

Міністерство освіти і науки України  
Національний технічний університет  
«Дніпровська політехніка»

Інститут електроенергетики  
Факультет інформаційних технологій  
Кафедра безпеки інформації та телекомунікацій

ПОЯСНЮВАЛЬНА ЗАПИСКА  
дипломної роботи

магістра

(назва освітнього рівня)

галузь знань	<u>17 Електроніка та телекомунікації</u> (шифр і назва галузі знань)
спеціальність	<u>172 Телекомунікації та радіотехніка</u> (код і назва спеціальності)
освітній рівень	<u>магістр</u> (назва освітнього рівня)
кваліфікація	<u>магістр з телекомунікацій та радіотехніки</u> (код і назва кваліфікації)

На  
тему: «Дослідження фрактальних моделей мережевого трафіку для  
управління передачею на основі короткострокового  
прогнозування»

Виконавець: студент 6 курсу, групи 172М-21-1

Матвієнко Інна Миколаївна

(підпис)

(прізвище ім'я по-батькові)

Керівники	Прізвище, ініціали	Оцінка	Підпис
проекту	к.ф.-м.н., проф. Гусєв О.Ю.		
розділів:			
спеціальний	к.ф.-м.н., проф. Гусєв О.Ю.		
економічний	к.е.н., доц. Романюк Н.М.		
Рецензент			
Нормоконтроль	к.ф.-м.н., проф. Гусєв О.Ю.		

Дніпро  
2022

Міністерство освіти і науки України  
Національний технічний університет  
«Дніпровська політехніка»

Інститут електроенергетики  
Факультет інформаційних технологій  
Кафедра безпеки інформації та телекомунікацій

**ЗАТВЕРДЖЕНО:**  
завідувач кафедри  
безпеки інформації та телекомунікацій  
д.т.н., професор \_\_\_\_\_ Корнієнко В.І.  
« \_\_\_\_\_ » \_\_\_\_\_ 20\_\_ року

**ЗАВДАННЯ**

на дипломну роботу магістра

спеціальність \_\_\_\_\_ *172 Телекомунікації та радіотехніка*  
(код і назва спеціальності)

студента \_\_\_\_\_ *172М-21-1*  
(група)

\_\_\_\_\_ **Матвієнко Інна Миколаївна**  
(прізвище ім'я по-батькові)

Тема дипломного проекту **«Дослідження фрактальних моделей мережевого трафіку для управління передачею на основі короткострокового прогнозування»**

Наказ ректора НТУ "ДП" від \_\_\_\_\_ № \_\_\_\_\_

Розділ	Зміст	Термін виконання
<i>Стан питання. Постановка задачі</i>	Аналітичний огляд літератури по темі проекту	Вересень 2022
<i>Спеціальна частина</i>	Принцип роботи Дослідження фрактальних моделей мережевого трафіку для управління передачею на основі короткострокового прогнозування. Оформлення пояснювальної записки.	Жовтень 2022
<i>Економічний розділ</i>	Розрахунок капітальних витрат	Листопад 2022

Завдання видав \_\_\_\_\_  
(підпис)

Гусев О.Ю.  
(прізвище, ініціали)

Завдання прийняв до виконання \_\_\_\_\_  
(підпис)

Матвієнко І.М.  
(прізвище, ініціали)

Дата видачі завдання: 03 вересня 2022 р.

Строк подання дипломного проекту до ДЕК:

## РЕФЕРАТ

Пояснювальна записка: с., рис., табл., додатків, джерел.

**Об'єкт дослідження:** мережі з пакетною передачею інформації

**Предмет дослідження:** імітаційне моделювання мережі з пакетною передачею та короткостроковим прогнозуванням навантаження

**Мета дослідження:** розробка моделі мережі з пакетним трафіком для керування передачею на основі короткострокового прогнозування

### **Постановка задачі:**

Метою цієї роботи є розробка моделі мережі з пакетним трафіком для управління передачею на основі короткострокового прогнозування. Для реалізації поставленої мети необхідно вирішити такі завдання:

- 1 Виконати налітичний огляд моделей мережевого трафіку
- 2 Дослідити методи прогнозування навантаження в мережах з комутацією пакетів
- 3 Розробити імітаційну модель мережі з короткостроковим прогнозуванням
- 4 Дослідити характеристики мережі та оцінити ефективність отриманих результатів

В першому розділі виконаний аналітичний огляд літературних джерел по темі дипломного проекту. Здійснено постановку задачі роботи.

У другому розділі досліджено фрактальні моделі мережевого трафіку для управління передачею на основі короткострокового прогнозування

У третьому розділі виконано розрахунок капітальних витрат на розробку та дослідження фрактальних моделей мережевого трафіку для управління передачею на основі короткострокового прогнозування

ДОСЛІДЖЕННЯ ФРАКТАЛЬНИХ МОДЕЛЕЙ МЕРЕЖЕВОГО ТРАФКУ, КОРОТКОСТРОКОВЕ ПРОГНОЗУВАННЯ, СИСТЕМИ ПЕРЕДАЧІ ДАНИХ, МОДЕЛЬ СИСТЕМИ, ІМІТАЦІЙНЕ МОДЕЛЮВАННЯ.

## ABSTRACT

Explanatory note: pages, figures, tables, appendices, sources.

Research object: networks with packet transmission of information

Research subject: network simulation with packet transmission and short-term load forecasting

Research objective: development of a packet traffic network model for transmission control based on short-term forecasting

Formulation of the problem:

The purpose of this work is to develop a packet traffic network model for transmission control based on short-term forecasting. In order to realize the set goal, it is necessary to solve the following tasks:

- 1 Perform an analytical review of network traffic patterns
- 2 Investigate load forecasting methods in packet-switched networks
- 3 Develop a network simulation model with short-term forecasting
- 4 Investigate the characteristics of the network and evaluate the effectiveness of the obtained results

In the first chapter, an analytical review of literary sources on the topic of the diploma project is performed. The work task has been set.

In the second section, fractal models of network traffic are investigated for transmission control based on short-term forecasting

In the third chapter, the calculation of capital costs for the development and research of fractal models of network traffic for transmission management based on short-term forecasting is carried out

STUDY OF FRACTAL MODELS OF NETWORK TRAFFIC, SHORT-TERM FORECASTING, DATA TRANSMISSION SYSTEMS, SYSTEM MODELING, SIMULATION MODELING.

## СПИСОК УМОВНИХ СКОРОЧЕНЬ

AR- Авторегресійна модель

QoS- набір методів для управління ресурсами пакетних мереж

TCP- Transmission Control Protocol, Протокол призначений для керування передаванням даних у комп'ютерних мережах, працює на транспортному рівні моделі OSI

Модель OSI -базова еталонна модель взаємодії відкритих систем,

ON/OFF - характерна модель Web-браузерів, АТМ- мережева високопродуктивна технологія комутації та мультиплексування; (англ. Automated Teller Machine) — банкомат

2G- друге покоління мобільних телекомунікацій (GSM, DAMPS, GPRS, TDMA).

3G- третє покоління мобільних телекомунікацій (WCDMA, IMT-2000, UMTS).

4G- четверте покоління мобільного радіозв'язку, наступник 3G та 2G. 4G відрізняється швидкістю передачі даних, яка перевищує показники 3G в 200—500 разів

5G-п'яте покоління мобільних мереж або п'яте покоління бездротових систем. Назва, яку використовують у деяких наукових працях та проектах для позначення наступних телекомунікаційних стандартів для мобільних мереж після стандартів 4G

VoIP- голос через IP) — технологія передачі медіа-даних у реальному часі за допомогою сімейства протоколів TCP/IP. IP-телефонія

IP- *Internet Protocol*; «інтернет протокол» «міжмережевий протокол»— протокол мережевого рівня для передавання датаграм між мережами.,

WWW- *World Wide Web*,— найбільше всесвітнє багатомовне сховище інформації в електронному вигляді: десятки мільйонів пов'язаних між собою документів, що розташовані на комп'ютерах, розміщених на всій земній кулі. Вважається найпопулярнішою і найцікавішою службою мережі Інтернет, яка дозволяє отримувати доступ до інформації незалежно від місця її розташування.

Web- Всесвітнє павутіння

LTE- маркетингова назва 4G LTE — назва мобільного протоколу передавання даних; проєкт 3GPP, стандарт з вдосконалення UMTS для задовільнення майбутніх потреб у швидкості.

HTTP- протокол передачі даних, що використовується в комп'ютерних мережах. Назва скорочена від *HyperText Transfer Protocol*, протокол передачі гіпертекстових документів

FTP- протокол передавання файлів) — стандартний мережевий протокол прикладного рівня, призначений для пересилання файлів між клієнтом та сервером в комп'ютерній мережі.

TELNET- мережевий протокол для реалізації текстового інтерфейсу по мережі (у сучасній формі — за допомогою транспорту TCP). Назву «telnet» мають також деякі утиліти, що реалізують клієнтську частину протоколу.

LAN- Локальна комп'ютерна мережа, є об'єднанням певного числа комп'ютерів на відносно невеликій території.

WAP- Протокол безпроводної передачі даних,— технологія, що використовується для запуску Інтернет-додатків на мобільних терміналах.

UDP— один із протоколів в стеку TCP/IP. Від протоколу TCP він відрізняється тим, що працює без встановлення з'єднання.

TES — це моделі нелінійної регресії з арифметикою за модулем 1.

MAC-адреса — це унікальний ідентифікатор, що зіставляється з різними типами устаткування для комп'ютерних мереж.

VPN- віртуальна приватна мережа

DNS- каталог системи доменних імен

P2P Peer-to-peer- однорангові мережі

IEEE 802.3- стандарти IEEE групи IEEE 802, які відносяться до функціонування комп'ютерних мереж. Родину цих протоколів також називають Ethernet.

ARIMA- У статистиці та економетриці, зокрема в аналізі часових рядів, модель авторегресійної інтегрованої ковзної середньої (ARIMA) є узагальненням моделі авторегресійної ковзної середньої (ARMA).

Ethernet- — найпопулярніший протокол кабельних комп'ютерних мереж, що працює на фізичному та каналному рівні мережевої моделі OSI. Станом на 2016 рік близько 85 % усіх комп'ютерів у світі були підключені до комп'ютерних мереж по протоколу Ethernet.

## ЗМІСТ

ВСТУП .....	10
1 СТАН ПИТАННЯ. ПОСТАНОВКА ЗАДАЧІ.....	12
1.1 Аналітичний огляд моделей мережного трафіка.....	12
1.1.1 Моделі, що базуються на процесах відновлення .....	13
1.1.2 Моделі, що ґрунтуються на марківських процесах .....	16
1.1.3 Авторегресійні моделі .....	19
1.1.4 Самоподібні моделі .....	24
1.1.5 Моделі трафіку додатків .....	29
1.2 Дослідження методів прогнозування навантаження в мережах з комутацією пакетів .....	32
1.2.1 Алгоритми прогнозування тимчасових рядів .....	32
1.2.3 Прогноз з урахуванням $AR(p)$ – моделі .....	45
1.2.4 Прогноз на основі $ARIMA(p,d,q)$ – моделі .....	49
1.2.5 Прогноз методом SSA («Гусениця») .....	52
1.2.6 Прогноз на основі ARFIMA моделі .....	56
1.3 Постановка задачі .....	57
1.4 Висновки .....	58
2 СПЕЦІАЛЬНА ЧАСТИНА .....	60



2.Імітаційне моделювання мережі з короткостроковим прогнозуванням.	60
2.1 Розробка імітаційної моделі мережі.....	67
2.2 Модель мережі з комутацією пакетів .....	68
2.3 Характеристики моделі .....	76
2.4 Оцінка ефективності моделі.....	79
2.5 Висновки .....	86
3 ЕКОНОМІЧНА ЧАСТИНА .....	87
3.1 Розрахунок капітальних витрат на розробку та дослідження фрактальних моделей мережевого трафіку для управління передачею на основі короткострокового прогнозування.....	87
3.2 Висновки .....	93
ВИСНОВОК.....	94
ПЕРЕЛІК ПОСИЛАНЬ.....	95
Додаток А.....	96
Додаток Б .....	97
Додаток В.....	98
Програми.....	100
Лістинг функцій пакету MATLAB.....	103
Лістинг скриптів пакета R .....	104

## ВСТУП

Розробка міцних і надійних мереж і мережевих послуг ґрунтується на розумінні характеристик трафіку в мережі. Протягом історії різні моделі мережевого трафіку були розроблені та використані для оцінки існуючих і пропонованих мереж і послуг.

Вимоги до комп'ютерних мереж не зовсім передбачувані. Моделювання продуктивності необхідне для визначення рівня якості обслуговування (QoS). Моделі продуктивності, у свою чергу, вимагають точних моделей трафіку, які мають здатність фіксувати статистичні характеристики фактичного трафіку в мережі. Багато моделей руху було розроблено на основі даних вимірювання трафіку. Якщо моделі трафіку, що лежать в основі, не фіксують характеристики фактичного трафіку, результатом може бути недооцінка або переоцінка продуктивності мережі. Це погіршує дизайн мережі. Таким чином, моделі трафіку є основним компонентом будь-якої оцінки продуктивності мереж, і вони повинні бути дуже точними.

«Теорія телетрафіку — це застосування математики для вимірювання, моделювання та контролю трафіку в телекомунікаційних мережах. Метою моделювання трафіку є пошук стохастичних процесів для представлення поведінки трафіку. Працюючи в Копенгагенській телефонній компанії в 1910-х роках, А. К. Ерланг характеризував телефонний трафік на рівні виклику за допомогою певних розподілів ймовірностей надходження нових викликів і часу їх утримання. Ерланг застосував моделі трафіку, щоб оцінити потужність телефонного комутатора, необхідну для досягнення заданої ймовірності блокування виклику. Формули блокування Ерланга мали величезний практичний інтерес для громадських операторів, оскільки телефонні засоби (комутація та передача) вимагали значних інвестицій. Протягом кількох десятиліть робота Ерланга стимулювала використання теорії масового обслуговування та

застосування ймовірності в цілому для проектування комутованої телефонної мережі загального користування. Теорія телетрафіку для пакетних мереж досягла значного прогресу в останні десятиліття. Було досягнуто значних успіхів у підходах із довгостроковою залежністю, вейвлетами та мультифрактальними підходами. У той же час моделювання трафіку продовжує стикатися з розвитком мережевих технологій і нових мультимедійних програм. Наприклад, бездротові технології забезпечують більшу мобільність користувачів. Мобільність повинна бути додатковою мірою для моделювання трафіку в бездротових мережах.

Метою цієї роботи є розробка моделі мережі з пакетним трафіком для управління передачею на основі короткострокового прогнозування. Для реалізації поставленої мети необхідно вирішити такі завдання:

- 1 Виконати налітичний огляд моделей мережевого трафіку
- 2 Дослідити методи прогнозування навантаження в мережах з комутацією пакетів
- 3 Розробити імітаційну модель мережі з короткостроковим прогнозуванням
- 4 Дослідити характеристики мережі та оцінити ефективність отриманих результатів

## 1 СТАН ПИТАННЯ. ПОСТАНОВКА ЗАДАЧІ

### 1.1 Аналітичний огляд моделей мережного трафіка

Як відомо, постановку експерименту та проведення дослідження можна виконати двома протилежними підходами. У першому випадку – задіяти реальне обладнання, досліджувати процес у часі. У другому випадку для проведення дослідження можна виконати чисельне моделювання експерименту, задіявши обчислювальні потужності комп'ютера та математичні моделі процесу, що цікавить. Найчастіше сукупність цих двох підходів, їх порівняльний аналіз дозволяють отримати достовірні результати дослідження. І якщо в джерелі експериментальних даних, отриманих за допомогою першого підходу, сумніватися не доводиться, то в другому підході дуже важливу роль при постановці експерименту відіграє адекватність математичної моделі процесу, що використовується.

У разі комп'ютерного моделювання мережі передачі даних (системи дискретних подій DES ) важливо використовувати відповідну модель мережного трафіку, що використовується у чисельному експерименті. Точність моделі, наскільки вона відповідає параметрам реальної мережі передачі даних, задає якість результатів експерименту.

У найпростішому випадку трафік може бути представлений як процес надходження дискретних сутностей (пакетів, повідомлень, одиничних сигналів і т.д.) і математично описаний як точковий процес, що містить послідовність сутностей, що надходять  $X_1, X_2, \dots, X_n$ , де  $X_0=0$ . В даному випадку точковим процесом може бути процес підрахунку сутностей, що надходять, і часу між їх надходженням.

Змішаний трафік міститиме більше одного елемента в сутності, що надходить  $X_n$ . Для опису змішаного трафіку використовується невід'ємна випадкова послідовність  $V_1, V_2, \dots, V_n$ , де  $n=1, 2, \dots, \infty$ , а  $V_n$  - число елементів у сутності, що надходить .

Моделі трафіку можуть бути класифіковані за характером процесу надходження сутностей та програмного забезпечення або додатку, що здійснює передачу даних. За характером процесу моделі можуть бути стаціонарними та нестаціонарними. Стаціонарні моделі, у свою чергу, можуть мати короткострокову і довгострокову залежність. До моделей з короткостроковою залежністю відносяться класичні регресійні та моделі, засновані на марківських процесах. Довгостроковою залежністю відрізняються фрактальні моделі. Також за додатком-джерелом трафік може бути класифікований як трафік web, peer-to-peer, потокового відео тощо.

### 1.1.1 Моделі, що базуються на процесах відновлення

Одним із перших розроблених моделей трафіку були моделі, засновані на теорії відновлення. Внаслідок своєї простоти вони знайшли широке застосування у дослідженнях перших мереж передачі даних. Інтервали часу між подіями у процесі відновлення є позитивними, незалежними та рівномірно розподіленими.

величинами. Процес відновлення можна визначити за допомогою процесу підрахунку  $\{N(t); t \geq 0\}$ , де  $N(t)$  - це число подій системи на інтервалі  $(0; t)$ . На кожному періоді настання подій  $S_n = X_1 + \dots + X_n$  з певною ймовірністю процес починається заново, тобто якщо  $n$ -е подія настає при  $S_n = \tau$ , тоді, починаючи з  $S_n = \tau$ ,  $j$ -я підпоследовність періоду наступу подій:  $S_{n+j} - S_n = X_{n+1} + \dots + X_{n+j}$ . Таким чином, при  $S_n = \tau$ ,  $\{N(r+t) - N(r); t \geq 0\}$  - вважає функція процесу відновлення з незалежними, рівномірно розподіленими інтервалами між подіями.

Процес відновлення нескладно використовувати, проте він має суттєвий недолік - функція автокореляції ряду  $\{X_n\}$  звертається у нуль для всіх ненульових лагів, що не відповідає результатам дослідження реального трафіку. Тобто аналіз АКФ у такому разі говорить про відсутність тимчасової залежності часового ряду. Більш того, позитивне значення автокореляції ряду  $\{X_n\}$  може пояснити наявність коротких спалахів мережної активності. Саме трафік змінного характеру, з періодами підвищеної активності, превалює в

комп'ютерних мережах, особливо ширококомовних, тому модель, що враховує автокорельованість даних, буде ближче відповідати реальній мережі.

### **Модель з урахуванням розподілу Пуассона.**

Модель трафіку на Основа розподілу Пуассона є однією з перших і найчастіше використовуваних. Вона застосовувалася в основному для дослідження телефонних мереж. Пуасонівський процес – це окремий випадок процесу відновлення, в якому час надходження подій експоненційно розподілено з параметром  $\lambda$ :  $P\{X_n \leq t\} = 1 - \exp(-\lambda t)$ . Розподіл Пуассона застосовно, коли трафік надходить від сукупності незалежних джерел, що відповідають вимогам розподілу. Середнє значення та дисперсія розподілу Пуассона визначаються параметром  $\lambda$ . Графічно розподіл Пуассона може бути представлений у вигляді обмеженого біномного розподілу. Розподіл має поряд математичних властивостей. По-перше, суперпозиція незалежних пуасонівських процесів дає новий пуасонівський процес з розподілом, що дорівнює сумі розподілів вихідних процесів. По-друге, властивість незалежних прирощень усуває тимчасові залежності ряду. По-третє, згідно з теоремою Пальма, пуасонівський процес часто використовується для моделювання сукупності незалежних джерел трафіку. Однак пізніше з'ясувалося, що агрегування трафіку який завжди призводить до розподілу Пуассона. Функція розподілу ймовірності для пуасонівського процесу

$$F(t) = 1 - e^{-\lambda t} \quad (1.1)$$

Функція щільності розподілу:

$$f(t) = \lambda e^{-\lambda t} \quad (1.2)$$

Найпростіший спосіб визначення того, що процес пуасонівський - графічний. Для цього достатньо визначити, що гістограма часу настання подій зменшується за експоненційним законом.

Варто зазначити, що у випадку, якщо модель трафіку на основі розподілу Пуассона залежить від часу, тобто  $\lambda$  не стала величина, то параметр розподілу виражається як функція від часу  $\lambda(t)$ .

Для моделювання глобальних мереж, в яких внесок у загальну картину трафіку одного абонента невеликий, сесії користувача можуть бути представлені, як пуассонівський процес. Пуассонівський процес підходить для моделювання TCP трафіку на сесійному рівні моделі OSI, коли сесії ініціюються користувачами, тобто TELNET та FTP програми.

### Модель з урахуванням розподілу Бернуллі.

Модель трафіку на основі розподілу Бернуллі – це дискретний аналог моделі Пуассона. Імовірність настання події в будь-який проміжок часу не залежить від інших подій. Для тимчасового проміжку відповідне число настання подій відповідає біномному розподілу:

$$P\{N_k = n\} = \binom{k}{n} p^n (1-p)^{k-n}, \quad (1.3)$$

де  $n$  набуває значення від 0 до  $k$ .

Час між настанням подій визначається параметром  $p$  з геометричним розподілом

$$P\{A_n = j\} = p(1-p)^{j-1}, \quad (1.4)$$

де  $j$  – позитивне ціле число.

Модель з урахуванням фазового процесу відновлення. Однією із моделей, заснованих на процесі відновлення, є модель трафіку так званого фазового типу. Фазовий процес наступу подій може бути змодельований як безперервний у часі марківський процес поглинання  $C = \{C(t)\}_{t=0}^{\infty}$  у просторі допустимих станів  $\{0, 1, \dots, m\}$ , де 0 відповідає стану поглинання, а решта стану - перехідні, і при цьому процес поглинання виконується протягом кінцевого проміжку часу. Для визначення  $X_n$  процес  $C$  запускається з початковим розподілом  $\pi$ . Коли

відбувається поглинання (тобто процес входить у стан 0), процес зупиняється. Минулий час буде відповідати  $X_n$ , що приведе до ймовірнісної комбінації сум експонент, потім процес виконується заново з початковим розподілом  $p_i$  і процедура повторюється незалежно для отримання  $X_{n+1}$ .

Використання моделі на основі фазового процесу відновлення дозволяє керувати характеристиками моделі, що моделюється, а також задовольняє умові апроксимації розподілів вхідних повідомлень.

### 1.1.2 Моделі, що базуються на марківських процесах

Моделі трафіку, засновані на марківських процесах, вводять залежність між елементами випадкову послідовність на відміну від моделей на основі процесів відновлення. Вважається, що ймовірність переходу системи в стан  $S_{n+1}$  залежить тільки від  $S_n$  і не залежить від інших станів  $S_i$ , де  $i < n$ . Це призводить до позитивної автокореляції  $\{S_n\}$ , що відповідає змінному характеру мережного трафіку, коли періоди підвищеної активності йдуть за періодами зниженої інтенсивності передачі даних. У таких моделях число станів звичайно. Чим воно вище, тим сильніша модель відповідає реальній мережі передачі даних, проте це викликає підвищення складності моделювання.

*Напівмарківські моделі* виходять, коли час між настанням станів підпорядковується випадковому розподілу ймовірностей. Якщо час між змінами станів моделі не враховується зовсім, то процес вважається дискретним марківським ланцюгом.

*ON-OFF та IPP моделі.* ON-OFF модель широко використовується для моделювання мереж передачі голосових даних. Модель використовується, коли необхідно врахувати скейлінг характер мережного трафіку. При цьому допускається лише два стани ON та OFF, а час переходу між станами розподілено за експоненційним законом. Для мережі, в якій  $N$  статистично ідентичних і незалежних ON-OFF джерел, кожне джерело характеризується  $L$  - середнім числом пакетів, переданих за ON період, піковим значенням  $S$  і середнім.  $r$



Рівноважна ймовірність джерела у разі може бути розрахована як  $\gamma=r/S$ .

В рамках IPP-моделі трафіку (Interrupted Poisson Process) мережа може бути лише у двох станах. У ON-стані мережа здійснює передачу даних відповідно до розподілу Пуассона, в OFF- стані передача даних не здійснюється.

**Модель з урахуванням марковського процесу відновлення.** В рамках моделі на основі процесу відновлення для мережі є два стани:  $S_1$  та  $S_2$ . Амплітуда трафіку у стані  $S_1$  дорівнює 0 та 1 - у стані  $S_2$ . Якщо середні інтервали часу переходу між станами прийняти рівними  $d_1$  і  $d_2$  відповідно, то ймовірність знаходження системи  $S_1$  дорівнює  $P_{S1}=d_1/(d_1+d_2)$ , а  $S_2$  - відповідно  $P_{S2}=d_2/(d_1+d_2)$ . При цьому суперпозиція незалежних процесів відновлення має біномний розподіл.

**Модель на основі марківського модульованого пуассонівського процесу (ММПП).** Через простоту реалізації, пуассонівський процес - сам по собі досить привабливий спосіб моделювання мережевого трафіку. Однак, очевидним є недолік такого підходу через використання постійної швидкості потоку  $\lambda$ . Якщо взяти реальний трафік мовних даних, швидкість потоку нічого очікувати однаковою, оскільки повідомлення починаються і закінчуються у випадкові моменти часу.

Прийmemo  $N$  голосових повідомлень за мультиплексований потік, а кожне окреме повідомлення – незалежний пуассонівський процес. Таким чином, базовий процес - пуассонівський зі швидкістю  $\lambda(t)$ . Швидкість потоку модулюється як  $\lambda(t)=n(t)\lambda$ , де  $n(t)$  - це кількість активних на даний момент передач голосових повідомлень. І тут,  $n(t)$  - це стан безперервної у часі ланцюга Маркова. ММПП зберігає деякі властивості відсутності тимчасової залежності пуассонівського процесу та може бути проаналізована в рамках марківської теорії.

Марківський модульований пуассонівський процес широко використовується при моделюванні трафіку завдяки високій гнучкості в якісному налаштуванні отриманої моделі, його також називають подвійним стохастичним процесом.

Простий приклад марковського модульованого пуассонівського процесу - це модель із двома станами: активним, з відповідним позитивним параметром розподілу Пуассона та вимкненим станом, при якому параметр Пуассона дорівнює нулю. Стан ON у разі відповідає передачі звуку, а стан OFF відповідає тиші. Така модель може бути поліпшена агрегування безлічі незалежних джерел, кожен з яких характеризується ММПП з індивідуальним модулюючим марківським процесом.

***Марківська модельована рідинна модель.***

Рідинні моделі визначають трафік як безперервний потік з параметром, визначальним швидкість цього потоку. Подібні моделі особливо підходять у випадках, коли вплив окремого пакета на мережу передачі незначний. Рідинні моделі від традиційних точкових тим, що ігнорують дискретну природу пакетів даних. Назва моделі виходить з аналогії із впливом однієї молекули рідини у трубі з водою.

Обробка рідинних моделей досить проста і вимагає високих обчислювальних потужностей. Найчастіше використовується марківська рідинна модель, при цьому поточний стан марківського ланцюга визначає швидкість потоку (трафіку). При моделюванні VBR відео використовується марківська модельована модель з постійною швидкістю, в якій стану  $S_k$  відповідає постійна швидкість  $\lambda_k$ . Вхідним змінним параметром мережі з урахуванням рідинної моделі є швидкість потоку, і натомість якої можна проводити дослідження поведінки окремих вузлів мережі. Найчастіше рідинні моделі враховують "ON-OFF" природу джерел повідомлень, при цьому OFF-період відповідає відсутності трафіку, а протягом ON-періоду повідомлення надходять детерміновано з постійною швидкістю. Періоди залежать друг від друга, які розподіл підпорядковується експоненційному закону. Простота моделювання та аналітичного трактування робить модель досить популярною.

### 1.1.3 Авторегресійні моделі

У рамках авторегрессионних моделей наступна випадкова величина в послідовності розраховується як явна функція від послідовності попередніх значень. Іншими словами, значення випадкової величини  $X_n$  засноване на наборі попередніх значень  $\{X_k\}$  де  $k < n$ . Авторегресійна модель порядку  $p$  позначається як  $AR(p)$ , а випадкова величина може бути виражена наступним чином:

$$X_k = r_1 X_{k-1} + r_2 X_{k-2} + \dots + r_p X_{k-p} + W_k, \quad (1.5)$$

де  $W_k$  – випадкова величина (білий шум),  $r_i$  – речові числа,  $X_t$  – кореловані випадкові величини.

Автокореляційна функція процесу  $AR(p)$  є загасаючою синусоїдою. Дискретна авторегресійна модель порядку  $p$  генерує стаціонарну послідовність випадкових величин з розподілом ймовірності та АКФ як у авторегресійної моделі порядку  $p$ .

Одним із способів налаштування подібної моделі є вибір розміру історичної вибірки визначення нових випадкових значень змінної.

**Авторегресійна лінійна модель.** Авторегрессионные моделі широко застосовуються при моделюванні VBR відео трафіку розробки систем управління навантаженнями у високошвидкісних мережах передачі мультимедіа даних. Подібна популярність виходить із характеру відео даних, де на 1 секунду припадає до 30 кадрів і, як наслідок, відмінності у цих кадрах найчастіше незначні. Істотні зміни між двома наступними кадрами відео ряду вносять зміни сцени, які спричиняють зростання обсягу переданих даних. Таким чином, відеоряд в рамках однієї сцени, без різких стрибків в обсязі трафіку, може бути змодельований за допомогою авторегресійних моделей, а для різких переходів між кадрами можна задіяти марківські ланцюги.

У роботі відеотрафік моделюється в рамках виразу:

$$X_n = Y_n + Z_n + V_n C_n \quad (1.6)$$

де  $Y_n$  та  $Z_n$  - два незалежні  $AR(1)$  процеси. Завдяки використанню одночасно двох процесів авторегресії вдається привести автокореляційну

функцію (АКФ) моделі, що вийшла до вигляду, відповідного реального відео трафіку. Твір  $V_n C_n$  - це стан марківського ланцюга та незалежної нормально розподіленої випадкової величини, що вводиться для обліку стрибкоподібного зростання навантаження при зміні сцен. Подібна модель застосовна для алгоритмів стиснення відео, у яких передаються лише зміни під час переходу між кадрами.

*Дискретна авторегресійна модель.*

Дискретна авторегресійна модель порядку  $p$  генерує стаціонарну послідовність дискретних випадкових величин з випадковим розподілом ймовірності і АКФ як авторегресійного процесу порядку  $p$  (AR( $p$ )).

Дискретний авторегресійний процес першого порядку - це окремий випадок DAR( $p$ ) процесу, який визначається на основі двох послідовностей незалежних випадкових величин  $\{V_n\}$  та  $\{Y_n\}$ . Випадкова величина  $V_n$  набуває одного з двох значень, 0 або 1 з ймовірністю  $(1-p)$  і  $p$  відповідно. Випадкова величина  $Y_n$  має дискретний набір станів, а матриці переходів  $P\{Y_n=i\}=\pi(i)$ . Випадкова величина  $X_n$  формується виходячи з виразу DAR-процесу першого порядку:

$$X_n = V_n X_{n-1} + (1-V_n) Y_n. \quad (1.7)$$

DAR процес першого порядку – це ланцюг Маркова з дискретним набором станів  $S$  та матрицею переходів

$$P = pI + (1-p) Q, \quad (1.8)$$

де  $I$  - це одинична матриця, а  $Q$  - матриця  $Q_j = \pi(j)$  для  $i, j \in S$ .

DAR-процес першого порядку має АКФ авторегресійного процесу першого порядку, а функція розподілу ймовірності - це функція від  $\pi$ .

DAR(1) процес має меншим, порівняно з ланцюгом Маркова, числом параметрів, а оцінка цих параметрів не викликає труднощів. Модель трафіку на основі DAR процесу легко піддається аналітичному аналізу, однак через експоненційне згасання функції автокореляції модель не застосовується до трафіку з АКФ, що повільно загасає.

*Авторегресійна модель ковзного середнього.* Авторегресійна модель ковзного середнього порядку (p,q) позначається як ARMA(p,q) і набуває наступного вигляду:

$$X_t = c + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_p \varepsilon_{t-p}. \quad (1.9)$$

Це еквівалентно наступному запису:

$$\varphi(B)X_t = c + \theta(B)\varepsilon_t, \quad (1.10)$$

де B - це оператор лага (оператор зсуву), такий що  $X_{t-1} = BX_t$ , а  $\varphi(B) = (1 - \varphi_1 B^1 - \dots - \varphi_p B^p)$ , c - константа.

Це еквівалентно фільтрації білого шуму  $\varepsilon_t$  лінійним фільтром, інваріантним до зсуву за часом, який має дробово-раціональну передатну функцію з p полюсами та q нулями, тобто:

$$H(z) = \frac{B_q(z)}{A_p(z)} = \frac{1 - \sum_{k=0}^q \theta_k z^{-k}}{1 - \sum_{k=1}^p \varphi_k z^{-k}}. \quad (1.11)$$

Автоковаріація ARMA(p,q) процесу може бути отримана

твором (11) та  $X_{t-k}$  з урахуванням математичного очікування та взаємної кореляції між  $\varepsilon_t$  та  $X_t$ . ARMA моделі широко використовуються для моделювання VBR трафіку. У цьому випадку тривалість відеокадру поділяється на рівні m інтервали. Число осередків  $n_i$ ,  $i=0, \dots, m-1$ , у часовому інтервалі моделюється за допомогою ARMA процесу:

$$X_n = \varphi X_{n-m} + \sum_{i=0}^{m-1} \theta_i \varepsilon_{n-i}. \quad (1.12)$$

Так як відеодані кожного кадру корелірованні між собою зі змінним коефіцієнтом кореляції, то функція автокореляції міститиме піки на лагах,

кратних  $m$ . У моделі AR-частина використовується для ефекту повторної кореляції, а  $\theta_k$  вводиться для підбору кореляції для інших затримок. Параметрична оцінка ARMA моделей складніша, ніж для AR моделей, оцінка  $\theta_k$  вимагає розв'язання безлічі нелінійних рівнянь. Аналітичний аналіз також досить скрутний.

***Інтегральна модель авторегресії - ковзного середнього.***

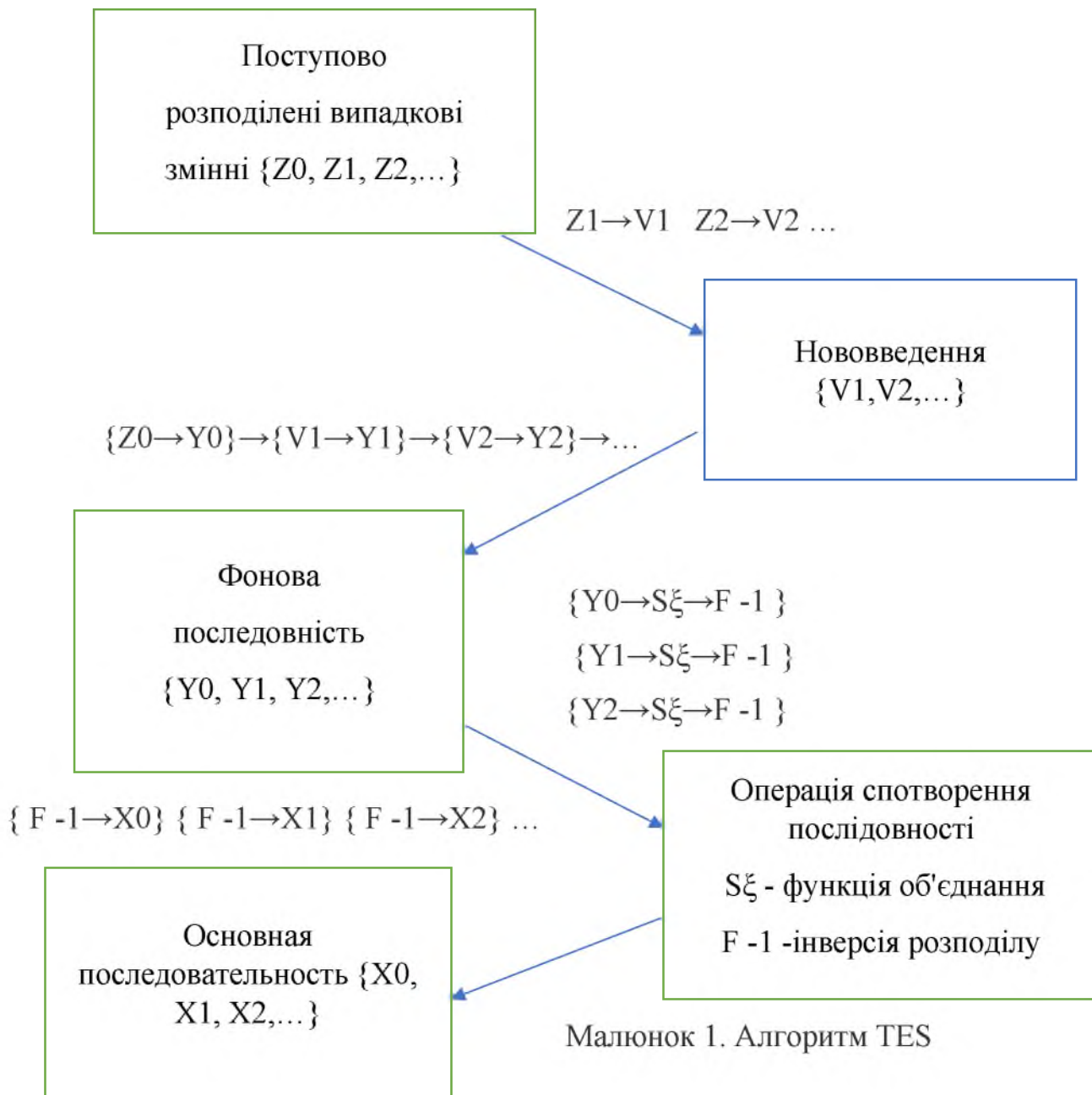
Авторегресійний інтегральний процес ковзного середнього порядку  $(p, d, q)$ , позначається як ARIMA  $(p, d, q)$  і будується на основі ARMA  $(p, q)$ . Тобто ARIMA $(p, d, q)$  може бути інтерпретована як ARMA $(p+d, q)$  - модель з  $d$  одиничним корінням, решта коренів многочлена лежить за межами одиничного кола. ARIMA $(p, d, q)$  процес може бути записаний таким чином:

$$\phi(B)\nabla^d X_t = \theta(B)\varepsilon_t, \quad (1.13)$$

де  $\nabla$  - оператор диференціювання, такий що  $(X_t - X_{t-1}) = \nabla X_t$  і  $\nabla X_t = (1-B)X_t$ , а  $\phi(B)$  - багаточлен від  $B$ . ARIMA $(p, d, q)$  процес використовується для моделювання нестационарних рядів, які виявляють однорідність на відміну їхнього локального рівня чи тренда. Як правило, використовують ARIMA з  $d$  рівним 1 або 0: якщо  $d=0$ , то процес має лінійний тренд, а якщо  $d>1$ , то поліноміальний.

***Модель розширення та перетворення вибірки.*** Модель розширення та перетворення вибірки (TES) – це нелінійна регресійна модель, яка націлена на відтворення наданої стаціонарної вибірки з дотриманням маргінального розподілу та структури АКФ. Іншими словами, TES генерує послідовність, керовану випадковим процесом, згладжену та адаптовану під данемаргінальний розподіл. Якість моделювання трафіку телекомунікаційних програм сильно залежить від того, наскільки змодельований часовий ряд відповідає основним характеристикам вхідної послідовності. Найбільш важливих характеристик три: маргінальний розподіл, структура АКФ, відповідність між графічним поданням

вихідних даних та змодельованого ряду. Загальний алгоритм моделювання часового ряду представлений малюнку 1.



У першому етапі відбувається ініціалізація низки  $\{Z_i\} \in [0;1]$ . Після чого  $Z_i$  використовується для вибору  $V_i \in [-0.5; 0.5]$  відповідно до заданої щільності. Величина  $V_i$  додається до фонового значення послідовності  $Y_{i-1}$  та формує таким чином  $Y_i = \{Y_{i-1} + V_i\} \bmod 1$ . У рамках алгоритму перетворення фонові послідовності відбувається двічі. По-перше, розрахунки за модулем 1  $Y_i$  призводять до деякого ефекту розриву, якщо припустити, що значення  $Y_i$

можуть бути досить близькими до 1. Також невеликі значення  $V+1$  можуть призвести до близькості  $Y_{i+1}$  нулю, тобто серйозних змін у фоновій послідовності. Подібні ефекти усуваються функцією об'єднання (віконною функцією)  $S$ , яка значення близькі до 0 і 1 встановлює близько 0:

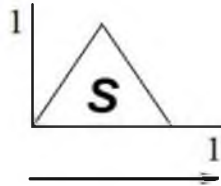


Рисунок 2. Функція об'єднання  $S$

Друге перетворення полягає в інверсії гістограми емпіричних даних, що наводить об'єднану послідовність ближче до діапазону значень вихідного ряду та встановлює необхідний маргінальний розподіл. Це останній етап моделювання емпіричного часового ряду, що генерує основну послідовність. Зазначимо, що однією з основних особливостей TES моделі є можливість подальших розрахунків лага АКФ спотворення та щільності розподілу нововведень.

#### 1.1.4 Самоподібні моделі

Нехай  $\{X_t\}$  - стаціонарний у широкому значенні стохастичний процес, що володіє стаціонарним середнім  $\mu = E[X_t]$ , стаціонарною та кінцевою дисперсією  $v = E[(X_t - \mu)^2]$  та стаціонарною АКФ  $\gamma_k = E[(X_t - \mu)(X_{t+k} - \mu)]$ , яка залежить тільки від  $k$  і не залежить від  $t$ . Зазначимо, що  $v = \gamma_0$ . Якщо прийняти  $\{X_k\}$  при лазі  $k$  до  $\rho k$ , то визначення  $\rho k = \gamma_k / \gamma_0$ .

Розділимо вихідний часовий ряд  $\{X_t\}$  на послідовності, що не перетинаються довжиною  $m$  і середнім їх, щоб отримати послідовність  $\{X_t^{(m)}\}$

$$X_j^{(m)} = m^{-1}(X_{jm-m+1} + \dots + X_{jm}), \quad (1.14)$$

де  $X_j^{(m)}$  - це вибіркова середня послідовність  $X_{jm-m+1} + \dots + X_{jm}$ . Позначимо  $v_m$  як дисперсію  $\{X_j^{(m)}\}$ :



$$\begin{aligned}
v_m &= E\left[\frac{1}{m}(X_{jm-m+1} + \dots + X_{jm})\right]^2 \left[E\frac{1}{m}(X_{jm-m+1} + \dots + X_{jm})\right]^2 = \\
&= \frac{v}{m} + \frac{2}{m^2} \sum_{k=1}^m (m-k)\gamma_k = \\
&= v \left[1 + 2 \sum_{k=1}^m \left(1 - \frac{k}{m}\right) \rho_k\right] m^{-1}.
\end{aligned} \tag{1.15}$$

У разі, якщо процес є білим шумом, то  $X_{jm-m+1} + \dots + X_{jm}$  будуть взаємно некоррельованими,  $\rho_k=0$  для  $k>0$  і  $v_m=v m^{-1}$ .

Для великих  $m$  можна усереднити:

$$v_m = v \left[2 \sum_{k=1}^m \rho_k\right] m^{-1} \tag{1.16}$$

Розглянемо випадок, у якому  $\rho_k \neq 0$  і  $\sum_{k=-\infty}^{\infty} \rho_k < \infty$ .

Тоді дисперсія вибіркового середнього асимптотично згасатиме і прагнеться до нуля пропорційно  $m^{-1}$ , тобто:

$$v_m \sim v c_p m^{-1}, \tag{1.17}$$

де  $c_p$  – константа.

Для більшості розглянутих вище моделей, таких як ARMA або моделей на основі марківських процесів, середня дисперсія вибірки згасає саме відповідно до (1.17).

У роботі для дослідженого трафіку вказується згасання середньої дисперсії вибірки навіть повільніше, ніж  $m^{-1}$ . Найбільш простим підходом у даному випадку було б прийняти згасання  $v_m$  пропорційно  $m^{-\alpha}$  для деякої  $\alpha \in (0,1)$ . Тоді  $\sum_{k=-\infty}^{\infty} \rho_k$  повинна бути пропорційна:  $m^{1-\alpha}$

$$\sum_{k=1}^m \rho_k \approx C m^{1-a}$$

(1.1  
8)

Оскільки  $\alpha < 1$ , то

$$\sum_{k=-\infty}^{\infty} \rho_k \rightarrow \infty$$

Таким чином АКФ згасає повільніше, тому що вона не підсумовується.

Процес  $\{X_t\}$  має короткострокову залежність, якщо

$$\sum_k p_k < \infty$$

Відповідно,  $v_m$  загасає асимптотично пропорційно  $m^{-1}$ , спектральна щільність потужності в нулі звертається в нескінченність, а усереднені послідовності  $\{X_t^{(m)}\}$  прагнуть чистого шуму при  $m \rightarrow \infty$

АКФ процесів, які мають короткострокову залежність, згасає експоненційно.

Процес  $\{X_t\}$  має довготривалу залежність, якщо

$$\sum_k p_k < \infty$$

Дисперсія середнього  $v_m$  згасає повільніше, ніж  $m^{-1}$ . При цьому довгострокова залежність лише визначає поведінка АКФ при великих лагах, але з визначає автокореляцію низки будь-якого фіксованого кінцевого лага.

Процес  $\{X_t\}$  самоподібний, якщо  $\rho_k^{(m)} = \rho_k$  всім  $m$  і  $k$ , тобто структура АКФ зберігається різних часових масштабах. Стохастичний самоподібний процес зберігає однакові статистичні показники на будь-якому діапазоні масштабів та задовольняє наступному виразу:

$$\{X_{at}\}^D = \alpha^H \{X_t\} \quad (1.19)$$

де  $D=$  означає рівність у розподілі, а  $H$  називається параметром Херста.

Якщо процес  $\{X_t\}$  має стаціонарний інкремент:  $Y_t = X_t - X_{t-1}$ , то АКФ:

$$\rho_k \rightarrow H(2H - 1)k^{2H-2}, \text{ якщо } k \rightarrow \infty \quad (1.20)$$

де  $H = 1 - \alpha/2$ , а при  $0 < H < 1$  і  $H \neq 1/2$ ,  $\sum_k \rho_k \rightarrow \infty$

**Фрактальна інтегрована модель авторегресії - ковзного середнього.**

Фрактальна інтегрована модель авторегресії ковзного середнього - ARFIMA(p,d,q) при  $0 < d < 1/2$  - це приклад моделі стаціонарного процесу з довготривалою залежністю. Вона являє собою розширення ARIMA (p, d, q) моделі та визначається наступним чином:

$$\phi(B)\nabla^d X_t = \theta(B)\varepsilon_t, \quad (1.21)$$

де  $0 < d < 1/2$ . Оператор  $\Delta^d = (1 - B)^d$  можна виразити через біномне розкладання

$$X(1 - B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k B^k \quad (1.22)$$

$$\binom{d}{k} = \frac{d!}{k!(d-k)!} = \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)} \quad (1.23)$$

де  $\Gamma(x)$  – гамма-функція.

**Модель фрактального броунівського руху.**

Броунівський рух це стохастичний процес для  $t > 0$ . Характерні такі характеристики:

- Інкремент  $(B_{t+t_0} - B_{t_0})$  має нормальний розподіл з нульовим середнім та дисперсією  $\sigma^2 t^{2H}$ .

- Інкремент на двох інтервалах, що не перетинаються  $[t_1, t_2]$  і  $[t_3, t_4]$ ,  $B_{t_4} - B_{t_3}$  і  $B_{t_2} - B_{t_1}$  - незалежні випадкові величини.

- $B_0 = 0$  та  $B_t$  безперервна при  $t = 0$ .

Фрактальний броунівський рух  $\{fB_t\}$  – це самоподібний гаусівський процес із параметром самоподібності  $0,5 < H < 1$ . Фрактальний броунівський рух

відрізняється від броунівського руху тим, що дисперсія інкременту дорівнює  $\sigma^2 t^{2H}$ . Визначимо дисперсію інкременту

$$\sigma^2 = E\{(fB_t - fB_{t-1})^2\} = E\{(fB_1 - fB_0)^2\} = E\{(fB_1)^2\} \quad (1.24)$$

Тоді

$$E\{(fB_{t_2} - fB_{t_1})^2\} = E\{(fB_{t_2-t_1} - fB_0)^2\} = \sigma^2\{(t_2 - t_1)^{2H}\} \quad (1.25)$$

Також

$$\begin{aligned} E\{(fB_{t_2} - fB_{t_1})^2\} &= E\{(fB_{t_2})^2\} + E\{(fB_{t_1})^2\} - 2E\{(fB_{t_2} - fB_{t_1})^2\} \\ &= \sigma^2(t_2)^{2H} + \sigma^2(t_1)^{2H} - 2\gamma\{(fB_{t_2}, fB_{t_1})\} \gamma\{(fB_{t_2}, fB_{t_1})\} = \\ &= \frac{1}{2\sigma^2(t_2)^{2H}} - (t_2 - t_1)^{2H} + (t_1)^{2H} \end{aligned} \quad (1.26)$$

Тому коваріація інкременту на двох інтервалах, що не перетинаються:

$$\begin{aligned} \gamma(fB_{t_4} - fB_{t_3}, fB_{t_2} - fB_{t_1}) &= \\ \gamma(fB_{t_4}, fB_{t_2}) - \gamma(fB_{t_4}, fB_{t_1}) - \gamma(fB_{t_3}, fB_{t_2}) + \gamma(fB_{t_3}, fB_{t_1}) &= \\ \frac{\sigma^2}{2(t_4-t_1)^{2H}} - (t_3 - t_1)^{2H} + (t_3 - t_2)^{2H} - (t_4 - t_2)^{2H} \end{aligned} \quad (1.27)$$

$$2 fB_{t_2}, fB_{t_1} = - (t_4 - t_1)^{2H} - (t_3 - t_1)^{2H} + (t_3 - t_2)^{2H} - (t_4 - t_2)^{2H} .$$

Фрактальний броунівський рух  $\{fB_t\}$  може бути отриманий з броунівського руху  $\{B_t\}$  інтегруванням:

$$\{fB_t\} = \int_0^t (t-u)^{H-0,5} dB(u) \quad (1.28)$$

таким чином, взаємозалежність між інкрементами фрактального броунівського руху може бути нескінченною.

У разі дискретного процесу автокореляція інкременту може бути розрахована заміною  $t_1, t_2, t_3, t_4$  на  $n, n+1, n+k, n+k+1$  та розподілом на  $\sigma^2$ :

$$\rho_k = \frac{1}{2} [(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}] \quad (1.29)$$

Процес інкрементування у разі - це фрактальний гаусовський шум. Автокореляція в останньому вираженні має довготривалу залежність,  $\rho^k \sim k^{2H-2}$ ,  $k \rightarrow \infty$ .

В одному із прикладів моделювання трафіку за допомогою фрактального гаусівського шуму, вказується на складність аналітичного аналізу

розподілу заповнення буфера. Тому наводиться приблизний аналіз поведінки хвостів розподілів. Показано, що для великих значень  $N$  зростання навантаження на мережу потребує значного збільшення обсягів пам'яті. Імовірність втрати пакету зростає алгебраїчно з розміром буфера, а не експоненційно, як у випадку з марковськими та ARMA моделями.

### 1.1.5 Моделі трафіку додатків

Очевидно, що процес передачі даних у мережі та характер самих даних визначається додатками, що використовуються. Більшість програм можна віднести до таких типів: Web, e-mail, peer-to-peer файлообмінники, потокове мультимедіа. У деякій роботі вказується, що понад 40% всіх даних, що передаються по мережі, відноситься до Web.

Цілком можливо застосувати розглянуті вище підходи для моделювання трафіку додатків, проте логічніше використовувати спеціальні моделі для отримання кращих результатів.

Web-трафік. У безлічі робіт згадується, що Web - трафік займає більшу частину всього трафіку мережі Internet, що не дивно з урахуванням того, що Web-браузери надають доброзичливий інтерфейс для роботи з електронною поштою, передачі файлів, віддаленої обробки даних, потокового медіа та ін.

Одні з перших досліджень Web-трафіку на самоподібність вказують коефіцієнт Херста, близький до 0,8. Крім характеру самого трафіку було зазначено, що Web клієнти можуть бути представлені як ON-OFF джерела повідомлень, що мають розподіл з важким хвостом. При цьому самоподібність трафіку виникає при агрегуванні множини потоків даних з розподілом з важким хвостом. Щільність розподілу ймовірності, відповідна РТХ, повинна мати статичний характер. «Тяжкість» хвоста означає, що екстремально великі значення ряду мають ненульову ймовірність на гістограмі розподілу. OFF - періоди можуть бути або активними, коли завантаження великих обсягів даних не здійснюється, але користувач працює з Web-браузером, або неактивними, коли користувач не використовує Web-браузер. Активні періоди OFF можуть

бути описані розподілом Вейбулла, а неактивні - Парето розподілом з важким хвостом.

Проводилося дослідження 1900 клієнтських Web-браузерів, для яких також була характерна ON/OFF модель. OFF-період описувався розподілом Вейбулла і відповідав період, під час якого користувач переглядав завантажений контент. ON-період відповідав завантаженню сторінки, проте процес розглядався детальніше. Завантаження основного коду сторінки та додаткових об'єктів описувалося окремими гамма-розподілами.

Проводився добовий збір трафіку з роутерів корпоративної мережі. Згідно з дослідженням, 42% всього потоку даних належало до Web-трафіку. Були розглянуті щільності розподілу числа серверів, з якими обмінювався даними кожен клієнт, і навіть кількість клієнтів, обслуговуваних одним сервером, обсяги трафіку, передані кожним клієнтом, обсяги переданих серверами даних. Одним із висновків став той факт, що вкрай складно виділити деяку типову поведінку окремого клієнта чи сервера.

Трафік однорангових мереж. Peer-to-peer (p2p) або трафік однорангових мереж часто порівнюють з Web-трафіком з тієї причини, що однорангові мережі є повною протилежністю клієнт-серверній архітектурі. На даний момент не так багато робіт присвячено дослідженню p2p трафіку.

Дослідження університетської мережі проводилося у роботі. Більшість даних, що передаються, відносилася до відео-форматів AVI і MPEG, а також MP3. У цьому найбільше навантаження давали кілька комп'ютерів у мережі, у яких були розміщені файли понад 700 Мб, доступ яких здійснювався іншими членами мережі. Таким чином, невелика кількість вузлів у мережі відповідає за основну частку навантаження.

В іншій роботі здійснювався збір p2p трафіку через роутери корпоративної мережі. Знову виявили, що найбільше навантаження в мережі створює кілька комп'ютерів.

Трафік відеоданих. Характеристики трафіку стислих відеоданих можуть відрізнятися залежно від характеру самого відеоряду і частоти зміни кадрів, і

навіть алгоритмів кодування записи. Для більшості алгоритмів кодування запису характерна реєстрація лише змін між кадрами, що послідовно йдуть, без дублюючого опису незмінної сцени. Тому динамічніше відео викликає більший обсяг переданих даних, а різка зміна сцен може бути відзначена сплесками трафіку.

Більшість джерел відео характеризуються швидкістю передачі кадрів. А моделі, як правило, враховують такі статистичні характеристики, як щільність розподілу ймовірності обсягу даних, що припадають на 1 кадр, характер АКФ ряду кадрів, середній час між зміною сцен, щільність розподілу ймовірності довжини файлів, що передаються.

Існує безліч відео - кодеків та алгоритмів кодування відеоданих, тому досить складно знайти єдину універсальну модель трафіку відео. Різним файлам будуть відповідати різні моделі, проте деякі принципи можна узагальнити :

- Для опису обсягу даних, що припадають на один кадр, зазвичай використовується логнормальний, гамма або розподіл Парето.

- АКФ нединамічного відео набуває експоненційного вигляду. АКФ для ширококомовного відео має більш складну структуру з швидко спадаючою залежністю на маленьких лагах і довготривалою залежністю на великих значеннях лага.

- Розподіл довжин сцен зазвичай відповідає розподілу Парето або Вейбулла, рідше – гамма-розподілу.

- Об'єм даних, що припадає на одну сцену, некорельований і відповідає розподілу Вейбулла або гамма-розподілу.

- Довжина потокової Web-передачі відео відповідає розподілу Парето з важким хвостом .

Розглянуто різні моделі трафіку, як розроблені досить давно і леглі в основу сучасних алгоритмів керування трафіком, так і актуальні моделі, що враховують нелінійно - динамічні властивості трафіку, розподіл з важким хвостом і залежність, що повільно спадає. Далі розглянемо реалізовані у

сучасних протоколах зв'язку алгоритми управління трафіком для подальшої розробки вдосконаленої методики боротьби з навантаженнями.

## **1.2 Дослідження методів прогнозування навантаження в мережах з комутацією пакетів**

### **1.2.1 Алгоритми прогнозування тимчасових рядів**

Завдання прогнозування нових значень часового ряду за історичною вибіркою не нова, проте дослідження в цій галузі не втрачають актуальності і постійно виникають нові алгоритми та методики прогнозування. Нещодавно дослідники почали застосовувати методи прогнозування часових рядів по відношенню до мережевого трафіку даних для управління трафіком, забезпечення QoS і боротьби з перевантаженнями. На даний момент робіт у цій галузі небагато, вони рідко спираються на дані щодо завантаженості реальної корпоративної мережі, а моделі прогнозування найчастіше авторегресійні, рідше – на основі нейромереж. До того ж алгоритмів прогнозування часових рядів та їх варіацій досить багато, а дослідження кожного з них є досить трудомістким завданням. Таким чином, завдання управління трафіком на основі короткострокових прогнозів стану мережі залишається дуже актуальним.

Формально завдання прогнозування часового ряду  $X(t)=X(1), X(2), \dots, X(T)$  у момент часу  $T$  полягає у визначенні значень ряду  $X(t)$  у моменти часу  $T+\Gamma, T+2, \dots, T+\Gamma'$ , де  $\Gamma'$  - час попередження чи горизонт прогнозу. Іншими словами, визначення майбутнього стану процесу ґрунтується на відомих значеннях ряду, а прогноз ґрунтується на моделі, що відображає функціональну залежність між майбутніми та минулими значеннями ряду. Безліч моделей прогнозування поділяється на статистичні та структурні. До статистичних моделей відносяться регресійні та авторегресійні, а також моделі згладжування. Моделі прогнозування на основі нейромереж, ланцюгів Маркова та класифікаційних дерев відносяться до структурних моделей.



*Алгоритми згладжування.* Різні алгоритми згладжування часового ряду або розрахунку середнього значення можуть використовуватися як проста і наочна модель, що прогнозує [85]. Найпростіший алгоритм ковзного середнього полягає у обчисленні наступного значення часового ряду на підставі середнього за період:

$$X_{t+1} = (X_t + X_{t-1} + X_{t-2} + \dots + X_{t-N}) / N, \quad (1.30)$$

де  $X_{t+1}$  - прогнозоване значення ряду на момент  $t+1$ ,  $X_{t-1}$  - значення ряду на момент  $t-1$ ,  $N$  - період згладжування. У цьому випадку важливо підібрати потрібне значення  $N$ . До того ж логічно припустити, що  $X_t$  значення ряду має сильніше впливати на прогноз, ніж  $X_{t-N}$ . Алгоритм експоненційного згладжування враховує цей недолік через деякий коефіцієнт:

$$X'_{t+1} = \alpha X_t + (1 - \alpha) X'_t, \quad (1.31)$$

де  $0 < \alpha < 1$  - ваговий коефіцієнт,  $X_t$  - реальне значення ряду в момент  $t$ , а  $X'_t$  - прогнозне значення ряду. Таким чином, ми отримуємо зважену ковзну середню, в якій наступне прогнозоване значення формується із прогнозованого та реального поточного значення ряду. При цьому коефіцієнт підбирається відповідно до характеру ряду - чим сильніше змінюються значення ряду, тим більше має бути коефіцієнт, що швидше пристосуватися до великих змін. І що вище шумова складова ряду, то менше має бути значення коефіцієнта, щоб згладжувати флуктуації. Для вирішення задачі підбору коефіцієнта розроблено алгоритм адаптивного згладжування.

Існує безліч варіацій алгоритму адаптивного згладжування, в самому простому випадку коефіцієнт підбирається наступним чином:

$$\alpha_{t+2} = |(X'_{t+1} - X_{t+1}) / X_{t+1}|. \quad (1.32)$$

У цьому якщо значення коефіцієнта виходить рівним чи більше 1, приймається  $\alpha = 0.99(9)$ , і якщо рівним 0, то  $\alpha = 0.0(0)1$ . Розглянуті алгоритми

згладжування часових рядів не враховують наявність у ряду тренду та сезонної складової.

Алгоритм експонентного згладжування з урахуванням тренду при згладжуванні враховує напрямок ряду:

$$\begin{aligned} L_t &= \alpha X_t + (1 - \alpha)(L_{t-1} + T_{t-1}) \\ T_t &= \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}, \end{aligned} \quad (1.33)$$

де  $L$ -рівень тренду, а  $T$ -тренд, при цьому  $0 < \alpha < 1$  і  $0 < \beta < 1$ . Тоді прогнозоване значення ряду на  $m$  кроків:

$$X_{t+\tau} = L_t + (T_t m)$$

$$X_{t+m} = L_t + (T_t m). \quad (1.34)$$

З урахуванням сезонної складової низки прогноз будується так :

$$\begin{aligned} L_t &= \alpha(X_t / SA_{t-c}) + (1 - \alpha)(L_{t-1} + T_{t-1}) \\ T_t &= \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \\ SA_t &= \gamma(X_t / L_t) + (1 - \gamma)(SA_{t-c}), \end{aligned} \quad (1.35)$$

де  $L$  - рівень тренду,  $T$ -тренд,  $SA_t$  - сезонна складова ряду на момент часу  $t$ ,  $C$  - розмір циклу для сезонної складової,  $0 < \alpha < 1$ ,  $0 < \beta < 1$ ,  $0 < \gamma < 1$ . Тоді прогнозоване значення ряду на  $m$  кроків уперед:

$$X_{t+m} = (L_t + (T_t m))SA_{t-c+m} \quad (1.36)$$

Незважаючи на спроби не просто усереднити ряд ковзним вікном, а врахувати його сезонну, трендову та циклічну складову, прогнози на основі

моделей, що згладжують, відфільтровують досить багато інформації про реальний процес. Разом із шумами усувається інформація, яка могла б бути використана під час прогнозування ряду. Розглянуті далі регресійні та авторегресійні моделі відрізняються більшою точністю, враховують більше факторів, що впливають на розподіл ряду.

*Регресійні та авторегресійні алгоритми.* У статистичній обробці даних регресійний аналіз застосовується для визначення залежності між вихідною залежною змінною та зовнішніми незалежними змінними, т.зв. регресорами. Найпростіша регресія

нова модель називається лінійною і передбачає наявність єдиного зовнішнього фактора, що впливає на процес з лінійним зв'язком:

$$X_t = \beta_0 + \beta_1 Y_t + \varepsilon_t, \quad (1.37)$$

де  $Y_t$  - незалежна змінна,  $\beta_0$  і  $\beta_1$  - коефіцієнти регресії,  $\varepsilon_t$  - помилка моделі. У разі множинної регресійної моделі зовнішніх факторів, що впливають на значення ряду – кілька:

$$X_t = \beta_0 + \beta_1 Y_t^1 + \beta_2 Y_t^2 + \dots + \beta_n Y_t^n + \varepsilon_t \quad (1.38)$$

Залежність між вихідним процесом та регресором не обов'язково має бути лінійною, можливий випадок, при якому залежність описується довільною функцією.

Найчастіше при дослідженні процесу складно виділити зовнішні фактори, що впливають, при цьому регресійний аналіз ґрунтується тільки на самих значеннях ряду, такі моделі називаються авторегресійними. Одна з найбільш відомих авторегресійних моделей - ARMA (Autoregressive Moving Average) - являє собою комбінацію авторегресійної моделі та ковзного середнього. Авторегресійна (AR(p)) модель може бути виражена наступним чином:

$$X_t = c + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} + \varepsilon_t, \quad (1.39)$$

де  $X_t$  - поточне значення ряду,  $\varepsilon_t$  - випадкова помилка,  $\varphi_i$  ( $i = 1, 2, \dots, p$ ) - вагові коефіцієнти, а  $c$  - константа. У записі AR(p),  $p$  – це порядок моделі. Трохи видозмінений запис представленої раніше моделі ковзного середнього MA(q):

$$X_t = \mu + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t, \quad (1.40)$$

де  $\mu$  - середнє значення вибірки,  $\theta_j$  ( $j=1, 2, \dots, q$ ) - параметри моделі та  $q$ - порядок моделі. Тоді ARMA(p,q) модель:

$$X_t = c + \varepsilon_t + \sum_{i=1}^p \varphi_i X_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} \quad (1.41)$$

ARMA модель працює з припущенням, що ряд стаціонарний. Операція усунення тренда у разі може призвести до стаціонарності низки чи навіть усунути сам тренд. Якщо як вихідні дані використовується не сам тимчасовий ряд, а різниця його компонентів, то в такому випадку модель називається ARIMA(p,d,q) -

У регресії введено рухоме значення, де  $d$  - порядок різниці між значеннями ряду. На основі ARMA моделі розроблено багато інших прогнозуючих моделей. FARIMA - одна з них, при якій операція детрендингу проводиться з дрібним порядком. SARIMA – модель, в якій враховано сезонну складову ряду.

### Алгоритми на основі нейромережевих моделей.

Нейронні мережі мають велику популярність у задачі прогнозування часових рядів через можливість імітації процесів з множинним зовнішнім впливом. Ефективні у завданнях класифікації об'єктів та розпізнавання образів. Принципово структура нейрона ANN схожа будовою людський нейрон і представлена малюнку 2.

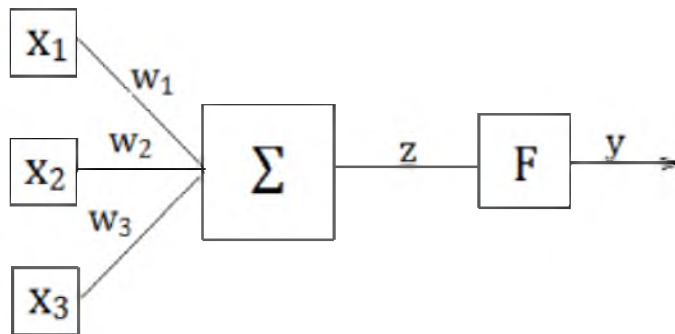


Рисунок 2. Принципова схема нейрона

На вхід штучного нейрона надходить вектор вхідних сигналів  $X_1, X_2, X_3, \dots, X_n$ , для яких вираховується середнє значення  $z$ , яке використовується в активаційній функції  $F$  для розрахунку вихідного значення  $Y$ . Активаційна функція може бути функцією одиничного стрибка, лінійною функцією та т.д, наприклад:

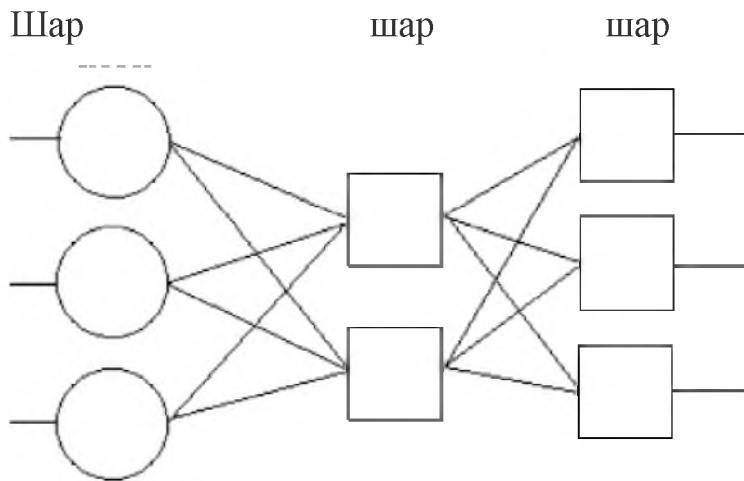
$$F(z) = \frac{1}{1 + e^{-z}} \quad (1.42)$$

Приклад ANN, що складається з 3 вихідних та 2 прихованих нейронів, представлений на малюнку 3. Вхідний шар служить для розподілу вхідного вектора значень і не містить нейронів. Зазначимо, що нейрони в повному обсязі нейрони обов'язково мають бути пов'язані між собою, тобто деяких зв'язків  $w$  може бути нульовим. Також нейронна мережа може мати зворотний зв'язок для переналаштування ваг. Вагові коефіцієнти зв'язків призначаються у процесі навчання, і навіть можуть перераховуватися залежно від результатів роботи мережі.

Вхідний

Прихований

Вихідний



Малюнок 3. Структурна схема нейронної мережі

Відповідно можна виділити основні особливості ANN (Artificial Neural Network):

- ANN можуть моделювати нелінійні залежності між вхідними та вихідними даними.
- ANN навчається на вхідних даних, тобто дані визначають модель між вхідними та вихідними даними.
- ANN можуть узагальнювати вхідні дані, що не призводить до зниження ефективності роботи мережі за зміни характеру даних.
- На відміну від статистичних моделей, ANN вносити припущення щодо розподілу вхідних даних.

У ході використання ANN виникає низка труднощів, які згідно виглядають наступним чином:

- Складність вибору оптимальної комбінації параметрів мережі, таких як швидкість навчання, кількість прихованих шарів, число прихованих нейронів на кожному шарі.
- Складність вибору алгоритму навчання та суворі вимоги до навчальної вибірки.
- Навчена нейронна мережа служить деякою «чорною скринькою», тобто досить складно виділити сукупність правил, за якими ANN приймає рішення.

Генетичні та еволюційні алгоритми. Дуже нетривіальною є завдання вибору відповідної моделі прогнозування, а також добору параметрів для

лінійної статистичної моделі або налаштування нелінійної моделі. Застосування генетичних алгоритмів може допомогти у вирішенні такого завдання.

Перші роботи у цьому напрямі описували використання генетичних алгоритмів у доборі коефіцієнтів для регресійних і авторегресійних моделей. У цьому популяція формувалася ковзним вікном по ряду даних, а коефіцієнти регресії служили вектором генів. Після подальших схрещувань та мутацій формувалася нова популяція, пристосованість якої визначалася точністю прогнозування низки.

У більш сучасних роботах генетичні алгоритми використовуються для налаштування штучних нейронних мереж, тобто вибору архітектури мережі: числа вхідних нейронів, числа вихідних нейронів, ваги синапсів (у тому числі нульові, якщо необхідно усунути зв'язок між деякими нейронами).

Алгоритм пошуку оптимальної ANN для деякого числового ряду наступний:

- Випадково генерується вихідна популяція (набір хромосом) нейронних мереж.
- Розраховується фенотип (архітектура ANN) та значення оптимізаційної функції кожної «особи» популяції.
- Застосовуються методи генетичних алгоритмів, такі як елітизм, відбір вдалих особин, кросовер та мутація для формування нової популяції.
- Попередні два кроки повторюються кілька разів до настання ефекту перенавчання мережі.

Методи локальної апроксимації застосовуються для прогнозування хаотичних і квазіперіодичних рядів, тобто процесів, у яких може бути відсутня глобальна лінійна складова. Тобто прогнозування ґрунтується на локальній підпоследовності низки, причому модель все ще може залишатися лінійною в рамках деякої локальної вибірки. Вибірка ж формується за тимчасової близькості значень низки, а, по близькості у просторі затримок.

Основна особливість методу у тому, що довжина прогнозу визначається можливостями моделі, а динамічними властивостями самого ряду. У цьому

методом локальної апроксимації оцінюється динаміка низки, а аналіз основних закономірностей здійснюється з допомогою інших методів, наприклад сингулярного спектрального аналізу (SSA).

Алгоритм побудови прогнозу на один крок уперед приблизно наступний:

- Побудова матриця затримок та вибір локального уявлення.
- Визначення числа сусідів.
- Оцінка параметрів вибраної моделі та побудова прогнозу у припущенні, що  $X$  змінюється за тим самим законом і з тими самими параметрами, що і його сусіди.

Прогнозування значень ряду методом локальної апроксимації може бути виконано одним із наступних способів:

- Ітеративний. При цьому коефіцієнти впливу відхилень у рамках кожної з матриць затримок та середнє значення прогнозів від кожного із сусідів розраховуються один раз, а спрогнозований вектор використовується як новий стартовий.
- Ітеративний із перерахунком. Відмінність від попереднього полягає в перерахуванні всіх параметрів після отримання прогнозу та виборі нових сусідів.
- Прямий. Полягає в тому, що після обчислення прогнозного значення воно не додається до вихідних даних у подальших розрахунках. Перевага такого підходу полягає в тому, що таким чином немає накопичення помилки прогнозу.

Методи локальної апроксимації мають багато спільного з широко відомими авторегресійними методами прогнозування рядів. Однак, за рахунок використання локально - шматкової лінійної апроксимації замість глобальної лінійної, методи ЛА дозволяють враховувати квазіперіодичну та хаотичну природу процесу, чого неможливо досягти за допомогою авторегресійних методів .

Методи оцінки точності прогнозу. Для визначення ступеня адекватності моделі прогнозування необхідно використовувати низку методик оцінки



точності прогнозу, зокрема - помилки прогнозування. Помилка прогнозування ET - це різниця між реалізацією низки та прогнозним значенням ряду:

$$e_t = X_{T+\tau} - X'_{T+\tau}, \quad (1.43)$$

де  $X_{T+\tau}$  – реальне значення ряду,  $X'_{T+\tau}$  – прогнозне значення ряду.

Розрахунок середнього значення помилки прогнозування (Mean Forecast Error) MFE - найпростіший спосіб оцінити точність прогнозування:

$$MFE = \frac{1}{n} \sum_{t=1}^n e_t \quad (1.44)$$

Дозволяє оцінити середнє відхилення прогнозних значень від реалізації низки, а також знак чи напрямок помилки. У цьому абсолютне значення MFE ще означає високої точності прогнозування. До того ж, MFE залежить від тимчасового масштабу ряду і слабо враховує екстремальні значення помилки прогнозу.

Абсолютна середня помилка прогнозування MAE (Mean Absolute Error) :

$$MAE = \frac{1}{n} \sum_{t=1}^n |e_t| \quad (1.45)$$

Дозволяє розрахувати середнє абсолютне відхилення прогнозу від реальних значень низки, оцінити сумарну помилку прогнозу. На відміну від MFE, негативні та позитивні за значенням помилки прогнозування не компенсують один одного, при цьому неможливо оцінити напрямок помилки. Також залежить від тимчасового масштабу вибірки та слабо враховує екстремальні значення помилок.

Середня абсолютна помилка у відсотках MAPE (Mean Absolute Percentage Error):

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{e_t}{X_t} \right| \times 100 \quad (1.46)$$

Оцінює середній відсоток помилки прогнозування без урахування знаку. MAPE не відображає екстремальних значень помилок, при цьому протилежні за знаком помилки не компенсують один одного.

Знакова середньоквадратична помилка SMSE (Signed Mean Squared Error):

$$SMSE = \frac{1}{n} \sum_{t=1}^n \left( \frac{e_t}{|e_t|} \right) e_t^2 \quad (1.47)$$

Показує середньоквадратичне відхилення помилки прогнозу з урахуванням знак.

Нормальна середньоквадратична помилка (Normalized Mean Squared Error) NMSE:

$$NMSE = \frac{1}{\sigma^2 n} \sum_{t=1}^n e_t^2,$$

$$\sigma^2 = \frac{1}{n-1} \sum_{t=1}^n (X_t - \bar{X})^2,$$

$$\bar{X} = \frac{1}{n} \sum_{t=1}^n X_t, \quad (1.50)$$

Досить ефективний спосіб оцінки точності прогнозування, у якому чим нижче значення NMSE, тим точніше модель.

Як правило, при оцінці точності методу прогнозування часового ряду використовується ряд метрик, що дозволяють розрахувати абсолютне значення помилки та її спрямованість.

Розглянуті алгоритми прогнозування процесів є загальновідомими і завдання дослідження часових рядів застосовуються досить давно. Однак для прогнозування реального процесу, в силу його характеру, деякі моделі підходять краще за інших. При цьому підбір відповідної моделі, що описує досліджуваний процес, є нетривіальним завданням, рішення якого необхідно базувати на точних статистичних та динамічних даних про природу самого процесу.

Згідно з відомою методологією Бокса-Дженкінса, аналіз та прогноз тимчасових рядів можна розділити на три етапи:

- Ідентифікація моделі, що найбільш точно описує процес:
  - про перевірку низки на стаціонарність;
  - аналіз повної та приватної АКФ.
- Проведення оцінки моделі та перевірка її адекватності:
  1. оцінка параметрів моделі, розрахованих першому етапі.
  2. Проперевірка моделі на відповідність вихідним даним.
- Розрахунок прогнозних значень ряду.

Найчастіше для нестационарних рядів властивість стаціонарності досягається операцією усунення тренда, тобто взяття різниці між сусідніми елементами (Малюнок4).

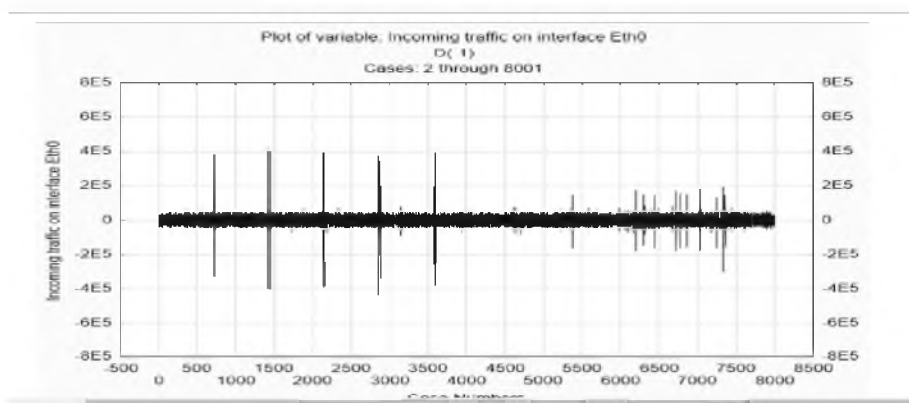


Рисунок 4. Графік інтенсивності вхідного трафіку (Eth0) після операції усунення тренда

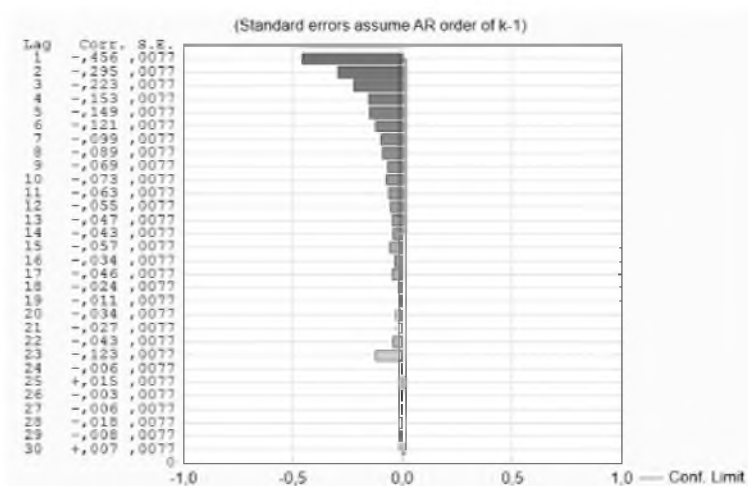


Рисунок 5 Графік ЧАКФ інтенсивності вхідного трафіку (Eth0) після операції усунення тренду

Порядок моделі авторегресії ( $AR(p)$ ) вибирається за графіком ЧАКФ (Малюнок 5), як останній ненульовий елемент, у разі  $p=23$ , тобто модель  $AR(23)$ . Порядок моделі ковзного середнього ( $MA(g)$ ) вибирається за графіком АКФ (Малюнок 6) як останній ненульовий елемент, тобто  $g=24$ , тобто  $MA(24)$ . Авторегресійна проінтегрована модель ковзного середнього у разі буде  $ARIMA(23,1,24)$ .

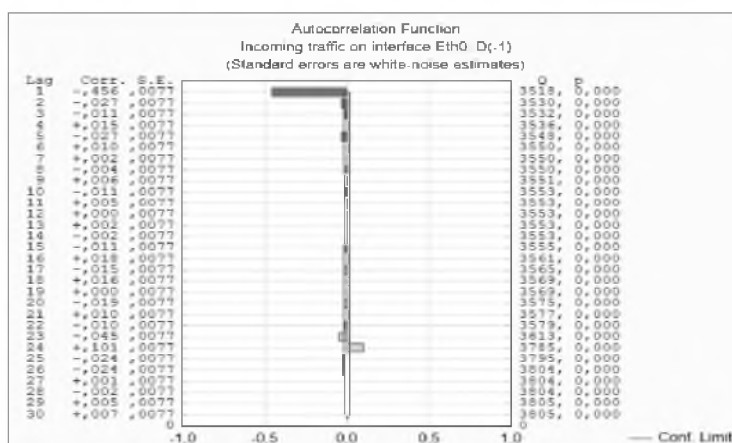


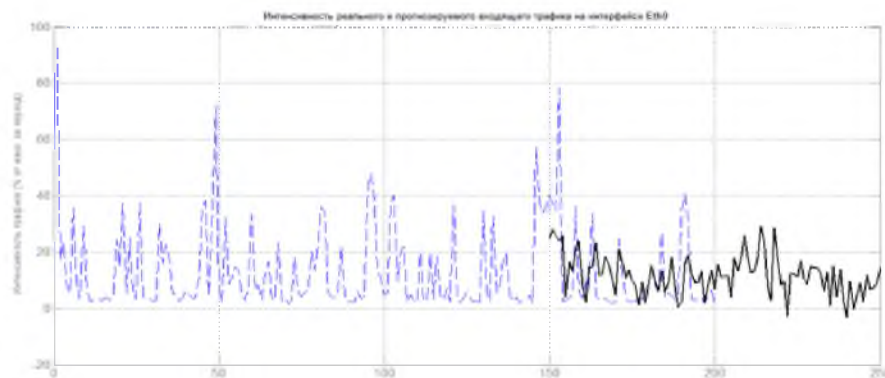
Рисунок 6. Графік ЧАКФ інтенсивності вхідного трафіку (Eth0) після операції

Надалі прогноз будувався згідно з викладеним алгоритмом на один крок вперед з метою виконати аналіз ефективності різних алгоритмів прогнозування стосовно досліджуваної області.

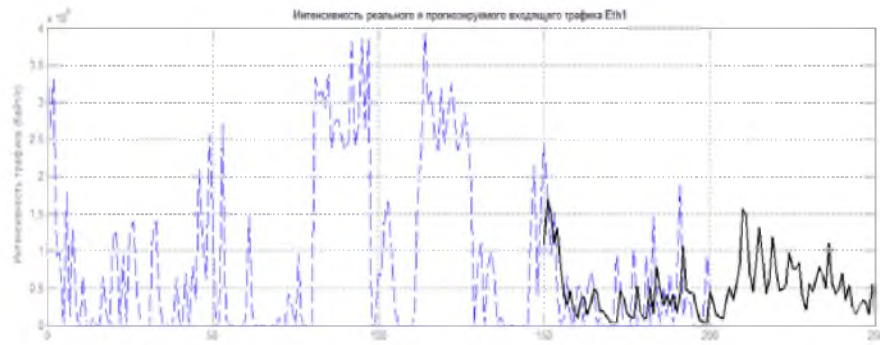
Варто відзначити, що оцінка точності прогнозу проводилася одним з найбільш простих способів через великі числові значення даних і різких короткочасних підвищення інтенсивності досліджуваних процесів, що при використанні інших способів оцінки прогнозу призводить до необ'єктивних кількісних значень.

### 1.2.3 Прогноз на основі $AR(p)$ – моделі

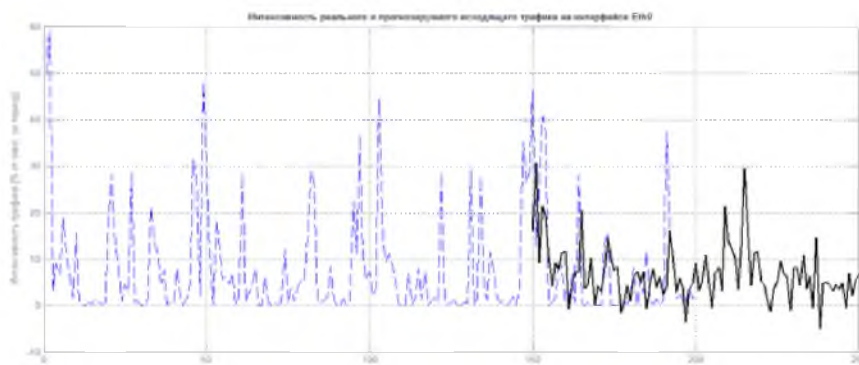
Розрахунки проводилися на зібраних під час експерименту даних, оцінювалася точність прогнозу, і навіть підбір значень параметра  $p$  моделі  $AR(p)$  як по АКФ, а й простим перебором значень порядку моделі. Графічно результати прогнозування представлені малюнки.



Малюнок 7. Інтенсивність реального та прогнозованого вхідного трафіку (EthO)

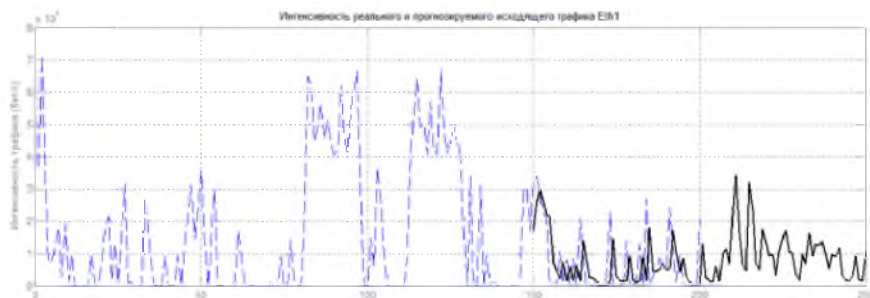


Малюнок 8. Інтенсивність реального та прогнозованого вхідного трафіку (Eth1)



Малюнок 9. Інтенсивність реального та прогнозованого вихідного трафіку (Eth0)

x Інтенсивність реального та прогнозованого вихідного трафіку Eth1



Малюнок 10. Інтенсивність реального та прогнозованого вихідного трафіку (Eth1)

Помилка прогнозу розраховувалася за відомою формулою:

$$ME = \frac{1}{n} \sum_{t=1}^n e_t, \quad (1.51)$$

де  $e_t = X_{T+t} - \hat{X}_{T+t}$  – це помилка прогнозування. Для процесів передачі трафіку MAE зазвичай приймала значення близько 12%, малюнки 39-42.

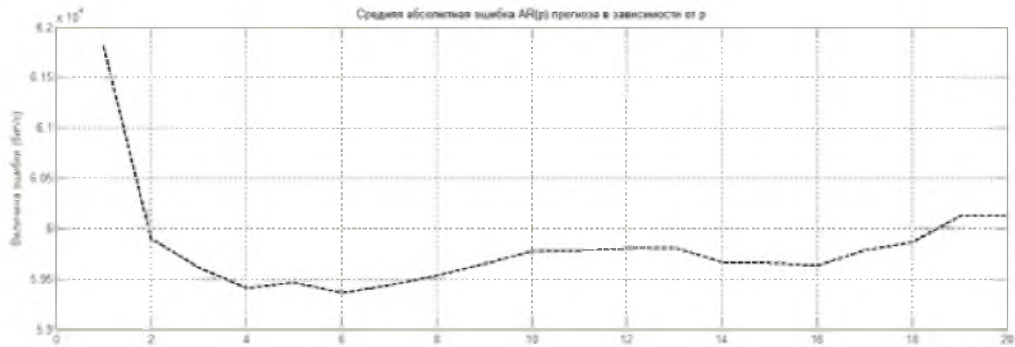


Рисунок 11. ME AR(p) прогнозу залежно від p для вхідного трафіку (Eth1)

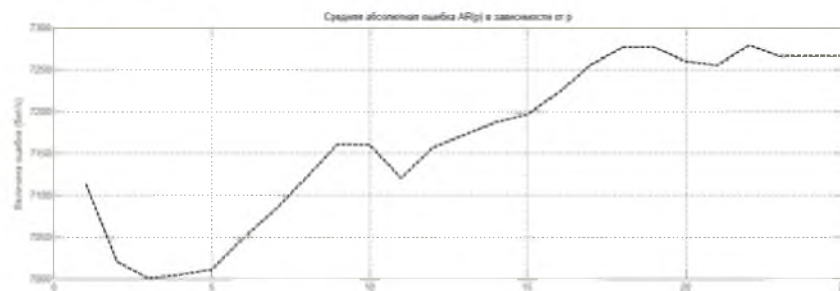


Рисунок 12. ME AR(p) прогнозу залежно від p для вхідного трафіку (Eth0)

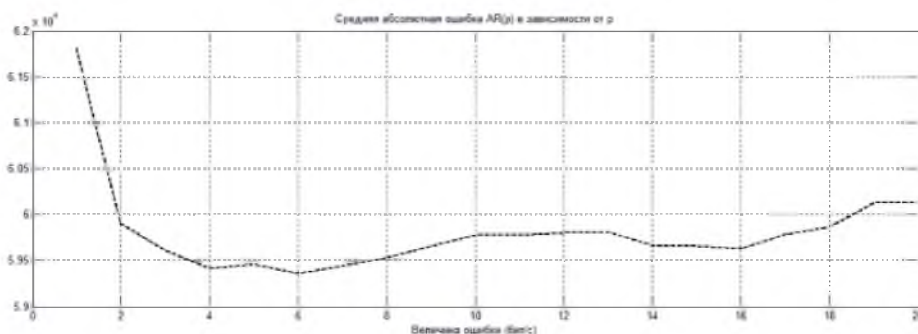


Рисунок 13. ME AR(p) прогнозу залежно від p для вихідного трафіку (Eth0)

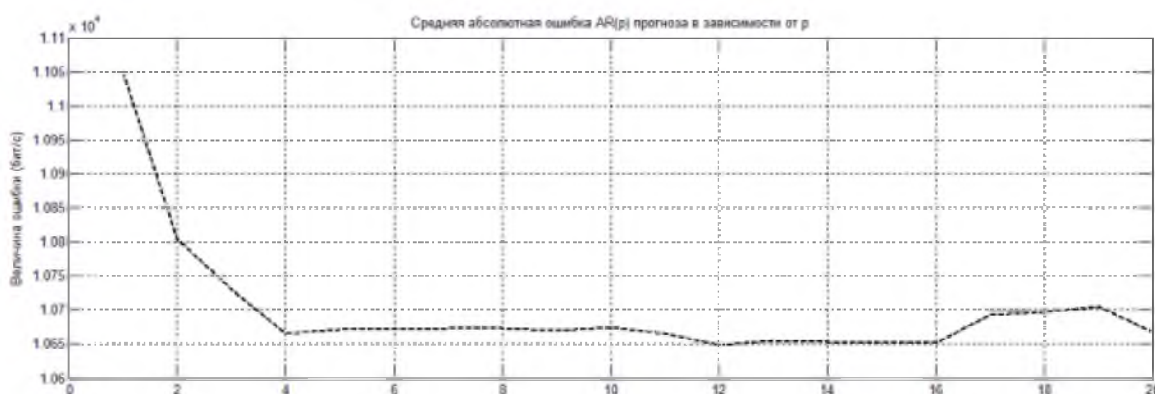


Рисунок 14. ME AR(p) прогнозу залежно від p для вихідного трафіку (Eth1)

Варто зауважити, що процеси розподілу апаратних ресурсів сервера відрізняються більшою лінійністю та наявністю періодичних складових, що було зазначено ще за статистичного аналізу. Для таких процесів AR(p) модель дає більшу точність та якість.

прогнозування. Наприклад, ME для процесу завантаження ЦП становила лише 1-3% від реального значення.

Реальне та прогнозоване значення завантаження ЦП





Рисунок 15. Реальне завантаження ЦП та прогнозоване значення

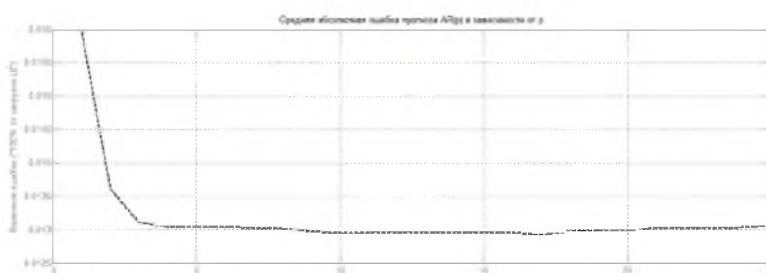


Рисунок 16. ME AR(p) прогнозу залежно від p для завантаження ЦП

Таблица 1. Помилка прогнозу AR(p) моделі

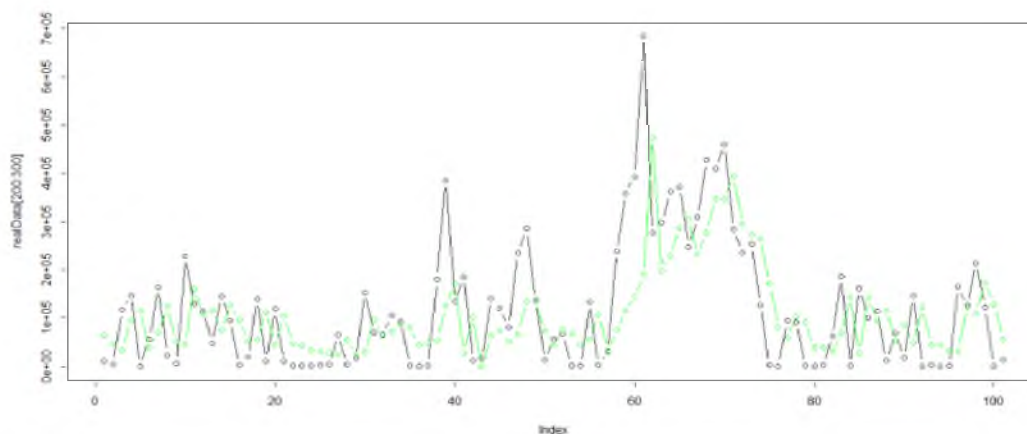
	Прогнозований процес	Помилка прогнозу (MAE)
1	Вхідний трафік Eth0	12,33%
2	Вихідний трафік Eth0	11,81%
3	Вхідний трафік Eth1	12,26%
4	Вихідний трафік Eth1	11,98%
5	Завантаження ЦП	6,14%
6	Об'єм вільної пам'яті	5,97%

#### 1.2.4 Прогноз на основі ARIMA (p, d, q) - моделі

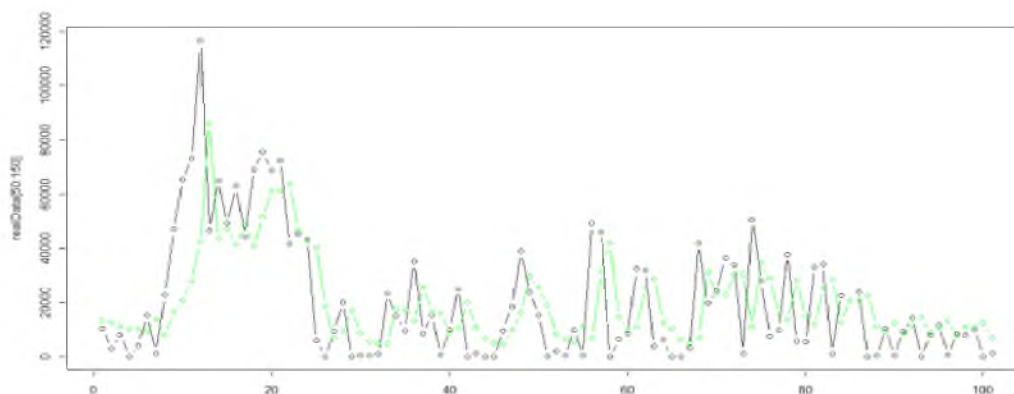
ARIMA(p,d,q) модель є сукупністю авторегресійної моделі та ковзного середнього, де p - порядок AR(p) моделі, q - порядок MA(q) моделі, d - порядок різниці між значеннями ряду (для усунення тренду). ARIMA(p,d,q) модель може бути представлена таким виразом:

$$X_t = c + \varepsilon_t + \sum_{i=1}^p \varphi_i X_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} . \quad (1.51)$$

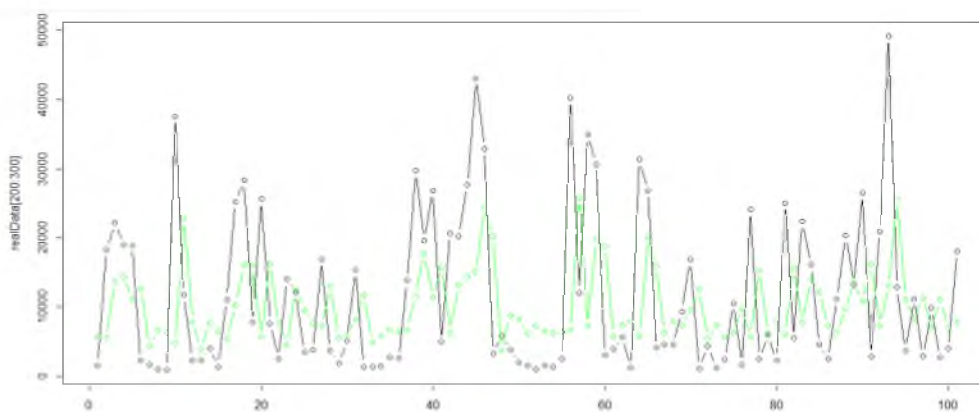
Для оцінки відповідних значень  $p$ ,  $d$  і  $q$  - порядків моделі було здійснено розрахунок помилки прогнозування з урахуванням  $ARIMA(p,d,q)$ , де  $p$  і  $q$  приймали значення від 1 до 10.



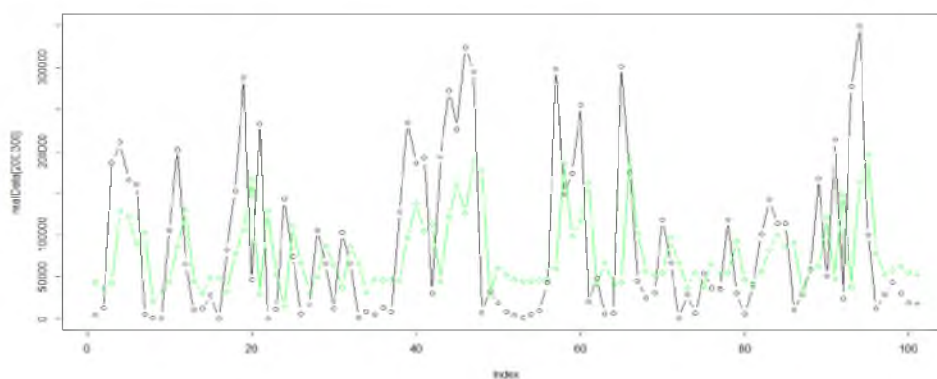
Малюнок 17. Інтенсивність реального та прогнозованого вхідного трафіку (Eth1)



Малюнок 18. Інтенсивність реального та прогнозованого вихідного трафіку (Eth1)



Малюнок 19. Інтенсивність реального та прогнозованого вхідного трафіку (Eth0)



Малюнок 20. Інтенсивність реального та прогнозованого вихідного трафіку (Eth0)

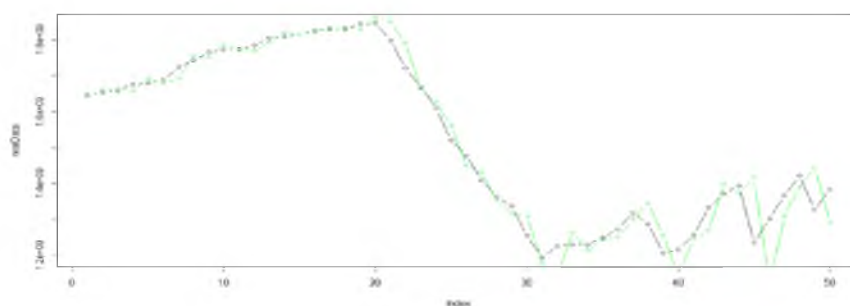
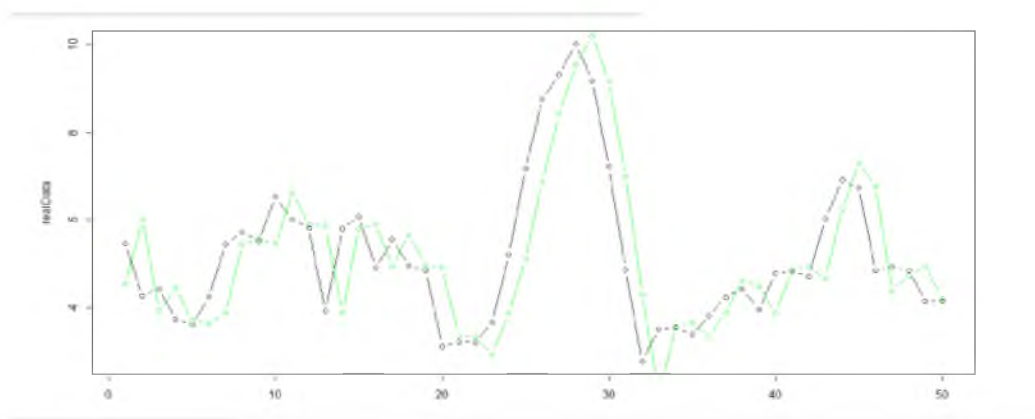


Рисунок 21. Реальний та прогнозований обсяг вільної пам'яті



Малюнок 22. Реальне та прогнозоване завантаження ЦП

Таблиця 2. Помилка прогнозу ARMA(p,d,q) моделі

	Прогнозований процес	Помилка прогнозу (MAE)
1	Вхідний трафік Eth0	10,56%
2	Вихідний трафік Eth0	9,92%
3	Вхідний трафік Eth1	10,47%
4	Вихідний трафік Eth1	10,25%
5	Завантаження ЦП	5,94%
6	Об'єм вільної пам'яті	5,13%

### 1.2.5 Прогноз методом SSA («Гусениця»)

Алгоритм Singular Spectrum Analysis (SSA) або «Гусениця» - це відносно новий і потужний спосіб обробки часових рядів, який увібрав у себе елементи класичних методів аналізу, багатовимірної статистики, багатовимірної геометрії та динамічних систем

Метод застосовний для аналізу будь-яких процесів, в яких потенційно може бути виявлена складна структура, шляхом розкладання вихідного часового ряду на композицію невеликої кількості незалежних, інтерпретованих компонентів, таких як тренди, що повільно змінюються, періодичні складові і шуми.

На першому етапі при обробці ряду  $X_1t^{\wedge}, X_N$  методом SSA вибирається вікно довжини  $L$  ( $1 < L < N$ ) і будується масив векторів  $X'$ :

$$X_i^* = (X_i, \dots, X_{i+L-1})^T, i = 1, 2, \dots, K = N - L + 1, \quad (1.52)$$

з якого формується матриця

$$X = (X_{i+j-1}^*)_{i,j=1}^{L,K} = [X_1^* : \dots : X_K^*] \quad \text{розміру } L * K, \quad (1.53)$$

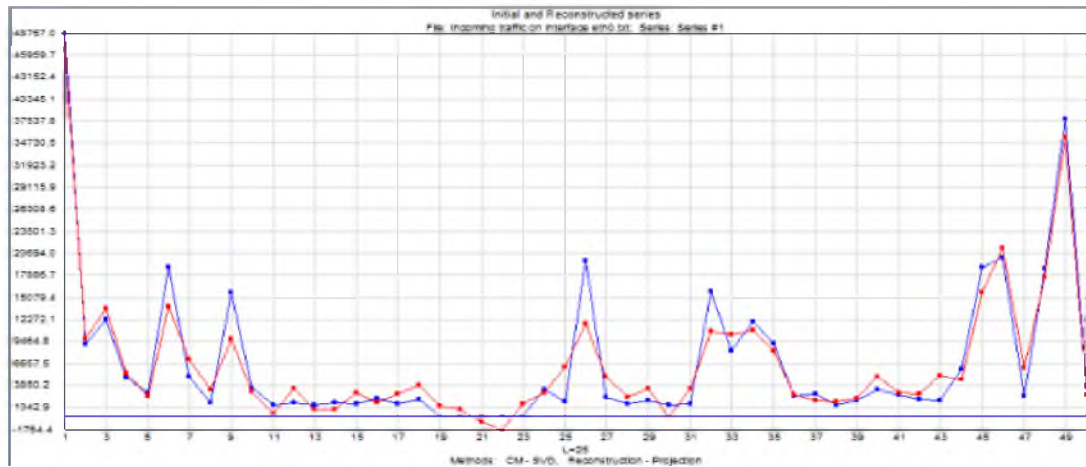
яку часто називають матрицею траєкторій. Це Ганкелева матриця, тобто всіх діагоналях, перпендикулярних головної, стоять однакові елементи (елементи  $i+j=\text{const}$ ).

Стовпці  $X_j^*$  матриці  $X$  приймаються за вектора, що лежать у  $L$ -розмірному просторі  $R^L$ . Сингулярне розкладання  $XX^T$  дозволяє отримати  $L$  сингулярних власних чисел та векторів. Комбінація деякого числа  $1 < L$  цих власних векторів визначає  $l$ -розмірний субпростір у  $R^L$ . Потім  $L$ -розмірний набір даних  $\{X_1, \dots, X_K\}$  проектується на  $l$ -розмірний субпростір з подальшим усередненням

за діагоналями, що дає деяку Ганкелеву матрицю  $X$ , яка є деяким наближенням матриці  $X$ . Ряд, отриманий відновленням  $X$  матриці, задовольняє деякому лінійному рекурентному співвідношенню, яке може бути використане для прогнозування майбутніх значень .

Крім прогнозування базовий алгоритм SSA може бути використаний для згладжування даних, фільтрації, усунення шуму, детрендингу, вилучення періодик. Існує ряд модифікацій алгоритму SSA, наприклад, для аналізу стаціонарних рядів, Монте-Карло SSA, алгоритми з модифікованою операцією сингулярного розкладання та обробки багатовимірних рядів.

Що стосується прогнозування трафіку SSA показав кращий результат, ніж ARIMA і AR.



Малюнок 23. Інтенсивність реального та прогнозованого вхідного трафіку (Eth0)

Таблиця 3 Помилка прогнозу моделі SSA

	Прогнозований процес	Помилка прогнозу (MAE)
1	Вхідний трафік Eth0	9,93%
2	Вихідний трафік Eth0	10,21%
3	Вхідний трафік Eth1	10,16%
4	Вихідний трафік Eth1	9,98%
5	Завантаження ЦП	5,54%
6	Об'єм вільної пам'яті	4,97%

### 1.2.6 Прогноз на основі ARFIMA моделі

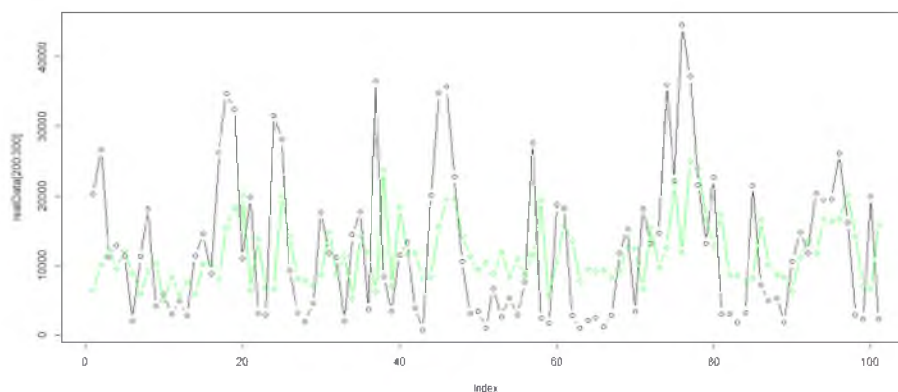
Як відомо, завдання прогнозування значень тимчасового ряду ґрунтується на моделі, яка описує процес. Чим достовірніше і точніше модель відображає процес, тим ближчі прогнозні значення до реальних даних. Раніше були наведені методи прогнозування рядів на основі авторегресії та ковзної середньої, а також згадувалося про відкриття ефекту самоподібності, що в сукупності є модель ARFIMA (Autoregressive fractionally integrated moving average). Моделювання процесів з урахуванням довгострокової залежності розглянуто на роботах , завдання прогнозування широко розглянуті в .

У цій роботі ARFIMA модель оптимізувалася на історичній вибірці на основі алгоритму Бройден-Флетчер-Гольдфарб-Шанно. БФГШ використовує наближення Тейлора цільової функції в околиці точки  $d$   $X$  :

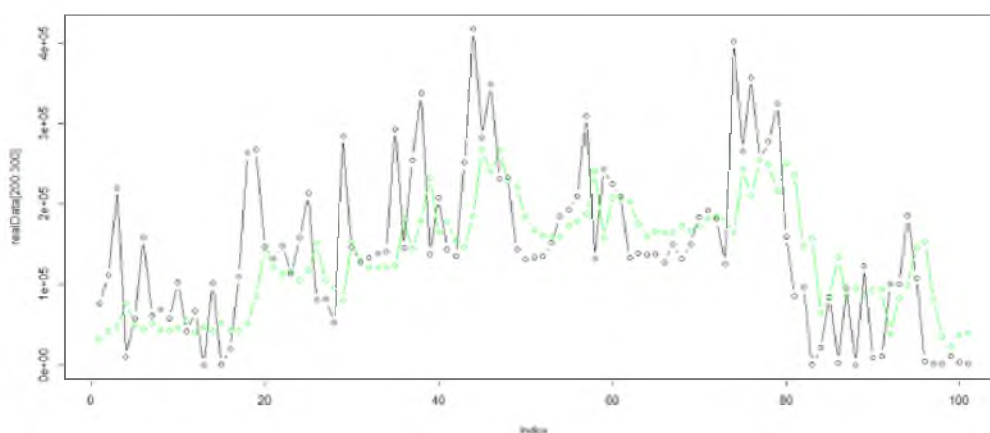
$$f(x+d) \approx q(d) = f(x) + d^T g(x) + \frac{1}{2} d^T H(x) d, \quad (1.54)$$

де  $g(x)$  – градієнт, а  $H(x)$  – матриця Гессе .

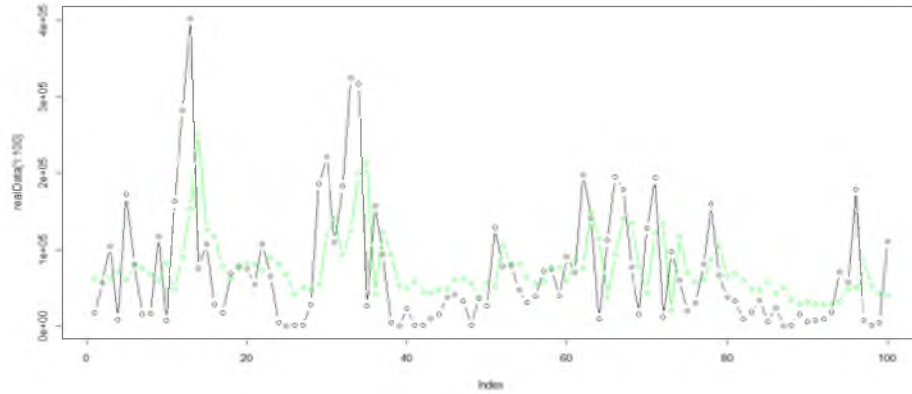
Після оптимізації моделі проводився прогноз на 1 крок уперед, після чого модель знову оптимізувалася, і розраховувався нове прогнозне значення і т.д.



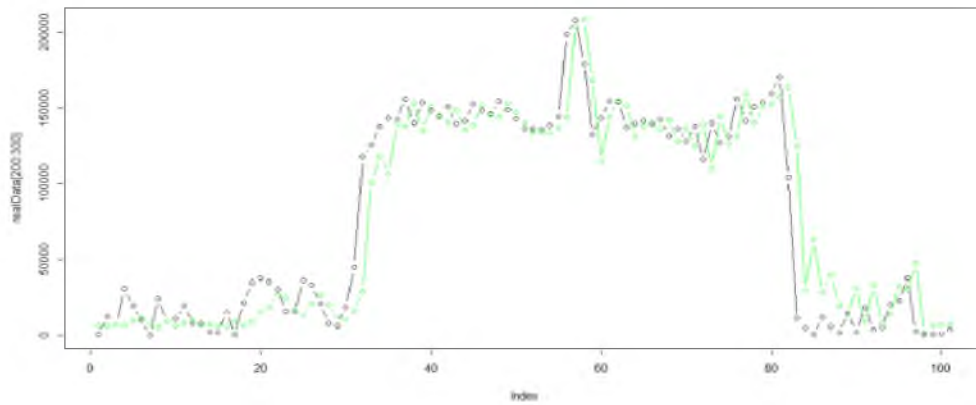
Малюнок 24. Інтенсивність реального та прогнозованого вхідного трафіку (Eth0)



Малюнок 25. Інтенсивність реального та прогнозованого вхідного трафіку (Eth1)



Малюнок 26. Інтенсивність реального та прогнозованого вихідного трафіку (Eth0)



Малюнок 27. Інтенсивність реального та прогнозованого вихідного трафіку (Eth1)

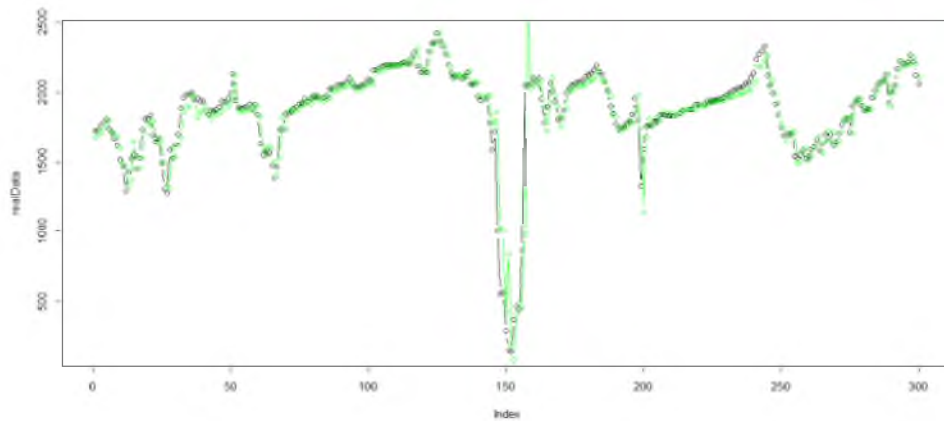
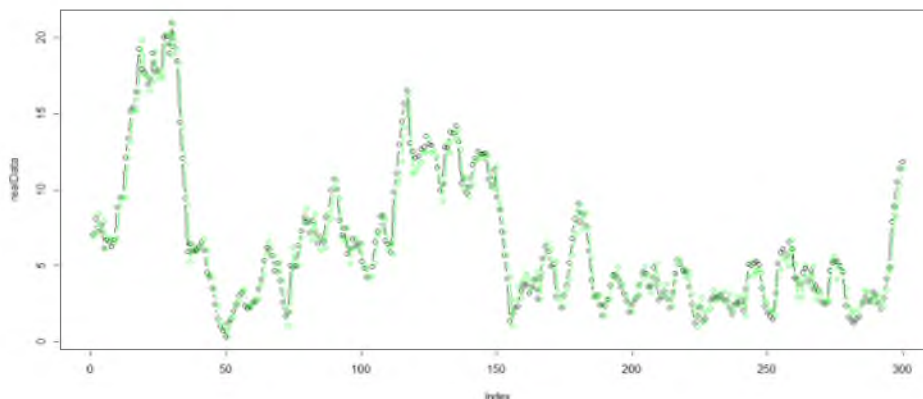


Рисунок 28. Реальний та прогнозований обсяг вільної пам'яті





Малюнок 29. Реальне та прогнозоване завантаження ЦП

Таблиця 4. Помилка прогнозу ARFIMA(p,d,q) моделі

	Прогнозований процес	Помилка прогнозу (MAE)
1	Вхідний трафік Eth0	8,85%
2	Вихідний трафік Eth0	9,21%
3	Вхідний трафік Eth1	8,46%
4	Вихідний трафік Eth1	8,95%
5	Завантаження ЦП	4,94%
6	Об'єм вільної пам'яті	4,57%

### 1.3 Постановка задачі:

Метою цієї роботи є розробка моделі мережі з пакетним трафіком для управління передачею на основі короткострокового прогнозування. Для реалізації поставленої мети необхідно вирішити такі завдання:

- 1 Виконати налітичний огляд моделей мережевого трафіку
- 2 Дослідити методи прогнозування навантаження в мережах з комутацією пакетів
- 3 Розробити імітаційну модель мережі з короткостроковим прогнозуванням
- 4 Дослідити характеристики мережі та оцінити ефективність отриманих результатів

## 1.4 Висновки

1) Для вибору відповідного методу прогнозування досліджуваних процесів було розглянуто найбільш відомі моделі часових рядів.

Постановка задачі:

Метою цієї роботи є розробка моделі мережі з пакетним трафіком для управління передачею на основі короткострокового прогнозування. Для реалізації поставленої мети необхідно вирішити такі завдання:

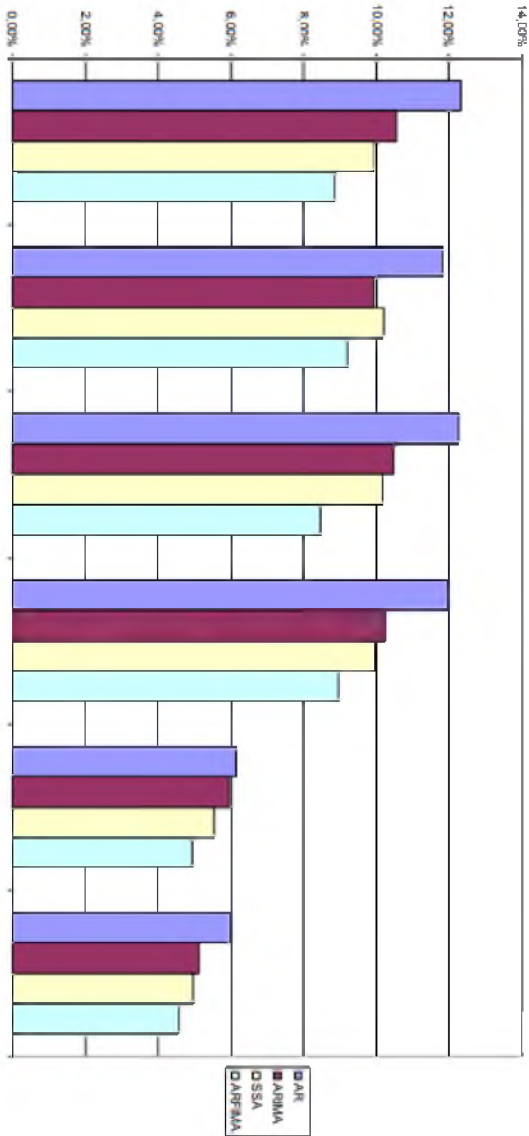
- 1 Виконати налітничний огляд моделей мережевого трафіку
- 2 Дослідити методи прогнозування навантаження в мережах з комутацією пакетів
- 3 Розробити імітаційну модель мережі з короткостроковим прогнозуванням
- 4 Дослідити характеристики мережі та оцінити ефективність отриманих результатів

2) Було проведено оцінку точності прогнозу основних характеристик сервера та потоку трафіку (Малюнок 58).

3) Модель ARFIMA(p,J,g) перевищує інші розглянуті за точністю прогнозування, що узгоджується зі специфікою моделі та самоподібним характером досліджуваних процесів.



Прогнозований процес
Вхідний трафік Eth0
Вихідний трафік Eth0
Вхідний трафік Eth1
Вихідний трафік Eth1
Завантаження ЦП
Об'єм вільної пам'яті



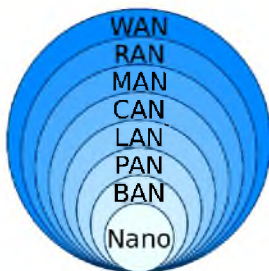
## СПЕЦІАЛЬНА ЧАСТИНА

### 2. Імітаційне моделювання мережі з короткостроковим прогнозуванням

#### Розробка моделі мережі з короткостроковим прогнозуванням навантаження

У ряді випадків вивчення реальної комп'ютерної мережі та окремих її вузлів може бути важко, пов'язане з великими тимчасовими та матеріальними витратами, а іноді й неможливе з міркувань безпеки. У такому разі виникає необхідність створення комп'ютерної моделі, що відображатиме реальні процеси з метою подальшого їх дослідження. Імітаційне моделювання - відомий спосіб вивчення характеристик системи без проведення реального експерименту, проте дозволяє робити висновки про властивості протікають процесів за умов адекватності моделі. З огляду на умови адекватності моделі логічно вести розробку останньої з урахуванням всіх відомих характеристик процесів, які у реальному об'єкті дослідження.

#### Типи комп'ютерних мереж за масштабом



Мал.31 Класифікація мереж даних за просторовим обсягом.

Нанорозмір ближнього поля (NFC) Тіло Персональний (PAN) Біля мене Локальний (LAN) Сховище (SAN) Бездротовий (WLAN) Віртуальний (VLAN) Дім (HAN) будівля Кампус (CAN) хребет Метрополітен (MAN) Муніципальний бездротовий зв'язок (MWN) Широкий (WAN) Хмара Інтернет Міжпланетний Інтернет

Мережі можуть характеризуватися багатьма властивостями або особливостями, такими як фізична ємність, організаційне призначення, авторизація користувачів, права доступу та інші. Іншим відмінним методом класифікації є фізичний протяжність або географічний масштаб. Нанорозмірна мережа Нанорозмірна мережа має ключові компоненти, реалізовані на нанорозмірі, включаючи носії повідомлень, і використовує фізичні принципи, які відрізняються від механізмів комунікації на макророзмірі. Нанорозмірна комунікація поширює комунікацію на дуже маленькі датчики та виконавчі механізми, такі як ті, що знаходяться в біологічних системах, а також прагне працювати в середовищах, які були б надто жорсткими для інших методів комунікації.

### **Персональна мережа**

Персональна мережа (PAN) — це комп'ютерна мережа, яка використовується для зв'язку між комп'ютерами та різними інформаційно-технологічними пристроями, розташованими поблизу однієї особи. Деякими прикладами пристроїв, які використовуються в PAN, є персональні комп'ютери, принтери, факсимільні апарати, телефони, КПК, сканери та ігрові консолі. PAN може включати дротові та бездротові пристрої. Радіус дії PAN зазвичай досягає 10 метрів. Дротова PAN зазвичай складається з підключеннями USB і FireWire, тоді як такі технології, як Bluetooth і інфрачервоний зв'язок, зазвичай утворюють бездротову PAN.

### **Локальна мережа**

Локальна мережа (LAN) — це мережа, яка об'єднує комп'ютери та пристрої в обмеженій географічній зоні, наприклад у домі, школі, офісній будівлі або групі будівель, розташованих близько. Дротові локальні мережі найчастіше базуються на технології Ethernet. Інші мережеві технології, такі як ITU-T G.hn, також забезпечують спосіб створення дротової локальної мережі за допомогою існуючої проводки, такої як коаксіальні кабелі, телефонні лінії та лінії електропередач. LAN можна підключити до глобальної мережі (WAN) за допомогою маршрутизатора. Визначальними характеристиками локальної

мережі, на відміну від глобальної мережі, є вищі швидкості передачі даних, обмежений географічний діапазон і відсутність опори на орендовані лінії для забезпечення з'єднання. Поточні технології Ethernet або інші технології локальної мережі IEEE 802.3 працюють при передачі даних швидкості до та понад 100 Гбіт/с, стандартизовані IEEE у 2010 році.

### **Домашня мережа**

Домашня мережа (HAN) — це домашня локальна мережа, яка використовується для зв'язку між цифровими пристроями, зазвичай розгорнутими вдома, як правило, невеликою кількістю персональних комп'ютерів і аксесуарів, таких як принтери та мобільні комп'ютерні пристрої. Важливою функцією є спільне використання доступу до Інтернету, часто широкопasmового доступу через кабельний доступ до Інтернету або постачальника цифрової абонентської лінії (DSL).

### **Мережа зберігання даних**

Мережа зберігання даних (SAN) — це виділена мережа, яка забезпечує доступ до консолідованого сховища даних на рівні блоків. Мережі SAN в основному використовуються для того, щоб зробити пристрої зберігання, такі як дискові масиви, стрічкові бібліотеки та оптичні музичні автомати, доступними для серверів, щоб сховище відображалось як локально підключені пристрої до операційної системи. SAN зазвичай має власну мережу пристроїв зберігання даних, які зазвичай недоступні через локальну мережу іншими пристроями. Вартість і складність SAN на початку 2000-х років впала до рівня, що дозволив ширше впровадження в середовищі як підприємств, так і малого та середнього бізнесу.

### **Мережа кампусу**

Кампусна мережа (CAN) складається із з'єднання локальних мереж у межах обмеженої географічної області. Мережеве обладнання (комутатори, маршрутизатори) і засоби передачі (оптичне волокно, кабелі Cat5 тощо) майже повністю належать орендареві або власнику кампусу (підприємство, університет, уряд тощо). Наприклад, мережа університетського містечка, ймовірно, з'єднає

різноманітні будівлі кампусу, щоб з'єднати академічні коледжі або факультети, бібліотеку та студентські гуртожитки.

### **Магістральна мережа**

Магістральна мережа — це частина інфраструктури комп'ютерної мережі, яка забезпечує шлях для обміну інформацією між різними локальними мережами або підмережами. Магістраль може об'єднувати різноманітні мережі в одній будівлі, у різних будівлях або на великій території. Під час проектування магістралі мережі продуктивність мережі та перевантаженість мережі є критичними факторами, які слід брати до уваги. Зазвичай пропускна здатність магістральної мережі більша, ніж пропускна здатність окремих мереж, підключених до неї. Наприклад, велика компанія може створити магістральну мережу для з'єднання відділів, розташованих по всьому світу. Обладнання, яке об'єднує мережі відділів, становить магістраль мережі. Іншим прикладом магістральної мережі є магістраль Інтернету, яка є масивною глобальною системою волоконно-оптичних кабелів і оптичних мереж, які передають основну частину даних між глобальними мережами (WAN), метро, регіональними, національними та трансокеанськими мережами.

### **Столична мережа**

Столична мережа (MAN) — це велика комп'ютерна мережа, яка зазвичай охоплює місто або велике кампус.

### **Глобальна мережа**

Глобальна мережа (WAN) — це комп'ютерна мережа, яка охоплює велику географічну територію, наприклад місто, країну, або охоплює навіть міжконтинентальні відстані. WAN використовує канал зв'язку, який поєднує багато типів носіїв, таких як телефонні лінії, кабелі та радіохвилі. Глобальна мережа часто використовує засоби передачі, що надаються звичайними операторами, наприклад телефонними компаніями. Технології WAN зазвичай функціонують на трьох нижніх рівнях еталонної моделі OSI: фізичному рівні, рівні каналу даних і мережевому рівні.



### **Корпоративна приватна мережа**

Корпоративна приватна мережа — це мережа, яку будує одна організація для з'єднання своїх офісів (наприклад, виробничих ділянок, головних офісів, віддалених офісів, магазинів), щоб вони могли спільно використовувати комп'ютерні ресурси.

### **Віртуальна приватна мережа**

Віртуальна приватна мережа (VPN) — це накладена мережа, у якій деякі зв'язки між вузлами здійснюються через відкриті з'єднання або віртуальні канали в деякій більшій мережі (наприклад, Інтернет), а не через фізичні дроти. Кажуть, що протоколи канального рівня віртуальної мережі тунелюються через більшу мережу, коли це так. Одним із поширених застосувань є безпечний зв'язок через загальнодоступний Інтернет, але VPN не обов'язково має явні функції безпеки, такі як автентифікація чи шифрування вмісту. VPN, наприклад, можна використовувати для розділення трафіку різних спільнот користувачів через базову мережу з потужними функціями безпеки. VPN може мати найкращу продуктивність або може мати певну угоду про рівень обслуговування (SLA) між клієнтом VPN і постачальником послуг VPN. Зазвичай VPN має складнішу топологію, ніж точка-точка.

### **Глобальна мережа**

Глобальна мережа (GAN) — це мережа, яка використовується для підтримки мобільного зв'язку в довільній кількості бездротових локальних мереж, зон покриття супутників тощо. Ключовою проблемою мобільного зв'язку є передача зв'язку користувача з однієї локальної зони покриття до іншої. У проєкті IEEE 802 це передбачає послідовність наземних бездротових локальних мереж.

### **Організаційні масштаби**

Зазвичай мережами керують організації, які ними володіють. Мережі приватних підприємств можуть використовувати комбінацію інтранетів і екстранетів. Вони також можуть надавати мережевий доступ до Інтернету, який

не має єдиного власника і дозволяє практично необмежене глобальне підключення.

### **Інтранет**

Інтранет — це набір мереж, які знаходяться під контролем одного адміністративного органу. Інтранет використовує протокол IP і інструменти на основі IP, такі як веб-браузери та програми для передачі файлів. Адміністративний суб'єкт обмежує використання внутрішньої мережі лише авторизованими користувачами. Найчастіше інтранет – це внутрішня локальна мережа організації. Велика інтрамережа зазвичай має принаймні один веб-сервер для надання користувачам організаційної інформації. Інтранет — це також усе, що знаходиться позаду маршрутизатора в локальній мережі.

### **Екстранет**

Екстранет — це мережа, яка також перебуває під адміністративним контролем однієї організації, але підтримує обмежене підключення до певної зовнішньої мережі. Наприклад, організація може надати доступ до деяких аспектів своєї внутрішньої мережі для обміну даними зі своїми діловими партнерами або клієнтами. Ці інші об'єкти не обов'язково є довіреними з точки зору безпеки. Мережне підключення до екстранету часто, але не завжди, здійснюється через технологію WAN.

### **Інтернет**

Мережа — це з'єднання кількох різних типів комп'ютерних мереж для створення єдиної комп'ютерної мережі шляхом накладання різного мережевого програмного забезпечення та з'єднання їх разом за допомогою маршрутизаторів.. Це глобальна система взаємопов'язаних урядових, академічних, корпоративних, державних і приватних комп'ютерних мереж. Він заснований на мережевих технологіях Internet Protocol Suite. Це наступник Мережі агентств передових дослідницьких проєктів (ARPANET), розробленої DARPA Міністерства оборони США. Інтернет використовує мідні комунікації та магістраль оптичної мережі, щоб увімкнути Всесвітню павутину (WWW), Інтернет речей, передачу відео та широкий спектр інформаційних послуг.

Учасники в Інтернеті використовують різноманітний набір методів із кількох сотень задокументованих і часто стандартизованих протоколів, сумісних із набором протоколів Інтернету та системою адресації (IP-адресами), якою керує Управління присвоєння номерів Інтернету та реєстри адрес. Постачальники послуг і великі підприємства обмінюються інформацією про доступність своїх адресних просторів через протокол прикордонного шлюзу (BGP), утворюючи резервну всесвітню сітку шляхів передачі.

### **Даркнет**

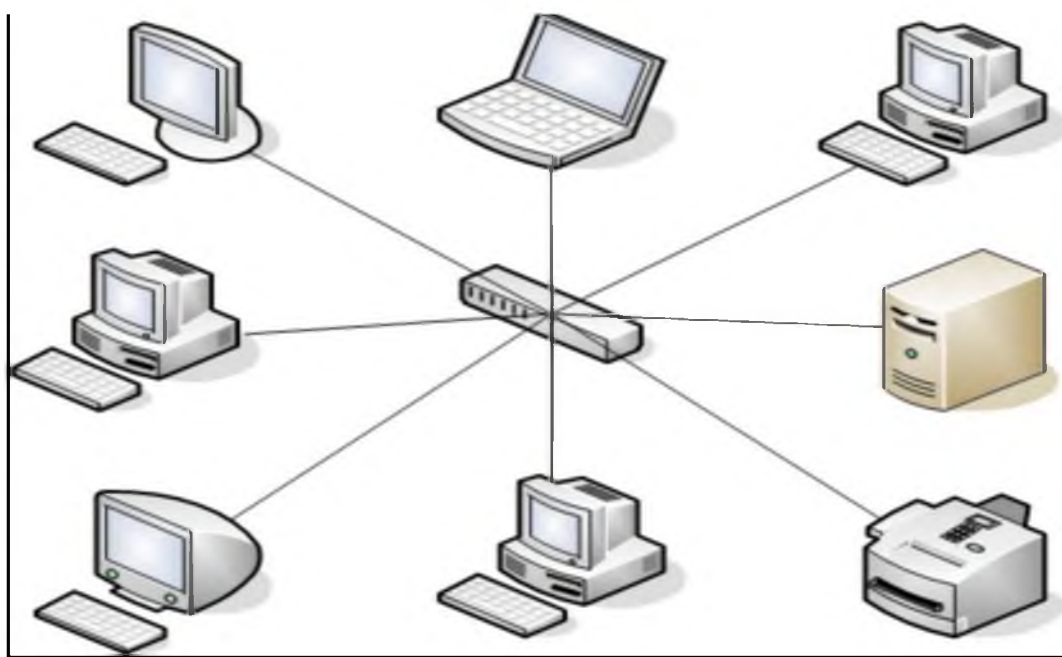
Даркнет — це накладна мережа, яка зазвичай працює в Інтернеті, доступ до якої доступний лише через спеціалізоване програмне забезпечення. Темна мережа — це анонімна мережа, де з'єднання здійснюються лише між довіреними одноранговими користувачами — іноді їх називають «друзями» (F2F) — за допомогою нестандартних протоколів і портів. Даркнет відрізняється від інших розподілених однорангових мереж, оскільки обмін є анонімним (тобто IP-адреси не є загальнодоступними), і тому користувачі можуть спілкуватися, не боячись урядового чи корпоративного втручання.

### **Мережевий сервіс**

Мережеві служби — це програми, розміщені на серверах у комп'ютерній мережі, щоб забезпечити певні функції для членів або користувачів мережі або допомогти самій мережі працювати. Всесвітня павутина, електронна пошта, друк і мережевий обмін файлами є прикладами добре відомих мережевих послуг. Мережеві служби, такі як DNS (система доменних імен), дають імена для IP- та MAC-адрес (люди запам'ятовують імена на зразок «`nm.lan`» краще, ніж числа на кшталт «`210.121.67.18`»), а також DHCP, щоб гарантувати, що обладнання на мережа має дійсну IP-адресу. Послуги зазвичай базуються на протоколі служби, який визначає формат і послідовність повідомлень між клієнтами та серверами цієї мережевої служби.

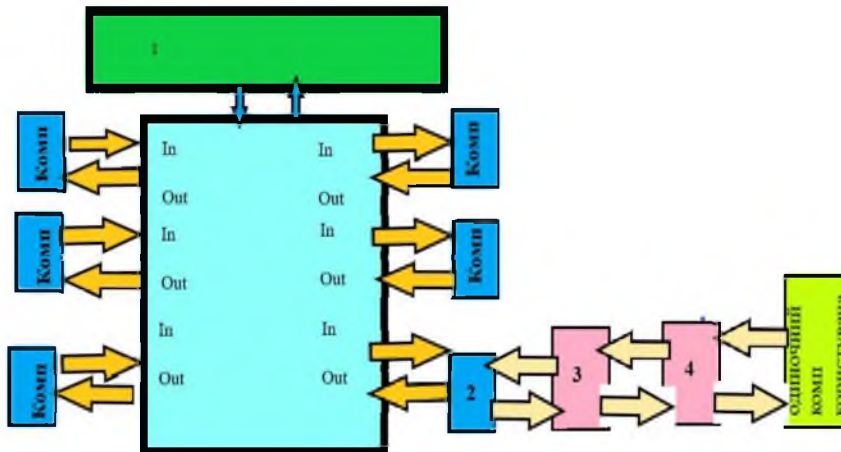
## 2.1 Розробка імітаційної моделі мережі

Створення моделі корпоративної мережі проводилося засобами бібліотеки SimEvents пакету MatLab. Принципово модель представлена малюнку 59. Це мережу з топологією типу «зірка» зі змінним числом серверів, підключених до комутатора. Блок Bandwidth дозволяє встановити пропускну здатність мережі, блоки Compl.CompN моделюють поведінку серверів, тобто генерують пакети повідомлень зі змінною швидкістю і різної довжини. Топологія "зірка" є найбільш популярною при організації мереж



Малюнок 32 Зіркоподібна топологія, що використовується в мережі

## 2.2 Модель мережі з комутацією пакетів



Малюнок 33. Загальний вигляд моделі

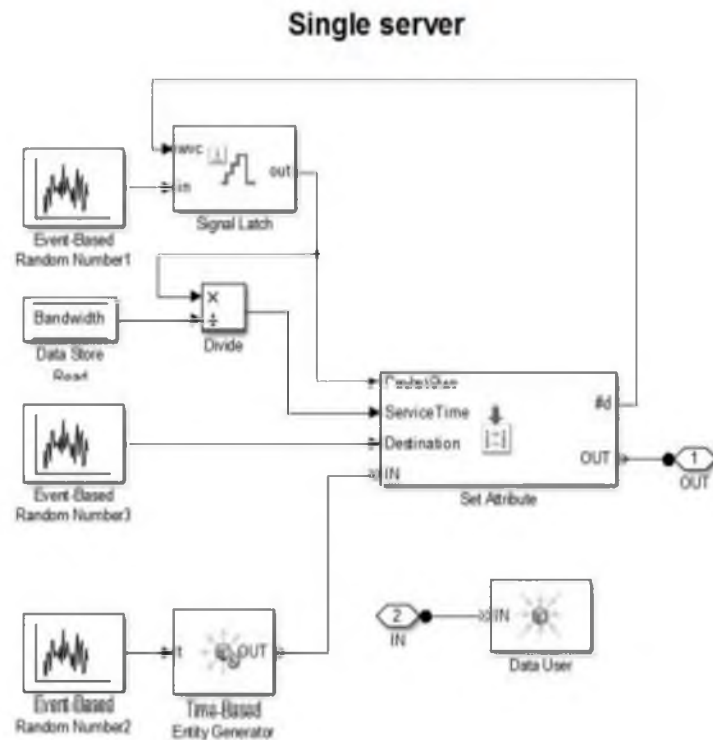
1- Bandwidth

2-Маршрутизатор

3-Роутер 1

4-Роутер 2 (в якості репітера)

Роль джерел повідомлень, тобто підключених до комутатора серверів грає модель, представлена малюнку. Із заданою швидкістю (пакетів/с) генеруються повідомлення, довжина яких у певному інтервалі - малюнок 34.



Малюнок 34. Модель сервера джерела пакетів даних

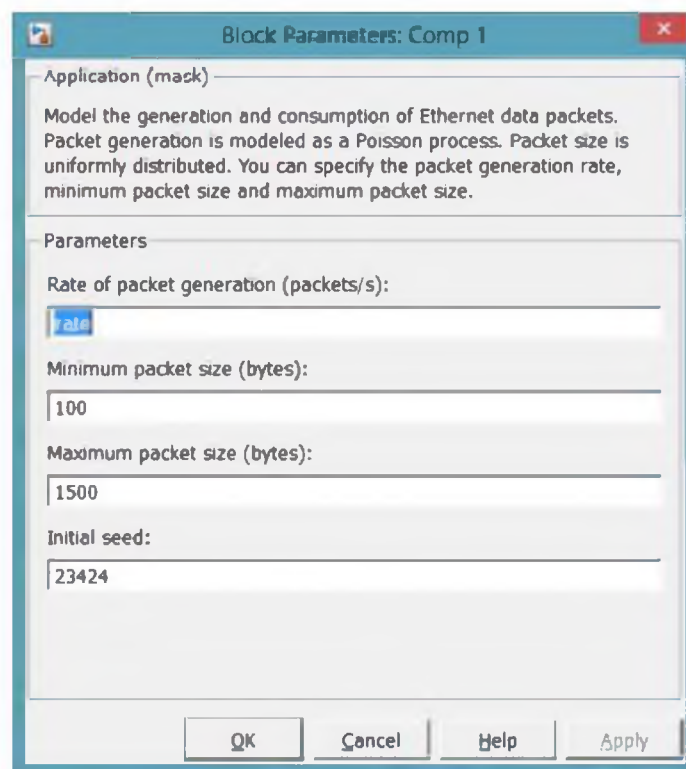
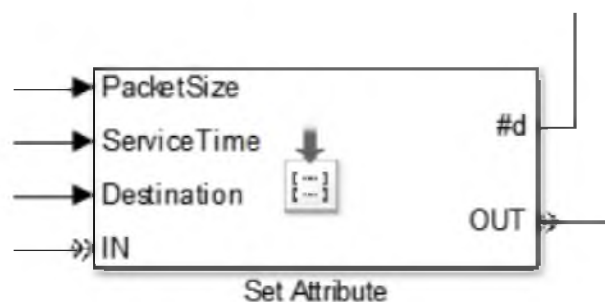


Рисунок 35. Основні характеристики джерела пакетів повідомлень

Кожному створеному пакету надається три параметри (Малюнок 36):

- PacketSize – довжина згенерованого пакета в байтах.
- ServiceTime – час обслуговування повідомлення.
- Destination – адреса призначення пакета.

Підключення до моделі комутатора здійснюється через з'єднання IN та OUT.

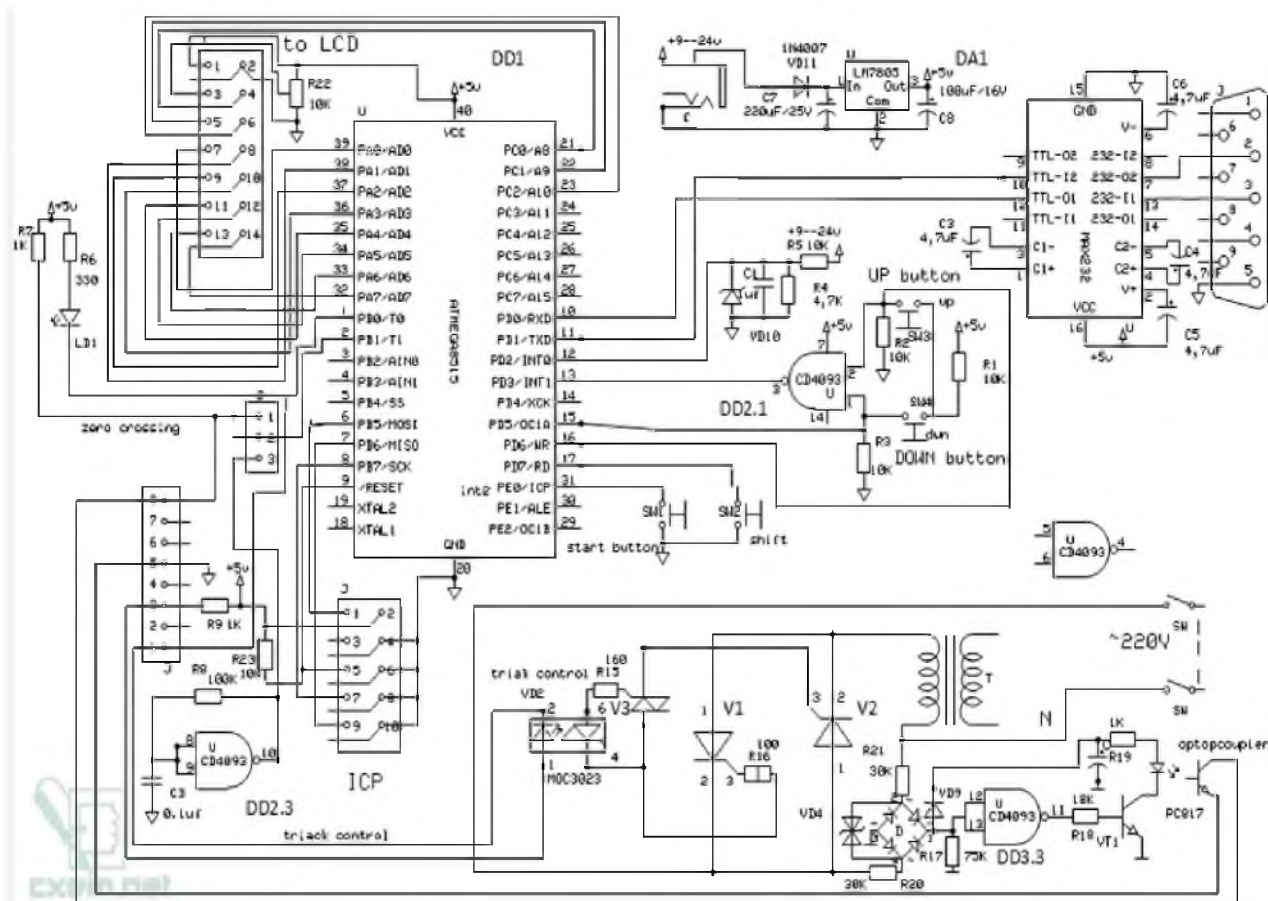


Малюнок 36. Блок Set Attribute моделі джерела повідомлень

Модель комутатора враховує дисципліну обслуговування заявок FIFO із контролем переповнення буфера. Дані про обсяг трафіку в комутаторі в кожний конкретний момент часу розраховуються з урахуванням довжини повідомлень у буфері та їх кількості, а потім передаються в середу MatLab для подальшої обробки.







Малюнок 38. Ethernet принципова схема

### Реалізовані алгоритми TCP забезпечення QoS

На сьогоднішній день TCP – це основний протокол мережі інтернет, який розроблявся з урахуванням низки методик управління трафіком та запобіганню перевантаженням у мережі передачі даних, базується на можливостях міжмережевого IP протоколу. Протокол до кожного повідомлення додає заголовок, структура якого представлена в таблиці 5.

Таблиця 5. Формат заголовка TCP-пакету

0	3	9	15	23	31
Порт джерела			Порт приймача		
Номер у послідовності					
Номер підтвердження					
Зміщення даних	Зарезервовано	URG; ACK; PSH; RST; SYN; FIN	Розмір вікна		
Контрольна сума			Показчик		
Додаткові дані заголовка					Дані вирівнювання

Досягти реалізації всіх алгоритмів передачі протоколу TCP - нетривіальне технічне завдання. Тому на даному етапі обмежимося найбільш важливими методиками TCP з управління навантаженнями. Реалізуємо динамічний контроль за розміром вікна передачі. У TCP використовується принцип «ковзного вікна», що полягає в тому, що кожна сторона передачі може відправляти максимум стільки байт, скільки було вказано в полі «розмір вікна» заголовка пакета, що підтверджує отримання попереднього банного блоку. Принцип «ковзного вікна» забезпечує випереджальну підтвердження посилку даних [109]. Для керування вікном передачі реалізуємо алгоритм повільного старту, полягає у відповідності інтенсивності передачі повідомлень джерелом з урахуванням обсягу буфера приймача. При цьому розмір вікна передачі поступово підвищується до виникнення повторних передач.

Реалізуємо також таймер повторної передачі. Таймер відраховує період до повторної передачі пакета, підтвердження якого не надійшло, позначається як RTO (Retransmission Timeout) і розраховується динамічно на основі часового інтервалу від моменту посилки повідомлення до отримання підтвердження - RTT (Round Trip Time):

$$\text{SRTT} = k * \text{SRTT} + (1 - k) * \text{RTT}, \quad (2.1)$$

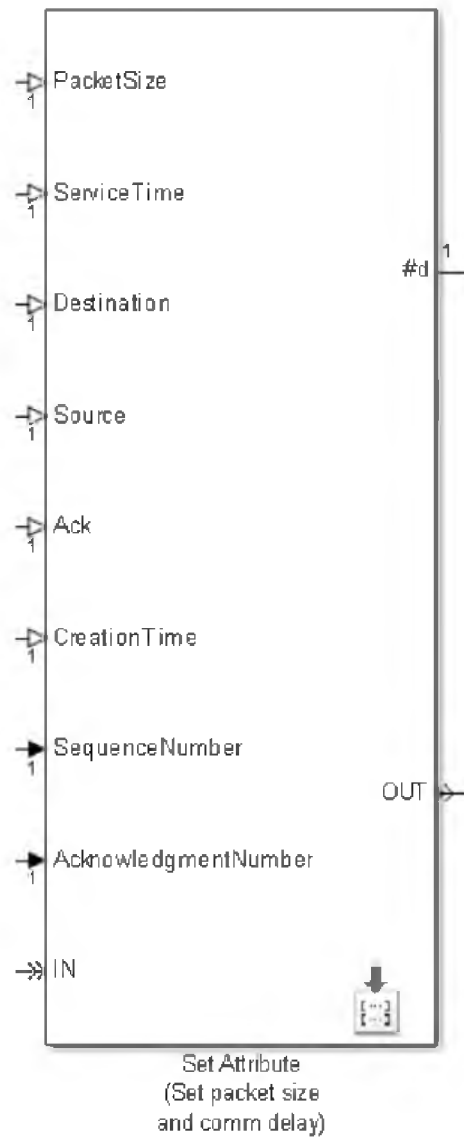
де SRTT (Smoothed RTT) - згладжене значення RTT, k - коефіцієнт, що згладжує. Тоді RTO обчислюється згідно з наступним виразом:

$$\text{RTO} = \min (U, \max (L, p * \text{SRTT})), \quad (2.3)$$

де U - обмеження зверху значення RTO, L – обмеження знизу значення RTO, p - деякий коефіцієнт. Якщо після зменшення таймера до нуля та повторної передачі підтвердження не приходить знову, то RTO збільшується і передача відновлюється.

### **Реалізація моделі мережі з комутацією пакетів з урахуванням алгоритмів TCP**

Для реалізації зазначених вище алгоритмів необхідно внести низку змін до розробленої раніше моделі. У першу чергу - додати кілька додаткових полів властивостей у кожен пакет, що генерується (рисунок 73). Поле Ack імітує службовий біт підтвердження заголовка TCP пакета. Поле CreationTime зберігає час створення пакета зручності розрахунку RTT і, відповідно, таймера повторної передачі. SequenceNumber служить визначення черговості пакетів. AcknowledgmentNumber використовується приймачем пакетів, щоб вказати відправнику номер наступного очікуваного пакета. Стан таймерів міститься у блоках Data Store. Після введення базових алгоритмів TCP, кількість джерел пакетів було збільшено, щоб залишалася можливість моделювати навантаження в мережі, при цьому залишаючи належну адекватність моделей абонентів.



Малюнок 39. Блок присвоєння параметрів пакету подальшої передачі  
Через значення, що зберігаються в блоках Data Store, проводиться  
розрахунок RTO і повторне відправлення неприйнятих пакетів (рисунок 40).

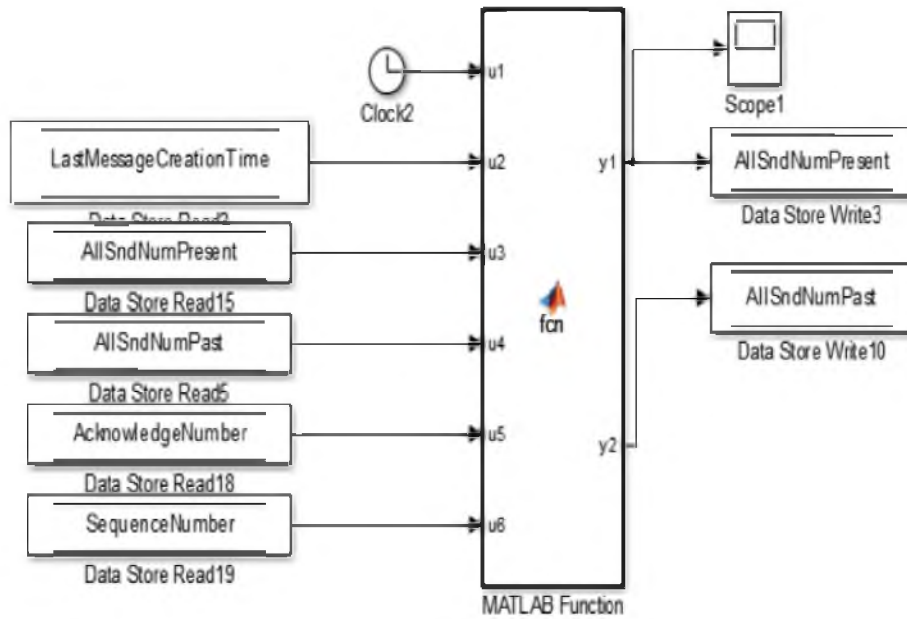
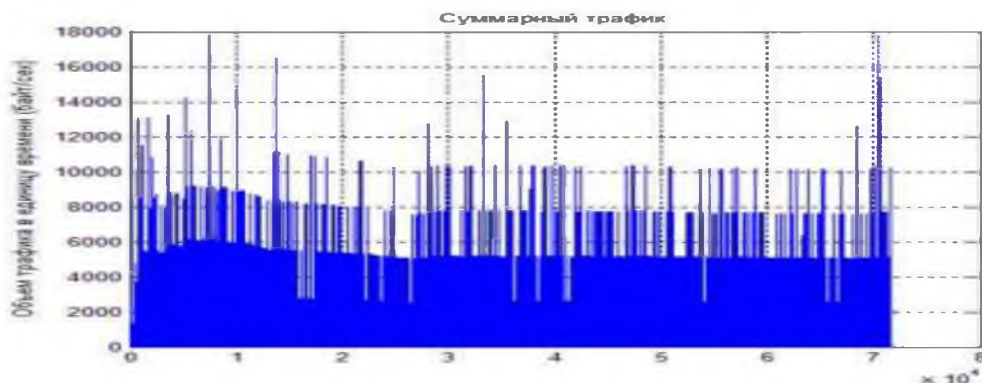


Рисунок 40. Блок розрахунку значень таймера RTO

### 2.3 Характеристики моделі

Для первинної оцінки працездатності моделі було проведено моделювання функціонування мережі протягом 30 секунд із змінною інтенсивністю трафіку. Для кожного джерела повідомлень встановлені однакові параметри, тобто довжина пакетів в діапазоні, що відповідає протоколу TCP/IP від 100 до 1500 байт і швидкість генерації пакетів 50, 100, 150 і 200 (пакетів/с) відповідно для малюнків б6а, б6б, б6в, б6г. На малюнку 42 представлена залежність між швидкістю генерування пакетів та сумарним трафіком, що проходить через комутатор. Вигляд графіка пояснюється кінцевим обсягом буфера комутатора, що обмежує пропускну здатність.



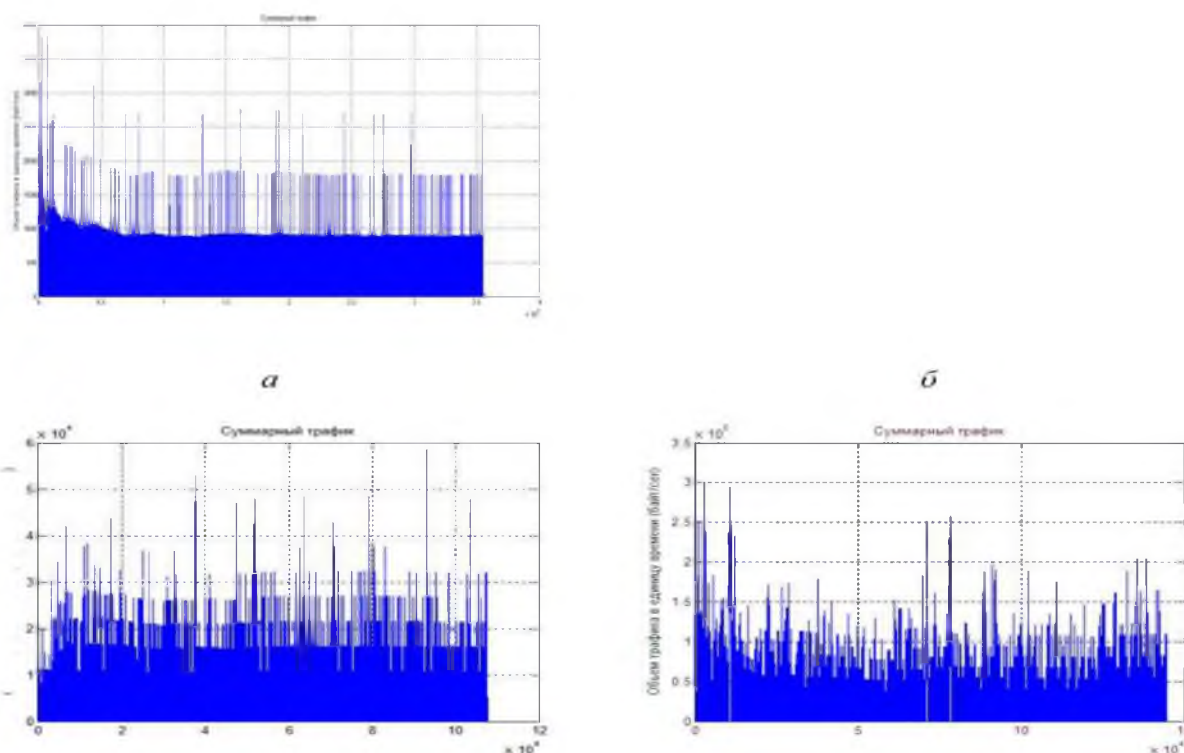
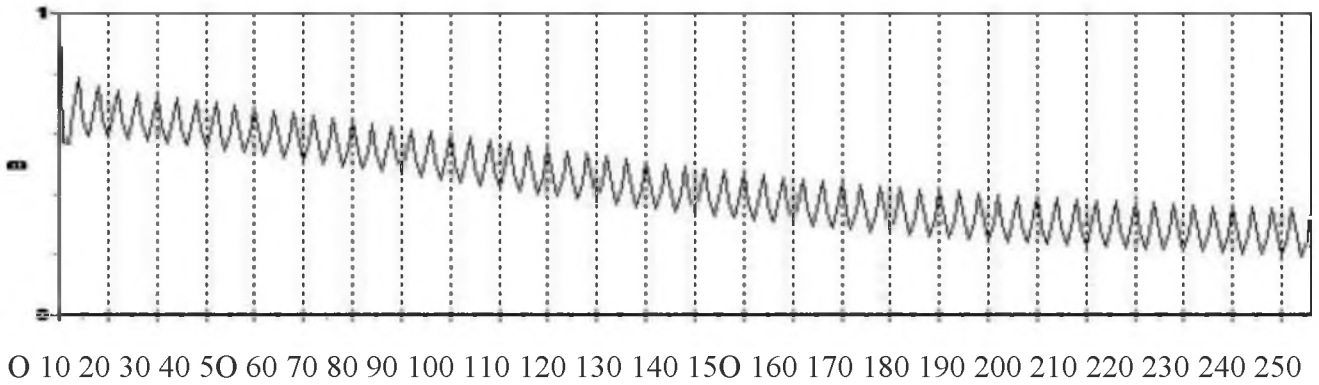


Рисунок 41. Сумарний трафік, що проходить через комутатор в одиницю часу при швидкості генерації повідомлень вузлами 50 пакетів в сек. (а), 100 пакетів у с. (б), 150 пакетів у с. (в), 200 пакетів в сек. (г)



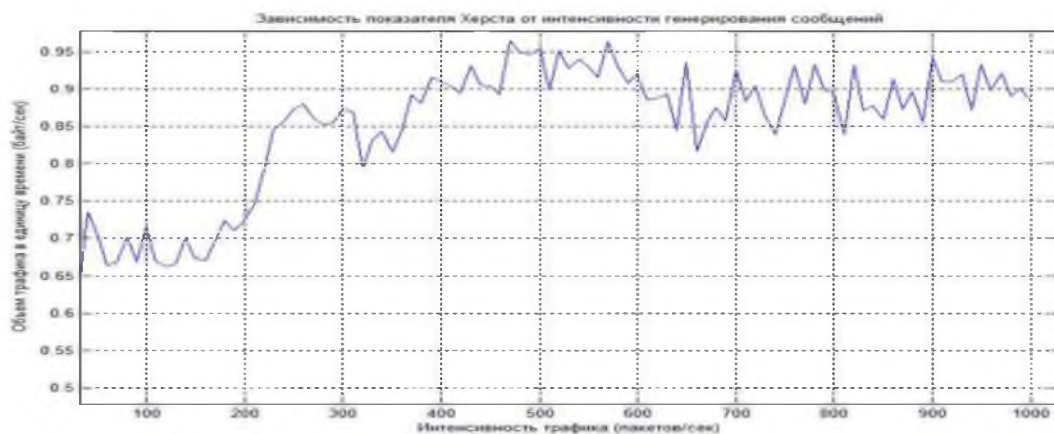
Рисунок 42. Залежність інтенсивності трафіку від швидкості генерування повідомлень.

На користь повільно спадної залежності в процесі формування трафіку говорить повільно загасаючий характер АКФ - рисунок .



Малюнок 43. АКФ трафіку

На малюнку 44 представлено графік зміни показника Херста від інтенсивності трафіку, результати узгоджуються з дослідженнями реального трафіку корпоративної мережі та вказують на пряму залежність. У разі збільшення інтенсивності трафіку показник Херста зростає експоненційно.



Малюнок44. Графік залежності показника Херста від інтенсивності трафіку

Розроблена імітаційна модель дозволяє задавати необхідні параметри джерел повідомлень та комутатора, таким чином стає можливим досягти більшої відповідності між реальною корпоративною мережею та моделлю. У подальших розрахунках використовувалися дані про стан мережі середній розмір пакета та кількість пакетів за секунду, зазначені в роботі (див. таблицю).

За результатами моделювання функція автокореляції агрегованого трафіку представлена малюнку 45. Повільне згасання говорить про довгостроковій залежності.

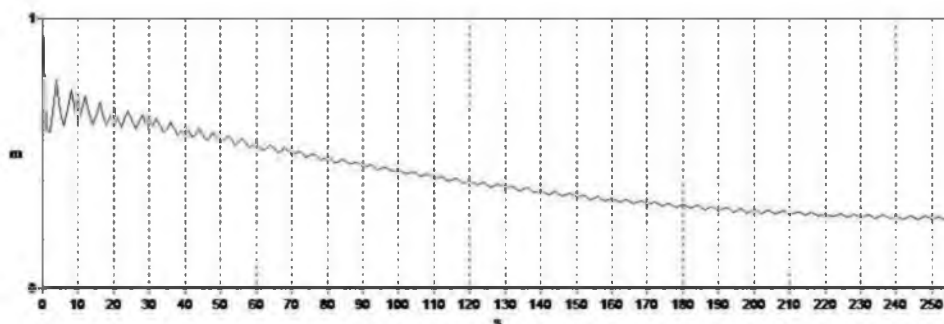
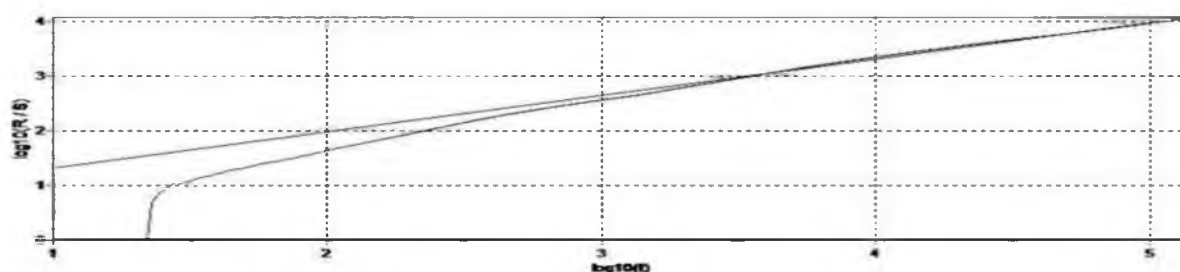


Рисунок 45. АКФ для моделі з параметрами реальної мережі

Для оцінки самоподібності процесу було обчислено показник Херста, який становив  $H=0.86$ , що говорить про самоподібність отриманих даних.



Малюнок 46. АКФ для моделі з параметрами реальної мережі

## 2.4 Оцінка ефективності моделі

Раніше нами було розглянуто різні алгоритми прогнозування рядів та виконано порівняльний аналіз застосування деяких математичних моделей до завдання прогнозування досліджуваних процесів. На основі реальних експериментальних даних було зроблено висновок щодо ефективності моделі ARFIMA. Тепер необхідно визначити, чи дозволяє використання методики короткострокового прогнозування обсягу трафіку досягти підвищення корисної пропускної спроможності мережі.

Як зазначалося вище, повне моделювання мережі передачі повідомлень за протоколом TCP - дуже нетривіальна технічна задача, що виходить за рамки поточної роботи, а розробка TCP з прогнозуванням заслуговує на окремі глибокі дослідження. На даному етапі обмежимося висновком про застосування ARFIMA в мережах з комутацією пакетів і реалізуємо один із можливих алгоритмів управління навантаженнями.



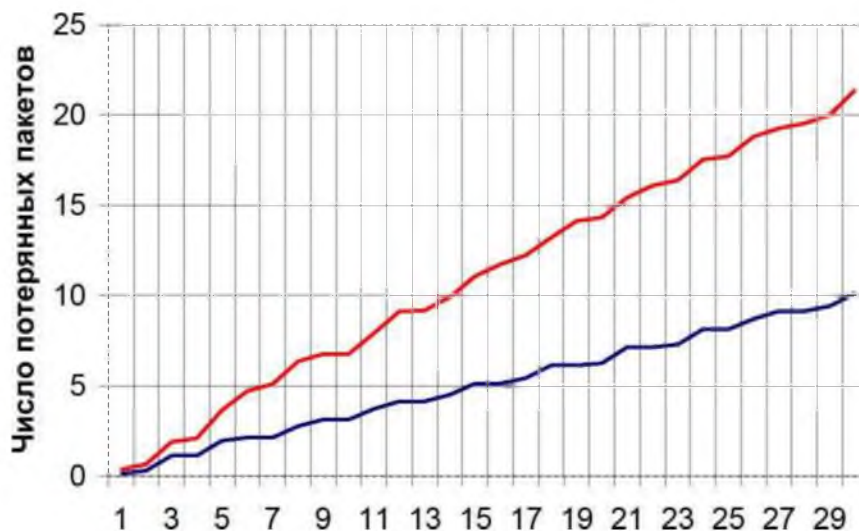
Додамо в розроблену модель зворотний зв'язок між пристроєм, що комує, і джерелами повідомлень для управління розміром вікна передачі. Додамо до моделі комутатора функцію підрахунку

Число унікальних джерел пакетів. В результаті використовуємо наступне вираз для розрахунку вікна передачі  $Win_i$  в момент часу  $i$  ( $0 < i < T$ ),  $T$  - повний час моделювання:

$$Win_i = \text{Min}[X_i/N, CW], i=0,1,\dots,T, \quad (2.3)$$

де  $N$  – число унікальних джерел пакетів,  $X_i$  – прогнозований доступний вільний обсяг буфера комутатора,  $CW$  – розмір вікна передачі, розрахований згідно з алгоритмом повільного старту.

Тепер змодельємо роботу мережі за неминучого виникнення навантаження, оцінимо втрати пакетів під час передачі і корисну пропускну спроможність.

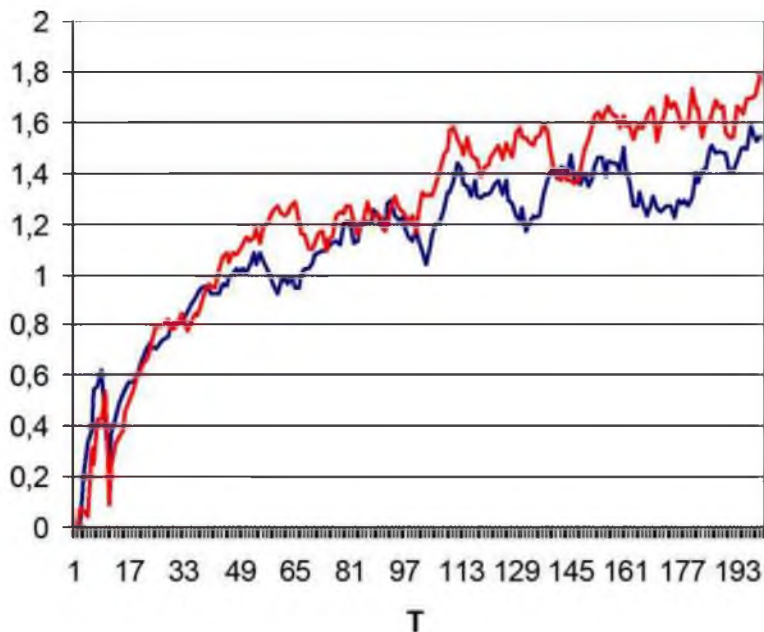


- Без зворотного зв'язку— Зі зворотним зв'язком і прогнозуванням  $T$

Рисунок 47. Залежність втрат пакетів від часу

Застосування механізму зворотного зв'язку, повідомляє джерелам повідомлень про завантаженість буфера, дозволяє знизити втрати пакетів в середньому від 9% до 12%, результат варіюється в залежності від параметрів джерела. Корисна пропускну здатність розраховувалася як відношення

максимальної заданої пропускної здатності (параметр моделі, що визначає час передачі повідомлення) до поточної реальної пропускної здатності. Застосування зворотного зв'язку між комутатором та джерелом повідомлень з урахуванням прогнозування дозволило підвищити корисну пропуску здатність на величину від 12 до 17%, при цьому значення також залежить від параметрів джерел повідомлень (Малюнок 48).



Без зворотного зв'язку

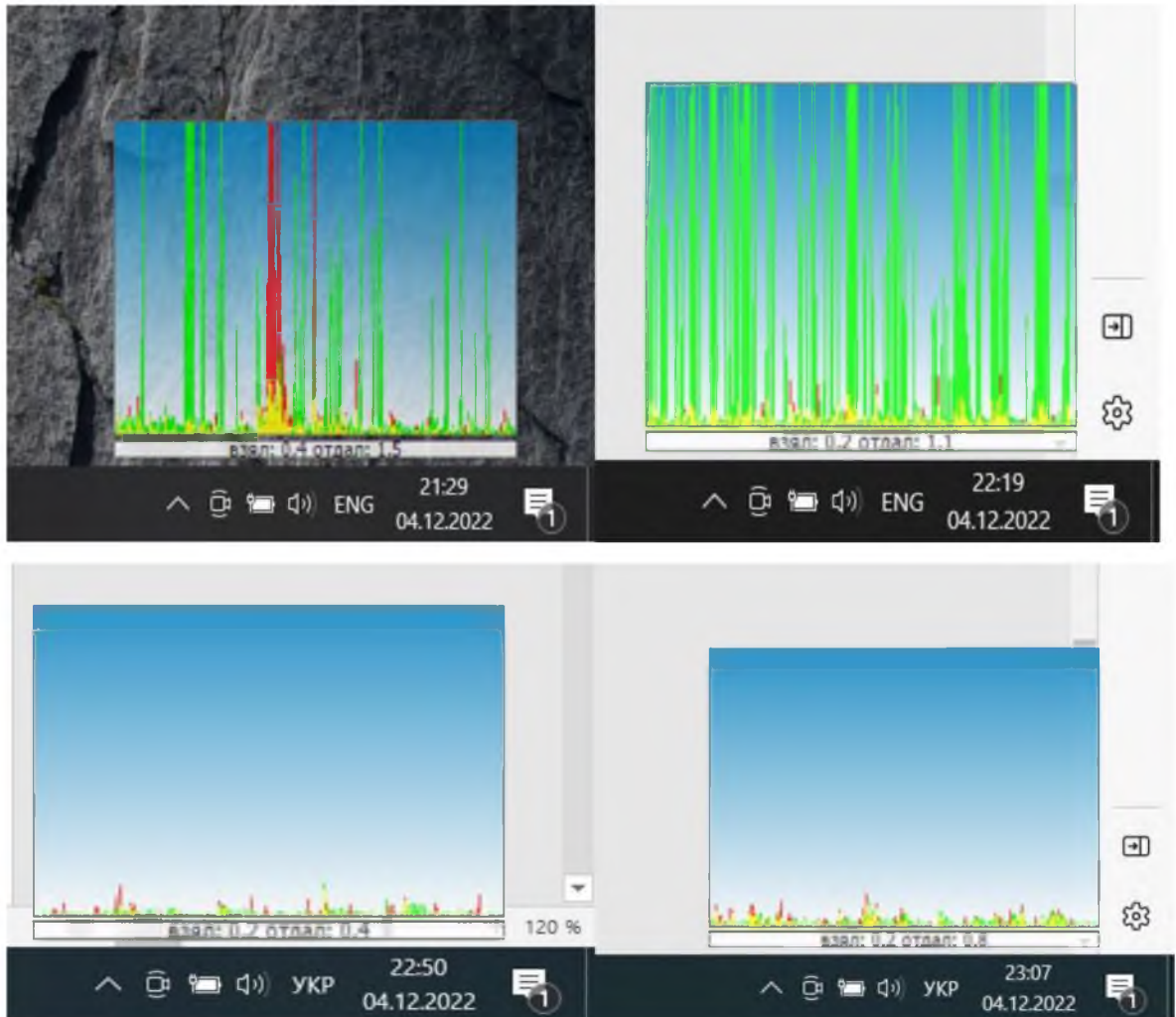
Зі зворотним зв'язком та прогнозуванням

Рисунок 48. Залежність корисної пропускної спроможності мережі від часу

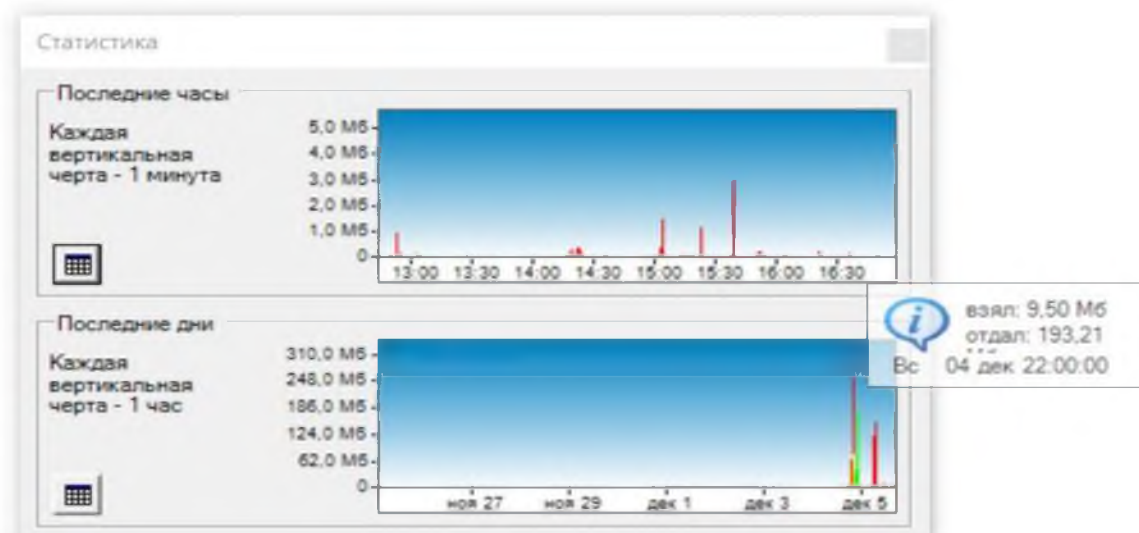
### Реальні показники трафіку одиночного користувача за допомогою програми BitMeter 2

*BitMeter 2* — це вимірник пропускної здатності, який дозволяє візуально контролювати ваше підключення до Інтернету, відображаючи прокручуваний графік, який показує вашу швидкість завантаження та завантаження з часом. BitMeter є безкоштовною програмою, тобто. він надається абсолютно безкоштовно, немає випробувального періоду, немає обмежень за часом функцій, Різні кольори на графіку відображають вашу швидкість завантаження

та завантаження, дисплей оновлюється щосекунди, а поточні швидкості також відображаються числово в нижній частині вікна.



Мал.49 Загальний вид пропускної здатності



Статистика

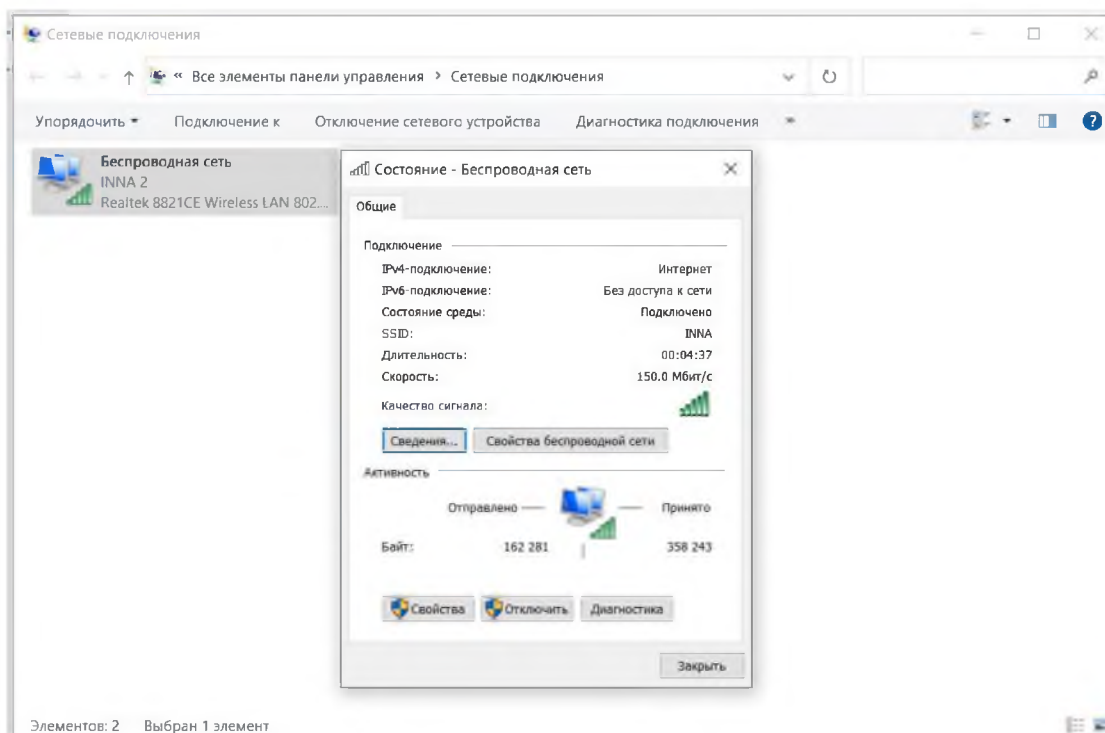
Общее Часы Дни Месяц

время	Дата	Скачано	Отдано	Сумма	Сред.скачка	Сред.отдача
14:00-15:00	Сейчас	1,78 Мб	800,59 Кб	2,56 Мб	517 байт/с	227 байт/с
13:00-14:00	Сейчас	203,56 Кб	228,92 Кб	432,48 Кб	57 байт/с	65 байт/с
12:00-13:00	Сейчас	3,53 Мб	850,36 Кб	4,36 Мб	1,00 Кб/с	241 байт/с
11:00-12:00	Сейчас	8,24 Мб	1,31 Мб	9,55 Мб	2,34 Кб/с	380 байт/с
10:00-11:00	Сейчас	272,81 Кб	139,80 Кб	412,61 Кб	77 байт/с	39 байт/с
9:00-10:00	Сейчас	0	0	0	0/с	0/с
8:00-9:00	Сейчас	0	0	0	0/с	0/с
7:00-8:00	Сейчас	163,96 Мб	3,45 Мб	167,41 Мб	46,64 Кб/с	1006 байт/с
6:00-7:00	Сейчас	128,06 Мб	3,87 Мб	131,94 Мб	36,43 Кб/с	1,10 Кб/с
5:00-6:00	Сейчас	0	0	0	0/с	0/с
4:00-5:00	Сейчас	0	0	0	0/с	0/с
3:00-4:00	Сейчас	0	0	0	0/с	0/с
2:00-3:00	Сейчас	0	0	0	0/с	0/с
1:00-2:00	Сейчас	0	0	0	0/с	0/с
0:00-1:00	Сейчас	0	0	0	0/с	0/с
23:00-0:00	04.12.2022	1,21 Мб	1,27 Мб	2,48 Мб	353 байт/с	370 байт/с
22:00-23:00	04.12.2022	9,50 Мб	193,21 Мб	202,71 Мб	2,70 Кб/с	54,96 Кб/с
21:00-22:00	04.12.2022	11,82 Мб	44,72 Мб	56,53 Мб	3,36 Кб/с	12,72 Кб/с
20:00-21:00	04.12.2022	658,60 Мб	85,05 Мб	743,65 Мб	187,33 Кб/с	24,19 Кб/с
19:00-20:00	04.12.2022	67,88 Мб	4,97 Мб	72,85 Мб	19,31 Кб/с	1,41 Кб/с
18:00-19:00	04.12.2022	4,58 Мб	528,25 Кб	5,10 Мб	1,30 Кб/с	150 байт/с

OK

Мал.50 Загальна статистика

**Реальні показники трафіку одиночного користувача за допомогою програм вбудованих в операційну систему Windows 10**



← Параметры

🏠 INNA

данных в этой сети

### Параметры IP

Назначение IP: Автоматически (DHCP)

### Свойства

SSID:	INNA
Протокол:	Wi-Fi 4 (802.11n)
Тип безопасности:	WPA2-Personal
Диапазон сети:	2,4 ГГц
Канал сети:	11
Скорость линии (прием и передача):	150/150 (Mbps)
Локальный IPv6-адрес канала:	fe80::cc3d:f1ff:7732:88ed%7
IPv4-адрес:	192.168.0.101
DNS-серверы IPv4:	192.168.0.1
Изготовитель:	Realtek Semiconductor Corp.
Описание:	Realtek 8821CE Wireless LAN 802.11ac PCI-E NIC
Версия драйвера:	2024.0.10.135
Физический адрес (MAC):	C0-E4-34-25-C8-F9



Мал.51 Загальна статистика показників трафіку

ТАБЛИЦА 8

14:00-15:00	Сейчас	1,78 Мб	800,59 Кб	2,56 Мб	517	
байт/с	227 байт/с					
13:00-14:00	Сейчас	203,56 Кб	228,92 Кб	432,48 Кб	57 байт/с	
65 байт/с						
12:00-13:00	Сейчас	3,53 Мб	850,36 Кб	4,36 Мб	1,00 Кб/с	
241 байт/с						
11:00-12:00	Сейчас	8,24 Мб	1,31 Мб	9,55 Мб	2,34 Кб/с	
380 байт/с						
10:00-11:00	Сейчас	272,81 Кб	139,80 Кб	412,61 Кб	77 байт/с	
39 байт/с						
9:00-10:00	Сейчас	0	0	0	0/с	0/с
8:00-9:00	Сейчас	0	0	0	0/с	0/с
7:00-8:00	Сейчас	163,96 Мб	3,45 Мб	167,41 Мб	46,64	Кб/с
1006 байт/с						
6:00-7:00	Сейчас	128,06 Мб	3,87 Мб	131,94 Мб	36,43	Кб/с
1,10 Кб/с						
5:00-6:00	Сейчас	0	0	0	0/с	0/с
4:00-5:00	Сейчас	0	0	0	0/с	0/с
3:00-4:00	Сейчас	0	0	0	0/с	0/с
2:00-3:00	Сейчас	0	0	0	0/с	0/с
1:00-2:00	Сейчас	0	0	0	0/с	0/с
0:00-1:00	Сейчас	0	0	0	0/с	0/с
23:00-0:00	04.12.2022	1,21 Мб	1,27 Мб	2,48 Мб	353 байт/с	370
байт/с						
22:00-23:00	04.12.2022	9,50 Мб	193,21 Мб	202,71 Мб	2,70 Кб/с	
54,96 Кб/с						
21:00-22:00	04.12.2022	11,82 Мб	44,72 Мб	56,53 Мб	3,36 Кб/с	
12,72 Кб/с						
20:00-21:00	04.12.2022	658,60 Мб	85,05 Мб	743,65 Мб	187,33	

<b>Кб/с 24,19 Кб/с</b>						
<b>19:00-20:00</b>	<b>04.12.2022</b>	<b>67,88 Мб</b>	<b>4,97 Мб</b>	<b>72,85 Мб</b>	<b>19,31</b>	
<b>Кб/с 1,41 Кб/с</b>						
<b>18:00-19:00</b>	<b>04.12.2022</b>	<b>4,58 Мб</b>	<b>528,25 Кб</b>	<b>5,10 Мб</b>	<b>1,30 Кб/с</b>	
<b>150 байт/с</b>						

## 2.5 Висновки

1) Для оцінки застосування алгоритмів прогнозування в задачі управління перевантаженнями була розроблена модель комп'ютерної мережі з комутацією пакетів.

2) Реалізовано основні алгоритми забезпечення QoS, закладені у TCP.

3) Адекватність моделі підтверджено відповідністю її основних статистичних характеристик реальному експерименту. Результати розрахунку нелінійно – динамічних властивостей змодельованого трафіку також відповідають реальному процесу передачі даних.

4) На рівні комутованого пристрою реалізована функція прогнозування завантаженості буфера та оповіщення джерел з подальшим керуванням вікном передачі.

5) Зроблено висновок про застосування методів прогнозування в задачі управління навантаженнями мережі та наведено кількісна оцінка ефективності алгоритму.

### 3 ЕКОНОМІЧНА ЧАСТИНА

В цьому проекті досліджено фрактальні моделі мережевого трафіку для управління передачею на основі короткострокового прогнозування

У економічному розділі розраховуються одноразові капітальні витрати на розробку та дослідження моделей цієї системи.

#### 3.1 Розрахунок капітальних витрат на розробку та дослідження фрактальних моделей мережевого трафіку для управління передачею на основі короткострокового прогнозування

##### 3.1.1 Визначення трудомісткості розробки моделі

Трудомісткість створення моделі визначається тривалістю кожної робочої операції, починаючи зі складання технічного завдання й закінчуючи оформленням документації (за умови роботи одного проектувальника):

$$t = tmз + tв + та + tnp1 + tnp2 + tonp + t∂ \text{ [год]}. \quad (3.1)$$

де:  $tmз$  – тривалість складання технічного завдання на впровадження методів;

$tв$  – тривалість вивчення технічного завдання (ТЗ) та літературних джерел за темою;

$та$  – тривалість розробки та дослідження моделей;

$tnp1$  – тривалість модулювання мережевого трафіку для управління передачею на основі короткострокового прогнозування ;

$tonp$  – тривалість опрацювання здобутих характеристик;

$t∂$  – тривалість підготовки технічної документації.

Вихідні дані для визначення трудомісткості на розробку та дослідження фрактальних моделей мережевого трафіку для управління передачею на основі короткострокового прогнозування приведені в таблиці 3.1.



Таблиця 3.1 – Тривалість розробки та дослідження

$t_{mз}$ , год	$t_{в}$ , год	$t_{a}$ , год	$t_{np 1}$ , год	$t_{onp}$ , год	$t_{\partial}$ , год
45	48	69	19	19	29

Розрахуємо трудомісткість розробки моделі за формулою (3.1):

$$t = 45+48+69+19+19+29=229 \text{ [год]}.$$

### 3.1.2 Розрахунок витрат на розробку моделі

Витрати на розробку моделі  $K_{ПЗ}$  складаються з витрат на заробітну платню розробника  $З_{ЗП}$  і вартості витрат машинного часу, що необхідний для опрацювання та дослідження фрактальних моделей мережевого трафіку для управління передачею на ПК  $З_{МЧ}$ .

$$K_{ПЗ} = З_{ЗП} + З_{МЧ} \text{ [грн]} \quad (3.2)$$

Заробітна плата виконавця враховує основну і додаткову заробітну плату, а також відрахування на соціальні потреби (пенсійне страхування, страхування на випадок безробіття, соціальне страхування тощо) і визначається за формулою:

$$З_{ЗП} = t \cdot З_{ПР} \text{ [грн]}. \quad (3.3)$$

де:  $t$  – трудомісткість створення моделі;

$З_{ПР}$  дорівнює 120 грн/год.

Розрахуємо заробітну платню проектувальника за формулою (3.3):

$$З_{ЗП} = 229 \cdot 120 = 27480,00 \text{ [грн]}.$$

Вартість машинного часу на ПК визначається за формулою:

$$З_{МЧ} = \left( t_a + t_{np 1} + t_{np 2} + t_{onp} + t_{\partial} \right) \cdot C_{мч} \text{ [грн]}. \quad (3.4)$$

де:  $C_{мч}$  – вартість 1 години машинного часу ПК, грн/година.

Вартість 1 години машинного часу ПК визначається за формулою:

$$C_{мч} = P_e \cdot t \cdot C_e + \frac{\Phi_{перв} \cdot H_a}{F_p} + \frac{K_{лмз} \cdot H_{анз}}{F_p} \text{ [грн/год]}, \quad (3.5)$$

де:  $P_e$  – встановлена потужність ПК;

$t$  – трудомісткість створення моделей;

$C_e$  – енерговитрати;

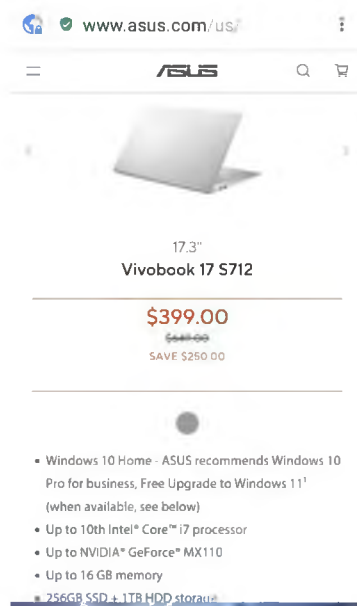
$\Phi_{перв}$  – первісна вартість ПК на початок року;

$H_a$  – річна норма амортизації на ПК;

$K_{лмз}$  – вартість ліцензійного програмного забезпечення;

$H_{анз}$  – річна норма амортизації на ліцензійне програмне забезпечення;

$F_p$  – річний фонд робочого часу (за 40-годинного робочого тижня).



Малюнок 3.1 ПК для розробки

Вартість ПК згідно курсу \$ дорівнює:  $399 \cdot 37,4531 = 14943.7869$  грн.

Енерговитрати розраховуються за формулою:

$$C_e = P_e \cdot C_{кВт} \text{ [грн/год]}, \quad (3.6)$$

де:  $C_{кВт}$  - тариф на електричну енергію. (приведено в таблиці 3.2)

Таблиця 3.2 – Тарифи на електричну енергію станом на 1.01.2022 року по областям для підприємств

Постачальник (енергорозподільча компанія)	Тариф (грн. за 1 МВт·год., без ПДВ)	
	1 клас напруги	2 клас напруги
Вінницька обл.		
ПАТ "Вінницяобленерго"	199,47	1257,26
Волинська обл.		
ПРАТ "Волинь- обленерго"	144,47	1102,19
Дніпропетровська обл.		
ПРАТ "ПЕЕМ "ЦЕК"	101,08	821,57
АТ "ДТЕК Дніпровські електромережі"	114,90	783,32
Донецька обл.		
АТ "ДТЕК Донецькі електромережі"	260,92	1286,28
Житомирська обл.		
АТ "Житомиробленерго"	247,28	1333,43
Закарпатська обл.		
ПРАТ "Закарпаття- обленерго"	326,79	1480,39

Розрахунок витрат на розробку та дослідження моделей мережевого трафіку для управління передачею, зводимо в таблицю 3.3

Таблиця 3.3 – Розрахунок витрат на розробку та дослідження

$P_e$ , кВт	$C_{кВт}$ кВт·год	$\Phi_{перв}$ , грн	$N_a$ , частка одиниці	$K_{лз}$ , грн	$N_{анз}$ , частка одиниці	$F_p$ , год
0,15	1.4804	14943.8	0,4	53402.4	0,4	1920,00

Тоді за формулою (3.6) отримаємо розмір енерговитрат:

$$C_e = 0,15 \cdot 1.4804 = 0.2221 \text{ [грн/год]}.$$

Річна норма амортизації, якщо використовується метод прискорення зменшеної вартості, визначається за формулою:

$$N_a = 2/T \cdot 100\% \quad (3.7)$$

де  $T$  – строк корисного використання ПК, дорівнює 5 років.

Розрахуємо річну норму амортизації за формулою (3.7):

$$H_a = 2/5 \cdot 100\% = 40\% = 0,40 \text{ [частки одиниці]}.$$

Строк корисного використання ліцензійного програмування дорівнює 5 років.

Річна норма амортизації на ліцензійне програмне забезпечення визначається за формулою (3.7):

$$H_{\text{апіз}} = 2/5 \cdot 100\% = 40\% = 0,40 \text{ [частки одиниці]}.$$

Ліцензійне програмне забезпечення, яке використовується в даному випадку представлено на мал. 3.1

The image shows a promotional offer for a software bundle. On the left, a box displays the bundle: 'Комплект Windows 10 Pro + Office 2021 + WinRAR + Avast'. The price is shown as ~~3700 грн.~~ and 2800 грн. Below this is a 'Додати в кошик' (Add to cart) button. On the right, a 'Pricing and Licensing' section for 'Simulink' is visible. It offers two options: 'Perpetual license' for EUR 3.150 and 'Annual license' for EUR 1.260. A 'Buy now' button is present. Below the pricing, it lists 'Required Products: MATLAB'. At the bottom, there is a note: 'Price applies for purchase and use in Netherlands. For pricing in other regions contact sales. Pricing excludes TAX/VAT.'

Малюнок 3.1 Промоакції на програмне забезпечення

Microsoft Windows 10 Professional. Його вартість 2800 грн.

Matlab & Simulink коштує 1.260 EUR (40,1606)=1260\*40,1606=50602.357 грн.

BitMeter2 є безкоштовною програмою, тобто. він надається абсолютно безкоштовно,

$$50602.36+2800+14943.8=68346.16 \text{ [грн]}$$

Вартість 1 години машинного часу ПК визначаються за формулою (3.5):

$$C_{мч} = 0,15 \cdot 229 \cdot 1.4804 + \frac{14943.8 \cdot 0,40}{1920} + \frac{53402.4 \cdot 0,40}{1920} = 86.4487 \text{ [грн/год]}$$

Розрахуємо вартість машинного часу за формулою (3.4):

$$Z_{мч} = (69+19+19+29) \cdot 86.4487 = 11757.0232 \text{ [грн]}.$$

Отже, підставивши отримані результати у формулу (3.2), отримаємо величину витрат на розробку моделі:

$$K_{ПЗ} = 27480.00 + 11757.03 = 39237.03 \text{ [грн]}.$$

### 3.1.3 Розрахунок капітальних витрат

Загальні капітальні витрати на розробку визначаються за формулою:

$$KЗ = K_{пз} + K_{навч} + K_n \text{ [грн]}, \quad (3.8)$$

де  $K_{навч}$  - витрати на навчання технічних фахівців і обслуговуючого персоналу;

$K_n$  - Витрати на встановлення обладнання та налагодження системи.

Дані про витрати на розробку моделі зводимо в таблицю 3.3

Таблиця 3.3 – Витрати на розробку моделі

<i>Киз,</i> грн	<i>Кнавч,</i> грн	<i>Кн,</i> грн
39237.03	400,00	6834,6

Отже, капітальні витрати становлять:

$$K3 = 39237.03 + 400 + 6834,6 = 50071.63 \text{ [грн]}.$$

### 3.2 Висновки

В економічному розділі було розраховано:

- 1 Трудомісткість розробки імітаційної моделі – 229 год;
- 2 Заробітна платня проектувальника – 27480,00 грн;
- 3 Витрати на розробку моделі – 39237.03грн;
- 4 Витрати на новий ПК з ліцензійним ПЗ –68346.16 грн;
- 5.Капітальні витрати на дослідження фрактальних моделей мережевого трафіку для управління передачею на основі короткострокового прогнозування- 50071.63 грн.
- 6.Для зменшення капітальних витрат рекомендовано продати після закінчення дослідження, моделювання та проектування ПК Асус зі встановленим ліцензованим програмним забезпеченням за 18274.53грн.

## Висновок

Основні результати дипломної роботи такі:

1) Виконано збір статистичних даних вхідного та вихідного трафіку, а також процесу розподілу апаратних ресурсів одного із фізичних серверів корпоративної мережі; проведено статистичний та динамічний аналіз зібраних даних, виявлено самоподібний та хаотичний характер процесів.

2) Визначено якісну та кількісну залежність між сумарним обсягом трафіку, що проходить через сервер у мережі та розподілом апаратних ресурсів вузла; проведено кореляційний та регресійний аналіз часових рядів.

3) Розроблено та протестовано імітаційну модель мережі в цілому та окремих процесів передачі даних, встановлено адекватність моделі; отримано модифіковану імітаційну модель комп'ютерної мережі з урахуванням короткочасного прогнозування (ARFIMA) завантаження буфера комутатора.

4) Обчислювальний експеримент показав, що запропонована методика на основі механізму зворотного зв'язку, повідомляє джерелам повідомлень про завантаженість буфера, дозволяє знизити втрати пакетів в середньому від 9% до 12%; застосування зворотного зв'язку між комутатором та джерелом повідомлень з урахуванням прогнозування дозволяє підвищити корисну пропускну здатність на величину від 12 до 17% залежно від параметрів джерел повідомлень.

Отримані результати можуть бути використані при вирішенні практичних завдань, що виникають при дослідженні функціонування різних типів мереж передачі даних, а також поліпшення їх техніко-економічних та експлуатаційних характеристик. Розроблені моделі та алгоритми в перспективі можуть бути поширені на вирішення задач виявлення аномалій та загроз у комп'ютерних мережах, у тому числі у комбінації з методами сигнатурного аналізу

## ПЕРЕЛІК ПОСИЛАНЬ

1 Ю.Ю. Громов, Н.А. Земской, О.Г. Иванова, А.В. Лагутин, В.М. Тютюнник. ФРАКТАЛЬНЫЙ АНАЛИЗ И ПРОЦЕССЫ В КОМПЬЮТЕРНЫХ СЕТЯХ. - Т.: Тамбовский государственный технический университет (ТГТУ), 2004,

2 Зембицкая А. С., Колисник Е. В. ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ ТРАФИКА ВЫСОКОСКОРОСТНЫХ СЕТЕЙ. - Институт компьютерных технологий Национального авиационного университета: Проблеми інформатизації та управління, 2(17)'2006

3 Колесников А. В. Моделирование сетевого трафика и алгоритмы борьбы с перегрузками на основе методов нелинейной динамики и краткосрочного прогнозирования временных рядов. - М.: «Московский государственный технический университет им. Н.Э. Баумана», 2015

4. <https://codebox.net/pages/bitmeter2>

5. Людмила Кириченко, Тамара Радивилова ФРАКТАЛЬНЫЙ АНАЛИЗ РЕАЛЬНЫХ ДАННЫХ International Journal "Information Content and Processing", Volume 5, Number 2, © 2018

6. <https://cyberleninka.ru/article/n/svoystva-fraktalnogo-trafika-pri-prohozhdenii-sistemy-massovogo-obsluzhivaniya-s-ocheredyu/viewer>

7. [https://www.researchgate.net/publication/321499373\\_Multifractal\\_Based\\_Network\\_Traffic\\_Modeling](https://www.researchgate.net/publication/321499373_Multifractal_Based_Network_Traffic_Modeling)

8. [https://www.researchgate.net/publication/322560117\\_Analysis\\_of\\_Self-Similar\\_Traffic\\_Models\\_in\\_Computer\\_Networks/citation/](https://www.researchgate.net/publication/322560117_Analysis_of_Self-Similar_Traffic_Models_in_Computer_Networks/citation/)

9. Mark E. Crovella and Azer Bestavros In Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes IEEE/ACM Transactions on Networking Vol 5, Number 6,

10. <https://tokzamer.ru/bez-rubriki/ethernet-principialnaya-shema>

11. <https://geoid.org.ua/skolko-elektroenergii-potreblyayet-noutbuk>



## ДОДАТОК А

## Відомість матеріалів дипломної роботи

№	Формат	Найменування	Кількість листів	Примітки
<i>Документація</i>				
1	A4	Реферат	2	
2	A4	Список умовних скорочень	3	
3	A4	Зміст	2	
4	A4	Вступ	2	
5	A4	Стан питання. Постановка задачі	46	
6	A4	Спеціальна частина	25	
7	A4	Економічний розділ	6	
8	A4	Висновки	1	
9	A4	Перелік посилань	1	
10	A4	Додаток А	1	
11	A4	Додаток Б	1	
12	A4	Додаток В	2	
13	A4	Додаток Г	5	
14		Матеріали дипломної роботи на оптичному носії		Оптичний диск



## Відгук керівника дипломної роботи

ВІДГУК  
на магістерську дипломну роботу

Студента(ки) \_\_\_\_\_ гр.

(прізвище, ім'я)

на тему: \_\_\_\_\_  
\_\_\_\_\_Актуальність теми \_\_\_\_\_  
\_\_\_\_\_Повнота розкриття теми \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_Теоретичний рівень \_\_\_\_\_  
\_\_\_\_\_Практична значущість \_\_\_\_\_  
\_\_\_\_\_Самостійність виконання роботи \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_Якість оформлення, загальна та спеціальна грамотність \_\_\_\_\_  
\_\_\_\_\_Переваги та недоліки роботи \_\_\_\_\_  
\_\_\_\_\_

---

---

---

---

Загальна оцінка роботи та висновок щодо рекомендації до захисту в ДЕК

---

---

---

Науковий керівник

к.ф.-м.н., професор

(посада)

(підпис)

Гусєв О.Ю.

(ініціали, прізвище)

«\_\_\_\_\_» \_\_\_\_\_ 2022 р.

## Програми

### Лістинг функцій пакету MATLAB

Код MATLAB функції побудови прогнозу з урахуванням AR(p) моделі з виведенням графіка залежності помилки прогнозування від порядку p.

```
function [ minRMSE ] = MyARfunc(IniArray,tuneSteps)
maxValue = max(IniArray);
s = size(IniArray);
arraySize = s(1);
normArray = zeros(1, arraySize);
transpArray = transpose(normArray);
for k = 1:arraySize
    transpArray(k) = (IniArray(k)/maxValue)*100;
end
MapeError = zeros(1, tuneSteps);
for i = 1:tuneSteps
    DAT = iddata(transpArray(1:arraySize),[],1);
    arModel=ar(DAT,i,'ls');
    YP=predict(DAT,arModel);
    MaperSum=0;
    data = DAT.OutputData;
    predictData = YP.OutputData;
    for k =1:arraySize
        temp = MaperSum;
        if(predictData(k) ~=0)
            MaperSum =((data(k)-predictData(k)) /data(k));
        end
        MaperSum =temp+MaperSum;
    end
end
```

```

MAPE=(MapeSum/arraySize)*100;
if (MAPE ~= 0)
MapeError(i) =MAPE;
end
minRMSE = min(MapeError);
if(minRMSE==MAPE)
p=i;
end
dif =data-predictData;
sumOfErrors(i) = sum(abs(dif));
mae(i)=sumOfErrors(i)/arraySize;
end
subplot(3,1,1);
plot(Yres);
subplot(3,1,2);
plot(dif);
subplot(3,1,3);
plot(Yres);
end

```

Код MATLAB функції побудови прогнозу з урахуванням ARIMA(p,d,q) моделі з виведенням графіка залежності помилки прогнозування (MAE) від порядку p.

```

function [ MAE ] = MyARfunc(IniArray,forecastSteps,tuneSteps,D,Q)
s = size(IniArray);
arraySize = s(1);
normArray = zeros(arraySize,1);
p = tuneSteps;
%avgValue = mean(IniArray);
for k = 1:arraySize

```

```

%transpArray(k) = (IniArray(k)/avgValue);
normArray(k) = IniArray(k);
end
for x=1:tuneSteps
arModel=arima(x,D,Q);
fittedArima = estimate(arModel,normArray);
Ymdl1pred=zeros(length(arraySize),1);
for i=p+1:length(normArray)-p;
[Ymdl1pred(i)]=forecast(fittedArima,forecastSteps,'Y0',normArray(1:i,1));
end;
tY = transpose(Ymdl1pred);
diff = zeros(size(tY),1);
diffSum =0;
for l=1:size(tY)
diff(l)=abs(normArray(l)-tY(l));
diffSum = diffSum + abs(diff(l))
end
MAE(x) =diffSum/l;
end
subplot(3,1,1);
plot(normArray);
subplot(3,1,2);
plot(Ymdl1pred);
subplot(3,1,3);
%plot(diff);
plot(MAE);

```

## Лістинг скриптів пакета R

**Код R скрипта побудови прогнозу з урахуванням ARIMA(p,d,q) моделі.**

```
remove(arimaNv);
library(forecast);
remove(newArrayArima);
remove(fit);
remove(ME);
remove(MAE);
remove(diffArray);
newArrayArima <-vector();
window <- 50;
initialIndex <- 500;
data<-CPU;
loopSize <- 50;
for (i in 1:loopSize)
{
from<-initialIndex+i;
to<-initialIndex+window+i;
fit <- arima(data[from:to], order=c(2,1,1));
arimaNv <- forecast(fit,h=1);
newArrayArima[i]<-as.numeric(arimaNv$mean[1]);
}
from<-initialIndex+window+2;
to<-initialIndex+window+loopSize+1;
realData <-data[from:to];
diffArray=(realData-newArrayArima);
ME=mean(diffArray);
remove(diffArray);
```



```

diffArray=(realData-newArrayArima);
MAE=mean(diffArray)/mean(realData)*100;
remove(diffArray);
diffArray=((realData-newArrayArima)/realData)^2;
RMSE=sqrt(sum(diffArray)/length(diffArray));
plot(realData,type="b");lines(newArrayArima,col="green",type="b");

```

**Код R скрипта побудови прогнозу з урахуванням ARFIMA(p,d,q) моделі.**

```

remove(farimaNv);
library(arfima);
library(forecast);
remove(newArrayFarima);
remove(fit);
remove(ME);
remove(MAE);
remove(diffArray);
newArrayFarima <-vector();
window <- 50;
initialIndex <-1000;
data<-(CPU);
loopSize <- 300;
for (i in 1:loopSize)
{
from<-initialIndex+i;
to<-initialIndex+window+i;
fit <- arfima(data[from:to],drange=c(0, 1));
farimaNv <- forecast(fit,h=1);
newArrayFarima[i]<-as.numeric(farimaNv$mean[1]);
}

```

```
from<-initialIndex+window+2;
to<-initialIndex+window+loopSize+1;
realData <-data[from:to];
diffArray=(realData-newArrayFarima);
ME=mean(diffArray);
remove(diffArray);
diffArray=abs(realData-newArrayFarima);
MAE=mean(diffArray)/mean(realData)* 100;
remove(diffArray);
diffArray=((realData-newArrayFarima)/realData)^2;
RMSE=sqrt(sum(diffArray)/length(diffArray));
plot(realData,type="b");lines(newArrayFarima,col="green",type="b");
```