

УДК 519.876.5: 004.738.5

І.І. Маракова<sup>1</sup>, М.В. Потапов<sup>2</sup><sup>1</sup>Вища національна телекомунікаційна школа Франції, Брест<sup>2</sup>Одеський Національний Політехнічний Університет, Одеса

## УДОСКОНАЛЕННЯ ТОЧНОСТІ ГРАВИТАЦІЙНОГО МЕТОДУ ОЦІНКИ ТРАФІКУ ІНТЕРНЕТ-МЕРЕЖИ

Запропоновано визначення реальних об'ємів інформації між парами абонентів на основі матриці трафіку. Розглянуто гравітаційну модель на прикладі однорангової (пірингової) мережі. Удосконалено гравітаційну модель шляхом врахування специфічної для мережі інформації у алгоритмі оцінки параметрів.

**Ключові слова:** матриці трафіку, пірингові мережі, транзитний піринговий трафік, зовнішній трафік, внутрішній трафік, вузли мережі, гравітаційна модель.

### Вступ

У теперішній час усе більшого значення набуває аналіз і прогнозування інформаційних потоків. Але обчислювальна складність обробки не дозволяє сподіватися на те, що в найближчому майбутньому буде можливо відслідковувати передачу інформації між абонентами. Практичну цінність захисту інформаційно-фінансових потоків, що існують у структурах типу банк-банк або банк-клієнт, а також всіх інших, котрі мають справу з проходженням фінансової інформації через канал зв'язку, важко переоцінити.

В останні п'ять років відбувається бурхливий розвиток алгоритмів, що здатні досить надійно оцінити розподіл кількості інформації, що пройшла між окремими парами абонентів, на основі тільки відсіків завантажень зв'язків у мережі. Детальні дані про розподіл об'ємів інформації у мережі мають дуже велике значення. По-перше, така інформація дозволяє раціональніше використовувати ресурси мережі. По-друге, подібний аналіз дозволяє визначити не тільки факт перевантажень мережі, а й їх причини. Це дає можливість провайдеру підвищити якість послуг. По-третє, повна картина розподілу інформації мережі у часі дозволяє більш вигідно планувати цінову політику провайдера, проводити прогнозування попиту на послуги.

Визначення реальних об'ємів інформації між парами абонентів практично неможливо здійснити прямим методом, особливо у великих мережах. Розглянута проблема у літературі отримала назву „оцінка матриці трафіку” (traffic matrix estimation) [1 – 4]. Під матрицею трафіку розуміють об'єм трафіка між кожною парою вузлів в мережі. Це може бути як поточна швидкість передачі, так і середня за деякий інтервал. Матриця може бути як одновимірною (вектор), так і двовимірною, у залежності від того, який варіант простіше використовувати у тому чи іншому алгоритмі її розрахунку. Існує кілька теоре-

тичних підходів рішення даної задачі, точність іноді досягає 90 – 95% [1 – 4]. Але більшість з них орієнтована на корпоративні мережі невеликих розмірів, точність цих методів так само оцінюється на основі експериментальних даних обмежених локальних мереж.

### Основна частина

У великих Інтернет-провайдерів більшість трафіку обмінюється між користувачами та вузлами мережі. Трафік користувачів та вузлів мережі якісно різний, і цей факт сильно впливає на матрицю трафіку. Більш того, трафік вузлів мережі має великий вплив на кожний аспект проектування та організації Інтернет-мережі, тому визначення матриці трафіку має дуже велике значення. З урахуванням цього, гравітаційну модель потрібно адаптувати під трафік користувачів та трафік між базовими вузлами. Отриману просту гравітаційну модель можна узагальнити введенням додаткової інформації, отриманої з матриці роутингу та завдяки класифікації каналів.

Розглянемо гравітаційну модель на прикладі пірингової мережі (рис. 1). Припустимо, що мережа має набір вузлів  $P_1, P_2, \dots$ , і відбувається обмін трафіком вузла  $P_i$  через набір зовнішніх зв'язків, виділених для вузла  $P_i$ . Таку ситуацію часто називають внутрішнім пірингом (private peering) – це найхарактерніший режим роботи базових вузлів. Нехай мережа має канали доступу користувачів  $a_1, a_2, \dots$ , та набір пірингових каналів  $p_1, p_2, \dots$ . Позначимо набір зовнішніх каналів, які переносять трафік до вузла  $P_i$  як  $\pi_i$ , а набір усіх пірингових каналів як  $\pi$ . Набір усіх каналів доступу користувачів позначимо за  $\Lambda$ . Заміри SNMP-завантаження дають значення трафіку на усіх зовнішніх каналах  $j: x_{link}^{in,out}(j)$ , де індекс  $in(out)$  вказує на трафік, який входить (виходить) з

базового вузла. Вхідний (вихідний) трафік відносно вузла  $P_i$  :

$$x_{peer}^{in,out}(P_i) = \sum_{p \in \pi_i} x_{link}^{in,out}(p). \quad (1)$$

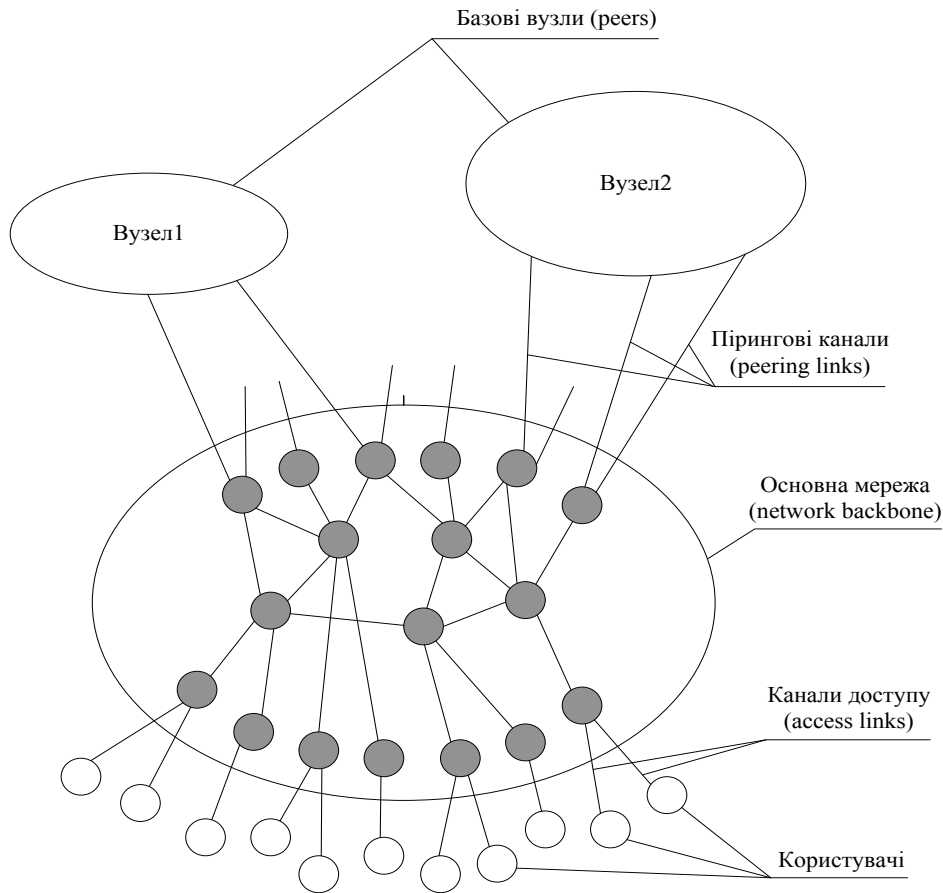


Рис. 1. Типова структура пірингової мережі

Основні поняття моделі: Транзитний піринговий трафік (transit peer traffic) – трафік, який проходить через центральний вузол з одної мережі в іншу. Його об’єм незначний, ним можна знехтувати. Трафік, що відправляється назовні (outbound traffic) – трафік з каналу користувача в піринговий канал. Будемо приймати до уваги те, що трафік, який приходить в піринговий канал з каналів доступу користувачів, пропорційний об’єму трафіка кожного з цих каналів. Позначимо вихідний піринговий канал для трафіка, який йде від каналу доступу  $a_i$  в базовий вузол  $P_j$  як  $X(a_i, P_j)$ . Ці дані можна отримати, маючи конфігурацію роутера. Припущення пропорційності на практиці частіше за все виконується, крім таких випадків, як короточасне врівноваження завантажень мережі (перерозподілення завантаження між каналами). У таких ситуаціях розглянутий нижче метод може бути доповнений доступною статистикою. Трафік, що надходить ззовні (inbound traffic) – трафік з пірингового каналу в канал користувача. Провайдер не має можливості в значній мірі контролювати трафік, який надходить у дані мережу

з інших пірингових мереж. Відповідно, ми будемо припускати, що трафік, який надходить з даного пірингового каналу розподіляється по каналам доступу користувачів пропорційно об’єму трафіка, що відправляється. Внутрішній трафік (internal traffic) – трафік з одного каналу доступу в інший канал доступу у межах даної пірингової мережі. Припустимо, що частка внутрішнього трафіка з даного каналу доступу  $a_i$  в інший канал доступу  $a_j$  пропорційна сумарному об’єму трафіка, який входить у мережу через  $a_i$ , і розрахуємо трафік між каналами шляхом нормалізації.

Враховуючи вище вказане, вихідний зовнішній трафік з каналу доступу  $a_i \in \Lambda$  до пірингового каналу  $p_m \in \pi_i$  є:

$$x_{outbound}(a_i, p_m) = \begin{cases} \frac{X_{link}^{in}(a_i)}{\sum_{a_k \in \Lambda} X_{link}^{in}(a_k)} X_{peer}^{out}(P_j), & \text{для } p_m = X(a_i, P_j), \\ 0, & \text{для } p_m = X(a_i, P_j). \end{cases} \quad (2)$$

Вхідний трафік з пірингового каналу  $p_i$  до каналу доступу  $a_j$  :

$$x_{inbound}(p_i, a_j) = X_{link}^{in}(p_i) \frac{X_{link}^{out}(a_j)}{\sum_{a_k \in \Lambda} X_{link}^{out}(a_k)}. \quad (3)$$

Внутрішній трафік з каналу доступу  $a_i$  до каналу доступу  $a_j$  :

$$x_{internal}(a_i, a_j) = \frac{X_{link}^{in}(a_i)}{\sum_{a_k \in \Lambda} X_{link}^{in}(a_k)} X_{internal}^{out}(a_j), \quad (4)$$

де

$$X_{internal}^{out}(a_j) = X_{link}^{out}(a_j) - \sum_{p_k \in \pi} X_{inbound}(p_k, a_j) = X_{link}^{out}(a_j) \left( 1 - \frac{\sum_{p_k \in \pi} X_{link}^{in}(p_i)}{\sum_{a_k \in \Lambda} X_{link}^{out}(a_k)} \right).$$

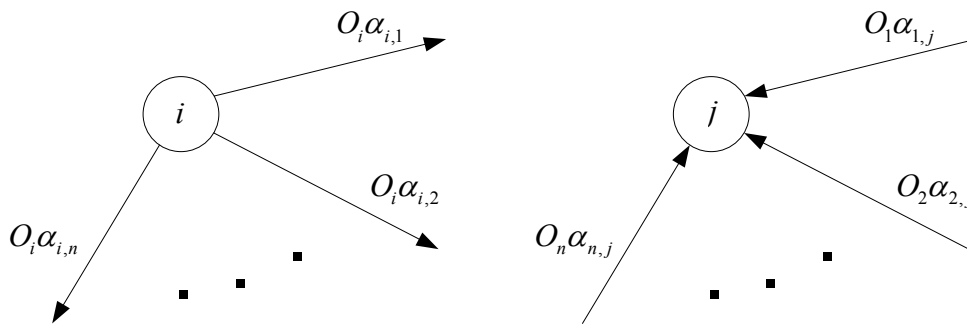


Рис. 2. Корегуючі коефіцієнти для вхідного/вихідного трафіка

Модель може бути розширена до загального випадку, коли маємо більше двох рішень. У такому разі ймовірність прийняття рішення  $j$  можна записати у вигляді:

$$P_C^i(j) = \frac{e^{V_j^i}}{\sum_{k \in C} e^{V_k^i}}, \quad (6)$$

де  $C$  – набір можливих рішень.

Ймовірність, записана у виразі (6), відповідає корегуючому фактору  $\alpha_{ij}$ , який необхідно було знайти. Враховуючи вираз (6), початкова гравітаційна модель прийме вигляд:

$$X_{ij} = O_i \frac{e^{V_j^i}}{\sum_{k \in S} e^{V_k^i}}, \quad (7)$$

Обчислювальна складність моделі (1) – (4) невелика.

Змінимо основний закон гравітаційної моделі. Запишемо вираз для долі загального трафіка  $X_{ij}$ , який виходить з вузла  $i$  :

$$X_{ij} = O_i \alpha_{ij}, \quad (5)$$

де  $O_i = \sum_j X_{ij}$  – загальний обсяг трафіку, який залишає вузол  $i$ ;  $\alpha_{ij}$  – параметр, який враховує коефіцієнт протиріччя, який регулює взаємодію між вузлами  $i$  та  $j$ .

Будемо вважати  $O_i$  фактором відштовхування, який відображає об'єм трафіка, який може згенерувати вузол. Цей об'єм відповідає сумі вихідного трафіка по усім каналам, які під'єднані до вузла  $i$ . Ми об'єднали фактори взаємного відштовхування та притягання в параметр  $\alpha_{ij}$ . На рис. 2 зображено принцип, яким чином вхідний та вихідний трафік відображається у рамках попередньо введених термінів.

Будемо вважати  $O_i$  фактором відштовхування, який відображає об'єм трафіка, який може згенерувати вузол. Цей об'єм відповідає сумі вихідного трафіка по усім каналам, які під'єднані до вузла  $i$ . Ми об'єднали фактори взаємного відштовхування та притягання в параметр  $\alpha_{ij}$ . На рис. 2 зображено принцип, яким чином вхідний та вихідний трафік відображається у рамках попередньо введених термінів.

де  $O_i$  – сумарний об'єм трафіка, що надходить з вузла  $i$  у мережу;  $S$  – набір можливих рішень.

Детермінована складова функції доцільності представляє відносну „зацікавленість” окремого вузла до набору потенційних поглинаючих вузлів. Найпростіший спосіб обчислити цю складову – припустити, що ймовірність вибору вузла  $j$  пропорційна сумарному об'єму трафіка, який надходить до нього. Позначимо цю величину за  $D_j$  :

$$D_j = \sum_i X_{ij}. \quad (8)$$

Відліки  $D_j$  можна легко розрахувати, так як вони відповідають сумі вхідного SNMP-трафіку для усіх каналів, під'єднаних до вузла  $j$  ( $O_i$ ) відповідає вихідному SNMP-трафіку). У цьому контексті доці-

льність вибору вузлом  $i$  вузла  $j$ , як об'єкта призначення, буде визначатися:

$$V_j^i = \mu_i D_j + \gamma_j, \quad (9)$$

де  $\mu_i$  – коефіцієнт врахування фактору  $D_j$ ;  $\gamma_j$  – масштабуючий коефіцієнт, враховуючий зацікавленість вузла  $i$  вузлом  $j$ , не враховану фактором  $D_j$ .

Фактично  $V_j^i$  – детермінована складова доцільності, і вона також представляє фактор зацікавленості у гравітаційній моделі.

Отже, у гравітаційну модель було введено додаткові параметри, які дозволяють отримати більш реалістичну, ймовірнісну модель функції доцільності. Введення додаткової інформації щодо частково відомої матриці трафіку суттєво підвищує точність розрахунку, дозволяючи провести необхідне масштабування результату. Використання ітераційних процедур методу правдоподібності та максимізації очікування майже на порядок підвищує точність алгоритму порівняно з простою гравітаційною моделлю.

Незважаючи на те, що на кожну пару джерел-одержувач необхідно оцінити невелику кількість параметрів, модель усе ж таки ускладнюється і підвищується ризик отримати неправильне рішення, але задовольняюче усім критеріям методу. Для деяких типів мереж обчислювальна важкість теж може стати проблемою.

Розширимо базову гравітаційну модель шляхом врахування специфічної для мережі інформації у алгоритмі оцінки параметрів. Введення такої інформації можна здійснити за рахунок моделювання детермінованих складових функції доцільності як комбінації  $m$  уточнювальних параметрів,  $E_{ik}$ ,  $k = 1, \dots, m$ , відповідно кожному вузлу. Уточнювальні параметри відповідають фактичним властивостям вузлів і включають такі характеристики, як кількість роутерів, кількість направлених каналів, кількість серверів даних та хостінгових серверів, кількість каналів користувачів, та інші.

З урахуванням вище вказаних параметрів, детермінована компонента функції доцільності для вузла  $i$  може бути представлена у наступному вигляді:

$$V_j^i = \mu_{i1} E_{i1} + \mu_{i2} E_{i2} + \dots + \mu_{im} E_{im} + \gamma_j, \quad (10)$$

де  $\mu_{ik}$  – вагові коефіцієнти, враховуючі відносну важливість кожного уточнювального параметра.

Параметри моделі, яка буде оцінюватися, тепер включає для кожного вузла параметри  $\mu_{ik}$  та  $\gamma_j$ . У цьому випадку усі раніше розглянуті моделі можуть використовуватися з параметром, представленим у такій формі.

Якщо маємо  $n$  вузлів та  $m$  уточнювальних параметрів, сумарна кількість параметрів, яку необхідно оцінити, складає  $n(m+1) = O(n)$ . Кількість параметрів зростає, але кількість параметрів залишилася лінійною по відношенню до кількості вузлів. Навіть при цьому складність по кількості параметрів не досягла складності цілком статистичних методів, яка складає  $O(n^2)$ .

Незважаючи на те, що гравітаційна модель не забезпечує високої точності результату, дослідження її на прикладі реальних даних є дуже актуальною задачею, адже ця модель використовується для ініціалізації багатьох алгоритмів.

У процесі дослідження гравітаційного методу вдалося дещо підвищити точність алгоритму за рахунок застосування кореляційного аналізу у гравітаційній моделі. Результати моделювання основної гравітаційної моделі та її модифікації з використанням моделі імітатора приведені у табл. 1.

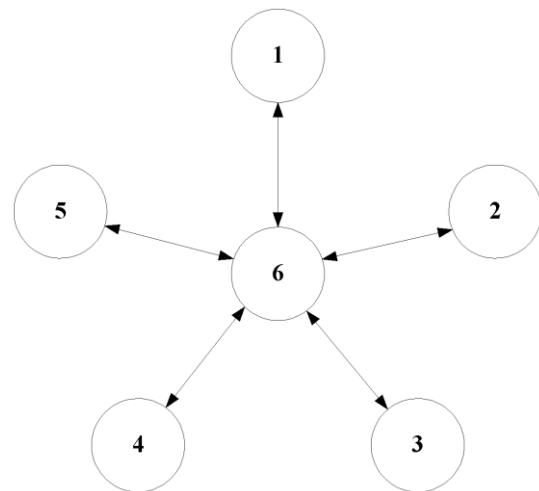


Рис. 3. Структура мережі для моделювання гравітаційної моделі

## Висновки

Аналізуючи результати, слід відзначити, що точність гравітаційного алгоритму фактичного не залежить від природи та структури Інтернет-трафіку.

Просту гравітаційну модель можна узагальнити введенням додаткової інформації, отриманої з матриці роутингу та завдяки класифікації каналів трафіку (46,36% та 47,66% для даних, наближених до реальних).

Модифікація алгоритму за рахунок введення кореляційного аналізу зміни швидкості трафіку призвела до деякого підвищення точності результату (на 20%) за рахунок зменшення пікових помилок алгоритму. Видно, що базовий алгоритм допускає помилки до 164%, у той час, як модифікований – 115%.

Дослідження точності гравітаційного алгоритму

| Номер пари ДО | Аналіз на основі випадкових даних |                       | Моделювання на основі даних імітатору |                             |
|---------------|-----------------------------------|-----------------------|---------------------------------------|-----------------------------|
|               | Типова Gravity-модель             | Типова Gravity-модель | Типова Gravity-модель                 | Модифікована Gravity-модель |
| 1             | 26,42%                            | 28,66%                | 27,06%                                |                             |
| 2             | 22,64%                            | 22,08%                | 15,66%                                |                             |
| 3             | 128,62%                           | 133,94%               | 115,31%                               |                             |
| 4             | 26,82%                            | 27,40%                | 16,87%                                |                             |
| 5             | 10,7%                             | 10,09%                | 103,43%                               |                             |
| 6             | 28,64%                            | 30,27%                | 5,79%                                 |                             |
| 7             | 45,68%                            | 42,48%                | 48,31%                                |                             |
| 8             | 8,65%                             | 10,25%                | 28,91%                                |                             |
| 9             | 49,67%                            | 47,81%                | 22,42%                                |                             |
| 10            | 29,68%                            | 32,06%                | 49,85%                                |                             |
| 11            | 35,42%                            | 37,81%                | 44,56%                                |                             |
| 12            | 154,68%                           | 164,33%               | 70,34%                                |                             |
| 13            | 31,31%                            | 33,79%                | 47,82%                                |                             |
| 14            | 68,12%                            | 66,65%                | 14,32%                                |                             |
| 15            | 71,33%                            | 72,27%                | 13,35%                                |                             |
| 16            | 21,33%                            | 25,47%                | 38,27%                                |                             |
| 17            | 13,64%                            | 10,26%                | 19,55%                                |                             |
| 18            | 98,63%                            | 107,26%               | 37,79%                                |                             |
| 19            | 44,79%                            | 42,60%                | 41,20%                                |                             |
| 20            | 10,56%                            | 7,74%                 | 31,13%                                |                             |
| Середнє:      | 46,36%                            | 47,66%                | 39,60%                                |                             |

Цей момент має істотне значення при ініціалізації інших алгоритмів. Деякі з них базуються на знаходженні екстремумів розрахункових функцій. Велика пікова помилка може призвести до того, що алгоритм змістить рішення у зону локального екстремуму і помилка ще більше зростає по відношенню до оціночного значення, отриманого гравітаційною моделлю.

### Список літератури

1. Vardi Y. *Network Tomography: Estimating Source-Destination Traffic Intensities From Link Data* / Y. Vardi // *American Statistical Association Journal*. – 1996. – № 443. – P. 365-377.
2. Tebaldi C. *Bayesian Inference on Network Traffic Using Link Count Data* / C. Tebaldi // *American Statistical Association Journal*. – 1998 – № 442. – P. 557-570.

3. Zhang Yin, Roughan M. *Fast Accurate Computation of large-scale IP traffic matrices from link loads* / Yin Zhang, M. Roughan // *AT&T labs, Shannon laboratory – Sigmetrics – 2003 – № 3 – P. 125-139.*

4. Medina Alberto. *Traffic matrix estimation. Existing techniques compared and new directions* / Alberto Medina // *Pittsburg – Sigconn – 2002. – № 8 – P. 98-112.*

Надійшла до редколегії 24.03.2009

**Рецензент:** д-р техн. наук, проф. В.А. Крисілов, Одеський національний політехнічний університет, Одеса.

### УСОВЕРШЕНСТВОВАНИЕ ТОЧНОСТИ ГРАВИТАЦИОННОГО МЕТОДА ОЦЕНКИ ТРАФИКА ИНТЕРНЕТ-СЕТИ

И.И. Маракова, Н.В. Потапов

Предложено определение реальных объемов информации между парами абонентов на основе матрицы трафика. Рассмотрена гравитационная модель на примере одноранговой (пиринговой) сети. Усовершенствованна гравитационная модель путем учета специфической для сети информации в алгоритме оценки параметров.

**Ключевые слова:** матрицы трафика, пиринговые сети, транзитный пиринговый трафик, внешний трафик, внутренний трафик, узлы сети, гравитационная модель.

### EXACTNESS IMPROVEMENT THE GRAVITY METHOD OF ESTIMATION INTERNET-NETWORK TRAFFIC

I.I. Marakova, N.V. Potapov

The real volumes determination of information is offered between the pair of subscribers on the basis of traffic matrix. It is considered gravity model on the example of unirank (peer) network. It is improved gravity model by the account of specific for a network information in the algorithm of estimation parameters.

**Keywords:** matrices of traffic, peer networks, transit peer traffic, external traffic, internal traffic, knots of network, gravitation model.