

УДК 621.391

Стрихалюк Б. М., канд. техн. наук, доцент (Тел.: +380 93 920 06 98. E-mail: bogdan_str@ukr.net)

Демидов І. В., канд. техн. наук., доцент (Тел.: +380 99 371 58 54 E-mail: demydov@lp.edu.ua)

Романчук В. І., канд. техн. наук., доцент (Тел.: +380 67 672 98 54 E-mail: romanchuk@lp.edu.ua)

Бешлей М. І., аспірант (Тел.: +380 93 710 32 44. E-mail: beshlembi@gmail.com)

(Національний університет «Львівська політехніка», м. Львів)

ДОСЛІДЖЕННЯ СТАТИСТИЧНИХ ПАРАМЕТРІВ ТА ХАРАКТЕРИСТИК ІНФОРМАЦІЙНИХ ПОТОКІВ ГЕТЕРОГЕННОЇ МЕРЕЖІ

Стрихалюк Б. М., Демидов І. В., Романчук В. І., Бешлей М. І. Дослідження статистичних параметрів та характеристик інформаційних потоків гетерогенної мережі. Проведено дослідження статистичних показників трафіку гетерогенної мультисервісної корпоративної мережі з метою формування адекватних імовірнісних моделей його розподілу для різних типів сервісу. Здійснено оцінку параметру Херста для різних типів трафіку, що характеризує властивості самоподібності потоків користувачів за відповідними типами сервісу.

Ключові слова: cloud-система, гетерогенна мережа, параметр Херста, мультисервісний трафік, статистичні моделі, самоподібний трафік.

Стрихалюк Б. М., Демидов І. В., Романчук В. І., Бешлей М. І. Исследование статистических параметров и характеристик информационных потоков гетерогенной сети. Проведено исследование статистических показателей трафика гетерогенной мультисервисной корпоративной сети с целью формирования адекватных вероятностных моделей его распределения для различных типов сервиса. Для разных типов трафика осуществлена оценка параметра Херста, характеризующего свойства самоподобия потоков пользователей по соответствующим типам сервиса.

Ключевые слова: cloud-система, гетерогенная сеть, параметр Херста, мультисервисный трафик, статистические модели, самоподобный трафик

Strykhalyuk B. M., Demydov I. V., Romanchuk V. I., Beshley M. I. Research of statistical parameters and characteristics of informational traffic in heterogeneous network. A research was carried out for statistical indexes of traffic in multiservice corporative heterogenous network to define adequate probabilistic models for different service types. There were Hurst parameters assessed for the variety of traffic types that characterizing respective typical self-similarity properties of consumers' service traffic.

Keywords: cloud-system, heterogeneous network, Hurst parameter, multiservice traffic, statistical models, self-similar traffic

Вступ. Сучасні розподілені сервісно-орієнтовані мережі є складними гетерогенними системами. В основі більшості з них на сьогодні лежать так звані хмарні технології. Більшість телекомунікаційних мереж (ТКМ) є гетерогенними, тобто складаються із різноманітних програмно-апаратних засобів під керуванням різних операційних систем, та забезпечують споживачам широкий спектр інформаційних послуг. Це стає суттєвою перешкодою для досягнення необхідної якості обслуговування, зокрема дотримання часових вимог, для багатьох інформаційних систем критичного застосування [1, 2].

Моделювання складних сервісних систем відбувається із застосуванням множини методів, в основі більшості з яких лежить теорія випадкових процесів у поєднанні з теорією телетрафіку. Із урахуванням властивості розподіленості, в таких системах особливого значення набувають методи керування інформаційним трафіком. Ефективне використання каналу зв'язку означає правильний розподіл його ресурсів, збільшення кількості послуг, відповідно і збільшення прибутку провайдера чи збільшення вигоди користувачу мережі.

Останні дослідження властивостей інформаційних потоків в гетерогенних телекомунікаційних системах показали, що використання моделей самоподібних процесів дозволяє більш точно описувати трафік, що передається в даних системах.

Оскільки в пакетному трафіку різних мереж було виявлено властивість самоподібності, виникла проблема побудови моделей, які могли б достатньо точно описати такий трафік та забезпечити збереження переданих даних або застосувати вже відомі моделі для опису подібних трафіків, що, зокрема, характерні для сервісів різного роду. Під час дослідження та

вивчення різних властивостей і характеристик систем передавання інформації виникає завдання побудови моделей вхідного трафіку, характеристики яких були б найближчими до характеристик реальних потоків даних заданого роду.

Постановка задачі. Характерною особливістю гетерогенних мереж можна вважати великі обсяги інформації, що циркулює в мережі та високі часово-імовірнісні вимоги до її передавання. В умовах істотного збільшення обсягів інформації у гетерогенних системах особлива роль приділяється вдосконаленню й розробці нових технологій передавання та розподілу інформації, що забезпечують оперативність обміну на основі інформаційних технологій теорії масового обслуговування та методів фрактального аналізу. Традиційні методи розподілу мережевого ресурсу припускають згладжування трафіка інформаційних потоків на основі статистичного мультиплексування. Існуючі методи керування перевантаженнями, які використовувались на критичних ділянках, також не враховують властивостей трафіка гетерогенних мереж [3]. Ця обставина визначає необхідність удосконалення класичного математичного інструментарію аналізу характеристик потоків, прогнозування навантаження та синтезу розподілених телекомунікаційних систем та мереж для підвищення оперативності передавання інформації.

Таким чином, проблема підвищення оперативності передавання інформації в умовах збільшення її обсягів на основі розробки математичних моделей і методів дослідження передавання інформаційних потоків в гетерогенних мережах має важливе науково-практичне значення та є актуальною задачею. В загальному випадку трафік деякої послуги представляється у вигляді випадкового процесу. Практичне використання такого методу опису є складним оскільки відсутній математичний апарат, що забезпечує оцінку параметрів якості нестационарного навантаження загального вигляду. Проте знання характеристик трафіку користувачів, який передаватиметься по мережі, є неодмінною умовою для проектування складної розподіленої гетерогенної мережі [4].

На Рис.1 ілюстративно зображено всі позитивні функції, якими володіють гетерогенні Cloud-системи у порівнянні зі звичайною ІТ системою.

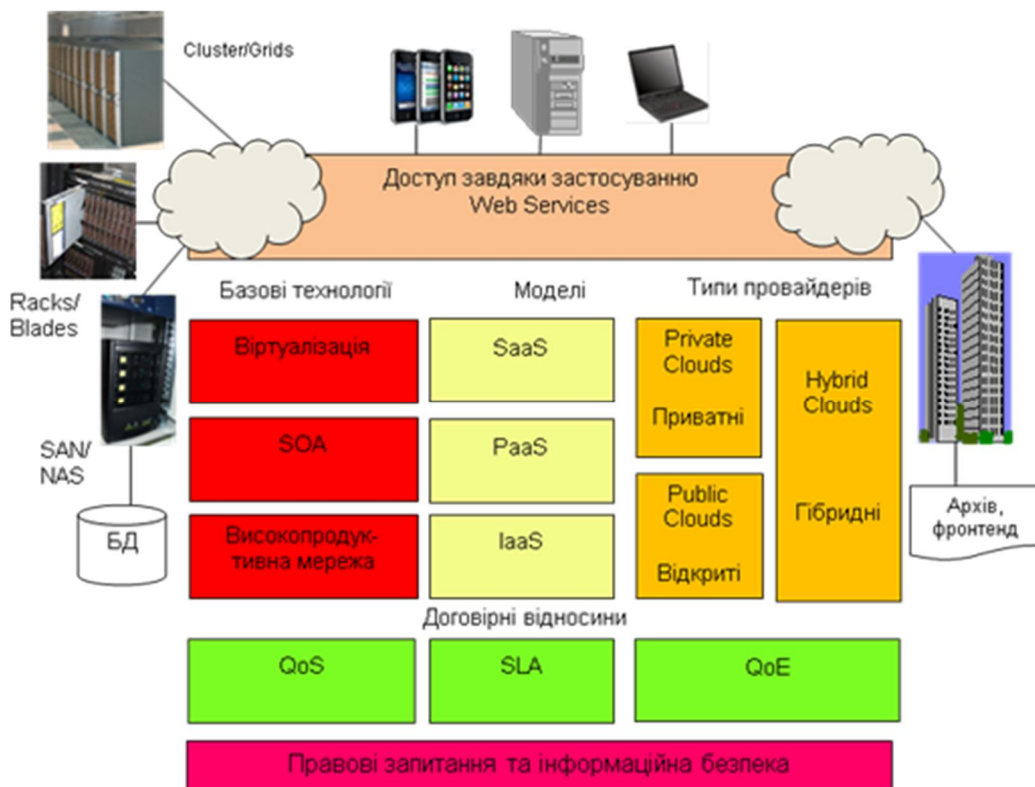


Рис. 1. Структура гетерогенної Cloud-системи

Важливо не забувати і про параметри надання сервісів у таких системах. Найважливішими критеріями щодо використання хмар виступають функції, які відповідають за знання характеристик та параметрів трафіку на основі яких розробляються моделі забезпечення гарантованої якості надання інформаційних потоків.

Опис досліджуваних мережних потоків. Опис і аналіз інформаційних потоків сучасних гетерогенних мереж ускладнюється такими причинами:

- широкий діапазон швидкостей передавання;
- різноманітні статистичні властивості мультимедійних інформаційних потоків;
- велика різноманітність мережних конфігурацій, безліч технологій і протоколів передавання;
- багаторівневе оброблення повідомлень, унаслідок чого якість обслуговування виявляється залежною від декількох рівнів обробки.

В даній роботі аналізувався реальний трафік корпоративної мережі із використанням пріоритетизації потоків [5]. В досліджуваній мережі існували групи користувачів, що прагнули отримати різнотипні послуги з найкращою якістю, та мережеве обладнання, маршрутизатори, комутатори, серверні ферми, щоб надати ці послуги та транспортувати дані до кінцевого користувача. Для того, щоб реально оцінити можливість якісного обслуговування мультисервісної мережі та вплив методів обслуговування різних типів потоків, дослід проведено при одночасній наявності всіх мережевих потоків, утворених множиною користувачів.

Отже, кожен кінцевий користувач бажає отримати наступні послуги:

- відео конференцзв'язок із іншими користувачами (Skype);
- голосовий зв'язок (IP-телефонія або VoIP);
- під'єднання до мережевої відео трансляції (IPTV);
- перегляд онлайн відео (VoD);
- інтернет-серфінг (Web Data);
- високошвидкісне завантаження вибраного файлу із локального файлового сховища (Web Download).

Для захоплення, запису та аналізу мережевого трафіку використано утиліту сніфер, а саме програмне забезпечення Wireshark Version 1.10.11, кінцевим результатом якої є створений дамп всіх мережевих пакетів присутніх на порті обслуговуючого пристрою за період спостереження. В даному дослідженні період запису дорівнює 2 хвилини. Цей параметр вибрано таким, виходячи з міркувань, який об'єм даних наявна обчислювальна техніка зможе обробити. Очевидно, що тривалі спостереження є досить ресурсоемними і потребують високих обчислювальних затрат. Отриманий дамп проаналізовано та визначено наявність шести мережевих потоків, що відповідають кожній з бажаних послуг кінцевих користувачів або, сервісам, які надаються мультисервісною гетерогенною мережею.

Основні параметрів мережевого потоку кожного типу трафіку Табл. 1

Тип трафіку	Протокол	Сер. довжина пакета L, байт	Смуга пропускання C, Кбіт/с	Пріоритет
SKYPE	RTSP	1370	3820	2
IPTV	UDP	1429	6044	4
Інтернет дані	HTTP	1514	7436	6
IP-телефонія	RTP	206	181	3
Відео за запитом (VoD)	TCP	666	1354	5
Службові дані	-	66	50	1
Загальний потік	-	1300	19122	-

На основі аналізу проведених досліджень [6, 7] запропоновано класифікацію трафіку за сервісами і виконано підбір кожному типу сервісів таких статистичних законів розподілу, що дозволяють найбільш повно описати поведінку трафіку на конкретному рівні моделі IETF.

Для опису трафіку в гетерогенних IP-мережах найбільш широко застосовується розподіл Парето. Крім того, для трафіку певних сервісів обмежено використовуються розподіли Вейбулла, логнормальний і пуассонівський. Важливо відзначити, що розподіл Парето, як і розподіл Вейбулла і логнормальний, відноситься до розподілів з важким хвостом і володіє нескінченною дисперсією в діапазоні $1 < a < 2$, де показник a характеризує “важкість” хвостів розподілу і визначає пачкову структуру процесу. Такі процеси відносяться до самоподібних процесів. Таким чином, поведінка трафіку IP-мережі характеризується різними законами розподілу. В якості математичної моделі СМО може бути обрана система із загальними розподілами вхідного потоку і процесів обслуговування типу G/G/1. СМО типу G/G/1 є найбільш загальним випадком одноканальних систем масового обслуговування, в яку надходить довільний потік заявок загального вигляду з функцією розподілу інтервалів між заявками $A(\tau)$. Тривалість обслуговування заявок в приладі розподілена по довільному закону $B(\tau)$. Завдання розрахунку систем виду G/G/1 для зазначених вхідних процесів і процесів обслуговування ускладнене тим, що дисперсії цих процесів нескінченні.

Розрахунок таких систем вимагає задання конкретних законів розподілів, що не дозволяє отримати аналітичне рішення в загальному вигляді. Аналітичне рішення можливо тільки для деяких розподілів, наприклад, експоненційного. Для більшості законів розподілів інтервалів між поступившими в систему заявками і тривалостями їх обслуговування неможливо отримати точне рішення в аналітичній формі.

У той же час, на практиці при дослідженні реальних систем рідко бувають відомі закони розподілів зазначених величин. Зазвичай при описі процесів надходження заявок в систему і їх обслуговування в приладі обмежуються кількома моментами відповідних розподілів, найчастіше – двома першими моментами, які задаються у вигляді математичного очікування і середньоквадратичного відхилення або коефіцієнта варіації шуканої випадкової величини. Однак при цьому виявляється неможливим отримання точного результату. Це обумовлено тим, що у випадку довільного (відмінного від найпростішого) потоку заявок, що надходять у систему, характеристики функціонування СМО, зокрема середній час очікування, залежать не тільки від двох перших моментів, але і від моментів більш високого порядку – третього, четвертого і т.д. Причому ця залежність тим менша, чим вище порядок числового моменту.

Таким чином, всі результати, отримані в аналітичній формі при заданні інтервалів між заявками, що надійшли до системи і тривалостями їх обслуговування двома першими моментами - середніми значеннями $a = 1/\lambda$ і $s = 1/\mu$ і коефіцієнтами варіації C_a і C_s , являють собою наближені залежності

Завдяки результатам дослідження, маючи кількість всіх пакетів, довжину кожного пакета, можна побудувати графік розподілу кількості пакетів відносно часу дослідження та відношення пропускної здатності (швидкості завантаження за одиницю часу) відносно часу дослідження. Через негабаритність даних, вхідні значення для графіків упускаються. Задля збереження якості подання інформації графіки наводяться в оригінальному вигляді, який отримується в програмі Wireshark, без умовних позначень осей абсцис та ординат, крім розмітки. Тривалість наведена у секундах, швидкість передавання – біти на секунду (bits per second).

Інтенсивність поступлення переданої/отриманої інформації в бітах на секунду до часу тривалості досліду показано на Рис. 2...6. Дослідження проводились в корпоративній мережі провайдерів НТЦ МТ Національного університету «Львівська політехніка», а також на мережних ресурсах Інтернет-провайдера Airbites у м. Львові.

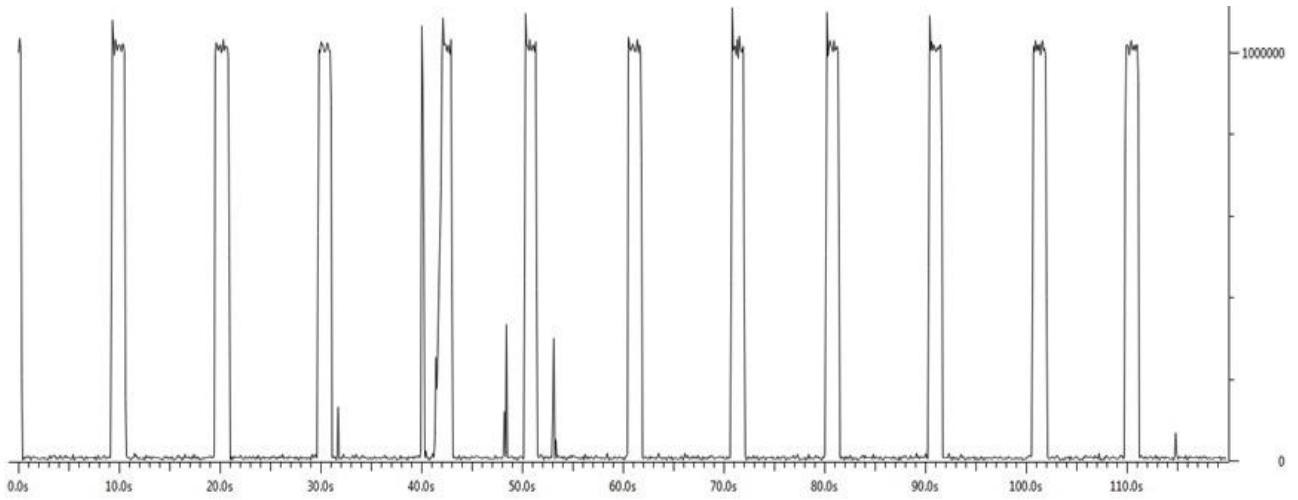


Рис. 2. Трафік при перегляді відео файлу он-лайн

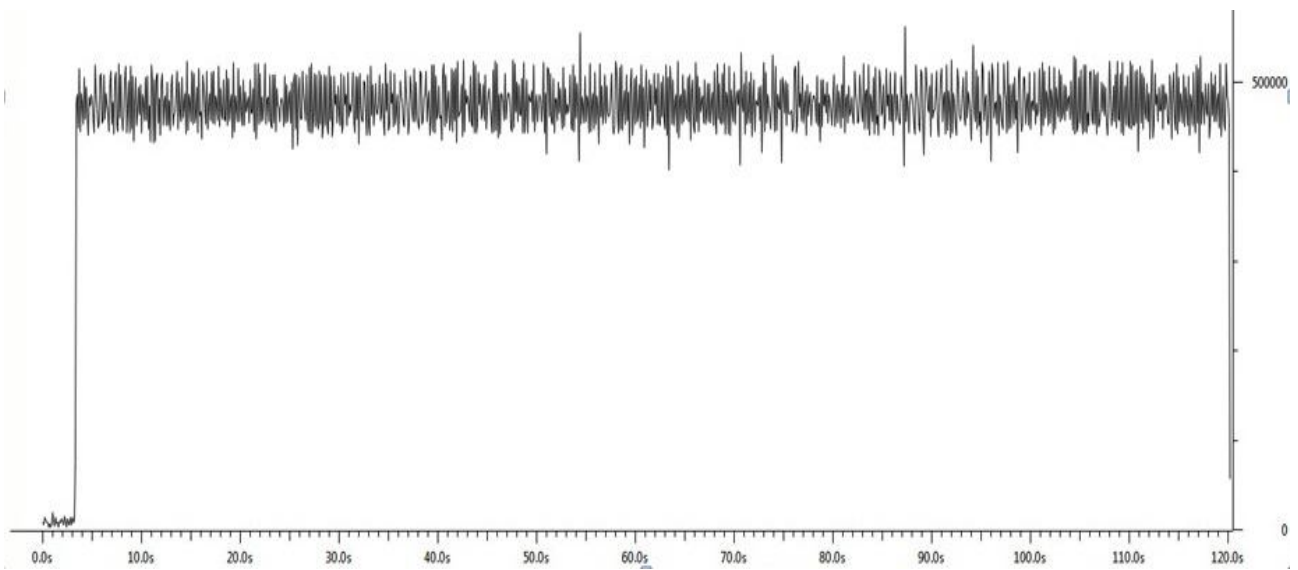


Рис. 3. Трафік при перегляді IPTV послуги

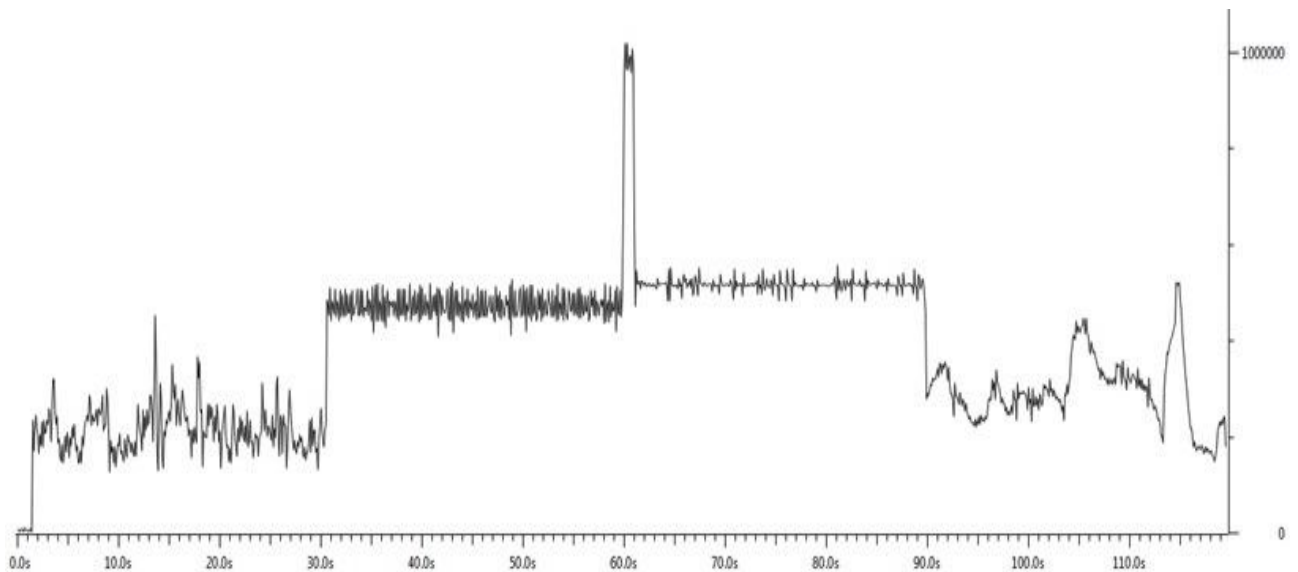


Рис. 4. Трафік при перегляді різних каналів IPTV послуги

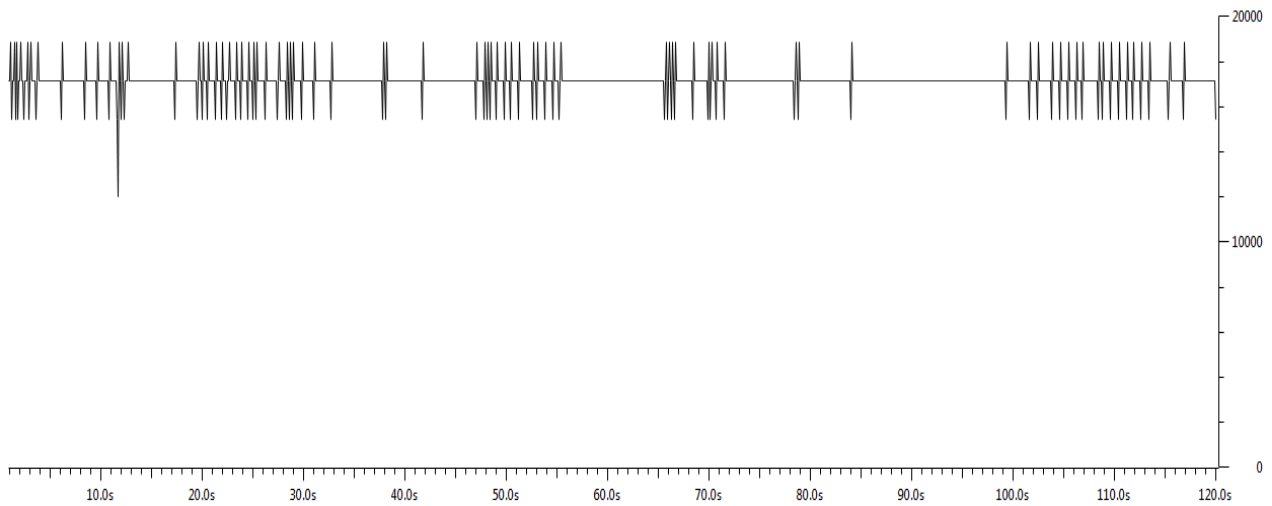


Рис. 5. Трафік при дзвінках послуги ІР-телефонії

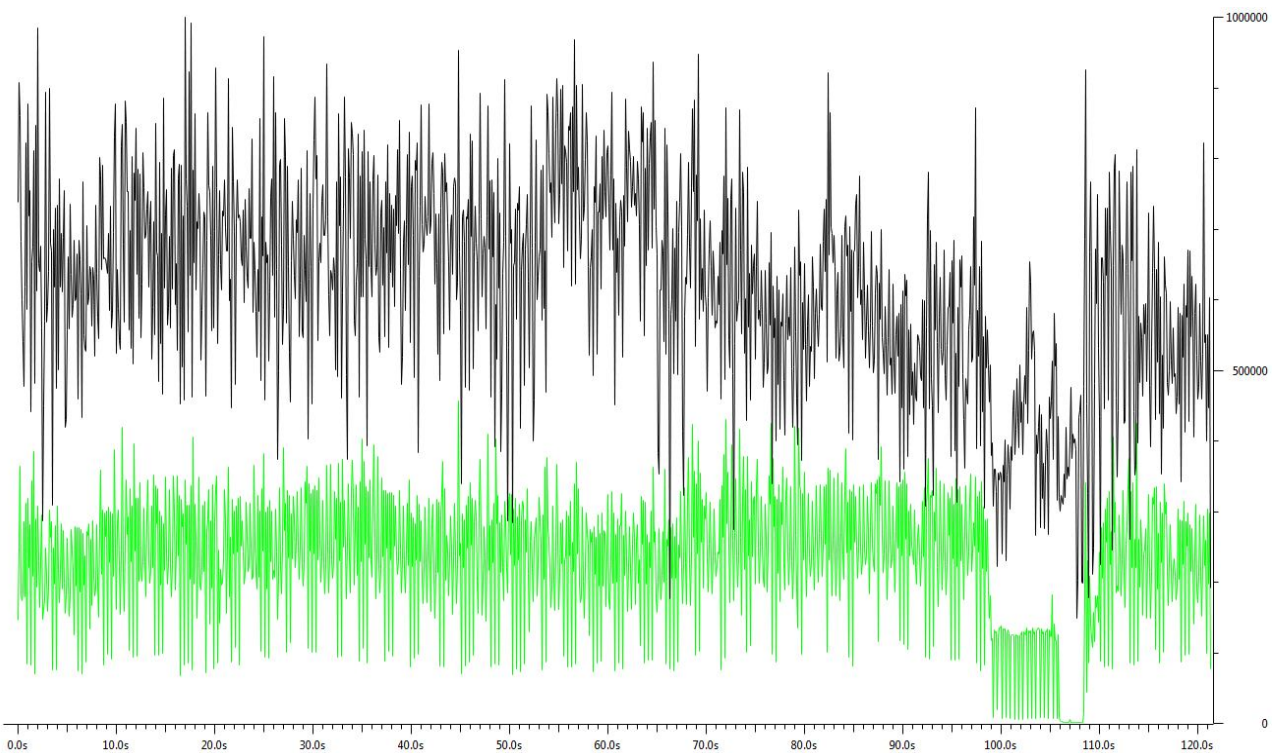


Рис. 6. Трафік SKYPE (світліший) та IPTV (темніший)

Підбір теоретичного закону розподілу і його параметрів. Для аналізу і вибору теоретичного розподілу, який найкраще характеризуватиме отриманий експериментальним шляхом розподіл, використаємо наступні закони розподілу:

Нормальний закон розподілу:

$$f(x) = \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma^2}} e^{-\frac{(x-m)^2}{\sigma^2}} ;$$

$$F(x) = \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{(x-m)^2}{\sigma^2}} dx .$$

Для даного розподілу $m = Mx$, $\sigma^2 = S^2$.

Показниковий закон розподілу:

$$f(x) = \lambda e^{-\lambda x} ;$$

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & x \geq 0. \end{cases}$$

Параметр λ визначається як $\lambda = \left| \frac{1}{M_x} \right|$.

Розподіл Пуассона:

$$f(x) = \frac{\lambda^x}{x!} e^{-\lambda} ;$$

$$F(x) = \frac{\Gamma([k+1], \lambda)}{[k]!},$$

де $\Gamma(x, y)$ – це неповна гама функція та $[k]$ – це ціла частка.

Для даного розподілу $\lambda = M_x$.

Релеєвський закон розподілу:

$$f(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}, \quad 0 \leq x \leq \infty ;$$

$$F(x) = 1 - e^{-\frac{x^2}{2\sigma^2}}, \quad 0 \leq x \leq \infty .$$

Параметр σ визначається як:

$$\sigma = |M_x| \cdot \sqrt{\frac{\pi}{2}} .$$

Логнормальний закон розподілу:

$$f(x) = \begin{cases} \frac{1}{x \cdot \sigma \cdot \sqrt{2 \cdot \pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, & x \geq 0 \\ 0, & x \leq 0 \end{cases} ;$$

$$F(x) = \frac{1}{2} - \frac{1}{2} \operatorname{erf} \left[\frac{\ln x - \mu}{\sqrt{2\sigma^2}} \right] .$$

Для агрегованого потоку навантаження характерною є статистична картина, що подана на Рис. 7. Встановлено, що найбільш придатним законом статистичного розподілу для даного випадку є релеєвський. Для даного розподілу $\mu = M_x$, а $\sigma^2 = S^2$.

З метою перевірки адекватності обраної статистичної моделі емпіричній вибірці висунуто статистичну гіпотезу H_0 про адекватність статистичної моделі обраній емпіричній вибірці. Для перевірки гіпотези використано статистичні критерії узгодженості Колмогорова, Пірсона та Колмогорова-Смірнова.

У Табл. 2...4 наведено результати дослідження обраної вибірки інтенсивності мережного трафіку, а на Рис. 7 наведено гістограму вибірки апроксимуючих кривих досліджуваних розподілів.

В проведених дослідженнях для визначення ступеня самоподібності трафіку використовувався параметр Херста, який може приймати наступні значення:

- $0 \leq H \leq 0,5$ – випадковий процес, випадковий ряд, який не володіє самоподібністю;
- $H = 0,5$ – повністю випадковий ряд, аналогічний до випадкових зміщень частинки при класичному броунівському русі;
- $H > 0,5$ – процес, який володіє довготривалою пам'яттю і є самоподібним.

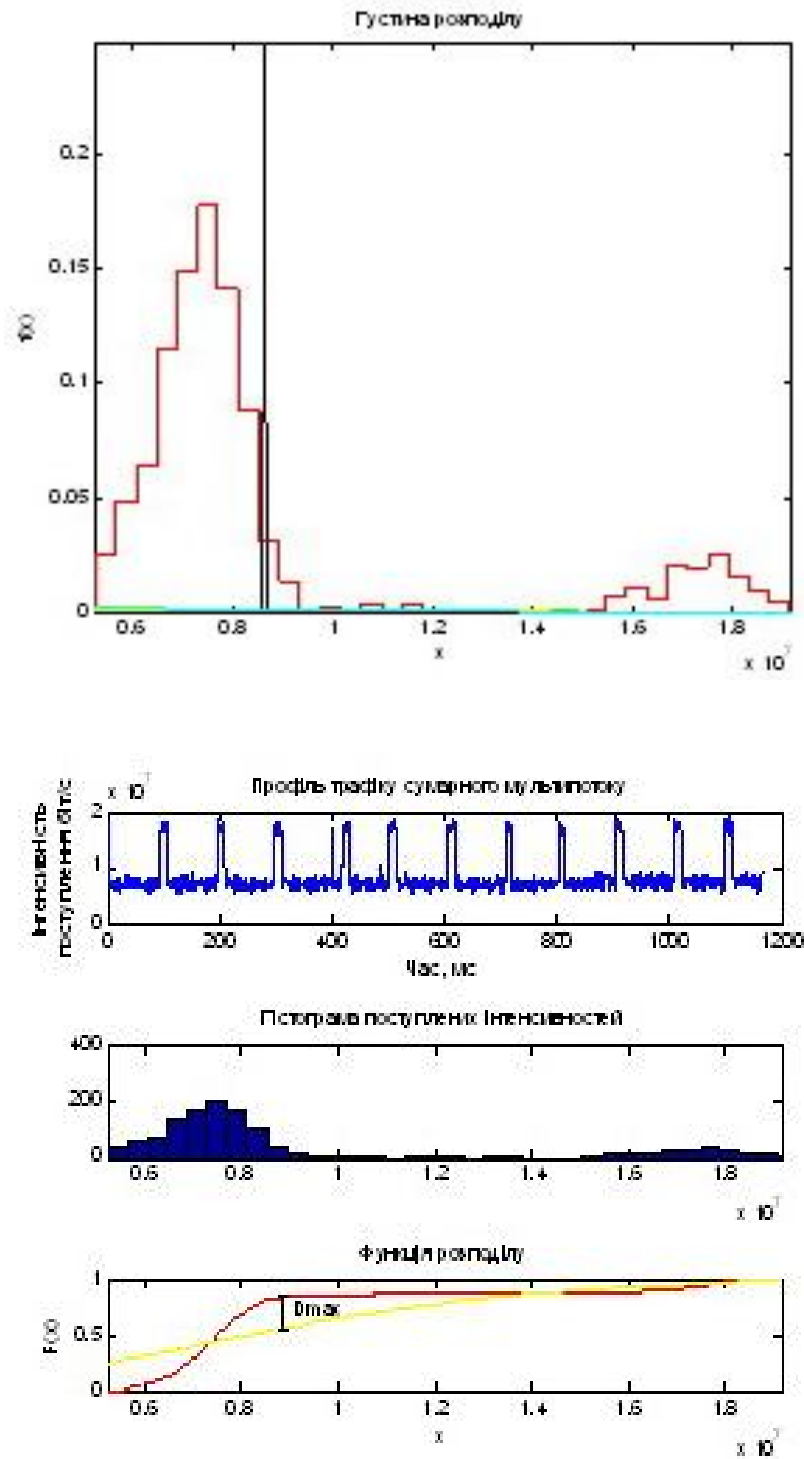


Рис. 7. Статистичний аналіз багатопотокового трафіку

Результати перевірки статистичної гіпотези за допомогою критерію Колмогорова із параметром значущості $p=0,05$

Табл. 2

Статистичний розподіл	λ кр.	λ емп.	H0
Нормальний	1,358	1,328	Приймається
Показниковий		1,323	Приймається
Пуассона		1,894	Відкидається
Релеєвський		1,635	Відкидається
Логонормальний		2,569	Відкидається

Результати перевірки статистичної гіпотези за допомогою критерію Колмогорова-Смірнова із параметром значущості $p=0,05$

Табл. 3

Статистичний розподіл	Ка кр.	Ка емп.	H0
Нормальний	0,5	0,1931	Приймається
Показниковий		0,6071	Відкидається
Пуассона		0,7844	Відкидається
Релеєвський		0,4974	Приймається
Логонормальний		0,7103	Відкидається

Результати перевірки статистичної гіпотези за допомогою критерію Пірсона із параметром значущості $p=0,05$

Табл. 4

Статистичний розподіл	χ^2 кр.	χ^2 емп.	H0
Нормальний	$0,59 \cdot 10^3$	$0,397 \cdot 10^3$	Приймається
Показниковий		$5,206 \cdot 10^3$	Відкидається
Пуассона		$6,028 \cdot 10^3$	Відкидається
Релеєвський		$3,672 \cdot 10^3$	Відкидається
Логонормальний		$2,631 \cdot 10^3$	Відкидається

Для оцінки значення параметра Херста випадкового ряду є багато методик. Найпростішою з них є RS-методика, яка, проте, має обмеження на застосування до процесів з малою дисперсією. Однак для потреб задач, вирішення яких пропонується в роботі, дана методика може бути використана.

Її суть полягає в наступному. Для випадкового ряду $X_k(k=1..N)$ визначаються математичне сподівання (1), дисперсія вибірки (2), інтегральне відхилення (3), рознесення випадкового процесу (4):

$$M_N = \frac{1}{N} \sum_{k=1}^N X_k ; \quad (1)$$

$$S_N^2 = \frac{1}{N} \sum_{k=1}^N (X_k - M)^2 ; \quad (2)$$

$$D_j = \sum_{k=1}^j X_k - jM, \quad j \in [1; N]; \quad (3)$$

$$R_N = \max_{1 \leq j \leq N} D_j - \min_{1 \leq j \leq N} D_j . \quad (4)$$

З встановленого Херстом співвідношення

$$\frac{R}{S} \approx \left(\frac{N}{2} \right)^H , \quad (5)$$

визначається параметр Херста H:

$$H = \frac{\log\left(\frac{R}{S}\right)}{\log\left(\frac{N}{2}\right)} . \quad (6)$$

Узагальнюючи отримані результати для кожного типу трафіку за видами сервісу згідно (1)...(6) можливо розрахувати відповідні типові значення параметру Херсту за RS методикою, що подані в Табл. 5.

Основні параметри мережного потоку для кожного типу трафіку Табл. 5

Тип трафіку	Коефіцієнт варіації	Параметр Херста
SKYPE	0,200	0,513
IPTV	0,204	0,346
Інтернет дані	0,207	0,685
IP-телефонія	0,370	0,981
Відео за запитом (VoD)	0,194	0,608
Службові дані	0,183	0,719
Загальний потік	0,39	0,639

Висновки. В даній роботі проведено статистичне дослідження мультисервісного трафіку в гетерогенній корпоративній cloud системі з метою аналізу показників його самоподібності для конкретних типів трафіку. Це, в свою чергу дозволить створювати та використовувати адекватні статистичні моделі для імітації процесів передавання інформаційних потоків, що відповідають заданим типам сервісу в гетерогенних мережних системах, отримувати їх усереднені статистичні параметри.

Отримані результати дослідження допоможуть оцінити придатність мережі до передавання мультисервісних даних заданого виду, виявити потенційні вузькі місця в гетерогенній cloud-системі з метою її модернізації, сформулювати пропозиції по зміні топології мережі та програмної архітектури активного агрегуючого обладнання, сприятимуть введенню в експлуатацію нових мережевих технологій, а також дозволять покращити параметри мультисервісної неоднорідної мережі при мінімальних економічних затратах.

Література

1. Klymash M. Service Quality Oriented Method of Multiservice Telecommunication Networks Design / Mykhailo Klymash, Orest Lavriv, Bohdan Buhyl, Yuriy Danik // 11th International Conference Modern Problems of Radio Engineering, Telecommunications and Computer Science Dedicated to the 60th Anniversary of Radio Department at Lviv Polytechnic National University TCSET'2012. February 21-24, 2012 Lviv-Slavske, Ukraine. – Publishing House of Lviv Polytechnic. – P. 235-236.
2. Strykhalyuk B., Service provisioning by using a structure stability algorithm in a virtualized data center based on cloud technology / Bogdan Strykhalyuk, Olga Shpur, Andriy Masiuk // Computational Problems of Electrical Engineering. – 2014. – Vol. 4, №1. – P.83-88.
3. Klymash M. A Novel Approach of Optimum Multi-criteria Vertical Handoff Algorithm for Heterogeneous Wireless Networks / M. Klymash, B. Stryhaluk, I. Demydov, M. Beshley, M. Seliuchenko // International Journal of Engineering and Innovative Technology (IJEIT). – November 2014. – Volume 4, Issue 5. – P. 42-52.
4. Beshley M. Research and Development the Methods of Quality of Service Provision in Mobile Cloud Systems / M. Beshley, T. Maksymyuk, B. Strykhaluk, M. Klymash // IEEE International Conference [Black Sea Conference on Communications and Networking (BlackSeaCom'2014)], May 27-30, 2014, Odessa, Ukraine. – 2014. – P. 165-169.
5. Klymash M. Model of network resources management on the basis of services priorities association / Mykhailo Klymash, Mykola Beshley, Orest Lavriv // Proceedings of international conference CADSM'2013. Polyana-Svalyava. – 2013. – P. 172-173.
6. Trang D. D. Fractal Analysis and Modeling of VoIP Traffic // D. D. Trang, B. Sonkoly, S. Molnar // Telecommunications Network Strategy and Planning Symposium. – 2004. – Issue №1. – P. 123-130.
7. Crovella M. E. Self-similarity in world wide web traffic: Evidence and possible causes / M. E. Crovella, A. Bestavros // IEEE/ACM Trans. Networking. –1997. – Vol. 5, № 6. – P. 835-846.