляти та впроваджувати нові модулі, працювати з додатками та модернізувати наявні модулі. Рівень управління послугами включає функції управління логікою послуг та додатків, надає інформаційно-комунікаційні послуги, створює та впроваджує нові послуги. Також рівень здатний включати багато незалежних підсистем, що базуються на різноманітних технологіях, мати абонентів, використовувати власні системи адресації, допускає вводити нові послуги, застосовувати одну програму логіки послуг, які є незалежними від типу транспортної мережі. В мережі NGN, крім послуг доставлення інформації, реалізована можливість підтримки надання розширених списків

послуг. Рівень доступу містить мережу абонентського доступу. Для цього використовують як традиційні мережі, клієнти яких отримують доступ до мультисервісної мережі через підключені до шлюзів вузли, так і інтегровані мережі доступу, які підключаються до крайових вузлів мультисервісної мережі.

Якість обслуговування є одним з основних аспектів, який потрібно брати до уваги при проектуванні мереж наступних поколінь (NGN). На відміну від мереж з комутацією каналів своєрідність пакетних мереж зводиться до того, що в інформаційному потоці може передаватися різнорідний трафік. Через пакетні мережі для передачі голосового трафіку існує поняття класи обслуговування, які дозволяють в пакетній мережі оцінити якість надання послуги. Якість обслуговування базується на методі експертних оцінок. Тобто немає гарантії, що при проектуванні мережі будуть закладені мережеві характеристики, які можуть дозволити забезпечити необхідну якість. Проте використання механізмів забезпечення якості обслуговування в процесі експлуатації впливає на надання послуг зв'язку, при цьому технічні характеристики дозволяють використати ці механізми.

Для оцінки якості послуг слід використовувати показники оцінки якості передачі мови та доставки інформації.

Найчастіше параметри, що фігурують у різноманітних визначеннях якості обслуговування, регламентують такі показники роботи мережі: – пропускну здатність; – затримки передачі пакетів; – рівень втрат і перекручувань пакетів А основними якісними характеристиками є сумарна затримка мовної інформації, час установлення з'єднання, ймовірність втрати пакетів та суб'єктивна оцінка якості мови користувачем по п'ятибальній системі MOS.

Якість обслуговування з точки зору користувача може бути виражено сукупністю параметрів (рисунок 1.2). Ці параметри описуються в термінах, зрозумілих як службі, так і користувачеві і не залежать від структури мережі.

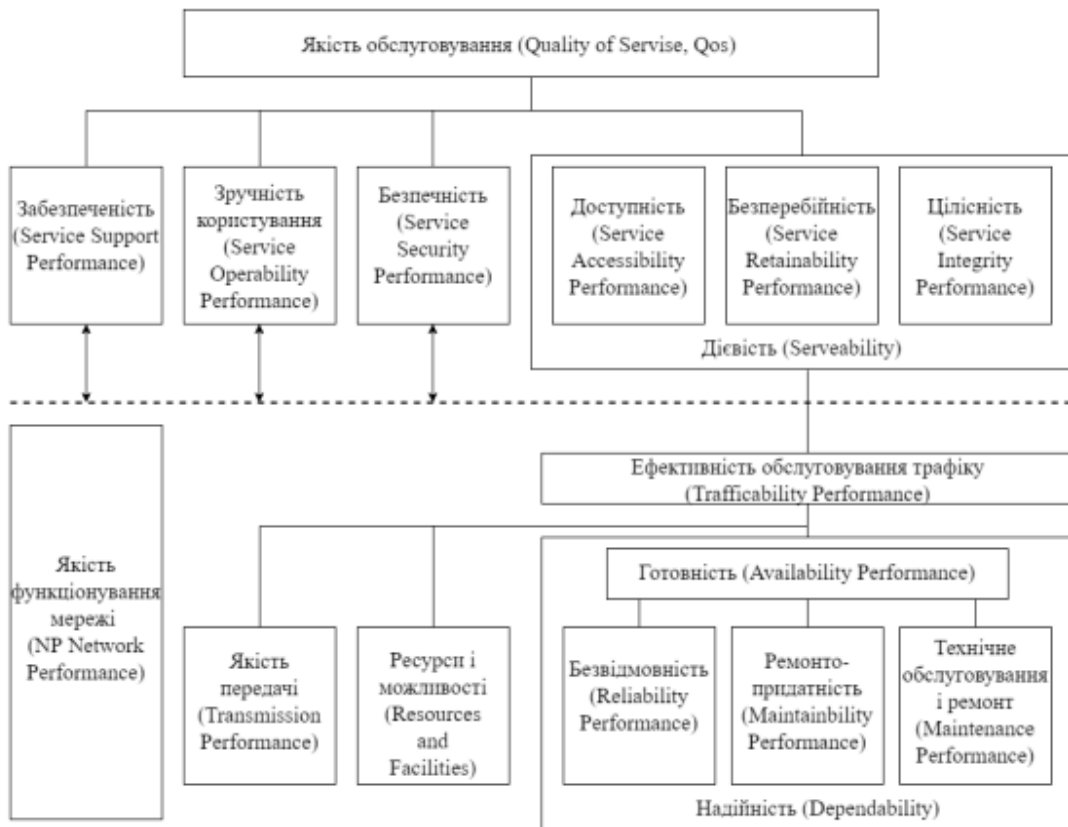


Рис. 1.2. Якість обслуговування телекомунікаційної мережі

1.4. Протокол SIP

SIP — це протокол сигналізації сеансового рівня моделі OSI, розроблений і стандартизований IETF, який використовується для встановлення, налаштування, зміни та завершення мультимедійного сеансу. Його можна порівняти з мовою, яка допомагає пристроям розуміти один одного і обмінюватися даними без помилок. Як і інші протоколи, такі як HTTP/SIP, він заснований на механізмі запит-відповідь. Цей протокол дає можливість встановлювати та завершувати сеанси обміну медіа-даними між двома кінцевими точками, визначаючи параметри, типи та формати даних, що передаються. Після встановлення сеансу відповідальність за передачу голосу та відео в реальному часі делегується транспортному протоколу RTP, який динамічно виділяє один порт, як правило, у діапазоні від 10 000 до 20 000. Він

використовується для різних цілей, напр. таких як IP-телефонія, відео- та аудіоконференції та навіть онлайн-ігри. Іншими словами, він працює за схемою «клієнт-сервер», постійно чергуючи запити і відповіді.

Протокол SIP забезпечує 5 запитів і відповіді на них. Сигналізація SIP дозволяє агентам користувачів і мережевим серверам знаходити, надсилати запити та керувати з'єднаннями:

- INVITE - запит запрошує користувача або сервіс взяти участь у сеансі підключення та містить опис параметрів підключення. За допомогою цього запиту користувач може визначити функціональність кінцевого пристрою партнера по зв'язку та розпочати сеанс зв'язку, використовуючи обмежену кількість повідомлень і підтвердження.
- ASK - запит підтверджує отримання від сторони, відповідальної за відповідь на команду INVITE, і завершує транзакцію.
- ОПЦІЇ - запит дозволяє отримати інформацію про функції агентів користувачів і мережесерверів. Однак цей запит не використовується для організації сеансів зв'язку.
- BYE - страждання та пожирання сторінок для знищення університету. Перед класифікацією знань програми для даного середовища агенти виправлять цю помилку на сервері, перед класифікацією знань, перед класифікацією знань.
- CANCEL дозволяє агентам користувачів і мережевим серверам скасовувати будь-який попередньо надісланий запит, якщо відповідь ще не отримана. Основна маршрутизація викликів SIP показана на рисунку 1.3.

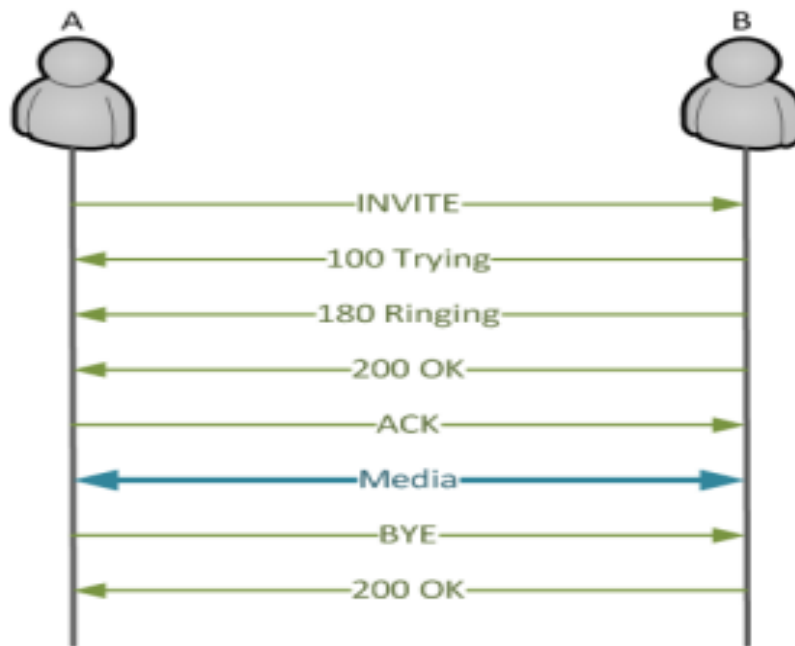


Рис. 1.3. Проходження дзвінка

На сьогоднішній день SIP є універсальним стандартом для обміну даними в мережі. Експерти вважають його найперспективнішим і найчастіше використовуваним. SIP-телефонія доступна там, де є інтернет зі швидкістю не менше 64 Кбіт/с. Тобто практично всюди.

IP-телефонія дозволяє прив'язати номер не до місця, а до конкретного користувача, що відкриває широкі можливості, наприклад, для аналітики. Адже так можна дізнатися про клієнта набагато більше. Зокрема, ця технологія лежить в основі відстеження викликів.

На рисунку 1.4. зображено архітектуру протоколу.

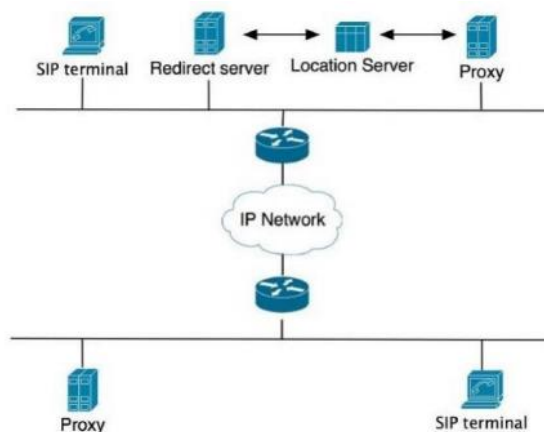


Рис. 1.4. Архітектура протоколу SIP

Він складається з таких елементів: – Термінал. За умови, що сервер і клієнт зосереджені в кінцевому середовищі і безпосередньо підключені і можуть взаємодіяти з самим користувачем мережі, вони стають клієнт-агентами і серверами, тобто клієнтом-агентом користувача (UAC) і агентом користувача. сервер (UAS), або Подібним чином, якщо пристрій містить агент клієнта та сервер, він стає агентом користувача, тобто. - User Agent (UA), працює як SIP термінал. – Проксі-сервер Керує завдяки іyumu paus і процесом процесу в мережі та процесом від імені самої інсталяції. Такий сервер фактично можна безпосередньо розділити на серверну частину та клієнтську, завдяки чому, крім обробки запитів, вони також можуть ініціювати та надсилати відповіді та підтвердження. Проксі-сервер має заголовок Via, до якого він може додати свою службову інформацію, не змінюючи вміст корисної інформації. Існує два типи проксі-серверів – (stateful) і (stateless). Перший тип зберігає всі запити та згенеровані ними процеси та відповіді у своїй пам'яті до кінця транзакції. Другий тип видаляє всю інформацію зі своєї пам'яті, як тільки запит отримано та оброблено, завдяки чому можна забезпечити виконання більш складних функцій і сервісів, таких як проксі-сервер statefull.

- Сервер переадресації використовується для переадресації виклику на поточне місце реєстрації користувача. Цей сервер не керує викликами та не генерує свої запити, а лише надає адресу запитуваного терміналу за допомогою запитів класу 3XX (301 Moved Permanently або 302 Moved Temporarily). Для виконання цих операцій сервер пересилання зв'язується з реєстратором SIP.

- Сервер реєстрації (реєстратор). Архітектура протоколу SIP забезпечує мобільність користувача, тобто абонент може змінити місце реєстрації в мережі, щоразу отримуючи поточну адресу. З цієї причини в SIP існує механізм, який повідомляє про нову адресу користувача, що робить призначений йому агент. Цей механізм називається «реєстрація».

Реєстраційний сервер використовується для запису та зберігання поточної адреси користувача і фактично є адресною базою даних, яка регулярно оновлюється. Запит на РЕЄСТРАЦІЮ надає реєстратору інформацію про адресу користувача: логін, пароль, IP-адресу та сам номер телефону. Сервер реєстрації може схвалити цей запит, якщо реєстраційні дані правильні, або відхилити запит на реєстрацію через неправильні дані або блокування введеної IP-адреси з його боку.

SIP-сервер є найважливішою частиною IP-АТС. Ця служба в основному керує SIP-комутаторами між SIP-клієнтами/кінцевими точками. У літературі SIP-сервер також може називатися SIP-проксі або SIP-реєстратор (Registrar).

Хоча цей сервер можна вважати найважливішою частиною телефонної системи на основі SIP, його завдання обмежуються керуванням і підтримкою сеансів SIP. Основні завдання SIP сервера:

- Створює сеанс зв'язку між двома (чи більше) кінцевими точками. Наприклад, голосова конференція може мати кілька кінцевих точок.
- Узгодження медіапараметрів для кожного сеансу між кінцевими точками за допомогою протоколу SDP
- Редагування або зміна параметрів сеансу після його початку. Наприклад, утримання дзвінка.
- Перехід від однієї кінцевої точки до іншої. Наприклад, переклад дзвінка учасника
- Завершення сесії SIP.

Сам SIP-сервер не обробляє медіа-потоки - це робить окремий медіа-сервер (Media server) за протоколом RTP. У реалізаціях IP-АТС SIP-сервер і супровідний медіа-сервер завжди розташовані на одному фізичному сервері.

Однак системи з великим навантаженням (наприклад, у великих провайдерів VoIP) можуть використовувати медіа-сервер, встановлений на іншому фізичному комп'ютері, для кращої обробки сеансу. Навантаження також можна розподілити між кількома медіасерверами.

Таким чином, SIP-сервер не бере безпосередньої участі в обробці аудіо та відео. Всі ці процеси лише керують і «знайомлять» учасників з правилами участі в бесіді.

Для класифікації SIP-серверів використовується кілька критеріїв:

- Фізичні характеристики (або варіант виконання). Існують апаратні та програмні SIP-servers. Перші виконуються у вигляді окремого фізичного пристрою (склад як у звичайного сервера: ПЗП, ОЗУ, спеціалізоване програмне забезпечення тощо). Програмні SIP-сервери - це спеціалізоване програмне забезпечення для встановлення на різні платформи. Програмне забезпечення розраховано високе навантаження. Апаратним плюсом є можливість розгорнути сервер практично на будь-якій платформі без величезних вкладень в апаратне забезпечення;
- Чисельна ємність. Це кількість передплатників (або іншими словами облікових записів), яку може обслуговувати сервер;
- Максимальна кількість одночасних підключень (не плутати з номерною ємністю). Його ще називають «шлейф». Визначає кількість учасників, які можуть брати участь у виклику одночасно. Залежно від сервера він може бути меншим або дорівнювати попередньому параметру;
- Кількість портів FXS і FXO. Перші використовуються для підключення зовнішніх телефонних ліній до сервера (наприклад, мережі фіксованого телефонного зв'язку, лінії сторонніх провайдерів тощо). FXO - для обслуговування внутрішніх ліній;

ВИСНОВКИ ДО РОЗДІЛУ 1

Таким чином, аналіз відкритих джерел показав, що за NGN майбутнє у своїй сфері і, що вона має надзвичайно гнучкий та багатофункціональний набір налаштувань порівняно з вже відомими рішеннями. Це стосується і протоколу SIP.

РОЗДІЛ 2

ТЕХНОЛОГІЇ NGN

2.1. Next Generation Network

NGN – технологія мультисервісних мереж нового покоління next generation network широкосмуговий концентратор.[12]

Сучасний ринок зв'язку знаходиться на такому етапі, коли оператори мають сприятливу можливість обійти всі труднощі конвергенції, притаманні мережам минулих років, і перейти до мереж наступного покоління на базі технології, яка отримала робочу назву NGN - «New Generation Network». Для того, щоб зробити цей прорив і приєднатися до високотехнологічних операторів, необхідні нові рішення в галузі створення та надання високопродуктивних послуг. NGN – технологія побудови мережі – призначена для надання послуг передачі даних та голосових сервісів. Вона знімає цілу низку обмежень і бар'єрів, що існують зараз, і в цьому полягає її економічна продуктивність.[12]

Інтеграція NGN-мережі дозволить:

- створити єдине інформаційне середовище підприємства;
- оптимізувати управління IT-інфраструктурою;
- використовувати сучасні послуги управління викликами;
- надавати мультисервісні послуги;
- керувати послугами в реальному часі;
- проводити моніторинг якості послуг і роботи мережевого обладнання.

Next Generation Network - мультисервісна мережа з пакетною комутацією, що складається з вузлів доступу, виконана за технологією Ethernet та IP MPLS (Multiprotcol Label Switching). Технологія NGN, на базі якої

побудована телекомунікаційна мережа SMARTNET, забезпечує підключення компаній на швидкостях до 1 Гбіт/с, за допомогою власного волоконно-оптичного кільця компанії SMARTNET в м. Алмати і володіє великою зоною покриття, вищим ступенем надійності та керованості а також унікальними можливостями по забезпечення якості обслуговування (QoS), класів обслуговування (CoS), продуктивності та відмовостійкості.[12]

NGN являє собою універсальну багатоцільову мережу, призначену для передачі мовлення, зображень та даних з використанням технології комутації пакетів. По суті, вона є результатом злиття Інтернету та телефонних мереж, поєднуючи в собі їх найкращі риси. На практиці це означає гарантовану якість голосового зв'язку та передачі даних у критично важливих додатках.[12]

Таким чином, NGN має ступінь надійності, характерну для СТОП (на противагу негарантованому якості зв'язку через Internet) і забезпечує низьку вартість передачі у розрахунку на одиницю обсягу інформації (наближеної вартості передачі даних по Internet, а не СТОП). Саме мережі наступного покоління - NGN забезпечать надання користувачам доступу до максимальної кількості сервісів. "Доступність будь-яких сервісів, завжди і скрізь" - так можна коротко висловити основну ідею та мету NGN. Мається на увазі, що така мережа має стати в широкому розумінні універсальним комунікаційним середовищем, яким передається трафік будь-якого типу (СТОП, Internet, бездротових мереж).[12]

Мережі NGN повинні забезпечувати надання необмеженого набору послуг з гнучкими можливостями щодо їх управління, персоналізації та створення нових послуг за рахунок уніфікації мережевих рішень. Виходячи з усього вищевикладеного, впливає, що NGN - це мультисервісна мережа, що забезпечує передачу всіх видів медіатрафіку та розподілене надання необмеженого спектру телекомунікаційних послуг, з можливістю їх додавання, редагування, розподіленої тарифікації з різними вимогами до якості обслуговування.[12]

Що отримує клієнт, який використовує послуги на базі мережі NGN?

- мінімізацію витрат на експлуатацію та обслуговування за рахунок створення єдиного мультисервісного (універсального) середовища;
- мінімізацію витрат на управління та підтримку ІТ-інфраструктури;
- Побудова корпоративної мережі в будь-яких масштабах;
- гнучкість в управлінні власною мережею та наданими йому ресурсами оператора;
- єдиний web-інтерфейс управління наданими послугами.

В основу технології покладено концепцію перебудови суспільства на принципах повнозв'язків, коли всі інформаційні ресурси стають загальнодоступними у будь-якому середовищі, вони можуть бути доставлені незалежно від того, де знаходиться людина. Зародок концепції повнозв'язку виник в Інтернеті, що дало поштовх величезної популярності мережі. Користувач, входячи до Інтернету, отримує доступ до всього світу.[12]

NGN з погляду споживача є хмарою чи «чорним ящиком», оскільки споживачеві не цікаво знати стеки, протоколи тощо. Вона дозволяє надавати такі послуги, як широкосмуговий доступ до Інтернету (100 Мбіт/сек), пакетну телефонію, VPN, «відео на запит» та виділені широкосмугові канали. Таким чином, NGN – мережа зв'язку наступного покоління (Next step generation) – гетерогенна мультисервісна мережа, що забезпечує передачу всіх видів медіатрафіку та розподілене надання необмеженого спектру телекомунікаційних послуг з можливістю їх додавання, редагування, розподіленої тарифікації.[12]

Мережа підтримує передачу трафіку з різними вимогами до якості обслуговування та забезпечує підтримку цих вимог. Пакетні технології обробки дозволяють запропонувати користувачу в такій мережі прозорі автоматизовані принципи розрахунків за приєднання, розрахунку за вхідний, вихідний трафік, вводити платежі за ініціалізацію, транзит та термінацію трафіку, розраховувати сигнальний трафік, виділяючи складову трафіку, пропущеного від стороннього оператора.[12]

У новій мережі застосовується передова технологія маршрутизації Riverstone. На відміну від традиційних мереж у структурі NGN утворено додатковий шар – управління комутацією транспортної мережі. Він організується за допомогою програмних комутаторів – «SoftSwitch», які мають підтримувати трансляцію основних протоколів VoIP до протоколів традиційних мереж. Тому для поєднання пакетних та традиційних телефонних мереж «SoftSwitch» повинен відповідати таким вимогам:[12]

- працювати з протоколами сигналізацій різної архітектури та взаємодіяти з медіашлюзами, що забезпечують передачу голосової, сигнальної інформації, даних, IP-телефонії та інших видів трафіку;
- підтримувати всю різноманітність сигналізацій - OKC-7, DSS1, ВКС та ін., оскільки з точки зору телефонної мережі він є транзитним комутатором та пунктом сигналізації OKC-7;
- підтримувати всі протоколи IP-телефонії (H.323, MGCP, H.248, SIP) та здійснювати їх конвертацію з одного протоколу в інший, оскільки для пакетних мереж він є пристроєм керування медіашлюзами та контролером сигналізацій.

Таким чином, обладнання програмної комутації в NGN відіграє роль універсального програмно-апаратного комплексу, конвертера сигналізації, який перетворює протоколи сигналізації як у мережі з комутацією каналів: OKC-7, DSS1, V5, CAS, так і в мережі пакетної комутації - протоколи IP-телефонії: H 323, SIP, MGCP, MEGACO/H.248.[12]

Програмний комутатор - це програмно-апаратний комплекс, «мозок» нової мережі, призначений керувати обробкою телефонних викликів, які у різних мережах, зокрема у мережах з комутацією пакетів, він акумулює весь інтелект мережі, інші елементи, розташовані на периферії, позбавлені інтелекту і повністю підконтрольні програмному комутатору, що в цілому сприяє кращій керованості та масштабованості мережі. Тому застосування програмних комутаторів з погляду побудови мережі є технологічною інвестицією у майбутнє, оскільки OKC-7 підтримує сьогодні роботу ТФОП, і

лише острівці нових пакетних мереж сигналізуються із застосуванням IP-протоколів.[12]

Слід зазначити, що немає чітко сформульованих вимог до мережі NGN. Поки що сформульовані деякі базові принципи, на які необхідно орієнтуватися при введенні мультисервісних мереж наступного покоління. Як зазначалося вище, ядром мережі нового покоління є програмний комутатор - мозок мережі, який акумулює весь інтелект мережі, а інші елементи, розташовані на рівнях доступу, шлюзів, транспорту, позбавлені інтелекту та повністю підконтрольні програмному комутатору, що в цілому сприяє кращій керованості та масштабованості мережі.[12]

У разі переважання на мережі аналогового обладнання, морально та фізично застарілого, можливо модернізувати мережу, використовуючи програмний комутатор як розподілений телефонний концентратор, що дозволить домогтися зниження витрат на будівництво та експлуатацію мережі за рахунок того, що замість двох роздільних мереж (з комутацією каналів та пакетів) будуватиметься одна мережа, функції вузлових (транзитних) АТС виконуватиме програмний комутатор, з'явиться можливість більш економічної організації нових послуг.[12]

2.2. Протоколи мереж NGN

Мережі NGN можна розглядати як мережеві рішення, що поєднують фрагменти різних існуючих мереж (Інтернет і ТФОП) із застосуванням властивих цих мереж технологій. Відповідно, в NGN застосовуються як протоколи Інтернет (наприклад протоколи стека TCP/IP), і протоколи ТфОП (наприклад протоколи інтерфейсу V5). Крім того, деякі протоколи NGN є перспективними, прямо чи опосередковано торкаючись принципів взаємодії мереж Інтернет та ТФОП у рамках створення мультисервісної мережі. Протоколи NGN з деякою часткою умовності можна класифікувати так:[14]

- базові протоколи Інтернету: IP, ICMP, TCP, UDP.

- транспортні протоколи: RTP, RTCP.
- сигнальні протоколи: SIP, H.323, SIGTRAN, MEGACO/H.248, MGCP, RSVP, SCTP, ISUP, BICC, SCCP, INAP.
- протоколи маршрутизації: RIP, IGRP, OSPF, IS-IS, EGP, BGP, IDRP, TRIP.
- протоколи інформаційних служб та управління: SLP, OSP, LDAP, SNMP.
- протоколи послуг: FTP, SMTP, HTTP, кодеки G.xxx, H.xxx, факс T.37, T.38, IRP, NNTP.

2.2.1. Базові протоколи стеку TCP/IP

Протоколи Інтернет можна використовувати для надсилання повідомлень через будь-який набір об'єднаних між собою мереж. Вони однаково придатні зв'язку, як у локальних, і у глобальних мережах. Комплект протоколів Інтернет включає не лише специфікації нижчих рівнів (наприклад, TCP і IP), але також специфікації для таких загальних застосувань, як пошта (SMTP), додатки гіпертекстових терміналів (HTTP) і передача файлів (FTP).[14]

2.2.2. Протокол IP

Маршрутизація за протоколом IP (Internet Protocol) визначає формат, адресацію та характер переміщення дейтаграм IP через об'єднані мережі (по одному пересиланню за раз). На початку дейтаграм весь їхній маршрут не відомий. Натомість на кожному проміжному вузлі обчислюється наступний пункт призначення шляхом зіставлення адреси пункту призначення, що міститься в дейтаграмі, із записом даних у маршрутній таблиці поточного вузла. Участь кожного вузла в процесі маршрутизації полягає в просуванні пакетів, що базується лише на внутрішній інформації, незалежно від того, наскільки успішним буде процес, і того, чи досягне пакет кінцевого пункту

призначення. Іншими словами, IP не забезпечує відправлення на вузол - джерело повідомлень про несправності, коли мають місце аномалії маршрутизації. Виконання цього завдання надано іншому протоколу Інтернету, а саме протоколу керуючих повідомлень Інтернет (Internet Control Message Protocol — ICMP).[14]

2.2.3. Протокол ICMP

ICMP виконує низку завдань у межах об'єднаної мережі IP. На додаток до основного завдання, для виконання якого він був створений (повідомлення джерела про відмови маршрутизації), ICMP забезпечує також:

- метод перевірки здатності вузлів утворювати в об'єднаній мережі повторне відлуння (повідомлення Echo та Reply ICMP);
- метод стимулювання більш ефективної маршрутизації (повідомлення Redirect ICMP – переадресація ICMP);
- метод інформування джерела про те, що якась дейтаграма перевищила призначений їй час існування в межах цієї об'єднаної мережі [ICMP-повідомлення Time Exceeded (час перевищено)];
- метод передачі інших корисних повідомлень.

Зроблений нещодавно доповнення до ICMP забезпечує для нових вузлів можливість знаходження маски підмережі, яка використовується на проміжній мережі в даний момент. В цілому, ICMP є інтегральною частиною будь-яких реалізацій IP, особливо таких, що використовуються маршрутизаторами.[14]

2.2.4. Протокол TCP

Transmission Control Protocol (TCP) забезпечує повністю гарантовані, з підтвердженням та керуванням потоком даних, послуги доставки для протоколів вищих рівнів. Він переміщує дані у безперервному неструктурованому потоці, у якому байти ідентифікуються за номерами послідовностей. TCP може підтримувати численні одночасні діалоги вищих рівнів.

2.2.5. UDP

Протокол UDP набагато простіше, ніж TCP. Він корисний у ситуаціях, коли потужні механізми забезпечення надійності протоколу TCP не є обов'язковими. Заголовок UDP має лише чотири поля; поле порту джерела (source port), поле порту пункту призначення (destination port), поле довжини (length) та полі контрольної суми UDP (checksum UDP). Поля порту джерела та порту призначення виконують ті самі функції, що у заголовку TCP. Поле довжини позначає довжину заголовка UDP та даних; поле контрольної суми забезпечує перевірку цілісності пакета. Контрольна сума UDP є факультативною можливістю.

Комплект протоколів Інтернет включає велику кількість протоколів вищих рівнів, що представляють найрізноманітніші застосування, у тому числі управління мережею, передача файлів, розподілені послуги користування файлами, емуляція терміналів та електронна пошта.

2.2.6. Сигнальні протоколи SIP та H.323

В даний час для встановлення мультимедійних дзвінків через мережі IP створено кілька протоколів, наприклад SIP (Session Initiation Protocol) та H.323. Поява даних стандартів відкриває широкі можливості децентралізації забезпечення телефонних послуг, причому послуги можуть керуватися з боку користувача.[14]

Протокол ініціювання сеансів зв'язку (SIP) призначений для організації, модифікації та завершення мультимедійних сеансів або дзвінків. Мультимедійні сеанси включають мультимедійні конференції, інтернет-телефонію та інші аналогічні програми. SIP є одним із ключових протоколів, що використовуються для реалізації передачі мови по мережах IP (Voice over IP - VoIP). Таким чином, SIP є спрощеним протоколом сигналізації, що має широке застосування в Інтернет-телефонії.[14]

SIP являє собою простий протокол сигналізації для встановлення, модифікування та руйнування мовних та мультимедійних з'єднань у сеансах IP-телефонії (VoIP) та мультимедійного конференц-зв'язку. SIP є протоколом типу «клієнт-сервер» і подібний до протоколу передачі гіпертексту (HyperText Transfer Protocol - HTTP) як з синтаксичної, так і з семантичної точок зору. Має текстові запити та відгуки, що містять поля заголовків, в яких передається інформація про обслуговування та характеристики з'єднання.[14]

Якщо для керування ресурсами середовища передачі MG між MGC та шлюзами середовища передачі (тобто транспортними або медіа-шлюзами) використовується протокол H.248/MEGACO (MGCP), то SIP або H.323 можуть застосовуватися для встановлення мовного з'єднання на ділянці між MGC і клієнтом VoIP.[14]

У разі необхідності встановлення телефонного з'єднання між транспортними шлюзами середовища передачі, не пов'язаними з тим самим контролером, зв'язок сигналізації виклику встановлюється між MGC, керуючим першим шлюзом, і MGC, керуючим другим шлюзом середовища передачі.[14]

Хоча і H.323, і SIP можуть використовуватися для здійснення такої сигналізації, необхідно зазначити, що жоден з них не розроблявся з розрахунку на підтримку послуг цього типу. Отже, очікується, що для підтримки зв'язку між MGC з метою надання послуг VoIP той та/або інший будуть відповідно оптимізовані.[14]

Протокол SIP-T (SIP for Telephones, RFC 3372) надає можливості інтеграції повідомлень традиційної телефонної сигналізації повідомлення протоколу SIP. SIP-T, таким чином, не є новим протоколом, а є набором механізмів узгодження традиційної телефонної сигналізації з сигналізацією SIP. Завданням SIP-T є виконання трансляції повідомлень протоколу та забезпечення прозорості транспортування їх властивостей через точки взаємозв'язку ТФОП-IP. Протокол призначений для використання там, де мережа VoIP має інтерфейс з ТФОП. Незважаючи на те, що SIP у типовому

варіанті застосовується поверх протоколів UDP або TCP, без внесення будь-яких технічних змін він може використовувати також можливості протоколів IPX, Frame Relay, AAL5/ATM або X.25.

2.2.7. Протокол MGCP

Протокол керування транспортним шлюзом (Media Gateway Control Protocol – MGCP) використовується між елементами розподіленого мультимедійного шлюзу. Розподілений мультимедійний шлюз включає агента виклику (Call Agent), який містить «інтелект» з управління викликом, і транспортний шлюз, який містить транспортні функції (наприклад, перетворення мовного каналу ІКМ в VoIP).[14]

Транспортні шлюзи містять кінцеві точки, у взаємодії з якими агент виклику може створити, змінити та знищити з'єднання для встановлення та керування транспортним сеансом зв'язку з іншими кінцевими точками. Також агент виклику може доручити кінцевим точкам виконання функцій виявлення деяких подій та генерування конкретних сигналів. Кінцеві точки автоматично повідомляють агенту виклику про зміни у стані обслуговування[14]

Базовий протокол MGCP може розширюватися додатковими функціональностями, що визначаються у про пакетах (package) — точно описаних розширеннях MGCP. Наприклад, можна згадати пакети різного призначення під назвою «MGCP-пакети 2ВСК». Ці пакети розширень призначені для роботи зі з'єднувальними лініями, на яких можуть використовуватись системи телефонної сигналізації з багаточастотним або імпульсним набором, інтерфейси з установчими АТС тощо.[14]

ВИСНОВКИ ДО РОЗДІЛУ 2

Тож, з наведеної вище інформації, можна сказати, що використання мереж нового покоління(NGN) є найбільш вигідним, як з економічної точки зору, так і з технологічної, а також з точки зору зручності використання.

РОЗДІЛ 3

SIP-СЕРВЕР. НАВАНТАЖЕННЯ НА SIP-СЕРВЕР.

3.1. SIP-сервер

Протокол ініціації сеансу SIP ґрунтується на стандарті RFC 2543. Він розроблений для подолання недоліків свого попередника H.323. У чому особливості стандарту:[8]

- незалежність від централізованих інформаційних серверів маршрутизації дзвінків;
- підтримка відео- та аудіовикликів;
- більш якісне з'єднання;
- підтримка одноадресних та багатоадресних передач;
- проста структура.

Згодом RFC 2543 замінили на нові версії протоколу. Однак він, як і раніше, актуальний як основа для інших стандартів.

Більшість SIP-адрес підключаються до унікального номера телефону. Таким чином, кожен користувач у мережі SIP може отримати точку доступу, що дозволяє йому здійснювати та приймати прямі дзвінки.

SIP-сервер є найважливішим компонентом IP-АТС. В основному цей сервіс управляє SIP-комутаціями між SIP-клієнтами/кінцевими точками (рис. 3.1). У літературі SIP-сервер також може називатися SIP-проксі або SIP-реєстратором (Registrar).[7]

Хоча цей сервер можна вважати найважливішою частиною телефонної системи на базі SIP, його завдання обмежуються керуванням та підтримкою SIP-сесій. Основні завдання SIP-сервера:[7]

Встановлює сеанс зв'язку між двома (або більше) кінцевими точками. Наприклад, голосова конференція може мати кілька кінцевих точок.[7]

- Узгодження параметрів медіа для кожної сесії між кінцевими точками, використовуючи протокол SDP[7]
- Підстроювання чи зміна параметрів сесії вже після її початку. Наприклад, постановка виклику утримання.[7]
- Перемикання з однієї кінцевої точки на іншу. Наприклад, переклад виклику абонента[7]
- Завершення SIP-сесії[7]

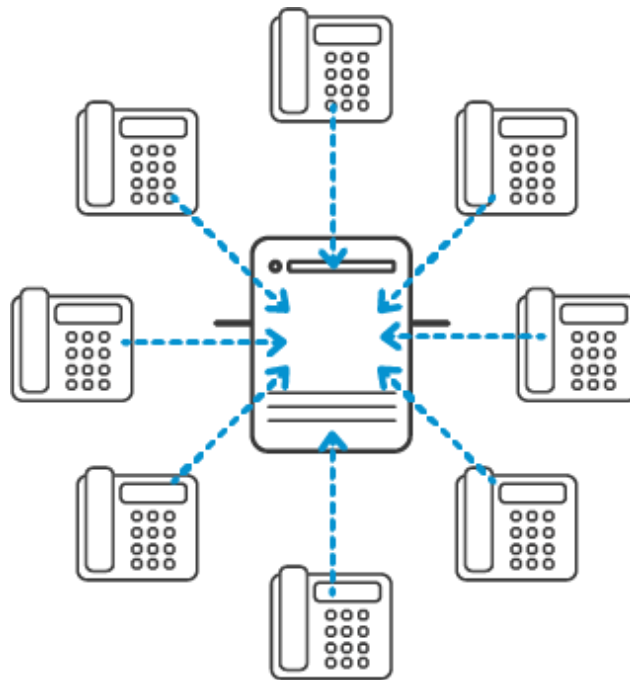


Рис. 3.1. SIP-сервер[7]

SIP-сервер сам не обробляє медіапотоки – це робить окремий медіа-сервер (Media server), використовуючи протокол RTP. У реалізаціях IP-АТС завжди SIP-сервер і його компаньйон медіа-сервер перебувають у одному фізичному сервері.[7]

Однак системи з великим навантаженням (наприклад, у великих VoIP-операторів) можуть використовувати медіа-сервер, встановлений на іншій фізичній машині, щоб краще справлятися з обробкою сесій. Також можливе розподілення навантаження на кілька медіа-серверів.[7]

SIP-сервер відповідає за передачу та завершення дзвінків через два типи проксі-серверів: без збереження стану(рис. 3.1.1) та зі збереженням(рис. 3.1.2).[8]



Рис. 3.2. Проксі сервер без збереження стану[8]

Це проксі-сервер, який не зберігає інформацію про дзвінки. Це означає, що немає записів, які зберігаються та використовуються для усунення несправностей та керування з'єднаннями. Проксі-сервери такого типу мають ряд переваг:[8]

- невисоке навантаження на пам'ять та ЦП, оскільки не потрібно зберігати інформацію про з'єднання;
- спрощене масштабування з погляду балансування навантаження;
- більш короткий час відповіді запиту, оскільки витрачається час на обробку запиту перед поверненням.



рис. 3.3. Проксі сервер зі збереженням стану[8]

Проксі-сервери без збереження стану лише маршрутизують пакети, а ось проксі-сервер SIP з відстеженням стану зберігає інформацію, пов'язану з кожним викликом. Якщо виникають проблеми зі з'єднанням, адміністратор

мережі телефонії може переглянути журнали дзвінків та відстежити точку неполадки.[8]

Крім того, якщо один власний агент SIP відключається, він може відновити з'єднання з іншою кінцевою точкою без необхідності ініціювати новий сеанс. Це заощаджує пропускну здатність для обох сторін, а також час, оскільки немає затримок при налаштуванні нових з'єднань.[8]

Недолік в тому, що проксі-сервери SIP з відстеження стану мають більше накладних витрат. Їхня вартість вища, тому що потрібно більше ресурсів для зберігання даних.[8]

3.1.1. Принцип роботи SIP-серверу та його переваги

SIP-сервер – це посередник усіх елементів, які підтримують зв'язок між двома чи більше кінцевими точками. SIP-сервер можна порівняти із сигнальником на залізниці, який стежить, щоб усі стрілки були переведені у правильному порядку. Якщо на маршруті сигнали зелені, повідомлення може бути надіслано.[8]

Проксі-сервер SIP також відомий як декодер, який гарантує отримання будь-якого повідомлення в будь-якому форматі. Як тільки з'єднання встановлено, сервер виконує дію – ставить абонента, що викликає, на утримання або переводить його на інший додатковий номер.[8]

Для миттєвого надсилання великих обсягів даних кожне повідомлення розбивається на кілька пакетів. Кожен пакет містить частину вихідного повідомлення. Коли пакети досягають місця призначення, вони знову збираються правильно.[8]

Під час підготовки до передачі пакета SIP-сервер спочатку надсилає сигнальні пакети. Як тільки мережа, що приймає, підтвердить отримання сигнальних пакетів, система розуміє, що може приступити до передачі передбачуваного повідомлення.[8]

Нарешті після завершення дзвінка SIP-сервер забезпечує правильне завершення сеансу між двома адресами.[8]

SIP-сервер офісної телефонії забезпечує ряд переваг, таких як:

- зниження витрат за пропускну спроможність;
- покращення обробки викликів;
- підвищення ефективності набору номера;
- підвищення продуктивності з допомогою балансування навантаження;
- менша затримка передачі.

В цілому наявність SIP-сервера допомагає бізнесу знизити витрати на телефонію за рахунок меншої кількості телефонних ліній, більшої мобільності, гнучкості використання різних пристроїв для зв'язку та підвищення безпеки.[8]

Особлива перевага SIP-серверів у тому, що вони підвищують безпеку мережі. SIP-сервер перевіряє особу користувача, перш ніж він відправить або отримає пакети даних. SIP-сервер також потрібен для автентифікації Message Digest.[8]

Автентифікація Message Digest (MD) надає механізм автентифікації однорангового вузла перед обміном сеансовими ключами. Обмін відбувається за допомогою HMAC-MD5 – коду автентифікації хеш-повідомлення як алгоритму, який використовує загальний секретний ключ для створення закодованого пакета повідомлень. Отримати пакет і декодувати його може тільки той бік, який має такий самий секретний ключ, як і відправник.[8]

Основною перевагою протоколу є його здатність забезпечувати цілісність та автентифікацію без необхідності відправляти паролі незахищеними каналами або зберігати їх на кожному пристрої.[8]

Продуктивність серверів обробки дзвінків визначається кількістю дзвінків, які здатні обслуговувати сервер в одиницю часу. Коли інтенсивність дзвінків перевищує пропускну здатність сервера, він переходить у режим навантаження, час очікування на сервері збільшується. Клієнт чекає на відповідь на запит і, не дочекавшись, надсилає повторний запит на встановлення з'єднання. Це призводить до ще більшого збільшення вхідного

навантаження на сервер і посилює навантаження. У цій ситуації за відсутності методу контролю навантаження продуктивність сервера різко погіршується. SIP-проксі-сервери, які є основними елементами, що виконують функції встановлення/підтримки/роз'єднання з'єднання мережах IP-телефонії, можуть бути схильні до різкого збільшення вхідного навантаження по низці причин: через збої на елементах мережі, під час проведення телевізійних шоу або при умисних DoS-атаках.

3.2. Перевантаження SIP-серверу

Продуктивність серверів обробки дзвінків визначається кількістю дзвінків, які здатні обслуговувати сервер в одиницю часу. Коли інтенсивність дзвінків перевищує пропускну здатність сервера, він переходить у режим навантаження, час очікування на сервері збільшується. Клієнт чекає на відповідь на запит і, не дочекавшись, надсилає повторний запит на встановлення з'єднання. Це призводить до ще більшого збільшення вхідного навантаження на сервер і посилює навантаження. У цій ситуації за відсутності методу контролю навантаження продуктивність сервера різко погіршується. SIP-проксі-сервери, які є основними елементами, що виконують функції встановлення/підтримки/роз'єднання з'єднання мережах IP-телефонії, можуть бути схильні до різкого збільшення вхідного навантаження по

низці причин: через збої на елементах мережі, під час проведення телевізійних шоу або при умисних DoS-атаках[15]

У специфікації протоколу SIP немає точних вказівок, як реагувати на навантаження. Протокол SIP передбачає базовий метод керування навантаженнями, який використовує код помилки 503 (Service Unavailable). У разі, якщо SIP-сервер не може обробити запит через тимчасове навантаження, він відхиляє запит із кодом помилки 503. Це повідомлення означає, що відправник повинен повторити відправку повідомлення через деякий час або направити запит на інший сервер. Перевантажений сервер також може додати

Retry-After заголовок у SIP-повідомлення, вказавши час у секундах, протягом якого сервер не слід звертатися. Відправник припиняє надсилання повідомлень на цей період часу, повідомлення надсилаються на альтернативний сервер, а відправлення повідомлень на перевантажений сервер відновлюється після закінчення інтервалу часу, вказаного в заголовку. Якщо альтернативний сервер відсутній, і заголовок Retry-After не вказано, сервер SIP повинен відправити відповідь 500 (Server Internal Error) агенту користувача.[15]

Хоча цей метод знижує навантаження на перевантажений проксі-сервер, але призводить до перенаправленню трафіку, який викликав перевантаження, на інший проксі-сервер, який потім також може перевантажитися.[15]

У режимі великого навантаження перевантажений сервер SIP може повідомляти про стан сусіднім серверам і просити їх зменшити навантаження. Але на практиці такі методи боротьби з навантаженням із зворотним зв'язком не завжди реалізовані.[15]

Тому дуже важливо, щоб сервери SIP могли локально захистити власні ресурси, необхідні для обробки повідомлень, за рахунок реалізації відповідних локальних методів боротьби з перевантаженням. Як правило, такі методи включають процедуру вимірювання навантаження на систему і порівняння з фіксованим значенням їжак. Коли граничне значення завантаження системи досягнуто або перевищено, запускається процедура боротьби з перевантаженням. Існує кілька локальних методів боротьби з навантаженням, описаних у літературі, які для вимірювання навантаження на систему можуть використовувати довжину черги буфера, завантаження процесора, кількість активних INVITE-транзакцій, середній час затримки пакетів у буфері сервера та інші параметри. У цих методах використовується стандартна схема обслуговування викликів, передбачена, коли в режимі високого навантаження все нові запити на з'єднання відкидаються відповіддю 503, хоча деякі з цих запитів можуть бути близькі до встановлення з'єднання.[15]

Перевантаження SIP-сервера виникає, коли кількість отриманих повідомлень більша, ніж сервер може обробити, через обмеження ресурсів, наприклад недостатню потужність процесора, ємність пам'яті, пропускну здатність мережі. Перевантаження може призвести до значного зменшення пропускну здатності до частки вихідного значення пропускну здатності. Стандарт протоколу SIP визначає базову відповідь на перевантаження пакетів за допомогою елемента відповіді 503 Server Error.

На основі принципів управління навантаженням запропоновано спосіб вирішення перевантаження на SIP-сервері, який враховує фазу встановлення з'єднання та пріоритет користувача. Цей метод, представлений у вигляді блок-схеми алгоритму обслуговування викликів на сервері SIP, відрізняється від стандартного алгоритму введенням блоку класифікатора та введенням детермінованої схеми маркування повідомлень.

Кожне нове повідомлення надсилається на процесор SIP-сервера.

Цей блок містить буфер, класифікатор повідомлень і модуль обробки. Кеш умовно розділений на дві частини для учасників сервера повідомлень і повідомлень таймера. У нормальній роботі SIP-сервер отримує повідомлення до служби в тому порядку, в якому вони надходять у чергу FIFO, а повідомлення в буфері таймера розташовуються в тому порядку, в якому вони знаходяться в таймерах. Повідомлення, що надходять на сервер, проходять через блок контролю перевантаження. На основі заздалегідь визначених критеріїв блок вирішує, чи поставити повідомлення в чергу, чи відхилити повідомлення. Після припинення служби повідомлення пересилається по мережі, а в буфері встановлюється таймер. Коли таймер закінчується, повідомлення ставляться в чергу для повторного надсилання.

У режимі завантаження повідомлення спочатку надходять у класифікатор, де вони поділяються на транзакції типу INVITE та транзакції, що не схожі на INVITE, і ставляться в чергу на основі пріоритету.

У режимі отримання повідомлення з високим пріоритетом мають більше шансів бути доставленими. Ми будемо використовувати динамічний

пріоритет, щоб вибрати повідомлення з черги, оскільки статичний пріоритет може призвести до недостатнього обслуговування черг з низьким пріоритетом. Для вибору черги ми використовуємо два параметри, які розраховуються для кожної черги.

- Кількість очікуваних запитів у черзі (N_i);
- Фактор важливості черги (U_i).

Тоді пріоритет черги можна визначити як $N_i \times U_i$. Коефіцієнт важливості черги вибирається залежно від важливості повідомлень у цій черзі. Значення цього коефіцієнта можна отримати зі статистичних даних, які описують вимоги користувача. Завдяки довжині пріоритетної черги ми скорочуємо час очікування при обслуговуванні повідомлень із черг із меншим пріоритетом.

Крім того, у режимі перевантаження повідомлення, що належать до транзакцій без INVITE, доставляються за принципом FIFO (першим увійшов – першим вийшов). Якщо повідомлення відноситься до транзакції INVITE, то в цьому випадку враховується поточна фаза встановлення з'єднання. Якщо це запит на нове підключення (фаза 1), повідомлення відхиляється з кодом 503. Якщо сервер INVITE вже обробляє запит (фази 2 і 3), повідомлення надходять на початок черги та обслуговуються за принципом LIFO (Last In First Out). Дисципліна обслуговування LIFO забезпечує кращу пропускну здатність і меншу затримку в черзі навантаження порівняно з FIFO.

Щоб зробити методи боротьби з перевантаженнями на SIP-серверах більш ефективними, розроблено алгоритм обробки викликів SIP-сервера з динамічним пріоритетом і дисципліною обслуговування LIFO в режимі навантаження з урахуванням поточної фази встановлення з'єднання.

Для максимізації пропускну здатності сервера для обробки клієнтських повідомлень, які приносять дохід оператору, рекомендується використовувати схему динамічного пріоритету. Щоб уникнути простої черги з низьким пріоритетом, була використана динамічна схема для вибору повідомлення з черги на основі довжини черги та фактора важливості повідомлення.

Щоб максимізувати загальну пропускну здатність сервера SIP, було запропоновано використовувати політику служби керування LIFO для перевантаження та FIFO для нормального навантаження. Щоб зменшити витрати ресурсів на генерацію випадкових чисел, було запропоновано використовувати детерміновану схему для позначення повідомлень, які не можуть бути обслуговані.

Використання можливостей пакету імітаційного моделювання CPN Tools оцінив продуктивність SIP-сервера з керуванням навантаженням і без нього для середнього часу підключення та пропускну здатності сервера.

ВИСНОВКИ ДО РОЗДІЛУ 3

Отже, судячи з аналізу проблеми навантаження на SIP-сервери можна зрозуміти, що без додаткового контролю навантаження у вигляді окремого балансувальника не може бути впевненості в коректності роботи серверу, чи кластера серверів при високих навантаженнях.

РОЗДІЛ 4

КОНТРОЛЬ НАВАНТАЖЕНЬ НА SIP-СЕРВЕРИ.

ДОСЛІДЖЕННЯ МЕТОДУ БАЛАНСУВАННЯ НАВАНТАЖЕННЯ ЗА КІЛЬКІСТЮ ТРАНЗАКЦІЙ НА СЕРВЕРІ

4.1. Аналіз схем балансування навантаження у кластері SIP серверів

Послуги передачі голосу через IP стають все більш популярними. Протокол SIP як найбільш використовуваний протокол сигналізації в IP-телефонії вона повинна забезпечувати надійність і масштабованість послуг. Одним із важливих компонентів мережі SIP є SIP проксі-сервер. Для того, щоб забезпечити стабільну роботу мережі в умовах зростаючого навантаження, було запропоновано підхід, спрямований на підвищення продуктивності сервера за рахунок кластеризації серверів. Тут постає питання забезпечення балансування навантаження між серверами в кластері.[6]

Балансування навантаження у кластері SIP-серверів полягає у розподілі вхідного трафіку між локальними серверами на основі наявних даних про завантаження ресурсів сервера, тим самим зменшуючи ймовірність того, що перевантаження відбудеться на одному із серверів. За замовчуванням частина операційних систем проксі серверів є модуль балансування навантаження для пом'якшення перевантаження, викликаного збільшенням активності абонентів або ненормальними умовами експлуатації.[13]

Існує багато різних алгоритмів та методів балансування навантаження. Вибираючи конкретний алгоритм, потрібно виходити, по-перше, зі специфіки конкретного проекту, а по-друге, з цілей, яких ми плануємо досягти.[5]

Серед цілей, для досягнення яких використовується балансування, слід виділити такі:

- справедливість – потрібно гарантувати, щоб на обробку кожного запиту виділялися системні ресурси і не допустити виникнення ситуацій, коли один запит обробляється, а решта чекає своєї черги;
- ефективність – усі сервери, які обробляють запити, мають бути зайняті на 100%; бажано не допускати ситуації, коли один із серверів простоює в очікуванні запитів на обробку (одразу ж обмовимося, що в реальній практиці ця мета досягається далеко не завжди);
- скорочення часу виконання запиту – потрібно забезпечити мінімальний час між початком обробки запиту (або його постановкою у чергу на обробку) та його завершення;
- скорочення часу відгуку – потрібно мінімізувати час відповіді запит користувача.[5]

Дуже бажано також, щоб алгоритм балансування мав такі властивості:

- передбачуваність: потрібно чітко розуміти, в яких ситуаціях і за яких навантажень алгоритм буде ефективним для вирішення поставлених завдань;
- рівномірне завантаження ресурсів системи;
- масштабування: алгоритм повинен зберігати працездатність при збільшенні навантаження.[5]

Алгоритм балансування навантаження є найважливішою частиною кластера SIP-серверів. Сучасні алгоритми балансування навантаження можна поділити на два типи: статичні та динамічні.[13]

Статичні алгоритми прості у реалізації, наприклад, випадковий розподіл навантаження Random, кругове балансування навантаження Round Robin, виважений розподіл навантаження Weighted Round Robin і т.д. Динамічні алгоритми використовують інформацію про поточний стан системи або вузла на основі різних метрик (наприклад, завантаження ЦП, доступна пам'ять і т.д.), щоб визначити спосіб розподілу навантаження.[13]

Розглянемо деякі найбільш використовувані алгоритми балансування навантаження SIP-серверів:[13]

4.1.1. За значенням ідентифікатора.

Багато користувачів розділено між M серверами у кластері залежно від першої літери ідентифікатора користувача.[13]

Сервер балансування навантаження розподіляє навантаження між серверами, залежно від ID користувача.[13]

4.1.2. Round Robin

Round Robin, або алгоритм кругового обслуговування, є перебором по круговому циклу: перший запит передається одному серверу, потім наступний запит передається іншому і так до досягнення останнього сервера, а потім все починається спочатку.[5]

Серед безперечних плюсів цього алгоритму слід назвати, по-перше, незалежність від протоколу високого рівня. Для роботи з алгоритму Round Robin використовується будь-який протокол, у якому звернення до сервера йде на ім'я. Балансування на основі алгоритму Round Robin ніяк не залежить від навантаження на сервер: DNS-сервери, що кешують, допоможуть впоратися з будь-яким напливом клієнтів.[5]

Алгоритм Round Robin має і низку істотних недоліків. Щоб розподіл навантаження за цим алгоритмом відповідало згаданим вище критеріями справедливості та ефективності, потрібно, щоб у кожного сервера був однаковий набір ресурсів. За виконання всіх операцій також має бути задіяна однакова кількість ресурсів. У реальній практиці ці умови здебільшого виявляються нездійсненними.[5]

Також при балансуванні за алгоритмом Round Robin не враховується завантаженість того чи іншого сервера у складі кластера. Уявімо таку гіпотетичну ситуацію: один із вузлів завантажений на 100%, тоді як інші — лише на 10 — 15%. Алгоритм Round Robin можливості виникнення такої

ситуації не враховує в принципі, тому перевантажений вузол все одно отримуватиме запити. Ні про яку справедливість, ефективність і передбачуваність у такому разі не може бути й мови.[5]

4.1.3. Weighted Round Robin

Це вдосконалена версія алгоритму Round Robin. Суть удосконалень полягає в наступному: кожному серверу присвоюється ваговий коефіцієнт відповідно до його продуктивності та потужності. Це допомагає розподіляти навантаження гнучкіше: сервери з великою вагою обробляють більше запитів. Однак усіх проблем з стійкістю до відмови це аж ніяк не вирішує. Більш ефективно балансування забезпечують інші методи, у яких при плануванні та розподілі навантаження враховується більша кількість параметрів.[5]

4.1.4. Destination Hash Scheduling ma Source Hash Scheduling

Алгоритм Destination Hash Scheduling був створений для роботи з кластером проксі-серверів, що кешують, але він часто використовується і в інших випадках. У цьому алгоритмі сервер, що обробляє запит, вибирається із статичної таблиці за IP-адресою одержувача.

Алгоритм Source Hash Scheduling ґрунтується на тих же принципах, що й попередній, тільки сервер, який оброблятиме запит, вибирається з таблиці за IP-адресою відправника.[5]

4.1.5. Sticky Sessions

Sticky Sessions – алгоритм розподілу вхідних запитів, при якому з'єднання передаються на той самий сервер групи. Він використовується, наприклад, у веб-сервері Nginx. Сесії користувача можуть бути закріплені за конкретним сервером за допомогою методу IP hash. За допомогою цього методу запити розподіляються на серверах на основі IP-адреси клієнта. Як зазначено в документації (див. посилання вище), «метод гарантує, що запити одного й того клієнта буде передаватися на один і той же сервер». Якщо

закріплений за конкретною адресою сервер недоступний, запит буде перенаправлено на інший сервер.

4.2. Розробка методу балансування навантаження з урахуванням транзакцій

Однією з головних особливостей балансування навантаження на SIP-серверах є те, що запити, що відповідають одному й тому виклику, надсилаються на той самий сервер. Вибір проксі-сервера сервером балансування здійснюється тільки для першого запиту про дзвінок. Усі наступні запити, відповідні цьому виклику, повинні йти на той же сервер. Запропонований у роботі метод враховує цю особливість розподілу дзвінків між серверами. Згідно з методом балансування навантаження з урахуванням транзакцій, алгоритм вибирає сервер з найменшою кількістю активних робіт, які на даний момент ще не завершено. Хоча концепція розподілу навантаження між серверами по найменшій кількості роботи, що залишилася, була застосована в інших дослідженнях, відповідь на питання про те, як це ефективно застосувати на практиці, часто доволі неочевидна. Системі необхідний метод, який дозволяє оцінити обсяг залишившихся робіт для кожного сервера та прийняти рішення про оптимальний розподіл навантаження.

У нашій системі сервер балансування навантаження може оцінити завантаження сервера в кластері SIP на основі запитів, які були направлені на цей сервер, та на основі відповідей, які він отримав від сервера. Усі відповіді від серверів до клієнтів спочатку проходять через балансувальник навантаження, який направляє відповіді до відповідних клієнтів. Відстежуючи ці відповіді, балансувальник навантаження може визначити, коли сервер завершив обробку запиту або виклику, і оновити дані про навантаження сервера в кластері серверів SIP. Сервер балансування навантаження фіксує момент надходження повідомлення INVITE, яке відповідає ініціації нового

виклику, та збільшує лічильник викликів на одиницю. У момент отримання відповіді 200 OK на запит BYE сервер балансування навантаження фіксує закінчення виклику та зменшує лічильник на одиницю.

Відповідно до RFC 3261 у протоколі SIP сесії поділяються на транзакції INVITE та non-INVITE (транзакції №1 і №2 на рис. 4.1. відповідно). У процедурі встановлення з'єднання створення нового сеансу ініціюється повідомленням INVITE, а її очікування – приходом відповіді 200 OK на пронос BYE, які відповідають тій самій сесії. Отримання відповіді при балансуванні навантаження за кількістю активних сеансів 200 OK — індикатор того, що проксі-сервер завершив обробку сеансу його ресурси були звільнені.

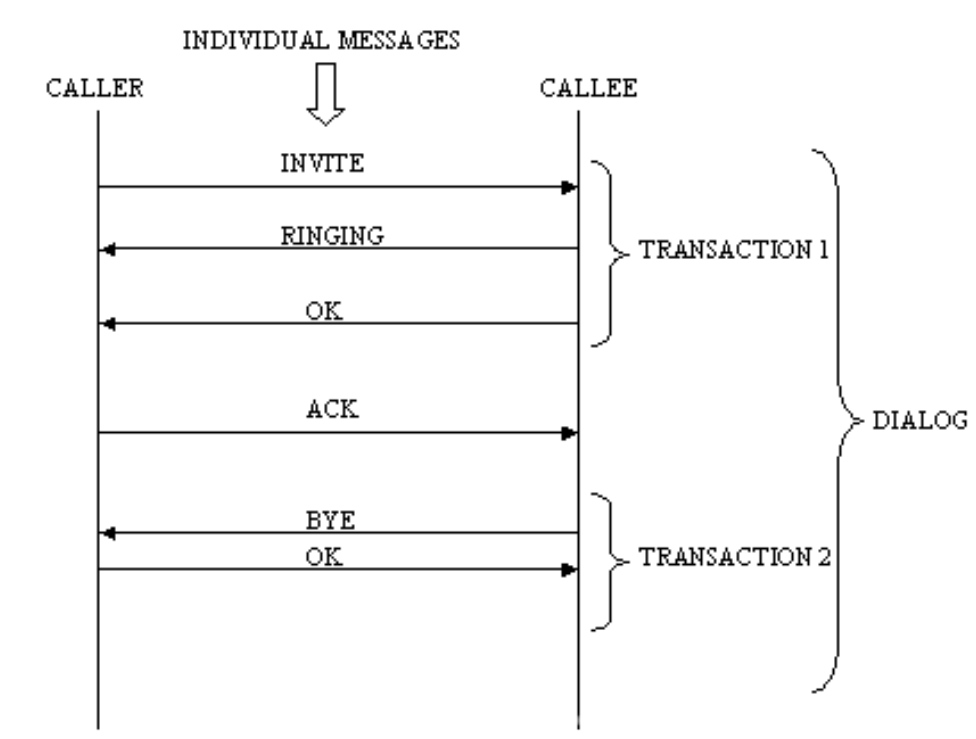


Рис. 4.1. Сценарій обміну сигнальними повідомленнями за протоколом SIP

Недоліком такого способу моніторингу навантаження на SIP сервер є те, що не завжди кількість викликів (сесій) відображає навантаження на сервер. Між транзакціями протягом сесії є тривалі періоди неактивності. Крім того, деякі сесії можуть включати різну кількість транзакцій, мати різну важливість, складність обробки, а отже – споживати різну кількість ресурсів сервер. Тому

пропонується використовувати як метод оцінки навантаження сервера підрахунок трансляційних коментарів сервер обслуговує в даний момент. Метод балансування навантаження на основі кількості активних транзакцій підраховує кількість запитів, які належать і обслуговуються на конкретному сервері в кластері. При отриманні запиту, наприклад, на встановлення нового INVITE з'єднання, або запиту на відключення, або запиту на реєстрацію, лічильник транзакцій на сервері балансує, навантаження збільшується на одиницю. Коли приходить відповідь 200 OK на відповідний запит, сервер балансування навантаження знижує лічильник до одиниці.

Наприклад, у протоколі SIP запит INVITE важливіший, ніж запит BYE. Термінальна машина для транзакцій INVITE набагато складніша, ніж для транзакцій без INVITE. Тому ці відмінності полягають у значущості обробки

для різних транзакцій, це необхідно враховувати при прийнятті рішення про балансування навантаження і для того, щоб врахувати ці особливості транзакцій протоколу SIP пропонується ввести вагу для кожної транзакції, яка буде враховуватись під час розрахунку кількості транзакцій на сервері балансування навантаження. Аналізуючи заголовок вхідного запиту, сервер присвоює цьому запиту вагу і збільшує його вміст до значення цієї ваги. Наприклад, сервер зараз обробляє транзакцію INVITE (з відносною вагою 1,75) і транзакцію BYE (з відносною вагою 1), тоді його завантаження становитиме 2,75. Нові виклики надсилаються на сервер з найменшим значенням лічильника транзакцій. Вага може бути введена не тільки для INVITE і BYE транзакцій, але і для всіх інших, передбачених у протоколі SIP.

Рівень завантаження SIP-сервера залежить від завантаженості ресурсів цього сервера:

$$E_i = 1 + \alpha - \sum_{j=1}^k r_j * \{r_{+j}r_2 \dots r_k\}^{-1} \left[\begin{array}{c} L(P_{1i}) \\ L(P_{2i}) \\ \dots \\ L(P_{ki}) \end{array} \right]^{-1}, \quad (4.1)$$

Де α – дуже маленьке число, наприклад 0,0001; r_j – зважені параметри ресурсу, які використовуються для моніторингу завантаження сервера ($j = 1, 2, \dots, k$, де k – число параметрів). Зважені параметри ресурсу налаштовуються адміністратором в залежності від функціональності сервера, його продуктивності. $L(P_{ki})$ – завантаження k -го ресурсу та i -го SIP-сервера. Для запропонованого способу визначення завантаження сервера здійснюється за кількістю активних транзакцій, що обробляються в даний момент з цим сервером, враховуючи відносну вагу цих транзакцій, ми маємо:

$$E_i = 1 + \alpha - r_i^{-1} [\sum_{j=1}^n L_{ji}(c_j, T_j)],^{-1}, \quad (4.2)$$

де c_j - відносна вага j -ї транзакції T_j ($j = 1, 2, \dots, n$, де n – кількість транзакцій, наприклад, транзакції INVITE, BYE, OPTION, REGISTER для мережі SIP).

Цей спосіб балансування навантаження за кількістю активних транзакцій дозволяє уникнути зворотного зв'язку та постійного обміну інформацією між серверами кластера та сервером балансування навантаження для збору поточних показників завантаження ресурсів, і базується лише на показаннях лічильники транзакцій на сервері балансування.

4.3. Дослідження ефективності методу

У першій частині було проведено аналіз методів балансування навантаження на серверах SIP. Для оцінки ефективності раніше відомих методів та запропонованого методу було проведено моделювання в середовищі CPN Tools. Моделювання проводилося на основі моделі, описаної вище у формі розфарбованих мереж Петрі. У ході моделювання припустили, що кластер серверів складається з 8 серверів з однаковими обчислювальними потужностями та з пропускною здатністю 37 дз/сек. Інтенсивність вхідного трафіку змінювалася від 1 до 37 дз/сек. В результаті метод балансування, що

розподіляє навантаження між серверами за показником активних транзакцій та враховує відносну вагу цих транзакцій, дав найкращі показники за часом встановлення з'єднання та пропускної спроможності кластера SIP-серверів, що відображено на рис.4.2., 4. 3.

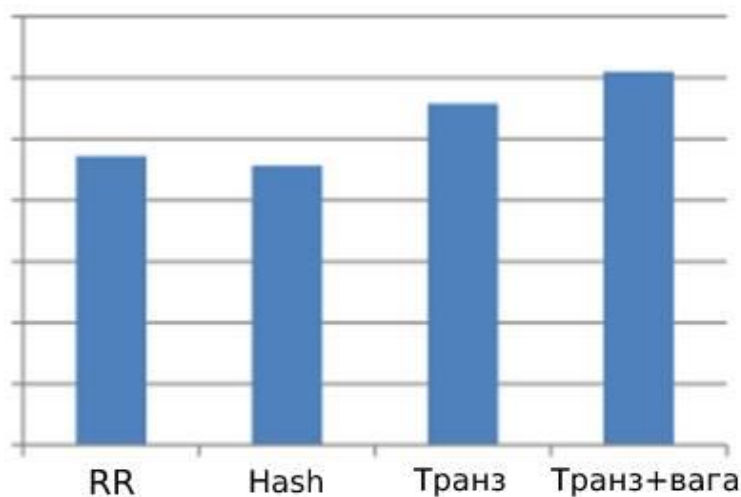


Рис. 4.2. Залежність пропускної спроможності кластера з 8 серверів від вхідного навантаження

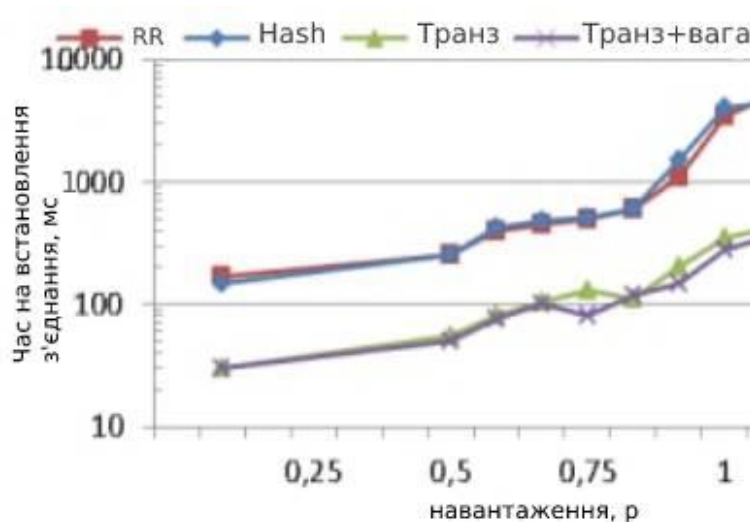


Рис. 4. 3. Залежність часу встановлення з'єднання кластера з 8 серверів від вхідного навантаження

ВИСНОВКИ ДО РОЗДІЛУ 4

Результати досліджень показують, що запропонований метод балансування завантаження за кількістю транзакцій на сервері має час

встановлення з'єднання в умовах високого навантаження в 10 разів менше, ніж інші підходи, а в умовах середнього навантаження – у 5-6 разів менше. Використання цього методу для балансування навантаження забезпечує на 25% кращу продуктивність серверів SIP-кластера, ніж стандартний метод, заснований на hash-алгоритмі, і на 14% кращу продуктивність, ніж циклічний метод розподілу алгоритму Round Robin (RR). Показник ефективної пропускної спроможності сервера при піковому навантаженні в кластері на 20% вище, ніж у разі застосування інших методів балансування навантаження.

РОЗДІЛ 5

ОХОРОНА ПРАЦІ

Охорона праці - це здійснення комплексу заходів технічного характеру (огородження небезпечних місць на виробництві, впровадження безпечного обладнання, зміна технологій з метою виключення робіт, небезпечних для життя і здоров'я людей) і санітарно-гігієнічних заходів (раціональне освітлення, охорона здоров'я людей на виробництві, огородження небезпечних місць, створення сприятливого мікроклімату у виробничих приміщеннях тощо), які забезпечують нормальні умови праці суб'єкта – працівника відділу ІТ.

Робоче місце радіотехніка розташоване в адміністративній будівлі, яка має гарне природне освітлення та обладнана лампами денного світла (ДБН В.2.5-28-2006 «Природне і штучне освітлення»). Комп'ютери налаштовують доставку, щоб зареєструвати технічне мислення заводів-виробників і гігієнічні стандарти правил включення пакувальних заводів. Робочі місця з дисплеями розташовують на відстані 2,5 м одне від одного. Приміщення мають хорошу природну вентиляцію, що відповідає прийнятим стандартам, та оснащені системою кондиціонування. Протягом робочого дня підтримується відносно стабільна температура (близько 19 °С), яка протягом робочого часу змінюється не більше ніж на 2 °С. Відносна вологість є історичною нормою в під'їзді (ДСН 3.3.6.042-99 Санітарні норми мікроклімату виробничих приміщень).

5.1. Опис робочого місця та умови праці співробітника ІТ-відділу.

Робоче місце для роботи з дисплеями влаштовують так, щоб вікна та освітлювальні прилади не поміщалися в поле зору працівника. Відеотермінали встановлюють під кутом 90-105 градусів до вікон і на відстані не менше 2,5-3 м від стіни з вікнами.

До поля зору працюючого з дисплеєм не повинно потрапляти поверхні, які мають властивість віддзеркалювання. Покриття столів матове з коефіцієнтом 0,25 - 0,4. Відстань між робочими місцями з ПК не меншою 1,5 м у ряду та не меншою 1 м між рядами. ПК розміщені не ближче 1 м від джерела тепла. Відстань від очей користувача до екрану становить 500 - 700 мм, кут зору - 10 - 20 градусів, але не більше 40 градусів, кут між верхнім краєм відео терміналу та рівнем очей користувача повинен бути меншим 10 градусів. Найбільш вигідне є розташування екрану перпендикулярно до лінії зору користувача.

5.2. Перелік шкідливих та небезпечних чинників, що діють на суб'єкта.

Шкідливі та небезпечні фактори, що виникають при роботі з обладнанням систем передачі даних, класифікуються згідно з ГОСТ 12.0.003-74. Такі фактори, які призводять до раптового погіршення здоров'я людини, називають небезпечними. Шкідливими вважаються ті чинники, які при контакті з організмом людини при порушенні вимог безпеки можуть спричинити нещасні випадки на виробництві, професійні захворювання або відхилення у стані здоров'я. (ДСанПіН 3.3.6.096-2002 Державні санітарні норми і правила при роботі з джерелами електромагнітних полів.)

Під час експлуатації обладнання системи передачі даних виникають наступні групи шкідливих і небезпечних виробничих факторів:

- мікроклімат робочого приміщення (температура, вологість, швидкість руху повітря, теплове випромінювання)
- збільшення сили електричного струму, який може пройти через людину і заповдіяти їй безпеку;
- підвищений рівень електромагнітного випромінювання;
- підвищений рівень шуму на робочому місці;
- небезпека ураження електричним струмом;

5.2.1. Мікроклімат (підвищена або занижена вологість, швидкість руху повітря)

Метеорологічні умови — це різні поєднання температури, вологості і швидкості руху повітря. Температура повітря у виробничих умовах залежить від кількості теплоти, переданої джерелами тепла і працівникам, об'єму приміщень, циркуляції повітря (природної чи штучної), інтенсивності тепла, передача через зовнішні стіни та огорожі. Регулюючи опалення, можна підтримувати потрібну температуру повітря в приміщенні.

Відносна вологість повітря під час відпустки їжі залежить від кількості води, що випаровується в приміщенні, метеорологічних умов поза приміщенням і може регулюватися вентиляцією та кондиціонуванням.

Швидкість повітря регулюється природною і штучною вентиляцією.

Тривале включення несприятливих погодних умов на ті самі продукти для програмування та добробуту, зниження продуктивності праці та визнання прийнятності радіо.

5.2.2. Підвищене значення електричного струму, що може проходити через людину і становити небезпеку для нього

Під час роботи з персональними комп'ютерами існує ризик ураження електричним струмом. Одним із способів захисту користувачів від ураження електричним струмом є заземлення пристрою. Тому доречно навести приклад розрахунку захисного заземлення.

Захисне заземлення є ефективним способом забезпечення безпеки людей, які працюють з електрообладнанням. Металеві корпуси електричних машин, апаратів, каркаси розподільних щитів та інших металоконструкцій, приєднаних до електроустановки, повинні бути заземлені. Штучний заземлювач являє собою замкнутий контур з 10 труб довжиною 2 м і діаметром 0,2 м, встановлених на глибині 1 м і з'єднаних заземлюючим поясом.

5.2.3. Електромагнітне випромінювання

Джерелами випромінювання електромагнітних полів (ЕМП) у апаратурі систем передачі даних є монітори комп'ютерів, елементи блоків живлення, принтери.

Систематичний вплив ЕМП на працюючого з рівнями, що перевищують припустимі, приводить до порушення стану їхнього здоров'я. При цьому можуть виникати зміни в нервовій, серцево-судинній, ендокринній і іншій системах організму людини (ДСН 3.3.6.039-99 Державні санітарні норми виробничої загальної та локальної вібрації) .

Ступінь шкідливого впливу ЕМП на організм людини визначається напруженістю електромагнітного поля, довжиною хвилі і тривалістю перебування організму в зоні діяльності ЕМП. (ДСН 3.3.6.037-99 Санітарні норми виробничого шуму, ультразвуку та інфразвуку.)

У залежності від відстані до джерела випромінювання працюючий може знаходитися або в зоні індукції, або в хвильовій зоні (зоні випромінювання).

Зона індукції :

розташовується поблизу джерела випромінювання на відстані

$$R \leq \frac{L}{2 \times \pi},$$

Хвильова зона:

розташовується на відстані

$$R > \frac{L}{2 \times \pi}, \text{ при } R \gg F^2 / L,$$

де L - довжина хвилі, $\pi=3,14159$,

F - геометричні розміри джерела випромінювання.

5.2.4. Розрахунок швидкості руху повітря в приміщенні ІТ-відділу, враховуючи основні санітарно-гігієнічні параметри

Загальний тепловий баланс суб'єкта характеризується рівнянням:

$$(1) \Delta Ql = Ql \pm Ql^k \pm Ql^r - Ql^v - Ql^m - Qf + q$$

де ΔQl – надлишок або недостача тепла в організмі;

Ql – теплопродуктивність організму;

Ql^k, Ql^r, Ql^v – складові теплообміну людини конвекцією, випромінюванням і за рахунок затрат тепла на випаровування вологи;

Ql^m – витрата тепла на механічну роботу;

Qf – тепло, затрачене на фізіологічні процеси (не перевищує 11,6 Вт);

q – сприйняте людиною променеве тепло сонячної радіації.

Для розрахунку складових теплообміну існує також багато емпіричних формул, складених лікарями-гігієністами.

Витрата тепла Ql^m зазвичай становить від 5 до 35 % від додаткових тепловиділень, що пов'язано звиконанням фізичної або розумової роботи. Із аналізу теплового балансу, прийнявши припущення, що люди знаходяться в стані спокою,

впливає:

– теплопродуктивність для стану спокою у сидячому положенні становить

$$Ql = 90 \text{ Вт}, \text{ в тому числі } Ql^v = 30-35 \text{ Вт.}$$

– конвективна тепловіддача людини

$$(2) Ql^k = Fd \cdot f \cdot (t_{od} - t_v) \cdot 12.1 \cdot V^{0.5},$$

де $Fd = 1,1 \text{ м}^2$ (поверхня дю-Буа в сидячому стані);

f – відношення площі поверхні одягненого й оголеного тіла, що для літнього одягу дорівнює 1,1;

t_{od} – температура одягу (35°C);

Ql' – променева тепловіддача суб'єкта (переважно становить 5–35 % від додаткових тепловиділень, пов'язаних з виконанням фізичної або розумової роботи).

Використовуючи $t_p = 0,5(t_v + t_r)$, визначаємо швидкість руху повітря V в приміщенні відділу технічної підтримки:

$$V = \frac{1}{106} \left(\frac{Q_{ia} + q}{35 - t_p} - 2.5 \right)$$

(3)

де $t_v = 22^{\circ} \text{C}$, а $t_r = 24^{\circ} \text{C}$

$V = 0.2 \text{ м/с}$

5.3. Пожежна безпека

Відповідно до НАПБ Б.03.002-2007 «Норми визначення категорій приміщень, будинків та зовнішніх установок за вибухопожежною та пожежною небезпекою» приміщення відділу пожежної безпеки відносяться до категорії Д, оскільки в них містяться негорючі рідини. . . і матеріали в холодному стані.

В приміщенні відділу ІТ знаходяться три переносні вогнегасники вуглекислотні типу ВВК-5, які відповідають вимогам НАПБ А.01.001-2004 Правил пожежної безпеки в Україні. При експлуатації вогнегасників необхідно керуватися Законом України "Про пожежну безпеку" (3745-12), національними та міжнародними стандартами, ДНАОП 0.00-1.07-94 та іншими нормативно-правовими актами.

В системі пожежної сигналізації використовуються бездротові димові сповіщувачі Ajax WS-501, три з яких встановлені в кімнаті технічної підтримки. Площа, яку охоплює один датчик Ajax WS-501, становить 20 м^2 .

Адресна пожежна сигналізація використовується в будівлі, де знаходиться ІТ-відділ. Це сучасна та високонадійна система, яка чітко визначає місце пожежі. Незважаючи на високу вартість комплектуючих, він часто обходиться дешевше у використанні.

5.3.1. Правила евакуації з приміщень.

У разі виникнення пожежі необхідно евакуювати обстежуваного з приміщення ІТ-відділу відповідно до плану евакуації на (рис. 5.1). У плані евакуації показано два маршрути: основний, за яким обстежуваний повинен рухатися по лівому крилу коридору до сходової клітки. Відповідно до маршруту екстреної евакуації обстежуваний повинен рухатися по правому крилу коридору до дверей запасного виходу, розташованих біля туалетів.



Рис. 5.1. План евакуації

ВИСНОВОК ДО РОЗДІЛУ 5

Інтенсивне навчання про хорошу роботу збільшило можливість розв'язування машинної участі, яка разом із продуктами продукції не буде її одягом безпечним для життя здоров'я працівників.

Протягом всієї своєї історії, людство намагається зробити життя комфортнішим, зігріватися, виробляти і водночас виробляти свою ефективність і точність. Особистий протест проти книги дуже часто заплямує цільову сторінку статті додатком до її проклятої, і її створення є її власною провиною.

Завдяки отриманим знанням з розвитком цивілізації рівень безпеки людей поступово зростає.

У розділі викладено вимоги до організації робочого місця користувача персонального комп'ютера, визначено чинники, що сприяють процвітанню працівників зв'язку, розкрито освітні заклади духовного безсмертя в пробах.

На виробництві широко застосовуються технічні засоби для запобігання нещасним випадкам, професійним захворюванням і нещасним випадкам.

Як правило, у процесі трудової діяльності дуже сильно проявляється комплекс несприятливих факторів виробничого середовища, внаслідок чого можливі виробничі травми та професійна праця.

Велика розмаїтість зазначених вище факторів, зріст їх кількості і зміна рівня впливу, з одного боку, і обмеженість можливостей систем захисту людини, багатогранність прояву її психічних особливостей дозволяє стверджувати, що досягнення абсолютної безпеки нереальне.

Комплексний характер впливу факторів виробничого середовища на працівників вимагає комплексного системного підходу у вирішенні завдань попередження нещасних випадків на виробництві та професійних захворювань.

Практика показує, що профілактика нещасних випадків і професійних захворювань, як явищу, не може бути забезпечена запровадженням одного,

навіть дуже ефективного заходу. Для цього необхідно використовувати можливості комплексного підходу а саме:

- безпека виробництва;
- організація безпеки;
- виробнича санітарія;
- психологія безпеки;
- економіка охорони праці;
- чинної законодавчої охорони праці.

Політика підприємства в галузі охорони праці повинна бути спрямована на послідовне зниження шкідливих виробничих факторів і усунення найбільш небезпечних з них, постійне утримання в справному стані технічного обладнання та електрообладнання. Покращуючи умови та дотримуючись усіх вимог, підвищиться продуктивність праці.

РОЗДІЛ 6

ОХОРОНА НАВКОЛИШНЬОГО СЕРЕДОВИЩА

Досягнення гармонійного поєднання навколишнього середовища, суспільства та сталого розвитку вважається необхідною умовою для задоволення потреб нинішнього покоління без шкоди здатності майбутніх поколінь задовольняти свої потреби. Сталий розвиток як мета досягається шляхом збалансування трьох основних компонентів сталого розвитку.

Суспільні очікування щодо сталого розвитку, прозорості та підзвітності розвивалися разом із дедалі суворішими законодавчими ініціативами, посиленням тиску на навколишнє середовище через забруднення, неефективне використання ресурсів, поводження з відходами, зміну клімату, деградацію екосистем і втрату біорізноманіття.

Це спонукало організацію прийняти системний підхід до екологічного менеджменту та запровадити системи екологічного менеджменту для підтримки екологічної складової сталого розвитку.

6.1. Завдання системи екологічного управління

Важливість природокористування сьогодні важко переоцінити, адже серед сучасних глобальних проблем людства екологічні проблеми є найважливішими. Охороні довкілля та раціональному використанню природних ресурсів нині приділяється особлива увага державних структур та міжнародної спільноти. На порядку денному стоїть питання екологічної безпеки держави.

Системний підхід до управління навколишнім середовищем може забезпечити краще управління інформацією, яка буде корисною для досягнення довгострокового успіху та використання можливостей, які сприяють сталому розвитку:

- збереження стану навколишнього природного середовища шляхом запобігання негативному впливу на нього або пом'якшення їх;
- пом'якшення потенційного несприятливого впливу умов зовнішнього середовища на організацію;
- допомога організації у виконанні обов'язкових вимог дотримання;
- підвищення екологічної ефективності;
- контроль (або вплив) на методи розробки продуктів і послуг організації, їх виробництва, розподілу, споживання та утилізації з урахуванням аспектів очікуваного життєвого циклу, що сприятиме запобігання впливу на навколишнє середовище від непередбачуваного переходу від так звані від однієї стадії життєвого циклу до іншої протягом життєвого циклу;
- досягнення фінансових вигод і операційних вигод, які можуть виникнути в результаті впровадження екологічно прийнятних альтернативних підходів, які зміцнюють позиції організації на ринку;
- обмінюванню екологічною інформацією з відповідними зацікавленими сторонами. Цей стандарт, як і інші стандарти, не призначено для збільшення чи змінення правових вимог щодо організації. [16]

Система управління

Сукупність взаємопов'язаних або взаємодійних елементів організації, для формування політик і встановлення цілей та процесів, щоб досягати цих цілей. [16]

Система екологічного управління

Частина системи управління яку використовують для керування екологічними аспектами, виконання обов'язкових для дотримання відповідності вимог та розв'язання питань, пов'язаних з ризиками та можливостями. [16]

Екологічна політика

Наміри та спрямованість організації стосовно своєї екологічної дієвості, офіційно проголошені її найвищим керівництвом. [16]

Організація

Особа чи група людей, які мають свої власні функції з відповідальністю, повноваженнями та взаємовідносинами, щоб досягати своїх цілей. [16]

Найвище керівництво

Особа чи група осіб, яка спрямовує та контролює діяльність організації на найвищому рівні. [16]

Зацікавлена сторона

Особа чи організація, яка може вплинути на рішення чи діяльність, піддана впливу чи сприймає себе такою, що піддана впливу рішень або діяльності. [16]

Приклад

Замовники, громадські об'єднання, постачальники, регуляторні органи, неурядові організації, інвестори та працівники. [16]

Терміни стосовно планування

Довкілля; навколишнє середовище

Середовище, у якому функціює організація, охоплюючи атмосферне повітря, водні об'єкти, земельні ділянки, природні ресурси, флору, фауну, людей, а також взаємозв'язки між ними. [16]

Екологічний аспект

Елемент діяльності організації або її продукції чи послуг, який взаємодіє чи може взаємодіяти з довкіллям. [16]

Умови довкілля; умови навколишнього середовища

Стан або характеристика довкілля (4.2.1), визначені в певний момент часу. [16]

Вплив на довкілля; вплив на навколишнє середовище

Зміна в довкіллі (4.2.1), несприятлива чи сприятлива, яку цілком або частково спричинено екологічними аспектами (4.2.2.) організації (4.1.4) [16]

Ціль

Результат, який має бути досягнуто. [16]

Екологічна ціль

Ціль (4.2.5), встановлена організацією (4.1.4) відповідно до своєї екологічної політики (4.1.3) [16]

Запобігання забрудненню

Використання процесів (4.3.5), методів, засобів, матеріалів, продукції, послуг або енергії для уникнення, зменшення чи регулювання (окремо чи в поєднанні) утворення, викидання чи скидання будь-якого виду забруднювальних речовин або відходів, щоб зменшити несприятливі впливи на довкілля (4.2.4). [16]

Вимога

Сформульовані потреба чи очікування, загальнозрозумілі чи обов'язкові. [16]

Обов'язкові для дотримання відповідності вимоги правові та інші вимоги

Правові вимоги (4.2.8), які організація (4.1.4) повинна дотримувати, а також інші вимоги, яких організація дотримує чи вирішує дотримувати. [16]

Ризик

Вплив невизначеності.

Вплив — це відхил, позитивний або негативний, від очікуваного.

Невизначеність — це стан нестачі навіть часткової інформації стосовно розуміння чи знання події, її наслідку чи ймовірності.

Ризик часто характеризують посиланням на можливі події (як визначено в ISO Guide 73:2009, 3.5.1.3) та можливі наслідки (як визначено в ISO Guide 73:2009, 3.6.1.3) чи на їх поєднання. [16]

Ризик часто подають з погляду поєднання наслідків події (охоплюючи зміни в обставинах) і ймовірністю її виникнення (як визначено в ISO Guide 73:2009, 3.6.1.1). [16]

Ризики та можливості

Потенційні несприятливі наслідки (загрози) та потенційні сприятливі наслідки (можливості). [16]

6.2. Система заходів захисту здоров'я користувачів корпоративною системою Інтернет мережі.

Комп'ютеризація та інформаційні технології останнім часом посідають дуже важливе місце в житті людей і стають основою нових якісних змін у світі.

Кінець ХХ століття характеризується кардинальними змінами в різних сферах життя. Це стосується політичних та економічних перетворень, а також ряду інших, які суттєво впливають на рівень життя населення світу. В першу чергу варто звернути увагу на те, що компанія, в якій досягнення промислових технологій завжди були на першому місці, першочергову увагу почала приділяти інформаційним технологіям. Слід підкреслити, що стрімкий розвиток індустріальних технологій тривав приблизно дві третини ХХ століття, а розвиток інформаційних технологій набув глобальних масштабів майже в останнє десятиліття.

Є показники, які чітко характеризують такий стан. Таким чином, кожні п'ять років обсяг ринку інформаційних систем буде подвоюватися. Темпи розвитку світового ринку інформаційних технологій і систем перевищують дванадцять відсотків на рік. Є дані, які підтверджують, що лише інформаційні технології та системи створюють близько десяти мільйонів робочих місць і забезпечують значне зростання внутрішнього національного продукту.

Важливо, що в США інформаційна сфера знаходиться під увагою президента Б. Клінтона і створена Програма національної інформаційної інфраструктури, яка підкреслює, що інформація є найважливішим національним ресурсом, найважливішим товаром, а також найважливішим чинник державного управління. Це можливо лише за умови використання сучасних інформаційно-телекомунікаційних технологій, і це є пріоритетним завданням інформатизації органів державної влади в США та багатьох розвинених країнах.

Очікуваним стратегічним результатом Програми є створення сучасної національної інформаційної структури, яка стане основою подальшого розвитку країни. Ця Програма визначає наступні стратегічні цілі: досягнення лідируючих позицій у більшості галузей інформаційних технологій, створення максимально відкритої демократії, підготовка молоді до переходу в 21 століття, створення нових робочих місць, використання інформаційних систем для розвитку освіти, культури, охорона здоров'я, створення умов для надання якісних та дешевих інформаційних послуг.

Слід зауважити, що на європейському континенті також значна увага приділяється проблемам інформатизації та інформаційних систем. Діє Програма технологічного розвитку Європейського Союзу, яка включає в себе Європейську програму стратегічної інформатизації, головною метою якої є скорочення технологічного відставання від США та Японії та зменшення залежності інформаційних технологій від цих країн. Його пріоритетними напрямками є: розробка та впровадження багатьох основних транс'європейських інформаційних послуг; підвищення технологічної та промислової ефективності у сфері комп'ютеризації та інформаційних систем; впровадження інформаційних технологій у соціальну сферу; розширення сфери використання інформаційних технологій; активізація розвитку нових інформаційних технологій.

Автори не можуть детально розглянути всі напрямки комп'ютеризації, тому для прикладу вибирають два з них, найважливіші.

Перший з них включає два основних проекти: «Інтернет для шкіл» та «Навчання в інформаційному суспільстві». У рамках реалізації цих проектів країни Європейського Союзу широко впроваджують нові підходи та методики, які орієнтовані на специфіку та можливості інформаційних технологій.

У багатьох європейських країнах реалізації цих проектів приділяють увагу на урядовому рівні. Наприклад, уряд Франції реалізує програму, розраховану на три роки. Він надає доступ до освітніх мультимедійних засобів

і електронну адресу всім студентам, школярам і вчителям. Супутниковий зв'язок забезпечить підключення до мережі всіх шкіл, розташованих у віддалених і важкодоступних районах. Французькі оператори зв'язку отримали пропозицію надати школам пільгові тарифи.

Іншим важливим напрямком впровадження інформаційних систем є електронна комерція. Прогнозується десятикратне збільшення використання електронної комерції в приватному секторі.

Досвід впровадження та використання інформаційних технологій і систем дозволяє говорити про великі потенційні можливості цієї сфери у вирішенні соціальних проблем суспільства. Україна – молода незалежна держава, яка не стоїть осторонь цих процесів. Можна стверджувати, що інформаційні системи все більше проникають у всі сфери життя суспільства.

Але слід підкреслити, що процеси створення та впровадження інформаційних систем відбувалися в нашій країні майже стихійно. Важливі кроки у напрямку ефективної координації цих робіт здійснено Національним агентством з інформатизації під керівництвом Президента України. Ці зусилля за підтримки державних, наукових, освітніх та недержавних установ завершилися перемогою. Верховна Рада України прийняла три закони України: «Про концепцію Національної програми інформатизації», «Про Національну програму інформатизації», «Про затвердження завдань Національної програми інформатизації на 1998-2000 роки». Ще в державному бюджеті України на 1998 рік були передбачені кошти на фінансування заходів щодо виконання завдань Національної програми інформатизації.

Окремо слід виділити рішення Київської державної адміністрації щодо фінансування та розширення робіт у сфері інформаційних систем та їх впровадження в різних галузях міста Києва. Цю роботу передбачалося розпочати Концепцією і здійснювати у 1999-2004 рр. Керівництво цими роботами керівництво міста Києва доручило Інституту проблем реєстрації інформації НАН України.

Розглядаючи реальний стан впровадження інформаційних систем і технологій в Україні, слід зазначити, що заплановане фінансування робіт досі не реалізовано, внаслідок чого процеси використання інформаційних технологій значно відстають. При цьому спостерігається така парадоксальна картина. Ще п'ять-сім років тому переважала думка, що ми суттєво відстаємо від розвинених країн світу. Наздогнати їх Україні вважалося неможливим. Але сьогодні можна стверджувати майже протилежне.

Наразі все ще існує затримка у впровадженні найсучасніших інформаційних систем і технологій, яка, однак, скоротилася до одного-двох років. Це не означає сферу впровадження інформаційних систем. За цим показником Україні важко конкурувати з розвиненими країнами. Але ми можемо навести декілька прикладів створення та впровадження інформаційних систем у різних галузях народного господарства України, які повністю відповідають світовим стандартам і є гордістю нашої країни.

Сучасною тенденцією розвитку корпоративних інформаційних систем та їх технічної (транспортної) бази – корпоративних мереж є перехід до інтегрованої передачі та трансляції даних (за оцінками експертів, інтегрована робота становитиме близько 84% від загальної кількості у 2000 р. порівняно з до 48% у 1996 році). Мотиви такого переходу очевидні. Це істотна економія при використанні найдорожчого ресурсу мережі - каналів зв'язку. По одному каналу передаються дані і голос (телефонні розмови), а також факси і відеоінформація, що гарантує багаторазове зниження вартості оренди каналів або їх прокладки.

Технічно це здійснюється шляхом мультиплексування інтегрованої передачі та подальшого демультиплексування окремих інформаційних потоків. Різні класи мультиплексорів дозволяють інтегрувати інформаційні потоки різного обсягу, що надходять від невеликих віддалених філій, а також великих регіональних філій по каналах від 9,6 Кбіт/с до 2,048 Мбіт/с і вище. В окремих випадках можливе використання додаткових механізмів, встановлених в мультиплексорах, що підвищують ефективність використання

пропускної здатності каналу зв'язку. Особливо з урахуванням різниці в характері денного і нічного трафіку: більше голосових каналів (телефонних розмов) вдень і каналів даних вночі.

Одночасно з інтеграцією вигідніше згущувати мовну інформацію. Наприклад, одна з найсучасніших технологій стиснення голосу дозволяє одночасно передавати до 13 телефонних дзвінків по одному стандартному каналу 64 Кбіт/с практично без втрати якості звуку мови.

Використання комплексної передачі інформаційних потоків забезпечує кожне робоче місце повним набором інформаційних послуг за виправданих витрат на їх підтримку. Зокрема, телефонні розмови між регіональними відділеннями перетворюються на розмови всередині установи, що забезпечує кращий контроль та безпеку, не кажучи вже про суттєве здешевлення міжміського зв'язку.

Якщо повернутися до питання про стан телекомунікаційної сфери України, то необхідно звернути увагу на відносно складні стартові умови для будівництва територіально розподілених мереж. Першою і найбільшою фундаментальною перешкодою є якість існуючих стандартних каналів зв'язку (тональних частотних каналів). Їх низька якість призвела до того, що ряд компаній, світових лідерів у сфері каналного обладнання, були змушені адаптувати свою продукцію для українського ринку. Це дозволило частково вирішити цю проблему. Це також стало поштовхом для впровадження альтернативних технологій, наприклад бездротового зв'язку.

Напрямки розвитку технологій побудови бізнес-мереж в Україні можна вважати визначеними. Українські фахівці грамотно аналізують накопичений досвід помилок і успішних рішень в інших країнах, розробляють власні концепції, не гірші за світові аналоги. Наразі можна сказати, що Україна має власний потенціал знань для побудови територіально розподілених мереж, здатних задовольнити потреби кожного корпоративного клієнта. В якості базових технологій використовуються Frame Relay, X. 25, ISDN, ATM.

Сучасні телекомунікаційні пристрої, багатофункціональні та «прозорі» для різних протоколів, дозволяють побудувати мережу компанії, використовуючи всі переваги цих протоколів. Додаткова гнучкість побудови характерна для так званої віртуальної приватної мережі, побудованої частково або повністю на основі оренди послуг публічних мереж.

Максимальна гнучкість рішень досягається за рахунок реалізації концепції накладених мереж. Ця концепція активно просувається на ринку, оскільки надає користувачам доступ до новітніх комунікаційних технологій в умовах елементарної відсутності як низькошвидкісних, так і міжміських каналів зв'язку. Деяке налаштоване телекомунікаційне обладнання (наприклад, мультиплексори) дозволяє отримувати послуги ISDN також по аналогових виділених лініях у приватній корпоративній мережі. Або ви можете з'єднати філії однієї компанії з накладеною мережею Frame Relay, не будуючи власної загальнонаціональної мережі того самого стандарту.

Вдалу побудову корпоративної мережі можна порівняти з успішним виконанням складної симфонії. Складових успіху може бути багато, але тут головним чином досвідчений диригент має представити свою концепцію і відомий твір.

Підхід до вибору технології передачі інформації є індивідуальним для країн з різним рівнем розвитку телекомунікаційної інфраструктури. Фактори, що впливають на цей вибір, мають економічний, географічний і політичний характер і пов'язані, насамперед, з політикою національних телекомунікаційних компаній. Наприклад, у Німеччині та Австрії, де оператори мереж зв'язку послідовно інвестували в розвиток послуг ISDN, корпоративні мережі будували за цією технологією. У країнах Латинської Америки, Іспанії та Португалії мережі будуються на цифрових лініях і пристроях X.25 з поступовим переходом на технологію Frame Relay. Ще один приклад – розвиток корпоративних банківських мереж в Україні. Узагальнюючи досвід кількох українських банків (Національний банк України, ПромІнвестБанк, Банк «Україна», Укрсоцбанк, Приватбанк),

зазначимо, що наразі ці мережі побудовані на аналогових лініях з модемним зв'язком за протоколом X.25 з переходом на супутник. канали приватних систем зв'язку на основі технології Frame Relay з інтегрованою передачею ширококомовних даних і практичною реалізацією концепції «віддаленого офісу».

Загалом, бізнес-мережа може бути побудована на різних каналах зв'язку - від виділених ліній (аналогових і цифрових) до комутованих цифрових, включаючи оптичні волокна, супутникові, радіо- та мікрохвильові канали та на основі різних протоколів і технологій ISDN, X 25, Frame Relay. і банкомат.

ВИСНОВКИ ДО РОЗДІЛУ 6

Нормативно-правова база системи управління навколишнім середовищем в Україні досить широка, охоплює всі сфери господарювання, але недостатньо деталізована та чітка, що дозволяє підприємствам уникати екологічної відповідальності. Після вивчення європейського досвіду впровадження екологічного менеджменту доцільним буде застосування більш жорстких стандартів контролю за виконанням зобов'язань, в окремих випадках також створення нових податків на викиди забруднюючих речовин у навколишнє середовище та інші види антропогенної діяльності. , щодо використання природних ресурсів і споживання матеріальних ресурсів, що стимулюватиме прискорення технічного прогресу, зменшення забруднення навколишнього середовища та споживання природних ресурсів. Крім того, недостатнє фінансування цієї системи в Україні суттєво обмежує розвиток регіонів, а тому доцільно впроваджувати заходи щодо залучення «зелених» інвестицій та підтримки зростання екологічної свідомості не лише в суспільстві, а й на підприємствах.

Встановлено, що підприємці не зацікавлені у впровадженні екоменеджменту, оскільки для такого бізнесу відсутні стимули та пільги (субсидії, пільги).

ВИСНОВКИ

Сьогодні SIP є універсальним стандартом обміну даними в мережі. Фахівці вважають його найперспективнішим і найбільш часто використовуваним. SIP-телефонія доступна там, де є інтернет зі швидкістю не менше 64 Кбіт/с. Тобто практично всюди.

IP-телефонія дозволяє присвоювати номер не локації, а конкретному користувачеві, що відкриває широкі можливості, наприклад, для аналітики.

SIP-сервер є найважливішою частиною IP-АТС. Ця служба в основному керує SIP-комутаторами між SIP-клієнтами/кінцевими точками. У літературі SIP-сервер також може називатися SIP-проксі або SIP-реєстратор (Registrar).

Хоча цей сервер можна вважати найважливішою частиною телефонної системи на основі SIP, його завдання обмежуються керуванням і підтримкою сеансів SIP. Сам SIP-сервер не обробляє медіа-потоки - це робить окремий медіа-сервер (Media server) за протоколом RTP. У реалізаціях IP-АТС SIP-сервер і супровідний медіа-сервер завжди розташовані на одному фізичному сервері.

Продуктивність серверів обробки дзвінків визначається кількістю дзвінків, які здатні обслуговувати сервер в одиницю часу. Коли інтенсивність дзвінків перевищує пропускну здатність сервера, він переходить у режим навантаження, час очікування на сервері збільшується. Клієнт чекає на відповідь на запит і, не дочекавшись, надсилає повторний запит на встановлення з'єднання. Це призводить до ще більшого збільшення вхідного навантаження на сервер і посилює навантаження. У цій ситуації за відсутності методу контролю навантаження продуктивність сервера різко погіршується. SIP-проксі-сервери, які є основними елементами, що виконують функції встановлення/підтримки/роз'єднання з'єднання мережах IP-телефонії, можуть бути схильні до різкого збільшення вхідного навантаження по

низці причин: через збої на елементах мережі, під час проведення телевізійних шоу або при умисних DoS-атаках[15]

У специфікації протоколу SIP немає точних вказівок, як реагувати на навантаження. Протокол SIP передбачає базовий метод керування навантаженнями, який використовує код помилки 503 (Service Unavailable). У разі, якщо SIP-сервер не може обробити запит через тимчасове навантаження, він відхиляє запит із кодом помилки 503. Це повідомлення означає, що відправник повинен повторити відправку повідомлення через деякий час або направити запит на інший сервер. Перевантажений сервер також може додати Retry-After заголовок у SIP-повідомлення, вказавши час у секундах, протягом якого сервер не слід звертатися. Відправник припиняє надсилання повідомлень на цей період часу, повідомлення надсилаються на альтернативний сервер, а відправлення повідомлень на перевантажений сервер відновлюється після закінчення інтервалу часу, вказаного в заголовку. Якщо альтернативний сервер відсутній, і заголовок Retry-After не вказано, сервер SIP повинен відправити відповідь 500 (Server Internal Error) агенту користувача.[15]

Хоча цей метод знижує навантаження на перевантажений проксі-сервер, але призводить до перенаправленню трафіку, який викликав перевантаження, на інший проксі-сервер, який потім також може перевантажитися.[15]

У режимі великого навантаження перевантажений сервер SIP може повідомляти про стан сусіднім серверам і просити їх зменшити навантаження. Але на практиці такі методи боротьби з навантаженням із зворотним зв'язком не завжди реалізовані.[15]

Тому дуже важливо, щоб сервери SIP могли локально захистити власні ресурси, необхідні для обробки повідомлень, за рахунок реалізації відповідних локальних методів боротьби з перевантаженням. Як правило, такі методи включають процедуру вимірювання навантаження на систему і порівняння з фіксованим значенням їжак. Коли граничне значення завантаження системи досягнуто або перевищено, запускається процедура боротьби з

перевантаженням. Існує кілька локальних методів боротьби з навантаженням, описаних у літературі, які для вимірювання навантаження на систему можуть використовувати довжину черги буфера, завантаження процесора, кількість активних INVITE-транзакцій, середній час затримки пакетів у буфері сервера та інші параметри. У цих методах використовується стандартна схема обслуговування викликів, передбачена, коли в режимі високого навантаження все нові запити на з'єднання відкидаються відповіддю 503, хоча деякі з цих запитів можуть бути близькі до встановлення з'єднання.[15]

Таким чином, виходячи з аналізу проблеми навантаження на SIP-сервери, можна зрозуміти, що без додаткового контролю навантаження у вигляді окремого балансувальника не може бути впевненості в коректності роботи сервера або кластера серверів при високих навантаженнях.

Однією з головних особливостей балансування навантаження SIP-сервера є те, що запити, які відповідають одному виклику, надсилаються на той самий сервер. Проксі-сервер вибирає буферний сервер лише для першого запиту виклику. Усі наступні запити, що відповідають цьому виклику, мають надходити на той самий сервер. Запропонований у роботі метод враховує цю властивість розподілу викликів між серверами. Відповідно до транзакційно-орієнтованого методу балансування навантаження, алгоритм вибирає сервер з найменшою кількістю активних завдань, які ще не завершені. Хоча концепція розподілу навантаження між серверами відповідно до найменшого обсягу роботи, що залишилася, застосовувалася в інших дослідженнях, відповідь на питання про те, як її ефективно реалізувати на практиці, часто залишається незрозумілою. Системі потрібен метод, який дозволить оцінити обсяг роботи, що залишився для кожного сервера, і прийняти рішення про оптимальний розподіл навантаження.

У нашій системі балансувальник навантаження може оцінити навантаження на сервер у кластері SIP на основі запитів, надісланих на цей сервер, і відповідей, отриманих від сервера. Усі відповіді від серверів до клієнтів спочатку проходять через балансувальник навантаження, який

направляє відповіді до відповідних клієнтів. Відстежуючи ці відповіді, балансувальник навантаження може визначити, коли сервер завершив обробку запиту або виклику, і оновити дані про навантаження сервера в кластері серверів SIP. Балансувальник навантаження зафіксує момент отримання повідомлення INVITE, що відповідає ініціації нового виклику, і збільшить лічильник викликів на одиницю. У момент отримання відповіді 200 OK на запит BYE сервер балансування навантаження фіксує закінчення виклику та зменшує лічильник на одиницю.

Як метод оцінки навантаження на сервер пропонується використовувати розрахунок трансляційних коментарів, які в даний момент обслуговує сервер. Метод балансування навантаження на основі кількості активних транзакцій підраховує кількість запитів, які належать і обслуговуються на певному сервері в кластері. Коли надходить запит, наприклад INVITE на створення нового підключення, запит на відключення або запит на реєстрацію, лічильник транзакцій на сервері вирівнюється, навантаження збільшується на одиницю. Коли на відповідний запит відповідає 200 OK, балансувальник навантаження зменшує лічильник до одиниці.

Результати дослідження показують, що запропонований метод балансування навантаження на основі транзакцій сервера має час встановлення з'єднання в 10 разів коротший за високого навантаження, ніж інші підходи, і в 5-6 разів коротший за середнього навантаження. Використання цього методу для балансування навантаження забезпечує на 25% кращу продуктивність для кластерних серверів SIP, ніж стандартний метод на основі хешування, і на 14% кращу продуктивність, ніж циклічний метод (RR). Показник ефективної пропускну здатності сервера при піковому навантаженні в кластері на 20% вище, ніж при використанні інших методів балансування навантаження.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. <https://uk.wikipedia.org/wiki/NGN> [Електронний ресурс]
2. <http://www.generallytech.ru/gentecs-629-1.html> [Електронний ресурс]
3. <https://studwood.net/1588251/tehnika/zadachi> [Електронний ресурс]
4. <https://www.znanius.com/3707.html> [Електронний ресурс]
5. <https://habr.com/ru/company/selectel/blog/250201/> [Електронний ресурс]
6. https://pt.nure.ua/wp-content/uploads/2020/01/142_kuzminykh_sip.pdf [Електронний ресурс]
7. <https://www.3cx.ru/voip-sip/sip-server/> [Електронний ресурс]
8. <https://new-tel.net/blog/chto-takoe-sip-server/> [Електронний ресурс]
9. <https://uk.wikipedia.org/wiki/VoIP> [Електронний ресурс]
10. <https://www.3cx.ru/voip-sip/voip-telephone/> [Електронний ресурс]
11. <https://dic.academic.ru/dic.nsf/ruwiki/18475> [Електронний ресурс]
12. https://studwood.net/2155279/tehnika/tehnologiya_multiservisnyh_setey_novogo_pokoleniya [Електронний ресурс]
13. https://pt.nure.ua/wp-content/uploads/2020/01/142_kuzminykh_sip.pdf [Електронний ресурс]
14. <https://studfile.net/preview/2918681/page:9/> [Електронний ресурс]
15. https://openarchive.nure.ua/bitstream/document/614/1/122_kuzmenyh_sip.pdf [Електронний ресурс]
16. https://osvita.ua/vnz/reports/econom_pidpr/21931/ [Електронний ресурс]