МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»

ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ

XIII ВСЕУКРАЇНСЬКА НАУКОВО-ПРАКТИЧНА КОНФЕРЕНЦІЯ СТУДЕНТІВ, АСПІРАНТІВ ТА МОЛОДИХ ВЧЕНИХ

ТЕОРЕТИЧНІ І ПРИКЛАДНІ ПРОБЛЕМИ ФІЗИКИ, МАТЕМАТИКИ ТА ІНФОРМАТИКИ

(21 – 23 травня 2015 р., м. Київ, Україна)

Том II

ФІЗТЕХ

МАТЕРІАЛИ КОНФЕРЕНЦІЇ

КИЇВ 2015
ВПІ ВПК «ПОЛІТЕХНІКА»

Матеріали XIII Всеукраїнської науково-практичної конференції студентів, аспірантів та молодих вчених **«Теоретичні і прикладні проблеми фізики, математики та інформатики»** (21 – 23 травня 2015 р., м. Київ, Україна)/ Уклад.: О. Д. Василенко, С. М. Пономаренко, П. О. Наказной, Д. О. Прогонов, К. І. Ільїн, П. В. Селюх, І. Ю. Мирошникова, С. В. Бех, А. В. Шевченко.

Матеріали конференції присвячені питанням перспективних фізичних технологій, перспективним напрямкам фізики енергетичних систем, теорії безпеки, криптографічного захисту інформації та криптоаналізу; захисту інформації в комп'ютерних мережах та комунікаціях, забезпеченню цілісності баз даних, захисту від витоку інформації по каналах побічного електромагнітного випромінювання, локальним мережам різної структури, технічного захисту об'єктів, а також науковим дослідженням фундаментального та прикладного характеру у сфері інформаційних наук.

Для студентів, аспірантів та фахівців, що цікавляться проблемами наукових досліджень у сфері фізичних, енергетичних та інформаційних технологій, а також захисту інформації.

Відповідальний за випуск: О. Д. Василенко
Редактор збірника: П. О. Наказной
Верстка в системі LaTeX 2_ε: С. М. Пономаренко

Офіційний сайт конференції: http://conf.pti.kpi.ua

Доповіді учасників конференції наведені в авторській редакції

©НТУУ «КПІ», 2015 р.

ОРГАНІЗАТОРИ КОНФЕРЕНЦІЇ

▷ Національний технічний університет України «Київський політехнічний інститут»
▷ Фізико-технічний інститут НТУУ «КПІ»

ПРОГРАМНИЙ КОМІТЕТ КОНФЕРЕНЦІЇ

Голова

Новіков О. М. – д.т.н., професор, проректор з науково-педагогічної роботи

Співголова

Мачуський Є. А. – д.т.н., професор, завідувач кафедри фізико-технічних засобів захисту інформації

Члени програмного комітету

Секція **«Актуальні питання сучасної фізики»**

Воронов С. О. – д.т.н., професор, завідувач кафедри прикладної фізики
Богорош О. Т. – д.т.н., професор кафедри прикладної фізики
Іванова В. В. – к.т.н., доцент кафедри прикладної фізики
Монастирський Г. Є. – к.ф.м.н., доцент кафедри прикладної фізики

Секція **«Фізика енергетичних систем»**

Халатов А. А. – академік НАНУ, д.т.н., професор, завідувач кафедри фізики енергетичних систем
Кобзар С. Г. – к.т.н., с.н.с., Інститут технічної теплофізики НАНУ
Доник Т. В. – к.т.н., Інститут технічної теплофізики НАНУ, асистент кафедри фізики енергетичних систем
Пономаренко С. М. – к.ф.-м.н., доцент кафедри фізики енергетичних систем

Секція **«Теоретичні та прикладні проблеми математики та інформатики»**

Смирнов С. А. – к.ф.-м.н., доцент кафедри інформаційної безпеки
Гомонай О. В. – д.ф.-м.н, професор кафедри інформаційної безпеки
Качинський А. Б. – д.т.н., професор кафедри інформаційної безпеки

Секція **«Теоретичні та прикладні проблеми криптографічного захисту інформації»**

Савчук М. М. – д.ф.-м.н., професор, завідувач кафедри математичних методів захисту інформації
Фаль О. М. – к.ф.-м.н, доцент кафедри кафедри математичних методів захисту інформації
Яковлєв С. В. – к.т.н., старший викладач кафедри кафедри математичних методів захисту інформації

Секція **«Проблеми безпеки інформації в інформаційно-комунікаційних системах»**

Архипов О. Є. – д.т.н., с.н.с., професор кафедри інформаційної безпеки
Грайворонський М. В. – к.ф.-м.н., в.о. завідувача кафедри інформаційної безпеки
Родіонов А. М. – к.т.н., доцент кафедри інформаційної безпеки

Секція **«Теорія та практика технічного захисту інформації»**

Мачуський Є. А. – д.т.н., професор кафедри фізико-технічних засобів захисту інформації
Луценко В. М. – к.т.н., доцент кафедри фізико-технічних засобів захисту інформації
Кущ С. М. – к.т.н., доцент кафедри фізико-технічних засобів захисту інформації

Вчений секретар

Наказной П. О. – старший викладач кафедри інформаційної безпеки, вчений секретар Ради ФТІ

ОРГАНІЗАЦІЙНИЙ КОМІТЕТ КОНФЕРЕНЦІЇ

Голова

Василенко О. Д. – заступник директора Фізико-технічного інституту

Заступники голови

Наказной П. О. – старший викладач кафедри інформаційної безпеки, вчений секретар Ради ФТІ
Пономаренко С. М. – к.ф.-м.н., доцент кафедри фізики енергетичних систем

Члени організаційного комітету

Бех С. В. – старший викладач кафедри прикладної фізики
Ільїн К. І. – аспірант кафедри інформаційної безпеки
Мейріс А. Ж. – аспірант кафедри фізики енергетичних систем
Мирошникова І. Ю. – аспірантка кафедри інформаційної безпеки
Прогонов Д. О. – асистент кафедри фізико-технічних засобів захисту інформації
Селюх П. В. – асистент кафедри математичних методів захисту інформації
Стремецька М. С. – студентка, голова профбюро ФТІ НТУУ «КПІ»
Шевченко А. В. – аспірант кафедри математичних методів захисту інформації

Зміст

Пленарні доповіді

1. Сучасні підходи до забезпечення кібернетичної безпеки 10
М. В. Грайворонський
2. Фундаментальні взаємодії та нова фізика 18
П. О. Наказной
3. Основные идеи рефлексивного анализа 24
С. А. Смирнов

Теоретичні та прикладні проблеми математики та інформатики

1. Використання технологій Semantic Web для управління метаданими в сховищах даних 31
Д. В. Анцибор, М. Ю. Терновой
2. Застосування методу навчання з підкріпленням для керування пулом ресурсів . 33
О. О. Будя, А. М. Родіонов
3. Модель менеджера кеш-пам'яті на основі навчання з підкріпленням 36
Д. М. Вдовичинський, А. М. Родіонов
4. Уклонение от столкновения двух точек, движущихся по кусочно-линейным траекториям 39
Е. В. Водолазский, В. И. Крюков
5. Дослідження інформаційних потоків, як динамічних мультиагентних систем . . 43
А. М. Грайворонська, Д. В. Ланде
6. Визначення асимптотичної обчислювальної складності за результатами виконання 46
А. П. Гуліта
7. Підхід до побудови зв'язних SPARQL запитів 49
А. С. Домарацький, М. Ю. Терновой
8. Модифікація методу аналізу ієрархій за допомогою інтервального оцінювання . 51
А. С. Дроботун, С. А. Смирнов
9. Вирішення проблеми повноти в композиційному програмуванні 55
Т. Л. Захарченко
10. Універсальна квантова схема для двокубітного перетворення 59
Л. Ю. Карасьова
11. Тестування неперервної моделі рефлексії другого рангу біполярного вибору . . 61
К. С. Коваленко, С. А. Смірнов
12. Нахождение оптимального решения проблемы в работе Пенсионного фонда . . 63
А. В. Красицкая, Т. В. Хомяк

13. Дослідження поведінки підсистеми вибору провайдера хмарного сервісу за умов динамічного середовища . . . 66
І. Ю. Мирошникова
14. Формалізація та моделювання багатошагових ігор трьох осіб . . . 68
К. Д. Мішина, С. А. Смирнов
15. Евклідова відстань як інтегральний показник могутності держави . . . 70
Д. Р. Молоченко
16. Опис алгоритму для побудови наївного байєсовського класифікатора . . . 74
Р. М. Онищук, І. М. Терещенко
17. Дослідження алгоритмів інформаційного пошуку у пірингових мережах . . . 76
І. С. Пазинич, Д. В. Ланде
18. Багатомодові розв'язки квазилінійної модифікації гіперболічного рівняння Бюргерса . . . 80
А. С. Попов, О. С. Макаренко
19. Розподіл ресурсів між пулами та розподіл часу між завданнями всередині пулу із застосуванням машинного навчання з підкріпленням . . . 83
К. В. Рева, А. М. Родіонов
20. Моделювання процесу призначення на посаду та її аналіз у будівельній галузі 86
І. О. Суіма, Т. В. Хом'як
21. Комбіновані модифікатори нечітких об'єктів та класів . . . 89
Д. О. Терлецький
22. Калькулятор абітурієнта . . . 93
К. С. Хабарлак
23. Разработка модели прогноза посетителей сайта на основе Яндекс-Метрики . . 96
Т. В. Хомяк, А. М. Ус
24. Веб-ресурси підтримки комунікаційного середовища функціонування кафедри ВНЗ . . . 98
В. А. Шуть
25. Застосування методу навчання з підкріпленням для керування вивільненням оперативної пам'яті . . . 102
Є. М. Юдін
26. Побудова розподіленого алгоритму захоплення робіт . . . 106
І. І. Ярошевська

Теоретичні та прикладні проблеми криптографічного захисту інформації

1. Ефективна методика отримання булевих функцій для криптографічних застосувань 110
Ю. М. Виноградов, І. М. Скабенок
2. Способи зменшення криптографічної стійкості шифру шляхом заміни його окремих компонент під видом збільшення стійкості . . . 112
О. Ю. Галій
3. Спосіб прискореного синтезу незворотних булевих перетворень для схем ідентифікації абонентів на основі концепції «нульових знань» . . . 114
М. М. Дігтяр
4. Обчислення верхніх меж диференціальних імовірностей для деяких класів блочних шифрів . . . 117
Є. В. Доброногов

УДК 004.94

ДОСЛІДЖЕННЯ ІНФОРМАЦІЙНИХ ПОТОКІВ, ЯК ДИНАМІЧНИХ МУЛЬТИАГЕНТНИХ СИСТЕМ

А. М. Грайворонська[1, a], Д. В. Ланде[2]

[1] *Національний технічний університет України «Київський політехнічний інститут»*
[2] *Інститут проблем реєстрації інформації НАН України*

Анотація

В сучасному інформаційному просторі публікується безліч інформаційних документів, і інтерес представляють закономірності формування та розвитку сукупності документів, що відповідають певній тематиці. В статті запропонована модель тематичного інформаційного потоку на основі агентів, які відповідають опублікованим документам. В результаті моделювання отримані часові ряди, які відображають зміни кількості агентів в системі і дослідженні їх властивості.

Ключові слова: інформаційний потій, мультиагентна модель, часовий ряд

Вступ

Сучасний інформацій простір є динамічною системою, яка складається з інформаційних елементів (документів) пов'язаних між собою за тематикою [1]. Багато інформаційних ресурсів, які публікують інформацію он-лайн, мають потокову структуру [2]. Прикладами таких ресурсів є соціальні мережі, стрічки новин, блоги. Публікацію наукових статей також можна розглядати як інформаційний потік.

Основним об'єктом моделювання є тематичний переріз інформаційного потоку – тобто послідовність документів, які відповідають певній тематиці. Такому тематичному інформаційному потоку ставлять у відповідність часовий ряд, який складається з кількості повідомлень або документів опублікованих за певний проміжок часу. Аналіз часових рядів є фундаментальною задачею математичної статистики. В сучасних дослідженнях також використовують методи дисперсійного, фрактального та вейвлет-аналізу. Інтерес представляють шаблони поведінки, виявлення аномалій та трендів, також актуальною є задача прогнозування [3].

На динаміку реального інформаційного потоку впливає безліч факторів. На невеликих часових проміжках зміну кількості документів за тематикою можна описувати простими лінійною або експоненціальною моделями. Проте більш актуальним є дослідження динаміки інформаційного потоку від появи нової тематики до його "старіння". Прикладом моделі зміни актуальності тематики є модель дифузії інформації, яка заснована на концепції клітинних автоматів [1]. Ще один сучасний підхід, який можна застосовувати для досліджень інформаційного простору є мультиагентні моделі, які будуть розглянуті далі в статті.

[a] nastya_graiv@ukr.net

1. Мультиагентні моделі

В мультиагентних моделях агент – це абстрактна сутність, для якої характерна автономна поведінка на основі правил. Принциповою особливістю моделей на основі агентів є децентралізованість. В таких моделях визначаються правила поведінки окремих агентів, їх локальна взаємодія і взаємодія із середовищем. Такий підхід є зручним в імітаційних моделях, коли глобальна динаміка системи дуже складна або недостатньо вивчена для безпосереднього моделювання.

В роботі [4] описані такі особливості мультиагентних моделей:

- Різнорідність (heterogeneity). Індивідуальність агентів може відрізнятися певним чином. Наприклад правилами прийняття рішень, місцем положення, пам'яттю, накопиченими ресурсами, генетично і.т.п.
- Автономність (autonomy). В мультиагентних моделях немає централізованого контролю індивідуальної поведінки.
- Визначений простір (explicit space). Еволюція моделі відбувається у певному просторі, де визначене поняття локального околу.
- Локальна взаємодія (local interactions). За звичай взаємодія відбувається між сусідніми (в сенсі простору) агентами.
- Обмежена раціональність (bounded rationality). Агенти використовують лише локальну інформацію і діють на основі простих правил поведінки.
- Нерівноважна динаміка (non-equilibrium dynamics).

2. Модель інформаційного потоку на основі агентів

В роботі представлена мультиагентна модель, у якій агенти відповідають інформаційним повідом-

лення (документам) з певною тематикою. В початковий момент часу ($t = 0$) в системі існує лише один агент. Далі в процесі еволюції системи агенти можуть народжуватися і помирати, що відповідає появі повідомлення в мережі та його видаленню або втраті актуальності. Події, які відбуваються з агентом протягом його існування, відповідають основним операціям із записами в соціальних мережах: like, dislike, repost. Далі розглянемо докладно правила, за якими відбувається еволюція системи.

Будемо розглядати модель з дискретним часом. Нові агенти можуть з'являтися в системі двома способами. По-перше в кожен момент часу існує ймовірність p_s самонародження агента. Це відповідає появі повідомлення із зовнішнього світу, наприклад публікації людиною нової статті. У випадку самонародження агент може дати посилання на існуючого агента. Другий спосіб появи агента – це копіювання існуючого, тобто операція repost.

Кожен агент народжується із початковою енергією E_0. З кожним часовим відліком енергія агента зменшується на 1. Якщо значення енергії стало меньше за 0, то агент помирає. Події, які відбуваються з агентом, змінюють значення енергії за наступними правилами:

- якщо поставили like, то енергія збільшується на 1;
- якщо поставили dislike, то енергія зменшується на 1;
- якщо зробили repost, то енергія збільшується на 2;
- якщо інший агент зробив посилання на даного агента, то енергія збільшується на 1.

Ймовірність того, що відбудеться певна подія, залежить від енергії агента. Нехай в n-ий момент часу енергія рівна E_n, тоді:

$$\mathbf{P}\,\{\text{при } t = n \text{ поставили like}\} \equiv P^n_{like} = P_{like_0}\varphi(E_n),$$

$$\mathbf{P}\,\{\text{при } t = n - \text{dislike}\} \equiv P^n_{dislike} = P_{dislike_0}\varphi(E_n),$$

$$\mathbf{P}\,\{\text{при } t = n \text{ зробили repost}\} \equiv P^n_{repost} = P_{repost_0}\varphi(E_n),$$

де $\varphi(E) = \begin{cases} \frac{E}{2(2E_0 - E)}, & E \in \left[0;\ \frac{4E_0}{3}\right]; \\ 1, & E > \frac{4E_0}{3} \end{cases}$

Величини P_{like_0}, $P_{dislike_0}$, P_{repost_0} є параметрами моделі. Функція φ – монотонна зростаюча, тому ймовірність будь-якої події є більшою для агента з більшим значенням енергії.

3. Результати моделювання

Для візуалізації роботи моделі створюється спеціальний граф. Рис. 2 є ілюстрацією основних позначень.

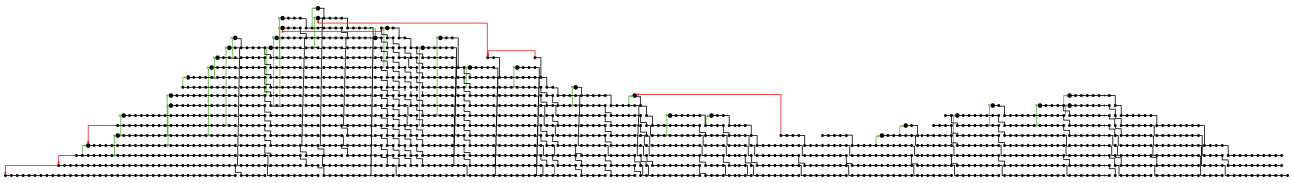

Рис. 1. Візуалізація евлюції системи агентів у вигляді графу

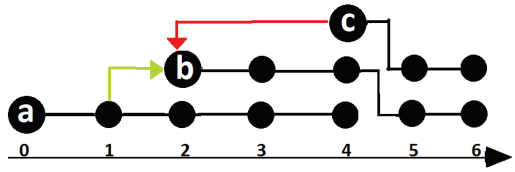

Рис. 2. Приклад візуалізації системи у вигляді графу з трьома агентами (a, b, c)

Внизу показана часова вісь, за якою видно, що в при $t = 0$ є один агент **a**. Агент **a** існує у проміжок часу від 0 до 4, що відображено його лінією життя. Також на Рис. 2 зображені агенти **b** і **c**, які народились при t рівне 2 і 4 відповідно, і померли в момент часу 6. Зеленою стрілкою показано, що в момент часу 1 зробили repost агента **a**, отже агент **b** є його копією. Агент **c** самонародився, і при цьому зробив посилання на агента **b**, що зображено червоною стрілкою.

Приклад роботи моделі з параметрами $P_{like_0} = 0.4$, $P_{dislike_0} = 0.05$, $P_{repost_0} = 0.1$ і $E_0 = 35$ зображений на Рис. 1.

У вступі було зауважено, що основним об'єктом дослідження є часовий ряд, що відповідає тематичному інформаційному потоку.

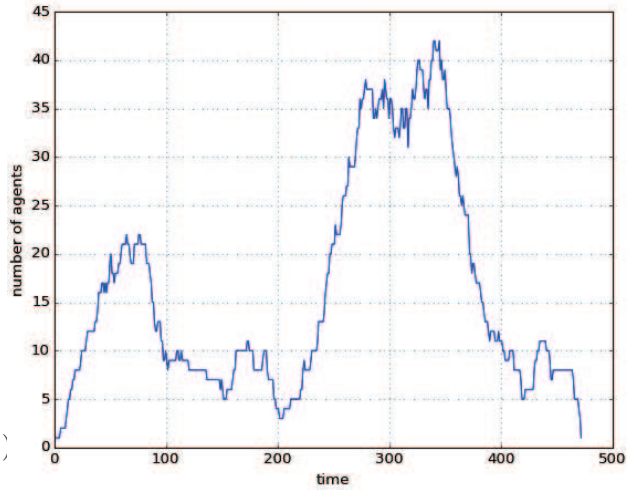

Рис. 3. Результат моделювання – часовий ряд

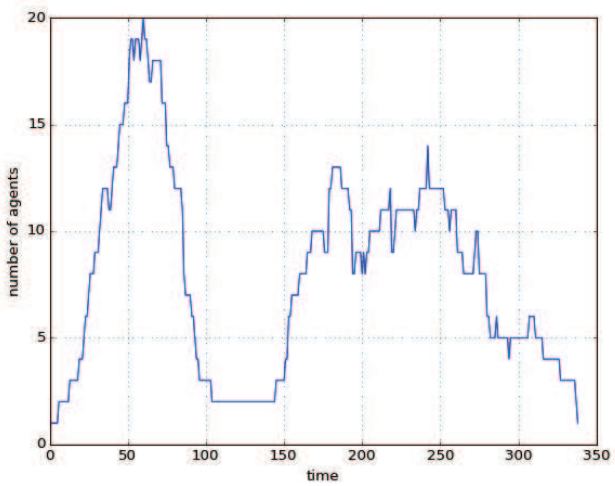

Рис. 4. Результат моделювання – графік залежності кількості агентів в системі від часу

У роботі [3] обґрунтовується, що для таких часових рядів характерні два основних локальних екстремуми, тобто два піки популярності тематики. За допомогою даною мультиагентної моделі вдається отримати часові ряди, які відповідають описаним характеристикам (Рис. 4, Рис. 3).

Реальним часовим рядам, які відповідають тематичним інформаційним потокам, притаманна хаотична поведінка. Поширеним методом аналізу самоподібності часових рядів є R/S-аналіз [1]. За допомогою цього методу обчислюється показник Херста (H), який характеризую різні типи кореляційної динаміки. Значення H<0.5 відповідає некорельованій поведінці значень ряду, а значення 0.5<H<1 визначають рівень автокореляції ряду. Показник R/S для заданого часового ряду $\{X(t),\ t = 1,\ 2, \cdots,\ N\}$ визначають через стандартне відхилення (S):

$$S = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(X(t) - < X >_N)^2},$$

де $< X >_N = \frac{1}{N}\sum_{t=1}^{N} X(t)$,

і розмах (R):

$$R = \max_{1\le t\le N} X(t,N) - \min_{1\le t\le N} X(t,N),$$

де $X(t,N) = \sum_{i=1}^{t}(X(i) - < X >_N)$.

Для багатьох часових рядів справедливе наступне співвідношення:

$$R/S = \left(\frac{N}{2}\right)^H$$

Для часових рядів, які були отримані в результаті моделювання, показник Херста має значення в межах [0.82 – 0.96]. Такі значення відбовідають високому ступеня автокореляцій в ряді.

Висновки

Для тематичних інформаційних потоків характерна складна динаміка, яка залежить від багатьох факторів. Для моделювання еволюції таких потоків застосований мультиагентний підхід. При цьому кожному агенту відповідає документ і зміна стану системи відбувається відповідно до локальних правил поведінки агентів. Розглянута модель також є ймовірнісною, тому дозволяє отримати різні типи еволюції інформаційного потоку.

В результаті моделювання отримані часові ряди, які відповідають кількості агентів в різні моменти часу. В багатьох випадках для отриманих рядів характерна наявність двох чітко виділених локальних максимумів. Також виявлений високий ступінь автокорельованість ряду. Вказані особливості також характерні для багатьох реальних інформаційних потоків.

Перелік використаних джерел

1. Горбулін В.П., Додонов О.Г., Ланде Д.В. Інформаційні операції та безпека суспільства: загрози, протидія, моделювання: монографія. — К.: Інтертехнологія, 2009. — 164 с.
2. Kleinberg J. Temporal dynamics of on-line information streams // Data Stream Management: Processing High-Speed Data Streams. — Springer, 2006. — 18 p.
3. Додонов О.Г., Ланде Д.В. Мультиагентная модель поведения тематических информационных потоков — Материалы 6-ой Всероссийской мультиконференции — 2013. — Т. 4, № 6. — с. 102–107.
4. J. M. Epstein Remarks on the foundations of agent-based generative social science — Handbook on Computational Economics — North Holland Press, 2005. — V. II.

УДК 004.942

ДОСЛІДЖЕННЯ АЛГОРИТМІВ ІНФОРМАЦІЙНОГО ПОШУКУ У ПІРИНГОВИХ МЕРЕЖАХ

І. С. Пазинич[1], Д. В. Ланде[2]

[1]*Національний технічний університет України «Київський політехнічний інститут»*
[2]*Інститут проблем реєстрації інформації НАН України*

Анотація

В цій статті розглядаються пірингові мережі як елемент глобальної мережі Інтернет та алгоритми пошуку інформації в них (BFS, RBFS, ISM та ін.). Також запропонований новий алгоритм інформаційного пошуку, приводяться графічні ілюстрації його використання на мережі, змодельованій за алгоритмом Барабаші-Альберта. Наведено результати аналізу ефективності роботи цього алгоритму.

Ключові слова: пірингові мережі, алгоритми інформаційного пошуку, BFS, RBFS, RWA, ISM, >RES

Вступ

Для початку наведемо базове визначення пірингових мереж. Пірінгові мережі (від англ. Peer-to-peer, P2P - один на один) – комп'ютерні мережі, засновані на рівноправності учасників. У таких мережах відсутні виділені сервери, а кожен вузол (peer) є як клієнтом, так і сервером. На практиці пирингові мережі складаються з вузлів, кожен з яких взаємодіє лише з деяким підмножиною інших вузлів [1]. Саме такі мережі й будуть об'єктом даного дослідження, а точніше алгоритми пошуку інформації в них. З визначення легко бачити, що основною характеристикою таких мереж є децентралізація, тобто тут практично відсутні будь-які натяки на центральний сервер, всі вузли мережі зв'язуються одне з одним напряму. Звісно, в реальному житті все виходить не зовсім так. Частіше все ж таки можна бачити пірингові мережі, розбиті на певні групи з вузлами, що є "центроїдами" даної групи. Такі "центроїди" пов'язані між собою, а усі інші вузли зв'язуються з цими "центроїдами". Відмінність у тому, що ці зв'язки не є жорсткими, користувачу ніхто не забороняє на пряму зв'язатись з будь-яким іншим вузлом, якщо в нього з'явиться така необхідність. Просто природньо користувачі групуються одне з одним за спільними інтересами, тематикою інформації, що їм необхідна й таке інше. Крім того досить часто пірингові мережі доповнюються виділеними серверами, несучими організаційні функції, наприклад авторизацію. Ця децентралізація є й основною перевагою пірингових мереж над традиційними й основним фактором, що створює складності при пошуку певної інформації. В нас немає центрального сервера, до якого можна звернутися з запитом де можна знайти те що нам потрібно. Потрібно використовувати більш гнучкі, адаптивні, можливо, інтелектуальні алгоритмі пошуку. На даний момент вже відомо й активно використовуються близько п'яти алгоритмів пошуку інформації у пірингових мережах. Їх огляд буде наведений далі.

1. Актуальність теми

Звертаючись до питання актуальності даної теми, слід зазначити, що близько 70% усього трафіку мережі Інтернет припадають саме на пірінгові мережі (рис. 1). Область їх застосування насправді досить

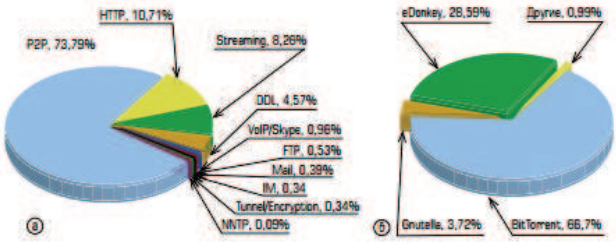

Рис. 1. Розподіл трафіку Інтернет за протоколами (а) та між P2P-мережами (б) (дані по Німеччині, 2007 р.) [1]

широка:

- Звичайно, в першу чергу на думку спадає те, задля чого такі мережі й були створені: обмін файлами. P2P виступають альтернативою FTP-архівів, які втрачають перспективу, зважаючи на значні інформаційні перевантаження.
- Розподілені обчислення. Сьогодні існує велика кількість проектів з різних областей науки, що пропонують користувачам на добровільній основі надавати ресурси своїх персональних комп'ютерів науковим організаціям для проведення розподілених обчислень певних природних явищ, обробки "Великих даних" (Big Data), що генеруються сьогодні з неймовірною швидкістю і т.д. Іноді, хоч і дуже рідко, за це користувачам навіть виплачуються певні дивіденди.
- Обмін повідомленнями. Наприклад, ICQ та Skype – це P2P-проекти.

- Групова робота. Зараз можна знайти багато різних сервісів, що надають функціонал по сумісній роботі над одним проектом цілим командам працівників. Один з найбільш відомих прикладів – програмне забезпечення TeamViewer.

Й це далеко не всі приклади використання методів P2P-мереж. Отже, розробка нових алгоритмів, більш ефективних з точки зору витрат часу та ресурсів комп'ютерів, дійсно є надзвичайно актуальним напрямком комп'ютерних наук та прикладної математики.

2. Огляд існуючих алгоритмів

З існуючих алгоритмів інформаційного пошуку слід виокремити такі:

- Метод широкого первинного пошуку (Breadth First Search, BFS) широко використовується в реальних файлообмінних P2P-мережах, таких як, наприклад, Gnutella. Основна його проблема полягає у надлишковому навантаженні на вузли мережі. Запит розсилається всім сусідам вузла, вони в свою чергу роблять те саме і т.д. поки необхідна інформація не буде знайдена. Базовий метод вирішення – використання параметру TTL (time-to-live), що визначає як далеко може зайти певний запит від місця генерації.
- Метод випадкового широкого первинного пошуку (Random Breadth First Search, RBFS) був запропонований як поліпшення "наївного" підходу BFS. Тут запит розсилається не всім сусідам а лише певній їх частині. Якій саме – це параметр RBFS. Отже метод імовірнісний.
- Ключова ідея методу випадкових блукань (Random Walkers algorithm, RWA) полягає в тому, що кожен вузол випадковим чином пересилає запит одному зі своїх сусідніх вузлів. Щоб скоротити час, необхідний для отримання результатів розсилається k запитів замість одного. Цей алгоритм нагадує метод RBFS, але в RBFS передбачається експоненціальне збільшення кількості запитів, а в методі випадкових блукань - лінійне. Обидва методи – і RBFS, і RWA – не використовують ніяких явних правил, щоб адресувати пошуковий запит до найбільш релевантному змістом
- Інтелектуальний пошуковий механізм (Intelligent Search Mechanism, ISM) складається з двох компонентів:
  1) Профайлу, який вузол q будує для кожного з сусідніх вузлів. Профайл містить останні відповіді кожного вузла.
  2) Механізму ранжирування профайлів вузлів (рангу релевантності). Ранг релевантності використовується, щоб вибрати сусідів, які даватимуть найбільш релевантні документи на запит.

  Механізм профайлів служить для того, щоб зберігати останні запити, а також кількісні характеристики результатів пошуку. Метод ISM ефективно працює в мережах, де вузли містять деякі спеціалізовані відомості.
- У методі «більшості результатів по минулій евристиці» (>RES) кожен вузол пересилає запит підмножині своїх вузлів, сформованій на підставі деякої узагальненої статистики. Запит в методі >RES є задовільним, якщо видається Z або більше результатів (Z – деяка константа). У методі >RES вузол q пересилає запити до k вузлів, що видав найбільші результати для останніх m запитів. В експериментах k змінювалося від 1 до 10, і таким шляхом метод >RES варіювався від BFS до підходу глибинного первинного пошуку (Depth-first-search). Метод> RES подібний до методу ISM, який розглядався раніше, але використовує більш просту інформацію про вузли.

У підсумку можна сказати, що на даний момент існують різноманітні алгоритми інформаційного пошуку, кожен з яких має свої переваги й недоліки, причому переваги в основному досягаються шляхом введення деякої допоміжної інформації, що зберігається на вузлах.

3. 11-кроковий алгоритм

Для початку введемо деякі базові визначення. Для моделювання нашої мережі використовується дещо адаптована векторно-просторова модель пошуку [2]. Кожен вузол мережі володіє певною кількістю документів, кожен документ в свою чергу описується вектором, в якому кожному терму, що присутні в документі, ставиться у відповідність одиниця у векторі, а тим, що не використовуються – нуль. Кількість термів (довжина вектора) n задана. Опис запиту, що відповідає необхідній користувачеві тематиці, також являє собою n-мірний вектор. Таким чином, для усіх вузлів мережі можна розрахувати дві характеристики: багатство та релевантність (близькість). Перша з них відповідає кількості термів, що ними володіє вузол у своїх документах. Звертаючись до суто математичного змісту – це сумарна по всім документам кількість одиниць на різних місцях. Цей показник може приймати значення від 0 (жодного терму, звичайно, на практиці таке неможливо) до n (вузол-володар усіх існуючих термів, також малоймовірна ситуація). Друга характеристика показує, наскільки документ за своїм змістом (термами що в ньому є) відповідає запиту користувача. Для оцінки близькості запиту і документа використовується операція логічного "Г". Тобто, чим більша кількість одиниці на однакових місцях у векторі запиту й у векторі документа, тим більша ймовірність, що цей документ слід повернути користувачу у відповідь на його запит. Суть нового пропонованого у цій статті алгоритму пошуку інформації у пірингових мережах полягає у плавному переході від прямування до абсолютного багатства до прямування до абсолютної релевантності. Приведемо пояснення на прикладі:

1) На першому кроці алгоритму серед усіх сусідів стартового вузла перехід відбувається до найба-

гатшого з них. Тобто, крок робиться абсолютно незалежно від змісту запиту.

2) На наступному кроці показник багатства сусідів поточного вузла множиться на 0,9, а релевантність на 0,1. Перехід відбувається до вузла з найбільшою сумою цих двох результатів множення.
3) На третьому кроці багатство сусідів множиться вже на 0,8 а релевантність на 0,2. Перехід відбувається аналогічним чином.
4) ...
5) На останньому, одинадцятому кроці, перехід відбувається до вузла, що є найбільш релевантним до запиту користувача серед усіх варіантів. Тобто, фактично багатство множиться на 0, а релевантність на 1.

Наведений алгоритм вигідно відрізняється від таких алгоритмів як ISM та >RES тим, що він не потребує зберігання відносно великих об'ємів інформації на вузлах для його роботи, але при цьому містить у собі залежність від змісту запиту та характеристик вузла, а не просто робить кроки, поки не знайде найліпшу відповідь як BFS, RBFS та RWA. З іншого боку, цей метод зовсім не гарантує 100% знаходження найкращого результату серед усіх можливих варіантів. Більш того, якщо не реалізовувати ніякого механізму запам'ятовування найкращого поточного результату в кінці можна отримати гірший результат ніж ті, що вже траплялися. Отримати вузол з документами з меншою релевантністю порівняно з вже переглянутими документами. Зрозуміло, що в залежності від масштабу мережі алгоритм також можна масштабувати шляхом збільшення кількості кроків на порядок: 101 крок, 1001 крок і т.д. Для тестування алгоритму використовувалася відносно невелика мережа на 50 вузлів, відповідно кількість кроків була обрана базова - одинадцять. Генерування мережі виконувалося за алгоритмом Барабаші-Альберта. Приведемо приклади результатів роботи алгоритму (рис. 2 - 6). На наведених малюнках червоним ко-

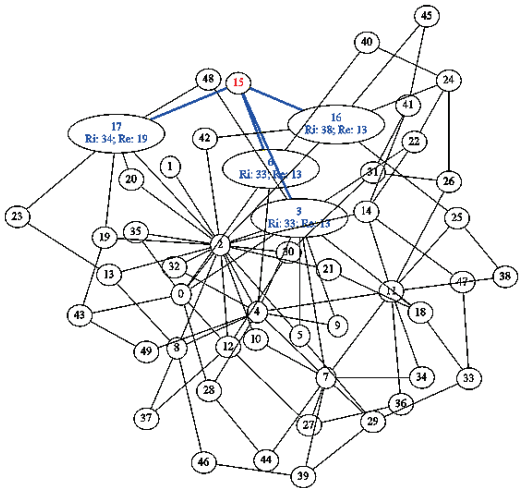

Рис. 2. Початок роботи алгоритму

льором виділяються номери поточних вузлів, синім – ребра до усіх сусідів та самі сусіди. Також в середині

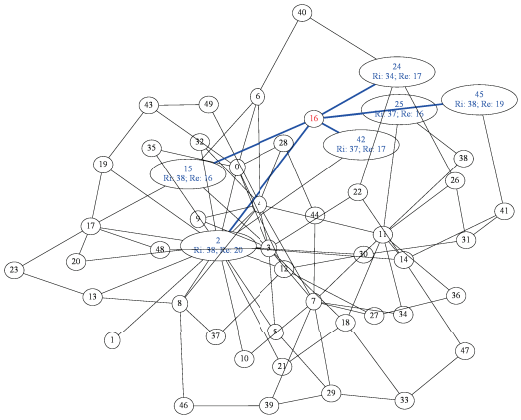

Рис. 3. Після виконання першого кроку

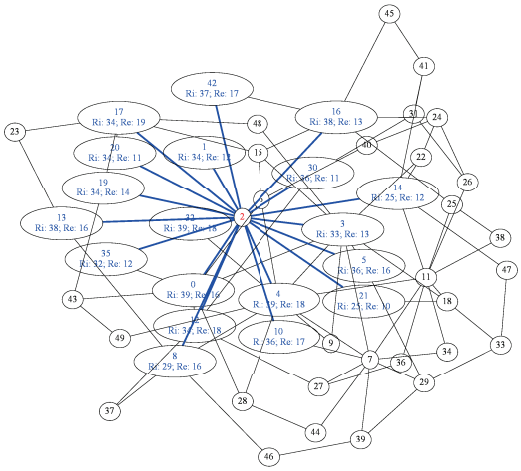

Рис. 4. Після виконання другого кроку

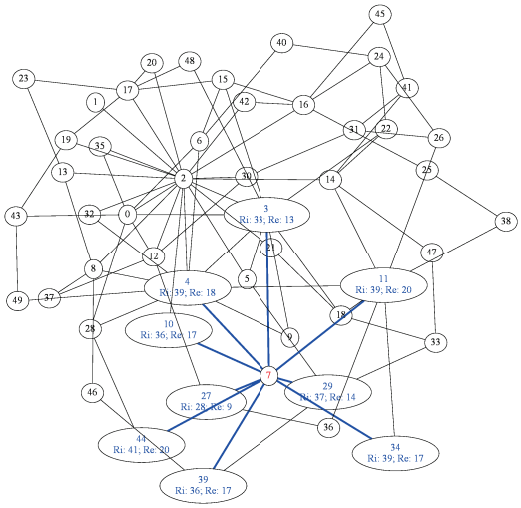

Рис. 5. Передостанній крок

вузла-сусіда вказуються його показники багатства (Ri від Richness) та релевантності (Re від Relevance). Легко бачити, що алгоритм реалізований коректно й надає досить релевантні результати, хоч й не найкращі. Загалом оцінка ефективності роботи алгоритму за релевантністю результату показала, що практично завжди результат виходить або найкращім (найбільша релевантність серед можливих) або другим за релевантністю. Середнє відхилення від найкращого результату – 6%. Однак слід зазначити одну осо-

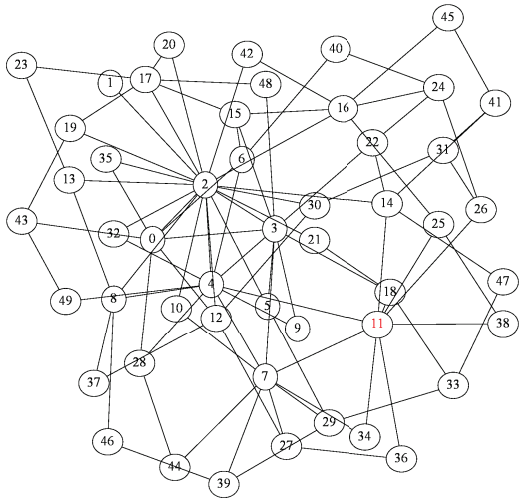

Рис. 6. Кінцевий результат

бливість роботи алгоритму, так зване “потрапляння в richclub”. На наведених скріншотах це не досить чітко видно (для цього потрібно проілюструвати всі 11 кроків, а ми обмежені розмірами статті), але практично завжди під час роботи алгоритму є період циклічного повторення кількох вузлів. Ці вузли характеризуються високими показниками порівняно з їх сусідами й при черговому виборі наступного вузла вибирається просто інший член “richclub”. В даному конкретному випадку “richclub” складали вузли 7 та 44. Слід зазначити, що явище “richclub” цілком відповідає реальному світу. У піринговій мережі природньо з’являються певні групи користувачів з найбільшою кількістю найкращих за певними критеріями даних.

Висновки

У межах даної статі було розглянуто поняття пірингових мереж та їх значної ролі у сучасному Інтернет. Також були наведені найбільш відомі та розповсюджені на даний момент алгоритми пошуку інформації в цих мережах. Основною суттю цього матеріалу є новий запропонований алгоритм пошуку інформації у відповідь на запит користувача, що враховує зміст запиту та певні характеристики вузлів мережі. Звісно, аналіз характеристик не такий суттєвий, як у алгоритмі ISM, наприклад, але він і не вимагає зберігання значної кількості додаткової інформації на вузлі (профайли у ISM). За наведеними ілюстраціями та деякими показниками ефективності можна побачити, що даний алгоритм в цілому є досить цікавим й, можливо з подальшими доробками, може бути досить актуальним й успішно використовуватись на реальних пірингових мережах.

Перелік використаних джерел

1. Ландэ Д. В. P2P - по секрету всему свету... О пиринговых сетях — 2008. — 104 с.
2. Интернетика. Навигация в сложных сетях. Модели и алгоритмы / Д. В. Ландэ, А. А. Снарский, И. В. Безсуднов— М. : Книжный дом «ЛИБРОКОМ», 2009. — 264 с.