МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Тернопільський національний економічний університет
Навчально-науковий інститут інноваційних освітніх технологій
Кафедра комп'ютерної інженерії

ЧЕЧЕТ Павло Павлович

Нейромережева система виявлення спамових повідомлень електронної пошти / Neural network system for spam e-mail messages detecting

спеціальність: 123 - Комп’ютерна інженерія
магістерська програма - Комп’ютерна інженерія

Магістерська робота

Виконав студент групи КІзм-21 П. П. Чечет
Науковий керівник: к.т.н., Л.О. Дубчак

Магістерську роботу допущено до захисту:

ТЕРНОПІЛЬ -2018

РЕФЕРАТ

Випускна кваліфікаційна робота на тему “Нейромережева система виявлення спамових повідомлень електронної пошти” на здобуття освітнього рівня “Магістр” зі спеціальності “Комп’ютерна інженерія” написана обсягом 107 сторінок і містить 26 ілюстрацій, 2 таблиці, 5 додатків та 91 джерело за переліком посилань.

Метою роботи є розробка методології проектування ефективної системи захисту інформації, що забезпечує фільтрацію спаму в організації.

Методи досліджень. В дипломній роботі використовувалися: методологія захисту інформації, методи системного аналізу, теорія множин, теорія ймовірності, теорія моделювання дискретних систем, теорія нейронних мереж, теорія багатоагентних систем. Для оцінки ефективності запропонованих рішень використовувалися методи математичного і імітаційного моделювання.

Запропонована нова концепція побудови автоматизованої багаторівневої багатоагентної системи виявлення спамових повідомлень, яка заснована на багаторівневій фільтрації спаму.

Запропоновано комбінований ієрархічний алгоритм формування бази знань, що дозволяє сформувати одночасно повну і достовірну базу знань системи фільтрації спаму.

Запропоновано ефективний алгоритм класифікації електронних повідомлень на основі когнітивного підходу і нейромережевого класифікатора, що дозволяє за допомогою використання бази знань ефективно вирішувати задачу класифікації вхідних електронних повідомлень на різних рівнях ієрархії організації.

Розроблена архітектура ієрархічної багатоагентної системи захисту інформації, що обробляється електронними поштовими системами від шкідливої дії спаму.

КЛЮЧОВІ СЛОВА: СПАМОВІ ПОВІДОМЛЕННЯ, ІНФОРМАЦІЙНА БЕЗПЕКА, ЕЛЕКТРОННА ПОШТОВА СИСТЕМА, ШТУЧНА НЕЙРОННА МЕРЕЖА, БАГАТОАГЕНТНА СИСТЕМА.

RESUME

Graduation work «Neural network system for spam e-mail messages detecting» on acquiring of educational qualification «Master» degree, from speciality «Computer engineering» with total volume 107 pages that contains 26 illustrations, 2 tables, 5 additions and 91 sources of information according to the list of references.

The object is to develop the methodology of designing an effective information security system that provides spam filtering in the organization.

Research methods. In the thesis work were used: the methodology of data protection, methods of system analysis, set theory, the theory of probability, the theory of discrete systems simulation, neural network theory, the theory of multi-agent systems. To evaluate the effectiveness of the proposed solutions were used mathematical and simulation modeling methods.

There are also proposed:

– a new concept of building an automated multi-agent detection system of spam messages that based on multi-level spam filtering;

– a combined hierarchical algorithm of knowledge base formation that allows us to form simultaneously a complete and accurate base of knowledge for spam filtering system;

– an efficient algorithm for the classification of messages based on the cognitive approach and neural network classifier, which allows by using the knowledge base effectively solve the problem of classification of incoming e-mail at different levels of the organization hierarchy.

The architecture of multi-agent hierarchical system of information protection is developed. That processed by electronic mail systems for protection from harmful spam effects.

KEY WORDS: SPAM MESSAGES, INFORMATION SECURITY, E-MAIL SYSTEM, ARTIFICIAL NEURAL NETWORK, MULTI-AGENT SYSTEM.

ЗМІСТ

ПЕРЕЛІК ПОЗНАЧЕНЬ І СКОРОЧЕНЬ

ФС – фільтрація спаму

ФЕП – фільтрація електронних повідомлень

НМК – нейромережевий категорізатор

БС – багатоагентна система

ЛОМ – локальна обчислювальна мережа

ВП – вхідні повідомлення

ВихП – вихідні повідомлення

ФВП – фільтр вхідних повідомлень

ФВихП – фільтр вихідних повідомлень

СФ – спам-фільтр

БЗ – база знань

ВСТУП

Актуальність роботи. Основною задачею досліджень в області захисту інформації є вдосконалення відомих і розробка нових методів, алгоритмів забезпечення безпеки інформації в процесі її збору, зберігання, обробки, передачі і розповсюдження. Одним з напрямів досліджень в цій області є розробка методів і алгоритмів фільтрації спаму (ФС). Надмірні не потрібні електронні листи порушують доступність інформаційних ресурсів, необхідних користувачам, оскільки споживають значні ресурси каналу вхідного зв'язку, а також можуть стати причиною порушення цілісності інформації у разі втрати повідомлення при ФЕП людиною, або програмою фільтрації. Разом із спамом можуть розсилатися шкідливі програми, здатні привести до повного або часткового знищення інформації або її спотворення. Ряд шкідливих програм може бути використаний для крадіжки персональних даних: номерів кредитних карт приватних користувачів, імен користувачів і паролів для доступу до систем віддаленого управління банківськими рахунками організацій. Крім того, конфіденційні дані можуть бути випадково або навмисно відправлені по електронній пошті. Залежно від прийнятої в організації політики безпеки, необхідно контролювати не тільки вхідний, але і вихідний трафік. Завдання пошуку відомостей, які складають державну або комерційну таємницю, у вихідному потоці електронної пошти аналогічна задача ФС. В цьому випадку для навчання системи можуть використовуватися не тільки не потрібні електронні листи, а конфіденційні документи, представлені в електронному вигляді. Не дивлячись на використання різних систем ФЕС, частка спаму в загальному поштовому трафіку все ще достатньо висока.

Питанням протидії спаму присвячено багато досліджень. В основному, це фільтри, побудовані на байєсівському підході, що, як відомо, не дозволяє враховувати семантику електронних повідомлень. При розробці систем фільтрації вхідних повідомлень недостатньо повно використовується системний підхід і сучасні технології штучного інтелекту для вирішення задачі класифікації. Тому,

задача розробки ефективних методів і алгоритмів ФС в організації є актуальною.

Мета і завдання дослідження. Метою дипломної роботи є розробка методології проектування ефективної системи захисту інформації, що забезпечує ФС в організації.

Для досягнення поставленої мети в роботі необхідно виконати наступні завдання:

– розробити концепцію побудови системи ФС в організації на основі методів штучного інтелекту;

– розробити багатоагентну архітектуру ієрархічної системи ФС в організації;

– розробити ефективний алгоритм класифікації електронних повідомлень з врахуванням семантики повідомлення;

– оцінити ефективність запропонованих підходів до ФС в організації.

Об'єкт дослідження – процес забезпечення ФС в організації.

Предмет дослідження – алгоритми ФС в організації на основі технологій штучного інтелекту.

Методи досліджень. В дипломній роботі використовувалися: методологія захисту інформації, методи системного аналізу, теорія множин, теорія ймовірності, теорія моделювання дискретних систем, теорія нейронних мереж, теорія багатоагентних систем. Для оцінки ефективності запропонованих рішень використовувалися методи математичного і імітаційного моделювання.

Наукова новизна одержаних результатів. Запропонована нова концепція побудови автоматизованої багаторівневої багатоагентної системи протидії шкідливій дії спам–розсилок на інформацію, яка заснована на багаторівневій ФС, що дозволяє підвищити доступність і забезпечити цілісність інформації, що обробляється в системах електронної пошти на різних рівнях ієрархії організації з врахуванням прийнятої політики безпеки. Запропоновано ефективний алгоритм класифікації електронних повідомлень на основі когнітивного підходу і нейромережевого класифікатора, що дозволяє за допомогою використання БЗ ефективно вирішувати задачу класифікації вхідних електронних повідомлень на різних рівнях ієрархії організації.

Практичне значення отриманих результатів. Практична цінність отриманих результатів полягає в підвищенні ефективності функціонування системи протидії розповсюдженню спаму в локальній обчислювальній мережі організації.

Розроблена архітектура ієрархічної багатоагентної системи захисту інформації, що обробляється електронними поштовими системами від шкідливої дії спаму, яка дозволяє будувати повну і достовірну БЗ, що відображає області інтересів користувачів системи в рамках ієрархії організації з врахуванням прийнятої політики безпеки.

Розроблений програмний прототип багатоагентної системи протидії розповсюдженню спаму в організації, що дозволяє оцінити ефективність запропонованого алгоритму.

Використання запропонованого підходу класифікації електронних повідомлень дозволяє врахувати в процесі аналізу семантичну компоненту повідомлення, тим самим понизити рівень помилкової класифікації на 5-10%.

У першому розділі виконаний аналіз різних видів спаму з погляду їх загроз захищеності інформації, а також розглянуті переваги і недоліки відомих підходів до протидії цим загрозам. Робиться висновок про необхідність розробки нової архітектури системи протидії розповсюдженню спаму, алгоритмів фільтрації, що дозволяють більш ефективно, в порівнянні з існуючими системами, забезпечувати ФС.

У другому розділі розглядається задача розробки концепції побудови автоматизованої ієрархічної системи протидії шкідливій дії спам-розсилок на інформацію, яка обробляється в системах електронної пошти, що полягає в багаторівневій ФС з використання БЗ, різних по повноті і достовірності. Виконаний аналіз основних потоків інформації в системі обробки повідомлень. Проаналізований вплив спаму на доступність і цілісність інформації. Виконаний порівняльний аналіз переваг і недоліків спам-фільтрів, що виконують централізовану і розподілену фільтрацію. Розглянуто три основні можливі способи формування БЗ корисних повідомлень і спаму для фільтрів, що здійснюють централізовану фільтрацію.

Пропонується процедура формування БЗ інтелектуальної системи боротьби із спамом, яка об'єднує в собі всі переваги серверних і персональних фільтрів. Розглядається вирішення задачі розробки алгоритму класифікації вхідного і вихідного потоку електронної кореспонденції на основі застосування парадигми нейронних мереж. З урахуванням специфіки задачі ФЕС, пропонується в якості спрощеної моделі текстового фрагмента використовувати представлення мінімальної семантичної одиниці – речення у вигляді семантичного графа. В якості класифікатора пропонується використовувати лінійний нейромережевий асоціатор, елементи вхідного вектора якого складені з елементів семантичної матриці фільтрованого повідомлення. Навчання асоціатора виконується на основі правила навчання Хебба. Архітектура розробленої системи багаторівневою ФС дозволяє реалізувати здатність системи до самонавчання.

У третьому розділі описується реалізація багаторівневої системи боротьби із спамом на основі багатоагентного підходу. Пропонована система боротьби із спамом є розподіленою, кожен системний компонент, керівник поштовими фільтрами, має лише інформацію, необхідну для вирішення задачі і може впливати на рішення задачі тільки на своїй ділянці. Це обумовлено специфікою складної, гетерогенної, розподіленої в просторі і непостійної по структурі системи, якою є інформаційна система сучасного підприємства. Ключовим моментом, що визначає вибір багатоагентної технології є необхідність забезпечення автономності системи ФС. Більш того, зміна структури інформаційної системи підприємства, додавання або видалення будь-якої призначеної для користувача робочої станції або навіть сервера окремого відділу, при використанні багатоагентної технології дозволяє враховувати ці зміни. Також це дозволяє підвищити ефективність роботи адміністратора мережі. Розроблена багатоагентная система ФС складається з трьох рівнів: верхнього, проміжного і нижнього. На верхньому рівні розташовується основний в організації поштовий сервер, на проміжному рівні знаходяться поштові сервера відділів, нижній рівень займають поштові клієнти, встановлені на робочих станціях користувачів.

1 АНАЛІЗ МЕТОДІВ ЗАБЕЗПЕЧЕННЯ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ ШЛЯХОМ АВТОМАТИЧНОГО ВИЯВЛЕННЯ СПАМОВИХ ПОВІДОМЛЕНЬ

1.1 Аналіз поняття спаму

Як відомо, захист інформації займається питаннями розробки, вдосконалення і застосування методів і різного роду засобів захисту інформації в процесі збору, зберігання, обробки, передачі і розповсюдження інформації [1]. Методи класифікації інформації в процесі отримання електронної пошти, її обробки на поштовому сервері, а також відправки вихідних поштових повідомлень, є однією з задач, що вимагає вирішення. Не потрібні електронні листи, що посилаються спамерами порушують доступність ресурсів, необхідних користувачам, оскільки можуть значно споживати ресурси каналу вхідного зв'язку. Фактично можна сказати, що має місце розподілена атака "Відмова в обслуговуванні" (DDoS – Distributed Denial of Service). В процесі фільтрації електронних повідомлень, як людиною, так і програмою, може бути допущена помилка другого роду – тобто потрібний лист може бути заблокованим, що порушує цілісність інформації. Також разом із спамом часто розсилаються віруси, які можуть знищити інформацію, що також порушує цілісність інформації. Ряд вірусів розроблений для крадіжки персональних даних: номерів кредитних карт приватних користувачів, імен користувачів і паролів для доступу до систем віддаленого управління банківськими рахунками організацій [2]. Крім того, конфіденційні дані можуть бути випадково або навмисно відправлені по електронній пошті. Задача пошуку відомостей, що складають державну або комерційну таємницю, у вихідному потоці електронної пошти аналогічна задачі фільтарації спаму, просто для навчання системи використовуються не спамові повідомлення листи, а конфіденційні документи.

Не потрібні електронні повідомлення були проблемою ще в 1975 році. Однією з перших спам-розсилок була спроба корпорації DEC розіслати рекламу на всі адреси мережі Arpanet. Загальноприйнятим це слово стало в 1994 році, коли 2 американських юристи розіслали велику кількість повідомлень, що закликали

скористатися їх послугами. Інші приклади ранніх розсилок спаму приведені в роботі Темплтона [3].

Запропоновані різні визначення спаму, частина з яких не зовсім розкриває суть спаму [4]. Чітке визначення спаму необхідне для збору статистики про спам і розуміння суті цього явища, а також може знадобитися для введення різних контрелементів, що стосуються комерційної поведінки, психології одержувача спаму, розширення законодавчих контекстів, економічної і технічної сторони справи. У більшості трактувань можна виділити загальні риси, які підходять під наступною визначення: "спам - розсилка по електронній пошті повідомлень, іноді систематична, не потрібної інформації у великих об'ємах особам, що не мають ніяких контактів з людиною, яка розсилає дані повідомлення і чиї електронні адреси були отримані незаконним шляхом." [4]; "зазвичай під терміном спам розуміється систематична розсилка електронних повідомлень комерційного характеру особами, що приховують або підроблюють свої справжні координати" [5]; "спам визначається як будь-які не потрібні електронні повідомлення, що приходять на електронну пошту у великих кількостях. Це визначення швидше характеризує масові електронні розсилки" [6].

Згідно даним Організації економічного співробітництва і розвитку (ОЕСР), у всіх визначеннях спаму можна виділити головні і додаткові його ознаки. До головних відноситься не затребуваність повідомлень, присланих у великій кількості, і їх комерційний характер. Будь-яка людина, що отримала величезну кількість листів, відразу зрозуміє, що це спам. У спаму також є додаткові ознаки, але не завжди лист, що володіє ними, є спамом: адреси користувачів використовуються без їх дозволу; розсипається непотрібна інформація; систематична розсилка; відправника не хвилює склад аудиторії, головне – кількість; Повідомлення приходять щодня; відправник скриває/підробляє свої справжні координати; рекламуються нелегальні товари; інформація в повідомленнях є помилковою.

Не дивлячись на складність виділення одного визначення спаму, точно і що чітко визначає це явище, за допомогою існуючих визначень спаму можна виділити його характерні риси: спам – це електронне повідомлення, отримане по

мережі Інтернет; спам приходить без попереднього запиту або згоди. Тобто якщо одержувач погоджується на розсилку, то це вже не спам; спам розсилається у великих кількостях. Тобто відправник розсилає багато однакових повідомлень, а одержувачі вибираються випадково.

Ці три межі характерні для масових не затребуваних електронних повідомлень; також це співпадає з визначенням, даним організацією Spam-House [7]. Дана робота оперує саме вищеназваним визначенням спаму, хоча тут можна виділити і четвертий пункт: спам повинен містити рекламу якого-небудь товару або послуги. Більшість повідомлень відносяться саме до цієї категорії і називаються рекламними повідомленнями.

1.2 Аналіз загроз інформаційній безпеці спамовими повідомленнями

Спам - повідомлення діляться залежно від цілей, що переслідуються спамером. Одні спамери роблять масові розсилки з метою отримання прибутку, наприклад, в повідомленні може міститися реклама товару або послуги або заклик брати участь в політичних кампаніях. Інші розсилають повідомлення шахрайського характеру або поширюють шкідливі програми (віруси і троянські коні).

Спам, розісланий з метою реклами товару або послуги, позначають як не затребувані комерційні електронні повідомлення. У більшості випадків такі повідомлення розглядаються компаніями як важливий метод залучення клієнтів, оскільки розсилка повідомлень поштою – найдешевший спосіб розповісти про товар або послугу. Проте більшість таких повідомлень розсилаються не самими компаніями, а спамерами, які отримують певну винагороду за розсилку спаму.

З погляду інформаційної безпеки, не затребувані комерційні електронні повідомлення впливають тільки на доступність інформації. Пропускна спроможність каналу зв'язку може повністю використовуватися для скачування цих повідомлень, і ресурс мережі Інтернет, до якого користувач намагається

дістати доступ, опиниться недоступний.

Не дивлячись на мізерну кількість відповідей на спам-повідомлення, замовники все одно отримують прибуток. Згідно OECP [4] 8% з опитаних признались, що придбали товари, що рекламуються через спам. Дослідження, показали, що навіть якщо відсоток тих, що відповіли складатиме 0.001%, реклама через електронні повідомлення у будь-якому випадку вигідна.

Рекламні повідомлення можуть містити політичну або релігійну пропаганду. Наприклад, в 2003 році члени Конгресу США розіслали виборцям сотні тисяч повідомлень [8]. По негативній дії на захищеність інформації, повідомлення некомерційного характеру аналогічні першим, тобто впливають тільки на доступність інформації.

Деякі спамери розсилають повідомлення з помилковою інформацією, тобто шахрайські листи. Листи помилкового змісту, послані з метою отримати яку–небудь інформацію, називають "скам". Прикладами можуть служити листи з проханнями перевести гроші на рахунок жертв природного лиха. Також сюди відносяться нігерійські листи, послані нібито від урядовців, що підтверджують, що вони вкрали мільйони доларів з фонду допомоги [9]. Одним з видів шахрайських листів є фішингові листи або бред-спуффінг, послані нібито від імені відомої компанії. Метою подібних листів є отримання у користувачів конфіденційних даних про паролі і коди доступу, наприклад, листа з банку з проханням підтвердити дані кредитної карти. І шахрайські, і фішингові листи окрім порушення доступності зовнішніх ресурсів, порушують ще і конфіденційність секретної інформації, такої як номери банківських карт або паролі доступу до систем дистанційного управління рахунком.

Помилкові повідомлення розсилаються з метою змусити одержувача повірити що яка-небудь помилкова подія є правдою, причому такі повідомлення часто супроводжуються проханням розіслати цей лист найбільшій кількості людей (ланцюгові листи). Деякі повідомлення попереджають про віруси, черв'яків або троянських коней, інші містять невірну інформацію про які-небудь політичні або суспільні події, іноді в повідомленнях міститися прохання про добродійність або пропозиції комерційного характеру, наприклад, в повідомленні може

знаходитися сертифікат на отримання безкоштовного подарунка від фірми. Таким чином, помилкові і ланцюгові повідомлення знижують тільки доступність інформації.

"Joe jobs" – це обманне повідомлення, послане від імені іншої людини з метою нанесення шкоди його репутації. Наприклад, "Joe jobber" може розіслати на тисячі адрес повідомлення з дитячою порнографією, при цьому зворотною адресою значитиметься, наприклад User@Company.Ru, щоб обурені одержувачі даного повідомлення закидали ящик User'а гнівними листами і репутація компанії Company була б підірвана. Назва joe jobs вперше була використана для опису подібної схеми, жертвою якої став хостер Джо Дол. Обліковий запис одного користувача був видалений з розсилки спаму для реклами своїх товарів; в помсту той користувач розіслав на мільйони адрес ще більше спаму, але зворотною адресою значилася адреса Джо Дола [10]. В даному випадку має місце розподілена атака "Відмова в обслуговуванні", де ролі тих, що атакують виконують одержувачі обманних повідомлень.

Шкідливі програми розробляються з метою нанесення шкоди комп'ютерній системі і розсилаються під виглядом нешкідливого додатку до повідомлення. Віруси, "черв'яки", "троянські коні", програми-шпигуни і рекламні програми вкладаються в листи і запускаються при відкритті вкладеного файлу. Між спамом і шкідливими програмами існує взаємозалежність [11]: через спам розсилаються шкідливі програми, вони завдають шкоди комп’ютеру, щоб контролювати його на відстані і розсилати ще більше спаму. Такі комп'ютери називаються "зомбі". Таким чином, електронні листи з шкідливими програмами порушують не тільки доступність зовнішніх ресурсів (за рахунок забивання вхідного каналу трафіком, що генерується при скачуванні спаму) і конфіденційність секретних даних (за рахунок можливості деяких вірусів знаходити і відправляти господареві номера банківських карт і паролів доступу), але і цілісність всієї інформації, збереженої на комп'ютері (за рахунок можливості деяких вірусів зашифровувати всі знайдені документи).

Негативна квитанція – це не доставлене повідомлення, що посилається назад відправникові. Згідно дослідженню Ironport [12], повідомлення, повернені

на підроблені зворотні адреси невинних третіх осіб, складають близько 9% від всього електронного трафіку, що відповідає 1,67 млрд. повідомлень, що повернулися, в день [13]. Негативна квитанція не є спамом як таким, але вона складає значну частину електронного трафіку, який збільшується із–за спаму. Таким чином, негативні квитанції знижують тільки доступність зовнішніх ресурсів.

Згідно Ferris Research [14] в 2010 році спам завдав американським компаніям збиток у розмірі $ 8,9 млрд, а європейським – $ 2,5 млрд. При цьому виявилося, що «закодування» непотрібною інформацією поштових скриньок не найстрашніше. Наприклад, 40% збитку викликано падінням продуктивності праці – співробітники змушені відволікатися на спам. Виявилося, що в середньому співробітники витрачають 4,5 секунди лише на видалення одного такого листа.

1.3 Аналіз методів боротьби зі спамовими повідомленнями

Методи можуть застосовуватися на різних стадіях відправки. Вони можуть бути задіяні на стадії роботи поштового клієнта, MTA поштового провайдера відправника, поштового вузла поза поштовим провайдером відправника і одержувача або клієнта одержувача [15]. На перших двох стадіях можна застосувати превентивні заходи. Бажано зупинити спам якомога раніше, щоб не витратилися такі ресурси як ширина смуги пропускання і час скачування. Проте, найпоширеніші технічні методи боротьби із спамом застосовуються на стороні одержувача.

Маршрути розсилки спаму можуть розрізнятися. Іноді спамери встановлюють власного агента пересилки (MTA) і розсилають повідомлення безпосередньо провайдерові поштових послуг (ESP – E–mail service provider) одержувача. Іншим способом може бути дослідження інфраструктури ESP, посилаючи повідомлення через MTA.

Методи боротьби із спамом можна розділити на коротко-, середньо- і довгострокові. Механізми фільтрації і блокування розглядаються як короткострокові методи, оскільки їх застосування може бути обмежене інфраструктурою місцевої поштової організації з незначними модифікаціями. Деякі методи, засновані на DNS, які впливають на структуру і вміст в записах DNS можуть запускатися місяцями або навіть роками. Методи, засновані на Інфраструктурі відкритих ключів (PKI – Public key infrastructure) і методи, засновані на ресурсах, можуть виявитися довгостроковими зважаючи на значні зміни і розширення інфраструктури. Вся різноманітність методів боротьби із спамом показана на рисунку 1.1.

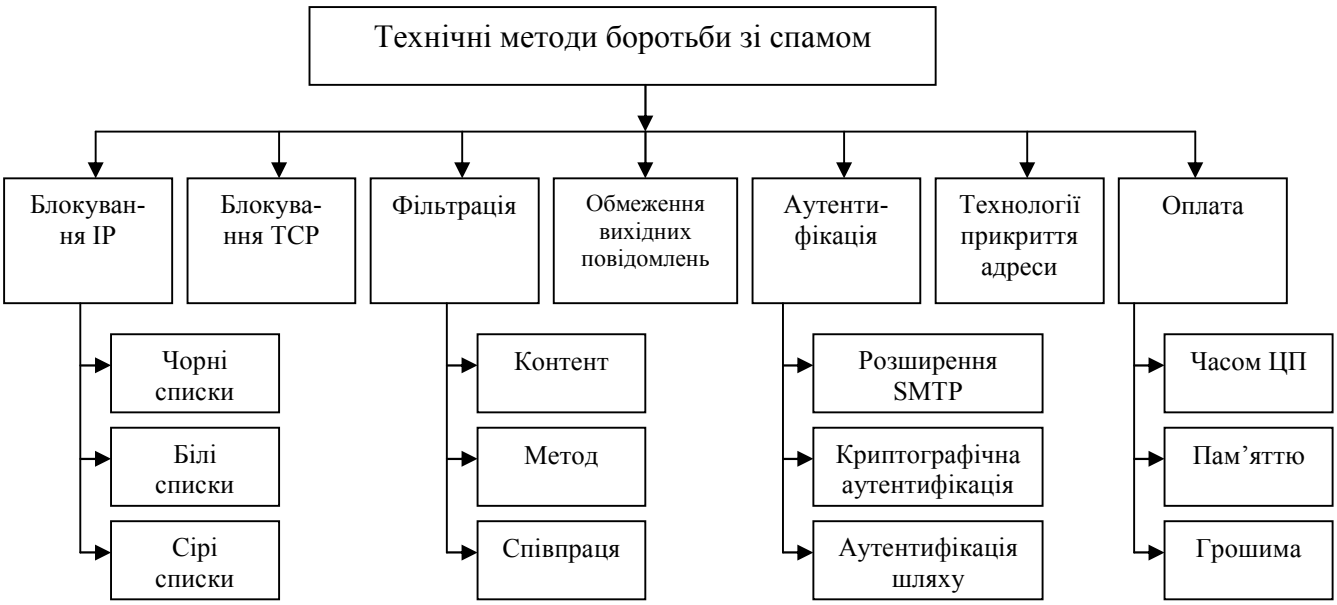


Рисунок 1.1 – Технічні методи боротьби із спамом

1.3.1 Метод блокування IP–адреси

Коли клієнт ініціює з'єднання SMTP, на мережевому або транспортному рівні встановлюється TCP/IP зв'язок з SMTP сервером. IP-адрес хоста відправника може бути легко визначений і він же буде першою інформацією про клієнта, яку отримає сервер. Якщо IP-адрес схожий на адрес клієнта, який розсилав спам у минулому, в з'єднанні може бути відмовлено ("чорний список"). Іноді, в чорний список заноситься цілий діапазон IP адрес, наприклад адреси певного домена або

ISP. Якщо IP–адрес належить надійному клієнтові, з'єднання буде встановлено ("білий список"). Термін "сірий список" описує такий спосіб, коли IP–адрес – це частина інформації, яка використовується для ухвалення або відхилення з'єднання. Кожна поштова транзакція спочатку відхиляється, а дані (адреса відправника, адреса одержувача і тема повідомлення) про цю первинну невдалу спробу зберігаються. Якщо в спеціальному тимчасовому вікні SMTP клієнт намагається знову виконати невдалу поштову транзакцію, сервер приймає цю транзакцію як таку, що підходить під збережені параметри. Занесення в сірі списки грунтується на припущенні, що більшість джерел спаму в цілях економії часу не пересилають повідомлення наново, вважаючи, що поштовий сервер став недоступним.

Блокування IP адреси легко здійснити, воно не вимагає величезної кількості ресурсів, оскільки рішення прийняти/відхилити ухвалюються на початковій стадії SMTP діалогу. Недоліками блокування IP є:

1. Вона даремна, якщо IP-адреса клієнта підроблена. Підробка IP-адреси – це проблема, властива протоколам TCP/IP; можливі атаки описані в роботах Tanase [16]. Проте IP-спуфінг не є серйозною проблемою, тому що: SMTP з'єднання засновані на TCP з'єднаннях з початковим трибічним "рукостисканням" (3–way handshake), тому реалізація IP-спуфінга проблематична; мережу можна захистити від IP-спуфінга за допомогою нескладних превентивних заходів; блокування IP – не єдиний засіб боротьби із спамом, що вживається сучасними MTA. На практиці рідко зустрічаються випадки, коли IP-спуфінг організовується з метою розповсюдження спаму.

2. Блокування IP працює логічно і, в принципі, воно може допустити помилки як першого, так і другого роду.

Чорні списки (BL – Black list) можуть відрізнятися по багатьом пунктам. Деякі організації, наприклад ISP підтримують приватні BL і не надають їх Інтернет співтовариству. Деякі організації надають доступ до приватних безкоштовно або за гроші в режимі реального часу, оскільки спамери міняють хости–відправники. База може містити IP-адреси відомих спамерів, відкритих трансляторів пошти, або незаконних посередницьких серверів (XBL –Exploit block

List). Прикладами бази даних Spamhaus block list (SBL), Arbitrary black hole list (ABL), Domain name system real time black list (DNSRBL), Open Relay Database (ORDB), MAPS Dial-up user list (MAPS-DUL) і Spam Prevention Early Warning System (SPEWS). Spamhaus надає списки XBL, що може допомогти відбитися від спамерских атак через відкриті проксі, а також блокувати черв'яків або вірусів з вбудованими механізмами розсилки спаму і інші типи троянських коней, що використовуються спамерами. Повний перелік чорних списків DNS надають DECLUDE Internet Security Software [17], Email–policy.com і InfoSec.

DNSBL можуть відрізнятися політикою, яка містить інформацію про відслідковування (MAPS – це компанія, яку легко знайти, SPEWS – абсолютно анонімна організація), про те, яким чином IP-адрес потрапляє в список (деякі приймають звіт і розслідують справу, деякі приймають звіт і відразу ж додають адресу в список) і про те, як адрес може бути видалений із списку (деякі публікують таку інформацію, критерії інших невідомі). Популярний метод занесення (тимчасово) IP-адреса в BL – це частотний аналіз ВП кожного хоста. Спамери іноді посилають за короткий проміжок часу величезну кількість повідомлень на особливий поштовий сервер, що приводить до виключно великої частоти. Хост, запідозрений в такій поведінці може бути заблокований, але при цьому додатково повинні використовуватися білі списки, щоб могла проходити "звичайна" масова пошта на зразок інформаційних бюлетенів.

Недоліками BL є: BL не можуть містити абсолютно всі адреси. Спамери мають звичай використовувати IP який–небудь короткий проміжок часу, наприклад 2 години, BL можуть не оновлюватися, а це приведе до помилкових відмов; часто BL можуть містити адреси або навіть діапазони адрес, які належать ESP або ISP, оскільки деякі спамери можуть скористатися інфраструктурою провайдера в своїх шахрайських цілях. Перш, ніж адміністрація компанії - провайдера дізнається про цей факт і вирішить проблему, електронні адреси тисяч або мільйонів користувачів можуть бути заблоковані; чорні списки DNS ведуть до збільшення трафіку і роблять DNS критичнішим ресурсом, оскільки він стає вразливішим в плані цілісності і достовірності (DNS спуфінг); варіантів BL з'явилося багато. "Right Hand Side Blacklist" зберігає не IP адреси, а доменні

адреси, які можна легко підмінити.

Чорні списки уніфікованих ідентифікаторів ресурсів відрізняються від "звичайних" списків тим, що вони використовуються для виявлення спамового вмісту в тексті повідомлення. Вони роблять можливим блокувати повідомлення, що мають в основній частині посилання на сайт, що використовує спам для залучення відвідувачів. Такий список надає surbl.org. На жаль, процедура витягування адреси сайту і перевірка його на включення в чорний список достатньо ресурсоємна. Більш того, в цьому випадку необхідно отримати і обробити всі повідомлення, а це вже фільтрація.

Також як і чорні списки, білі списки (WL — White list) можуть підтримуватися на місцевому рівні, або надаватися для загального користування. Якщо вони опубліковані через DNS, вони, аналогічно чорним спискам, називаються білі списки системи доменних імен (DNSWL). DECLUDE Internet Security Software [17] надає DNSWL для загального користування. На відміну від BL, білі списки не потрібно оновлювати в реальному часі. WL без додаткових заходів не є ефективним способом боротьби із спамом, оскільки рівень помилок другого роду буде дуже високий. WL слід застосовувати разом з іншими методами, але при цьому вони повинні вважатися методом першого рівня, оскільки повідомлення з хостів, занесених в WL не вимагають перевірки іншими методами боротьби із спамом.

Сірі списки успішно застосовуються для захисту поштових серверів від спаму. Проте і тут є деякі недоліки, що зменшують ефективність списків.

По–перше, для кожної поштової транзакції зберігається IP адреса і деяка інформація з конверта повідомлення. Таким чином, SMTP транзакція не може бути прийнята або відхилена до надходження всіх даних. На відміну від WL і BL, цей метод вимагає трохи більше часу для ухвалення рішення. Крім того, потрібний простір для зберігання.

По-друге, сірі списки приводять до збільшення поштового трафіку, оскільки повідомлення необхідно пересилати. По-третє, списки не вирішують проблеми зловживання звичайними користувачами. По-четверти, великі поштові системи іноді складаються з декількох серверів-відправників, які можуть по черзі

пересилати відхилені повідомлення. Оскільки сервера мають різні IP-адреса, повідомлення може загубитися.

1.3.2 Метод фільтрації електронних повідомлень

Методи фільтрації є евристичними (як і методи блокування IP), вони класифікують повідомлення на 2 категорії: спам і нормальна пошта (по англ. ham). Фільтр може шукати ключові слова або особливу структуру (наприклад, застосування HTML MIME) або перевіряти використовувану мову (англійська мова розглядається як підозріла). Фільтри можуть застосовуватися клієнтом ESP, самим ESP або одержувачем повідомлення.

Методи фільтрації широко застосовуються і можуть варіюватися залежно від об'єкту і методу перевірки. Деякі фільтри перевіряють тільки заголовок або основну частину повідомлення, інші перевіряють і те, і інше. Є фільтри, які перевіряють також параметри конверта. Існує також велика кількість особливих методів фільтрації. Це стало можливим, тому що можуть застосовуватися всі алгоритми для класифікації тексту. ФС – це задача, що включає 2 стадії: "навчання" і "класифікація". Таким чином, область навчання машини можуть забезпечити потрібні алгоритми. Даний підрозділ включає деякі методи, що самі діють, не зважаючи на завершеність: фільтрація може грунтуватися на правилах, на підписі або на статистиці (в основному, байєсовська класифікація). Фільтрація також може застосовувати моделі Support Vector Machines, Boosting Trees, Artificial Neural Networks і Markov Random Field.

Деякі фільтри можуть бути розподіленими, тобто не централізованими, а такими, що включають багато серверів, які надають інформацію про спам–повідомлення. Нижче розглядаються сумісні системи, засновані на підписах.

Щоб фільтри були ефективними, повинні виконуватися наступні умови: фільтри потрібно постійно перенастроювати оскільки спамери часто міняють структуру і зміст своїх повідомлень; фільтри повинні настроюватися індивідуально, оскільки організації або користувачі можуть використовувати різну термінологію; фільтр повинен бути надійним і стійким до помилок.

Наприклад, спамери зараз активно працюють над способами поводження фільтрів і один з варіантів таких способів може полягати в дробленні і неправильному написанні слів, щоб фільтри не змогли їх розпізнати. Більш того, вони повинні бути здатними вирішити проблему за умови, що спам і ham стають все більш і більш схожими.

Фільтрація як спосіб захисту від спаму володіє декількома істотними недоліками, не залежними від методу фільтрації. По-перше, відсутня повна гарантія визначення спаму. В якості прикладу помилкового дозволу можна привести слова Graham [18]: "Вид спаму, який мені було особливо важко фільтрувати – це такі повідомлення, як наприклад, з Болгарії з рекламою послуг контрактного програмування. Вони проходять крізь фільтр оскільки я сам є програмістом, а спамові повідомлення, що приходять на мій ящик містять ті ж слова, що і нормальна пошта". Помилка другого роду небезпечніша помилок першого роду: у 2003 році фільтри заблокували закон про згвалтування, посланий членам Палати общин Британського Парламенту; обговорення закону було зірване. Androutsopoulos і ін. [19] представили теоретичну модель розсилки спаму, яку вони пропонують використовувати для визначення оптимального співвідношення між помилковим доступом і відмовою при фільтрації. По-друге, методи фільтрації, особливо засновані на аналізі основної частини повідомлення, ресурсоємні. ESP повинні рівномірно розподіляти процесорний час для перевірки мільйонів повідомлень в день. Призначені для користувача фільтри вимагають попереднього скачування всього повідомлення, у тому числі і незапрошувані, що може виявитися достатньо проблематичним при використанні модему. По-третє, чим більше спам стає схожим на нормальні повідомлення, тим менш ефективними стають існуючі фільтри. По-четверте, механізми фільтрації зменшують ймовірність попадання спаму до одержувача. Але небезпека полягає в тому, що це дозволяє спамерам посилати ще більше повідомлень в надії обійти фільтри. Отже, фільтри посилюють проблему ресурсоємкості, яка виникає із-за спаму [20].

Варто відмітити, що деякі системи фільтрації використовують різноманітні методи фільтрації і підсумовують часткові результати до загального результату. Наприклад, організація "SpamAssassin" застосовує аналіз тексту, фільтрацію

Байєса і об'єднані бази даних фільтрації і призначає бали для кожного повідомлення для визначення ймовірності спаму. Кожен поштовий сервер може встановити і використовувати власні межі для того, щоб відрізнити спам і нормальну пошту.

Якщо ФС здійснюється за допомогою правил (rule-based filtering), їх можна створити вручну або автоматично. Просте правило може мати наступний вигляд і відноситися як до теми повідомлення, так і до його змісту. Наприклад, якщо тема повідомлення містить "ВІАГРА" і тіло повідомлення містить "Шановні пані та панове", то це спам. Докладніше обговорення настройки правил фільтрації представлене в роботах Cohen [21] і Crawford [22]. Головний недолік правил полягає в тому, що спамер може їх обійти, злегка змінивши написання слів в повідомленні, наприклад, слово "VIAGRA" можна написати "V1AGRA".

При ФС на базі підписів (signature-based filtering) повідомлення зменшується до розміру підпису, наприклад, за допомогою хеш-функції. Проте для ефективності даного методу важлива його стійкість до незначних змін повідомлень, наприклад, вказівка імені у вітанні, а також важлива частота оновлення бази підписів, оскільки спамові повідомлення міняються день від дня. Загальна процедура для відсіву спамового повідомлення – це створення підпису повідомлення і його порівняння з вже наявними в базі підписів.

VipuFs Razor – це поширена, об'єднана мережа по виявленню і ФС. "За допомогою користувачів створюється постійно оновлюваний каталог спаму, до якого підключаються поштові клієнти. Визначення спаму відбувається статистичними і випадковими підписами, що ефективно розпізнають змінний зміст спам-повідомлень. Якщо ж спам-повідомлення пройшло через фільтр, то користувач, знайшовши його, може створити новий підпис і відправити його до каталога. Якщо від інших користувачів поступлять аналогічні підписи, то довіра системи до вибору даного користувача зростає, інакше система поступово перестає довіряти оцінкам даного користувача. Damiani і ін. [23] пропонують серверний спам-фільтр, що працює в одноранговій мережі. Для кожного повідомлення, на яке поступила скарга, розраховується 256-бітовий дайджест, стійкий до типових змін повідомлень. Вважається, що два повідомлення є

однаковими, якщо їх дайджест розрізняється максимум 74 бітами. Автори використовують трирівневу структуру, що складається з рівня користувача, середнього рівня з поштовим сервером і верхнього рівня, де збираються скарги. Сервери обмінюються інформацією (дайджестами спам-повідомлень) за посередництва серверів верхнього рівня, які також обмінюються інформацією. Zhou і ін. [24] також використовують однорангову мережу. Замість дайджестів створюється набір контрольних сум файлу кожного повідомлення і розповсюджується через розширену систему Decentralized object location and routing system (DOLR). Організація Distributed Checksum Clearinghouse (DGG) використовує багато відкритих серверів, що підтримують БД контрольних сум повідомлень.

При фільтрації, заснованій на алгоритмі Байєса, кожне повідомлення представляється вектором x = $\{x_1, x_2, x_3, ..., x_n\}$, де x_i – значення атрибутів X_1, ..., X_n. Кожен з цих атрибутів є певним словом. Якщо атрибут рівний одиниці, слово присутнє, якщо нулю – відсутнє. Для вибору зі всіх атрибутів листа лише деяких, які характеризують його найкраще, проводиться розрахунок характеристичної інформації *MI* (1.1):

$$MI(X;C) = \sum_{x\in\{0;1\},\ C\in\{\text{спам, ллегітимн}\}} P(X=x,\ C=c)\cdot\log\frac{P(X=x,C=c)}{P(X=x)\cdot P(C=c)}. \quad (1.1)$$

Потім вибираються атрибути з найбільшими значеннями MI. Ймовірності оцінюються, як частота, з якою зустрічалися атрибути в інформації, на основі якої відбувалося навчання системи.

По теоремі Байєса і теоремі складання ймовірності, ймовірність приналежності документа *d* категорії *c* при заданому векторі атрибутів $x = \{ x_1, x_2, x_3, ..., x_n\}$ визначається по формулі (1.2):

$$P(C=c\left|\vec{X}=\vec{x}\right.) = \frac{P(C=c)\cdot P(\vec{X}=\vec{x}\left|C=c\right.)}{\sum_{k\in\{\text{спам, ллегітимн}\}} P(C=k)\cdot P(\vec{X}=\vec{x}\left|C=k\right.)}. \quad (1.2)$$

Практично ймовірності $P(\vec{X}|C)$ неможливо обчислити безпосередньо у зв'язку з великою кількістю значень $\vec{X}$. Простій Байєсовський класифікатор робить спрощене допущення, що $X_1,...,X_n$ незалежні від категорії C. Тоді:

$$P(C=c|\vec{X}=\vec{x}) = \frac{P(C=c)\cdot\prod_{i=1}^{n} P(\vec{X}=\vec{x}|C=c)}{\sum_{k\in\{\text{спам, ллегітимн}\}} P(C=k)\cdot\prod_{i=1}^{n} P(\vec{X}=\vec{x}|C=k)}, \quad (1.3)$$

де $P(X_i|C)$ і $P(C)$ можуть бути оцінені як ймовірність зустрічі атрибуту в листах певної категорії і ймовірність зустрічі листів, що належать до тієї або іншої категорії в масиві навчальних даних відповідно.

Проста Байєсовська класифікація дуже.ефективна, не дивлячись на те, що допущення, що лежать в її основі, є такими, що надмірно спрощують.

Помилкова класифікація легітимного повідомлення як нелегітимного зазвичай завдає більшої шкоди, чим пропуск нелегітимного повідомлення (класифікованого як легітимне). Нехай L→S і S→L позначають ці дві помилки відповідно (також вони ще називаються помилками першого і другого роду).

Припускаючи, що L→S в λ раз гірше S→L , повідомлення класифікується, як спам, якщо:

$$\frac{P(C=\text{спам}|\vec{X}=\vec{x})}{P(C=\text{легітимне}|\vec{X}=\vec{x})} > \lambda. \quad (1.4)$$

Оскільки класифікатор, що використовує цей критерій, досягає високих результатів, приведені вище допущення цілком розумні, обчислена ймовірність достатньо точна. В даному випадку:

$$P(C = \text{спам} \left| \vec{X} = \vec{x} \right.) = 1 - P(C = \text{легітимне} \left| \vec{X} = \vec{x} \right.), \quad (1.5)$$

що дозволяє сформулювати приведений вище критерій інакше:

$$P(C = \text{спам} \left| \vec{X} = \vec{x} \right.) > t, \quad (1.6)$$

де t – поріг.

$$t = \frac{\lambda}{1+\lambda}. \quad (1.7)$$

Встановлення порогу t = 0,999 означає, що блокування одного легітимного повідомлення так само погано, як пропуск 999 спам–повідомлень. Такий високий поріг необхідний, коли заблоковані повідомлення знищуються без подальшої обробки, оскільки для більшості користувачів втрата легітимного повідомлення представляється неприпустимою. Можлива установка нижчого порогу, якщо замість видалення повідомлення, наприклад, просити підтвердження відправки.

У ФС застосовуються і інші методи класифікації тексту: Support Vector Machines [25], Metzger і ін. В [26] запропонували сумісне фільтрування, Boosting Trees [27], Artificial Neutral Networks [28] і Markov Random Field Models [29]. На даний момент ефективність даних методів не вивчена.

1.3.3 Метод блокування TCP

На відміну від блокування IP, блокування TCP направлене на запобігання розсилці спаму з боку відправника. Оскільки SMTP повідомлення направлені на 25–й порт, ISP часто блокують весь вихідний з цього порту TCP трафік, що блокує розсилку спаму від клієнтів SMTP безпосередньо на хост MX. Блокування TCP ефективне, в разі встановлення спамером механізмів розсилки на своїх власних або заражених ПК. Проте блокування 25-го порту може створити

проблеми для тих клієнтів, кому з будинку необхідно відправляти повідомлення через корпоративний поштовий сервер [30]. Щоб дозволити клієнтам зв'язуватися з SMTP сервером, в якості механізмів аутентифікації пропонуються передача повідомлень через порт TCP 587 і SMTP–AUTH.

1.3.4 Аналіз механізмів аутентифікації

Механізми аутентифікації діляться на 3 категорії. До першої відносяться розширення SMTP, друга категорія заснована на криптографічній аутентифікації і відноситься до захисту шляху передачі повідомлень. Третя категорія направлена розпізнавання шляху передачі повідомлень, який визначає домен останнього мережевого сегменту або останнього MTA. Ця категорія включає протоколи, які називаються LMAP – Lightweight MTA Authentication Protocols.

Розширення протоколу, на зразок SMTP–AUTH [31], "SMTP після POP" і "SMTP після IMAP" застосовуються для аутентифікації користувачів або клієнтів SMTP. SMTP–AUTH визначає сервісне розширення SMTP, за допомогою чого клієнт SMTP може вказати серверу механізм аутентифікації, виконати обмін параметрами протоколів аутентифікації і вибірково погоджувати деякі параметри механізму аутентифікації. Це розширення – профіль Simple Authentication and Security level (SASL) [32] дозволяє користувачам виконувати аутентифіковане з'єднання між їх MUA і сервером SMTP, наприклад, використовуючи ім'я користувача і пароль. "SMTP після POP" і "SMTP після IMAP" засновані на імені користувача і паролі і аутентифікують користувача через успішне з'єднання по протоколах POP і IMAP. Як тільки з'єднання відбулося, користувачеві дозволяється посилати повідомлення на якийсь певний проміжок часу, наприклад на 10 хвилин. За допомогою цих методів можна боротися із спуфінгом імен відправника і/або хоста, проте імена користувачів і паролі не захищаються, і легко можуть бути компрометовані.

Методи криптографічної аутентифікації направлені на боротьбу із спуфінгом в загальному сенсі. Цифровий підпис додається в повідомлення і підтверджує особу відправника. Підписи розрізняються по мірі достовірності:

деякі засновані на імені користувача або його адресі (підписує зазвичай MUA), інші вимагають підтвердження домена або ESP (підписує MTA).

Несиметрична криптографія включає протоколи S/MIME [33], PGP [11], META Signatures [34] IIM [35], Domain Keys [36], Microsoft Postmarks [37] і ін. Комітет з інженерних питань Інтернет (IETF) створив робочу групу "Стандарти Підписів для Аутентифікації Повідомлень". У Domain Keys і PGP відкритий ключ зберігається в спеціальних БД (DNS запису і KeyServer відповідно), таких, що мають відношення до відправника. У інших схемах відкритий ключ є частиною підпису, що дозволяє одержувачеві перевірити підпис без доступу в Інтернет, але в даному випадку проблематично перевірити автентичність підпису. Достовірність підпису може бути підтверджена третьою стороною, наприклад центром сертифікації [38].

Метод криптографічної аутентифікації має наступні недоліки. По-перше, метод бореться із спуфінгом і не застосовується, якщо повідомлення не містить підроблених даних. Також метод не ефективний у разі компрометації ключа. По-друге, з боку одержувача розпізнати повідомлення з підробленими даними можуть або MTA провайдери, або MUA користувача тільки після повного завантаження, обчисленні хеш-кодування і порівнянні з розшифрованим хеш-кодуванням. По-третє, деякі посередники перетворюють повідомлення, наприклад, змінюють кодування або додають інформацію про проходження листа через їх сервер [38]. По-четверте, у разі підтвердження на рівні домена можливе звинувачення в розсилці спаму організацію, а не людину, хто розіслав спам. При підтвердженні на рівні користувача необхідне застосування криптографії на кожному ПК, що збільшує обчислювальне навантаження. По-п’яте, необхідне стандартизувати формати даних ключів і сертифікатів, криптографічних алгоритмів і вирішити проблеми інфраструктури. Для успішного застосування інфраструктури відкритих ключів (PKI), вона повинна бути об'єднаною (щоб жоден провайдер не зміг обмежувати можливості ринку), розподіленою і такою, що точно дублюється. Гарфінкел (Garfinkel) [39] стверджує, що причиною повільного застосування PKI невідповідність її можливостей потребам клієнта.

Для того, щоб уникнути аутентифікації заснованою на криптографії, яка

вимагає деякі види PKI і застосовує наскрізну аутентифікацію, були запропоновані механізми аутентифікації шляху (в основному на основі DNS). Теорія аутентифікації шляху полягає в тому, що якщо пункт призначення підтверджує попередня транзитна ділянка (SMTP клієнт) і може довіряти результатам і якщо попередня транзитна ділянка підтверджує дійсного відправника, тоді дійсний відправник повідомлення може вважатися підтвердженим і авторизованим [40]. Сукупність (заснованих на DNS) методів аутентифікації шляху для боротьби із спамом називаються Lightweight Message Authentication Protocol (LMAP), вказаний в Internet draft [41]. Протокол LMAP бореться з проблемою підробки повідомлень перевіряючи, чи має хост, з якого було послано повідомлення, право розсилати повідомлення, використовуючи домен в конверті або заголовку. Наприклад, він перевіряє, чи дане повідомлення, точно послане з User@Mail.Com було послано MTA Mail, що діяв від імені організації.Com. Інакше повідомлення вважається підробленим, або ж воно було послане через проміжний MTA, який використовувався як зовнішній поштовий транслятор (його недоліки обговорюються вище). LMAP грунтується на 2 принципах: публікація ідентифікаційних даних доменом (в більшості випадків за допомогою записів DNS) і використання тих даних одержувачем (MTA). Таким чином, це застосовується в протоколах SMTP (RFC2821) і DNS (RFS 1034).

Коли повідомлення посилається через SMTP, у одержувача MTA є багато способів аутентифікувати поштового відправника: IP-адрес, аргумент HELO/EHLO, зворотний шлях і заголовки повідомлення. Всі перераховані способи можуть використовуватися для різного роду ідентифікацій. Це привело до появи різноманітних реалізацій протоколу LMAP що відрізняються видом підтверджуваної інформації про відправника, даними з якими асоціюється ця інформація, джерело мережі і тип запису DNS (якщо використовується DNS). До найпоширеніших реалізацій LMAP відносяться SPF [42], RMX [43], DMP [44], MS–Sender–ID [45], CSV [46] і MTAMARK [47]. Аналіз і порівняння реалізацій LMAP представлені Лейбзоном (Leibzon) [38] Шраєном (Schryen) і Хове–ном (Hoven) [48]. Але стандартизація при цьому не була досягнута. Більш того, робоча група MARID була розпущена в 2004 році.

Методи, засновані на використанні протоколу LMAP, мають ряд недоліків. По–перше, протокол LMAP призначений для боротьби із спуфінгом, а не із спамом. Спам–повідомлення, що не містять підроблені дані, не виявлятимуться. По–друге, багато серверів, що наприклад розсилають вітальні листівки, дозволяють розсилати повідомлення з вказівкою зворотної адреси, причому існування зворотної адреси не перевіряється. По–третє, спамери можуть придбати домен, розіслати спам від його імені і кинути його. По–четверти, більшість версій LMAP використовують DNS, що робить DNS критичним ресурсом. Ненадійність в DNS може дозволити ворожим сторонам пересилати підроблену аутентифікуючу інформацію. Атаки "відмова в обслуговуванні" проти DNS серверів можуть зробити неможливим для клієнтів LMAP отримання аутентифікуючих даних.

1.3.5 Аналіз схеми верифікації відправника

Оскільки спамери розсилають мільйони повідомлень, вони ігнорують повідомлення про неможливість доставки листів, що використовують механізми верифікації для блокування спаму. У схемі верифікації, жодне повідомлення не посилається поки відправник, або організація-відправник не буде занесена в білий список. Якщо відправник намагається доставити повідомлення в захищений ящик, дане повідомлення ставиться в чергу на карантин і повертається запит. Щоб користувач міг надалі відправляти листи, йому пропонується відповісти на повідомлення, виконати математичні обчислення або розпізнати текст з картинки (повністю автоматизований відкритий тест Тюрінга по розпізнаванню людей і машин CAPTCHA [49]). Як тільки завдання коректно вирішене, адреса відправника заноситься в білий список одержувача і первинне повідомлення відправляється. Оскільки виклик одержувача вимагає відповіді відправника, ця процедура також називається "виклик-відгук". Прикладом є Sender Address Verification Extension (SAVE), запропонована Bless і ін. [50]: для адреси відправника, не занесеного в білий список, агент пересилки повідомлень організації відправника створює повідомлення, що містить 2 ребуси: один для

людини, наприклад, картинка з комбінацією цифр, інший для машини, наприклад, злом хеш-кодування. Методи, засновані на процедурі "виклик-відгук" володіють наступними недоліками: процес передачі поштових повідомлень стає складнішим, із-за процедури "виклик-відгук" збільшується Інтернет трафік, масова розсилка звичайних повідомлень типу інформаційних бюлетенів стає неможливою, людині може бути складно розпізнати об'єкт, якщо він представлений дуже недбало, відгуки, орієнтовані на завдання, що виконуються вручну, можуть піддаватися технічним атакам. Якщо завдання спамера ідентифікувати текст на картинці, він може обманним шляхом заманити користувача відвідати сайт з картинкою і змусити його вирішити цю задачу, обіцяючи взамін доступ, наприклад, до ресурсів для дорослих.

1.3.6 Методи , які засновані на оплаті вихідних повідомлень

Дані методи відносяться до поштових систем, що створюють економічні перешкоди для розсилки спаму. Поштові сервери вимагають невеликої оплати за доставку повідомлення або за ухвалення до доставки. Ця плата не відчувається при відправці одиничних листів, але стає відчутною при відправці багато листів [51]. Але виникає проблема розсилки великої кількості повідомлень. Способом оплати може бути час центрального процесора або об'єм пам'яті, конвертована або віртуальна валюта. Оплата часом часто позначається як процедура "з доказом роботи".

Основу методу оплати часом центрального процесора складає система "доказ роботи", яка припускає використання параметрів роботи центрального процесора (CPU) для обчислення ціни за вихідні повідомлення. Функції ціноутворення, запропоновані Дворком і Наором (Dwork and Naor) [52], штампи "HashCash" [53] і цифрові штампи спам-фільтра "Camram" [54] є найпродуманішими. Синтія Дворк і Моні Наор запропонували поштову систему, яка пропонує комп'ютеру відправника перед відправкою кожного листа вирішити обчислювальну задачу. За наявності додаткової інформації підрахунок може значно спрощується, що дозволяє розсилати звичайні масові повідомлення. Кожен

користувач може задати список відправників, повідомлення від яких приймаються без перевірок. Автори методу пропонують застосовувати функцію ціноутворення не до повідомлення, а тільки до значення хеш-кодування функції. Даний метод не позбавлений недоліків. По-перше, витрачаються ресурси відправника. CPU заражених комп'ютерів завантажуються на 100%. По-друге, час обчислення залежить від швидкості CPU, що ставить користувачів в нерівне положення. По-третє, якщо одержувач погодився отримувати повідомлення від певного відправника, то необхідні механізми аутентифікації на рівні користувача.

Абаді та ін. [55] запропонував сукупність помірно складних функцій, заснованих на пам'яті. Підхід оплати часом пам'яті полягає в тому, щоб змусити відправника обчислити деяку цільову функцію, для чого відправникові необхідно дістати доступ до множини елементів великого масиву даних в непередбачуваній послідовності. Розмір масиву даних можна вибрати так, щоб він був значно більшим, чим найбільша кеш-пам'ять, але, в той же час, значно меншим об'єму ОЗП, встановленого в достатньо сильно застарілих комп'ютерах. Дворк і ін. показали, що в середньому, відправник повідомлення повинен виконати багато операцій доступу до пам'яті, а одержувач, щоб перевірити правильність обчислення цільової функції, повинен виконати значно меншу кількість доступів. В принципі підхід, заснований на пам'яті, має ті ж недоліки, що і метод оплати часом центрального процесора.

У грошовому методі (Monetary) відправник платить певну суму за кожне повідомлення, якщо одержувач не заніс відправника в білий список. Як плата може використовуватися реальна готівка (відправник передає третій стороні заставу і втрачає її, якщо одержувач вважає, що був відправлений спам) або віртуальні гроші. У системі оплати готівкою використовуються електронні марки, що прикріплюються до кожного повідомлення, що відправляється. Якщо незнайомець намагається відправити повідомлення без марки, воно повертається з вимогою прикріпити її. Коли одержувачеві приходить лист з маркою, він може обналічить марку в своєму банку. Систему можна спростити, якщо мережеві вузли, а не користувачі, прикріплюватимуть марки. У роботах Темплтона (Templeton) [56] розглядаються проблеми розробки і застосування такої системи.

Loder та інші [57] показали, що заставний механізм може бути ефективнішим за ідеальний фільтр. Згідно Fahmann [58], кожен одержувач встановлює власну ціну і вирішує прийняти гроші або відхилити. Система оплати реальною готівкою володіє наступним недоліком [56]: оскільки інформація про відправника із заголовка може бути підроблена, потрібний додатковий механізм аутентифікації. Застосування несиметричної криптографії має наступні недоліки: по-перше, необхідна PKI для обслуговування ключів шифрування, по-друге, повідомлення необхідно завантажити перед перевіркою підпису; по-друге, з'являться віруси, що посилають без відома користувачів листа на фіктивні адреси, а шахраям залишиться тільки зняти гроші; по-третє, система електронних марок складна технічно і організаційно. Багато дослідників пропонували такі інфраструктури у минулому, але жодна з них не була успішною.

Тернер і Хеві (Turner and Havey) [20] запропонували інфраструктуру з псевдо валютою (Lightweight Currency Protocol) [59]. Суть в тому, що кожен поштовий сервер використовуватиме псевдо валюту LCP для оплати відправлених повідомлень і отримуватиме гроші при отриманні повідомлення. Кожна організація може випустити свою власну валюту. Подібна структура підтримувала б баланс між ВП і ВихП. Коли спамери розішлють мільйони повідомлень зі своїх доменів, вони не отримають сумірну кількість відповідей, тобто у них не буде псевдо валюти, необхідної для відправки нової партії листів. У цього підходу є наступні недоліки і обмеження: по-перше, спамери можуть зловживати послугами ISP, створюючи облікові записи у провайдерів, чия валюта повсюдно приймається; по-друге, залишається можливість розсилки спаму черех заражених комп'ютерів; по-третє, утруднюється організація потрібної розсилки.

1.3.7 Метод обмеження вихідних повідомлень

Останнім часом все більше спаму почали розсилати заражені ПК. З цим зіткнулися як ISP, які обслуговують анонімних користувачів, так і ISP, що вимагають обов'язкову аутентифікацію. Провайдери прагнуть не дати своїм клієнтам розсилати спам мільйонам користувачів [60]. Для цього окрім

обмеження частоти відправки листів, необхідно заборонити автоматичну реєстрацію облікових записів. На жаль, навряд чи можливо запровадити і контролювати реалізацію даного методу по всьому світу.

1.3.8 Аналіз технології приховування адреси

Технології захована адреси направлені на те, що перешкоджання попаданню електронних адрес користувачів в бази спамерів. Хол (Hall) [61] запропонував концепцію віртуальних каналів, через які доступна поштова адреса. Кожен канал має чітку структуру і містить в собі ім'я облікового запису і криптостійкий, псевдовипадковий ланцюжок (визначник типу каналу). Кожному законному користувачеві дозволяється дізнатися одну з цих адрес каналу. Власникові облікового запису доступні прості механізми управління, що дозволяють відкривати нові канали, закривати старі, перемикати канали з подальшим повідомленням про це вибраних. Адреса каналу має наступну структуру: Ім'я_користувача–ідентіфікатор_каналу–@адрес_сервер, наприклад: User–947539–@Mail.Ru. Канали діляться на *send–only* (тільки відправка, прийом повідомлень заборонений), *private channel* (закритий канал, прийом повідомлень тільки від певних користувачів) і *public channel* (загальнодоступний канал, прийом повідомлень від всіх користувачів). Недоліками є: по–перше, ускладнена поштова комунікація і обмін адресами; по–друге, віруси, що заразили ПК, можуть скомпрометувати вміст адресної книги.

Концепція Габбер (Gabber) [62] відрізняється прив'язкою поштової адреси до певного користувача. Концепція одноразових адрес (SPA – Single–Purpose Address), запропонована Ioannidis, спирається на складну систему зашифрованих правил.

1.3.9 Аналіз підходів, заснованих на репутації відправника

Дані методи припускають, що рішення прийняти або відхилити повідомлення одержувач приймає, керуючись репутацією відправника. Коаліція

постачальників поштових послуг (Email Service Provider Coalition) [63] пропонує реєструвати і сертифікувати легітимних масових відправників. Spam–house пропонує створити новий домен верхнього рівня, наприклад .Mail. Будь–який домен повинен володіти доменним ключем (наприклад, Company.Ru.Mail), що видається за наявності достовірної і підтвердженої інформації "Who–Is", наявність ефективних методів для захисту від спаму, і якщо з моменту реєстрації домена пройшло більше 6 місяців. Також домен повинен надати інформацію про IP–адрес і ім'я хоста відправляючого поштового сервера. Інфраструктура повинна стати ефективною, але невідомі способи розробки ефективних технічних методів захисту від спаму.

1.4 Постановка задачі дослідження

З урахуванням вищевикладеного, в роботі формулюється концепція побудови перспективних систем ФС, заснована на застосуванні наступних основоположних системних принципів:

1. Принцип ієрархічної організації, що означає побудову системи ФС в класі багаторівневих ієрархічних систем захисту інформації з розділенням (декомпозицією) її на рівні, що відрізняється вибором мети в рамках політики безпеки.

2. Принцип комплексування моделей, методів і алгоритмів ФС, що полягає в застосуванні як класичних методів фільтрації, так і методів інтелектуальної фільтрації на основі нейронних мереж.

3. Принцип побудови відкритих інформаційних систем в якості основи інтелектуалізації і стандартизації технологій обробки інформації.

Показано, що для коректної постановки і вирішення задач ФС в рамках запропонованої концепції потрібна розробка моделі спаму і корисних повідомлень, алгоритмів і методів ФС, що забезпечують зниження рівня помилок першого і другого роду.

Метою дипломної роботи є розробка методології проектування ефективної системи захисту інформації, що забезпечує ФС в організації.

Для досягнення поставленої мети в роботі необхідно виконати наступні завдання:

– розробити концепцію побудови системи ФС в організації на основі методів штучного інтелекту;

– розробити багатоагентну архітектуру ієрархічної системи ФС в організації;

– розробити ефективний алгоритм класифікації електронних повідомлень з врахуванням семантики повідомлення;

– оцінити ефективність запропонованих підходів до ФС в організації.

Отже, в першому розділі дипломної роботи виконаний аналіз різних видів спам–повідомлень з погляду їх загроз захищеності інформації. Розглянуті переваги і недоліки відомих методів до протидії загрозам захищеності інформації. Робиться висновок про необхідність розробки нової архітектури системи протидії розповсюдженню спам-повідомлень та нових методів і алгоритмів фільтрації, що дозволяють ефективніше, в порівнянні з існуючими системами, забезпечувати ФС в потоці електронної пошти.

2 РОЗРОБКА АЛГОРИТМІВ ВИЯВЛЕННЯ СПАМОВИХ ПОВІДОМЛЕНЬ

2.1 Архітектури системи виявлення спамових повідомлень

Всі існуючі на даний момент системи фільтрації спам–повідомлень за місцем установки діляться на два класи: серверна система (рисунок 2.1, а), коли фільтр встановлений на сервері електронної пошти, і персональна (рисунок 2.1, б), коли фільтр встановлений на персональному комп'ютері кожного користувача електронної пошти [64].

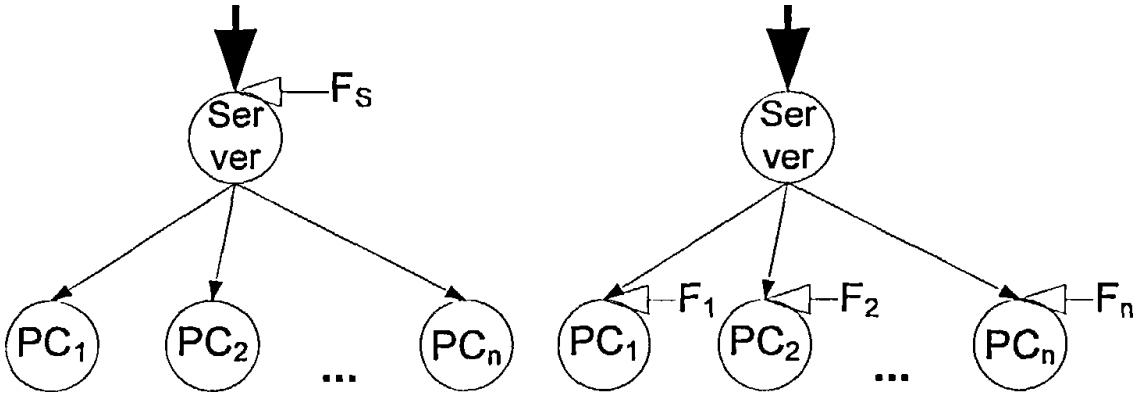


Рисунок 2.1 – Схема розташування фільтрів

де M – потік вхідних електронних повідомлень, призначених якомусь окремому користувачеві або всім користувачам поштової системи;

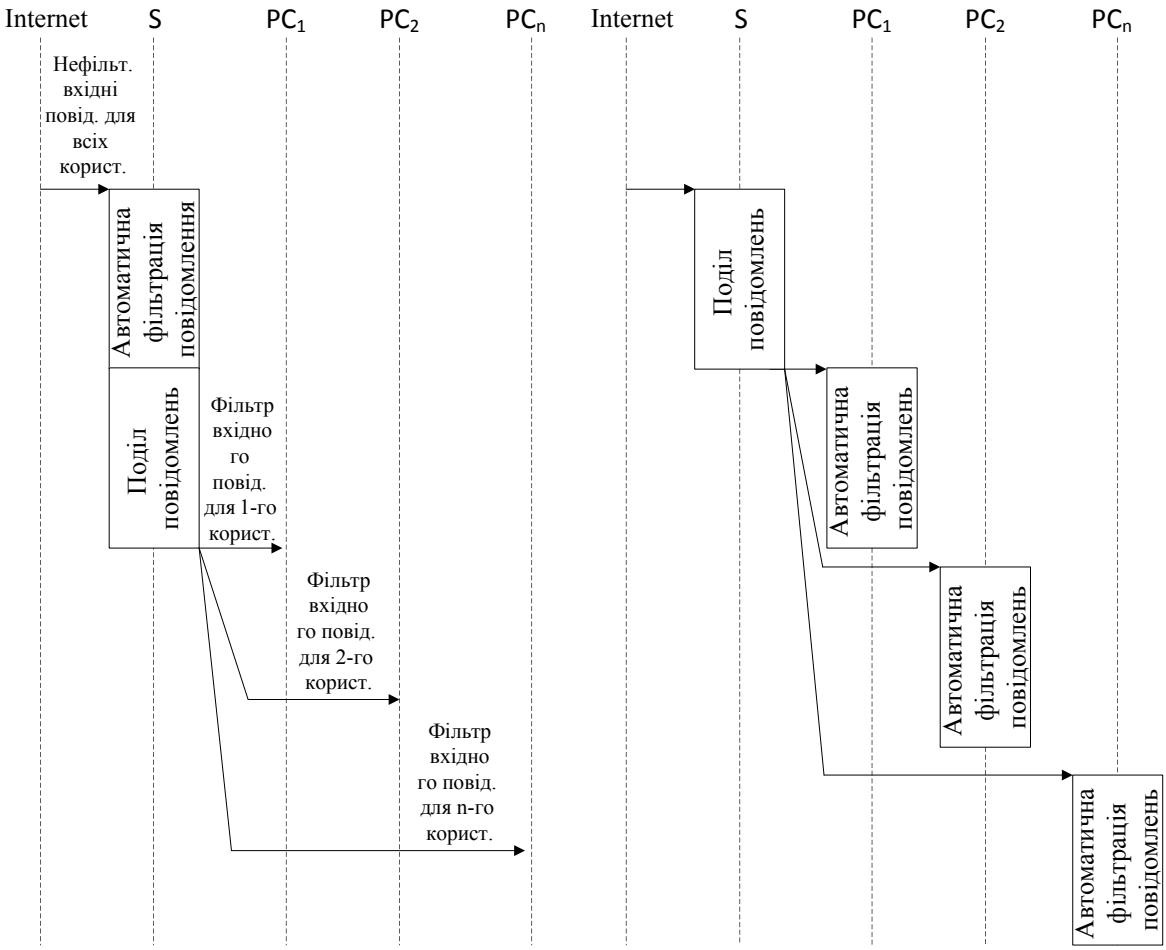
S – сервер електронної пошти;

F_s – серверний класифікатор електронних повідомлень;

PC_j – персональний комп'ютер j-го користувача системи електронної пошти;

F_j – персональний класифікатор, індивідуально налаштований та навчений j-м користувачем.

Діаграма взаємодії для випадку розташування спам–фільтра на поштовому сервері показана на рисунку 2.2 (а). Діаграма взаємодії для випадку розташування спам–фільтрів на персональному комп'ютері кожного користувача електронної пошти показана на рисунку 2.2 (б).



а) б)

Рисунок 2.2 – Діаграма взаємодії у разі установки фільтра спаму на поштовому сервері (а) і на персональному комп'ютері (б)

У разі установки фільтра спаму на сервер електронної пошти з'являються наступні переваги: з'являється можливість досить швидкого виявлення і відсікання масових розсилок однакових повідомлень, надісланих на електронні поштові адреси різних користувачів; з'являється можливість оперативно пристосовування до винайдених методів і способів обману поширених спам–фільтрів за рахунок централізованого адміністрування системи фільтрації та збору статистичної інформації на основі скарг користувачів на отримані електронні повідомлення, що містять спам; економиться пропускна здатність каналу зв'язку від сервера електронної пошти до персонального комп'ютера користувача (так звана, остання миля), що особливо відчутно при роботі в Інтернеті через звичайний dial-up модем.

До недоліків серверних фільтрів спаму можна віднести неможливість обліку

області інтересів кожного конкретного користувача, що може привести до прийняття помилкового рішення про пропуск спам-повідомлення (помилки першого роду, що не так критично) або блокуванні корисного листа (помилка другого роду, що практично неприпустимо). Інформаційні потоки в системах з розташування фільтрів спаму на різних вузлах мережі представлені на рисунку 2.3.

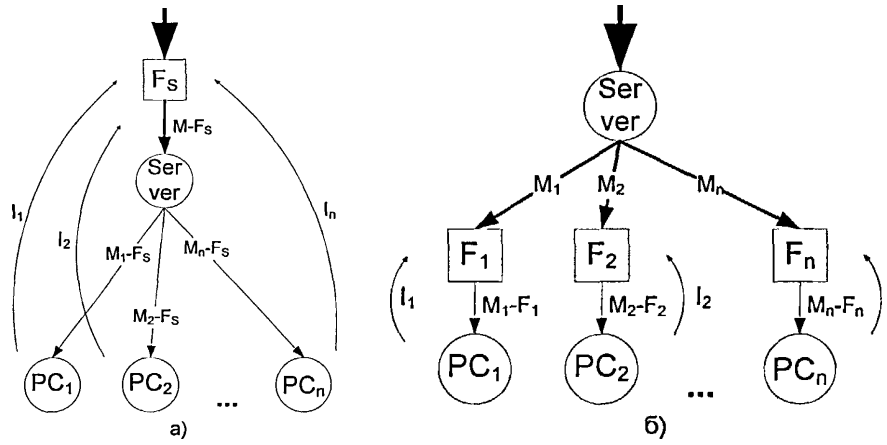


Рисунок 2.3 – Інформаційні потоки в системах з розташування фільтрів спаму на різних вузлах мережі

В окремому випадку кожне повідомлення призначене тільки одному користувачу поштової системи, тобто для а) і б) маємо відповідно:

$$M - F_S = (M_1 + F_S) + (M_2 + F_S) + ... + (M_n + F_S) = \sum_{k=1}^{n} [(M]_k + F_S \,, \qquad (2.1)$$

$$M = M_1 + M_2 + ... + M_n = \sum_{k=1}^{n} M_k \,. \qquad (2.2)$$

У більш загальному випадку відправник електронного повідомлення може вказати в якості одержувачів двох і більше користувачів поштової системи, тоді для випадків а) і б) маємо відповідно:

$$M - F_S = (M_1 - F_S) \cup (M_2 - F_S) \cup ... \cup (M_n - F_S) = \bigcup_{j=1}^{k} (M_j - F_S), \quad (2.3)$$

$$M = M_1 \cup M_2 \cup ... \cup M_n = \bigcup_{j=1}^{k} M_j . \quad (2.4)$$

Навчальна множина формується користувачами поштової системи з надісланих електронних повідомлень, які зазнали ручної класифікації, тому маємо для випадків розташування фільтра спаму на поштовому сервері і на персональному комп'ютері користувача відповідно [65]:

$$I_1 \in (M_1 - F_S), I_2 \in (M_2 - F_S), ..., I_n \in (M_n - F_S), \quad (2.5)$$

$$I_1 \in (M_1 - F_1), I_2 \in (M_2 - F_2), ..., I_n \in (M_n - F_m), \quad (2.6)$$

де M – потік вхідних електронних повідомлень, призначених якомусь окремому користувачеві або всім користувачам поштової системи;

S – сервер електронної пошти;

F_S – серверний класифікатор електронних повідомлень;

PC_j – персональний комп'ютер j-го користувача системи електронної пошти;

F_j – персональний класифікатор, індивідуально налаштований і навчений j користувачем системи електронної пошти;

$M–F_S$ – потік вхідних електронних повідомлень, призначених якомусь окремому користувачеві або всім користувачам системи електронної пошти та відфільтрованих серверним класифікатором F_S;

$M_j–F_S$ – потік вхідних електронних повідомлень, призначених j-му користувачеві системи електронної пошти та відфільтрованих серверним класифікатором F_S;

I_j – електронні повідомлення, вручну класифіковані j-м користувачем поштової системи і призначені для навчання серверного класифікатора F_S або персонального класифікатора F_j;

M_j – потік вхідних електронних повідомлень, призначених j-му користувачеві поштової системи;

M_j–F_j – потік вхідних електронних повідомлень, призначених j користувачеві поштової системи та відфільтрованих персональним фільтром F_j.

Діаграма взаємодії в системі з користувацьким фільтром спаму і навчанням системи фільтрації показана на рисунку 2.4. Діаграма взаємодії в системі з системою з серверним фільтром спаму і навчанням системи фільтрації показана на рисунку 2.5.

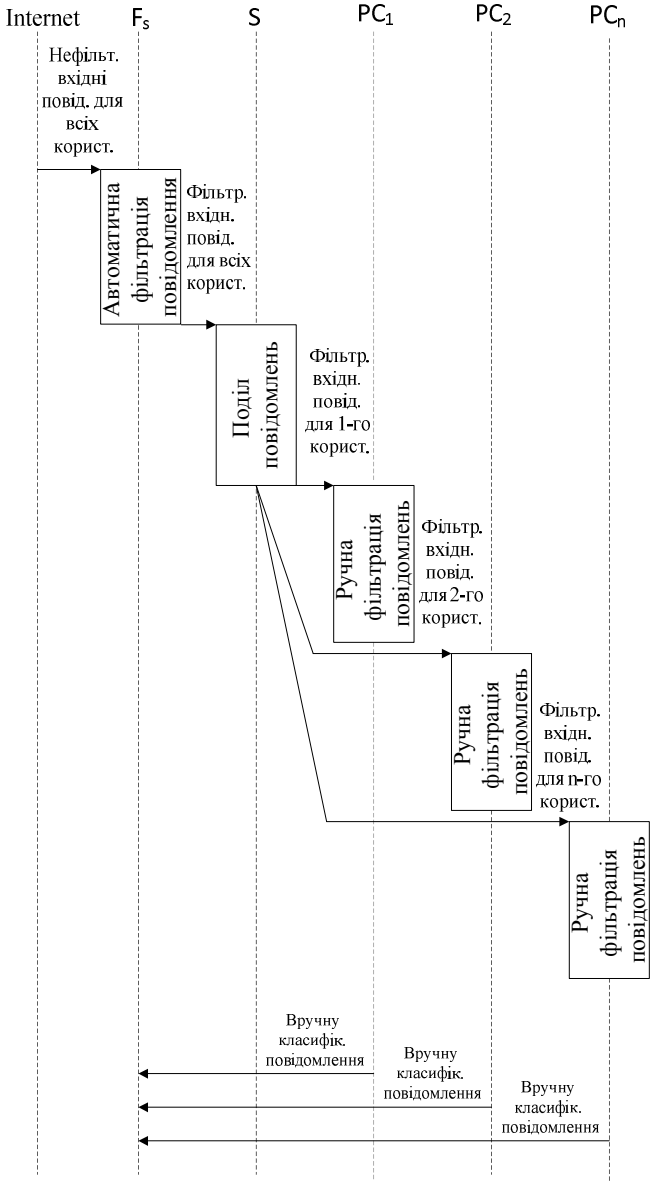


Рисунок 2.4 – Діаграма взаємодії в системі з користувацьким фільтром спаму

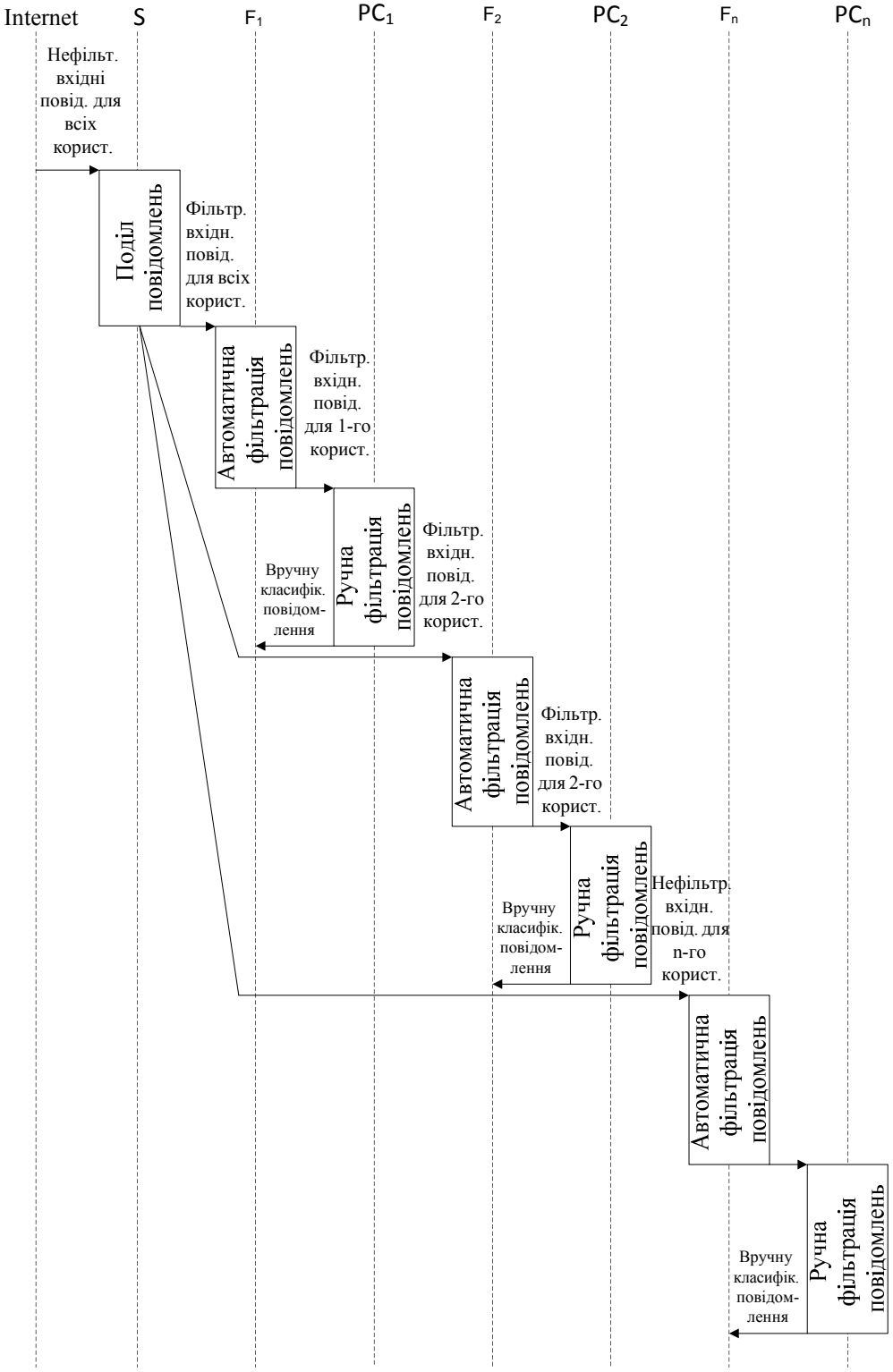


Рисунок 2.5 – Діаграма взаємодії в системі з серверним фільтром спаму

Способи формування БЗ серверних спам-фільтрів діляться на три види, кожен з яких розглянуто нижче.

Просте перетинання, коли для формування БЗ використовуються повідомлення, помічені як спам всіма користувачами поштової системи (рисунок 2.6). У цьому випадку рівень помилок першого роду підвищується, а

рівень помилок другого роду знижується. Метод використовується досить широко.

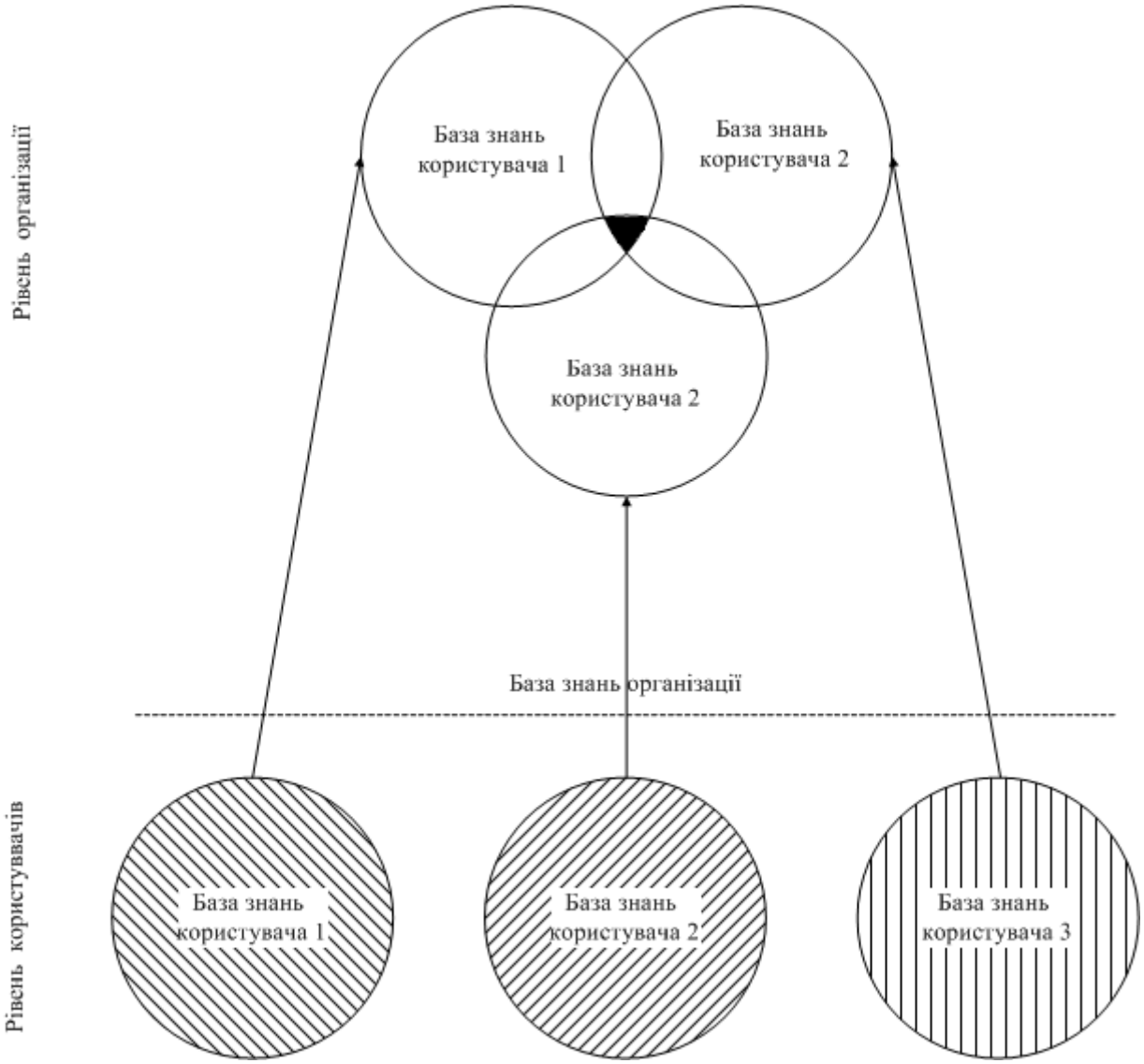


Рисунок 2.6 – Спосіб формування БЗ підприємства шляхом простого перетину БЗ окремих користувачів

Даний спосіб формування баз знань можна виразити таким чином:

$$B = b_1 \cap b_2 \cap \ldots \cap b_n = \bigcap_{k=1}^{n} b_k \,, \tag{2.7}$$

де B – БЗ організації;

b_k – БЗ k–го користувача поштової системи;

n – кількість користувачів системи електронної пошти в організації.

Складне перетинання, коли повідомлення, позначене визначенним користувачем як спам, в свою БЗ поштова система приймає на основі рейтингу користувача, який, в свою чергу, обчислюється з того, скільки з точки зору інших користувачів, даний користувач повідомлень позначив вірно, і скільки невірно (рисунок 2.7). У більшості випадків оцінка і кожного користувача відбувається в неявному вигляді: система самостійно порівнює, скільки одержувачів одного і того ж повідомлення помітили його як спам. У цьому випадку рівень помилок першого роду підвищується, а рівень помилок другого роду знижується. Метод використовується не дуже часто.

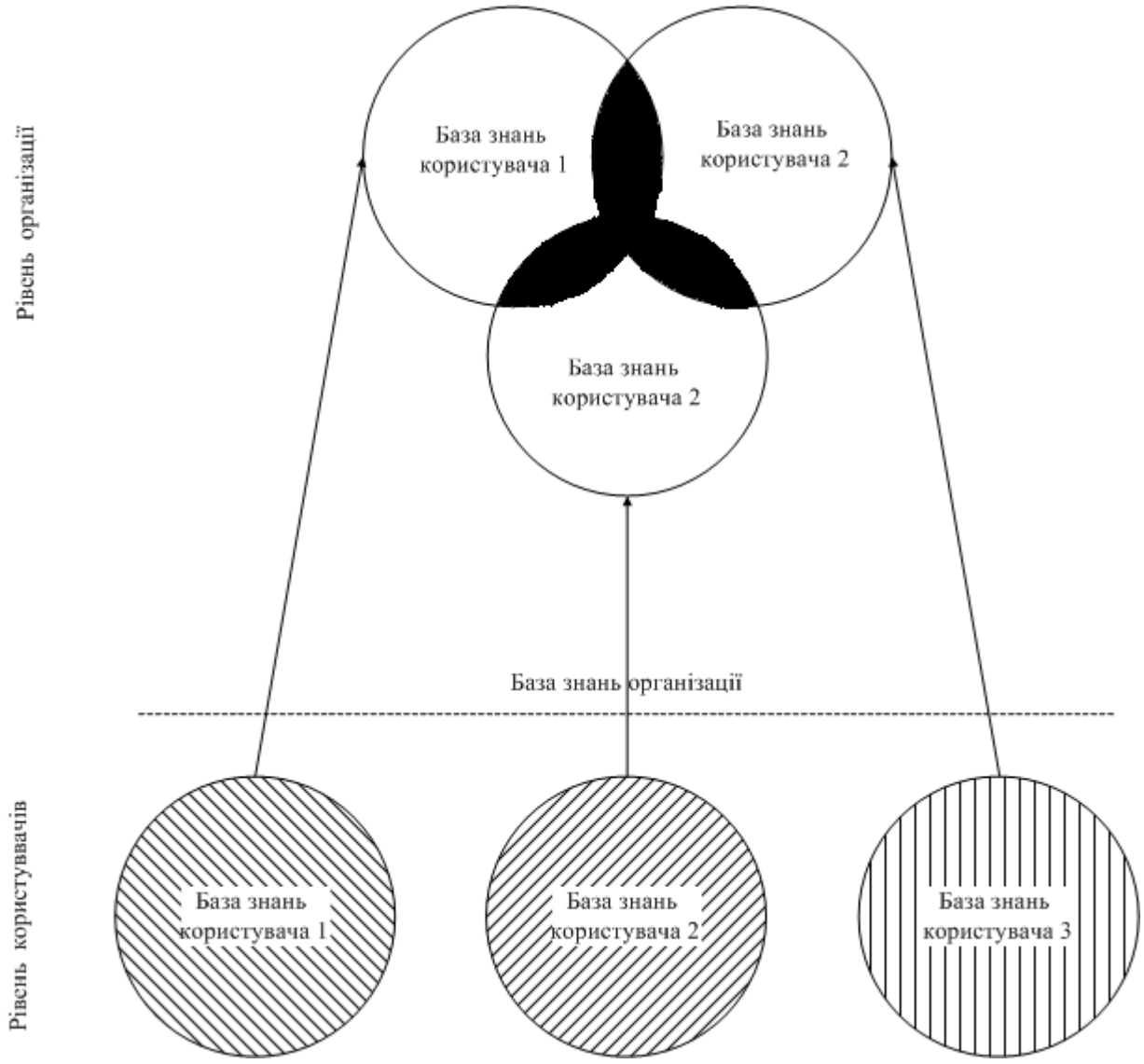


Рисунок 2.7 – Спосіб формування БЗ підприємства шляхом складного попарного перетину баз знань окремих користувачів

У спрощеному вигляді, даний спосіб формування баз знань можна виразити наступним чином:

$$B = (b_1 \cap b_2) \cup (b_1 \cap b_2) \cup \ldots \cup (b_1 \cap b_n) \cup (b_2 \cap b_3) \cup \ldots$$
$$\cup (b_2 \cap b_n) \cup \ldots \cup (b_{n-1} \cap b_n) = \bigcup_{k=1, j=1}^{k=n, j=n, k \neq j} (b_k \cap b_j) \quad (2.8)$$

Об'єднання, коли для формування БЗ використовуються повідомлення, помічене як спам хоча б, одним користувачем системи електронної пошти (рисунку 2.8). У цьому випадку рівень помилок першого роду знижується, а рівень помилок другого роду підвищується. Метод використовується вкрай рідко.

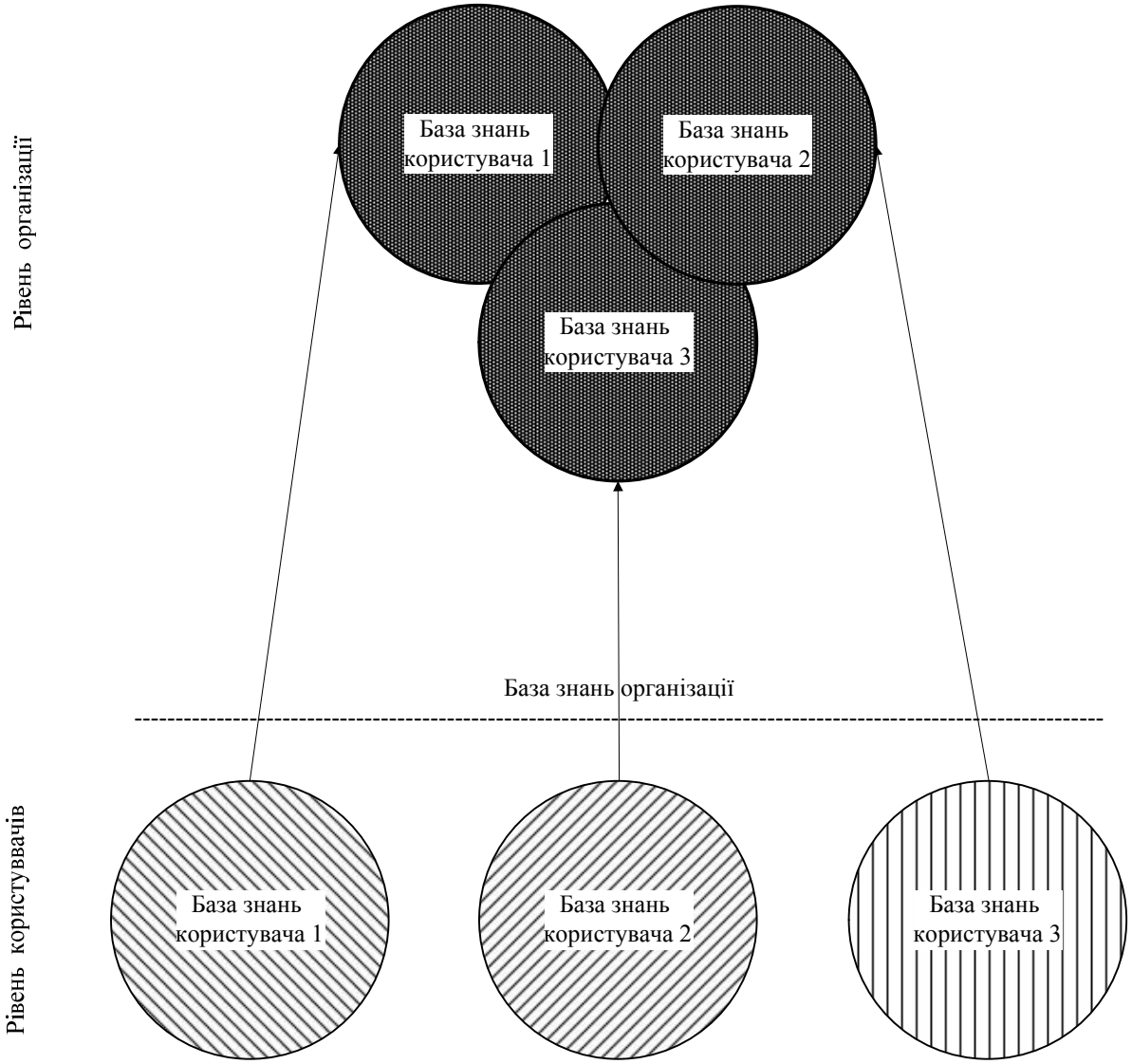


Рисунок 2.8 – Спосіб формування БЗ підприємства шляхом об'єднання баз знань всіх користувачів

Даний спосіб формування БЗ можна виразити таким чином:

$$B = b_1 \cup b_2 \cup ... \cup b_n = \bigcup_{k=1}^{n} b_k / \qquad (2.9)$$

Персональні фільтри дозволяють враховувати область інтересів конкретного користувача і точніше фільтрувати вхідні повідомлення. Недоліками персональних спам-фільтрів є: неможливість використання досвіду інших користувачів. Тобто при отриманні спам–повідомлення, що пройшло через персональний фільтр, кожен користувач повинен вказати своєму фільтру на помилку, щоб надалі при отриманні подібних повідомлень спам-фільтр виявляв і відтинав їх; витрата пропускної здатності каналу зв'язку "останньої милі" на скачування спам-повідомлень.

Розглядається задача розробки процедури формування БЗ інтелектуальної системи боротьби зі спамом, що об'єднує в собі всі переваги серверних та персональних фільтрів і позбавленою здебільшого описаних вище недоліків [66].

Умовно організація, яка застосовує систему боротьби зі спамом, ділиться на три рівні: рівень кінцевого користувача; рівень відділу або підрозділу; рівень організації в цілому.

На нижньому рівні кожен користувач всю отриману кореспонденцію ділить на корисну і на спам. Також користувач може вказати папки, в котрих зберігаються корисні документи. Після цього система виділяє слова, більш властиві спам-повідомленнями, і слова, частіше зустрічаються в корисних повідомленнях. Потім слова-маркери передаються на рівень відділу, де всі списки об'єднуються, оскільки область професійних інтересів всіх співробітників одного відділу приблизно збігаються. На останньому етапі формування БЗ система виділяє слова-маркери, виділені усіма відділами, і відкидає слова, невиділені хоча б одним відділом. Це пов'язано з можливим розходженням області інтересів різних відділів однієї організації [66]. Процес формування БЗ системи боротьби зі спамом зображений на рисунку 2.9.

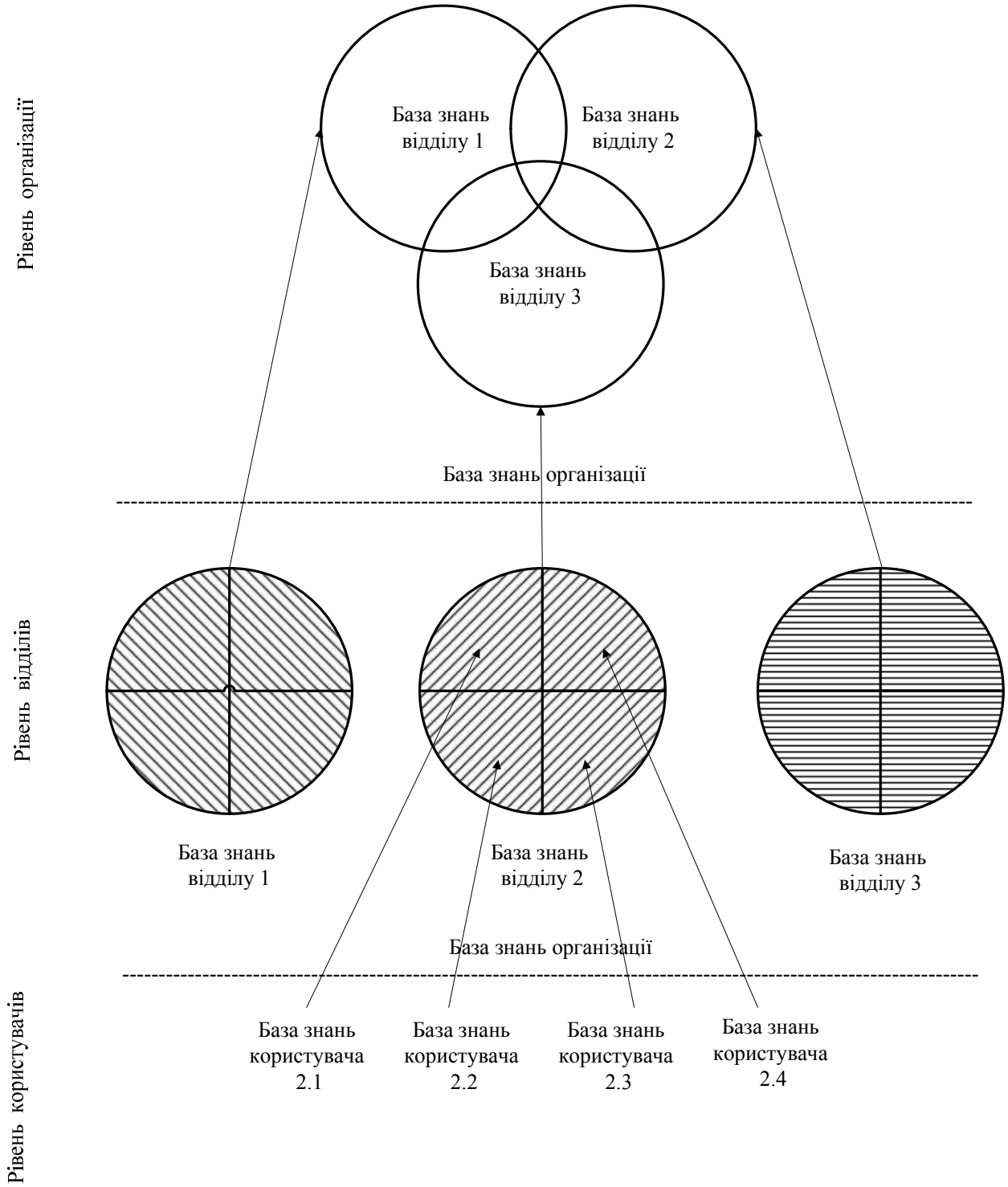


Рисунок 2.9 – Ієрархічне формування БЗ багаторівневої системи боротьби зі спамом

При такій схемі досягається досить висока ефективність роботи спам-фільтра за рахунок збільшення обсягу навчальної вибірки (при створенні БЗ на

рівні відділу) при збереженні рівня помилок на низькому рівні за рахунок використання з навчальної вибірки випадково попавших маркерів (на рівні організації в цілому).

Даний спосіб формування баз знань можна виразити таким чином:

$$B_1 = b_{11} \cup b_{12} \cup ... \cup b_{1n} = \bigcup_{k=1}^{n} b_{1k}\,, \tag{2.10}$$

$$Base = B_1 \cap B_2 \cap ... \cap B_p = \bigcap_{k=1}^{p} B_k\,. \tag{2.11}$$

Тоді:

$$\begin{aligned} Base = (b_{11} \cup b_{12} \cup ... \cup b_{1n}) \cap (b_{21} \cup b_{22} \cup ... \cup b_{2m}) \cap ... \\ \cap (b_{p1} \cup b_{p2} \cup ... \cup b_{po}) = \bigcap_{k=1}^{p} \left(\bigcup_{j=1, j\neq k}^{r} b_{k,j} \right) \end{aligned}\,, \tag{2.12}$$

де p – кількість відділів в організації;

n – кількість працівників в першому відділі організації;

m – кількість працівників у другому відділі організації;

o – кількість співробітників в p-ому відділі організації;

b_{po} – БЗ, зформована o-м співробітником p-го відділу організації;

b_k – БЗ k-го відділу організації;

$Base$ – БЗ всієї організації;

Схема інформаційних потоків в системі, що розробляється для боротьби зі спамом показана на рисунку 2.9.

Діаграма взаємодії в розробленій системі боротьби зі спамом показана на рисунку 2.10.

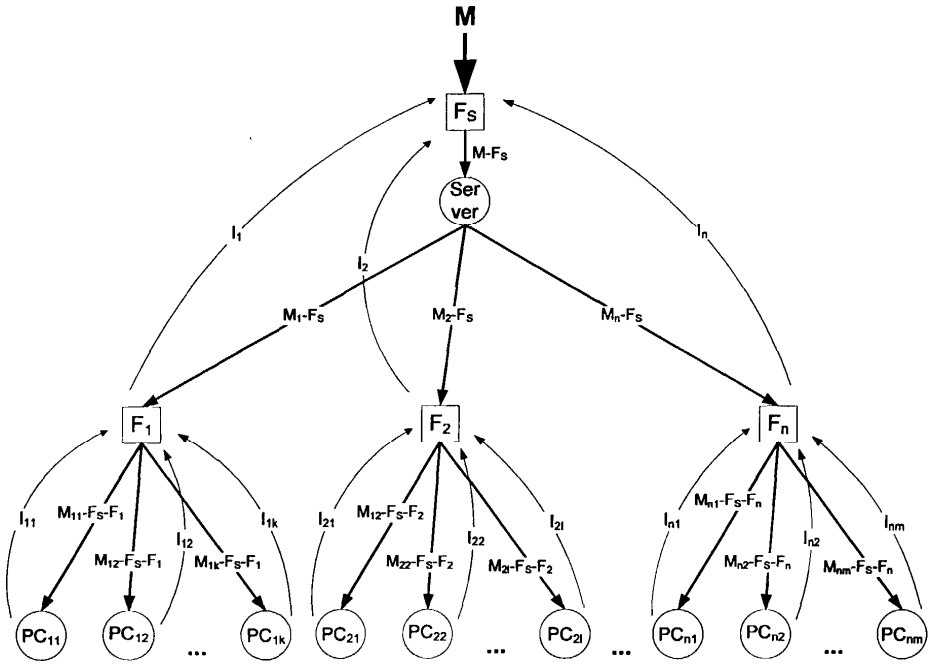


Рисунок 2.10 – Інформаційні потоки в баторівневій системі боротьби зі спамом

В окремому випадку, коли кожне вхідне електронне повідомлення призначено тільки для одного користувача, маємо:

$$M - F_S = (M_1 - F_S) + (M_2 - F_S) + ... + (M_n - F_S) = \sum_{k=1}^{n} (M_k - F_S), \qquad (2.13)$$

$$\begin{aligned} M - F_S &= (M_{k,1} - F_S - F_k) + (M_{k,2} - F_S - F_k) + ... \\ &+ (M_{k,m} - F_S - F_k) = \sum_{j=1}^{m} (M_{k,j} - F_S - F_k) \end{aligned} . \qquad (2.14)$$

З (2.1) і (2.2) отримуємо:

$$M - F_S = \sum_{k=1}^{n} \left(\sum_{j=1}^{m} (M_{k,j} - F_S - F_k) \right), \qquad (2.15)$$

$$I = I_1 + I_2 + ... + I_n = \sum_{k=1}^{n} I_k . \qquad (2.16)$$

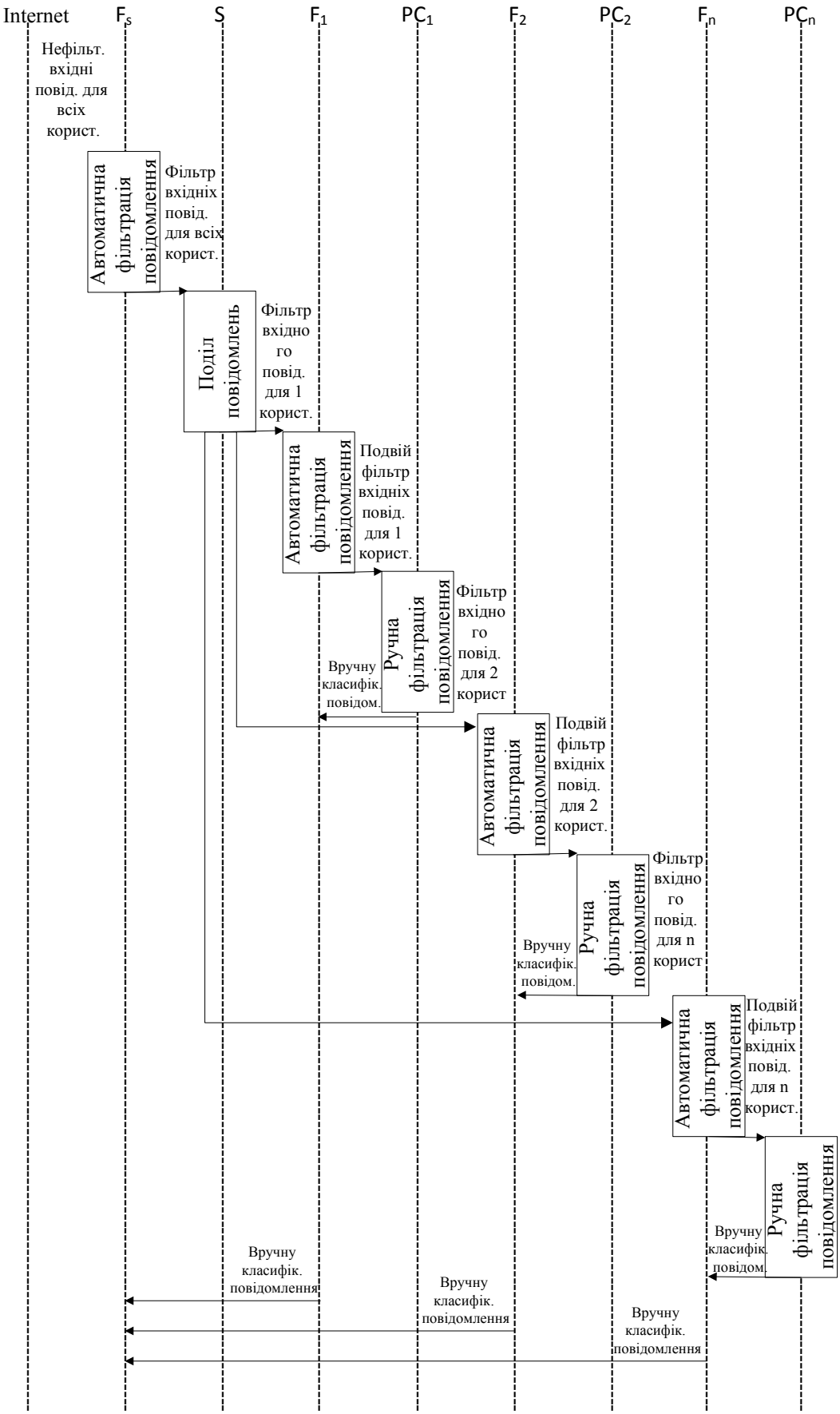


Рисунок 2.11 – Діаграма взаємодії багаторівневої системи боротьби зі спамом

$$I_k = I_{k,1} + I_{k,2} + ... + I_{k,n} = \sum_{j=1}^{m} I_{k,j} \,. \tag{2.17}$$

З (2.16) і (2.17) отримуємо:

$$I = \sum_{k=1}^{n} \left(\sum_{j=1}^{m} I_{k,j} \right). \tag{2.18}$$

У більш загальному випадку відправник електронного повідомлення може вказати в якості одержувачів двох і більше користувачів поштової системи, тому маємо:

$$M - F_S = (M_1 - F_S) \cup (M_2 - F_S) \cup ... \cup (M_n - F_S) = \bigcup_{k=1}^{n} (M_k - F_S), \tag{2.19}$$

$$\begin{aligned} M - F_S = (M_{k,1} - F_S - F_k) \cup (M_{k,2} - F_S - F_k) \cup ... \\ \cup (M_{k,m} - F_S - F_k) = \bigcup_{j=1}^{m} (M_{k,j} - F_S - F_k) \end{aligned} \,. \tag{2.20}$$

З (2.7) і (2.8) отримуємо:

$$M - F_S = \bigcup_{k=1}^{n} \left(\bigcup_{j=1}^{m} (M_{k,j} - F_S - F_k) \right), \tag{2.21}$$

$$I = I_1 \cap I_2 \cap ... \cap I_m = \bigcap_{j=1}^{n} I_k \,, \tag{2.22}$$

$$I_k = I_{k,1} \cup I_{k,2} \cup ... \cup I_{k,m} = \bigcup_{j=1}^{m} I_{k,j} \,. \tag{2.23}$$

З (2.10) і (2.11) отримуємо:

$$I = \bigcap_{k=1}^{n} \left(\bigcup_{j=1, j \neq k}^{m} I_{k,j} \right), \tag{2.24}$$

де M – потік вхідних електронних листів, призначених якому-небудь окремому користувачеві або всім користувачам;

S – поштовий сервер;

F_S – серверний класифікатор;

PC_k – персональний комп'ютер k-го користувача;

F_n – класифікатор підрозділе, налаштований і навчений користувачами n-го підрозділу;

$M_n - F_S$ – потік вхідних електронних листів, призначених користувачам n-го підрозділу і відфільтрованих серверним класифікатором F_S;

$M_n - F_S - F_n$ – потік вхідних електронних листів, призначених m-му користувачу з n-го підрозділу і відфільтрованих спочатку серверним класифікатором F_S, а потім фільтром n-го підрозділу F_n;

I_{nm} – повідомлення, вручну класифіковані m-м користувачами з n-го підрозділу і призначені для навчання класифікатора n-го підрозділу F_n;

I_n – повідомлення, вручну класифіковані всіма користувачем n-го підрозділу і призначенні для навчання серверного класифікатора F_S.

2.2 Алгоритм заповнення бази знань системи фільтрації

Алгоритм заповнення БЗ окремого користувача представлений в додатку А. На початку роботи користувачеві необхідно вказати папки з текстовими і табличними документами, які система може використовувати для поповнення БЗ корисних повідомлень [67]. Потім система приступає до формування БЗ. Кожен

документ розглядається окремо з метою економії обчислювальних ресурсів користувальницького комп'ютера. Після побудови індексу одного документа система переходить до побудови індексу наступного документа і так до тих пір, поки не будуть побудовані індекси всіх документів. В процесі побудови індексу документа системі необхідно виділити лексичні одиниці, які в даному випадку є словами, і підрахувати частоту зустрічання виділених слів в межах аналізованого документа. При виділенні окремого слова з аналізованого документа система відкидає його закінчення, щоб об'єднати слова, які вжиті в різних відмінках [66].

Можна було б створити словники, в яких вказати всі корені слів і все вживаються з кожним із слів закінчень, однак пошук по даній базі зайняв би надто тривалий час, тому система буде оцінювати величину завершення виходячи з довжини слова: чим довше слово, тим більше з кінця символів відкидається.

Для підрахунку частоти слів, що зустрічається в межах розглянутого документа всі слова попередньо упорядковуються за алфавітом, а потім подраховуються повтори кожного з слів.

2.3 Алгоритм фільтрації документа

Узагальнено, робота системи складається з трьох етапів (рисунок 2.12): автоматична класифікація надісланого поштового повідомлення, ручна класифікація вхідного електронного листа і навчання системи.

Якщо після автоматичної класифікації надійшло повідомлення впевненість системи в приналежності повідомлення одному з класів перевищує певний поріг, заздалегідь заданий користувачем, то повідомлення використовується для навчання системи. Якщо ж система не впевнена в класі повідомлення, то вона перед тим, як навчитися на даному повідомленні, чекає, поки користувач вручну не вкаже клас повідомлення.

Рисунок 2.12 – Граф станів багаторівневої системи боротьби зі спамом

Процедура автоматичної фільтрації спам-повідомлень проводиться на сервері з метою зниження обчислювальної навантаження на клієнтські комп'ютери (в тому числі і "тонкі"), а також економії пропускної здатності мережі. Алгоритм роботи системи фільтрації спаму в режимі перевірки зображений в додатку Б.

На першому етапі перевірки лист перевіряється за формальними ознаками, заданим користувачем, наприклад:

1. Чи входить адресу відправника в чорні або білі списки, задані самим користувачем і в публічні чорні списки.

2. Наявність в темі повідомлення певних слів.

3. Наявність в полі «Кому» адреси користувача (часто спамери в поле «Кому» вказують тільки одну адресу, а адреси інших одержувачів спаму заносять в полі «прихована копія», тому що більшість фільтрів блокують листи, адресовані великій кількості одержувачів).

У разі відповідності електронного повідомлення одному або декільком формальним ознаками користувач може задати такі дії системи фільтрації спаму: лист блокується або пропускається без подальшої перевірки іншими способами, а

його текст включається в БЗ системи фільтрації користуватися спаму і використаний для навчання системи; ймовірність нелегітивності електронного листа збільшується або зменшується на певну користувачем величину, після чого лист перевіряється іншими способами.

На другому етапі система перевіряє, чи реально існує домен, з якого лист був нібито надіслано (іноді спамери вказують неіснуючі домени, справжні ж користувачі ніколи не підробляють зворотний адрес, тому даний метод в окремих випадках вкрай ефективний). Якщо вказаного домену не існує, система збільшує оцінку листа.

На третьому етапі лист аналізується за розробленим унікальним алгоритмом в два кроки, на кожному з яких використовується окрема БЗ.

На першому, більш важливому кроці, при аналізі використовується БЗ, отримана при узагальненні баз знань всіх відділів організації. На другому кроці аналізу надісланого повідомлення використовується БЗ відділу, отримані при об'єднанні баз знань співробітників одного відділу. Ймовірність занесення повідомлення не в ту БЗ при формуванні БЗ всього підприємства зводиться до нуля, оскільки в її формуванні брали участь безліч експертів. Тому оцінка вірогідності приналежності отриманого повідомлення до спаму, отримана на першому кроці, сильніше впливає на підсумкову оцінку повідомлення, ніж оцінка, отримана на другому кроці.

Після того, системи фільтрації спаму оцінила підсумкову вірогідність приналежності електронного листа до спаму, воно може бути: поміщено в папку "Вхідні" (якщо підсумкова оцінка нижче нижнього порогового значення); поміщено в "корзину" (якщо підсумкова оцінка перевищує верхнє порогове значення); поміщено в папку "Сумнівні" (у випадку, коли підсумкова оцінка лежить в інтервалі від нижнього до верхнього порогового значення).

Після того, як користувач відкриє надіслане повідомлення, він може позначити його як "спам" або як "не спам", після чого воно буде використано системою для навчання.

Діаграма варіантів використання клієнтської частини системи інтелектуальной фільтрації спаму наведена на рисунку 2.13. Користувач може або

читати отримані електронні листи, або налаштовувати систему. Режим налаштування системи включає в себе завдання параметрів налаштування системи і ручна класифікація надійшовших повідомлень.

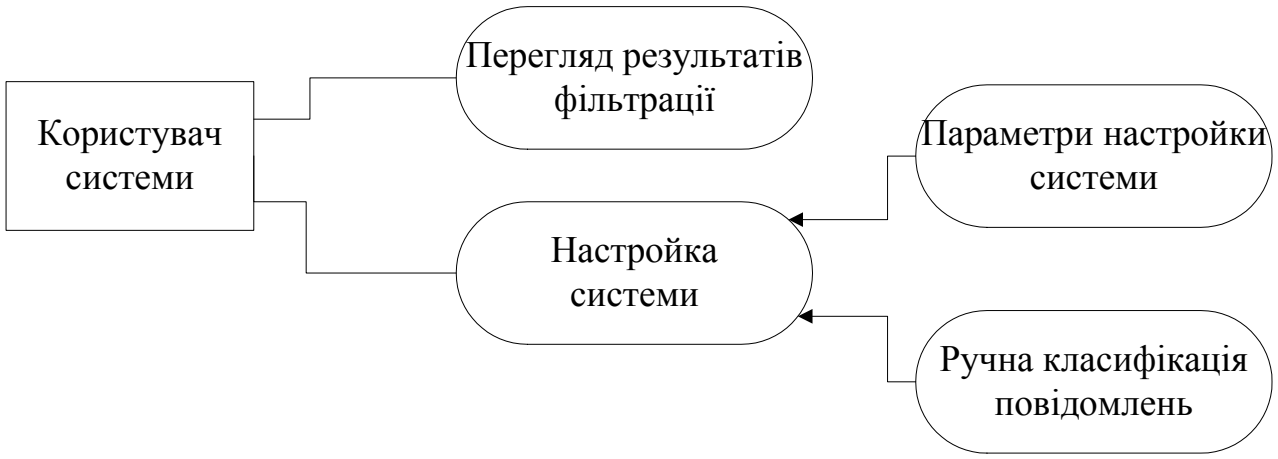


Рисунок 2.13 – UML–діаграма варіантів використання клієнтської частини багаторівневої системи боротьби зі спамом

Користувач може налаштувати:

1. Продовжувати чи ні аналіз в разі виявлення адреси відправника в користувальницькому списку білих адрес.
2. Величину, на яку зменшується оцінка не легітимності листа в разі виявлення адреси відправника в користувальницькому списку білих адрес.
3. Продовжувати чи ні аналіз в разі виявлення адреси відправника в користувальницькому списку чорних адрес.
4. Величину, на яку збільшується оцінка не легітимності листа в разі виявлення адреси відправника в користувальницькому списку чорних адрес.
5. Продовжувати чи ні аналіз в разі виявлення адреси відправника в публічному списку чорних адрес.
6. Величину, на яку збільшується оцінка не легітимності листа в разі виявлення адреси відправника в публічному списку чорних адрес.
7. Продовжувати чи ні аналіз в разі виявлення в темі повідомлення слів або фраз, визначених користувачем.
8. Величину, на яку зменшується оцінка не легітимності листа в разі

виявлення в темі повідомлення слів або фраз, визначених користувачем.

9. Продовжувати чи ні аналіз в разі відсутності адреси користувача в полі "Кому" або "Копія".

10. Величину, на яку збільшується оцінка не легітимності листа в разі відсутності адреси користувача в полі "Кому" або "Копія".

11. Продовжувати чи ні аналіз у випадку, якщо домен відправника не існує.

12. Величину, на яку збільшується оцінка не легітимності листа у випадку, якщо домен відправника не існує.

13. Величину не легітимності, при досягненні якої електронного листа поміщають в папку "Сумнівні".

14. Величину не легітимності, при досягненні якої електронного листа поміщають в папку "Спам".

Функціональна схема клієнтської частини програмного продукту зображена на рисунку 2.14.

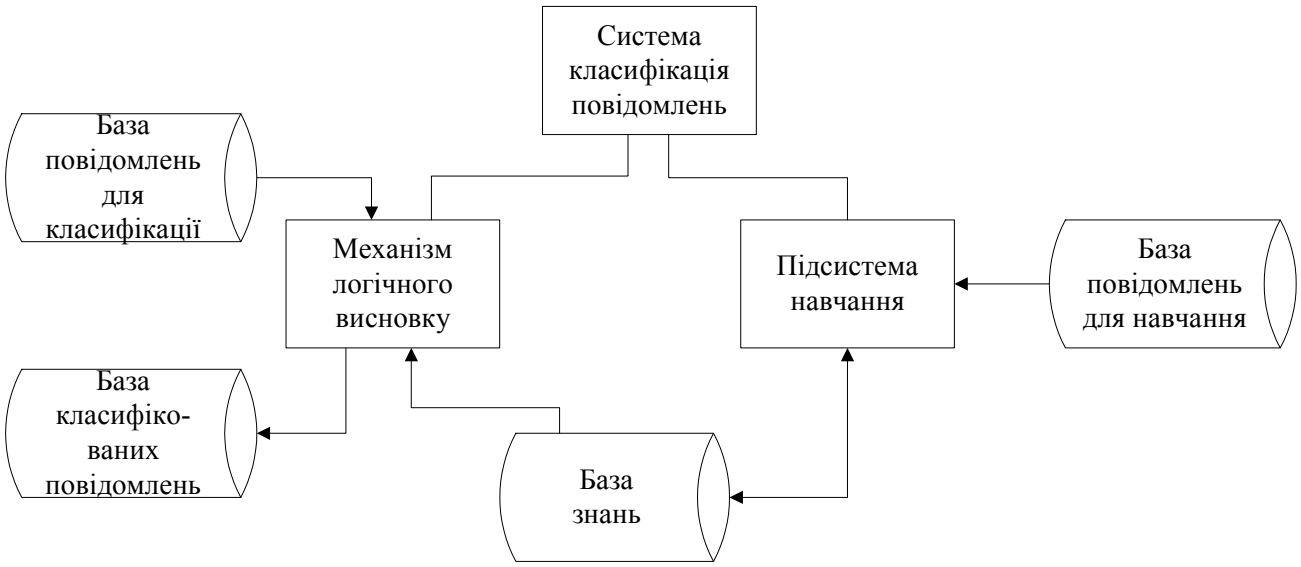


Рисунок 2.14 – Функціональна схема клієнтської частини багаторівневої системи боротьби зі спамом

Клієнтська частина системи класифікації електронних повідомлень складається з наступних модулів: модуль механізму логічного висновку; модуль підсистеми навчання; модуль БЗ; модуль бази класифікуються повідомлень; модуль бази класифікованих повідомлень; модуль бази повідомлень для навчання.

Механізм логічного висновку завантажує електронного листа з бази класифікуються повідомлень, запитує з БЗ необхідну для класифікації інформацію, виробляє класифікацію електронного послання, присвоює йому оцінку не легітимності і поміщає бази класифікованих повідомлень. Схема механізму логічного висновку приведена на рисунку 2.15. Підсистема навчання завантажує нове електронне повідомлення з бази повідомлень для класифікації, потім будується матриця суміжності для завантаженого повідомлення, після чого розраховується ймовірність не легітимності завантаженого повідомлення і визначається клас електронного повідомлення: "спам", "не спам" або "підозріле".

Рисунок 2.15 – Схема механізму логічного висновку

Після того, як користувач вручну перевірить коректність визначення класу завантаженого електронного повідомлення, лист використовується для навчання системи і поповнює БЗ.

2.4 Нейромережевий класифікатор спамових повідомлень

Проблема автоматичного розуміння тексту в тій чи іншій мірі вирішена в багатьох класах систем, в кожному з яких використовуються власні методи [68]: системи машинного перекладу: Promt, Socrat та ін.; система автоматичного індексування: Yandex Desktop Search, Google Desktop Search та ін.; системи аналізу документів: Searchlnform та ін.

Задача представлення інформації, що міститься в текстовому повідомленні є нетривіального внаслідок неоднозначності зв'язку між словами, а також великого числа значень, яке може приймати одне і теж слово залежно від контексту його використання в реченні [69]. В англійській мові слова в реченні жорстко впорядковані, внаслідок чого представити інформацію, виражену англійською мовою набагато простіше, ніж інформацію, виражену українською або російською мовами, оскільки в українській або російській мовах значення слова залежить від того, яке воно займає місце в реченні і від змісту цілого абзацу [70]. У більшості відомих систем будується інформаціоно-лінгвістична модель, в процесі побудови якої з'єднуються лінгвістичні механізми розуміння, що прагнуть до точності і зберігають тотожність при будь-яких перетвореннях, з інформаційними механізмами, які усувають відомості, непотрібні користувачеві. Процес аналізу закінчується побудовою багатовимірної структури, що відображає різні варіанти прочитання заданого фрагмента тексту.

Стандартна система автоматичного розуміння тексту є достатньо складною і громіздкою структурою. Вона включає [68]: лінгвістичні знання, в т.ч. словники і граматики; спеціальні знання в предметній області, в т.ч. структури ситуацій і тезауруси. Процес автоматичного розуміння тексту досить тривалий за часом і вимогливий до обчислювальних ресурсів. Він включає наступні етапи [69]: досинтаксичний аналіз, в т.ч. первинний аналіз тексту і морфологічний аналіз; синтаксичний аналіз; семантичний аналіз, в т.ч. локальний семантичний аналіз і глобальний семантичний аналіз. Етап первинного аналізу тексту включає наступні підетапи [70]: структуризація, сегментація, графоматичний аналіз.

В процесі структуризації кожен текстовий документ, вибраний для аналізу, оголошується самостійним і забезпечується ідентифікатором. В процесі сегментації текстовий документ ділиться на фрагменти, кожен з яких несе закінчене смислове навантаження, наприклад речення або цілий абзац. В процесі графоматичного аналізу кожен текстовий фрагмент ділиться на мінімальні лінгвістично значущі елементи, наприклад, слова або фразеологізми. Результатом роботи первинного аналізу текстового документа є добре структурований масив, який служить основою для подальшого аналізу тексту.

Метою морфологічного аналізу є для кожного мінімального лінгвістично значущого елементу, виділеного на етапі графоматичного аналізу, отримання повної морфологічної характеристики і основної форми слова [71]. Результатом роботи морфологічного аналізу є морфологічне представлення тексту. Існують два підходи до проведення морфологічного аналізу тексту [72]: словниковий, в т.ч. з використанням словника словоформ і з використанням словника основ; безсловниковий.

Використання словника основ при проведенні морфологічного аналізу дозволяє економічно витрачати пам'ять, проте словники словоформ простіші і універсальніші. При безсловниковому методі задаються тільки можливі закінчення, що дозволяє проводити аналіз набагато швидше, хоча і не так достовірно, як при словниковому морфологічному аналізі.

Результатом синтаксичного аналізу текстового фрагмента є синтаксична структура або синтаксичне представлення для кожного речення [73]. Синтаксичний аналізатор є найважливішим, але в той же час, і найскладнішим компонентом систем автоматичного розуміння тексту [74].

Семантичний компонент є головним компонентом систем автоматичного розуміння тексту, він повинен погоджувати три мови [121]: мова лінгвістичних структур, побудованих на попередніх етапах аналізу тексту; мова предметної області, до якої відноситься текстовий документ; мова користувача, для якого текстовий документ і повинен бути проаналізований.

Результатом локального семантичного аналізу є семантичне представлення речення, яка не співпадає з синтаксичним представленням того ж самого речення.

Локальний семантичний аналіз включає наступні етапи [76]: пряма інтерпретація одиниць синтаксичного представлення речення; аналіз сильних лексичних зв'язків; інтерпретація слабких зв'язків; інтерпретація всіх одиниць простих висловів як елементів ситуаційного представлення.

В процесі глобального семантичного аналізу будуються вузли структури з матеріалу різних речень [77]. Результатом глобального семантичного аналізу є єдиний зв'язаний граф, зіставлений цілому текстовому документу. Часто в процесі глобального семантичного аналізу проводиться стиснення лексичного матеріалу для видалення з нього надмірної інформації.

Таким чином, стають очевидними недоліки класичних систем автоматичного розуміння тексту: високе обчислювальне навантаження на комп'ютер користувача, великий об'єм попередньої роботи, великий об'єм службових даних, складність реалізації, достатньо великий розмір аналізованого текстового фрагмента.

Враховуючи описані вище недоліки класичних систем автоматичного розуміння тексту, в якості моделі пропонується використовувати представлення мінімальної семантичної одиниці – речення у вигляді спрощеного семантичного графа (рисунок 2.16):

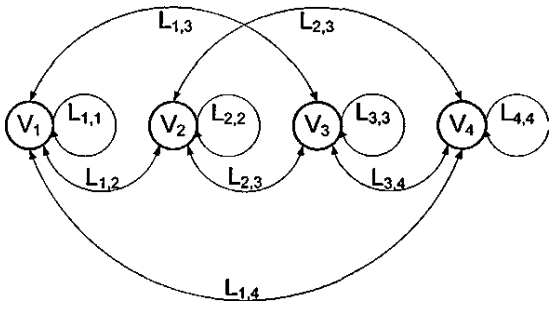


Рисунок 2.16 – Спрощений семантичний граф

де V – леми речення;

L – норма близькості.

Дане представлення дозволяє швидко і при мінімальних обчислювальних витратах побудувати просту структуру, яка достатньо повно відображає суть текстового повідомлення. При повнішому аналізі обчислювальні навантаження і

тимчасові витрати різко зростають, а інформативність моделі представлення збільшується все повільніше і повільніше.

Представлення мінімальної семантичної одиниці – речення у вигляді спрощеного семантичного графа є наглядним для людини, але незручним для штучної нейронної мережі, тому виникла задача розробки представлення спрощеної семантичної моделі повідомлень.

Представленням, яке найбільше відповідає для аналізу за допомогою штучних нейронних мереж була вибрана таблиця, в якій заголовками рядків і стовпців були леми, на головній діагоналі були представлена частота повторення кожної леми у всьому аналізованому текстовому документі, в інших комірках знаходилася частота повторення двох лем в межах одного повідомлення (таблиця 2.1). Наприклад, лема *a* зустрічалася у всьому текстовому документі 4 рази, лема *b* зустрічається у всьому текстовому документі 6 разів, а в рамках одного речення обидві леми *a* і *b* зустрічаються 3 рази.

Таблиця 2.1 – Приклад семантичної моделі повідомлення

	a	b	C	d	e	f	g	h	i	j	k	l	m	N	o	p	q	r	s	t
a	4	3	1	0	2	0	2	0	0	0	1	0	1	0	0	1	0	1	0	0
b	3	6	0	3	0	0	5	4	0	0	5	0	0	2	0	0	0	0	0	0
c	1	0	5	0	4	0	0	0	0	0	4	0	0	0	0	4	0	0	3	0
d	0	3	0	4	3	0	3	0	3	3	2	3	2	1	0	0	0	2	3	0
e	2	0	4	3	6	0	0	3	0	0	5	3	0	0	0	0	0	0	5	0
f	0	0	0	0	0	7	0	0	0	3	0	0	0	2	0	0	2	0	6	0
g	2	5	0	3	0	0	6	5	3	0	1	0	0	0	3	0	0	0	0	0
h	0	4	0	0	3	0	5	9	3	0	6	0	4	0	0	0	3	0	0	0
i	0	0	0	3	0	0	3	3	4	0	0	0	0	3	2	0	0	0	0	1
j	0	0	0	3	0	3	0	0	0	5	4	2	0	0	0	0	2	0	2	0
k	1	5	4	2	5	0	1	6	0	4	7	0	5	2	0	0	3	0	4	0
l	0	0	0	3	3	0	0	0	0	2	0	4	0	0	3	0	0	0	3	0
m	1	0	0	2	0	0	0	4	0	0	5	0	6	0	3	0	0	0	0	0
n	0	2	0	1	0	2	0	0	3	0	2	0	0	4	0	0	2	0	0	2
o	0	0	0	0	0	0	3	0	2	0	0	3	3	0	4	3	0	0	0	0
p	1	0	4	0	0	0	0	0	0	0	0	0	0	0	3	5	0	3	0	0
q	0	0	0	0	0	2	0	3	0	2	3	0	0	2	0	0	4	3	1	0
r	1	0	0	2	0	0	0	0	0	0	0	0	0	0	0	3	3	6	5	0
s	0	0	3	3	5	6	0	0	0	2	4	3	0	0	0	0	1	5	7	0
t	0	0	0	0	0	0	0	0	1	0	0	0	0	2	0	0	0	0	0	3

Лінійна асоціативна мережа (лінійний асоціатор – ЛАМ) представлена на рисунку 2.17.

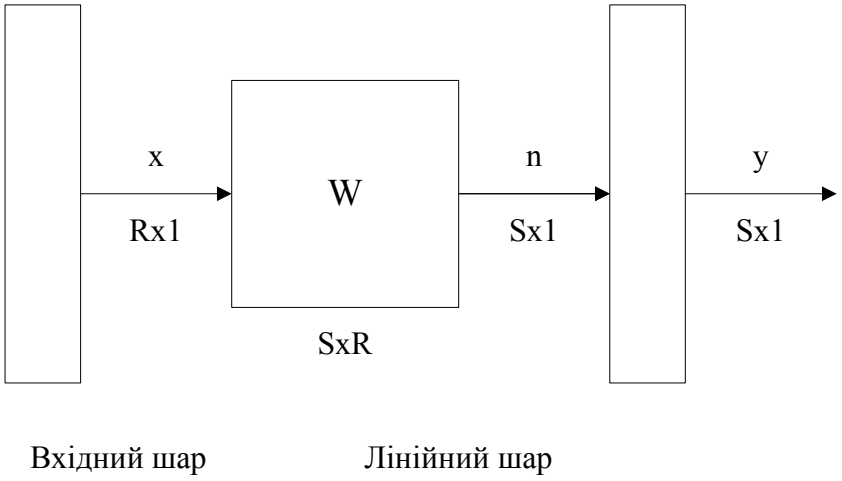


Рисунок 2.17 – Структура лінійного асоціатора

де R – розмірність вхідного вектора V;

S – розмірність вихідного вектора.

Вихідне значення для ЛАМ:

$$a_i = \sum_{j=1}^{R} w_{ij} p_j , \tag{2.25}$$

де $i = 1 \div S,\ j = 1 \div R$

Навчальна вибірка:

$$\{p_1, t_1\}, \{p_2, t_2\}, ..., \{p_Q, t_Q\}, \tag{2.26}$$

де p_i – вхідний образ;

$i = 1 \div Q$;

t_j – вихідний образ;

$j = 1 \div Q$.

Правило навчання Хебба:

$$w_{ij}^{new} = w_{ij}^{old} + \alpha f_i(a_{iq}) g_j(p_{jq}), \tag{2.27}$$

де a_{iq} – вихідний сигнал;

p_{jq} – вхідний сигнал;

α – коефіцієнт.

Спрощена матрична форма при коефіцієнті $\alpha = 1$:

$$W^{new} = W^{old} + t_q p_q^T. \tag{2.28}$$

Пакетна форма навчання:

$$W = t_1 p_1^T + t_2 p_2^T + ... + t_Q p_Q^T = \sum_{q=1}^{Q} t_q p_q^T. \tag{2.29}$$

Матрична форма представлення (2.29):

$$W = \left[t_1 t_2 ... t_Q\right] \begin{bmatrix} p_1^T \\ p_2^T \\ ... \\ p_Q^T \end{bmatrix} = TP^T, \tag{2.30}$$

де $P = \left[p_1 p_2 ... p_Q\right]$ – вектор вхідних сигналів;

$T = \left[t_1 t_2 ... t_Q\right]$ – вектор вихідних сигналів.

З врахуванням цього на виході маємо:

$$a = Wp_k = \left(\sum_{q=1}^{Q} t_q p_q^T \right) p_k = \sum_{q=1}^{Q} t_q (p_q^T p_k) \tag{2.31}$$

1) Випадок ортогональних вхідних векторів:

$$(p_q^T p_k) = 1, q = k, \tag{2.32}$$

$$(p_q^T p_k) = 0, q \neq k, \tag{2.33}$$

На виході ЛАМ отримує бажаний результат:

$$a = Wp_k = t_k \tag{2.34}$$

2) Вхідні вектора нормалізовані, але не ортогональні:

$$a = Wp_k = t_k + \sum_{q \neq k} t_q (p_q^T p_k) = t_k + \varepsilon, \tag{2.35}$$

Під нормалізацією розуміємо:

$$V_i = \frac{V_i}{\|V_i\|} = \frac{V_i}{\sqrt{V_1^2 + V_2^2 + \ldots}}. \tag{2.36}$$

Появляється помилка, яка пов’язана з величиною кореляції вхідних векторів.

Навчання на основі псевдо обернення:

$$Wp_q = t_q, q = 1,2,\ldots,Q \tag{2.37}$$

У випадку, якщо не виконується (2.37), то використовується (2.38) в якості критерію якості:

$$F(W) = \sum_{q=1}^{Q} \left\| t_q - W p_q \right\|^2 \to \min \tag{2.38}$$

Матрична форма (2.37):

$$WP = T\,, \tag{2.39}$$

$$T = \left[t_1 t_2 ... t_Q \right],\ P = \left[p_1 p_2 ... p_Q \right]. \tag{2.40}$$

Тоді необхідно мінімізувати:

$$F(W) = \left\| T - WP \right\|^2 = \left\| E \right\|^2, \tag{2.41}$$

де $\left\| E \right\|^2 = \sum_i \sum_j e_{ij}^2$.

Нехай дано:

$$WP=T. \tag{2.42}$$

Необхідно мінімізувати:

$$F(W) = \left\| T - WP \right\|^2 = \left\| E \right\|^2, \tag{2.43}$$

якщо зворотна матриця для P існує, $F(W) \to 0$

$$W = TP^{-1}. \tag{2.44}$$

Якщо немає зворотної матриці для P, то можна використовувати псевдоінверсію:

$$W = TP^{+}, \tag{2.45}$$

$$P^{+} = (P^{T}P)^{-1}P^{T}. \tag{2.46}$$

Зв'язок з правилом Хебба:

$$W = TP^{T}. \tag{2.47}$$

Правило псевдозвертання:

$$W = TP^{+}, \tag{2.48}$$

$$P^{+} = (P^{T}P)^{-1}P^{T}. \tag{2.49}$$

Якщо вхідні вектора ортогональні, то

$$P^{T}P = I, \tag{2.50}$$

де I – одинична матриця, тоді:

$$P^{+} = (P^{T}P)^{-1}P^{T} = P^{T}. \tag{2.49}$$

Отже , у разі ортогональності векторів з навчальної вибірки, точність лінійного класифікатора зростає.

Як наголошувалося раніше, при вирішенні задачі класифікації спаму в рамках відомих підходів використовується синтаксичний аналіз при формуванні словників лексичних одиниць: текст d_i розбивається на лексеми, які потім використовуються як атрибути при класифікації. Далі з їх допомогою приймається рішення про приналежність повідомлення до якоїсь категорії. Недоліком таких фільтрів є неможливість врахування семантичної складової

електронних повідомлень.

Розроблені системи витягування знань з великих масивів даних на основі алгоритмів DataMining, що дозволяють витягувати семантичну компоненту, які можуть бути використані при побудові систем інтелектуальної фільтрації спаму [78]. Проте, як відомо, вони відрізняються високою складністю, що утрудняє їх використання для вирішення задачі фільтрації спаму на комп'ютері користувача, так і на сервері поштових повідомлень.

В рамках пропонованого методу при вирішенні задачі побудови інтелектуального класифікатора пропонується оцінювати не тільки частоту появи лексем в повідомленні, але і визначати міру близькості лексем в рамках прийнятої мінімальної семантичної структури повідомлення. За мінімальну семантичну структуру в текстових повідомленнях можна, наприклад, прийняти речення. З врахуванням невеликого об'єму електронних листів це дозволяє витягувати частину семантичної інформації, яка може бути використана в процесі категоризації повідомлень.

Розглянемо основні етапи процесу категоризації повідомлень на основі пропонованого підходу. На першому етапі агентом-класифікатором з отриманого текстового повідомлення формується множина $F=\{f_k\}$, де f_k – задана для системи фільтрації мінімальна семантична структура, $i = 1 \div m$. При цьому із структури видаляються надлишкові елементи, але зберігається послідовність лексем. В результаті структура f_k може бути представлена у вигляді семантичного графа s_k, що визначає зв'язок лексем в f_k. При оцінці сили зв'язку враховується лексикографічний порядок в f_k. Після обробки всього повідомлення отримуємо семантичний граф $G =< S,V >= \bigcup_1^n s_k$, який представляє собою семантичну модель повідомлення, де S – множина вершин, що містить лексеми повідомлення, що пройшли попередню фільтрацію на основі заданого словника; V – множина ребер, вагові значення яких визначають силу зв'язку лексем в f_k. Граф G задається матрицею суміжності Z (семантична матриця). Далі, з врахуванням векторів атрибутів повідомлень на базі G будуються дві узагальнені семантичні матриці Z_s (для спаму) і Z_e (для легітимних повідомлень). В результаті задача класифікації

нових повідомлень зводиться до побудови бінарного класифікатора *N*, що визначає приналежність повідомлення до однієї з представлених семантичними матрицями категорій.

В якості класифікатора вибраний лінійний нейромережевий асоціатор вигляду $y = w^T x$, де x – вхідний вектор (елементи вектора складені з елементів семантичної матриці фільтрованого повідомлення), w – вектор вагів класифікатора (див. рисунок 2.16, де R – розмірність вектора x).

Навчання асоціатора виконується на основі правила навчання Хебба [6]. До переваг цього класифікатора слід віднести простоту його навчання, що є визначальною характеристикою при виборі його в якості ядра системи інтелектуальної фільтрації спаму. На рисунку 2.17 представлена діаграма зміни частки спаму і легітимних повідомлень в процесі навчання багатоагентної системи фільтрації [79].

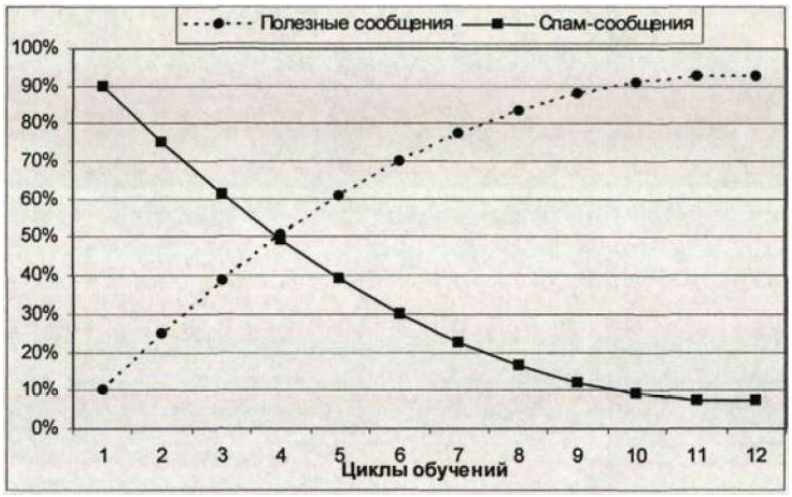


Рисунок 2.17 – Зміна частки спаму і корисних повідомлень

На рисунку 2.17 представлена зміна частки спаму і легітимних повідомлень в процесі навчання системи фільтрації. На рисунку 2.18 приведені результати порівняння ефективності розробленої системи з бейсівським фільтром.

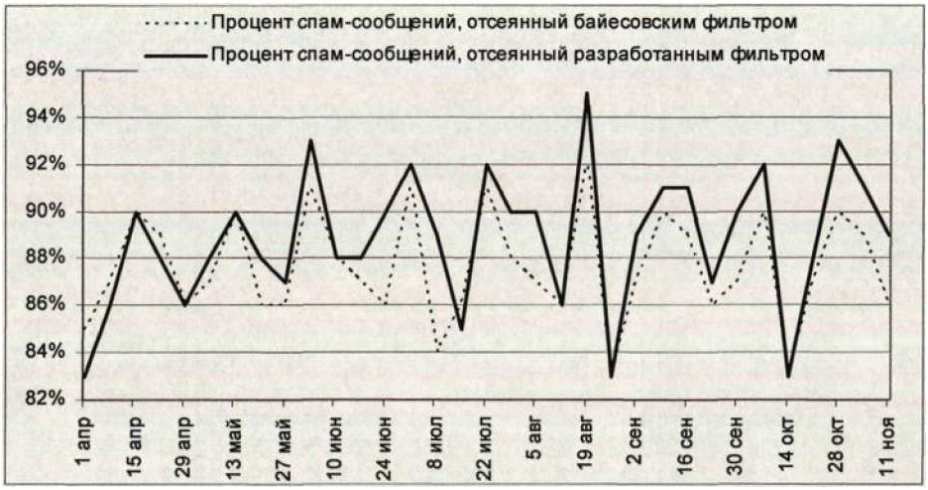


Рисунок 2.18 – Порівняння ефективності роботи фільтрів

У другому розділі в рамках запропонованої концепції системи фільтрації спаму запропоновано комбінований ієрархічний алгоритм формування БЗ, що дозволяє сформувати одночасно повну і достовірну БЗ системи фільтрації спаму. Розроблена архітектура ієрархічної системи фільтрації, яка дозволяє підвищити ефективність фільтрації за рахунок використання баз знань, різних за повнотою та достовірності на різних рівнях ієрархії ЛОМ. Розроблено алгоритм багатоетапної класифікації повідомлень, який дозволяє підвищити достовірність класифікації та знизити час затримки. Запропонована семантична модель електронного повідомлення, заснована на врахуванні лексикографічного впорядкування лексем у вибраній мінімальній семантичній одиниці повідомлення: речення, абзац, повідомлення. Відмітною особливістю запропонованої моделі є її простота і можливість використання її в системах фільтрації реального часу. Запропоновано представлення БЗ системи фільтрації спам–повідомлень у вигляді семантичної матриці, діагональні елементи якої зберігають частоту появи лексеми в повідомленнях, що належать тому або іншому класу, а недіагональні елементи відображають зв'язок лексем у вибраній семантичній одиниці електронного повідомлення. Запропонований алгоритм класифікації повідомлень на основі семантичної матриці і нейронних мереж. В якості нейронної мережі вибраний лінійний асоціатор, що дозволяє забезпечити навчання нейронної мережі в реальному масштабі часу. Результати експериментів показали достатньо високу ефективність вирішення задачі класифікації електронних повідомлень.

3 РЕАЛІЗАЦІЇ СИСТЕМИ ВИЯВЛЕННЯ СПАМОВИХ ПОВІДОМЛЕНЬ

3.1 Архітектура системи виявлення спамових повідомлень

Для програмної реалізації розробленої системи фільтрації спам–повідомлень у потоці електронної кореспонденції була використана багатоагентна технологій (кожен агент є компонентом системи боротьби зі спамом різного рівня). Концепція багатоагентних систем, що припускає наявність в достатній мірі незалежних сутностей, здатних до комунікації і складної взаємодії один з одним, найкраще відповідає вимогам до розробленої системі боротьби зі спамом.

Пропонована система боротьби зі спамом на підприємстві є розподіленою, кожен системний компонент, керуючий поштовими фільтрами, має лише частину інформації, необхідної для вирішення проблеми, і може впливати на рішення проблеми тільки на своїй ділянці.

Це обумовлено складною специфікою, гетерогенної, розподіленої в просторі і непостійною за структурою системи, якою є інформаційна система сучасного підприємства. Побудова системи боротьби зі спамом є хорошим прикладом задачі, ефективно розв'язуваної за допомогою багатоагентної системи. Повне рішення даного завдання забезпечується за рахунок спілкування, взаємодії агентів.

Ключовим моментом, що визначає вибір саме багатоагентної технології, є забезпечення автономності системи. Зміна структури інформаційної системи підприємства, додавання або видалення будь-якого користувача робочої станції з поштовим клієнтом або навіть сервера електронної пошти окремого підрозділу, при вибраній архітектурі тягне за собою зміни і в структурі багатоагентної системи боротьби зі спамом в потоці електронної пошти. Це дозволяє підвищити ефективність роботи адміністратора безпеки локальної обчислювальної мережі підприємства.

Таким чином, програмна реалізація системи фільтрації електронних спам-повідомлень представляє собою мобільну, легкомасштабовану багатоагентну

систему.

Далі наведено опис структури багатоагентної системи боротьби зі спамом і одночасно, процесу її проектування.

При проектуванні багатоагентної системи боротьби зі спамом можна виділити наступні основні етапи:

1. Аналіз цілей і завдань функціонування багатоагентної системи фільтрації спам-повідомлень.

2. Вибір загальної архітектури багатоагентної системи боротьби зі спамом. На даному етапі відбувається визначення структури багатоагентної системи фільтрації спам-повідомлень, зв'язок її з структурою локальної обчислювальної – мережі підприємства та інформаційних потоків (електронної пошти) і типів використаних агентів.

3. Розгляд сценаріїв роботи агентів.

4. Розмежування ролей агентів відповідно до виконуваних ними функціями.

5. Розробка угод про спілкування агентів (створення системи повідомлень).

6. Розробка структури окремих агентів: розробка інформаційної моделі агентів; розробка набору функцій агентів.

7. Вибір платформи для програмної реалізації багатоагентної системи боротьби зі спамом.

8. Програмна реалізація багатоагентної системи фільтрації спам–повідомлень у потоці електронної пошти.

У свою чергу, програмна реалізація системи боротьби зі спамом, також є досить складним завданням, що складається з декількох етапів і вирішування за допомогою певних методів (методології розробки програмного забезпечення, шаблони проектування і т. д.), але в даній роботі наведена розробка тільки прототипу багатоагентної системи боротьби зі спамом з метою ілюстрації розробленого підходу фільтрації електронних поштових повідомлень.

Етапи проектування системи показані на рисунку 3.1

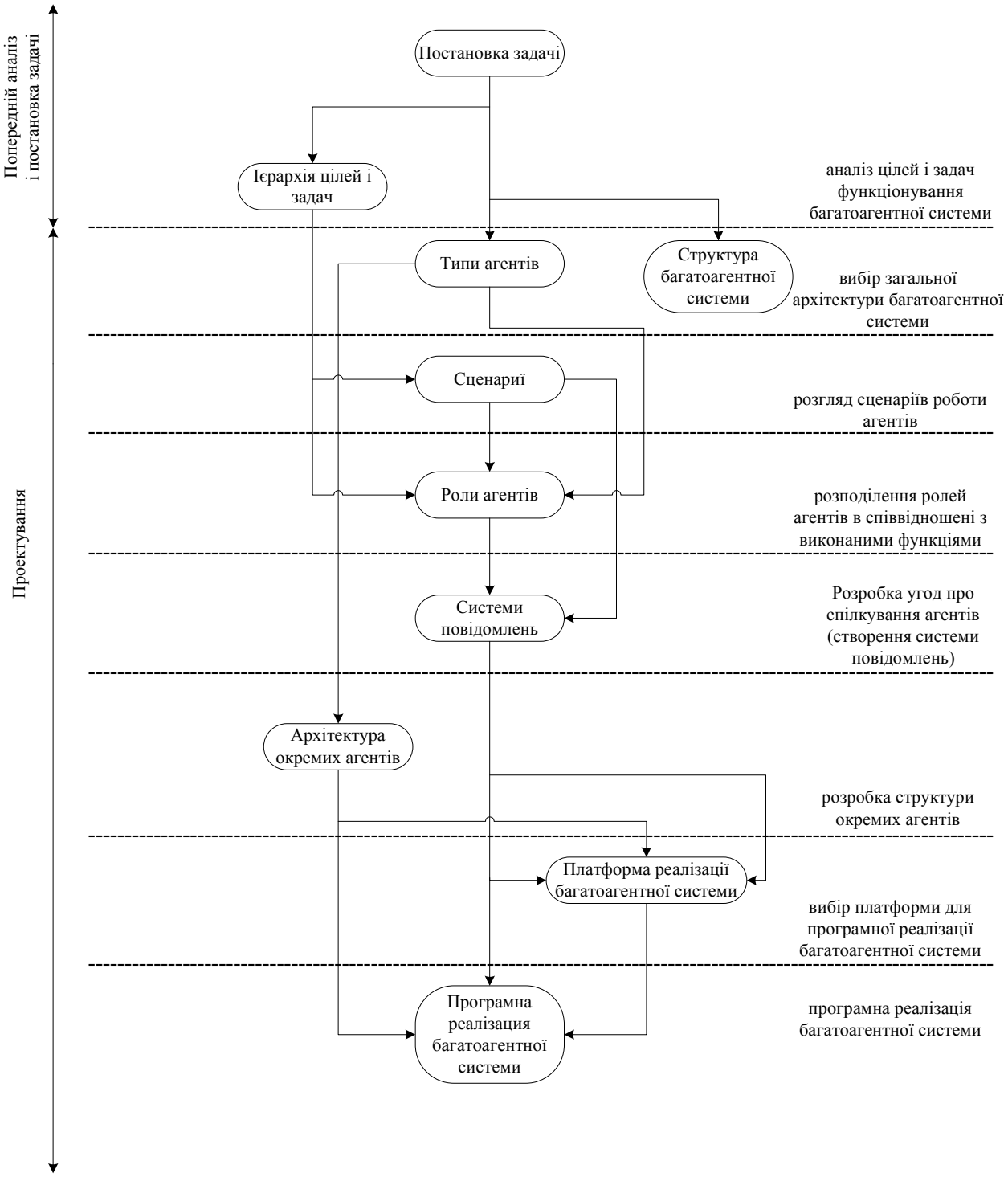


Рисунок 3.1 – Етапи проектування системи

Перший етап проектування – аналіз цілей і завдань функціонування багатоагентної системи – можна вважати вже виконаним в попередніх частинах роботи. Очевидно, що метою функціонування системи є боротьба з небажаною електронною поштою (спамом). Для досягнення цієї мети багатоагентна система (БС) вирішує такі завдання (рисунок 3.2): управління поштовими фільтрами

поштових клієнтів і серверів; поширення критеріїв віднесення електронних листів до спаму з тією чи іншою мірою достовірності всередині багатоагентної системи (самонавчання системи); підтримання відповідності структури багатоагентної системи боротьби зі спамом і структури цікавлячих нас інформаційних потоків (електронної пошти) на підприємстві, за рахунок механізмів клонування (відтворення) та міграції агентів; оперативне поширення налаштувань, заданих адміністратором, по всій багатоагентній системі.

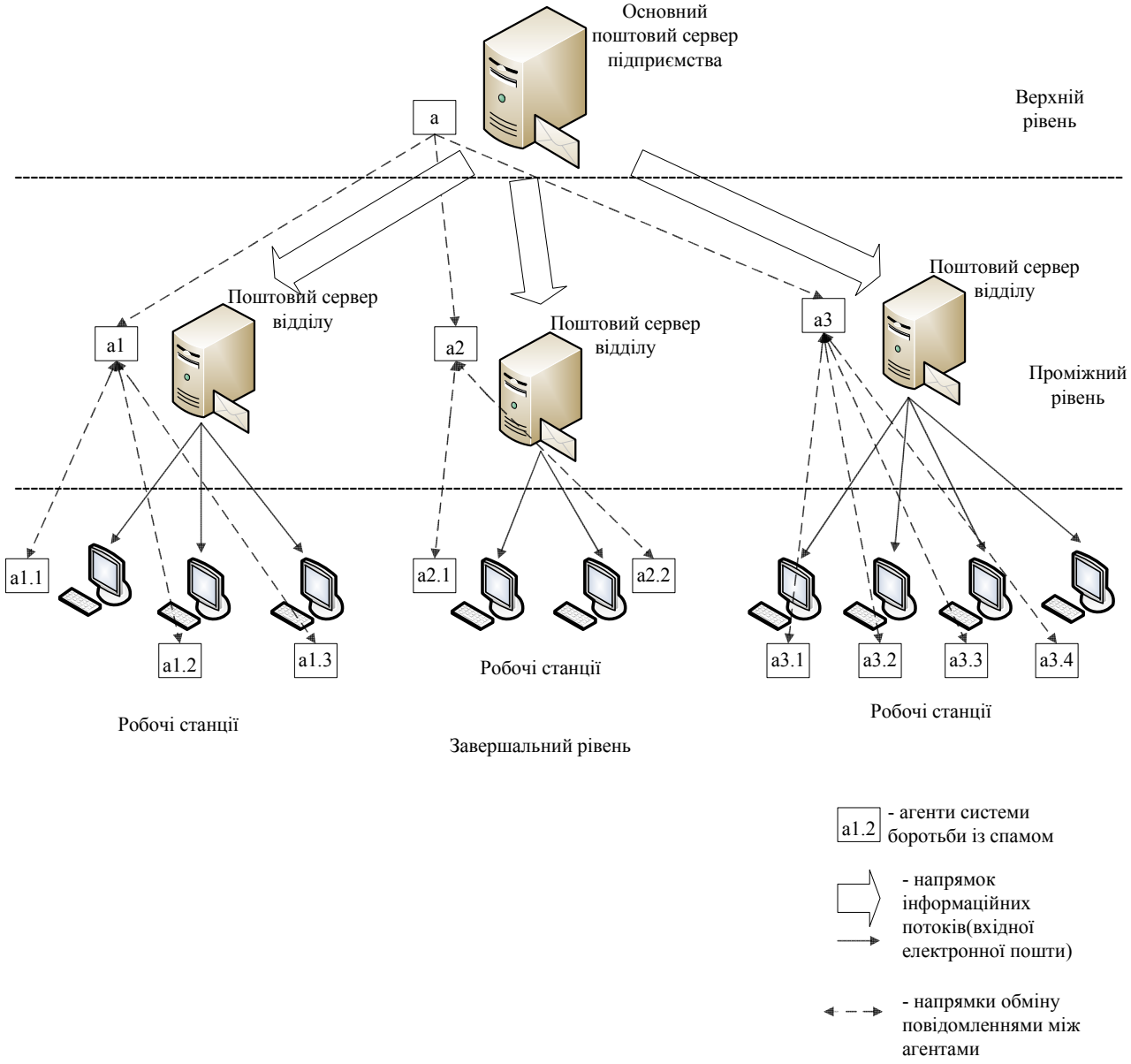


Рисунок 3.2 – Структура багатоагентної системи боротьби зі спамом

Перший етап проектування, що полягає в аналізі цілей і завдань функціонування багатоагентної системи боротьби зі спамом, можна вважати вже

виконаним в попередніх розділах даної роботи. Очевидно, що метою функціонування системи є фільтрація небажаної кореспонденції (спам–повідомлень). Для досягнення цієї мети БС вирішує наступні завдання: управління поштовими фільтрами поштових клієнтів, встановлених на робочих станціях, і серверів електронної пошти; поширення критеріїв віднесення електронних листів до спаму з тим або іншим ступенем достовірності всередині багатоагентної системи (самонавчання системи); підтримання відповідності структури багатоагентної системи боротьби зі спамом і структури інформаційних потоків (електронної кореспонденції) на підприємстві, за рахунок механізмів клонування (відтворення) та міграції агентів на робочих станціях та сервері (рисунок 3.3); оперативне розповсюдження параметрів налаштувань, встановлених адміністратором безпеки локальної обчислювальної мережі по всій багатоагентній системі фільтрації спам-повідомлень.

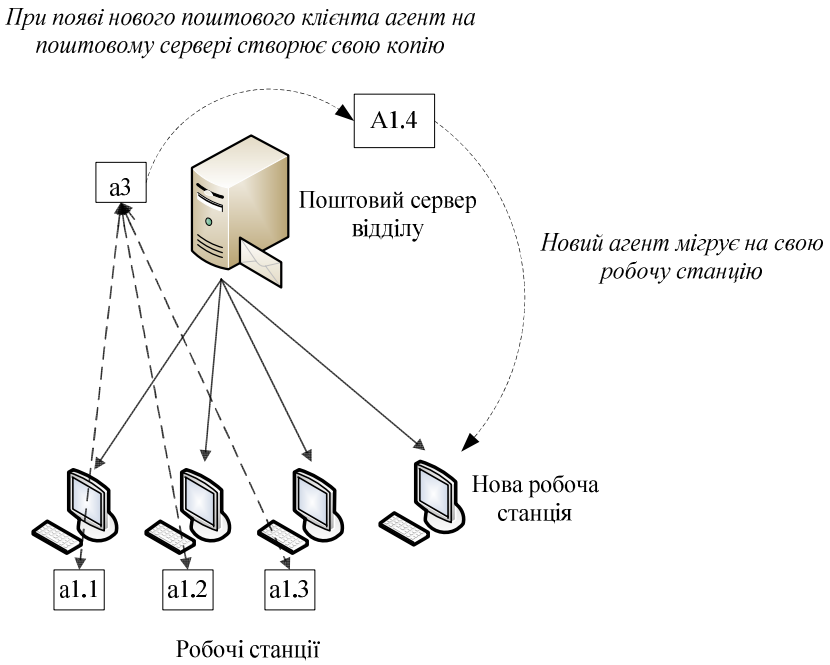


Рисунок 3.3 – Процедура створення нового агента при підключенні до локальної мережі нової робочої станції

Запропонований підхід до розгортання дозволяє системі боротьби зі спамом протягом усього періоду експлуатації мати необхідну для роботи структуру й адаптуватися до змін в інформаційній системі підприємства без втручання

системного адміністратора (за умови функціонування механізмів відстеження змін в структурі інформаційних потоків, що входить електронної пошти).

3.2 Реалізація агентів системи

При роботі агентів системи боротьби зі спамом можна виділити наступні основні сценарії: робота з електронними повідомленнями, що включає в себе управління поштовими фільтрами і самонавчання); поширення змін в параметрах настройках системи фільтрації спам–повідомлень у потоці електронної кореспонденції (зміни виробляються адміністратором безпеки мережі або системним адміністратором); спостереження за змінами в структурі потоків електронної пошти і реагування на них, в тому числі адаптації структури багатоагентної системи боротьби зі спамом.

Діаграма взаємодії в багатоагентній системі боротьби зі спамом в цих сценаріях показана на рисунку 3.4.

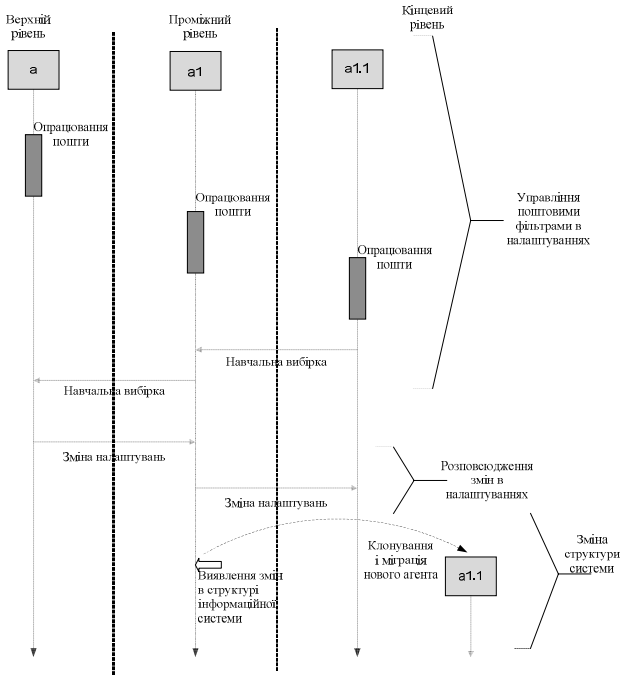


Рисунок 3.4 – Діаграма станів багатоагентної системи боротьбі зі спамом при різних сценаріях

Як вже було сказано, БС боротьби зі спамом має ієрархічну структуру, що складається з трьох рівнів: верхнього, проміжного і кінцевого. На верхньому рівні розташовується основною в організації сервер електронної пошти, на проміжному рівні знаходяться поштові сервера підрозділів, останній рівень займають поштові клієнти, встановлені на робочі станції користувачів. Ролі агентів залежать від місця, займаного ними у цій ієрархії. У різні моменти часу один і той же агент може виконувати різні ролі, в залежності від ситуації, що склалася та архітектури локальної обчислювальної мережі.

Функції кожного агента визначаються його ролями: управління поштовими фільтрами на своєму рівні; відправка навчальних вибірок агенту більш високого рівня (для агентів кінцевого і проміжного рівнів); прийом навчальних вибірок від агентів попереднього рівня, формування БЗ (по алгоритму, що визначається виходячи з місця агента в ієрархії), а також самонавчання (для агентів проміжного і верхнього рівнів); спостереження за структурою інформаційної системи підприємства на своєму рівні, реагування на зміни в цій структурі, виражене в появі і зникненні робочих станцій і серверів електронної пошти (для агентів проміжного і верхнього рівнів); поширення службових повідомлень (наприклад, змін параметрів налаштувань) з верхнього рівня на нижні.

Спілкування між агентами відбувається в двох напрямках: спілкування "знизу вгору" – від кінцевих агентів до агентам верхнього рівня. Дане направлення забезпечує навчання системи боротьби зі спамом; спілкування "зверху вниз" – від агентів на верхньому рівні до крайовим агентам. Цей напрямок використовується для централізованого управління параметрами настройками системи фільтрації спам–повідомлень у потоці електронної пошти.

Система повідомлень в багатоагентній системі боротьби зі спамом має наступну структуру: повідомлення з навчальними вибірками для БЗ; повідомлення, що регулюють параметри настройки в багатоагентній системі; запити про зміни інформаційної структури підприємства (додавання або видалення серверів або робочих станцій, на яких відбувається обробка електронної кореспонденції) і відповіді на них.

Архітектура окремого агента показана на малюнку 3.5.

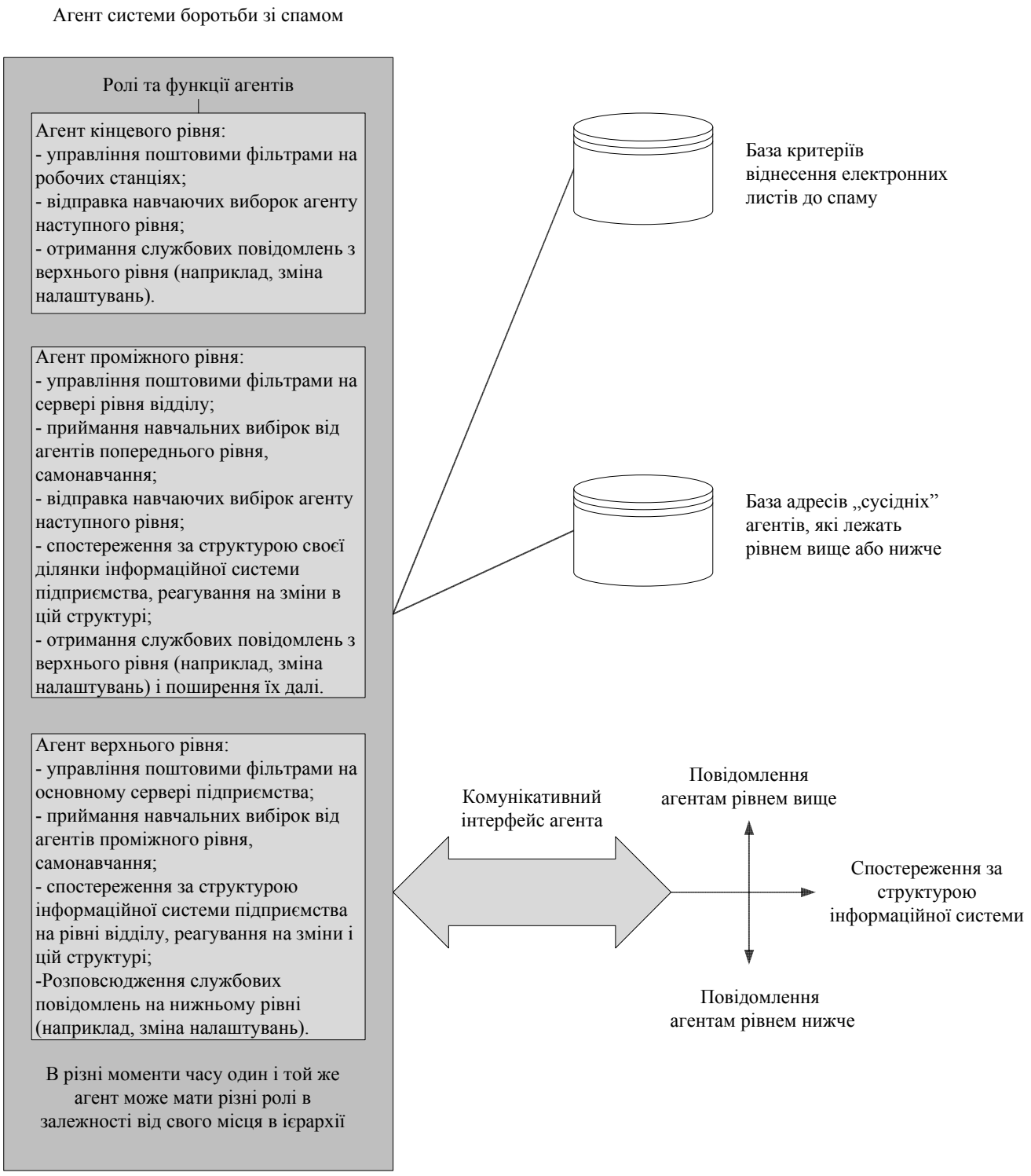


Рисунок 3.5 – Архітектура окремого агента

Інформаційна модель агента включає в себе БЗ з критеріями віднесеними електронних поштових відправлень до одного з класів і базу адресів сусідніх агентів – одного агента-батька, розташованого на рівень вище, і декількох агентів-нащадків, розташованих на рівень нижче (займають «підлегле» положення в ієрархії). Адреси агента-батька, розташованого на рівень вище, задається при

створенні агента (як копії «батька»).

Поведінка агента визначається його поточної роллю [80]. Кожен агент має комунікаційний інтерфейс, який служить для обміну повідомленнями з іншими агентами і запитів про зміни в структурі інформаційної системи до служби каталогів підприємства.

3.3 Програмна реалізація системи

Для ілюстрації пропонованого підходу був розроблений програмний прототип багатоагентної системи на базі платформи Java Agent Development Framework (JADE).

Попередньо був проведений огляд існуючих платформ для розробки багатоагентних систем. Переваги віддавалися безкоштовним засобів з відкритими вихідними кодами, оскільки це дозволяє переконатися у відсутності не документованих можливостей. Як вже було сказано, для розробки була обрана платформа JADE (Java Agent DEvelopment Framework), є програмним забезпеченням середньої ланки (middle-ware) для створення багатоагентних систем різного рівня складності.

Порівняльна характеристика різних засобів розробки наведена нижче (таблиця 3.1). Поряд з платформою Java DEvelopment Framework (JADE) в порівнянні брали участь: Cougaar Agent Architecture [80]; agentTool [81]; MultiAgent Systems Tool [82]; FIPA–OS (Foundation Intelligence Platform Agents–Open Source) [83].

Перевагами платформи JADE є [84]: безкоштовність і відкриті вихідні тексти (ліцензія LGPL – Lesser General Public License); розподілений характер платформи, що забезпечує високий ступінь надійності, живучості і масштабованості розробленої системи; підтримка «легковагових» пристроїв (таких, як мобільні телефони, смартфони, кишенькові персональні комп'ютери тощо); досить високий рівень вбудованих засобів забезпечення безпеки;

підтримка міжнародного набору стандартів на інтелектуальні багатоагентні системи FIPA; можливість використання довільних алгоритмів кодування для передачі повідомлень, що дозволяє здійснювати туннелювання через віртуальні приватні мережі і т. п.; наявність вбудованих засобів адміністрування багатоагентної системи, наявність вбудованих засобів налагодження багатоагентних систем.

Таблиця 3.1 – Порівняння засобів розробки багатоагентних систем

Критерії	Jade	Cougaar	Agent Tool	MultiAgent Systems Tool	FIPA–OS
Продуктивність	висока	висока	низька	висока	середня
Масштабованість	висока	середня	низька	низька	середня
Підтримка мобільності агентів	+	+	–	–	–
Підтримка міжнародних стандартів в області МАС	так	обмежено	ні	обмежено	так
Вбудовані засоби ЗІ	+	+	–	–	–
Вбудовані засоби адміністрування системи	+	–	–	+	+

JADE включає дві основні частини [85]: платформу для запуску та налагодження агентів; прикладну бібліотеку JADE API (мовою Java) для розробки багатоагентних систем.

Той факт, що дана платформа написана на Java, означає також, що вона має всі основні переваги даної технології, а саме кросплатформеність і підвищення надійності, за рахунок гарантій проектування (використання керованого коду, строгий контроль типів і т. д.)

Значна частина сучасних бізнес–додатків експлуатують виртуальну машину

Java або оболонку. NET. Платформа Java фактично стала відгалузевим стандартом.

Основними елементами багатоагентної системи, реалізванной на основі JADE є платформи та контейнери [86].

Платформа може містити кілька розподілених по мережі контейнерів. Один з цих контейнерів є головним (MainContainer), він містить основні службові компоненти платформи JADE, що забезпечують функціонування платформи.

В багатоагентній системі фільтрації спам–повідомлень у потоці електронної пошти використовуються наступні компоненти [87]:

– AMS – Agent Management System (система управління агентами). Являється службовим агентом, що здійснює основний контроль над доступом і використанням платформи. На кожній платформі може бути не більше одного агента AMS. AMS надає сервіси, що забезпечують життєвий цикл агентів, службу "white-pages", що зберігає дані про розташування агентів, забезпечує монтування каталогу ідентифікаторів агентів AID (Agent ID), забезпчує підтримку станів агентів ("запущений", "призупинено" і т. п.);

– DF – Directory Facilitator (служба каталогу). Також є службовим агентом, надає довідкову інформацію, що забезпечує спілкування агентів всередині платформи. При необхідності може бути замінена компонентом аналогічного призначення власної розробки;

– ACC (MTP) – Agent Communication Channel (Message Transport System) – канал спілкування агентів (система передачі повідомлень). Програмний компонент, який контролює весь обмін повідомленнями в рамках платформи, включаючи повідомлення інших платформ.

Всі перераховані вище служби автоматично завантажуються при запуску платформи. Платформа, як вже було сказано, має розподілений характер. На кожному з хостів створюється віртуальна машина Java (JVM – Java Virtual Machine), що забезпечує середовище виконання для агентів відповідного контейнера. На головному контейнері (містить AMS і DF) реєструється служба Java RMI (Remote Methods Invoking). Решта контейнери взаємодіють з головним, використовуючи цю технологію [88].

Агенти, що розробляються на базі JADE, з точки зору програмування, є об'єктами класів-нащадків класу Agent JADE API. У цьому класі реалізовані службові методи, що забезпечують роботу агентів в платформі – реєстрація, конфігурування, віддалене управління і т.д., а також методи, що забезпечують посилку і прийом повідомлень ACL {Agent Communication Language – мова спілкування агентів).

В ході свого життєвого циклу агент може знаходитися в різних станах [89].

INITIATED (ініційований) – об'єкт агента створено, але ще не зареєструвався в AMS, не має імені та адреси і не може спілкуватися з іншими агентами.

ACTIVE (активний) – об'єкт агента зареєструвався в AMS, має коректне ім'я та адресу і має доступ до всіх необхідних можливостей, що надаються платформою (враховуючи, природно, обмеження, що накладаються поточною політикою безпеки).

SUSPENDED (призупинено) – об'єкт агента припинений і не виконує ніяких дій.

WAITING (очікування) – об'єкт агента блокований, чекаючи поки будуть виконані деякі зазначені умови (наприклад, прийде повідомлення і т. д.).

DELETED (вилучено) – агент припинив своє функціонування, його реєстрація в AMS анульована.

TRANSIT (транзит) – мобільні агенти переходять в цей стан, коли мігрують з хоста на хост.

Поведінка агентів визначається призначеними їм класами поведінки. Всі класи поведінки успадковані або від класу *Behaviour* JADE API, або від уже реалізованих його нащадків.

Існують наступні класи діяльності агентів [90]: разові дії; періодичні дії; складні дії, в т.ч. паралельні; послідовні, керовані кінцевим автоматом станів.

Успадкуємо від класу *Agent* новий клас, *AntiSpamAgent*, в методах якого реалізуємо весь практичний функціонал системи боротьби зі спамом, що описаний раніше.

Функції агентів реалізуються за допомогою окремих класів–спадкоємців

класу *Behaviour*. Набір функцій вже був описаний вище.

Для реалізації запропонованого механізму розгортання системи використані механізми клонування та міграції агентів платформи JADE (клас *Agent* надає методи *Agent.migrate()* і *Agent.clone()*).

На рисунку 3.6 показаний вигляд розгорнутого (на віртуальних машинах) прототипу багатоагентної системи боротьби зі спамом в адміністративному інтерфейсі JADE (JADE Remote Agent Management GUI).

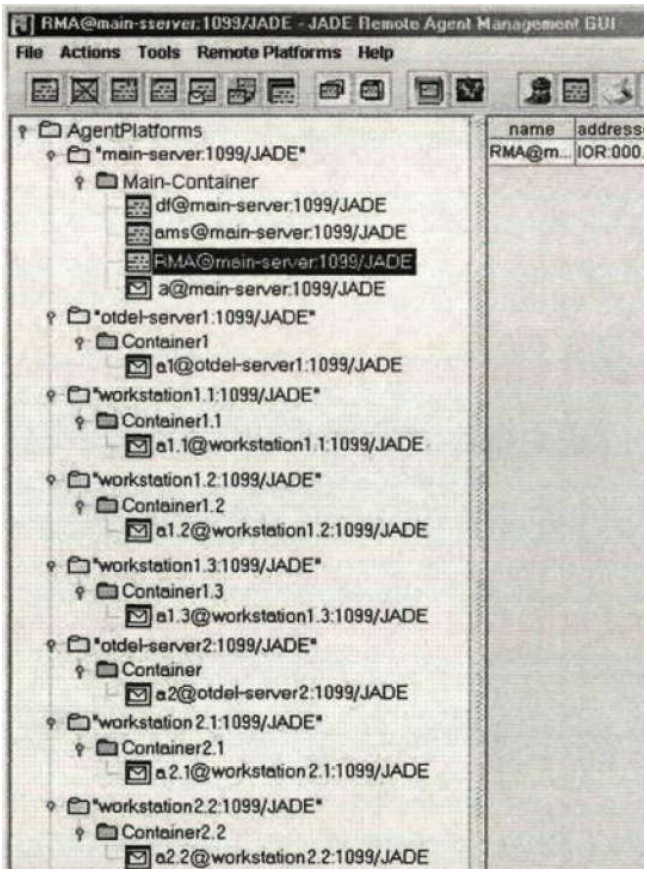


Рисунок 3.6 – Прототип багатоагентної системи боротьби зі спамом

Фрагмент коду прототипу багатоагентної системи класифікації спаму представлено в додатку В.

Фрагмент коду програми побудови бази знань представлено в додатку Г.

Довідка про використання результатів дипломної роботи представлена в додатку Д.

В третьому розділі роботи запропоновано використання багатоагентної технології для реалізації багаторівневої системи фільтрації спаму, що дозволяє

забезпечити в рамках запропонованої концепції масштабованість, властивість самоорганізації та інтелектуалізації процесу фільтрації спаму. Розроблено архітектуру багатоагентної системи фільтрації спаму, що відрізняється від відомих можливістю гнучкої реконфигурації системи з врахованими зміни конфігурації ЛОМ організації. Розроблено моделі агентів, виділено їх основні цілі та задачі, які вони вирішують. Розроблено прототип багатоагентної системи на базі програмно–технологічної платформи JADE, що дозволяє забезпечити кросплатформеність розробленої системи.

ВИСНОВКИ

В роботі поставлена і вирішена задача розробки системи фільтрації спаму в організації. При вирішенні даної задачі отримані такі результати:

1. Розроблено концепцію побудови автоматизованої ієрархічної системи протидії шкідливому впливу спам-розсилок на інформацію, що обробляється в системі електронної пошти, яка полягає в багаторівневої фільтрації спаму з використанням баз знань, різних по повноті.

2. Реалізація розробленої концепції в організації дозволяє підвищити точність класифікації електронних повідомлень на різних рівнях ієрархії системи фільтрації і забезпечити цілісність і доступність інформації в рамках прийнятої в організації політики безпеки.

3. Розроблена ієрархічна архітектура системи захисту електронної поштової інформації в класі розподілених систем обробки інформації на основі багатоагентного підходу, яка дозволяє забезпечити масштабованість і властивості самоорганізації для гнучкої політики безпеки, прийнятої в організації.

4. Розроблено ефективний алгоритм класифікації електронних повідомлень на основі когнітивного підходу та нейромережевого класифікатора, що дозволяє вирішувати задачу класифікації входящих електронних повідомлень на різних рівнях ієрархії організації з частковим врахуванням семантики повідомлення.

5. Розроблено методику проектування багатоагентної системи протидії поширенню спаму в організації із застосуванням однотипного нейромережевого класифікатора на всіх ієрархічних рівнях обробки інформації, використання якої дозволяє реалізувати елементи запропонованої автоматизованої системи фільтрації спаму в організації в вигляді програмних модулів.

6. Як показали результати моделювання функціонування розробленої системи, рівень помилок першого і другого роду знизився на 5–10%.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Стрельцов А.А. Обеспечение информационной безопасности России. – М.: МЦНМО, 2002. – 290 с.
2. Горелик А. Л., Скрипкин В. А. Методы распознавания. – М.: Высшая школа, 2004. – 261 с.
3. Templeton, B.: n.d.b, Origin of the term "spam" to mean net abuse, http://www.templetons.com/brad/spamterm.html, p. 54.
4. OECD: 2004a, Background Paper for the OECD Workshop on Spam, p. 40–41.
5. Gauthronet, S. and Etienne, D.: 2001, Unsolicited Commercial Communications and Data Protection, p. 128.
6. NOIE: 2003, Final Report of the NOIE Review of the Spam Problem and How It Can Be Countered, Technical report. www.noie.gov.au/publications/NOIE/spam/final report/SPAMreport.pdf, p. 1.
7. Spamhaus: n.d., The Definition of Spam, http://www.spamhaus.org/definition.html, p. 1–2.
8. The New York Times: 2003, We Hate Spam, Congress Says (Except Ours), http://www.nytimes.com, p. 3.
9. The Government of Hongkong: n.d., Nigerian letters, http://www.mfo.gov.hk/police/pda/con–tricks/con7.htm, p. 2–3.
10. Doll, J.: n.d., Spam Attack, http://www.joes.com/spammed.html, p. 3.
11. OECD: 2004b, Report of the 2nd OECD Workshop on Spam, p. 17–18.
12. Ironport: 2006, Internet Email Traffic Emergency: Spam Bounce Messages are Compromising Networks, http://www.ironport.com/bouncereport/, p. 3.
13. Evett, D.: 2006, Spam Statistics 2006, http://spam–filterreview.toptenreviews.com/spam–statistics.html, p. 5–6.
14. Ferris Research: 2005, The Global Economic Impact of Spam, 2005. Report #409, p. 172–173.
15. Bundesamt f.ur Sicherheit in der Informationstechnik (BSI): 2005, Antispam – Strategien: Unerw.unschte E–mails erkennen und abwehren, p. 75.
16. Tanase, M.: 2003, IP Spoofing: An Introduction,

http://www.securityfocus.com/infocus/1674, p. 1–2.
17. DECLUDE Internet Security Software: 2005, List of All Known DNSbased Spam Databases, http://www.decmde.com/Articles.asp?ID=97, p. 3.
18. Graham, P.: 2003, Better Bayesian Filtering, Proceedings of the 2003 Spam Conference, p. 117–119.
19. Androutsopoulos, I., Magirou, E. and Vassilakis, D.: 2005, A Game Theoretic Model of Spam E–Mailing, Proceedings of the 2nd Conference on Email and Anti–Spam (CEAS 2005), p. 12–15.
20. Turner, D. A. and Havey, D. M.: 2004, Controlling Spam through Lightweight Currency, Proceedings of the 37th Annual Hawaii International Conference on System Sciences, p. 70–72.
21. Cohen, W.: 1996, Learning rules that classify e–mail, Papers from the AAA! Spring Syposium on Machine Learning in Information Access, p. 18–25.
22. Crawford, E., Kay, J. and McCreath, E.: 2001, Automatic induction of rules for e–mail classification, Proceedings of the Sixth Australasian Document Computing Symposium, p. 34–35.
23. Damiani, E., De Capitani di Vimercati, S., Paraboschi, S. and Samarati, P.: 2004, P2p–based collaborative spam detection and filtering, Proceedings of the Fourth International Conference on Peer–to–Peer Computing, p. 176–183.
24. Zhou, F., Zhuang, L., Zhao, B., Huang, L., Joseph, A. and Kubiatowicz, J.: 2003, Approximate object location and spam filtering on peer–to–peer systems, Proceedings of ACMAFIP/USENIX International Middleware Conference, p. 22–26.
25. Drucker, H., Wu, D. and Vladimir N. Vapnik: 1999, Support Vector Machines for Spam Categorization, IEEE Transactions on Neural Networks 10(5), p. 1048–1054
26. Metzger, J., Schillo, M. and Fischer, K.: 2003, A Multiagent–Based Peer–to–Peer Network in Java for Distributed Spam Filtering, in V. Mark, J. Mller and M. Pechoucek (eds), Proceedings of the 3rd International Workshop of Central and Eastern Europe on Multi–Agent Systems (CEEMAS), p. 616–625.
27. Carreras, X. and Mrquez, L.: 2001, Boosting Trees for Anti–Spam Email Filtering, in T. Chark (ed.), Proceedings of the International Conference on Recent Advances

in Natural Language Processing, p. 126.

28. Drawes, R.: 2002, An artificial neural network spam classifier, Technical report. www.interstice.com/drewes/cs676/spam–nn/spam–nn.html, p. 2.
29. Chhabra, S. and Siefkes, C: 2004, Spam Filtering using a Markov Random Field Model with Variable Weighting Schemas, Proceedings of the 4th IEEE International Conference on Data Mining, p. 347–350
30. Anti–Spam Technical Alliance (ASTA): 2004, Anti–spam technical alliance technology and policy report, Technical report, p. 223–226.
31. Myers, J.: 1999, SMTP Service Extension for Authentication, RFC 2554, IETF Network Working Group, p. 92.
32. Myers, J.: 1997, Simple Authentication and Security Layer (SASL), RFC 2222, IETF Network Working Group, p. 117–119.
33. Ramsdell, B.: 2004, Secure/Multipurpose Internet Mail Extensions (S/MIME) Version 3.1 Message Specification, RFC 3851, IETF Network Working Group, p. 52.
34. Leibzon, W.: 2005b, META Signatures (Message Enhancements for Transmission Authorization), Technical report, http://www.metasignatures.org/meta signatures proto–col.htm, p. 1–2.
35. Fenton, J. and Thomas, M.: 2005, Identified Internet Mail, J. fenton and m. thomas, IETF Network Working Group, pp. 165–167.
36. Delany, M.: 2005, Domain–based Email Authentication Using Public–Keys Advertised in the DNS (DomainKeys), Internet draft, IETF Network Working Group, p.202–204.
37. Microsoft: 2004, Email Postmarks, p. 119–121.
38. Leibzon, W.: 2005a, Email Security Anti–Spoofing Protection with Path and Cryptographic Authentication Methods, http://www.metasignatures.org/path and cryptographic authentication.htm, p. 2.
39. Schryen, G.: 2004b, Approaches addressing spam, Proceedings of the HHCCП, p.41–43.
40. Garfinkel, S.: 2003, Email–Based Identification and Authentication: An Alternative to PKI?, IEEE Security & Privacy 1(6), p. 20–26.
41. DeKok, A.: 2004, Lightweight MTA Authentication Protocol (LMAP) Discussion

and Applicability Statement, Internet draft, IETF Network Working Group, p. 65–67.

42. Wong, M. and Schlitt, W.: 2005, Sender Policy Framework (SPF) for Authorizing Use of Domains in E–MAIL, version 1, Internet draft, IETF Network Working Group, p. 52–54.
43. Danish, H.: 2004, The RMX DNS RR and method for lightweight SMTP sender authorization, Internet draft, IETF Network Working Group, p. 117–119.
44. Fecyk, G.: 2003, Designated Mailers Protocol, Internet draft, IETF Network Working Group, p. 91–93.
45. Lyon, J.: 2005, Purported Responsible Address in E–Mail Messages, Internet draft, IETF Network Working Group, p. 29–31.
46. Crocker, D., Leslie, J. and Otis, D.: 2005, Certified Server Validation (CSV), Internet draft, IETF Network Working Group, p. 58–59.
47. Stumpf, M. and Hoehne, S.: 2005, Marking Mail Transfer Agents in Reverse DNS with TXT RRs, Internet draft, IETF Network Working Group, p. 15–19.
48. Schryen, G. and Hoven, R.: 2004, Appropriateness of Lightweight MTA Authentication Protocols for Fighting Spam, Proceedings of IPSI International Conference on Advances in the Internet, Processing, Systems, and Interdisciplinary Research, p. 98–99.
49. von Ahn, L., Blum, M. and Langford, J.: 2004, Telling Humans and Computers Apart Automatically, Communications of the ACM 47(2), p. 57–60.
50. Bless, R., Conrad, M. and Hof, H.–J.: 2005, Spam Protection by using Sender Address Verification Extension (SAVE). http://doc.tm.uka.de/2005/SAVE.pdf, p. 1.
51. Tompkins, T. and Handley, D.: 2003, Giving E–mail Back to the Users: Using Digital Signatures to Solve the Spam Problem, FirstMonday 8(9), p. 34–36.
52. Dwork, C. and Naor, M.: 2002, Procing via Processing or Combatting Junk Mail, in D. Boneh (ed.), Proceedings of the 22rd Annual International Cryptology Conference (CRYPTO 2002), number 740 in LNCS, Springer, p. 137–147.
53. Back, A.: 2002, Hashcash – A Denial of Service Counter–Measure, http://www.hashcash.org/papers/hashcash.pdf, p. 2.
54. Johansson, E.: n.d., Camram, www.camram.org, p. 2–3.

55. Abadi, M., Burrows, M,, Manasse, M. and Wobber, T.: 2003, Moderately Hard, Memory–Bound Functions, Proceedings of the 10th Annual Network and Distributed System Security Symposium, p. 25–39.
56. Templeton, B.: n.d.a, E–stamps, http://www.templetons.com/brad/spam/estamps.html, p. 1–3.
57. Loder, T., van Alstyne, M. and Walsh, R.: 2004, Information Asymmetry and Thwarting Spam, Technical report, University of Michigan, p. 52–54.
58. Fahmann, S.: 2002, Selling interrupt rights: A way to control unwanted e–mail and telephone calls, IBM Systems Journal 41(4), p. 759–766.
59. Turner, D. and Ross, K.: 2003, A Lightweight Currency Paradigm for the P2P Resource Market, p. 107–109.
60. Anti–Spam Technical Alliance (ASTA): 2004, Anti–spam technical alliance technology and policy report, Technical report, p. 62–63.
61. Hall, R.: 1996, Channels: Avoiding Unwanted Electronic Mail, Proceedings of the DIMACS Symposium on Network Threats, p. 135–137.
62. Gabber, E., Jakobsson, M., Matias, Y. and Mayer, A. J.: 1998, Curbing Junk E–Mai via Secure Classification, Proceedings of the Second International Conference on Financial Cryptography, p. 198–213.
63. Email Service Provider Coalition: 2003, Project Lumos: A Solutions Blueprint for Solving the Spam Problem by Establishing Volume Email Sender Accountability, Technical report, p. 76–78.
64. Nikitin A. P. Multi–Agent Anti–Spam System. P. 227–229. Proceeding of the Workshop on Computer Science and Information Technologies (CSIT'2005), Ufa, September 18–21, 2005. Volume 2, p. 227–229.
65. Никитин А. П. Система классификации электронных сообщений. Зимняя школа аспирантов 2006–2007. Уфа 2007, с. 55–56
66. Никитин А. П. Система интеллектуальной фильтрации спама. Труды I международной заочной научно–технической конференции «Актуальные проблемы безопасности информационных технологий» (АПроБИТ–2007) Красноярск –2007., с. 40–44.
67. Nikitin A.P., Valeev S.S. Intellectual Anti–Spam System. 9th International Work-

shop on Computer Science and Information Technologies CSIT'2007, Ufa–Krasnousolsk, Russia, 2007, p. 57–59.

68. Городецкий Б.Ю. Компьютерная лингвистика: моделирование языкового общения // Новое в зарубежной лингвистике. – М., 1989. — Вып. 24, с. 125–127.
69. Демьянков В.З. Основы теории интерпретации и ее приложения в вычислительной лингвистике. – М., 1985, с. 136–138.
70. Никитин А. П. Валеев С. С. Интеллектуальная система фильтрации сообщений. Стр. 246–247. Восьмая Международная научно–техническая конференция «Проблемы техники и технологии телекоммуникации», 26–28 ноября 2007 г. Уфимск. гос. авиац. техн. ун–т. – Уфа, 2007 –382 с.
71. Фрайн В. С. Распознавание образов и машинное понимание естественного языка.–М., 1987, с. 89–91.
72. Кобзарева Т.Ю. Принципы сегментационного анализа русского предложения // Московский лингвистический журнал. – М., 2004. – Т. 8. – № 1, с. 23–24.
73. Ножов И.М. Прикладной морфологический анализ без словаря // Труды конференции по искусственному интеллекту–2000. – М., 2000. – Т. 1. – с.424–429.
74. Анношкина Ж.Г. Морфологический процессор русского языка // Альманах "Говор". – Сыктывкар, 1995. – с. 17–23.
75. Волкова И.А., Мальковскии М.Г., Одинцев Н.В.Адаптивный синтаксический анализатор *II* Труды Международной конференции ДИАЛОГ–2003. – Протвино, 2003.–с. 476–480.
76. Ножов И.М. Реализация автоматической синтаксической сегментации русского предложения. — М., 2003, с. 46.
77. Богуславский И.М., Цинман Л.Л. Семантический компонент лингвистического процессора // Семотика и информатика. – М., 1190. – Вып. 30. – с. 5–30.
78. Новиков А.И. Семантика текста и ее формализация. – М., 1983, с. 144–146.
79. Рубашкин В. Ш. Семантический анализ текста: Модели и методы // Материалы конференции CORPORA–2004. – СПб., 2004, с. 67–69.
80. David P. Buse, Qing–Hua Wu: IP Network–based Multi–agent Systems for Industrial Automation: Information Management, Condition Monitoring and Control of

Power Systems, Springer, 2006, p. 78–80.

81. Hans–Dieter Burkhard, Gabriela Lindemann, Rineke Verbrugge: Multi–Agent Systems and Applications V: 5th International Central and Eastern European Conference on Multi–Agent Systems, CEEMAS 2007, Leipzig, Germany, (Lecture Notes in Computer Science), Springer, 2007, p. 144–146.
82. Katsumi Inoue, Ken Satoh, Francesca Toni: Computational Logic in Multi–Agent Systems: 7th International Workshop, CLIMA VII, Hakodate, Japan, May 8–9, 2006, Revised Selected and Invited Papers (Lecture Notes in Computer Science), Springer, 2007, p. 150.
83. Rafael H. Bordini, Mehdi Dastani, Jurgen Dix, Amal El Fallah Seghrouchni: Programming Multi–Agent Systems: 4th International Workshop, ProMAS 2006, Hakodate, Japan,May 9, 2006, Revised and Invited Papers (Lecture Notes in Computer Science), Springer, 2007, p. 56–59.
84. Ngoc Thanh Nguyen, Adam Grzech, Lakhmi C. Jain: Agent and Multi–Agent Systems: Technologies and Applications: First KES International Symposium, KES–AMSTA 2007, Wroclaw, Poland, May 31–June 1, 2007, Proceedings (Lecture Notes in Computer Science), Springer, 2007, p. 94–97.
85. Chris van Aart: Organizational Principles for Multi–Agent Architectures (Whitestein Series in Software Agent Technologies), Birkhauser, 2005, p. 37–39.
86. Lin Hong: Architectural Design of Multi–Agent Systems: Technologies and Techniques , IGI Global, 2007, p. 123–125.
87. Moreno Antonio, Pavon Juan: Issues in Multi–Agent Systems (Whitestein Series in Software Agent Technologies and Autonomic Computing), Birkhauser Basel, 2007, p. 112–114.
88. Ron Sun: Cognition and Multi–Agent Interaction : From Cognitive Modeling to Social Simulation, Cambridge University Press, 2005, p. 103.
89. Shamma Jeff: Cooperative Control of Distributed Multi–Agent Systems, Wiley–Interscience, 2008, p. 89–91.
90. Cavedon Lawrence, Maamar Zakaria, Martin David, Benatallah Boualem: Extending Web Services Technologies: The Use of Multi–Agent Approaches, Springer, 2005, p. 57–59.

91. Методичні рекомендації до виконання дипломної роботи з освітньо–кваліфікаційного рівня “Магістр”. Спеціальність „Комп’ютерні системи та мережі” / О.М. Березький, Р.Б. Трембач, Г.М. Мельник / Під ред. О.М. Березького – Тернопіль: ТНЕУ, 2012.– 42 с.