

ДІАГНОСТИКА УПЕРЕДЖЕНОСТІ ТА ПЕРЕПІДГОНКИ АЛГОРИТМІВ МАРШРУТИЗАЦІЇ В ОПОРТУНІСТИЧНИХ МЕРЕЖАХ

Запропоновано підхід до аналізу роботи протоколів маршрутизації в опортуністичних мережах як алгоритмів, отриманих в результаті машинного навчання. Розглянуто кілька таких протоколів та діагностовано в них упередження або перепідгонку.

Ключові слова: опортуністичні мережі, маршрутизація, машинне навчання.

Вступ

Сьогоднішній бурхливий розвиток мобільних технологій відкриває великі перспективи та нові шляхи у розвитку мереж передачі даних. Зокрема, з кожним роком популярнішою стає нова парадигма опортуністичних мереж, основаних на спонтанних можливостях з'єднання між вузлами. Дані можуть передаватися в мережі, навіть коли між адресатом і відправником раніше не існувало маршруту. Маршрутизація у таких мережах часто виконується без знання про повну топологію, таким чином мережі існують на основі тимчасових випадкових з'єднань між пристроями.

Останнім часом такі мережі дуже часто розглядають у контексті руху людей – носіїв мобільних телефонів. Коли пристрої опиняються на порівняно невеликій відстані один від одного, між ними може бути встановлене з'єднання. Для реалізації таких опортуністичних мереж дуже важливою є задача побудови адекватної моделі переміщення людей. У зв'язку з великою актуальністю цієї проблеми у цьому напрямі ведеться багато досліджень, огляд яких можна знайти в [2].

Особливість роботи протоколів маршрутизації в опортуністичних мережах полягає в тому, що алгоритми повинні через відсутність будь-якої статичної структури вивчати ті чи інші шаблони у тимчасових з'єднаннях для прийняття рішень про передачу повідомлення наступному вузлу. Це означає, що така маршрутизація передбачає навчання кожного з вузлів та мережі в цілому як системи. А для відлагодження роботи цієї техніки актуальним є завдання діагностики причин незадовільних результатів навчання: упередженості та перепідгонки. Перепідгонкою називатимемо явище, коли отриманий в результаті навчання алгоритм добре працює для даних, на яких здійснювалося навчання, але погано на тес-

товій вибірці. Упередженістю називатимемо явище, коли вибрана статистична модель має замалу складність для того, щоб отриманий алгоритм працював добре на навчальній вибірці.

Алгоритми маршрутизації

У цій статті змодельовано роботу п'яти алгоритмів маршрутизації в опортуністичних мережах. Усі вони виконують завдання передачі повідомлення від початкового до кінцевого вузла у мережі. При цьому кожен із вузлів може як генерувати нові повідомлення для передачі, так і тимчасово зберігати повідомлення, призначені для інших вузлів. Розглядається лише одностороння (*unicast*) передача повідомлень. Часто дані протоколи описано через імовірності передачі та реплікації повідомлення [6].

Алгоритм прямої доставки

Це один із примітивних алгоритмів, який передбачає, що передача повідомлення між початковим та кінцевим вузлами можлива лише за умови прямого контакту між ними. Тобто коли вони опиняться на такій відстані один від одного, що стане можливим утворення бездротового з'єднання між даними вузлами. Однак для багатьох мереж імовірність такої події може бути близькою до нуля.

Лавиноподібний (епідемічний) алгоритм

Алгоритм передбачає лавиноподібне поширення повідомлень. Вузол передає копію повідомлення кожному, з ким утворюється з'єднання, доки повідомлення не досягне адресата [8]. Це досить просте рішення потребує великої кількості ресурсів. По-перше, надмірна передача даних може швидко вичерпати заряд батареї, від якої живиться вузол мережі. По-друге, вузли мають обмежений об'єм пам'яті для збереження копій повідомлень, і неефективне використання

вузла може призвести до швидкого вичерпання ресурсів мережі. Тому, щонайменше, повідомлення, які передавалися між вузлами більше певної кількості разів, повинні відкидатися під час чергової спроби передачі.

Випадковий алгоритм

Ідея випадкового алгоритму лежить десь по середині між попередніми двома. Полягає вона в тому, що повідомлення передається вузлу, з яким встановлено з'єднання, але який не є адресатом, з певною заданою імовірністю – імовірністю передачі. Досить часто, моделюючи маршрутизацію з цим алгоритмом, відключають можливість реплікації повідомлення [2]. Тобто імовірність того, що в мережі існуватиме більше однієї копії повідомлення, дорівнює нулю.

Алгоритм першого контакту

Цей примітивний алгоритм підтримує не більше однієї копії повідомлення в мережі та передає його першому вузлу, з яким відбувся контакт.

Алгоритм PROPHET

Ідея епідемічного поширення повідомлень у мережі могла б бути прийнятною, якби ймовірність зустрічей (утворення з'єднань) між вузлами мала б рівномірний розподіл. Однак характер переміщення «носіїв» даних (*Data MULEs* [5]) дає протилежний результат. У багатьох випадках можна говорити про те, що ймовірність зустрічі з певним вузлом у майбутньому є більшою, якщо зустріч із ним вже відбувалася в минулому. На такому припущенні й побудовано роботу алгоритму PROPHET [4]. Він враховує історію попередніх зустрічей та передач для обчислення ймовірності доставки повідомлення заданому вузлу через певний проміжний вузол.

Коли вузол i зустрічає вузол j , імовірність доставки для вузла i вузлу j змінюється наступним чином:

$$p'_{ij} = (1 - p_{ij}) p_0 + p_{ij}$$

де p_0 – початкова імовірність доставки, параметр для певної мережі.

Якщо вузол i не зустрічає вузол j протягом певного часу, то імовірність доставки зменшується:

$$p'_{ij} = a * k * p_0,$$

де a – фактор старіння, $a \in (0;1)$, k – кількість одиниць часу, який проминув із моменту останньої зустрічі.

Транзитивні ймовірності передачі повідомлень за допомогою проміжних вузлів обчислю-

ються після обміну даними між вузлами. Наприклад, імовірність доставки вузлу z через вузол i обчислюють так:

$$p'_{iz} = p_{iz} + (1 - p_{iz}) p_{ij} p_{iz} \beta,$$

де β – параметр, який задається під час проектування мережі та визначає вплив транзитивності на імовірності доставки.

Алгоритм Spray and Wait

Розробники алгоритму *Spray and Wait* [1] мали на меті значно зменшити кількість передач повідомлень порівняно з епідемічним підходом, при цьому зберігаючи рівень кількості успішних доставок близьким до оптимального. Алгоритм складається з двох фаз:

- фаза поширення, під час якої L копій згенерованого повідомлення від відправника передають (можливо через проміжні вузли) L різними приймачам;
- фаза очікування, під час якої повідомлення, якщо воно не було доставлене під час поширення, не передається далі, окрім як самому адресатові, тобто очікує прямої передачі.

Постановка задачі та моделювання

Нетривіальні алгоритми маршрутизації в опортуністичних мережах передбачають навчання на основі даних про попередні зустрічі вузлів. Так, періоди між подіями появи з'єднання між вузлами тим, чи іншим чином впливають на прийняття рішення про передачу повідомлення в алгоритмах PROPHET та *Spray and Wait*. Варто також зазначити, що це навчання відбувається в режимі он-лайн, оскільки нові дані постійно з'являються під час функціонування мережі. Таким чином, властивості алгоритмів маршрутизації можна розглянути з точки зору навчання машин.

За низької точності роботи інтелектуальних інформаційних систем, які розв'язують задачі передбачення або класифікації, після виконання навчання здійснюють аналіз, від чого саме страждає система: упередженості чи перепідгонки. У першому випадку (упередженості) для покращення точності роботи системи необхідно змінювати модель функції, обраної для передбачення, оскільки поточна не може покрити набір вхідних даних. Збільшення об'єму вхідних даних не зможе поліпшити точність роботи системи. У другому випадку (перепідгонка), навпаки, більша кількість даних, на основі яких здійснюють навчання, позитивно позначиться на якості результатів роботи системи. Зокрема для опортуністичних мереж це означатиме, що з часом імовірність успішної доставки повідомлень зростатиме.

Для такого аналізу користуються кривими навчання, які показують залежність помилки роботи системи від об'єму вхідних тестових даних і даних для перехресної перевірки. Якщо ці криві близькі одна до одної, то система зазнає упередженості, інакше – перепідгонки [1].

Отже, для аналізу алгоритмів маршрутизації в опортуністичних мережах потрібно побудувати для них графіки залежностей імовірності неуспішної доставки повідомлення від об'єму вхідних даних, які використовують для навчання, та даних для перехресної перевірки.

Для побудови таких графіків було проведено моделювання роботи алгоритмів за допомогою симулятора ONE [3], у який внесено необхідні зміни та доповнення. У таблиці нижче наведено параметри моделювання.

Таблиця. Параметри моделювання алгоритмів маршрутизації

Модель переміщення людей	ShortestPathMapBasedMovement
Об'єм буфера повідомлень у вузлів	5 Мб
Радіус передачі повідомлення	10 м
Швидкість передачі повідомлень	250 Кб/с
Період генерації нових повідомлень	25–30 с

Використана модель переміщення людей (*ShortestPathMapBasedMovement*) передбачає випадковий вибір кінцевої точки для кожного з вузлів і подальший рух цього вузла до неї за найкоротшим маршрутом на заданій карті.

Будуючи графіки для даних, які використовували для навчання, за основу брали передачі повідомлень між вузлами, які вже зустрічалися. А при побудові графіків для даних перехресної перевірки за основу слугували передачі між тими вузлами, між якими контакту ще не було.

Результати моделювання

На рис. 1 показано зміну кількості нових зустрічей між вузлами опортуністичної мережі з часом. Для мережі, у якій кількість вузлів залишається незмінною протягом усього часу, цей графік матиме плато, значення якого рівне C_N^2 , де N – кількість вузлів у системі.

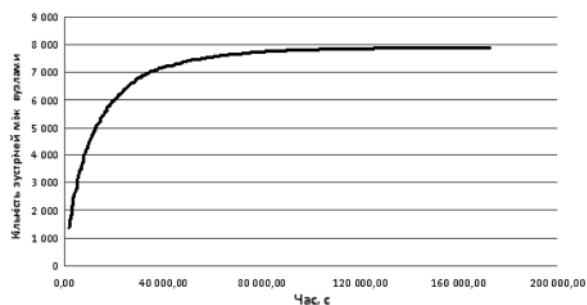


Рис. 1. Динаміка появи нових зустрічей між вузлами

Це справедливо і для моделі робочого дня, використаної у цій статті. Слід зазначити, що для більшості реальних мереж кількість вузлів не є сталою. Оскільки вузли мобільні, то вони можуть залишити мережу або ж до неї можуть увійти нові учасники. Кількість вузлів може зменшитися також через появу у них несправностей або розрядження батареї.

На більшості графіків, які будуть наведені, імовірність неуспішної передачі повідомлення для вузлів, між якими ще не було прямого контакту, є меншою за імовірність неуспішної передачі для вузлів, між якими вже відбулася зустріч. Враховуючи той факт, що передачі між «незнайомими» вузлами інтерпретують як дані перехресної перевірки, варто зауважити, що описана ситуація досить нетипова. Однак це можна пояснити за допомогою графіка динаміки генерації нових повідомлень між вузлами (рис. 2).

На рис. 2 показано кількість згенерованих повідомлень окремо для пар вузлів, між якими вже відбувався контакт, та окремо для тих, де ще не було зустрічей. Сума цих двох кривих дає лінійну функцію, оскільки повідомлення генерувалися зі сталим періодом, а відправник та адресат визначалися випадковим чином із рівномірним розподілом.

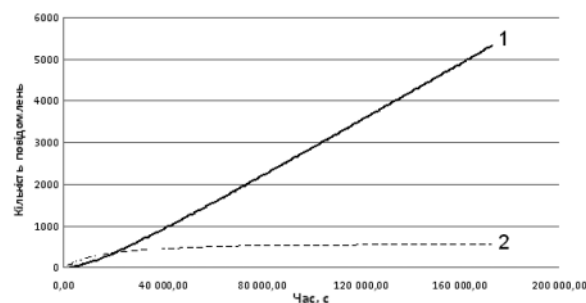


Рис. 2. Динаміка генерації повідомлень між вузлами: 1 – вузлами, з якими вже був контакт, 2 – вузлами, з якими не було контакту

Цей графік може пояснити високу імовірність успішної передачі між «незнайомими» вузлами порівняно з невеликою кількістю повідомлень, які генеруються між такими адресатом та приймачем.

На рис. 3–7 зображено криві навчання для п'яти розглянутих алгоритмів. Суцільною лінією на них показана імовірність помилки при доставці повідомлення вузлу, з яким відбувався контакт. Пунктиром – імовірність помилки для вузлів, з якими зустрічей ще не було.

Епідемічний алгоритм, алгоритми прямої доставки та першого контакту мають невелику різницю між помилками передачі для «знайомих» і «незнайомих» вузлів. Це свідчить про те, що ці алгоритми зазнають упередженості.

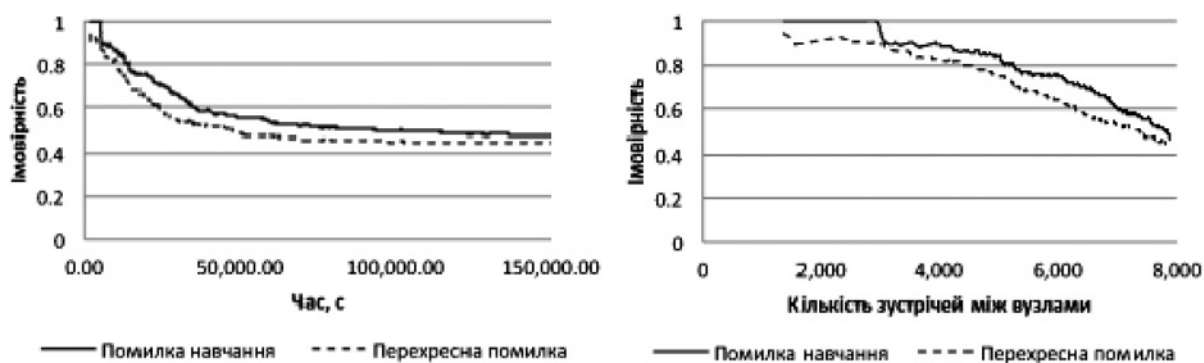


Рис. 3. Криві навчання для алгоритму прямої доставки

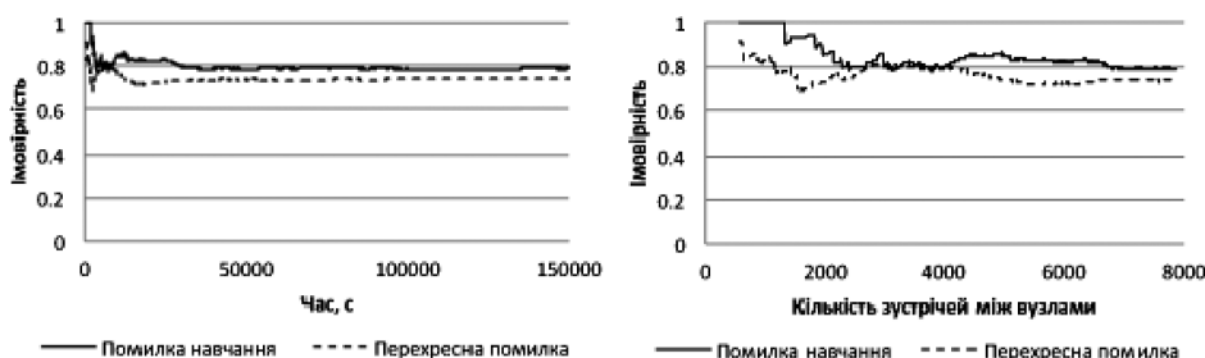


Рис. 4. Криві навчання для епідемічного алгоритму маршрутизації

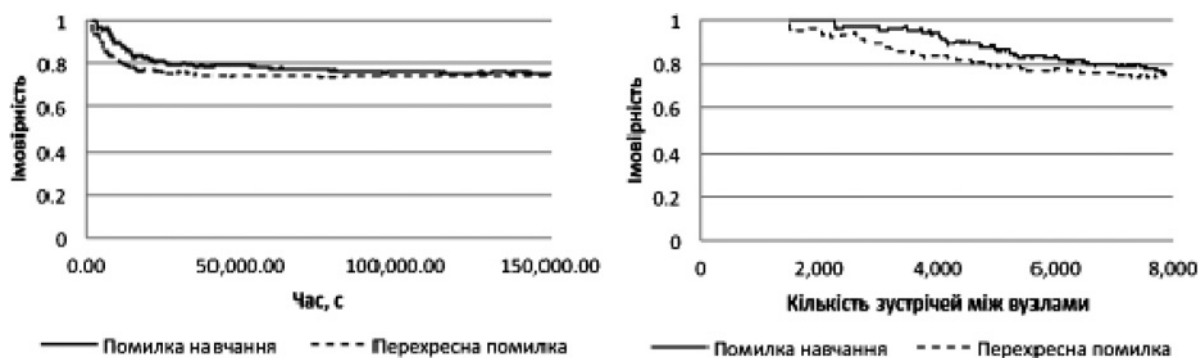
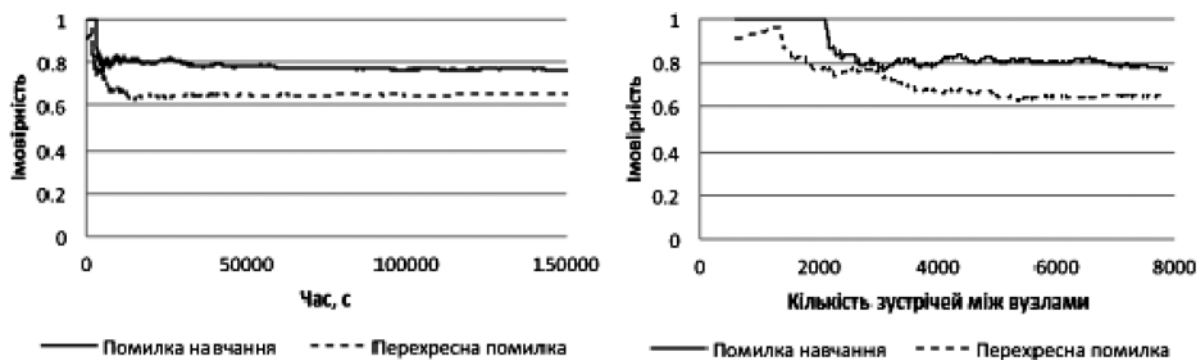
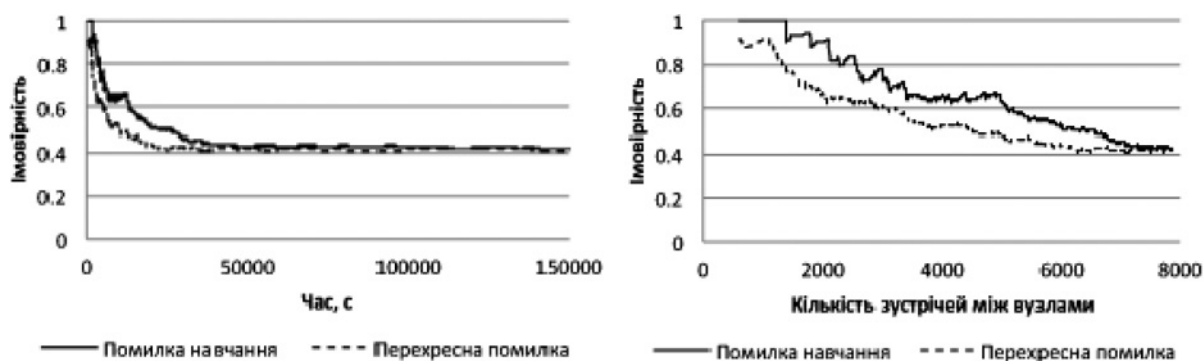


Рис. 5. Криві навчання для алгоритму першого контакту

Рис. 6. Криві навчання для алгоритму *PROPHET*

Рис. 7. Криві навчання для алгоритму *Spray and Wait*

Графіки алгоритму *Spray and Wait* показують, що для нього справджується велика різниця між помилками передачі для «знайомих» та «незнайомих» вузлів, що свідчить про те, що він страждає від перепідгонки. На кривих також видно, що з продовженням навчання імовірність помилки зменшується, а також зменшується і різниця між різними типами помилок.

Графіки алгоритму PROPHET мають дещо суперечливий характер. З одного боку, на них видно порівняно велику різницю між різними типами помилок, що притаманно алгоритмам, які зазнають перепідгонки, з другого – із продовженням навчання ймовірність помилки не зменшується. Це разом із фрагментом кривих навчан-

ня з практично рівними імовірностями помилок свідчить про упередженість.

Висновки

У більшості алгоритмів маршрутизації в опортуністичних мережах здійснюється навчання на даних про попередні зустрічі з вузлами. Це актуалізує завдання діагностики в них упередженості або перепідгонки. П'ять алгоритмів проаналізовано з точки зору машинного навчання. Для визначення упередженості чи перепідгонки використано підхід порівняння рівнів помилок для набору даних, на якому здійснювалося навчання, та для набору даних перехресної перевірки.

Література

- Guyon I. A practical guide to model selection / I. Guyon ; [in J. Marie, editor] // Machine Learning Summer School. – Springer. – 2009.
- Karamshuk D. Human mobility models for opportunistic networks / D. Karamshuk, C. Boldrini, M. Conti, A. Passarella // IEEE Commun. Mag. – 2011.
- Keränen A. The ONE Simulator for DTN Protocol Evaluation / A. Keränen, J. Ott, T. Kärkkäinen // SIMUTools'09 : 2nd International Conference on Simulation Tools and Techniques. – Rome. – 2009.
- Lindgren A. Probabilistic Routing in Intermittently Connected Networks / A. Lindgren, A. Doria, O. Schele'n // SIGMOBILE Mobile Computing Comm. Rev. – 2003. – Vol. 7, No. 3. – P. 19–20
- Shah R. C. Data MULEs: Modeling a Three-tier Architecture for Sparse Sensor Networks / R. C. Shah, S. Roy, S. Jain, W. Brunette // IEEE SNPA Workshop. – 2003.
- Song L. Routing in Mobile Opportunistic Networks / Libo Song, David F. Kotz // Mobile Opportunistic Networks. – CRC Press, 2011. – P. 1–24.
- Spyropoulos T. Spray and wait : an efficient routing scheme for intermittently connected mobile networks / T. Spyropoulos, K. Psounis, C. S. Raghavendra // In Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking (WDTN '05). – ACM, New York, NY, USA. – P. 252–259.
- Vahdat A. Epidemic routing for partially-connected ad hoc networks / Amin Vahdat, David Becker // Technical Report CS-2000-06. – Duke University, 2000.

S. Gorokhovskiy, R. Mazur

BIAS AND OVERFITTING DIAGNOSTICS IN OPPORTUNISTIC NETWORKS ROUTING ALGORITHMS

The paper suggests an approach to opportunistic networks routing protocols analysis as algorithms derived from machine learning. Several such protocols are treated, and either bias or overfitting is detected.

Keywords: opportunistic networks, routing, machine learning.

Матеріал надійшов 17.04.2012