УДК 681.324

А. Г. Шепеленко, студ.;
І. В. Дегтяренко, к. т. н., доц.

СПОСІБ ОЦІНКИ ПАРАМЕТРА ХЕРСТА ДЛЯ ТРАФІКУ ІР-МЕРЕЖ

З урахуванням того, що Інтернет трафік та трафік ІР-мереж виявляє самоподібність та довгострокову залежність, запропоновано заст осувати модифіковану дисперсію Алана (МДА) для визначення параметра Херста довгостроково залежних потоків даних. Метод було застосовано до псевдо випадкових самоподібних послідовностей даних, що були сформовані для заданих значень параметра самоподібності Н. Також було досліджено поведінку МДА для найпоширеніших детермінованих сигналів. Даний спосіб застосовано до реального каналу ІР-трафіку, результати чого свідчать про можливість використання МДА для аналізу трафіку.

Вступ

Стрімкий розвиток телекомунікаційних систем визначає нові підходи і погляди на аналіз трафіку. Неадекватність традиційних моделей ставить задачу розробки нових методів оцінки характеристик трафіку в сучасних мережах зв'язку. Те, що потік даних не описується класичною Пуассонівською моделлю, вперше вдалося помітити в мережах Ethernet [1]. Спираючись на велику кількість експериментальних даних і ретельний статистичний аналіз [2], було доведено, що трафік ІР-мереж і Інтернет трафік, зокрема, характеризується такими кореляційними властивостями, як самоподібність і довгострокова залежність. В ході нещодавніх досліджень, проведених за підтримки Лос-Аламосської національної лабораторії [3], виявилося, що потоки даних, які спочатку не проявляють властивостей самоподібності, пройшовши обробку на вузлових серверах і активних мережевих елементах, починають подавати яскраво виражені ознаки самокорреляції. Протилежно припущенням класичної моделі Пуассона, ці властивості підкреслюють довгострокову залежність у часі між надходженнями пакетів.

Основною *метою роботи* є розробка дослідження властивостей самоподібності та довгострокової залежності трафіку ІР-мереж, розробка алгоритму оцінки параметра Херста. Актуальним на даний момент є вироблення рекомендацій та визначення напрямків подальшого застосування результатів дослідження.

Випадковий процес $X(t)$ $(t \in \Re)$ вважається самоподібним з масштабуючим параметром самоподібності або параметром Херста $H > 0$, якщо

$$X(t) \overset{d}{=} a^{-H} X(at) \tag{1}$$

для всіх $a > 0$, де $\overset{d}{=}$ означає рівність для всіх кінцевих розподілів. Іншими словами, статистична характеристика процесу $X(t)$ не змінюється при масштабуванні за амплітудою на a^{-H} і за часом на a.

Довготривала залежність процесу визначається показником ступеня загасання автокореляційної або еквівалентної функції спектральної густини потужності [4]. Процес $Y(t)$ характеризується довгостроковою залежністю, якщо його спектральна густина потужності (СГП) спадає асимптотично

$$S_\gamma(f) \sim c_2 |f|^{-\gamma} \text{ для } f \to 0, \tag{2}$$

де $0 < \gamma < 1$ — показник ступеня загасання, причому

$$\gamma = 2H - 1. \tag{3}$$

Стосовно вимірювання трафіку, параметр Херста для довгостроково залежної послідовності даних (біт в одиницю часу) фактично характеризує самоподібність початкової послідовності біт. Якщо розглядати мережевий трафік, то під самоподібністю мається на увазі повторюваність розподілу навантаження у часі у різних масштабах. Таким чином, IP-трафік виглядає майже однаково як на великих інтервалах (година, доба), так і на малих відрізках часу (секунди, мілісекунди).

Постановка задачі

Задачею роботи є розробка алгоритму оцінки параметра самоподібності H (параметра Херста) стосовно аналізу трафіку IP-мереж. Пропонується спосіб оцінки параметра Херста для довгостроково залежного трафіку за допомогою модифікованої дисперсії Алана (МДА).

Нехай дана нескінченна послідовність відліків $\{x_k\}$ сигналу $x(t)$, рівномірно розподілених у часі з кроком дискретизації τ_0. Якщо виділити кінцевий набір із N відліків $\{x_k\}$, з періодом дискретизації τ_0, оцінка МДА може бути виконана з використанням стандартної оціночної функції, що її рекомендовано ITU-T [5]

$$\mathrm{Mod}\ \sigma_y^2(n\tau_0) = \frac{1}{2n^4\tau_0^2(N-3n+1)} \sum_{j=1}^{N-3n+1} \left[\sum_{i=j}^{n+j-1} \left(x_{i+2n} - 2x_{i+n} + x_i \right) \right]^2, \tag{4}$$

де $n = 1,\ 2,\ \ldots, [\mathrm{N}/3]$.

Існує рекурсивний алгоритм швидкого розрахунку даної оціночної функції, що зменшує кількість необхідних для обчислення МДА операцій для всіх можливих значень n до $\sim N^2$, замість $\sim N^3$ [6].

Розглянемо випадковий процес, СГП якого задається як модель

$$S_x(t) = \begin{cases} \sum_{i=1}^{P} h_{\alpha_i} f^{\alpha_i}, \text{ якщо } 0 < f \le f_h; \\ 0, \text{ якщо } f > f_h, \end{cases} \tag{5}$$

де P — кількість типів ступенів шуму, що розглядаються в моделі; α_i і h_{α_i} — параметри моделі ($\alpha_i,\ h_{\alpha_i} \in \Re$); f_h — верхня межа частоти.

Для будь-якого значення з ряду збіжності МДА підкоряється ступеневому закону на інтервалі спостереження τ

$$\mathrm{Mod}\ \sigma_y^2(\tau) \sim A_\mu \tau^\mu, \tag{6}$$

де $\mu = -3 - \alpha$.

Якщо $P > 1$, то (6) можна узагальнити складанням ступенів $A_{\mu_i}\tau^{\mu_i}$.

Важливо зрозуміти поведінку МДА, як і для будь-якого іншого способу аналізу довгостроково залежних і самоподібних процесів, коли процес $x(t)$ не може бути промодельовано за допомогою формули (5). Розглянемо поведінку МДА, коли $x(t)$ включає детерміновані компоненти такі як, наприклад, зміщення щодо нульового середнього значення, лінійний або квадратичний зсув, періодичні складові та імпульси.

1) *Зсув і зміщення*. Поява в початковому інформаційному сигналі додаткових складових може бути викликана перешкодами в каналі передачі даних. Припустимо, що сигнал включає зміщення середнього значення і лінійний та квадратичний зсув, тобто $x(t) = A + At + Ct^2$. При підстановці в (4) отримаємо

$$\mathrm{Mod}\ \sigma_y^2(\tau) = 2C^2\tau^2. \tag{7}$$

Таким чином, $\mathrm{Mod}\ \sigma_y^2(\tau)$ не залежить від постійного зсуву і лінійного зміщення, але виявляє квадратичний зсув у вхідних даних, який пропорційний $\sim\tau^2$.

2) *Періодичний сигнал*. Припустимо, що $x(t)$ — чиста синусоїда з частотою f_m, тобто $x(t) = A\sin 2\pi f_m t$, із СГП $S_x(f) = \left(A^2/2\right)\delta\left(f - f_m\right)$. Тоді підставляючи в (4), отримаємо (для для $n \to \infty,\ n\tau_0 = \tau$)

$$\mathrm{Mod}\ \sigma_y^2(\tau) = A^2 \frac{\sin^6 \pi\tau f_m}{\left(\pi\tau f_m\right)^4}. \tag{8}$$

Отже, графік МДА є пульсаціями з періодом $2/f_m$.

3) *Імпульси*. Найпоширеніші приклади нестаціонарності трафіку IP-мереж — раптові зміни середньої бітової швидкості, наприклад, через перемаршрутізацію трафіку або регулювання пропускної спроможності каналу. Припустимо, що $x(t)$ — імпульс в момент $t = 0$, $x(t) = Au(t)$, отримаємо

$$\mathrm{Mod}\ \sigma_y^2(\tau) = 0. \tag{9}$$

Одиночний імпульс в сигналі $x(t)$ виявляє нульову дію на $\mathrm{Mod}\ \sigma_y^2(\tau)$, визначену за усереднювання на нескінченному тимчасовому відрізку.

Розглянемо довгостроково залежний процес із СГП (5), що характеризується параметром Херста $0,5 < H < 1$. У загальному випадку, згідно [7], параметр Херста для серії IP-пакетів $\{x_k\}$ можна визначити таким чином:

1) розрахувати $\mathrm{Mod}\ \sigma_y^2(\tau)$ за допомогою оцінної формули (3) на підставі послідовності даних $\{x_k\}$ для зростальних цілих значень $\{1 \le n \le N/3\}$;

2) визначити середній кут нахилу в подвійному логарифмічному масштабі для $n > 4$ шляхом апроксимації кривої прямою лінією;

3) перевірити, що $-3 < \mu < -2$ і отримати значення параметра Херста як $H = (\mu/2) + 2$.

Достовірність і точність запропонованого методу була перевірена на моделі. Спосіб оцінки параметра Херста на основі МДА був застосований до псевдовипадкових довгостроково залежних послідовностей даних $\{x_k\}$ завдовжки N, для генерації яких використовувався алгоритм [8]. Він оснований на спектральному перетворенні: вектор випадкових комплексних чисел із середньою амплітудою, рівною квадратному кореню з бажаного значення $S_x\left(f_k\right)$, і фазою, рівномірно розподіленою на інтервалі $[0,\ 2\pi]$, піддається зворотному перетворенню Фур'є.

Для кожного значення параметра Херста з ряду $\{H_i\} = \{0{,}50;\ 0{,}55;\ 0{,}60;\ \ldots;\ 1{,}00\}$ сформовані псевдовипадкові послідовності завдовжки 1024 відліків. До кожної з послідовностей було застосовано запропонований алгоритм і обчислено значення H, а також похибку методу Δ^H як відношення різниці між бажаним (вхідним) і фактичним (отриманим) значеннями до вхідного значення параметра самоподібності H. Як показано на рис. 1, відносна похибка обчислення в середньому не перевищує 5 %.

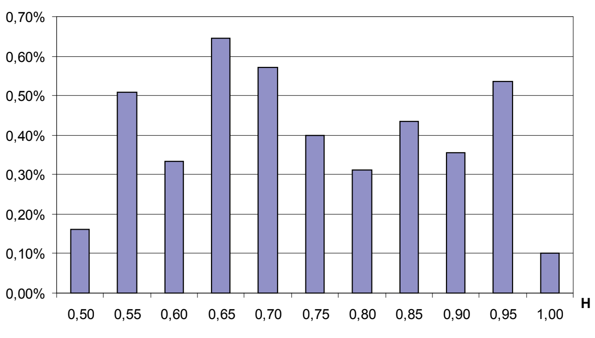

Рис. 1. Похибка методу МДА

З метою вивчення основних статистичних властивостей реального трафіку проведено аналіз експериментальних даних. Об'єктом вивчення було обрано магістраль 100 Мбіт/с Інтернет провайдера.

Статистичні дані роботи трансокеанічного Інтернет-каналу за перший квартал 2006 року отримано за допомогою online-проекту MAWI (Measurement and Analysis on the WIDE Internet) [9]. Основною методикою отримання початкових даних було перехоплення IP-пакетів, що надходять (за допомогою утиліти tcpdump).

За вхідну послідовність даних взято кількість

IP-пакетів, переданих в одиницю часу (рис. 2), розраховано МДА і визначено середній нахил у подвійному логарифмічному масштабі (рис. 3).

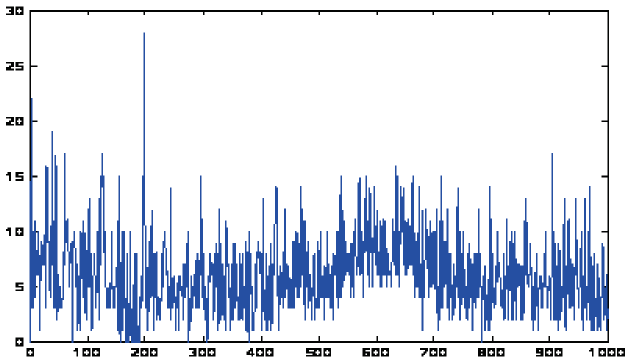

Рис. 2. Число IP-пакетів в одиницю часу (1мс)

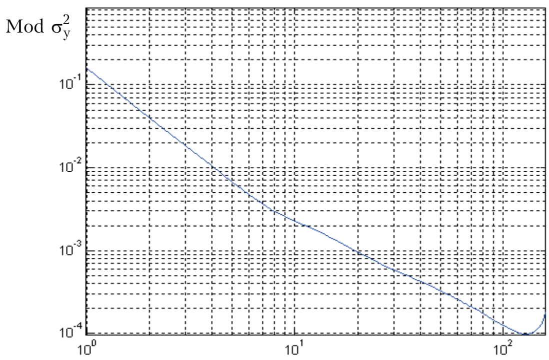

Рис. 3. МДА для реального трафіку

Визначення параметра Херста підтверджують гіпотезу про самоподібність — $H = 0,8$.

Висновки

Отже, розподіл трафіку на магістральному вузлі провайдера Інтернет близький до Пуассонівського, але таким не є. Потік даних є самоподібним. Самоподібність можна розцінювати як фундаментальну статистичну властивість мережевого трафіку, яку необхідно враховувати на практиці.

Практичні результати можуть бути використані в різних інженерних аспектах, наприклад:

— моделювання роботи телекомунікаційних мереж під час їх проектування;

— розрахунок пропускної здатності каналів зв'язку та обсягу буферів комутаторів;

— визначення показників якості обслуговування в мережах передавання даних (QoS);

— виявлення аномалій роботи мереж.

На відміну від інших методів (наприклад, за формулою Уїттла [8]), спосіб оцінки параметра Херста на основі МДА дозволяє точніше обчислити значення H. Крім того, даний метод дозволяє визначити наявність компонент фракційного шуму в трафіку IP-мереж. У підсумку можна зазначити, що даний спосіб може бути використаний спільно з іншими алгоритмами, такими як метод логарифмічної діаграми на основі вейвлет-перетворень.

СПИСОК ЛІТЕРАТУРИ

1. Leland W., Taqqu M., Willinger W., Wilson D. On the Self-Similar Nature of Ethernet Traffic (Extended Version). // IEEE/ACM Transactions on Networking, February 1994.
2. Столингс В. Современные компьютерные сети. — С-Пб.: Питер, 2003. — 783 с.
3. Митилино С. «Фрактальная катастрофа» TCP/IP / Режим доступа: http://itc.ua/article.phtml?ID=5571&IDw=49&pid=19
4. Крылов В. В, Самохвалова С. С. Теория телетрафика и ее приложения. — С-Пб.: ВХВ — Петербург, 2005. — 288 с.
5. ITU-T Rec. G.810. Definitions and Terminology for Synchronization Networks. — Geneva, 1996—2003.
6. Bregni S. Chapter 5 — Characterization and Modeling of Clocks // Synchronization of Digital Telecommunications Networks. Chichester, UK, 2002, PP. 203—281.
7. Allan D. W., Barmes J. A. A Modified Allan Variance with increased Oscillator Characterization Abitity. – Proc. 35th Annual Frequency Control Symposium, 1981.
8. Paxson V. Fast Approximation of Self-Similar Network Traffic // ACM/SIGCOMM Computer Communication Review. Vol. 27, No. 7, Oct. 1997, P. 5—18.
9. MAWIproject / Режим доступу: http://tracer.csl.sony.co.jp/mawi/.

Матеріали статті рекомендовані до опублікування оргкомітетом III Міжнародної науково-технічної конференції «Сучасні проблеми радіоелектроніки, телекомунікацій та приладобудування (СПРТП-2007)» (31.05—2.06.2007 р.)

Надійшла до редакції 30.09.07
Рекомендована до друку 04.10.07

Шепеленко Анна Геннадіївна — студентка факультету комп'ютерних інформаційних технологій і автоматики, ***Дегтяренко Ілля В'ячеславович*** — доцент кафедри автоматики та телекомунікацій

Донецький національний технічний університет