

УДК (UDC) 004.942

**Стрілець  
Вікторія Євгенівна**

*к.т.н., доцент кафедри теоретичної та прикладної системотехніки  
Харківський національний університет імені В. Н. Каразіна, майдан  
Свободи, 6, м. Харків, 61022  
e-mail: viktoria.strilets@karazin.ua  
<https://orcid.org/0000-0002-2475-1496>*

**Дорошенко  
Максим Ігорович**

*студент факультету комп'ютерних наук  
Харківський національний університет імені В. Н. Каразіна, майдан  
Свободи, 6, м. Харків, 61022  
e-mail: xa11867769@student.karazin.ua  
<https://orcid.org/0000-0003-3645-252X>*

## Аналіз і прогнозування характеристик комп'ютерної мережі

Більшість застосунків та пристроїв сьогодні працюють у взаємодії з мережами передачі даних, тому важливим є розробка й удосконалення технологій і методів для кращого розуміння, контролю, управління або прогнозування поведінки і стану комп'ютерних мереж та їх характеристик. Тому задачі, пов'язані з розробкою моделей та методів оцінки й прогнозування параметрів трафіку комп'ютерної мережі, є важливими для управління комп'ютерними мережами. Метою роботи була розробка моделі виявлення трендів характеристик комп'ютерних мереж та їх подальше оцінювання для підвищення якості прогнозування. У роботі запропоновано розглядати трафік комп'ютерних мереж з точки зору часових рядів. Розглянуті та проаналізовані моделі тренду часових рядів і критерії виявлення та методи оцінки тренду. Обраний метод оцінки тренду трафіку на основі тесту Манн-Кендалла, для інтерпретації результатів використаний метод консенсусу. Також була розглянута задача прогнозування трафіку комп'ютерної мережі із врахуванням показників тренду. Дана задача була успішно розв'язана з використанням моделі прогнозування на основі ковзного середнього, та удосконалена із застосуванням градієнтного бустингу. Окремою задачею були збір та попередня обробка набору вхідних даних, який описує роботу комп'ютерної мережі, його формалізація й подальший кількісний та якісний аналіз. Був створений унікальний набір даних, сформований шляхом парсингу логів (системних файлів) із даних моніторингу стану трафіку комп'ютерної мережі. Саме цей набір даних був використаний для створення моделі виявлення тренду і подальшого прогнозування характеристик комп'ютерної мережі. Отримані результати показали, що розроблені моделі і методи можна використовувати у практичній діяльності при розв'язанні задач моніторингу і управління комп'ютерними мережами.

**Ключові слова:** комп'ютерна мережа, трафік, аналіз тренду, задача прогнозування, машинне навчання.

**Як цитувати:** Стрілець В. Є., Дорошенко М. І. Аналіз і прогнозування характеристик комп'ютерних мереж. *Вісник Харківського національного університету імені В.Н.Каразіна, сер. «Математичне моделювання. Інформаційні технології. Автоматизовані системи управління».* 2022. вип. 55. С.49-57. <https://doi.org/10.26565/2304-6201-2022-55-05>

**How to quote:** V.Y. Strilets, and M.I. Doroshenko, "Analysis and forecasting of computer network characteristics." *Bulletin of V.N. Karazin Kharkiv National University, series "Mathematical modelling. Information technology. Automated control systems,* vol. 55, pp. 49-57, 2022. <https://doi.org/10.26565/2304-6201-2022-55-05>

### 1 Вступ

Останнє десятиліття можна охарактеризувати швидким зростанням обсягів трафіку комп'ютерних мереж. Така тенденція пов'язана зі стрімким розвитком таких комп'ютерних технологій, як хмарні обчислення, хмарні сховища, «потоківі» сервіси для фільмів, музики чи ігор. Все це змушує інтернет-провайдерів вирішувати очевидне завдання збільшення та модернізації існуючих комп'ютерних мереж. Основним завданням при модернізації комп'ютерної мережі є аналіз мережевого трафіку, а саме його структури та обсягу. Саме для розв'язання цієї задачі пропонується застосувати методи виявлення трендів трафіку комп'ютерних мереж і розробити модель прогнозування показників трафіку комп'ютерної мережі.

В сучасній теорії аналізу даних і статистики існує багато різних методів прогнозування. Значна їх частина відноситься до прогнозування часових рядів. Особливістю прогнозування часових рядів є те, що аналізуються лише дані спостережень без додаткової інформації, без аналізу впливу зовнішніх сил. Звичайно, такий аналіз виглядає досить неповним, але доволі часто прогнози

часових рядів є точними. Набір методів аналізу досить широкий. Деякі методи є універсальними, інші – спеціалізованими та вимагають подальшого розроблення й апробування. Велика різноманітність наявних методів, іноді недостатня обізнаність фахівців про особливості використання тих чи інших методів, складність застосування математичного апарату створюють для аналітиків труднощі та навіть можуть призводити до формулювання неправильних висновків [1, 2].

Розв’язання будь-якої задачі з аналізу і прогнозуванню часового ряду починається з побудови графіка досліджуваного показника. На етапі графічного аналізу можна досліджувати компонентний склад часового ряду, а також зробити перші кроки до вибору моделі для опису динаміки і подальшого прогнозування. Під час аналізу велику увагу приділяють виявленню тенденцій або трендів. Це допомагає підвищити якість прогнозування результату деякого явища часового ряду [3].

На сьогоднішній день існують різні способи для моніторингу та аналізу мережевого трафіку. Основні з них:

- за допомогою програм-аналізаторів (у тому числі з допомогою спеціальних протоколів у маршрутизаторах);
- статистичні методи;
- методи на основі нейронних мереж.

Всі ці методи мають свої переваги і недоліки і застосовуються в залежності від цілей.

## 2 Постановка задачі

Трафік комп’ютерних мереж можна розглядати як часовий ряд. Тоді задачу аналізу мережевого трафіку можна представити як пошук закономірностей в часових рядах, формулювання гіпотез і доведення їх статистичної значущості.

Метою дослідження є вдосконалення моделі прогнозування характеристик комп’ютерної мережі з врахуванням параметрів тренду.

Для досягнення мети необхідно вирішити такі задачі:

- формування та формалізація вхідних даних для побудови моделі прогнозування, а також їх підготовка та обробка для покращення ефективності методу.
- розробка (або модифікація існуючого) методу виявлення тренду даних.
- розробка моделі прогнозування часового ряду з використанням сучасних засобів моделювання.

покращення моделі за допомогою методів машинного навчання, а також порівняння ефективності моделей до та після застосування обраного методу машинного навчання.

## 3 Аналіз тренду мережевого трафіку

Виділення тренду мережевого трафіку є дуже важким і водночас дуже важливим завданням, оскільки його вирішення дозволяє здійснювати прогноз, у якому випадкова складова часового ряду використовується з метою оцінки точності прогнозу стану трафіку. В даний час найчастіше використовуються два методи виділення тренду [4].

Перший метод полягає в тому, що за емпіричними даними часового ряду підбирається крива (математична модель), яка з найбільшою точністю описує часовий ряд. При цьому в якості математичних моделей використовуються різні функції: рівняння прямої та експоненти, парабола (квадратична, кубічна і більш високих ступенів), логістична крива, крива Гомперца та ін.

Другий метод виділення тренду полягає в згладжуванні ряду за методом ковзного середнього. При цьому зазвичай знаходять середнє значення трьох (або п’яти) перших членів, далі беруться наступні три члени зі зміщенням на одиницю і середнє. Таким чином, вдається зменшити випадкову складову.

При аналізі часового ряду є доцільним виявити трендову складову. Для цього існують критерії виявлення трендів:

- критерій Аббе-Лінника;
- критерій Кокса-Стюарта;
- критерій Фостера-Стюарта;

– критерій Манн-Кендала.

Кожен з цих критеріїв використовується індивідуально для розглядуваного часового ряду, в залежності від властивостей даних, на яких цей критерій буде застосовуватись.

Для оцінки наявності трендової компоненти в роботі було використано метод Манн-Кендала.

Статистичний тест Манна-Кендалла (МК) на тенденцію використовується для оцінки того, чи набір значень даних збільшується з часом чи зменшується з часом, і чи є тенденція в будь-якому напрямку статистично значущою. МК не оцінює величину зміни [6].

Тест можна використовувати для індикаторів із різними одиницями вимірювання та періодами часу та не потребує довірчих інтервалів (які доступні не для всіх індикаторів). Крім того, перевірку можна виконати, навіть якщо в наборі відсутні значення.

Щоб обчислити тестову статистику  $S$ , треба порівнювати кожне значення з усіма наступними значеннями періоду часу для індикатора. Для кожної пари порівняння присвоюється оцінка «+1», якщо останнє значення перевищує перше.

Якщо останнє значення нижче за перше, то порівнянню присвоюється оцінка «-1». Потім усі бали підсумовуються для обчислення тестової статистики  $S$ . Позитивне значення  $S$  означає, що тенденція зростає, а від'ємне значення  $S$  означає, що тенденція зменшується.

Тест МК перевіряє, чи слід відхилити нульову гіпотезу ( $H_0$ ) і прийняти альтернативну гіпотезу ( $H_a$ ), де

- $H_0$ : немає монотонної тенденції;
- $H_a$ : монотонна тенденція присутня.

Початкове припущення тесту МК полягає в тому, що дані є правдивими та що дані мають бути переконливими поза розумним сумнівом, перш ніж вони будуть відхилені та прийняті [6, 7]. Регресійний аналіз вимагає, щоб залишки від підігнаної лінії регресії були нормально розподілені, припущення, яке не вимагається тестом Манн-Кендалла, тобто є непараметричним (без розподілу) тестом.

Обчислення  $S$  статистики відбувається за формулою:

$$S = \sum_{k=1}^{N-1} \sum_{l=k+1}^N \text{sign}(x_l - x_k).$$

Далі обчислюється  $\text{VAR}(S)$ :

$$\text{VAR}(S) = \frac{1}{18} \left[ \begin{aligned} & n(n-1)(2n+5) - \sum_{p=1}^g t_p(t_p-1)(2t_p+5) - \sum_{q=1}^h u_q(u_q-1)(2u_q+5) + \\ & + \frac{\sum_{p=1}^g t_p(t_p-1)(2t_p-2) - \sum_{q=1}^h u_q(u_q-1)(2u_q-2)}{9n(n-1)(n-2)} + \frac{\sum_{p=1}^g t_p(t_p-1) \sum_{q=1}^h u_q(u_q-1)}{2n(n-1)} \end{aligned} \right],$$

де  $g$  – кількість груп зв'язаних даних,  $t_p$  – кількість зв'язаних даних у  $p$ -й групі,  $h$  – кількість разів вибірки, які містять кілька даних,  $u_p$  – кількість множинних даних за  $q$ -й період часу.

$Z$  – статистика задається рівнянням:

$$Z = \begin{cases} \frac{S-1}{\sqrt{\text{VAR}(S)}}, & \text{якщо } S > 0; \\ 0, & \text{якщо } S = 0; \\ \frac{S+1}{\sqrt{\text{VAR}(S)}}, & \text{якщо } S < 0. \end{cases}$$

Метод був обраний за основу для аналізу тренду, оскільки має декілька переваг перед іншими, а саме:

- він не передбачає розподіл даних відповідно до якогось конкретного правила, тобто, наприклад, не вимагає, щоб дані розподілялися нормально;
- на результат не впливають відсутні дані, окрім факту, що кількість точок вибірки зменшено і, отже, може негативно вплинути на статистичну значущість;
- на результат не впливає нерівномірний розподіл часових точок вимірювання;
- на результат не впливає довжина часового ряду.

Окрім тесту Манн-Кендала використовується також його модифікація, запропонована Хамедом та Рао. На основі модифікованого значення дисперсії статистики тренду Манна-Кендала використовується модифікований непараметричний тест тренду, який підходить для автокорельованих даних. Точність модифікованого тесту з точки зору рівня його емпіричної значущості є вищою, ніж оригінальний тест тренду Манна-Кендала без будь-якої втрати потужності.

Для перевірки гіпотези про віднесення часового ряду до класу стаціонарних (щодо лінійного тренду) чи нестаціонарних процесів був використаний тест Дікі-Фуллера (ADF). Результат тесту ADF (р-значення нижче 0,05) свідчить про те, що нульову гіпотезу про наявність одиничного кореня можна відхилити з довірчим рівнем 95%. Отже, якщо р-значення нижче 0,05, ЧР є стаціонарним [5].

У роботі для перевірки наявності тренду було запропоновано метод, при якому загальна оцінка наявності тренду складається на основі як мінімум трьох тестів: тест Манн-Кендала, модифікований тест Манн-Кендала (Ю. Ванг) та модифікований тест Манн-Кендала (Х. Рао). Результат методу інтерпретується як ціле число в інтервалі [-1;1]. Якщо результат є від'ємне число – то тренд існує, та є низхідним. В випадку, коли результат 0 – це свідчить про те, що тренду не існує. Якщо результат дорівнює 1 – це говорить про те, що в даних є тренд, який має висхідний напрям. Таким чином можна зробити комплексну оцінку наявності тренду, та отримати об'єктивні результати аналізу. Схему роботи методу можна побачити на рис. 1.

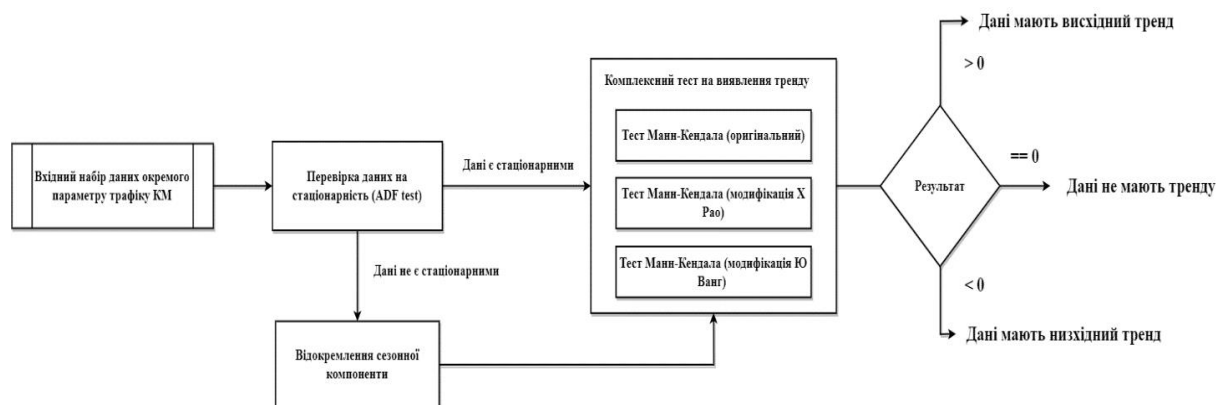


Рис. 1. Схеми роботи методу виявлення тренду на основі тесту Манн-Кендала

#### 4 Метод прогнозування мережевого трафіку

Прогнозування часових рядів – це процес аналізу даних часових рядів із використанням статистики та моделювання для прогнозування та прийняття стратегічних рішень. Це не завжди точне передбачення, і ймовірність прогнозів може різко відрізнятися, особливо коли йдеться про типові коливання змінних у даних часових рядів, а також фактори, які ми не контролюємо. Однак прогнозування розуміння того, які результати є більш імовірними або менш імовірними, ніж інші потенційні результати [5].

Прогнозування рядів часто використовується в поєднанні з аналізом часових рядів. Аналіз часових рядів включає розробку моделей для розуміння даних, щоб зрозуміти основні фактори впливу.

Моделі зазвичай оцінюються за допомогою середньоквадратичної помилки (MSE) або середньоквадратичної помилки (RMSE) [6].

Серед найбільш відомих і використовуваних моделей прогнозування часових рядів можна відмітити: авторегресійні моделі, моделі ковзного середнього й інтегровані моделі.

Інтегрована модель авторегресії та ковзного середнього, запропонована Боксом і Дженкінсом включає як параметри авторегресії, так і параметри ковзного середнього [7]. Саме, є три типи параметрів моделі: параметри авторегресії ( $p$ ), порядок різниці ( $d$ ), параметри ковзного середнього ( $q$ ). В позначеннях Бокса і Дженкінса модель записується як ARIMA ( $p, d, q$ ). Наприклад, модель (0, 1, 2) містить 0 (нуль) параметрів авторегресії ( $p$ ) і 2 параметри ковзного середнього ( $q$ ), які обчислюються для ряду після взяття різниці з лагом, що дорівнює 1.

Методологія Бокса-Дженкінса підбору ARIMA-моделі для певного ряду спостережень складається з чотирьох етапів [8]:

- ідентифікація моделі – процес вибору моделі, що найкраще відповідає аналізованому реальному процесу;
- оцінювання моделі – використання регресійних методів для отримання оцінок параметрів, включених в модель;
- тестування моделі – перевірка основних передумов використання регресійного аналізу, перевірка адекватності моделі з використанням тестів на нормальність залишків (тест Жарка-Бера), на автокореляцію залишків (тест Дарбіна Вотсона), на сталість дисперсій випадкових залишків (критерії Кохрана та Голдфалда Кванта) якості специфікації моделі (F-тест);
- використання моделі для прогнозування.

На етапі ідентифікацією порядку моделі необхідно вирішити, як багато параметрів авторегресії ( $p$ ) і ковзного середнього ( $q$ ) має бути присутнім в ефективній та економічній моделі процесу (економічність моделі означає, що в ній є найменша кількість параметрів та найбільша кількість ступенів свободи серед усіх моделей, що підганяються до даних). Насправді дуже рідко буває, що число параметрів  $p$  чи  $q$  більше 2.

Наступний, після ідентифікації, крок полягає в оцінюванні параметрів моделі. Отримані оцінки параметрів використовуються на останньому етапі, щоб обчислити нові значення ряду і побудувати довірчий інтервал для прогнозу. Процес оцінювання проводиться за перетвореними даними (підданим застосуванню оператора різниці). До побудови прогнозу необхідно виконати зворотню операцію (інтегрувати дані). Таким чином, прогноз методології порівнюватиметься з відповідними вихідними даними.

Додатково моделі ARIMA можуть містити константу, інтерпретація якої залежить від моделі, що підганяється. Саме, якщо у моделі немає параметрів авторегресії, то константа є середнє значення ряду, якщо параметри авторегресії є, то константа є вільним членом. Якщо бралася різниця ряду, то константа є середнім або вільним членом перетвореного ряду. Наприклад, якщо бралася перша різниця (різниця першого порядку), а параметрів авторегресії в моделі немає, то константа є середнім значенням перетвореного ряду  $i$ , отже, коефіцієнт нахилу лінійного тренду вихідного.

В роботі була використана модель прогнозування ARIMA, оскільки вона зазвичай застосовується для прогнозування даних стаціонарних часових рядів та враховує трендову компоненту ряду, завдяки чому має високі показники точності серед усіх моделей даного типу.

## 5 Опис набору даних, які характеризують мережевий трафік

За основу для набору даних було обрано логи комерційної компанії, яка спеціалізується на аналізі мережевого трафіку для крупних клієнтів [9]. З отриманих логів за допомогою розробленого програмного забезпечення було сформовано вибірку (з використанням парсингу текстових файлів), яка налічує більше 300 000 рядків. Об'єкти набору даних мають 9 параметрів (рис. 2):

1. Max packet size (bytes) – відповідає за показник максимального розміру пакету, що передається по мережі;
2. Send latency (usec) – показник затримки відправника у мілісекундах;
3. Recv latency (usec) – відповідає за показник затримки отримувача у мілісекундах;
4. Min packet spacing (usec) – відповідає за часову різницю між пересиланням пакетів (інтервал передачі пакетів) у мілісекундах;
5. Max rate (Mbps) – відповідає за показник максимальної швидкості даних в мережі на момент заміру;

6. ADR (Mbps) – показник адаптивного бітрейту мережі;
7. Grey bandwidth resolution – показник сірого трафіку мережі;
8. Measurement date – дата, коли було сформовано дані вимірювання показників трафіку мережі;
9. Available bandwidth (Mbps) – наявна пропускна здатність мережі.

	A	B	C	D	E	F	G	H	I
1	Max packet size (bytes)	Send latency (usec)	Recv latency (usec)	Min packet spacing (usec)	Max rate (Mbps)	ADR (Mbps)	Grey bandwidth resolution	Measurement date	Available bandwidth (Mbps)
2	1472	7	1		14857.14	0.65	0.03	06/03/2009	0.65
3	1472	33	1		66181.82	0.10	0.00	06/02/2009	0.10
4	1464	169	1		33835.31	0.25	0.01	06/01/2009	0.25
5	1472	90	1		18066.67	0.30	0.01	06/01/2009	0.30
6	1472	22	1		44272.73	0.54	0.03	06/01/2009	0.54
7	1464	39	1		78153.03	0.28	0.01	06/01/2009	0.28
8	1442	47	1		94125.11	0.30	0.02	06/01/2009	0.30
9	1472	43	1		86139.53	5.82	0.29	06/01/2009	2.93
10	1472	43	1		86139.53	6.29	0.31	06/01/2009	3.16

Рис. 2. Приклад перших 10 записів сформованого датасету

### 6 Результати оцінки тренду і прогнозування мережевого трафіку

В якості вхідних даних для моделі було обрано два параметри з найбільшим коефіцієнтом кореляції – Grey bandwidth resolution і Available bandwidth (Mbps). Усі дані було згруповано помісячно. Попередня обробка даних проводилась з використанням статистичних методів (усі N/A комірки були замінені на середнє значення всього набору), це дозволило отримати валідні дані для побудови моделі. Ці два показники також були вибрані через наявність чіткої трендової компоненти, модель тренду має логістичний тип (змішаний).

Наступним кроком був аналіз ряду на стаціонарність за допомогою теста Діка-Фулера. Цей тест показав, що відібрані 2 набори даних є стаціонарними: р-показник для параметру Grey bandwidth resolution дорівнює 0.00079, а для параметру Available bandwidth – 0.00701.

Для додаткового аналізу датасету також є доцільним провести декомпозицію часового ряду. Результати декомпозиції можемо побачити на рис. 3, 4.

Як можна побачити, сезонна компонента не є очевидною у декомпозиційному графіку обох параметрів, що вказує також на їх стаціонарний характер. Щодо тренду можна сказати, що трендова компонента присутня, але не є яскравовираженою.

При виборі параметрів моделі прогнозування ARIMA  $p, d, q$  було застосовано ітеративний метод підбору цих параметрів на основі критерію Акаїке. Для обох наборів вхідних даних кращій результат показала модель ARIMA(2,0,0)(0,0,0) (показник критерію Акаїке – 264.068).

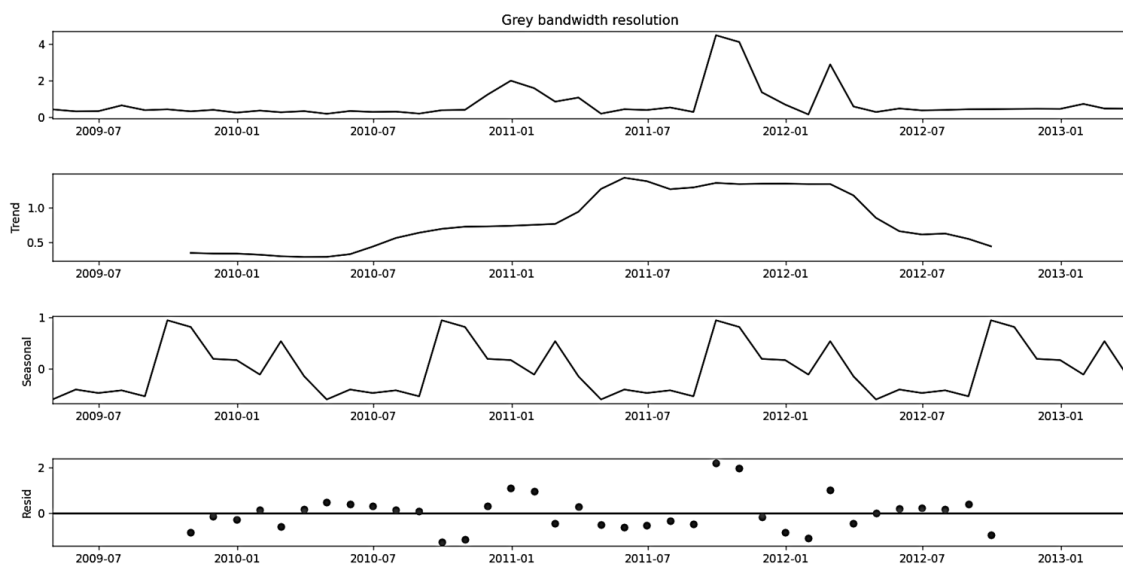


Рис. 3. Декомпозиція часового ряду параметру Grey bandwidth resolution

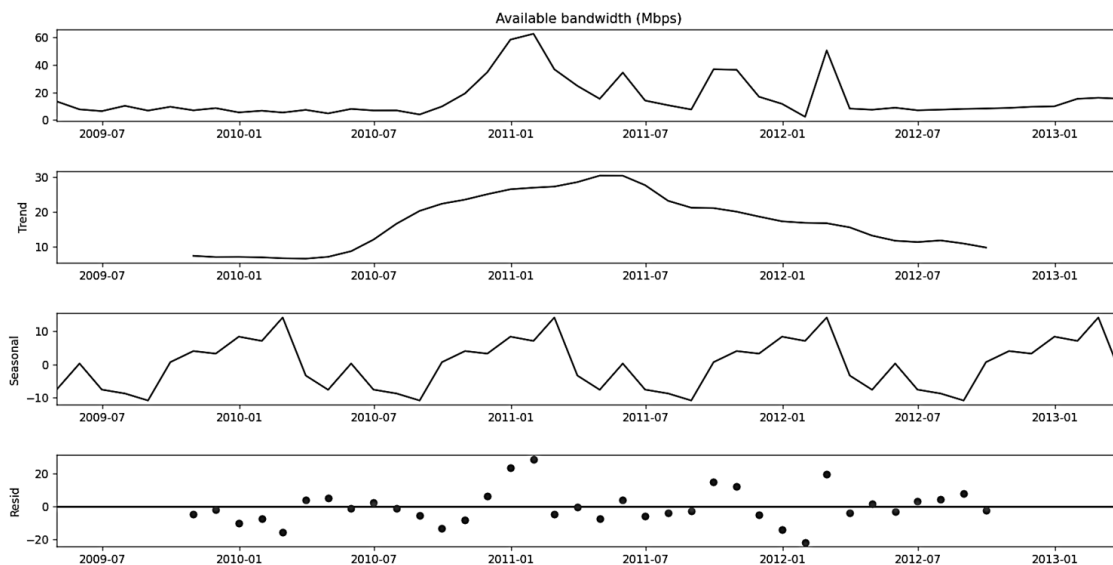
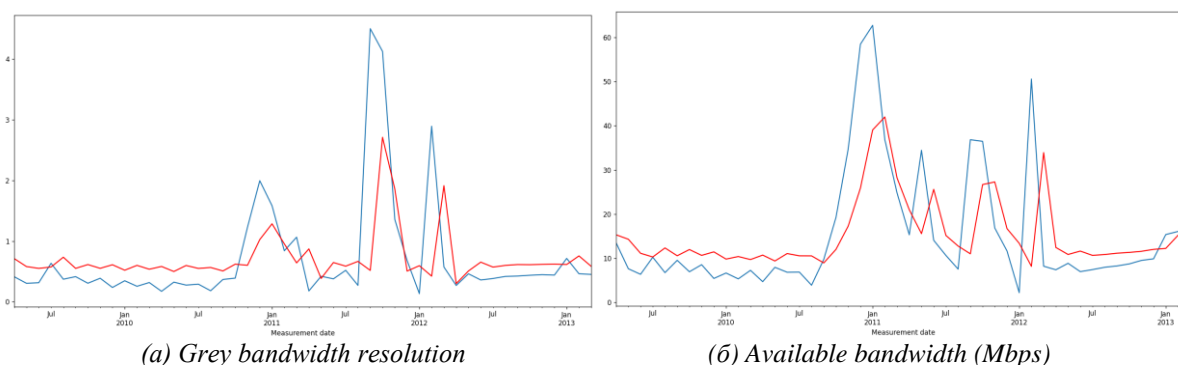


Рис. 4. Декомпозиція часового ряду параметру Available bandwidth (Mbps)

Після побудови моделі було отримані результати, графічне представлення яких можна побачити на рис. 5. Синім кольором на графіку відокремлені актуальні дані, червоним – результати моделі прогнозування. Як можна побачити за попередньою оцінкою, точність прогнозування не є високою.



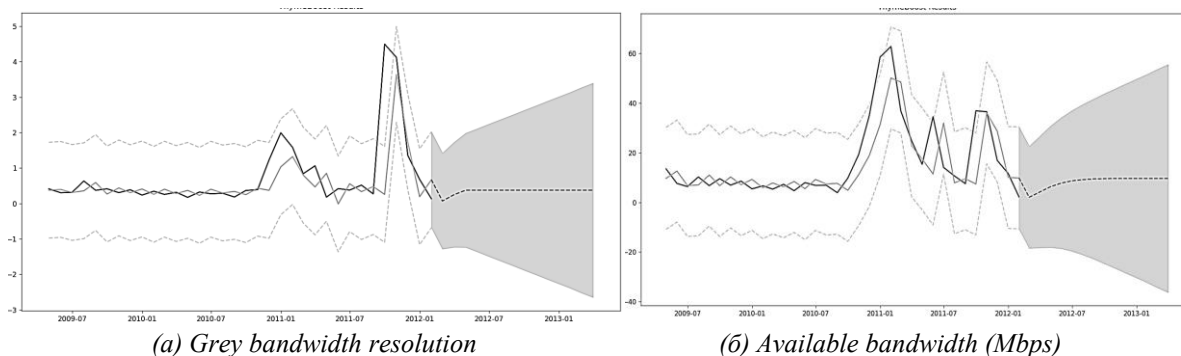
(a) Grey bandwidth resolution

(б) Available bandwidth (Mbps)

Рис. 5. Порівняльні графіки результатів моделі ARIMA

Запропоновано покращити результати роботи моделі прогнозування параметрів трафіку з використанням методів машинного навчання, а саме методу градієнтного бустінгу [10, 11]. Даний метод зазвичай використовується для зменшення відхилень та помилок. Модель авторегресії ковзного середнього включає в себе екзогенні змінні, завдяки яким можна використовувати екзогенні регресори (Exogenous Regressors), що і було зроблено в даній роботі.

На рис 6. можна побачити результати роботи моделі. Світло-сірим кольором позначено результати прогнозування значень.



(a) Grey bandwidth resolution

(б) Available bandwidth (Mbps)

Рис.6. Порівняльні графіки результатів покращеної моделі ARIMA

Для того, щоб упевнитися в тому, що результати моделі дійсно було покращено, було проведено порівняння актуальних метрик моделей. В якості набору характеристик точності моделі було використано показники MAE (Mean absolute error), MSE (Mean squared error) та RMSE (Root mean squared error). Результати порівняння наведено в таблиці 1.

Таблиця 1. Порівняння показників метрик точності моделі

Назва моделі	Grey bandwidth resolution			Available bandwidth (Mbps)		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE
ARIMA	0.454	0.749	0.728	7.973	13.626	0.618
Boosted ARIMA	0.296	0.766	0.475	5.71	13.33	0.44

Інтерпретуючи результати порівняння можна сказати, що дійсно, після застосування методу градієнтного бустінгу якість роботи моделі вдалося покращити. Значення всіх типи помилок вдалося покращити в середньому на 30%.

## 7 Висновки

На сьогоднішній день існують різні методи виявлення та оцінки трендів у трафіку комп'ютерних мереж. Побудова моделей аналізу трафіку комп'ютерних мереж за допомогою дослідження трендів часових рядів, а також лінійних алгоритмів регресії дозволять підвищити точність та якість аналізу стану комп'ютерних мереж. У роботі був сформований набір даних, який характеризує комп'ютерну мережу, для його подальшого аналізу, виявлення тренду і побудови моделі прогнозу.

Результати можуть бути використані в різних типах задач, таких як задачах аналізу трафіку комп'ютерних мереж, прогнозування станів технічних систем, розробка моделей комп'ютерних мереж, оцінювання функціонування комп'ютерних мереж, а також використані для аналізу та оцінки поточного трафіку комп'ютерних мережі для подальшої оптимізації, пошуку перспектив для подальшого розвитку у структурі мережі, виявлення проблем, що негативно впливають на систему. Використання сучасних методів аналізу та оцінки трафіку відіграють велику роль у розвитку комп'ютерних мереж, тож необхідне постійне вдосконалення існуючих методів та впровадження нових.

## ЛІТЕРАТУРА

1. Gray K. L. Comparison of Trend Detection Methods. *Graduate Student Theses, Dissertations & Professional Papers*. 2007. 228 p.
2. IBM Corporation Documentation. Trend detection. 2021. URL: <https://www.ibm.com/docs/en/siffs/2.0.3?topic=learning-trend-detection>.
3. Андриенко В.М., Арсирий Е.А. Интеллектуальный анализ временных рядов со стохастическим трендом. *Восточно-европейский журнал передовых технологий*. 2011. № 4/4 (52). С. 4-8.
4. Ковтун Н.В. Теорія статистики: Курс лекцій, практикум. К.: ІмексЛТД, 2007. 276 с.
5. Баклан І.В., Степанкова Г.А. Імовірнісні моделі для аналізу та прогнозування часових рядів. *Искусственный интеллект*. 2008. № 3. С. 505-515.
6. Романчук В.І., Лаврів О.А., Червенець В.В., Бак Р.І. Дослідження імовірнісних властивостей трафіку корпоративної мультисервісної мережі. *Радиоелектроніка та телекомунікації* : [збірник наукових праць] / відповідальний редактор Б. А. Мандзій. Львів : Видавництво Львівської політехніки. 2011. С. 128-134. (Вісник Національного університету «Львівська політехніка», № 705).
7. Robert Nau. Statistical Forecasting: notes on regression and time-series analysis. ARIMA Models for time-series forecasting. 2020.
8. Рилова Н., Оксанич І. Синтез ARIMA-моделей для прогнозування коефіцієнтів виходу кондиційних напівпровідникових матеріалів. *Системи обробки інформації*. 2015. Вип. 5 (130). С. 102-107.
9. Real World Networking datasets. URL: <https://gist.github.com/stefanbschneider/96602bb3c8b256b90058d59f337a0e59>.



10. Dekel Shai, Oren Elisha, Ohad Morgan. Wavelet Decomposition of Gradient Boosting. CoRR abs/1805.02642. 2018. P. 1-13.
11. Yanru Zhang, Ali Haghani. A gradient boosting method to improve travel time prediction. *Transportation Research Part C: Emerging Technologies*. 2015. Vol. 5, Part B. P. 308-324. Doi: <https://doi.org/10/1016/j.tr.2015.02.019>.

## REFERENCES

1. Gray K. L. Comparison of Trend Detection Methods. *Graduate Student Theses, Dissertations & Professional Papers*. 2007. 228 p.
2. IBM Corporation Documentation. Trend detection. 2021. URL: <https://www.ibm.com/docs/en/siffs/2.0.3?topic=learning-trend-detection>
3. Andriienko V.M., Arsiy Y.A. Intelligence analysis of time series with stochastic trend. *East-European Journal of Advanced Technology*. 2011. № 4/4 (52). P. 4-8.
4. Kovtun N.V. Theory of Statistics: Course of lectures, practice. Kyiv: ImeksLTD, 2007. 276 c.
5. Baklan I.V., Stepankova H.A. Probabilistic models for time series analysis and forecasting. *Artificial intelligence*. 2008. № 3. P. 505-515.
6. Romanchuk V.I., Lavriv O.A., Chervenets V.V., Bak R.I. Study of probabilistic properties of corporate multiservice network traffic. *Radio electronics and Telecommunications* : [collection of scientific papers] / editor B.A. Mandziy. Lviv : Publishing house of Lviv Polytechnic. 2011. P. 128-134.
7. Robert Nau. Statistical Forecasting: notes on regression and time-series analysis. ARIMA Models for time-series forecasting. 2020.
8. Rylova N., Oksanych I. ARIMA-models synthesis for forecasting the yield factors of conditioned semiconductor materials. *Information Processing Systems*. 2015. Vol. 5 (130). P. 102-107.
9. Real World Networking datasets. URL: <https://gist.github.com/stefanbschneider/96602bb3c8b256b90058d59f337a0e59>.
10. Dekel Shai, Oren Elisha, Ohad Morgan. Wavelet Decomposition of Gradient Boosting. CoRR abs/1805.02642. 2018. P. 1-13.
11. Yanru Zhang, Ali Haghani. A gradient boosting method to improve travel time prediction. *Transportation Research Part C: Emerging Technologies*. 2015. Vol. 5, Part B. P. 308-324. Doi: <https://doi.org/10/1016/j.tr.2015.02.019>.

**Strilets Viktoriia** *Ph.D, associate professor of the theoretical and applied system engineering department V.N. Karazin Kharkiv National University, 6 Svobody sq., Kharkiv, Ukraine, 61022*

**Doroshenko Maksym** *student of the computer science faculty V.N. Karazin Kharkiv National University, 6 Svobody sq., Kharkiv, Ukraine, 61022*

## Analysis and forecasting of computer network characteristics

Today most applications and devices interact with the data networks, so it is important to develop and improve technologies and methods to better understand, control, manage or predict the behavior and state of computer networks and their characteristics. Therefore, the tasks related to the development of models and methods for evaluating and forecasting computer network traffic parameters are important for computer network management. The paper considers the traffic of computer networks in terms of the time series. Trend models of the time series, trend detection criteria and assessment methods are reviewed and analyzed. The selected method for evaluating the traffic trend is based on the Mann-Kendall test, and the consensus method has been used to interpret the results. The task of forecasting computer network traffic taking into account trend indicators also has been considered. This problem has been successfully solved by using a forecasting model based on a moving average, and improved by using gradient boosting. A separate task was to collect and pre-process set of input data describing the operation of a computer network, to formalize it and perform subsequent quantitative and qualitative analysis. A unique data set has been created by parsing logs (system files) from monitoring the state of computer network traffic. It is this set of data that has been used to create a trend detection model and further forecast the characteristics of the computer network. The obtained results shows that the developed models and methods can be used on practice solving problems of monitoring and managing computer networks.

**Keywords:** *computer network, network traffic, trend analysis, forecasting problem, machine learning.*