

В.В. Гнатушенко, Б.М. Дейнека

**ДОСЛІДЖЕННЯ СТАТИСТИЧНОЇ МОДЕЛІ АНАЛІЗУ  
МЕРЕЖЕВОЇ АКТИВНОСТІ КОРИСТУВАЧІВ**

*Анотація. У статті розроблено статистичну модель аналізу мережевої активності користувачів, яка є масштабованою і інваріантною в часі. Класичні моделі, що використовують пуасонівські потоки, не враховують розподілу і кореляційні властивості інтервалів між запитами користувачів в агрегованому трафіку. Запропоновано модель на основі використання статистичного закону Ципфа, яка дозволяє проводити оцінку активності користувачів і прогнозувати параметри росту трафіку, використовувати її для виявлення і локалізації аномалій мережевого трафіку, включаючи DDoS атаки.*

*Ключові слова: трафік, модель, аналіз, користувач.*

**Постановка проблеми**

В даний час широкого поширення набули мережеві додатки, що викликало збільшення інформації, яка передається через комп'ютерні мережі. З іншого боку передача інформації по мережах пред'являє до інфраструктури жорсткі вимоги і виникає необхідність аналізу навантаження комп'ютерних мереж і розрахунку їх характеристик.

Існуючі моделі та алгоритми мережевого трафіку в основному відображають технічні аспекти передачі інформації в мережах на різних рівнях, при цьому часто не враховується вплив на динаміку трафіку спільної активності користувачів, при цьому цей фактор надає зростання впливу в зв'язку з інтенсифікацією обміну інформацією в мережевих спільнотах. Технічні рішення для мереж різного масштабу і протоколів різного рівня істотно розрізняються, в той час як динаміка активності користувача залежить від розмірів і зв'язності спільноти користувачів. Це призводить до необхідності дослідження можливості опису динаміки звернень користувача до мережевих ресурсів.

**Аналіз останніх досліджень**

Математичне моделювання поведінки трафіку комп'ютерних мереж знаходить застосування від технічної організації процесу передачі даних до алгоритмів маршрутизації та управління потоками даних. В даний час запропоновано і програмно реалізовано велику кількість різних моделей і алгоритмів, що імітують потоки даних. Виділимо кілька напрямків розвитку моделювання. Першу групу складають моделі, що імітують технічні аспекти передачі даних в мережах різної організації на різних рівнях. Дані моделі відображають особливості мережових протоколів, імітують відповідно до використовуваних алгоритмами розбиття елементів трафіку на пакети, їх подальшу організацію на різних мережових рівнях. На основі цих моделей можна оцінити характеристики потоків даних, що виникають при передачі заздалегідь відомого набору інформації, оцінити за результатами моделювання характеристики якості обслуговування (Quality of Service, QoS) [1,2].

Моделі, що описують типові характеристики потоків звернень до ресурсів мережі, які формуються кінцевими користувачами і спільнотами кінцевих користувачів, є основою другого напрямлення. У моделях, які засновані на пуасонівській динаміці, враховано угруповання (інкапсуляція) пакетів, неоднорідність потоку. З розвитком інформаційних мереж стало очевидним, що дані моделі не відображають справжньої динаміки потоків даних, в першу чергу, за рахунок того, що вони вважають передачі окремих елементів даних незалежними.

Окремо слід відзначити клас імітаційних моделей, які націлені на відтворення типової поведінки кінцевих користувачів мережі. Запропоновано моделі на основі агентного підходу, які формують реалізації по статистичних властивостях близькі до емпіричних даних потоків запитів в мережі [3,4].

Розвиток засобів інформаційного обміну приводить до більш вираженої нерегулярності динаміки звернень до тих чи інших мережових ресурсів, пов'язаних з поширенням посилань на ресурси через соціальні мережі. Потоки даних, безпосередньо пов'язані з використанням соціальних мереж займають домінуючу частку трафіку. Опис спільної поведінки кінцевих користувачів пов'язано з моделями по-

ширення інформації в мережевих спільнотах, які переміщуються в соціальних мережах.

У зв'язку з цим затребуваною і актуальною є задача аналізу поширення інформації в соціальних мережах, що може спростити процес розуміння механізмів поширення інформації, розробку методів передбачення поширення інформації та впливу обміну інформації між користувачами.

### Формулювання цілей статті (постановка завдання)

Метою даного дослідження є розробка моделі мережі для дослідження та аналізу активності поведінки користувачів на основі статистичної моделі даних.

### Основна частина

Для надійної роботи комп'ютерної мережі необхідна організація управління інформаційним обміном - узгодження певних умов таких, як режим обміну інформацією, максимальний обсяг інформації, який передається за один раз, формат даних, перелік дій в разі виникнення помилки і т.д. Розподіл вхідного потоку залежить від характеру фізичних процесів, які протікають в об'єкті, що моделюється. Найчастіше для моделювання використовуються експоненціальне, Ерланга k-го порядку, Релея, нормальне і рівномірне розподілення. Вхідні потоки є нерегулярними і простими.

У роботі створено імітаційну модель мережі, яка складається з 10 робочих станцій, сервера, 2 комутаторів та роутера, що забезпечує обмін даними між обчислювальними пристроями (рисунок 1).

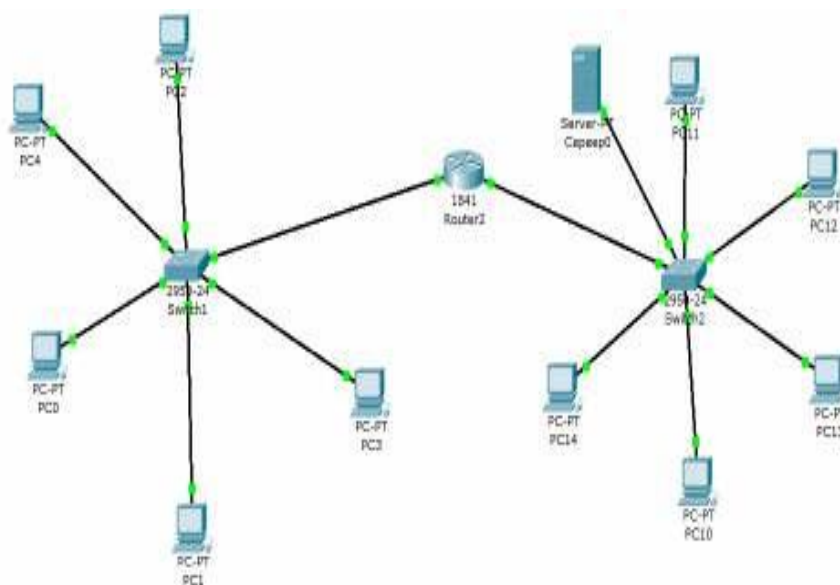


Рисунок 1 – Імітаційна модель мережі

В ході роботи була отримана статистика кількості пакетів по підмережам: кількість пакетів, переданих в дану підмережу за кожен досліджуваний день.

Таблиця 1

Статистика пакетів у підмережах

підмережа	вхідні пакети даних	вихідні пакети даних	інтервал часу
192.168.0.250			
192.168.0.1	102558	237364	04.12.18
192.168.0.2	120547	148695	04.12.18
192.168.0.3	148952	145254	04.12.18
192.168.0.4	65871	445285	04.12.18
192.168.0.5	157789	458688	04.12.18
168.192.1.250			
168.192.1.1	128456	148589	04.12.18
168.192.1.2	125489	18528	04.12.18
168.192.1.3	145585	141512	04.12.18
168.192.1.4	412356	158821	04.12.18
168.192.1.5	1445251	184558	04.12.18

Проведено аналіз досліджуваних даних на відповідність закону Ципфа.

$$f(k, s, N) = \frac{1 / k^s}{\sum_{n=1}^N (1 / n^s)},$$

де  $N$  — це число елементів,  $k$  — ранг елемента,  $s$  — параметр, який характеризує розподіл. У нашому випадку  $N$  — параметр, який характеризує середній час,  $s$  - загальна кількість вхідних та вихідних пакетів,  $N$  - кількість елементів у підмережі.

Для цього підмережі були розсортовані за величиною їх трафіку за день і по часових інтервалах. Підмережі мають подібну характеристику і відповідають закону Ципфа. Це говорить про існування статичної феноменологічної моделі для даної системи, яка представляє параметри трафіку в залежності від кількості активних користувачів. Також на основі проведених експериментів зроблено висновок про те, що даний потік не є пуасонівським. Таким чином підтверджений недолік класичних моделей, які передбачають пуасонівський характер розподілу інтервалів між запитами.

Для апроксимації досліджуваних даних необхідно вибрати модель, якій вони будуть найбільш відповідати. Як було показано вище дані відповідають закону Ципфа, отже доцільно використовувати ступеневу модель розподілу. У класичних моделях використовується розподіл Пуасона або Ерланга. Однак при великих обсягах трафіку і обмежених вибірці в кінці графіка виникає експоненціальний завал. Для його врахування статистична функція замінюється на добуток статистичної і експоненційної функції. Для отримання універсальної феноменологічної моделі дані апроксимації піддані регресійному аналізу.

### **Висновки та перспективи подальших досліджень**

Класичні моделі трафіку, які використовують пуасонівські потоки, істотно недооцінюють ряд статистичних показників: розподіл і кореляційні властивості інтервалів між призначеними для користувача запитами в агрегованому трафіку. У роботі запропонована модель на основі використання статистичного закону Ципфа. Проведений аналіз показав застосовність даної моделі для феноменологічного опису трафіку на різних рівнях мережі. Дана модель має властивість масштабованості на рівні підмереж і по всій мережі в цілому, інваріантна до часу. Таким чином, на основі даній моделі можна робити оцінку активності користувачів і прогнозувати параметри росту трафіку, а також використовувати її в якості одного з інструментів для виявлення і локалізації аномалій мережевого трафіку, включаючи DoS атаки.

### **ЛІТЕРАТУРА**

1. Шелухин О.И., Тенякшев А.М., Осин А.В. Моделирование информационных систем. Учебное пособие. — М.: Радиотехника, 2005. — 368 с.
2. Таненбаум Э.С. Компьютерные сети [пер. с англ.]. — СПб.: Издательский дом "Питер" 2012. — 960 с.
3. Universal model for collective access patterns in the internet traffic dynamics: A superstatistical approach / A. Tamazian, V.D. Nguyen, M. Bogachev, O. Markelov // EPL. -2016.- Vol.115. - Pp. 10008(1-7).
4. Поршнева С.В. Математические модели информационных потоков в высокоскоростных магистральных интернет-каналах. Учебное пособие для вузов. — М., 2016. — 323 с. ISBN 978-5-9912-0508-5