УДК 004.72

А. В. САГУН

Черкаський державний технологічний університет

ОПТИМІЗАЦІЯ МОДЕЛІ РОЗПОДІЛУ МЕРЕЖНОГО ТРАФІКУ В ІНФОРМАЦІЙНІЙ СИСТЕМІ З ВРАХУВАННЯМ ПОЛІТИКИ ДОСТУПУ

Запропоновано оптимізацію моделі розподілу мережного трафіку для інформаційної системи, яка має критичні обмеження щодо його проходження. Запропонований метод спирається на застосування теореми Байєса та оцінки класифікатора в QoS моделі. Розглянуто та запропоновано рішення задачі невідомих декад, здійснено постановку обчислювального експерименту, викладено принципи формування ймовірнісного простору та класифікатору. В результаті аналізу стеку мережного протоколу запропоновано функціональну схему реалізації розподілу та фільтрації мережного трафіку та запропоновано прикладне застосування QoS–фільтрації для роутера Cisco 3550–3560.

Ключьові слова: *теорема Байєса, наївний класифікатор, фільтрація трафіку, політики доступу, ймовірності, якість обслуговування.*

Вступ

В сучасних умовах стрімкого збільшення кількості користувачів інформаційного трафіку в інформаційно-комунікаційних системах і мережах власник такої мережі не завжди може задовольнити потреби в швидкості доступу та якості каналу для всіх споживачів. Особливо актуальна така проблема є в сільській місцевості через нерозвиненість мережних комунікацій. Одним зі шляхів вирішення цієї проблеми є вдосконалення моделі фільтрації QoS-трафіку на базі Байєсівської моделі.

Прикладним застосування ймовірнісних моделей займаються В. А. Попов, Н. П. Кондратенко, Н. Н. Гора. Використання математичного апарату Байєса для моніторингу навантаження локальної обчислювальної мережі досліджували в своїх роботах А. В. Скатов, Д. Ю. Воронін, Д. Н. Данільчук. Методики оцінки оптимальності розподілу ресурсів інформаційно–комунікаційних систем та мереж на основі критерію якості обслуговування розглянуто в роботі [1]. Проблему ідентифікації узагальнених параметрів математичної моделі комп'ютерної мережі в задачі забезпечення QoS сформульовано в [2]. З огляду на недоліки існуючих в ряді операційних систем технологій підтримки інтелектуального розподілу мережного трафіку, є необхідність створити оптимізовану систему, що виконує таку задачу, максимально враховуючи потреби користувача.

Постановка задачі

Суб'єкти всієї розподіленої автоматизованої системи розподілу мережного трафіку: насосна станція; вузол IP–телефонії, інформаційна мережа доступу до локальних ресурсів сільради. Оптимізацію інформаційного трафіку необхідно провести з врахуванням правил політики доступу, використовуючи простий та прозорий метод фільтрації та розподілу на базі байєсівського математичного апарату та заданої угоди про трафік (QoS).

Метою роботи є модифікація методу фільтрації пакетів з врахування шаблонів правил політики доступу. Класифікація трафіку являє собою окрему задачу.

Виклад основного матеріалу

Для більшості випадків якість зв'язку визначається чотирма параметрами: пропускна смуга, затримка при передачі пакета, коливання затримки при передачі пакетів, втрата пакетів.

В умовах сільської ради, з обмеженим зовнішнім трафіком, коли передача даних зіштовхується із проблемою «вузького місця», для прийому й відправлення пакетів на роутерах зазвичай використається метод FIFO: перший прийшов — перший пішов (First In — First Out). Але наявність великої кількості сервісів буде створювати інтенсивний трафік, що стає причиною заторів, які вирішуються вкрай простим чином: всі пакети що не ввійшли до черги FIFO ігноруються роутером, і відповідно губляться безповоротно. Така ситуація в даному випадку є неприпустимою, адже є задачі, обслуговування яких повинно бути пріоритетним.

Алгоритм використання теореми Байеса для забезпечення необхідної якості обслуговування (QoS) в такому випадку можна сформулювати так:

1. Перше застосування – обчислення ймовірності того, що декада адреси небажана, знаючи, що дана декада з'являється в цьому пакеті.

2. Друге – для обчислення ймовірності того, що пакет небажаний, враховуються всі його декади (або відповідні їх підмножини).

3. Інколи застосовується в третій раз, коли зустрічаються повідомлення з рідкісними декадами (обчислення ймовірності того, що пакет, який містить дану декаду, є небажаним).

Припустимо, що шуканий пакет містить в заголовку декаду 213. Більшість фільтрів, які отримують такі декади, наприклад, знають, що вони розташовані в сегменті 213.xxx.xxx.xxx, і, відповідно, цей пакет, швидше за все, буде небажаним. Програма виявляє його, однак, не «знає», про той факт, що він є небажаним. Все, що вона може зробити – обчислити ймовірності.

Вираз, що буде використаний в програмному забезпеченні, отриманий з теореми Байєса, і формули повної ймовірності [3, 4]:

$$\Pr(S|W) = \frac{\Pr(S|W) * \Pr(S)}{\Pr(W)} = \\ = \frac{\Pr(W|S) * \Pr(S)}{\Pr(W|S) * \Pr(S) * \Pr(W|H) * \Pr(H)}, \quad (1)$$

де $\Pr(S|W)$ — умовна ймовірність того, що пакет підлягає фільтрації, за умови, що декада адреси з зовнішнього діапазону в ньому знаходиться;

$\Pr(S)$ — повна ймовірність того, що довільний пакет підлягає фільтрації;

$\Pr(W|S)$ — умовна ймовірність того, що небажана декада з'явиться в повідомленні, якщо вона підлягає фільтрації;

$\Pr(H)$ — повна ймовірність того, що довільна декада є небажаною;

$\Pr(W|H)$ — умовна ймовірність того, що декада ззовні з'явиться в пакеті, якщо він відносяться до категорії небажаних.

Оцінка ймовірностей $\Pr(S)$ и $\Pr(S|W)$ відбувається на навчальній виборці. Ймовірність появи окремого класу можна оцінити виразом:

$$P(c) = D_c D, \quad (2)$$

де D_c – кількість пакетів, що належать до класу C, D – загальна кількість пакетів у вибірці для навчання.

Оцінку ймовірності появи пакета в класі можна здійснювати декількома шляхами. Наприклад, за методом multinomial bayes model [5]:

$$P(W_i|C) = \frac{W_{ic}}{\sum_{i' \in V} W_{i'c}}, \quad (3)$$

де W_{ic} — кількість появи i-го пакету в пакету класу C;

V — список всіх унікальних пакетів.

Отже, чисельник описує частоту появи декади в пакетах класу (повтори рахуються), а знаменник – сумарна кількість декад у всіх пакетах даного класу.

Проблема невідомих декад

Проблема невідомих декад відома в формулюванні байєсівського класифікатора [3, 7].

Теоретично розв'язати цю проблему можна за рахунок обробки на процесі навчання великої кількості пакетів. Для випадку протоколу IPv4, при можливих значеннях кожної декади від 0 до 255 кількість пакетів, необхідних для повного навчання становитиме 4294967296, а для протоколу IPv4-281474976710656. Це дуже велике значення.

Розв'язок даної проблеми запропонований в [5]. Це так зване адитивне згладжування (згладжування Лапласа). Такий підхід зсуває оцінку ймовірностей в бік менш ймовірних розв'язків. Отже, декади, які були невідомі на початок навчання отримують в класифікаторі відмінну від нуля ймовірність гіпотез.

Наведемо такий приклад на практиці: нехай в результаті навчання маємо декади (табл. 1).

Таблиця 1

Частота появи декад

Декада	Частота появи
192.168.0.255	3
192.168.0.1	2
192.168.0.15	1

Припустимо, що на етапі класифікації, з'являється декада «2» в пакеті 192.168.0.2. На етапі навчання вона була відсутня. Отже оригінальна і зсувна за Лапласом оцінка ймовірностей буде мати вигляд (рис. 1).

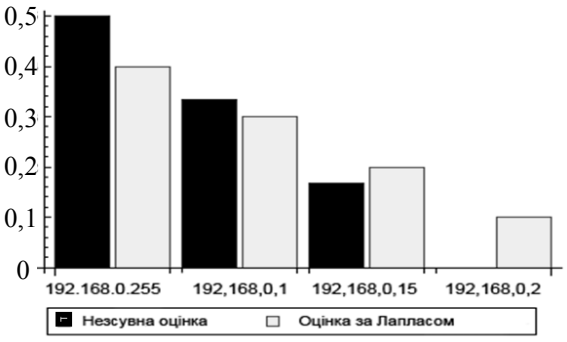

Рис. 1. Оцінка ймовірностей

З графіку на рис. 1 видно, що зсувна оцінка ніколи не буває нульовою. Це захищає від проблеми невідомих декад.

Хоча на практиці в сегменті мережі, що маршрутизується однім комутатором діапазон IP–адрес знаходиться в межах 65355, тобто кількість пакетів для повного навчання буде 65535. Обробити таку вибірку нескладно.

При відносно великій масці у фільтрі декад пакету необхідно перемножувати велику кількість мало розрядних чисел. При цьому обов'язково виникатиме проблема арифметичного переповнення знизу [5]. Для розв'язку цієї проблеми скористаємось властивостями логарифма добутку:

$$\log ab = \log a + \log b. \quad (4)$$

В зв'язку з тим, що логарифм–функція монотонна, то її застосування до обох частин виразу (4) змінить її чисельне значення, але не параметри, при яких дана функція досягає максимуму. При цьому логарифм від числа, близького до нуля, природно буде числом від'ємним, але тільки за модулем більшим ніж поточне значення [6]. Це робить логарифмічне значення ймовірностей більш зручним до аналізу. Отже, можна стверджувати, що пріоритетна маршрутизація запишеться як:

$$C_{map} = \arg\max c \in C\left[\log P(c) + \sum i\right] = \\ = \ln\log P\left(W_i \middle| C\right). \quad (5)$$

Для наведеного вище виразу основа логарифму значення не має. Тобто, можна використовувати всі види логарифмів. Підставивши всі обрані оцінки у вираз (5) отримаємо кінцеву формулу, за якою працює класифікатор:

$$C_{map} = \arg\max c \in C\left[\log D_c D + \sum i\right] = \\ = \ln\log W_{ic} + i * |V| + \sum i' \in V * W_{i'c}.$$

Для реалізації Байєсівського класифікатора необхідна навчальна вибірка. В ній слід проставити відповідності між пакетами та їх класами. Потім, необхідно зібрати статистику з вибірки, яка буде використана на етапі класифікації, тобто:

1) сумарну кількість декад в пакеті кожного класу;

2) відносні частоти декад в межах одного класу;

3) розмір словника вибірки.

Сукупність даної інформації є моделлю класифікатора. На етапі класифікації необхідно для кожного класу розрахувати значення наступного виразу і обрати клас з максимальним значенням:

$$\log D_c D + \sum i \in Q \log W_{ic} + i|V| + L_c,$$

де D_c – кількість пакетів в навчальній виборці, що належать класу с;

D – загальна кількість пакетів в навчальній виборці;

$|V|$ – кількість унікальних декад у всіх пакетах навчальної вибірки;

L_c – сумарна кількість декад в усіх пакетах класу с в навчальній виборці;

W_{ic} – кількість зустрічань і-ої декади в пакетах класу с в навчальній виборці;

Q – множина декад класифікованих пакетів, включаючи повтори.

Припустимо, що маємо 3 пакети, для яких відомі їх класи. Корисний (USF) означає, що даний пакет є корисним і відфільтровуватися не буде, а небажаний (NUSF) означає, що даний пакет повинен бути відсіяний з даного рівня і додатково прокласифікований (таким пакетам буде надаватися найнижчий 3-й рівень пріоритету). Тобто, маємо пакети:

1) [NUSF] 192.168.0.5;

2) [NUSF] 192.168.0.2;

3) [USF] 198.168.1.4.

Для формування моделі класифікатора заповнимо табл. 2.

Таблиця 2

Статистичні дані аналізу корисності пакетів

Категорія	небажаний	корисний
Частота класів	2	1
Сумарна кількість декад	6	3

Далі класифікуємо пакети (табл. 3)

Таблиця 3

Дані класифікації пакетів

Декада	небажаний	корисний
0	1	0
5	1	0
0	1	0
2	1	0
1	0	1
4	0	1

Тепер класифікуємо пакет, що складається з наступних декад. Припустимо, що потрібно класифікувати комбінацію декад 192.168.1.8. Розрахуємо значення виразу для класу NUSF:

$$\log 23 + \log 18 + 4 + \log 28 + 4 + \log 18 + 4 \approx 17,3.$$

Далі класифікуємо ту саму комбінацію декад для класу USF:

$$\log 13 + \log 28 + 2 + \log 28 + 2 + \log 18 + 2 \approx 11,26.$$

В даному випадку клас USF отримав мінімальне значення і не класифікується, як небажаний. Надалі

відфільтровані пакети потрапляють до вторинного класифікатору. Далі необхідно формування ймовірнісного простору.

В елементарному випадку оператор обирає клас, який отримав максимальну оцінку. Оцінки, які видає алгоритм, не задовольняють двом формальним властивостям, яким повинні задовольняти всі ймовірнісні оцінки, зокрема [7]: всі вони повинні знаходитися в діапазоні від 0 до 1; їх сума дорівнює 1.

Розв'язок даної задачі полягає у формуванні ймовірнісного простору з логарифмічних оцінок. Іншими словами, це означає необхідність позбавитися логарифмів та нормувати суму поодинці:

$$P(c|d) = eqc + \sum c' \in Ceqc',$$

де qc – логарифмічна оцінка алгоритму для класу c;

e – основа натурального логарифму.

Піднесення до ступеню e основи натурального логарифму, використовується для того, щоб позбутися від логарифму ($a \log ax = x$) [6].

Таким чином, якщо в розрахунках використовується не натуральний логарифм, а десятковий логарифм, необхідно використовувати не число e, а 10.

Для приклада, що наведений вище, ймовірність того, що повідомлення класу NUSF дорівнюватиме:

$$e-0,352e-0,352+e-2,5906=0,73,$$

тобто даний пакет є небажаним з ймовірністю 73 %.

При формулюванні моделі класифікатора розділимо всіх споживачів каналу трафіку на три основні групи в залежності від пріорітету таким чином:

1. Споживачі з середнім пріорітетом доступу (система VoIP–телефонії, бухгалтерія).
2. Споживачі з високим пріорітетом доступу (автоматизована система центрального водопостачання).
3. Споживачі з низьким пріорітетом доступу (система телевідеоконференцій, персональні споживачі web-ресурсів).

Таким чином, при аналізі конкретного споживача трафіку виділимо гіпотези виду θ_i для споживачів, що належать i –групі, де $i \in (1,2,3)$. Відомо, що 50 % споживачів трафіку мали середній пріорітет доступу QoS, 30 % – високий, 20 %– низький.

Використовуючи ці дані, визначимо апріорні ймовірності гіпотез:

$$\Pr(\theta_1)=0,5;\ \Pr(\theta_2)=0,3;\ \Pr(\theta_3)=0,2.$$

Будемо вважати, що ознаками підвищеного пріорітету на споживання є необхідність задовольнити потреби населення (y_1) та критерій збитків населеного пункту при невиконанні (y_2). З подальшого аналізу відомо, що ознаки необхідності задоволення потреб населення виникають у 40 % споживачів з середнім пріорітетом, 80 % – з високим, 30 % – з низьким. Звідки можна записати умовні ймовірності виду:

$$\Pr(y_1|\theta_1)=0,4;\ \Pr(y_1|\theta_2)=0,8;\ \Pr(y_1|\theta_3)=0,3.$$

Також відомо, що критеріальний рівень збитків безпеці населеного пункту при невиконанні запиту виникає у 70 % споживачів з середнім рівнем пріорітету доступу, 90 % – з високим і 0 % – з низьким. Звідки можна записати умовні ймовірності виду:

$$\Pr(y_2|\theta_1)=0,7;\ \Pr(y_2|\theta_2)=0,9;\ \Pr(y_2|\theta_3)=0.$$

Занесемо всі відомі апріорні та умовні ймовірності в табл. 4.

Таблиця 4

Апріорні та умовні ймовірності

I	1	2	3
$\Pr(\theta_i)$	0,5	0,3	0,2
$\Pr(y_1\|\theta_i)$	0,4	0,8	0,3
$\Pr(y_2\|\theta_i)$	0,7	0,9	0,0

Слід відмітити, що умовні ймовірності протилежних свідчень можна визначити з очевидної умови виду: $\Pr(y_c|\theta_i)=1-\Pr(y|\theta_i)$. В процесі збирання інформації (навчання) моделі ймовірності гіпотез будуть підвищуватися, якщо факти їх підтверджують, та зменшуватися – якщо спростовуються. Після формування гіпотез ймовірності проаналізуємо стек мережного протоколу для побудови моделі.

В полях заголовків різних мережних протоколів присутні спеціальні поля, виділені для маркування трафіку [8]. Поле Class of Service (CoS) в мережному пакеті має розмірність 3 біта. Це дозволяє розділити трафік на 8 потоків с різним маркуванням IP. В старому стандарті існувало поле ToS (8 біт), з якого, в свою чергу, виділялись 3 біта під назвою IP Precedence. Це поле копіювалось в поле CoS Ethernet заголовка. У відповідності до діючого стандарту поле ToS було перейменовано в DiffServ, і додатково відводиться 6 біт для поля Differencial Service Code Point (DSCP), в якому можна передавати параметри, необхідні для обслуговування даного типу трафіку (рис. 2). Тобто, теоретично кількість ознак для фільтра, можна довести до 64, що цілком достатньо для розв'язку поставленої вище задачі.

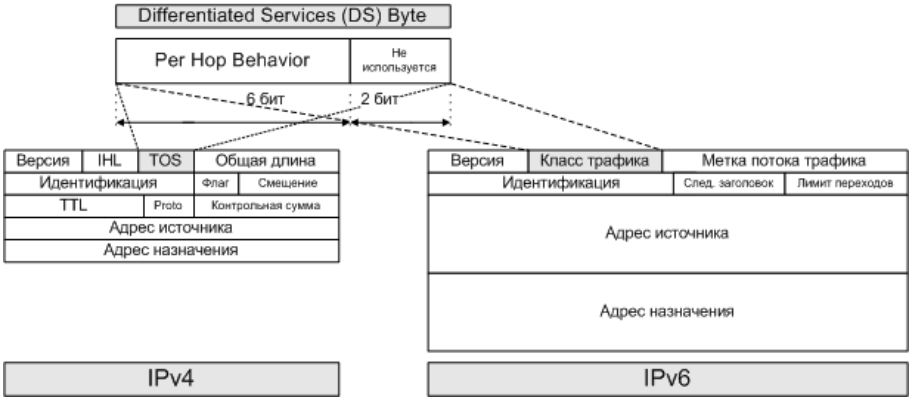

Рис. 2. Структура стеку протоколу TCP/IP

Всі дані, які необхідно піддавати QoS–фільтрації описаним вище методом, слід маркувати в пакеті близько до джерела даних. Так, наприклад, більшість IP-телефонів самостійно додають в IP-заголовок голосових пакетів поля DCSP = EF або CS5. Схожа технологія застосовується в пірінгових мережах [9].

На основі існуючої топології організації сільської ради, яка являє собою дворангову Ethernet–мережу, та сформульованого математичного апарату (4), формулювання моделі класифікатору для сільської ради (табл. 2), аналізу стека протоколу TCP/IP для QoS та його структури (см. рис. 2) можна побудувати функціональну схему модифікованого методу фільтрації QoS–трафіку (рис. 3).

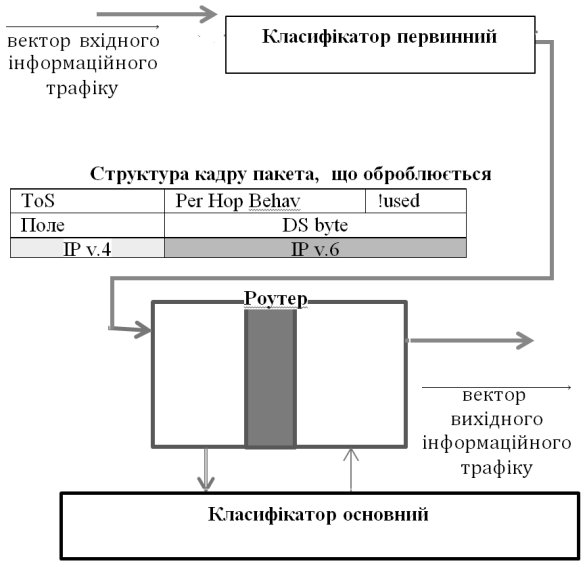

Рис. 3. Функціональна схема модифікованого методу фільтрації QoS–трафіку

Наведена функціональна схема являє собою практичну модифікацію існуючих методів QoS і ToS моделей фільтрації мережного трафіку, застосованою до задач планування та реалізації правил політики доступу в автоматизованих системах 2 і 3 класу [10].

Розроблена технологія може бути реалізована в складі Class-based shaping. CB shaping може бути налаштовано для вихідних пакетів і застосовано до фізичного інтерфейсу або підінтерфейсу.

При налагодженні її на роутері необхідно вказувати shaping rate, Bc та Bci параметри можуть не вказуватися, а параметр Tc не може бути заданий напряму. Відповідно, CB –shaping розраховує невказані значення. Такі значення обраховуються по–різному, залежно від значення shaping rate, зокрема для роутерів Cisco 3550-3560 [11].

Висновки

В ході проходження трафіку, на основі проведених досліджень вдалося запропонувати оригінальний механізм класифікації трафіку, який можна вважати варіантом зваженого справедливого обслуговування. Цей механізм вільний від такого недоліку останніх, як поділ на декілька класів для кожного з яких ведеться окрема черга пакетів та зв'язується з кожною чергою не її пріоритету, а відсотку пропускної спроможності вихідного інтерфейсу, що гарантується даному класу трафіку при перевантаженнях цього інтерфейсу.

Література

1. Стрюк, А. Ю. Методика оценки оптимальности распределения ресурсов инфокоммуникационных сетей на основе показателя воспринимаемого качества обслуживания [Текст] / А. Ю. Стрюк, И. Н. Понамарев, А. В. Соловьева // Радіоелектронні і комп'ютерні системи. – 2009. – № 6 (40). – С. 20–25.

2. Славко, О. Г. Ідентифікація узагальнених параметрів математичної моделі комп'ютерної мережі в задачах забезпечення QoS [Текст] / О. Г. Славко // Радіоелектронні і комп'ютерні сис-

теми. – 2010. – № 3 (44). – С. 68–74.

3. Теорема Байеса [Електронний ресурс]. – Режим доступу: http://ru.wikipedia.org/wiki/%D0%A2%D0%B5%D0%BE%D1%80%D0%B5%D0%BC%D0%B0 %D0%91%D0%B0%D0%B9%D0%B5%D1%81%D0%B0. – 20.01.2014.

4. Text miming explained. Практические советы по реализации систем извлечения информации [Электронный ресурс]. – Режим доступа: http://krondix.blogspot.com/2006/10/blog-post.html. – 20.01.2014.

5. Логарифм [Электронный ресурс]. – Режим доступа: http://ru.wikipedia.org/wiki/%D0%9B%D0%BE%D0%B3%D0%B0%D1%80%D0%B8%D1%84%D0%BC. – 20.01.2014.

6. Действия с логарифмами и степенями [Электронный ресурс]. – Режим доступа: http://allmath.ru/logarifm-formules.htm. - 20.01.2014.

7. Смирнова, Н. В. Байесовские сети и их приложения [Электронный ресурс] / Н. В. Смирнова. – ИПУ РАН, 2011. – Режим доступа: http://www.slideshare.net/indra-uolles/2-10311879. – 20.01.2014.

8. TCP [Электронный ресурс]. – Режим доступа: http://uk.wikipedia.org/wiki/TCP. - 21.01.2014.

9. Понятие пиринговые сети. Блог [Электронный ресурс]. – Режим доступа: http://kyzminov-trecker.admin007.net/t14-topic. - 22.01.2014

10. НД ТЗІ 2.5-005-99. Класифікація автоматизованих систем і стандартні функціональні профілі захищеності оброблюваної інформації від несанкціонованого доступу [Текст]. – Чинний з 1999-07-01. – Київ : ДСТСЗІ СБ України, 1999. – 16 с.

11. Configuring Class-Based Shaping. Cisco IOS Quality of Service Solutions Configuration Guide, Release 12.2 [Электронный ресурс]. – Режим доступа: http://www.cisco.com/c/en/us/td/docs/ios/12_2/qos/configuration/guide/fqos_c/qcfcbshp.html. – 24.01.2014.

Надійшла до редакції 24.01.2014, розглянута на редколегії 12.02.2014

Рецензент: д-р техн. наук, проф., професор кафедри інтелектуальних систем прийняття рішень С. В. Голуб, Черкаський національний університет ім. Б. Хмельницького, м. Черкаси.

ОПТИМИЗАЦИЯ МОДЕЛИ РАСПРЕДЕЛЕНИЯ СЕТЕВОГО ТРАФИКА В ИНФОРМАЦИОННОЙ СИСТЕМЕ С УЧЕТОМ ПОЛИТИКИ ДОСТУПА

А. В. Сагун

Предложена оптимизация модели разделения сетевого трафика для информационной системы с критическими ограничениями. Предложенный метод базируется на применении теоремы Байеса и оценки классификатора в QoS модели. Рассмотрено и предложено решение задачи неизвестных декад, проведена постановка вычислительного эксперимента, изложены принципы формирования вероятностного пространства и классификатора. В результате анализа стека сетевого протокола предложена функциональная схема реализации распределения и фильтрации сетевого трафика и прикладное применение QoS–фильтрации для роутера Cisco 3550–3560.

Ключові слова: теорема Байеса, наивный классификатор, фильтрация трафика, политики доступа, вероятности, качество обслуживания.

OPTIMIZATION OF DISTRIBUTION OF NET TRAFFIC IN INFORMATION SYSTEM POLITICS OF ACCESS BASED

A. V. Sagun

The optimization of model of net traffic's divide for information system where proposed. Method proposed is based on application of Bayes theorem and evolution of a classifier at QoS model. The solution of the task of unknown decades was considered and proposed. The performance of calculation experiment was carry out. Principles of forming of probability's environment and classifier were expressed. As a result of net protocol's stack analyze, functional scheme of distribution and filtration of net's traffic, applications of QoS-filtration for Cisco 3550-3560 router was proposed.

Key words: Bayes theorem, naïve classifier, traffic filtration, politic of access, probabilities, quality of service.

Сагун Андрей Викторович – канд. техн. наук, доцент кафедры информатики и информационной безопасности Черкасского государственного технологического института, Украина, e-mail: avd29@ukr.net.