МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЧЕРКАСЬКИЙ ДЕРЖАВНИЙ ТЕХНОЛОГІЧНИЙ УНІВЕРСИТЕТ

ДЄЄВ КОСТЯНТИН СЕРГІЙОВИЧ

УДК 004.77:519.2

ДОСЛІДЖЕННЯ МЕРЕЖЕВОЇ ВЗАЄМОДІЇ ЗА ДОПОМОГОЮ СИСТЕМИ ГЛИБОКОГО АНАЛІЗУ ПАКЕТІВ

05.13.05 – комп’ютерні системи та компоненти

Автореферат
дисертації на здобуття наукового ступеня
кандидата технічних наук

Черкаси — 2018

Дисертацією є рукопис.

Робота виконана в Київському національному університеті імені Тараса Шевченка Міністерства освіти і науки України.

Науковий керівник: кандидат фізико-математичних наук, доцент
Бойко Юрій Володимирович,
Київський національний університет
імені Тараса Шевченка,
начальник Інформаційно-обчислювального центру.

Офіційні опоненти: доктор технічних наук, професор
Мусієнко Максим Павлович,
Чорноморський національний університет імені Петра Могили, професор кафедри комп'ютерної інженерії;

доктор технічних наук, професор
Бараннік Володимир Вікторович,
Харківський національний університет
Повітряних Сил імені Івана Кожедуба, начальник кафедри бойового застосування та експлуатації АСУ.

Захист відбудеться « » _____ 2018 року о _____ годині на засіданні спеціалізованої вченої ради К 73.052.04 Черкаського державного технологічного університету за адресою: 18006, м. Черкаси, бул. Шевченка, 460.

З дисертацією можна ознайомитися в бібліотеці Черкаського державного технологічного університету за адресою: 18006, м. Черкаси, бул. Шевченка, 460.

Автореферат розісланий « » _____ 2018 року

Учений секретар
спеціалізованої вченої ради Е. В. Фауре

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. В останні роки популярність додатків, які працюють у мережах рівноправних вузлів, та створюваний ними трафік у мережах загального призначення зростають швидкими темпами. Такі додатки отримали назву P2P-додатків (Peer-to-Peer, P2P). Збільшення в об`ємі трафіку, використаного розподіленими програмними реалізаціями, крім доброякісних додатків, спостерігається також через розповсюдження шкідливого програмного забезпечення, яке функціонує з застосуванням схожого архітектурного підходу та являє собою формування систем бот-мереж. Класифікація такого трафіку є складною задачею, тому що традиційні методи, які в основному аналізують номери портів або властивості та значення заголовків протоколів, стають неефективними поміж додатків, що використовують випадкові порти або шифрування. Необхідність проведення точної класифікації бажана не лише з огляду на управління внутрішніми процесами у мережі та вирішення проблем безпеки, але й для виявлення розподілених бот-мереж, які використовуються для створення розподілених атак чи розповсюдження шкідливих додатків. Використовуючи можливість класифікації трафіку рівноправних учасників зазначених мереж, незвичайні інформаційні потоки можуть бути виявлені та, залежно від їх шкідливості чи дозволеної в кожному конкретному випадку поведінки, виділені в окремий сервісний клас, разом із тим встановлюючи інструмент контролю якості (*Quality of Service, QoS*). Складніші підходи у класифікації використовують методи інспекції корисного навантаження, які узагальнено називаються методами глибокої інспекції пакетів (*Deep Packet Inspection, DPI*).

Необхідність глибшого вивчення природи мережевого трафіку неодноразово висвітлювалась у роботах Shiravi A., Rahbarinia B., Sen S., Karagiannis T. Саме P2P-мережі та об’єднання їх у системи-учасники вивчались переважно Wang J., Zhou L., Nguyen T., Zuev D., Gupta P. На пострадянському просторі як найвагоміший внесок можна відзначити роботи, виконані Васильєвим А., Гетьманом І., Бугаєвим А., Дорт-Гольцем А., Семеновим Ю. Передові досягнення в сфері ідентифікації мережевих загроз обговорювались Євстроповим Є. та Маркіним Ю. у їх авторських працях.

Незважаючи на це, залишаються невирішеними ряд задач з виявлення трафіку, створеного в однорангових мережах з урахуванням принципів взаємодії рівноправних учасників. Дисертаційна робота спрямована на підвищення ефективності керування такими інформаційними потоками в мережах загального призначення. Виходячи із зазначеного вище, тема дослідження є актуальною і становить науковий та практичний інтерес.

У роботі запропоновано використовувати як метод класифікації однорангової взаємодії алгоритми J48 і REPTree у реалізації AdaBoost. Останні досягнення в сфері машинного навчання класифікаційних нейромереж підтверджують його

ефективність для аналізу статистичних характеристик потоків, які є одночасно незалежними від номеру порту чи специфіки корисного навантаження. Особлива ефективність класифікації досягається за умови використання абстрактного мережевого пакетного фільтра та математичної моделі магістрального каналу зв'язку на основі агрегованих заголовків протоколів.

У цьому дослідженні процедура знаходження P2P-трафіку забезпечується шляхом його класифікації за типом, згідно з якою він являє собою однорангову взаємодію або не є нею, незалежно від того, містить він шкідливі вкладення чи ні. Як вже зазначалося, основною проблемою класифікації P2P-трафіку є складність застосування традиційних технік ідентифікації, які головним чином базуються на аналізі номерів портів або типу навантаження. Вони стають неефективними проти додатків, що вимагають випадкових номерів портів чи шифрують тіло своїх запитів. Це дослідження, завдяки поєднанню двох підходів класифікації та додаванню гнучкості у виборі кожного з них до відповідного профілю трафіку з використання методів машинного навчання як уточнюючого фактора (з застосуванням *Group method of data handling, GMDH* та алгоритмів J48 і REPTree у реалізації AdaBoost), досягає високої ефективності, результатів якої достатньо для створення сервісної карти мережі та виявлення підозрілої активності. Водночас завдяки цьому досягається повна незалежність від типу протоколів чи характеру навантаження.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційна робота виконана в рамках таких науково-дослідних робіт: «Дослідження можливостей створення обчислювального кластеру на платформі WINDOWS Compute Cluster Server 2003 (CCS) на базі класу персональних комп'ютерів» у рамках договору № 07-008/М, укладеного між ТОВ «Майкрософт Україна» та Київським національним університетом імені Тараса Шевченка (2006 – 2007 рр.); «Засоби забезпечення надійності збереження даних та керування ресурсами в інфраструктурі для наукових та освітніх установ України на базі технології IBM Mainframe» (номер держреєстрації 0111U005489, 2011 – 2012 рр.), в яких автор був виконавцем.

Мета та завдання дослідження. Основною метою дисертаційного дослідження є підвищення ефективності контролю мережевого трафіку через виявлення взаємодії однорангових додатків типу P2P для забезпечення надійної роботи пакетної мережі загального призначення.

Для досягнення поставленої мети у роботі сформульовано і вирішено наступні **науково-прикладні задачі**:

1) проаналізувати, якими моделями можуть бути представлені системи класифікації мережевого трафіку у випадку однорангової взаємодії. Побудувати моделі зазначених взаємодій з урахуванням характерних особливостей магістральних каналів зв'язку. Обґрунтувати механізми визначення побудованих моделей за рахунок застосування альтернативних засобів мережевої класифікації;

2) створити метод визначення однорангової мережевої взаємодії, використовуючи підходи ідентифікації та контрольованого навчання фільтра класифікатора. Встановити, який із методів аналізу чи яке їх поєднання забезпечить виявлення та ідентифікацію рівноправних мережевих додатків з подальшою класифікацією останніх і можливістю виділення їх в окремий сервісний клас;

3) провести експериментальне виявлення однорангової мережевої взаємодії за допомогою встановлених методів і підходів. Оцінити похибку класифікації та ефективність використаних моделей з отриманих практичних результатів та прогнозованих теоретичних обчислень. Формалізувати та втілити в програмному забезпеченні використані методи ідентифікації однорангової мережевої взаємодії на основі застосування технік автоматичного машинного навчання.

Об'єктом дослідження є процес ідентифікації взаємодії рівноправних вузлів в одноранговій розподіленій мережевій архітектурі з визначенням оптимального алгоритму навчання моделі фільтра класифікатора для вирішення задачі мережевої класифікації.

Предметом дослідження є математичні моделі систем класифікації мережевого трафіку, методи виявлення та ідентифікації однорангової взаємодії в мережах загального призначення.

Методи дослідження. Метод системного підходу до дослідження мережевої взаємодії типу P2P як цілісної множини однорангових протоколів використовується в роботі для аналізу та синтезу сукупності відношень і зв'язків між елементами мережі та наступної класифікації такої взаємодії; під час моделювання пакетного трафіку використовувалися апарат теорії ймовірності та випадкових процесів, метод групового врахування аргументів та математичної статистики, імітаційне моделювання; в процесі аналізу ефективності роботи класифікатора використовувалися методи автоматичного машинного навчання класифікаційної мережі для дослідження процесів ідентифікації цілісних інформаційних потоків рівноправних додатків.

Наукова новизна одержаних результатів полягає в наступному:

1) вперше розроблено математичну модель абстрактного мережевого пакетного фільтра з можливістю класифікації однорангової взаємодії типу точка-точка, що дало можливість проводити ідентифікацію мережевих додатків, детальніше контролювати процеси, що відбуваються в мережі, та реагувати на мережеві аномалії швидше, тим самим підвищуючи відмовостійкість системи;

2) вперше розроблено та обґрунтовано математичну модель магістрального каналу зв'язку на основі тільки заголовків транзитних мережевих пакетів та протоколів обміну даними, що дало змогу суттєво розширити сферу застосування мережевого класифікатора шляхом його встановлення на пограничних мережевих елементах. Використовуючи лише заголовки мережевих пакетів, об'єм зібраних статистичних даних на порядок менший загального інформаційного потоку через інтерфейс класифікатора;

3) удосконалено підхід в уточненні запропонованих моделей шляхом комбінаційного поєднання методів ідентифікації однорангової мережевої взаємодії та контрольованого навчання фільтра класифікатора, що дало можливість проводити аналіз мережевих пакетів на швидкостях, які раніше були доступними лише для спеціалізованого апаратного забезпечення.

Практичне значення одержаних результатів:

1) розроблено комп'ютерну модель мережевого пакетного фільтра з можливістю класифікації однорангової взаємодії типу точка-точка, реалізовано програмні засоби та впроваджено в роботу систему класифікації мережевого трафіку, що забезпечує ідентифікацію та визначення параметрів таких взаємодій;

2) досліджено сценарії застосування механізму класифікації для створення гнучких тарифних політик, основні напрацювання яких використано для побудови системи ідентифікації на основі визначення типу додатку, що використовує користувач. Завдяки цьому досягається повна незалежність від типу протоколів чи характеру навантаження;

3) вдосконалено та реалізовано у вигляді програмного забезпечення формалізовані методи ідентифікації однорангової мережевої взаємодії на основі технік автоматичного машинного навчання, що дало змогу з точністю в 95-98 % ідентифікувати взаємодію окремих додатків та було використано для створення сервісної карти мережі і виявлення підозрілої активності.

Розроблені та уточнені в процесі роботи моделі та методи реалізовано у вигляді самостійних програмних модулів, які можуть бути використані в подальших практичних дослідженнях в сфері управління та оптимізації роботи комп'ютерних мереж. Метод ідентифікації однорангової мережевої взаємодії на основі застосування технік автоматичного машинного навчання дає можливість провести навчання мережевого класифікатора за окремими випадками передачі даних з їх наступним віднесенням до відповідних груп, які формують кластери аналогічних характеристик трафіку. Практична цінність результатів дисертаційної роботи полягає у суттєвому підвищенні можливостей комплексної діагностики мереж загального призначення для забезпечення їх надійної та ефективної роботи. Таким чином, на основі отриманих даних класифікованих однорангових взаємодій та з використанням запропонованих методів аналізу стає можливим блокування небажаних додатків та/або шкідливого програмного забезпечення.

Практична цінність роботи підтверджена актами впровадження та висновками експертних комісій у комерційних установах та організаціях: СП «International Telecommunication Company», ПрАТ «МТС Україна». Крім того, результати дисертаційного дослідження впроваджено в навчальний процес ЧНУ імені Богдана Хмельницького.

Застосовані підходи, розроблені моделі та методи реалізовано у вигляді самостійного модуля системи класифікації мережевого трафіку OpenDPI. Документація та вихідні коди програмного забезпечення розміщено на загальнодоступному ресурсі git://gitlab.com/kdeev.

Особистий внесок здобувача. Дисертація є самостійно виконаною завершеною роботою здобувача. Усі результати, що представлені в дисертаційній роботі, отримані автором самостійно. Із робіт, опублікованих у співавторстві, до дисертації увійшли лише ті результати, що належать автору. У роботах, опублікованих у співавторстві, автором одержано такі результати: [1] – проведено аналіз існуючих проблем мережевої класифікації та зроблено порівняння трьох методів для класифікації P2P-додатків; [3] – висвітлено один із методів підвищення продуктивності у системах мережевої класифікації, використовуючи попередню компіляцію правил фільтрації; [4] – створено та реалізовано гнучкий набір інструментів для обробки мережевих пакетів; [5] – описано метрики оцінки класифікації мережевих пакетів при формуванні гнучких політик у системах контролю якості обслуговування у мережах загального призначення, використовуючи рівень додатків як ідентифікатор; [6] – сформовано основні засади методу визначення P2P-взаємодії за допомогою задання правил класифікації у вигляді регулярних виразів; [7] – проаналізовано можливість застосування альтернативних методів аналізу мережевої взаємодії.

Апробація результатів дисертації. Результати досліджень та основні положення дисертаційної роботи було представлено та обговорено на:

- Міжнародній конференції студентів, аспірантів та молодих вчених з прикладної фізики, Київ, Україна, 2013 р.;
- Міжнародній науковій конференції студентів, аспірантів та молодих вчених «Шевченківська весна», Київ, Україна, 2014 р.;
- Міжнародній науковій конференції «Електроніка та прикладна фізика», Київ, Україна, 2015 р.
- Міжнародній науково-практичній конференції «FOSS Lviv 2016», Львів, Україна, 2016 р.

Публікації. Основні результати дисертаційної роботи опубліковано в 7 статтях у фахових виданнях (з них дві публікації у виданнях, включених до наукометричної бази Index Copernicus), 9 тезах доповідей на науково-практичних конференціях, основні з яких наведено в авторефераті.

Структура дисертаційної роботи. Дисертація складається зі вступу, чотирьох розділів, висновків, що містять основні результати роботи, списку використаних джерел, що містить 224 посилання, та трьох додатків. Повний обсяг дисертації становить 183 сторінки, із них 134 сторінки основного тексту. Робота містить 42 рисунки та 10 таблиць.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У **вступі** обґрунтовано актуальність обраної теми дисертації, сформульовано мету та задачі дослідження, визначено зв’язок з науковими програмами і темами, сформульовано наукову новизну та практичне значення отриманих результатів, наведено інформацію щодо впровадження результатів роботи, відомості про апробацію роботи, а також основні положення, що виносяться на захист.

У **першому розділі** детально розглянуто принципи роботи розподілених комп’ютерних мереж. Окрема увага приділяється механізмам управління мережевим трафіком у таких системах. Зроблено огляд загальних принципів вирішення задач пакетної класифікації в системах передачі інформації в комп’ютерних мережах. Застосовані методи забезпечення надійності класифікації на основі існуючих засобів аналізу та управління не враховують специфічних особливостей функціонування одноранговых мереж. У розділі обґрунтовується необхідність проведення класифікації пакетів, що поширюються мережею загального призначення. Проведено аналіз моделей представлення і методів опису реальних інформаційних потоків у мережах. Досліджено методи ідентифікації потоків даних одноранговых мереж.

У найближчій перспективі ефективнішим є застосування одноранговых мереж, на противагу класичній клієнт-серверній архітектурі. В одноранговых мережах кожний додаток як рівноправний учасник виступає в ролі клієнта та/або в ролі сервера, що дає можливість розподілити навантаження між усіма учасниками і, таким чином, потенційно зменшити об’єми трафіку в каналах передачі та суттєво змінити напрямок інформаційних потоків у мережі обміну даними. Встановлено, що реальний трафік у комп’ютерній мережі корелює з типом та архітектурою реалізації додатку, що його породив, і має властивість післядії, тобто є самоподібним процесом. Проведено огляд засобів і технічних рішень класифікації мережевих пакетів зважаючи на важливість аналізу та ідентифікації саме одноранговых додатків для комп’ютерних мереж. Досліджено апаратні комплекси мережевого аналізу та аналогічні за функціоналом програмні реалізації класифікаторів.

Проведений аналіз виявив недоліки відомих методів аналізу класифікації мережевих пакетів, за умови їх застосування до взаємодії одноранговых додатків, що дало можливість окреслити напрямок і сформулювати задачі дослідження. Необхідність створення ефективної системи класифікації одноранговых додатків у мережах загального користування визначається стрімким поширенням розподіленої архітектури рівноправних мережевих додатків, проте до роботи такої системи висувається ряд вимог, таких як висока точність виявлення взаємодій типу точка-точка, надійність, простота експлуатації та можливість до розширення і вдосконалення з часом. З позиції цих критеріїв виконано аналіз існуючих механізмів класифікації P2P-трафіку, виявлені можливі шляхи оптимізації застосування останніх на магістральних каналах зв’язку мереж загального призначення.

У **другому розділі** з використанням апарату методу групового врахування аргументів (МГВА) розроблено математичну модель абстрактного мережевого пакетного фільтра з можливістю класифікації одноранговой взаємодії.

За схемою реалізації алгоритми МГВА схожі на процес навчання систем розпізнавання зображень, що загалом використовують формалізовані залежності між параметрами системи за допомогою персептронів чи нейромереж. Зазвичай для простих моделей можна отримати точний аналітичний розв'язок, але для складних

систем, якими є мережеві класифікатори, отримати аналітичне рішення інколи взагалі неможливо. У цьому випадку використовується імітаційне моделювання.

Методи математичного моделювання припускають заміну досліджуваної системи або процесів відповідною математичної моделлю зі збереженням основних її характеристик. Для визначення структури і параметрів моделі отриманої за МГВА використовують динамічні алгоритми уточнення коефіцієнтів, коли структура моделі попередньо відома можуть застосовуватися алгоритми параметричної ідентифікації.

Алгоритми побудови моделей дозволяють створювати моделі досліджуваних процесів з високою точністю, але ці моделі є нефізичними, їх структура заздалегідь невідома і в процесі побудови моделі може постійно змінюватися. Для нефізичних моделей мережевого трафіку у розглянутому випадку з одноранговою взаємодією та відповідною класифікацією МГВА забезпечує найефективніше використання обчислювальних ресурсів. Індуктивні методи дають унікальну можливість автоматично знаходити взаємозалежності в даних, вибирати оптимальну структуру моделі, підвищувати точність класифікації у застосованих алгоритмах тощо. Ключовою відмінністю алгоритмів МГВА порівняно з аналогічними підходами для побудови ефективної системи мережевої класифікації є можливість знаходження оптимальної складності структури моделі в залежності від наявних апаратних ресурсів. Отже, перевагою таких алгоритмів є можливість побудови моделей з урахуванням суб'єктивних особливостей конкретної системи і умов її функціонування, автоматичний вибір структури моделі і її висока ступінь точності. Недоліками таких моделей є відсутність явно вираженого фізичного сенсу, наявність особливостей у проведенні аналізу та труднощі у використанні таких моделей для проведення мережевої класифікації через необхідність застосування аналізу стійкості в режимі реального часу. Таким чином, у роботі в якості еволюційного алгоритму було обрано до використання нейронну мережу з автоматичним навчанням та ряд алгоритмів з самоорганізацією структури зв`язків. У запропонованій моделі будь-які ознаки мережевої взаємодії, що можуть мати вплив на вихідний результат класифікації, використовуються як вхідні аргументи. Інтерпретаційні взаємозв’язки у протокольних заголовках визначаються ще до аналізу даних, формулюючи тим самим набір вхідних змінних. Реалізований алгоритм має багаторядну структуру, завдяки чому можливе використання паралельних обчислень в їх програмній реалізації. На наступний рівень класифікаційного відбору передається не один, а декілька найкращих результатів класифікації мережевої взаємодії, що може використовуватися для підвищення точності моделі чи механізму їх імітаційного моделювання. Отримання моделі абстрактного мережевого пакетного фільтра з можливістю класифікації однорангової взаємодії типу точка-точка складається з кількох етапів, кожен із яких починається означенням ряду селекції, а закінчується формуванням групи найточніших моделей.

Запропонована математична модель виконана в формі поліноміальних функцій представлення пакетного заголовка та протокольних повідомлень мережевого рівня. В цьому випадку обробка граничних умов щодо ідентифікації однорангової взаємодії відбувається шляхом застосуванням непараметричних методів та вибором структури моделі. У непараметричних методах, завданням яких є задача ідентифікації, визначення вхідних параметрів перестає бути ключовим, натомість функцій трансформації представляють удосконалену модель. В цьому випадку відомо, що рівняння лінійної стаціонарної моделі (1) можна виразити за допомогою лінійного оператора, врахувавши наявність точок мінімуму на тестовій послідовності навчальної вибірки:

$$y(t) = \int_0^t h(\varepsilon) \cdot x(t-\varepsilon) d(\varepsilon), \text{ де } h(\varepsilon) \text{ – перехідна функція} \quad (1)$$

при значеннях заголовків кожного окремого мережевого пакета $x(t)$ та функції ідентифікації $y(t)$ навчальної вибірки $t \in T$, T – інтервал між вибірками.

Таким чином розглянута модель може бути виражена у вигляді диференціального рівняння (2):

$$\sum_{i=0}^{n} a_i \cdot y^i(t) = \sum_{j=0}^{m} b_i \cdot y^j(t), m < n \quad (2)$$

або вона може бути представлена у вигляді еквівалентного рівняння з передаточною функцією $H(p)$ (3):

$$H(p) = \frac{\sum_{j=0}^{n} b_j \cdot p^j(t)}{\sum_{i=0}^{n} a_i \cdot p^i(t)}. \quad (3)$$

Якщо в моделі передбачені протокольні параметри за наперед визначеними зміщеннями в пакетних заголовках та вони задовольняють умові (2), то задача ідентифікації визначається наступними параметрами: $a_1, a_2, a_3, \ldots, a_n$ та $b_1, b_2, b_3, \ldots, b_m$. У випадку, коли існує інформація про структуру математичної моделі, буде використана ідентифікація з параметрами. На противагу цьому, у загальному випадку, коли не існує ніякої інформації про модель, використовується ідентифікація без параметрів. Визначення значущих параметрів відомої моделі легше і простіше, ніж ті самі визначення, але для невідомої моделі.

Точною моделлю ідентифікованого об'єкта $y = F(x)$, як правило, вважається безперервний оператор трансформації, в кожному випадку точність ідентифікованої моделі може бути гарантована лише для мережевих пакетів, які мало відрізняються від попереднього тестового набору. Лінійні ідентифіковані моделі можна записати в загальному вигляді у відповідності до функціональної теорії. Для нелінійних операторів немає ніякого загального вигляду. Проте кожне нелінійне трансформування є інваріантним і безперервний оператор може бути наближеним з заданою точністю за допомогою функціонального полінома (4):

$$y = h_0 + \sum_{i=0}^{N} h_i(\varepsilon_1, \ldots, \varepsilon_i) \prod_{r=1}^{i} x(t-\varepsilon_r) d\varepsilon_1, \ldots, d\varepsilon_i), \quad (4)$$

де $h_i(\varepsilon_1, \ldots, \varepsilon_i)$ задовольняє умові $h_i(\varepsilon_1, \ldots, \varepsilon_i) = 0$, якщо $\varepsilon_j < 0$.

При $J = 1 \div 3$ ці функції називаються ядрами Вольтерра. Завдання ідентифікації однорангової взаємодії полягає у визначенні параметрів цих змінних із наведених тестових вибірок, використовуючи багатовимірне перетворення Лапласа (5) функціональним поліном Вольтерра, що може бути записане в такому вигляді:

$$y = h_0 + \sum_{i=1}^{N} \left(\frac{1}{2\pi j}\right) \cdot \int_{-\infty}^{+\infty} H_i\,(p_1, \ldots, p_i) \cdot \prod_{r=1}^{i} X(P_r) e^{\sum_{r=1}^{i} P_r} dp_1, \ldots, dp_i)\,, \qquad (5)$$

де образи ядер Вольтерра $H_i(p_1, \ldots, p_i)$ відображають багатовимірні функцій передачі і записуються у вигляді параметра (6):

$$H_i(p_1, \ldots, p_i) \;=\; \frac{\sum_{r_1=0}^{m} \cdots \sum_{r_i=0}^{m} b_{r_1} \ldots b_{r_i} \cdot p_1^{r_1} \ldots p_i^{r_i}}{\sum_{r_1=0}^{n} \cdots \sum_{r_i=0}^{n} b_{r_1} \ldots b_{r_i} \cdot p_1^{r_1} \ldots p_i^{r_i}}\,. \qquad (6)$$

Завданням класифікації мережевих пакетів за типом є ідентифікація цих параметрів і необхідність визначення коефіцієнтів чисельника і знаменника відповідно до експериментальних даних. Слід зазначити, що ця задача вимагає складного розрахунку (з урахуванням багатьох параметрів). У зв`язку з цим доцільним є виконання спрощень щодо нелінійності моделі та граничних умов застосування фільтру мережевого класифікатора. Оптимальний критерій був обраний із наступних функціональних рівнянь, записаних у вигляді (7):

$$Q(y, \tilde{y}) = \overline{y(t) - \tilde{y}(t)^2} \rightarrow \; min \qquad (7)$$

Особливістю підходу самоорганізації є його успішне функціонування в умовах навантажень, які в кілька разів перевищують показники при стаціонарному режимі роботи класифікатора. Досягнення мінімуму ансамблю критеріїв селекції при формуванні математичної моделі сигналізує про отримання точної моделі. В ансамбль критеріїв селекції включають різні критерії селекції. Ці критерії мають недоліки, які компенсуються їх спільним використанням. Ансамбль критеріїв селекції дозволяє зробити вибір моделі однозначним. Відбір вихідних параметрів здійснюється за критерієм максимальної точності для попереднього ряду селекції, що був переведений на наступний рівень. Кількість моделей визначається за наступною формулою (8):

$$C_N^K = \frac{N!}{K! * (N-K)!} \qquad (8)$$

Оскільки МГВА є еволюційним методом, він реалізує підхід самоорганізації автоматично. Опис досліджуваної системи (9) замінюється власними описами вигляду (10):

$$\varphi = f_1(x_1, x_2, x_3, \ldots, x_i) \qquad (9)$$

$$Y_1 = f_1(x_1, x_2),\; y_1 = f_2(x_2, x_3), \ldots, y_m = f_1(x_{n-1}, x_n), \text{ де } m = C_n^2 \qquad (10)$$

$$Z_1 = f_1(y_1, y_2),\ z_1 = f_2(y_2, y_3), \ldots, z_p = f_1(x_{m-1}, x_m),\ \text{де}\ p = C_m^2$$

Способи конструювання опису досліджуваної системи в алгоритмах МГВА відрізняються за типом базисних функцій. Найбільш поширеними є алгоритми, в яких використовуються поліном другого ступеня, лінійні поліноми, а також імовірнісні алгоритми. Поліном другого ступеня використовують в алгоритмах, призначених для побудови моделей складних систем. Алгоритм із лінійним поліномом з чотирма аргументами має вигляд (11):

$$Z = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + \ldots + a_{16}x_{16},\ \text{де} \qquad (11)$$

$x_1 = x_1, x_2 = x_2, \ldots, x_5 = x_1x_2, x_6 = x_1x_3, \ldots, x_{16} = x_1x_2x_3x_4$

Можлива заміна вихідного полінома (11) рядами часткових лінійних поліномів (12):

$$\begin{aligned} Y_1 &= b_0^1 + b_1^1x_1 + b_2^1x_2 \\ Y_2 &= b_0^2 + b_1^1x_3 + b_4^2x_4 \\ &\vdots \\ Y_8 &= b_0^8 + b_1^8x_{15} + b_2^8x_{16} \end{aligned} \qquad (12)$$

Вхідна вибірка даних являє собою набір агрегованих пакетних заголовків мережевої взаємодії, яка містить N значень (записів) спостережень множини з P параметрів. Вихідні змінні визначаються наперед та залежать від сервісного профілю мережевого трафіку досліджуваної мережі. На наступному рівні перебираються всі моделі відповідно до схеми, зображеної на рис. 1.

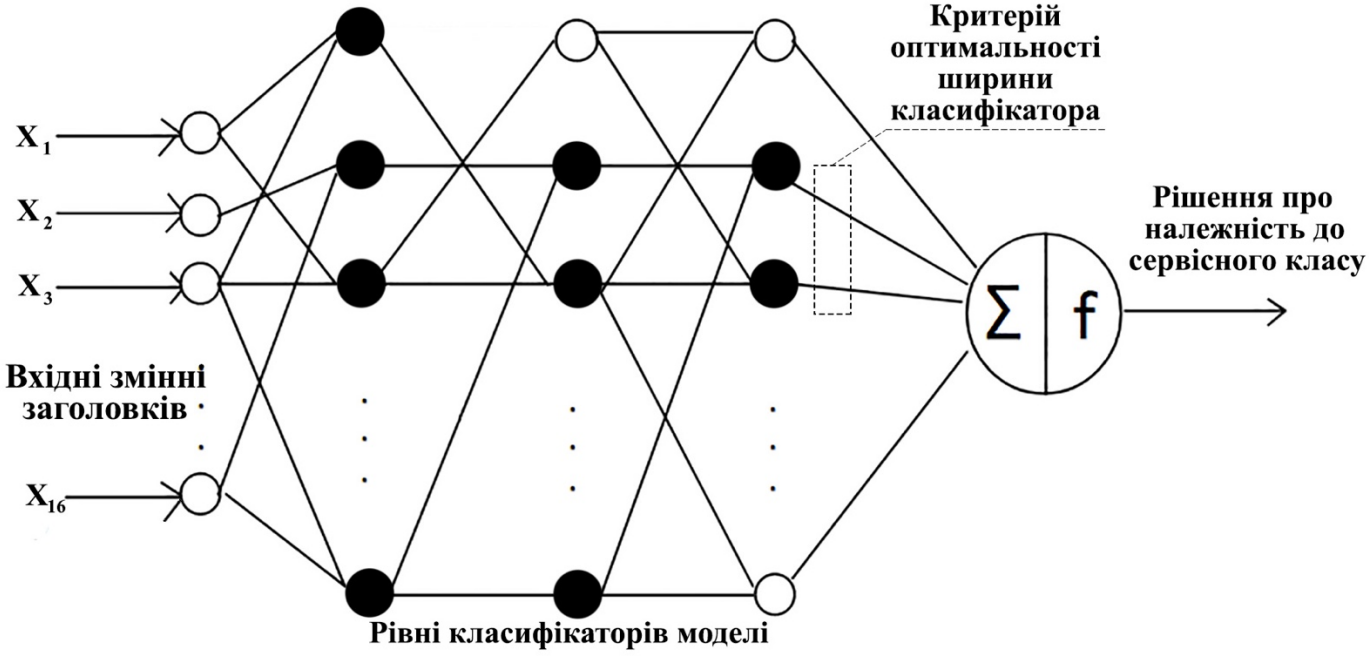

Рис. 1. Схема перебору нелінійних змінних у вхідній вибірці заголовків

У стандартний ансамбль критеріїв селекції зазвичай включають наступні критерії:

- Критерій мінімуму зміщення (13) або критерій несуперечності. Цей критерій дозволяє вибирати модель, яка буде співпадати з моделлю, отриманою за даними іншого вимірювального інтервалу

$$n^2_{\text{зм.}} = \frac{1}{n}\sum_{t\in N}(y_t^A - y_t^B)^2 \rightarrow min \qquad (13)$$

- Критерій регулярності (14) обчислює середньоквадратичне відхилення моделі на тестовій перевірочній вибірці:

$$\Delta^2(B) = \frac{\sum_{t\in N}(y_t^P - y_t)^2}{\sum_{t\in N} y_t^2} \rightarrow min \qquad (14)$$

Критерій регулярності може бути застосований при побудові моделей, які використовуються для короткострокового прогнозу. Вплив втрачених змінних можна в наступних розрахунках врахувати через інші змінні. На особливу увагу заслуговує комбінаторний алгоритм МГВА, який використано для побудови системи класифікації мережевого трафіку та виділення взаємодії однорангових додатків в окремий сервісний клас.

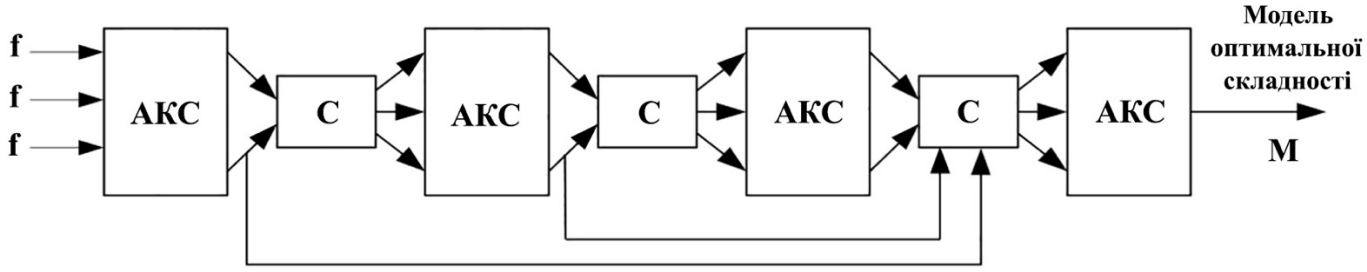

Рис. 2. Визначення оптимальної моделі класифікації мережевих пакетів за МГВА

На рис. 2 зображено схему уточнення моделі магістрального каналу зв'язку на основі заголовків IP за допомогою використання альтернативних засобів аналізу мережевої взаємодії, f – базисні функції, АКС – ансамбль критеріїв селекції, С - механізм комбінування моделей. Результатом такого моделювання є модель оптимальної складності **M**, яка для трафіку комп'ютерної мережі загального призначення та однорангової взаємодії може ідентифікувати такий обмін і віднести його до відповідного сервісного класу.

У **третьому розділі** запропоновано шляхи вдосконалення методів виявлення та ідентифікації однорангової мережевої взаємодії на основі класифікації трафіку за типом. Сформовано формалізований алгоритм тренувального навчання класифікатора без учителя в умовах однорангової мережі та трафіку типу точка-точка. Для формалізації виконання етапів аналізу мережевої взаємодії застосовано модель магістрального каналу зв'язку на основі заголовків IP (розглянута раніше, з урахуванням параметричних особливостей однорангових мереж, які використовуються в умовах скінченності кількості учасників інформаційного

обміну). Нейромережевий класифікатор однорангової взаємодії являє собою нейромережу Кохонена і використовує механізми конкуренції для автоматизованого навчання. При подачі на вхідний шар нейромережі вектора взаємодій перемагає той нейрон, вектор ваги якого найбільш подібний до вхідного вектора досліджуваної взаємодії. Для нейрона-переможця виконується співвідношення (15):

$$d(x, w') = \min_{1<i<n} d(x, w_i), \tag{15}$$

де n – кількість нейронів, w_i – вектор ваги нейрона-переможця,

$d(x, w_i)$ – відстань між векторами x та w_i .

Як міра відстані використовується Евклідова міра (16):

$$d(x, w_i) = \|x - w_i\| = \sqrt{\sum_{j=1}^{n}(x - w_{ij})^2}. \tag{16}$$

Дискретний характер вихідної змінної персептрона, що вказує на належність мережевого пакета до того чи іншого класу, унеможливлює чіткіше урахування точності ідентифікації для вибору структури самого персептрона. Після дискретизації вхідних векторів отримання нефізичних моделей неможливе, тому що тільки неперервні змінні дають змогу знайти мінімум зовнішнього критерію, який визначає оптимальну структуру нефізичної моделі. Структура нейромережі тим простіша, чим менша дисперсія завад. У нашому випадку під завадами може розглядатися бажання однорангового додатка замаскувати свою взаємодію чи використання динамічно визначених діапазонів портів транспортного рівня під час інформаційного обміну. В той самий час збільшення кількості вхідних нейронів рівнозначне зменшенню перешкод. Структура нефізичної моделі при зростанні вибірки наближається до структури фізичної моделі. Мережевий трафік може бути представлений великою кількістю нефізичних моделей, що будуть залежати від дисперсії перешкод та довжини вибірки. Якщо розглядати механізм ідентифікації конкретного програмного додатка однорангової мережевої взаємодії та отриману модель такого обміну, отримання класифікаційної нейромережі можливе не лише шляхом виключення деяких одиничних пакетів та невідомих особливостей реалізації транспортних протоколів, а й випадково, так щоб для конкретної взаємодії деяких P2P-додатків отримати глибший мінімум зовнішнього критерію визначеної моделі. Радіус навчання визначає, яка кількість нейронів, крім самого нейрона-переможця, буде брати участь у навчанні, тим самим змінюючи свою вагу при конкретній ітерації. Радіус навчання набуває найбільшого значення на першій ітерації і поступово зменшується зі збільшенням кількості ітерацій таким чином, що наприкінці навчання тільки нейрон-переможець коригує свою вагу. Вага нейрона-переможця та всіх нейронів, розташованих в області радіуса навчання, визначається (17), виходячи з правила Кохонена:

$$w_i^{(k+1)} = w_i^{(k)} + v_i^{(k)}\left[x - w_i^{(k)}\right], \quad (17)$$

де $i = \overline{1,n}$, x – вхідний вектор, k – номер циклу навчання, $v_i^{(k)}$ – коефіцієнт навчання i-го нейрона з радіуса навчання в k-му циклі.

Коефіцієнт швидкості навчання $v_i^{(k)}$-го нейрона в k-му циклі навчання розбивається на дві частини (15): функцію сусідства $G_i(d_i, k)$ і функцію швидкості навчання $v(k)$:

$$v_i^{(k)} = G_i(d_i, k) \cdot v(k). \quad (18)$$

Функція сусідства дає змогу досягти незмінності ваги нейронів, що знаходяться за межами радіуса навчання. Як функція сусідства використовується Гаусівська функція (19):

$$G_i(d_i, k) = e^{-d_i/2\sigma(k)}. \quad (19)$$

Тут d_i – відстань між векторами ваги i-го нейрона та нейрона-переможця.

При цьому $\sigma(k)$ є спадною функцією від номера циклу навчання. Будемо використовувати функцію $\sigma(k) = 1/k$, монотонно спадну від номера циклу навчання. Визначено також функцію швидкості навчання $\varepsilon(k)$, яка також являє собою функцію, що спадає від номера циклу навчання. Будемо використовувати функцію виду $\varepsilon(k) = e^{-k}$. Застосування функції $v_i^{(k)}$ дає можливість домогтися того, що всі вектори з навчальної вибірки роблять приблизно однаковий внесок у результат навчання. Навчання складається з двох основних етапів: на першому етапі воно проводиться з досить великими значеннями швидкості і радіуса, що дає змогу розташувати вектори ваги нейронів відповідно до розподілу прикладів у навчальній вибірці. На другому етапі необхідно провести точне налаштування ваги нейронів для значень параметрів швидкості навчання, що набагато менші від початкових. Навчання триває до того моменту, поки похибка квантування при вхідних векторах не стане досить малою величиною (w' – вектор ваги (20) нейрона-переможця):

$$E = \frac{1}{p}\sum_{i=1}^{p}\|x_i - w'\|^2. \quad (20)$$

При навчанні мережі Кохонена існує проблема незадіяних нейронів. Однією з особливостей будь-якого конкуруючого шару є те, що деякі нейрони виявляються незадіяними, тому нейрони, в яких початкові вектори ваги значно віддалені від векторів входу, ніколи не виграють конкуренції незалежно від тривалості навчання. Внаслідок цього вхідні вектори будуть інтерпретуватися меншою кількістю нейронів, а похибка квантування буде збільшуватися. Тому необхідно налаштувати мережу класифікатора так, щоб міг перемагати кожен з нейронів мережі. Для цього алгоритм навчання було модифіковано таким чином, щоб нейрон-переможець час від часу втрачав активність. Механізм обліку активності нейронів виражається в

підрахунку потенціалу кожного нейрона в процесі навчання. Спочатку нейронам присвоюється потенціал p_i, де $p_i(0) = 1/n$ і n відповідає кількості нейронів (кластерів). Значення потенціалу змінюється кожного разу після подачі вхідного вектора x.

В k-му циклі навчання для нейрона-переможця потенціал визначається за (21):

$$p'(k) = p'(k-1) - p_{min}, \tag{21}$$

де $p'(k)$ – потенціал нейрона-переможця в k-му циклі навчання, p_{min} – мінімальний потенціал, що допускає участь у конкурентній боротьбі, він задається в межах від 0 до 1. Для всіх інших нейронів потенціал визначається за правилом (22):

$$p_i(k) = p_i(k-1) + \frac{1}{n}, \tag{22}$$

де n – кількість нейронів, i – порядковий номер нейрона.

Якщо значення потенціалу $p_i(k)$ опускається нижче рівня p_{min}, то нейрон не розглядається. Переможець визначається серед решти нейронів, для яких $p_i \geq p_{min}$. Вибір конкретного значення p_{min} дає змогу встановити поріг готовності нейрона до конкурентної боротьби. При $p_{min} = 0$ нейрони не виключаються з боротьби, що призводить до появи «мертвих» нейронів. При $p_{min} = 1$ нейрони перемагають по черзі, тому що в кожному циклі навчання тільки один із них готовий до боротьби. Практичні дослідження показують, що хороший результат виходить при $p_{min} \approx 0{,}75$. У мережі Кохонена вхідні значення необхідно нормувати. Для цього використовується формула (23):

$$x_{norm.} = \frac{x_i}{\sqrt{\sum_{j=1}^{n} x_j^2}}, \tag{23}$$

де x_i – нормований компонент вхідного вектора, n – кількість нейронів.

Другий шар нейромережі ідентифікації мережевої взаємодії на основі застосування технік машинного навчання здійснює кластеризацію вхідного простору образів класів, у результаті чого утворюються кластери, кожному з яких відповідає свій нейронний вихідний елемент. Третій шар складається з двох лінійних нейронних елементів, які використовують лінійну функцію активації. Цей шар здійснює процедуру остаточного рішення про належність мережевого пакета даних до взаємодії типу точка-точка в одноранговій мережі.

У **четвертому розділі** проведено реалізацію методів і моделей взаємодії одноранговых мереж та виконано експериментальне дослідження їх ефективності. Теоретичне і практичне дослідження методів та способів підвищення ефективності проведення класифікації в одноранговых мережах для взаємодій типу P2P було використано для підрахунку метрик класифікації таких взаємодій, а також для оцінювання похибок у роботі фільтруючого модуля класифікації. Було приділено

увагу розробці програмних реалізацій, документуванню опрацьованих результатів та їх публікації у відкритих джерелах для подальшого практичного використання. В табл. 1 показано узагальнені метрики підрахунку ефективності роботи нейромережі класифікатора однорангової мережевої взаємодії.

Таблиця 1

Результати метрик класифікації однорангової мережевої взаємодії та порівняння ефективності попередніх і запропонованих методів та моделей

Додаток	Правильність	Похибка	Точність	Повнота	Попередні дослідження	Зміна, %
uTorrent	0.982	0.099	0.956	0.974	0.927	+2.1
BtSync	0.935	0.090	0.947	0.957	0.905	+0.5
Viber	0.990	0.097	0.969	0.950	0.858	+6.1
Skype	0.985	0.088	0.965	0.979	0.917	+1.6
eMule	0.967	0.102	0.966	0.975	0.902	+6.5
Storm	0.999	0.099	0.971	0.982	0.985	+3.6
Waledac	1.000	0.084	0.987	0.986	0.944	+5.6
Zeus	0.982	0.094	0.967	0.975	0.967	+1.5

Проведено експериментальне порівняння використаних методів дослідження мережевої взаємодії та оцінювання їх ефективності для ряду найпопулярніших однорангових додатків (табл. 1). Детально описано впроваджену систему виділення трафіку рівноправних додатків типу P2P в окремий сервісний клас в однорангових мережах загального призначення, при проектуванні та розробленні якої було використано результати досліджень попередніх розділів і результати експериментальних досліджень, проведених на підприємствах та в організаціях. У ході проведення експерименту та при ідентифікації однорангової мережевої взаємодії на основі класифікації трафіку за типом додатку всі з восьми додатків тестового навчального набору було правильно класифіковано та віднесено до відповідного класу однорангової взаємодії. Відповідно до отриманих результатів зроблено висновок про те, що ефективність розроблених моделей та їх поєднання з оптимізованими методами ідентифікації мережевої взаємодії на основі застосування технік машинного навчання за МГВА становить близько 95 %. Появу помилок класифікації й ідентифікації можна пояснити наявністю неоднорідностей та сплесків мережевої активності поза областю аналізу, а також частковою модифікацією окремих протокольних обмінів і помилково визначених системних інформаційних взаємодій.

У додатках до дисертаційної роботи наведено лістинги розроблених програмних засобів, акти впровадження та висновки експертних комісій.

ВИСНОВКИ

У дисертаційній роботі вирішено актуальну науково-технічну задачу, яка полягає в розробці сукупності моделей і методів аналізу мережевого класифікатора

для виявлення та виділення в окремий сервісний клас однорангового P2P-трафіку мережі в автоматичному режимі. Згідно з розробленими моделями визначено оптимальне поєднання методів та підходів кожного з функціональних рівнів, що надає системі модульності та можливості до подальшого розширення. Дослідження, проведене в рамках дисертаційної роботи, дало змогу створити мережевий класифікатор для ідентифікації взаємодії однорангових додатків, що досягає рівня точності, достатнього для створення сервісної карти мереж загального призначення та виявлення підозрілої активності в них. У той же час досягається повна незалежність від типу використаних при взаємодії протоколів чи характеру навантаження.

Найбільш значущі наукові і практичні результати роботи полягають у наступному:

1) проаналізовано моделі представлення трафіку у випадку однорангової взаємодії в системах мережевої класифікації. Обґрунтовано механізми підвищення точності побудованих моделей за рахунок застосування альтернативних засобів класифікації, зокрема, автоматичного навчання класифікатора на тестовій вибірці. Похибка теоретичних результатів порівняно з практичними повністю задовольняє вимоги, які висуваються до аналогічних програмно-апаратних систем, а в деяких випадках – для окремих мережевих додатків, демонструє 3-5 % підвищення точності класифікації;

2) внаслідок виконаних досліджень визначено оптимальне поєднання методів виявлення та класифікації трафіку в одноранговій мережі з домінуючими рівноправними мережевими додатками. Програмний код модуля мережевого класифікатора забезпечує ідентифікацію безпосереднього додатка, що його породив;

3) підвищення точності класифікації отримано за рахунок використання методів автоматизованого машинного навчання по попередньо зібраному тестовому набору даних. Для забезпечення однозначності отриманих у роботі результатів було проведено експериментальне виявлення однорангової мережевої взаємодії за допомогою розроблених моделей та методів. Також було проведено експериментальне виявлення рівноправних мережевих додатків, що дало можливість виділення їх в окремий сервісний клас;

Практична цінність роботи полягає у спрощенні комплексної діагностики мереж загального призначення для забезпечення контролю і підтверджується актами впровадження на підприємствах та в організаціях.

ОСНОВНІ ПОЛОЖЕННЯ ДИСЕРТАЦІЇ ОПУБЛІКОВАНО У НАСТУПНИХ ПРАЦЯХ:

[1] Бойко Ю. В., Дєєв К. С. Методи покращення ефективності для систем високошвидкісної класифікації пакетів // Вісник Харківського національного університету. Серія: математичне моделювання, інформаційні технології, автоматизовані системи управління. Харків, 2014. Вип. 1131. С. 5–12.

[2] Дєєв К. С. Вивчення характеру взаємодії типу точка-точка для класифікації мережевого трафіку // Автоматизовані системи управління та прилади автоматики. Харків, 2015. Вип. 163. С. 94–101.

[3] Дєєв К. С., Бойко Ю. В. Аналіз вмісту IP-пакетів на магістральних каналах зв'язку. Проблема класифікації потоків // Науковий вісник КУЕІТУ. Серія: дизайн, обробка і використання інформації [Нові технології]. Кременчук, 2014. Вип. 1-2 (43-44). С. 64–70.

[4] Дєєв К. С., Бойко Ю. В. Аналіз методів та засобів реалізації пакетної фільтрації для глибокого аналізу мережевих пакетів // Вісник Вінницького політехнічного інституту. Розділ: інформаційні технології та комп'ютерна техніка. Вінниця, 2014. Вип. 6. С. 84–90.

[5] Дєєв К. С., Бойко Ю. В. Аналіз метрик якості обслуговування в комп'ютерних мережах в залежності від характеристик IP взаємодії // Вісник Хмельницького національного університету. Серія: технічні науки. Хмельницький, 2015. Вип. 3. С. 147-152.

[6] Дєєв К. С., Бойко Ю. В. Визначення мережевої взаємодії типу точка-точка за допомогою регулярних виразів // Праці Одеського політехнічного університету. Одеса, 2015. Вип. 2 (46). С. 119–123.

[7] Дєєв К. С., Бойко Ю. В. Аналіз альтернативних апаратних засобів при дослідженні мережевої взаємодії // Вісник Черкаського національного університету ім. Б. Хмельницького. Серія: прикладна математика, інформатика. Черкаси, 2015. Вип. 38 (331). С. 3–11.

[8] Deev K. S. Characterization of CDMA2000 cellular data network traffic // XIII International young scientists' conference on applied physics: ICAP'13 (Kiev, 12–15 June 2013). P. 220–222.

[9] Deev K. S. Developing class specific traffic matrices for packet classification quality-of-service in IP/MPLS networks // IX International scientific conference: Electronics and applied physics (Kiev, 20–22 October 2013). P. 134–135.

[10] Deev K. S. Service-oriented network as result of IP header space analysis // International scientific conference of students and young scientists: Shevchenkivska Vesna 2014 (Kiev, 25–28 March 2014). P. 39–41.

[11] Deev K. S. Considering approaches for flexible packet matching using CAM structures // XIV International young scientists' conference on applied physics: ICAP'14 (Kiev, 14–16 June 2014). P. 199–200.

[12] Deev K. S. Predictive admission for guaranteed QOS data flows based on traffic classes // XIV International young scientists' conference on applied physics: ICAP'14 (Kiev, 14–16 June 2014). P. 201–202.

[13] Deev K. S. Approaches of heuristic traffic classification by packets parameter weight determination // X International scientific conference: Electronics and applied physics (Kiev, 22–25 October 2014). P. 67–70.

[14] Deev K. S. Consider using Cisco OnePK framework for packet headers manipulation using DPSS // International scientific conference of students and young scientists: Shevchenkivska Vesna 2015 (Kiev, 2–3 April 2014). P. 59–61.

[15] Deev K. S. Applying deep packet inspection methods for optimizing network // XIII International young scientists' conference on applied physics: ICAP'15 (Kiev, 10-13 June 2015). P. 128-129.

[16] Deev K. S. VyOS Introduction // VI Scientific conference free open-source software: FOSS'16, (Lviv, 19–22 April 2016). P. 22–24.

АНОТАЦІЯ

Дєєв К. С. Дослідження мережевої взаємодії за допомогою системи глибокого аналізу пакетів. – На правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.05 – комп'ютерні системи та компоненти. – Черкаський державний технологічний університет, Черкаси, 2018.

У дисертації запропоновано методи забезпечення ідентифікації мережевої активності однорангових додатків для наступної їх класифікації та виділення в окремий сервісний клас такої взаємодії з метою проведення гнучкої тарифікації у мережі оператора Інтернет-послуг. Реалізація створених моделей і методів у програмних засобах та їх впровадження у тестових сегментах розподілених мереж операторів послуг дало можливість значною мірою підвищити ефективність надання послуг користувачам та використовувати оптимальні політики управління мережевим трафіком. У роботі окреслено підходи до стандартизації та реалізації у програмних платформах функцій аналізу пакетного навантаження згідно з їх представленням у вигляді багаторівневої моделі взаємодії відкритих систем OSI. Визначено функціональний рівень та надано рекомендації щодо оптимального розміщення комплексу аналізатора в операторській мережі загального призначення. Розглянуто можливості часткової віртуалізації окремих компонентів системи з метою підвищення загальної пропускної здатності.

Ключові слова: глибока інспекція мережевих пакетів, аналіз трафіку, класифікація пакетів, система запобігання вторгненням, засоби моніторингу мережі.

АННОТАЦИЯ

Деев К. С. Исследование сетевого взаимодействия с помощью системы глубокого анализа пакетов. – На правах рукописи.

Диссертация на соискание ученой степени кандидата технических наук по специальности 05.13.05 – компьютерные системы и компоненты. – Черкасский государственный технологический университет, Черкассы, 2018.

В диссертации предложены методы обеспечения идентификации сетевой активности одноранговых приложений для последующей их классификации и выделения в отдельный сервисный класс такого взаимодействия с целью проведения гибкой тарификации в сети оператора Интернет-услуг. Реализация созданных моделей и методов в программных средствах и их внедрение в тестовых сегментах распределенных сетей операторов услуг позволило в значительной степени повысить эффективность предоставления услуг пользователям и использовать оптимальные политики управления сетевым трафиком. В работе обозначены подходы к стандартизации и реализации в программных платформах функций анализа пакетной нагрузки согласно их представлению в виде многоуровневой модели взаимодействия открытых систем OSI. Определен функциональный уровень и даны рекомендации по оптимальному размещению комплекса анализатора в операторской сети общего назначения. Рассмотрены возможности частичной виртуализации отдельных компонентов системы с целью повышения общей пропускной способности.

Ключевые слова: глубокая инспекция сетевых пакетов, анализ трафика, классификация пакетов, система предотвращения вторжений, средства мониторинга сети.

ABSTRACT

Deev K. S. The study of networking interactions by using deep packet inspection system. – As a manuscript.

Ph.D. thesis on specialty 05.13.05 – computer systems and components. – Cherkasy State Technological University, Cherkasy, 2018.

The Thesis introduces models of representing Peer-to-Peer networking interactions and proper methods of conducting IP packet header and payload analysis. This area has been already heavily investigated by many scientists, however there are few questions still open. As Internet is growing and becoming more popular, the number of concurrent data flows starts to increase, which makes sense in amount of bandwidth requested and that should be analyzed respectively.

In this work, the methods for ensuring identification of network activity of Peer-to-Peer applications for their subsequent classification and proper allocation to separated service class are offered. Such interaction by being classified provides ability to implement flexible charge policy in service-provider network.

That should considerably increase user's experience and lower overall capacity load on backbone infrastructure links.

Service-providers and corporate customers need the ability to identify Peer-to-Peer interactions, because they are generally not directly related to workflow and lead to premature exhaustion of the available bandwidth of external links.

This Thesis represents the principles of building system, which searches for Peer-to-Peer interaction in live network traffic and then places such conversation into formerly

marked QoS class with bandwidth constraints. The implementation of the created models and used methods in software has significantly increased the efficiency of provided services. To ensure high quality service to all its subscribers it is desirable to create the system that carries identification of such flows based on classes of service with different priorities. It was tested in specific segments of service-provider's distributed networks and have shown that optimal policies for managing network traffic considerably simplifies management of complex setup.

With consistently increasing number of packets per second that should be investigated, the analysis using standard server's hardware-based solutions is challenging, as it is necessary to distribute the load over multiple systems. Therefore, the best way is to use special software-defined complex rather than hardware implementations. Software will distribute the load in the internals of the complex, using the principles and approaches, in particular, described in this paper. Throughput of the system configured in the same manner was analyzed though.

The paper outlines the approach with standardization and implementation packet payload analysis functions in software platforms according to their representation in the form of a multilevel OSI model. The functional level is determined and recommendations for the optimal placement of the analyzer complex in service provider network are given.

Outlined methods and approaches in the implementation of flexible network packet classifying system are based on deep packet inspection technique. Highlighted approach is analyzed, its benefits are determined to approximate the value of suggested improvements in terms of throughput. Regular expressions matching can balance classified packet payload and could be used for parallel execution on multiple specialized nodes. The possibilities of partial virtualization of individual components of the system with the purpose of increasing the overall throughput are also considered and recommendations are provided.

The Thesis presents a flexible approach to match network packet via search engine using relaxed regular expressions for whole network layer headers. By using such mechanism, software and hardware composition that might be used as a detector of anomalies in the network has been created finally.

Further improvements to the scope of network classification will be performed based on created method of applications interaction identification which rooted on supervised automated machine learning techniques coupled with specific composed training data publicly available for consideration.

The results of Thesis are helpful in terms of practical experience, which can be applied to development of scalable packet classifying system with limited budget on set of available hardware.

Keywords: deep packet inspection, traffic analysis, packet classifying methods, intrusion detection system, network monitoring tools.