МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Харківський національний університет радіоелектроніки
Факультет Центр післядипломної освіти
(повна назва)

Кафедра Програмної інженерії
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти другий (магістерський)

Дослідження моделей фільтрування інформації в коммунікаційних мережах
(тема)

Виконав:
Студент 2 курсу, групи ІПЗздм-19-1
Ангелін А.І.
(прізвище, ініціали)

Спеціальність 121 Інженерія програмного забезпечення
(код і повна назва спеціальності)

Тип програми освітньо-наукова

Керівник проф. Шостак І.В.
(посада, прізвище)

Допускається до захисту
Зав. кафедри _____ З.В. Дудар
(підпис) (прізвище, ініціали)

2021

Харківський національний університет радіоелектроніки

Факультет Центр післядипломної освіти
(повна назва)

Кафедра Програмної інженерії
(повна назва)

Рівень вищої освіти другий (магістерський)

Спеціальність 121 Інженерія програмного забезпечення
(код і повна назва спеціальності)

Тип програми освітньо-наукова
(освітньо-професійна або освітньо-наукова)

Освітня програма Інженерія програмного забезпечення
(повна назва)

ЗАТВЕРДЖУЮ:
Зав.кафедри _____
(підпис)
« ____ » _____ 2021 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

студента Ангеліна Андрія Ігоровича
(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження моделей фільтрування інформації в коммунікаційних мережах

затверджена наказом університету від 26.03.2021 № 34 Стз

2. Термін подання роботи до екзаменаційної комісії 11 05 2021р.
3. Вихідні дані до роботи *проаналізувати існуючі алгоритми, що використовуються для вимог підтримки прийняття рішень, мови розробки програмного забезпечення*
4. Перелік питань, що потрібно опрацювати в роботі *мета роботи, аналіз проблемної галузі і постановка задачі, опис запропонованих варіантів оптимізації, використовувані методи та алгоритми, опис розробленої програмної системи, опис застосованих програмних рішень, аналіз можливих застосувань*
5. Перелік графічного матеріалу із зазначенням креслеників, схем, слайдів, ілюстрацій (п.5 включається до завдання за рішенням випускової кафедри)

Мета завдання, обґрунтування доцільності розробки, постановка задачі, базові моделі, методи й алгоритми, структурно-логічна схема взаємодії даних, інтерфейс програмної системи, результати дослідної експлуатації програмної

системи, висновки

6. Консультанти розділів роботи (п.6 включається до завдання за наявності консультантів згідно з наказом, зазначеним у п.1)

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата
спецчастина	проф. Шостак І.В.		

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1.	Аналіз предметної галузі	26 березня 2021 р.	виконано
2.	Огляд існуючих методів	31 березня 2021 р.	виконано
3.	Розробка алгоритмів, проектування та розробка ПЗ	15 квітня 2021 р.	виконано
4.	Підготовка пояснювальної записки	28 квітня 2021 р.	виконано
5.	Спецчастина	30 квітня 2021 р.	виконано
6.	Підготовка презентації та доповіді	05 травня 2021 р.	виконано
7.	Попередній захист	10 травня 2021 р.	виконано
8.	Нормоконтроль, рецензування	10 травня 2021 р.	виконано
9.	Занесення роботи в електронний	11 травня 2021 р.	виконано
10.	Допуск до захисту в зав. кафедри	12 травня 2021 р.	виконано

Дата видачі завдання ____ _____ 2021р.

Студент _____
(підпис)

Керівник роботи _____ проф. Шостак І.В.
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ /ABSTRACT

Пояснювальна записка до кваліфікаційної роботи магістра: 90 с, 5 табл., 46 рис., 6 дод., 37 джерел

ІНФОРМАЦІЙНО-ТЕЛЕКОМУНІКАЦІЙНІ МЕРЕЖІ, ЗАБОРОНЕНИЙ КОНТЕНТ, БЕЗПЕКА ДАНИХ, СЕМАНТИЧНИЙ АНАЛІЗ.

Об’єктом дослідження є інформаційно-телекомунікаційні мережі, що перебувають під впливом загрози поширення забороненої інформації.

Мета роботи полягає в підвищенні точності прогнозування загрози поширення забороненої інформації в інформаційно-телекомунікаційних мережах.

Результат – розроблене алгоритмічне і програмне забезпечення, що автоматизує процес пошуку вузлів – потенційних розповсюджувачів забороненої інформації у великомасштабних інформаційно-телекомунікаційних мережах, що й дозволяє скоротити час пошуку таких вузлів.

INFORMATION AND TELECOMMUNICATIONS NETWORKS, PROHIBITED CONTENT, DATA SECURITY, SEMANTICAL ANALYSIS.

The object of the study is information and telecommunication networks, which are under the influence of the threat of dissemination of prohibited information.

The purpose of this work is to increase the accuracy of forecasting the threat of dissemination of prohibited information in information and telecommunications networks.

The result is a developed algorithmic and software that automates the process of finding nodes – potential distributors of prohibited information in large-scale information and telecommunications networks, which reduces the search time for such nodes.

Я, Ангелін Андрій Ігорович, студент гр. ІПЗздм-19-1, здобувач вищої освіти на другому (магістерському) рівні кафедри «Програмна інженерія», заявляю: моя кваліфікаційна робота на тему «Дослідження моделей фільтрування інформації в коммунікаційних мережах», що буде представлена в екзаменаційну комісію для публічного захисту, виконана самостійно, в ній не містяться елементи плагіату і вона може бути опублікована в електронному архіві відкритого доступу ElAr KhNURE. Всі запозичення з друкованих та електронних джерел мають відповідні посилання.

Я ознайомлен з діючим положенням «Про протидію академічному плагіату в ХНУРЕ», згідно з яким виявлення плагіату є підставою для відмови в допуску кваліфікаційної роботи до захисту та застосування дисциплінарних заходів.

ЗМІСТ

ВСТУП

Інформаційно-телекомунікаційні мережі (ІТКМ) забезпечують практично повний спектр можливостей для обміну інформацією між користувачами – мережевими абонентами. Сучасною проблемою таких систем є їхній низький рівень інформаційної безпеки. Для забезпечення захисту інформації в телекомунікаційних мережах, включаючи Інтернет, розроблено багато методів і засобів. Проте, ефективного захисту абонентів від загроз поширення забороненої інформації, зокрема в умовах широкого використання індивідуально-орієнтованих сервісів і пов’язаних з ними протоколів і технологій (SOAP, CORBA, REST і ін.), не існує. Серед великої кількості функцій захисту, принциповою відносно даних систем є функція попередження прояву забороненої інформації. Вона реалізується за рахунок механізмів прогнозування загрози поширення й розсилання повідомлень із попередженнями про наслідки дій із забороненим контентом. Використання інших функцій (попередження, виявлення, локалізації й ліквідації загрози) припускає наявність повного контролю над системою, що в сучасних умовах неможливо [1].

Одним з підходів до прогнозування загрози поширення забороненої інформації (ЗПЗІ) є моделювання, наприклад, з використанням моделей впливу, моделей просочування й зараження. Дані моделі, як правило, не враховують топологічні особливості мережі (розподіл степенів зв’язності, кластерний коефіцієнт, середня довжина шляху). Взаємодія між абонентами в рамках цих математичних моделей описується переважно гомогенним графом, що при моделюванні великомасштабних мереж (більш ніж 10 млн. вузлів) може дати похибку прогнозування ЗПЗІ понад 30%. Крім того, такі підходи носять в основному теоретичний характер, практика їх використання не виходить за рамки експериментів. Таким чином, дослідження, спрямовані на створення моделей і алгоритмів ЗПЗІ, актуальні й мають теоретичне й практичне значення в

розв’язанні проблеми забезпечення інформаційної безпеки в системах і мережах телекомунікацій [2].

Об’єктом дослідження є інформаційно-телекомунікаційні мережі, що перебувають під впливом загрози поширення забороненої інформації.

Предметом дослідження є моделі загрози поширення забороненої інформації в інформаційно-телекомунікаційних мережах.

Мета роботи полягає в підвищенні точності прогнозування загрози поширення забороненої інформації в інформаційно-телекомунікаційних мережах.

Для досягнення мети роботи необхідно розв’язати наступні завдання:

– провести інформаційний огляд і експерименти для виявлення істотних характеристик об’єкта й зовнішніх факторів, що впливають на процес реалізації ЗПЗІ. Виконати аналіз основних підходів до моделювання ЗПЗІ;

– розробити імітаційну модель ЗПЗІ в ІТКМ;

– синтезувати й показати адекватність аналітичної моделі ЗПЗІ в ІТКМ;

– розробити методику формування топології ІТКМ;

– змоделювати процес реалізації ЗПЗІ на топології реальної великомасштабної ІТКМ із використанням розробленого програмного забезпечення та провести експериментальне дослідження отриманих результатів.

Розроблена імітаційна модель реалізації ЗПЗІ в ІТКМ, що враховує середній степінь зв’язності вузлів, середню довжину шляху мережі, коефіцієнт кластерізації мережі, а також особливості інформаційної взаємодії абонентів як людино-машинних систем, що й дозволяє підвищити точність представлення процесів забезпечення інформаційної безпеки у великомасштабних ІТКМ.

Розроблена інформаційна модель реалізації ЗПЗІ, що відрізняється від класичної епідеміологічної моделі Кермака та МакКендрика переліком характеристик уразливості ІТКМ, що й дозволяє підвищити точність оперативного прогнозу, особливо в умовах неповноти вихідних даних про топологію мережі.

Розроблена методика формування топології великомасштабної ІТКМ, що включає:

– алгоритм формування графа доступної частини мережі, що дозволяє здійснити збір даних про топологію з будь-якого вузла абонента;

– алгоритм формування повного графа мережі, що дозволяє в умовах неповноти вихідних даних спрогнозувати топологію відсутньої частини мережі.

Застосування методики дозволяє підвищити точність представлення моделі топології ІТКМ.

Розроблене програмне забезпечення, що автоматизує процес пошуку вузлів – потенційних розповсюджувачів забороненої інформації у великомасштабних інформаційно-телекомунікаційних мережах, що й дозволяє скоротити час пошуку таких вузлів.

Розроблено алгоритми й програмне забезпечення аналізу топології великомасштабної інформаційно-телекомунікаційної мережі, яка дозволяє підвищити захищеність організації за рахунок скорочення часу розслідування інцидентів у рамках ліквідації наслідків порушення конфіденційності.

1 БЕЗПЕКА В ІНФОРМАЦІЙНО-ТЕЛЕКОМУНІКАЦІЙНИХ МЕРЕЖАХ

1.1 Аналіз об’єкта дослідження

ІТКМ забезпечують практично повний спектр можливостей для обміну інформацією між користувачами – мережевими абонентами. ІТКМ надає різні сервіси для організації соціальних взаємовідношень між користувачами (абонентами). На сьогоднішній день найбільш популярним з них є соціальні мережі.

У світі існує величезна кількість різних соціальних мереж, але практично в кожній країні або регіоні їх популярність може відрізняєтися. У США це «Facebook», «Myspace», «Twitter» і «Linkedin»; «Nexopia» – у Канаді; «Bebo» – у Великобританії; «Facebook», «dol2day» – у Німеччині. На рисунку 1.1 зображена динаміка росту [3].

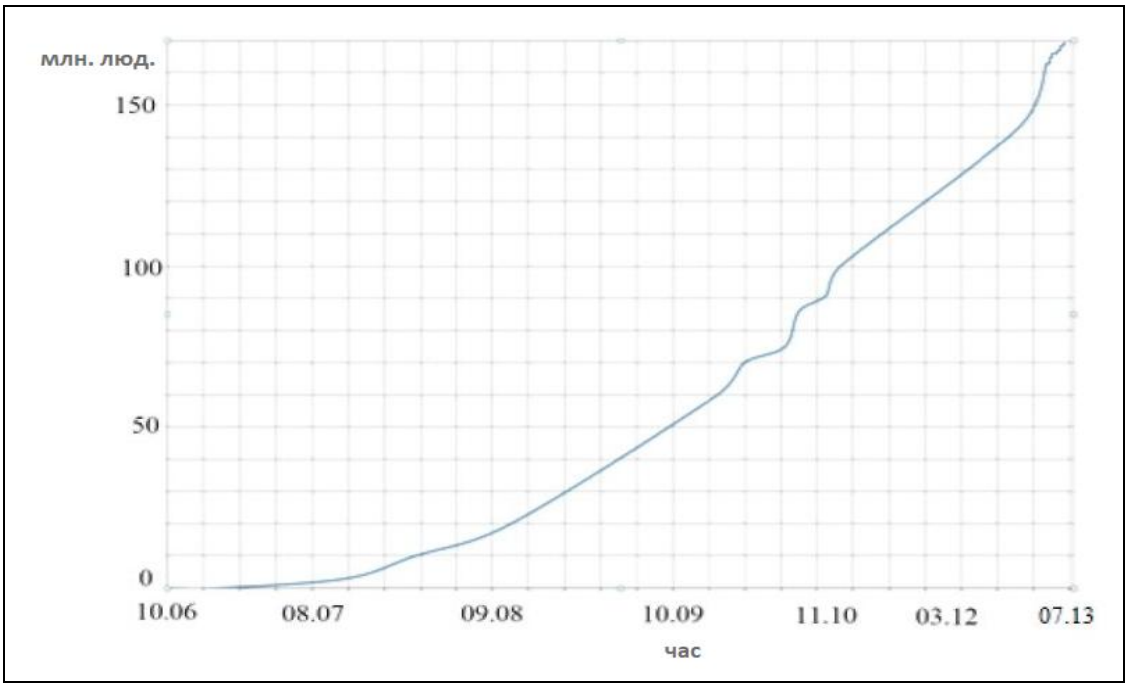

Рисунок 1.1 – Динаміка росту користувачів соціальних мереж [3]

З бурхливим ростом числа користувачів ІТКМ виникають і проблеми безпеки в них.

Узагальнена структурна схема ІТКМ наведена на рисунку 1.2.

Рисунок 1.2 – Структурна схема ІТКМ

Її склад у загальному випадку утворюють наступні функціональні елементи абоненти (А). Під абонентом розуміється людино-машинна система, що складається з обладнання, через яке здійснюється доступ до мережі, і безпосередньо користувача ІТКМ. Абоненти можуть бути окремими вузлами мережі (якщо користувач використовує свій домашній комп'ютер), або можуть бути об'єднані в корпоративну обчислювальну мережу (КОМ) (якщо абонент використовує робочий комп'ютер), містять у собі модулі (інформаційного) захисту (МЗ) і програмне забезпечення (браузер) для взаємодії з керуючим елементом:

– мобільні абоненти (МА) – користувачі, що використовують мобільні пристрої (смартфони, планшети й інше), для доступу до мережі. Також використовують програмне забезпечення (спеціальний додаток) і МЗ;

– сервери (С) – у КОМ перебувають інформаційні сервери різного функціонального призначення, що здійснюють інформаційну взаємодію (наприклад, проксі-сервер);

– КОМ містить у собі крім абонентів і серверів, також засоби маршрутизації, комутації й адміністрування (МКА), систему безпеки (СБ), що включає механізми захисту для всієї корпоративної мережі;

– засоби телекомунікації, що забезпечують взаємодію між собою абонентів;

– керуючий елемент технічно являє собою сукупність комутуючого й серверного обладнання, що реалізує основні функції системи. Містить у собі сервери, що містять у загальному випадку: балансувальники навантаження (БН), елемент бізнес-логіки (БЛ), бази даних (БД), інфраструктурні системи (ІС) (системи статистики, конфігурації, моніторингу тощо).

Архітектура типового керуючого елемента представлено на рисунку 1.3. Ця багатошарова архітектура містить «Презентаційний шар» , на цьому шарі приймаються HTTP-запити від абонентів, зазвичай веб-браузерів, і видаються їм HTTP-відповіді, як правило, разом з HTML-сторінкою, зображенням, файлом, медіа-потоком або іншими даними. Тут же здійснюється розподіл і балансування навантаження, ведення журналу звернень абонентів до ресурсів [4].

Шар бізнес сервісів призначений для відбору й обробки даних.

Персистентний шар виконує обслуговування й керування базою даних і відповідає за цілісність і сінхоронність даних, а також забезпечує операції введення-виведення при доступі абонента до інформації.

На шарі загальних інфраструктурних систем розміщуються системи протоколювання статистики, конфігурації додатків, моніторингу.

SSO (Single Sign-On, технологія єдиного входу) – технологія, при використанні якої користувач переходить із одного розділу порталу в інший без повторної аутентифікації.

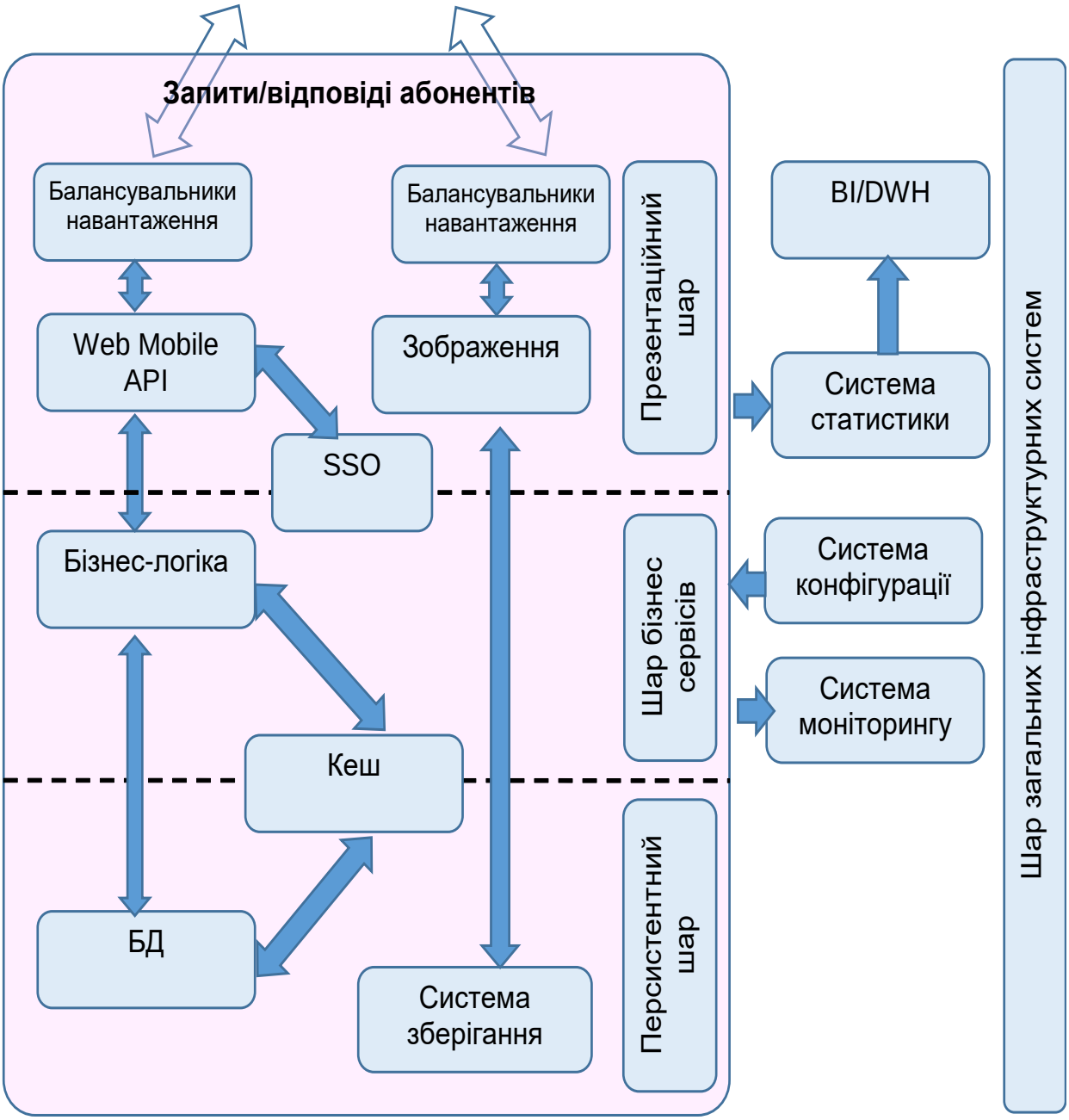

Рисунок 1.3 – Архітектура типового керуючого елемента

BI (Business intelligence, бізнес-аналіз, бізнес-аналітика) [5]– методи й інструменти для побудови інформативних звітів про поточну ситуацію в системі.

DWH (Data Warehouse, сховище даних) [6] – предметно-орієнтована інформаційна база даних, спеціально розроблена й призначена для підготовки звітів і бізнесу-аналізу.

1.2 Проблеми інформаційної безпеки в ІТКМ

Основні проблеми інформаційної безпеки в ІТКМ, які актуальні для даного дослідження. Використання глобальної мережі Інтернет як розподіленої інформаційно-телекомунікаційної системи.

Найбільш уразливими компонентами системи, які через це найчастіше атакуються, є:

- сервери;
- робочі станції;
- середовища передачі інформації;
- вузли комутації.

Типові інформаційні впливи зловмисників описано нижче.

Прослуховування мережевого трафіка. Для прослуховування трафіка (sniffing) мережевий адаптер переводиться в «безладний» режим. У даному режимі адаптер перехоплює всі мережеві пакети, що проходять через нього, а не лише призначені даній адресі, як у нормальному режимі функціонування - технології – ARP Spoofing (Arp-poisoning), MAC Flooding і MAC Duplicating [7]. Перехоплення здійснюється з використанням мережевих моніторів, з яких найбільш функціональними є Sniffer Pro від компанії Sniffer Technologies [8], IRIS Network Traffic Analyzer [9] і TCP Dump [10]. Таким чином, сучасні мережеві протоколи (TCP/IP, ARP, HTTP, FTP, SMTP, POP3 й інші) не мають механізмів захисту (передаються у відкритому вигляді).

Зловмисник, що перехоплює трафік між сервером і будь-яким вузлом мережі, може заволодіти аутентифікаційними даними користувача (отримати пароль).

Протидія – відомо кілька методів визначення наявності запущеного сніффера в мережі, наприклад, метод пінгу, метод ARP, метод DNS і метод пастки [11].

Сканування вразливостей. Результатом роботи сканера є інформація про систему, що включає список мережевого устаткування, комп'ютерів із запущеними на них службами, версіями мережевого ПО (а значить і вразливостей, властивих даному ПО), облікові записи користувачів. Сканування вразливостей, звичайно, є етапом, що випереджає атаку. Саме результати сканування дозволяють точно підібрати експлойти для здійснення безпосередньо НСД.

Виявлення**.** Само по собі сканування не є незаконним. Однак, якщо сканування з боку зовнішньої, стосовно системи, мережі звичайне явище, то сканування комп'ютерів із внутрішньої мережі – безумовно, інцидент безпеки, що вимагає негайної реакції з боку мережевого адміністратора. Виявити сліди сканування можна, вивчаючи журнали реєстрації МЕ. Однак такий підхід не дозволяє вчасно реагувати на подібні інциденти. Тому сучасні МЕ й СОВ мають модулі [12], що дозволяють виявити сканування в режимі реального часу. Деякі сканери вразливостей використовують оригінальні методи, що дозволяють робити сканування максимально скритно. Наприклад, в системі Nmap [13] існують можливості, що дозволяють значно ускладнити виявлення сканування для СОВ.

Протидія – використання мережевих СОВ, або періодичне вивчення журналів реєстрації МЕ.

Мережеві атаки можна розділити на [14]:

– атаки, засновані на переповненні буферу (overflow based attacks). Вони використовують уразливість системи, що полягає в некоректній програмній обробці даних, при цьому з'являється можливість виконання шкідливого коду з підвищеними привілеями;

– атаки, спрямовані на відмову в обслуговуванні (Denial Of Service attacks), атаки не обов'язково використовують вразливості в ПЗ системи, що атакується.

Порушення працездатності системи відбувається через те, що дані, які їй посилають, приводять до значної витрати ресурсів системи. Найпростішим прикладом атаки цього типу є атака «Ping Of Death» [15]. Сутність її в наступному: на машину жертви посилається сильно фрагментований ICMP-пакет

великого розміру (64KB). Реакцією ОС Windows на одержання такого пакета буде повне зависання.

Атаки, що засновані на використанні вразливостей у ПЗ мережевих додатків – експлойти (exploit) [16]. Даний клас атак заснований на експлуатації різних дефектів у ПЗ. Експлойти представляють собою шкідливі програми, що реалізують відому уразливість в ОС або прикладному ПЗ, для одержання НСД до вразливого хосту або порушення його працездатності. Для експлойтів характерна наявність функцій придушення антивірусних програм і МЕ. Наслідки застосування експлойтів можуть бути самими критичними. У випадку одержання зловмисником дистанційного доступу до системи, він має практично повний (системний) доступ до комп'ютера. Наступні дії та збиток від них можуть бути такими: впровадження троянської програми, впровадження набору утиліт для приховання факту компрометації системи, несанкціоноване копіювання зловмисником даних із твердих і змінних носіїв інформації системи, створення на віддаленому комп'ютері нових облікових записів з будь-якими правами в системі для наступного доступу як віддалено, так і локально, крадіжка файлу з хешами паролів користувачів, знищення або модифікація інформації, здійснення дій від імені користувача системи.

Протидія**.** МЕ й СОВ, встановлені на системі, що атакується, у ряді випадків не в змозі відбити дію експлойтів [17]. Для успішного відбиття атак експлойтів необхідно оновлювати засоби захисту, оскільки механізм виявлення вторгнень заснований на розпізнаванні сигнатур уже відомих атак. Хоча існують розробки, здатні, за заявами розробників, відбивати невідомі атаки, практика показує, що вони все ще не ефективні.

Шкідливі програми (Шпр). Шпр – це комп'ютерна програма чи переносний код, призначений для реалізації загроз інформації, що зберігається в мережі, або для прихованого нецільового використання ресурсів чи іншого впливу, що перешкоджає нормальному функціонуванню мережі. До Шпр відносяться комп'ютерні віруси, трояни, мережеві хробаки й т.ін. [18].

Типовим методом протидії є використання антивірусних засобів, що працюють у режимі реального часу (моніторів). Для виявлення троянських програм існує спеціалізоване програмне забезпечення.

1.3 Аналіз проблем забороненого контенту

.

Залежно від законодавства країни різні матеріали можуть вважатися нелегальними. У більшості країн заборонені: матеріали сексуального характеру за участю дітей і підлітків, порнографічний контент, описи насильства, у тому числі сексуального, екстремізм і розпалення расової ненависті.

Аналогічно з концепцією забезпечення комплексного захисту об’єкта інформатизації, можна сформувати повний набір функцій захисту від забороненої інформації.

Під функцією захисту (ФЗ) розуміється сукупність однорідних у функціональному відношенні заходів, регулярно здійснюваних в автоматизованих системах різними засобами й методами з метою створення, підтримки й забезпечення умов, об’єктивно необхідних для надійного захисту інформації [19].

Перелік повного набору функцій захисту від забороненої інформації в соціальних мережах:

– попередження умов виникнення забороненої інформації. Функція реалізується за допомогою нормативно-правових актів. Вона не може повністю виключити загрозу поширення забороненої інформації в соціальних мережах, тому що, в цілому, ситуація з дотриманням законів незадовільна, а в інтернет-просторі загострюється через технічні складності;

– попередження безпосереднього прояву забороненої інформації – функція реалізується за рахунок механізмів прогнозування поширення забороненої інформації в соціальній мережі. Більш докладно дана функція буде розглянута нижче.

– виявлення забороненої інформації, що з'явилася – функція пов'язана з моніторингом ІТКМ на предмет забороненої інформації на сторінках абонентів. Як правило, для реалізації даного захисту використовується різні СОРМ. Дана ФЗ пов'язана із проблемами контекстного пошуку, а також необхідністю контролю над усією системою.

– попередження впливу на абонентів забороненої інформації, що з'явилася – функція може бути реалізована за допомогою автоматичного розсилання повідомлень із попередженням про відповідальність за поширення забороненої інформації, аж до блокування абонента.

На рисунку 1.4 наведені всі комбінації подій, які потенційно можливі при здійсненні всіх ФЗ.

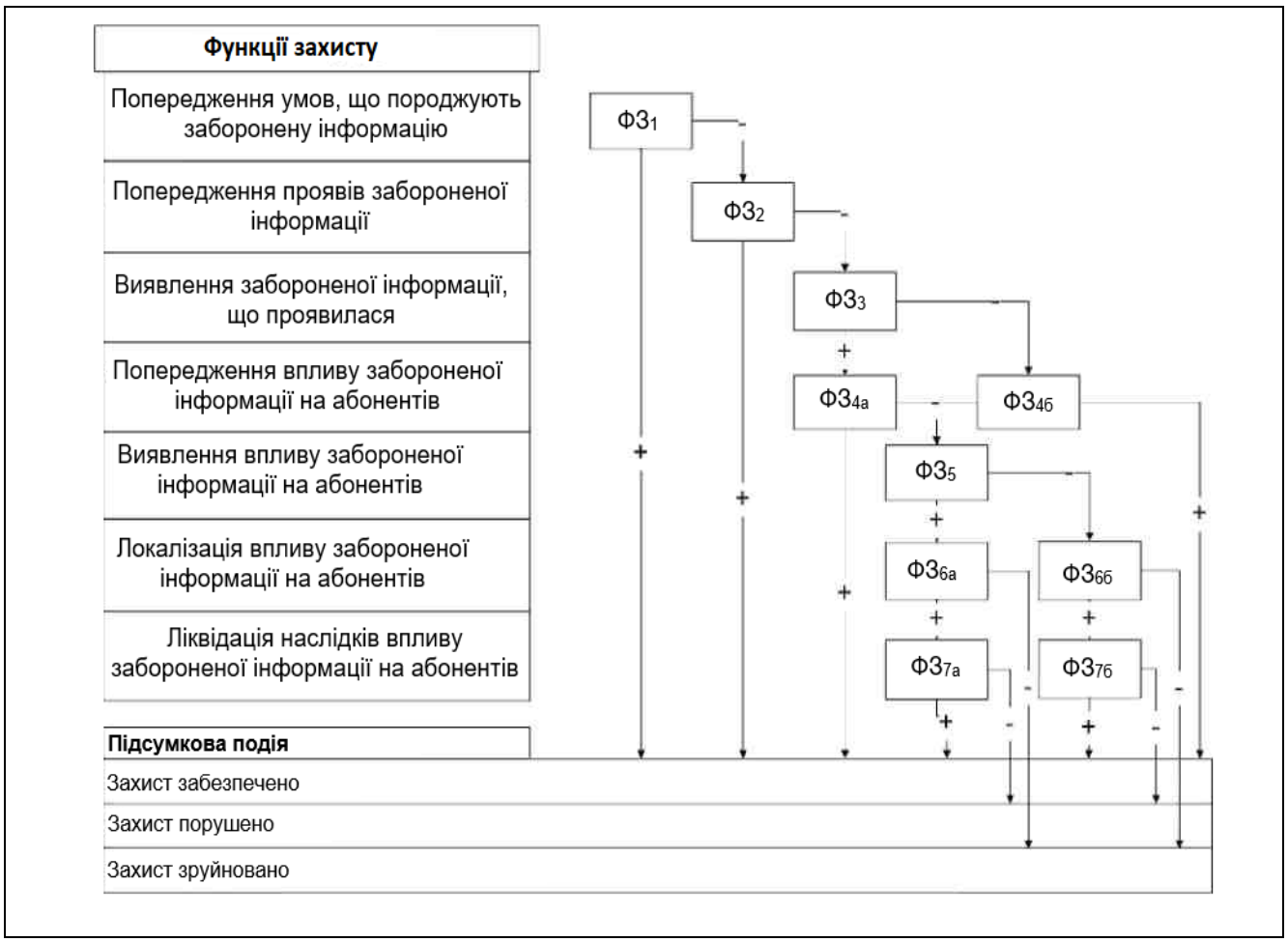

Рисунок 1.4 – Функції захисту від забороненої інформації в ІТКМ

Блокування може здійснюватися легітимними засобами при наявності доступу до управління системи й нелегітимними – при його відсутності (злом

акаунту). ФЗ ділиться на дві функції ($ФЗ_{4а}$ й $ФЗ_{4б)}$. Перша пов'язана з попередженням абонентів, на сторінках яких була знайдена заборонена інформація, а друга – з розсиланням попереджень потенційним одержувачам забороненої інформації. Виявлення впливу забороненої інформації на абонентів. Функція пов'язана безпосередньо з фіксацією процесу поширення забороненої інформації, може бути реалізована через контекстний аналіз повідомлень. Властиві такі ж недоліки, як і для $ФЗ_3$ [20].

Локалізація, обмеження впливу забороненої інформації на абонентів. Функція реалізується через блокування абонентів, що поширюють заборонену інформацію ($ФЗ_{6а}$), або абонентів – потенційних розповсюджувачів ($ФЗ_{6б)}$. Дана ФЗ спирається на попередні функції й для її ефективної реалізації необхідний контроль над системою.

Ліквідація наслідків виявленого впливу забороненої інформації на абонентів. Функція пов'язана з видаленням забороненої інформації із системи. Для реалізації даної функції також необхідний контроль над системою.

З аналізу функцій захисту видно, що найбільш ефективні функції – це перші функції, тому що вони забезпечують захист на ранніх етапах. Усі функції мають свої недоліки.

Найбільш перспективною ФЗ інженерно-технічного напрямку є $ФЗ_2$. На даному етапі, маючи інформацію про топологію ІТКМ і потенційних розповсюджувачах забороненої інформації, можливе прогнозування процесу її поширення [21] .

Створення моделей і алгоритмів поширення загрози забороненої інформації – одне із ключових завдань у даному напрямку. При його вирішенні виникають проблеми, пов'язані із властивостями розглянутої інформаційно-телекомунікаційної системи, а саме:

– відсутність перевірки реальності даних про вузол системи, дуже часто абоненти ІТКМ вказують недостовірну інформацію про себе;

– закритість системи – структура й інформація про керування системою є конфіденційною інформацією;

– проблема збору інформації – неможливо одержати повну інформацію про топологію ІТКМ. Існує можливість для звичайного абонента збору інформації про структуру мережі (функції API), але ця можливість має багато обмежень (налаштування приватності, часовий інтервал) [22].

У рамках даної роботи розглядається тільки обмін повідомленнями між абонентами, тому концептуальна математична модель інформаційної взаємодії представляється графом, вузлами якого є абоненти, а ребрами – зв’язки між ними. Перелічимо властивості графа, принципові для даного дослідження:

– велика розмірність – система містить мільйони елементів;

– гетерогенність – у графі, який відображає взаємозв’язок елементів у системі, вершини мають різну кількість прилягаючих ребер;

– динаміка зв’язків – у системі протягом часу відбуваються зміни зв’язків;

– динаміка вузлів – протягом часу змінюється кількість вузлів (елементів) системи;

– наявність груп вузлів, що мають велику кількість зв’язків усередині й невелике – між групами.

Граф, що представляє систему, має певну кластеризацію. Для таких систем характерно, що два вузли, що мають зв’язки до якого-небудь вузла, часто також мають зв’язок між собою.

1.4 Моделювання ІТКМ

Найбільш ефективне прогнозування поширення загрози забороненої інформації здійснюється за допомогою моделювання даного процесу. Таким чином, ми приходимо до завдання моделювання ІТКМ за допомогою її математичної моделі (графів).

Один з основних способів вивчення ІТКМ – моделювання, яке прийнято розглядати у двох аспектах. Перший стосується моделювання топології

(структури інформаційних зв'язків між вузлами мережі) ІТКМ, а другий – проблеми вивчення процесів, що проходять у ній. У нашому випадку це загроза поширення забороненої інформації (ЗПЗІ).

З погляду топології ІТКМ відносять до складних мереж [23]. Складні мережі (комплексні мережі, complex networks) – це існуючі в природі мережі, що володіють нетривіальними топологічними властивостями.

Наведено одну із прийнятих класифікацій топологічних моделей мереж (див. рисунок 1.5) [24]. В овалах зазначені класи мереж, а в прямокутниках конкретні речники-моделі-представники. Описуються їх характеристики: розподіл степенів зв'язності вузлів мережі, кластерний коефіцієнт і середня довжина шляху мережі.

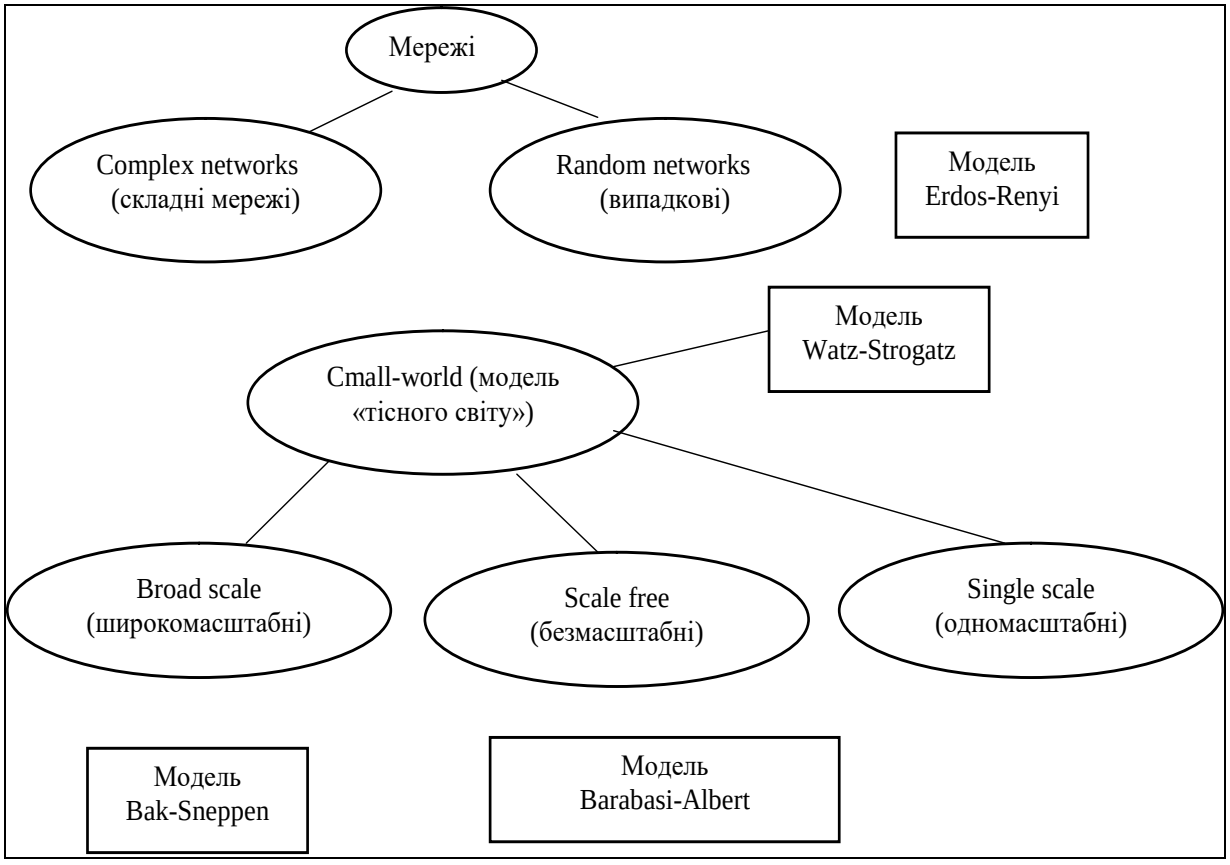

Рисунок 1.5 – Класифікація мереж [24]

Тема випадкових графів (мереж) розкрита в [25] описуються основні моделі мереж і їх основні характеристики. Проводиться дослідження топології популярних ІТКМ і здійснюється пошук найбільш адекватної топологічної

моделі. Представлений огляд канонічних робіт з даної тематики. Виділені головні сучасні тенденції в області аналізу топології ІТКМ:

– дослідження топологічних характеристик ІТКМ;

– дослідження еволюції ІТКМ;

– вивчення й розробка методів для обчислення характеристик великомасштабних ІТКМ, розв'язання проблеми одержання репрезентативної вибірки з ІТКМ.

ІТКМ із погляду маркетингових стратегій на їхній основі часто відносять до scale-free (SF) мережам [26]. Перші роботи належать Barabasi і Albert та присвячені однойменній моделі. Вони порівнюють існуючі моделі й свою модель, докладно її описують та наводять її сильні й слабкі сторони.

Рішення такої задачі полягає в тому, що знаходиться розподіл зв'язності вузлів і деякі інші пов'язані з ним параметри мережі. Також показане, що масштабованість, яка виникає, дійсна не тільки для даної моделі, але й для широкого класу зростаючих мереж. Також може бути вирішенням і отримання універсальних відношень масштабування, які описують властивості scale-free мереж, що розвиваються, і вказують межі їх дії. Дано доказ того, що основні властивості SF мереж, що розвиваються, можуть бути описані в рамках аналітичної моделі.

В наш час розглядається не тільки топологія ІТКМ як SF мережи, але аналізується процес поширення епідемії на таких мережах [27]. Отримані дані по комп'ютерних вірусах і виявлено такі їхні параметри, як середня «тривалість життя» вірусу й стійкість до знищення. Описано динамічну модель поширення інфекції в мережах. Навели метод визначення наявності епідемічного порога в мережі.

Ще однієї точкою зору на тип топології ІТКМ, що ІТКМ топологічно являє собою клас small-world мереж. Модель Watts-Strogatz, яка відноситься до класу small-world мереж, імітує структуру ІТКМ. Розглянута проблема перколяції вузлів на small-world мережах. Цей підхід дозволяє розглядати просту модель поширення захворювань (SIS) і одержати апроксімований вираз для порога

перколяції. Усі аналітичні результати підтверджуються чисельними розв'язаннями моделі. Small-world мережі розглядаються в багатьох роботах не тільки з точки зору топології мережі, але і як основа для епідеміологічних моделей [28].

Broad Scale мережі розглядаються як аналіз Bak-Sneppen моделі. Даний вид складних мереж найменш привабливий при моделюванні топології ІТКМ.

Аналіз наукових праць, у яких розглядаються різні підходи до моделювання топології ІТКМ, показує, що при розв'язанні даного завдання, як правило, використовуються small-world і scale free мережі.

1.5 Моделювання процесів інформаційної взаємодії в ІТКМ

При розгляді питань, що стосуються моделювання процесів, які протікають в ІТКМ, основним підходом є застосування моделей впливу, інформаційного керування й протиборства [25]. Моделі впливу найбільш адаптивні до розв'язуваних задач. На рисунку 1.6 представлена узагальнена класифікація моделей впливу. Коротко охарактеризуємо представлені класи моделей впливу.

Граничною моделлю є будь-яка модель, у якій є граничне значення або набір граничних значень, що використовуються при зміні станів. Класичні моделі з порогами були розроблені Schelling, Axelrod і Granovetter для моделювання колективної поведінки [25].

Моделі незалежних каскадів (Independent Cascade Model) належить категорії моделей так званих «систем взаємодіючих часток» (Interacting Particle Systems). Вузол мережі (агент) визначається аналогічно до моделі, що описана вище. Коли агент i стає активним у деякий момент часу, він отримує шанс активувати на наступному (і тільки на наступному) кроці кожного зі своїх сусідів j з імовірністю p_{ji} (причому j можуть намагатися незалежно активувати й інші агенти) [26].

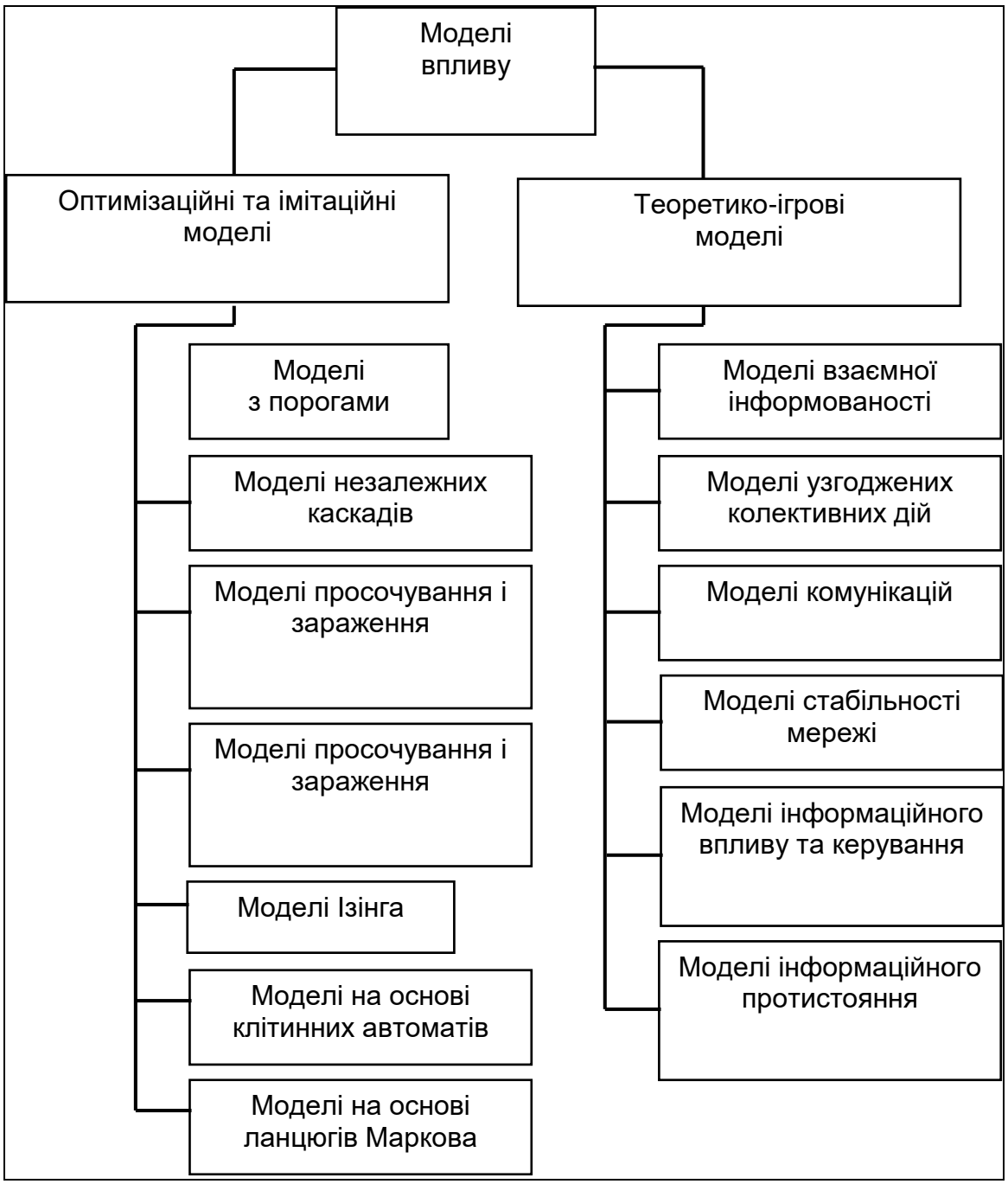

Рисунок 1.6 – Класифікація моделей впливу

Моделі просочування й зараження є популярним способом вивчення поширення інформації й інновацій у соціальних системах.

Модель Ізінга – це математична модель, що описує виникнення намагнічування матеріалу. Конформність або незалежність у великій соціальній групі може моделюватися за допомогою моделі Ізінга; вплив найближчих сусідів є визначальним, а аналогом температури є готовність групи мислити творчо, готовність прийняти нові ідеї. Зовнішнім полем для соціальної групи є вплив

«авторитету» або керування. Більш складні моделі, що описують ІТКМ на термодинамічних аналогіях, розглядалися в [20].

Для опису процесів поширення інформації в ІТКМ останню можна розглядати як складну адаптивну систему, що складається з великої кількості агентів, взаємодія між якими приводить до масштабної, колективної поведінки, яку складно передбачити й аналізувати. Для моделювання й аналізу таких складних систем іноді використовуються клітинні автомати. Клітинний автомат складається з набору об’єктів (у цьому випадку агентів), що скінчено утворюють регулярні решітки. Стан окремо взятого агента в кожний дискретний момент часу характеризується деякою змінною. Стани синхронно змінюються через дискретні інтервали часу відповідно до незмінних локальних імовірнісних правил, які можуть залежати від станів найближчих сусідніх агентів в окраїні даного агента, а також, можливо, від стану самого агента.

Також представлена модель ланцюгів Маркова, у якій вивчається вплив у команді (групі агентів). Запропонована модель є динамічною байєсовською мережею (Dynamic Bayesian Network – DBN) із дворівневою структурою: рівнем індивідів (моделюються дії кожного агента) і рівнем групи (моделюються дії групи в цілому).

Моделі взаємної поінформованості [26] – є агент, що входить у деяку соціальну мережу. Агент інформований про поточну ситуаційну обстановку (діях і представленнях інших агентів, параметрах середовища – так званому стані природи (state of nature) й ін.). Ситуаційна обстановка впливає на наявний у агента набір цінностей, установок і представлень, пов’язаних у такий спосіб: цінності впливають на установки, а ті, у свою чергу, призводять до схильності представлень того або іншого рівня, із схильностями погоджена, що перебуває «у пам’яті» агента ієрархічна система уявлень про світ. Схильність до тем або іншим представленням і ситуаційна обстановка (наприклад, дії інших агентів) призводять до формування нових або модифікації старих представлень. Відповідно до цих представлень і встановленою метою агент ухвалює рішення й

виконує дію. Результати дій приводять до зміни як самої ситуаційної обстановки, так і внутрішніх цінностей, установок і представлень.

Моделі погоджених колективних дій. Ключове значення тут мають соціальні зв'язки. З одного боку, соціальні зв'язки можуть забезпечити ефективний локальний соціальний контроль для стимулювання участі в колективній дії (у силу тиску з боку своїх сусідів, довіри до них, соціального схвалення, необхідності збереження позитивних відносин і відповідності очікуванням, емоційній прихильності, збереження своєї репутації, ототожнення себе із сусідами й інше). Так, наприклад, поведінка сусідів агента вплине на його власну поведінку. З іншого боку, соціальні зв'язки забезпечують агента інформацією про наміри й дії інших агентів у мережі й формують його (неповні) уявлення, на основі яких агент ухвалює свої рішення. І, нарешті, у межах соціальних зв'язків агенти можуть прикладати спільні зусилля по створенню локального суспільного блага й спільно користуватися ім. Тому структура ІТКМ впливає на рішення агентів про прийняття участі в колективній дії.

ІТКМ може розглядатися як комунікаційна, за допомогою якої агенти повідомляють один одному про свою готовність взяти участь у колективній дії. Кожний агент інформований про готовність тільки своїх найближчих сусідів і на основі цього локального знання ухвалює рішення про участь, використовуючи правило прийняття рішень «я візьму участь, якщо візьмеш участь ти» (механізм координації). Тобто розглядається координаційна гра з неповною поінформованістю. Комунікаційна мережа сприяє координації, і основний інтерес становить те, які властивості таких мереж, що допускають колективну дію. Розглядаються мінімально достатні мережі, які вишиковують агентів в ієрархію соціальних ролей/щаблів: «провідні» (initial adopters), «послідовники» (followers) і т.ін. до «пізніх послідовників» (late adopters). Такі мережі сприяють координації в наступний спосіб:

- інформуючи кожний щабель про більш ранні щаблі;
- формуючи загальне знання в межах кожного щабля.

Таким чином забезпечується розуміння ролі (локально) загального знання в колективній дії й співвідношення між структурою соціальної мережі й загальним знанням.

Рівновага стабільної мережі (stable network equilibrium) [27] – ситуація, у якій не існує агента, для якого будь-яка комбінація зміни його дії й зміни його зв'язків приведе до кращого результату. Тільки рівноваги з повною участю або повною неучастю – є рівновагами стабільної мережі.

1.6 Аналіз алгоритмів епідеміологічних моделей

Дослідники проводять аналогію між біологічними й комп'ютерними вірусами й розглядають адаптацію методів математичної епідеміології до вивчення комп'ютерних вірусів. Розглядаються стандартні епідеміологічні моделі на орієнтованому графі, використовується моделювання для вивчення поширення вірусів. Значна увага приділяється вивченню критичного порога епідемії. Розглядаються моделі поширення інфекційних захворювань серед населення, проводиться їхній математичний аналіз і застосування до конкретних захворювань. Розглядається класична епідеміологічна SIR модель Кермака-МакКендрика, MSEIR і SEIR ендемічні моделі. Також розглядається епідеміологічні моделі поширення вірусів і боротьби з ними. Представлена нова модель, яка може бути використана для прогнозування процесу поширення шкідливих програм і оцінки ефективності протидії їм. Показано, як застосовується модель для аналізу динаміки системи, інфекційних спалахів і інших процесів, пов'язаних з поширенням вірусів.

В представленні аналізу динаміки розвитку епідемії в складних гетерогенних мережах наводяться аналітичні й чисельні результати, розглядається вплив початкових умов і актуальність статистичних результатів дослідження, що

стосується гетерогенних мереж. Представлені теоретичні відомості становлять великий інтерес і можуть дати корисну інформацію для розробки стратегій, спрямованих на адаптивне стримування епідемії. Також розглядають вірусний маркетинг в ІТКМ.

Вірусний маркетинг – загальна назва різних методів поширення реклами, що характеризуються поширенням у прогресії близької до геометричної, де головним розповсюджувачем інформації є самі одержувачі інформації. Здійснюється даний підхід шляхом формування вмісту повідомлення, таким чином, який здатний залучити нових одержувачів інформації за рахунок яскравої, творчої, незвичайної ідеї. Також ефективність повідомлення ґрунтується на використанні природних довірливих відносинах між одержувачем і відправником.

У межах розв’язуваних завдань для нас найбільше підходять оптимізаційні й імітаційні моделі. З них розглянемо моделі просочування й зараження (клас епідеміологічних моделей), тому що дані моделі найбільше точно відбивають специфіку розглянутих нами проблем. Даний клас моделей є дуже розповсюдженим при дослідженнях процесів взаємодії в ІТКМ.

1.7 Постановка завдань дослідження

ІТКМ є великомасштабними мережами з постійно зростаючим числом абонентів. З бурхливим ростом числа користувачів ІТКМ виникають проблеми інформаційної безпеки й захисту інформації в них. Аналіз проблем інформаційної безпеки виявив, що окрім проблем, пов’язаних з використанням глобальної мережі Інтернет як розподіленої інформаційно-телекомунікаційної системи, які досить добре відомі й розв’язані, існує маловивчена проблема забороненого контенту.

Створення моделей і алгоритмів поширення загрози забороненої інформації – один із ключових підходів при вирішенні даного завдання. Проведений аналіз

публікацій по даній тематиці показує, що існуючі рішення малоефективні. Зазвичай при моделюванні поширення загрози забороненої інформації не враховується топологія ІТКМ (модель мережі – повнозв'язний граф).

Якщо топологія враховується, то, як правило, використовується найпростіша SIS модель, а структура мережі відображається SF мережею. При моделюванні ЗПЗІ важливо мати топологію, що відображає структуру зв'язків реальної мережі, а також використовувати адекватну модель інформаційної взаємодії вузлів.

Ще однією важливою проблемою є великомасштабність ІТКМ, яка заважає одержати дані з імітаційної моделі за прийнятний час. Розв'язання цього завдання полягає в створенні аналітичної моделі ЗПЗІ в ІТКМ.

Після проведення аналізу предметної області в рамках даної роботи були поставлені наступні завдання дослідження:

Створити імітаційну модель поширення загрози забороненої інформації в ІТКМ:

– розробити алгоритм ЗПЗІ в ІТКМ;

– на основі розробленого алгоритму створити імітаційну модель ЗПЗІ в ІТКМ;

– провести моделювання дослідження імітаційної моделі ЗПЗІ в ІТКМ.

Створити аналітичну модель поширення загрози забороненої інформації в ІТКМ:

– на основі експериментальних даних за імітаційною моделлю створити аналітичну модель ЗПЗІ в ІТКМ;

– провести дослідження аналітичної моделі, перевірити адекватність моделі.

Розробити алгоритм формування топології ІТКМ:

– алгоритму формування графа доступної частини мережі;

– алгоритму формування повного графа.

Змоделювати процес поширення загрози забороненої інформації на реальній великомасштабної ІТКМ:

– розробити алгоритм формування топології великомасштабної ІТКМ;

– реалізувати алгоритми у вигляді ПЗ;

– розробити ПЗ під розподілену обчислювальну систему для моделювання ЗПЗІ на топології великомасштабної ІТКМ;

– провести експериментальне дослідження імітаційної моделі ЗПЗІ на топології великомасштабної ІТКМ із використанням розробленого ПЗ;

– провести експериментальне дослідження з отриманих результатів.

2 ОПИС ПРОВЕДЕНИХ ТЕОРЕТИЧНИХ ДОСЛІДЖЕНЬ

2.1 Алгоритми імітаційного моделювання

За результатами огляду предметної області були поставлені завдання створення імітаційної й аналітичної моделей поширення загрози забороненої інформації в ІТКМ. Імітаційна модель необхідна для одержання експериментальних результатів для синтезування аналітичної моделі. Необхідність створення аналітичної моделі обґрунтовується тим, що для імітаційного моделювання на топології існуючих ІТКМ (десятки мільйонів вузлів) необхідні значні часові витрати. Не враховуючи час на збір інформації про топологію мережі, який може становити більше тижня, безпосереднє моделювання ЗПЗІ займає кілька годин навіть при використанні розподілених обчислювальних ресурсів. Аналітична модель може дати прогноз ЗПЗІ майже миттєво. З її допомогою можна одержати актуальні дані (до того моменту, коли кількість атакуючих абонентів буде максимальним) за динамікою ЗПЗІ.

Процес ЗПЗІ характеризується наступними особливостями [1], [7]: у мережі існують вузли трьох типів. Перший тип – атакуючі вузли – це вузли, що поширюють заборонену інформацію. Другий тип – захищені вузли, що характеризуються тим, що не беруть участь у поширенні забороненої інформації й ніколи не будуть цим займатися. Третій тип – потенційно вразливі. Вузли такого типу не беруть участь у процесі поширення загрози, але можуть бути піддані негативному впливу з боку атакуючих вузлів і можуть почати поширювати заборонену інформацію.

Нехай дано: N – кількість вузлів, рівна числу абонентів мережі, I_0 – кількість абонентів-зловмисників – постійних джерел загрози, R_0 – кількість абонентів постійно несприйнятливих до атакуючих впливів, β -параметр, що відображає потужність загрози, імовірність здійснення атаки, γ -параметр, що відображає ступінь протидії загрозі, ймовірність захисту абонента β та γ у визначені як константи, але можуть бути виражені як функції, що залежать від профілів

абонентів ІТКМ), φ – коефіцієнт топологічної уразливості мережі, що відображає внутрішню властивість ІТКМ, заснований на характеристиках її топології, який сприяє поширенню забороненої інформації , t – час процесу (в умовних одиницях часу).

Потрібно розробити аналітичну модель динаміки атаки $I(t)$ і захисту вузлів $R(t)$:

$$\begin{cases} I(t) = f(N, \beta, \gamma, \varphi, t) \\ R(t) = g(N, \beta, \gamma, \varphi, t) \end{cases}.$$

Наведено алгоритм реалізації ЗПЗІ, що ґрунтується на описі процесів, що протікають у реальних ІТКМ. Схема реалізації загрози представлено на рисунку 2.1.

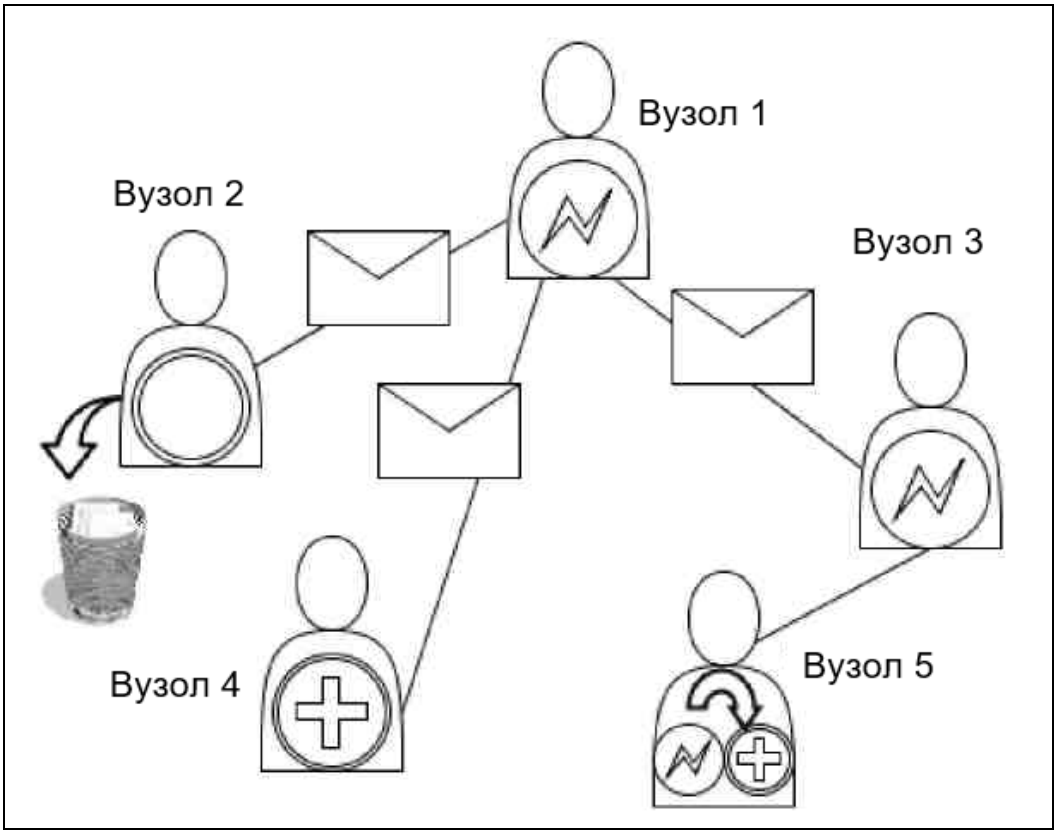

Рисунок 2.1 – Схема реалізації ЗПЗІ

Методика розробки аналітичної моделі містить у собі послідовність наступних дій:

– формування імітаційної моделі для дослідження характеру й параметрів процесу ЗПЗІ;

– синтез аналітичних залежностей параметрів процесу;

– проведення експериментів з метою перевірки точності (адекватності) моделі.

Алгоритм ЗПЗІ в ІТКМ.

Крок 1. Поширення забороненої інформації (ЗІ) (далі процес «атаки») ініціює який-небудь абонент-зловмисник (на рисунку – вузол 1), поширюючи повідомлення із ЗІ (реалізує загрозу) за його списком контактів. Атаку може починати один зловмисник або група.

Крок 2. Абоненти-ідержувачі (вузли 2,3,4), прийнявши повідомлення із ЗІ, читають його й включаються в процес атаки, поширюючи її далі за своїм списком контактів (вузол 3), або ігнорують чи взагалі видаляють повідомлення (вузол 2), тобто в атаці не беруть участь. Процес атаки звичайно йде лавино подібно. Абоненти, що атакують, не закінчують атаку, один раз передавши повідомлення із забороненою інформацією. Вікно атаки, як правило, триває протягом досить значного проміжку часу й залежить від типу подачі ЗІ в повідомленні, зацікавленості абонента й інше.

Крок 3. Абоненти можуть перестати сприймати й, відповідно, поширювати ЗІ (вузол 5) (далі процес «захисту»), внаслідок впливу механізмів захисту (наприклад, попередження про неї), тому повідомлення із ЗІ від атакуючих абонентів будуть постійно відкидатися.

Крок 4. Процес триває доки в мережі є абоненти-зловмисники, або є потенційно вразливі вузли, якщо відсутній процес захисту.

Таким чином, ЗПЗІ в ІТКМ являє собою складний динамічний процес, що складається із двох протиборчих підпроцесів атаки й захисту вузлів мережі.

На основі описаного алгоритму може бути побудовано імітаційну модель ЗПЗІ в ІТКМ, яка складається з розробленої програми Modelgraph і даних, які можуть бути сгенеровані за допомогою ПЗ Pajek [36].

Імітаційна модель ЗПЗІ. Вхідні дані: N, k – середній ступінь зв'язності вузлів, α – параметр, що відображає середню довжину шляху й рівень мережевої кластеризації, β , γ (у моделі вважається, що β, γ однакові для кожного абонента), I_0, R_0.-

Вихідні дані: $I(t)$, $R(t)$, $S(t)$ – чисельні масиви даних, що описують динамічний процес реалізації ЗПЗІ (кількості атакуючих, захищених і потенційно вразливих вузлів у кожну умовну одиницю часу відповідно).

Крок 1. Створення топології ІТКМ – граф $G_{sw} = <V, E>$, де G_{sw} – граф small-world мережі (на основі моделі Watts-Strogatz), $V = \{v_i\}$ – множина вершин, $E = \{e_{ij}\}$ – множина ребер, $i=1...N$, $j=1...N$. Даний крок здійснюється з використанням вільно розповсюджуваної програми Pajek, адаптованої для заданого завдання, за рахунок топологічних параметрів N, k, a, що задаються.

Крок 2. Сформувати множину $V=\{V^I, V^S, V^R\}$, де $V^I = \{v^I_i\}$ – множина атакуючих вузлів ($|V^I| = I_0$, $V^R = \{v_i^R\}$ – множина захищених вузлів ($|V^R|=R_0$), $V^S = \{v_i^S\}$ – множина потенційно вразливих вузлів ($|V^S| = N - I_0 - R_0$).

Крок 3. Для кожного v_i^I якщо існує e_{ij} й $v_j \in V^S$ $j = 1...N$, то з імовірністю β виконати: $V^S \backslash v_j$ і $V^I \cup v_j$; з імовірністю γ виконати: $V^I \backslash vi$, $V^R \cup v_i$.

Крок 4. Якщо $V^I = \varnothing$ або $\gamma=0$ і $V^S=\varnothing$, то кінець алгоритму, інакше перейти до кроку 3.

Програма Modelgraph –для імітаційного моделювання ЗПЗІ в ІТКМ [8]. Даний програмний продукт є однопотоковим додатком. Програма складається з файлу, що виконується, Modelgraph.exe і бібліотеки chartdir50.dll для побудови графіків. Після вибору типу мережі й уведення її параметрів відбувається імітаційне моделювання за алгоритмом 2.1. Потім результати відправляються у функцію побудови графіків для представлення результатів у графічному вигляді. Програма написана в середовищі розробки Microsoft Visual Studio .NET 2018. Вихідними даними для гетерогенної мережі є файл формату .net, визначений у програмі Pajek.

ПЗ Pajek представляє собою програму, для ОС MS Windows, призначену для аналізу й візуалізації великих мереж. Дана програма перебуває у вільному доступі й призначена для некомерційного використання.

Проаналізовано підпроцес атаки без захисту, провівши ряд експериментів (експеримент 1-3) з використанням імітаційної моделі , якщо φ –коефіцієнт топологічної вразливості мережі).

Вплив сили атаки на процес. Експеримент проводився для наступних значень параметрів: N=1000, φ =20, I_0=1, β = 0,1…0,9 (рисунок 2.2).

Вплив значення середнього ступеня зв'язності вузлів у мережі на процес.

Експерименти проводилися за наступними значеннями параметрів: N=1000, φ = 0,5.. 60, I_0=1, β=0,5.

Вплив кількості споконвічно атакуючих вузлів на процес.

Експерименти проводилися за наступними значеннями параметрів: N=1000, φ=20, I_0=1…40, β=0,5 .

Кожний із трьох типів експериментів проводився 100 разів, бралися усереднені значення.

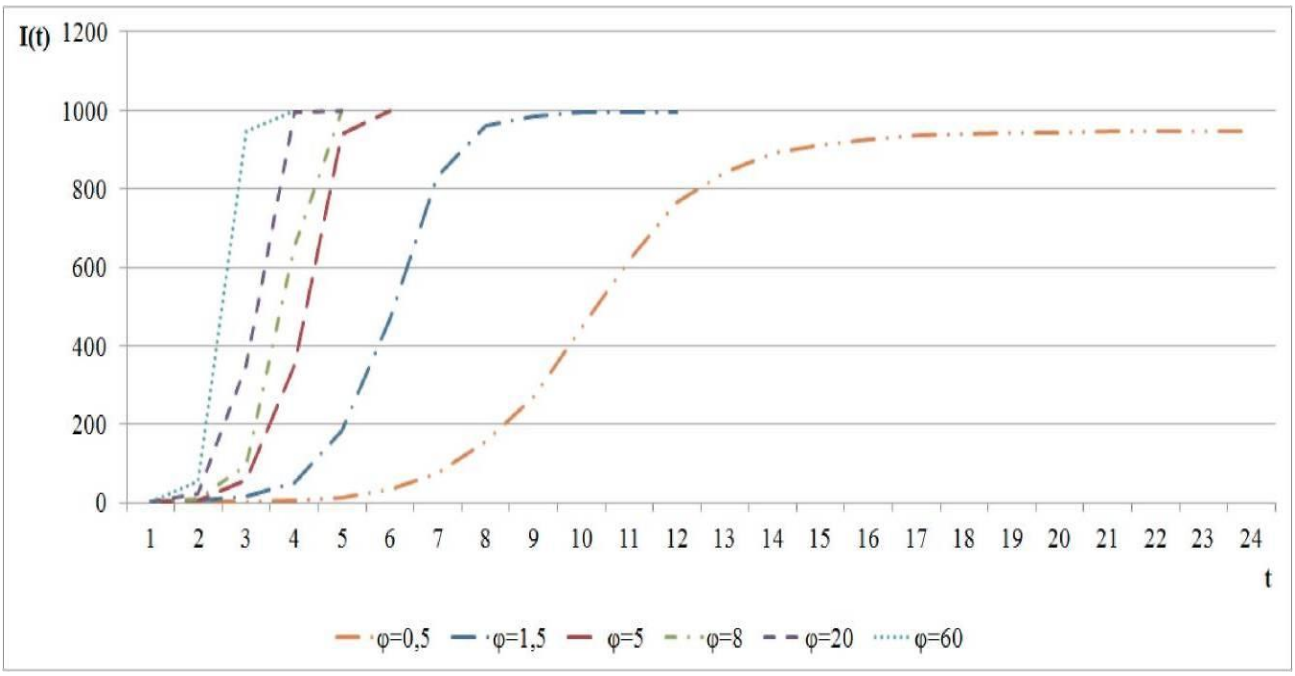

Рисунок 2.3 – Вплив φ на процес атаки

За результатами аналізу алгоритму можна зробити наступні висновки:

– процес атаки $I(t)$ має експонентну залежність;

– при збільшенні значень φ, I_0, β зростає динаміка зараження вузлів (інтенсивність атаки);

– при рості ймовірності проведення атаки β від 0,1 до 0,9, час процесу знижується у два рази (з 8 до 4 умовних одиниць часу) (див. рисунок 2.3);

– коефіцієнт топологічної уразливості φ має найбільший вплив (у порівнянні з I_0, β) на тривалість процесу. Наприклад, при $\boldsymbol{\varphi}$ = 0,5 (низька вразливість) атака триває 24 умовні одиниці часу, а при $\boldsymbol{\varphi}$ = 60 – 4;

– велика кількість вузлів, що атакують спочатку I_0 знижує час, за який відбувається зараження всіх вузлів у мережі. Наприклад, при I_0=40 тривалість процесу становить 3 умовні одиниці часу.

Додавши підпроцес захисту, який залежить від початкової кількості захищених вузлів R_0 та ймовірності захисту β.

Вплив ймовірності захисту. Експерименти проводилися за наступними значеннями параметрів: $N=1000$, $\varphi = 20$, $I_0=1$, $\beta=0,5$, $\gamma =0,1..0,9$, $R_0 =0$. (рисунок 2.4).

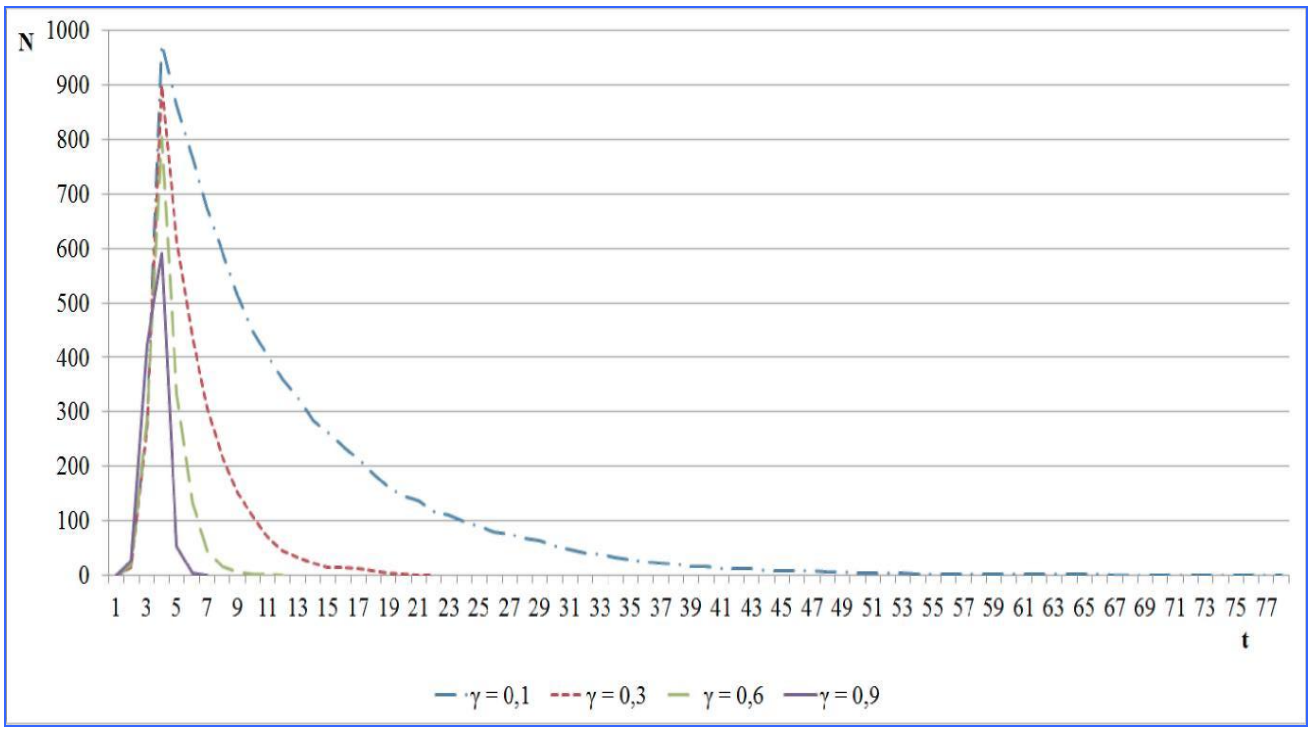

Рисунок 2.4 – Розрахунок впливу γ на процес атаки

Вплив початкової кількості захищених вузлів. Експерименти проводилися за наступними значеннями параметрів: N=1000, $\varphi = 20$, $I_0=1$, $\beta=0,5$, $\gamma=0,5$, $R_0=0..200$ (див. рисунок 2.5).

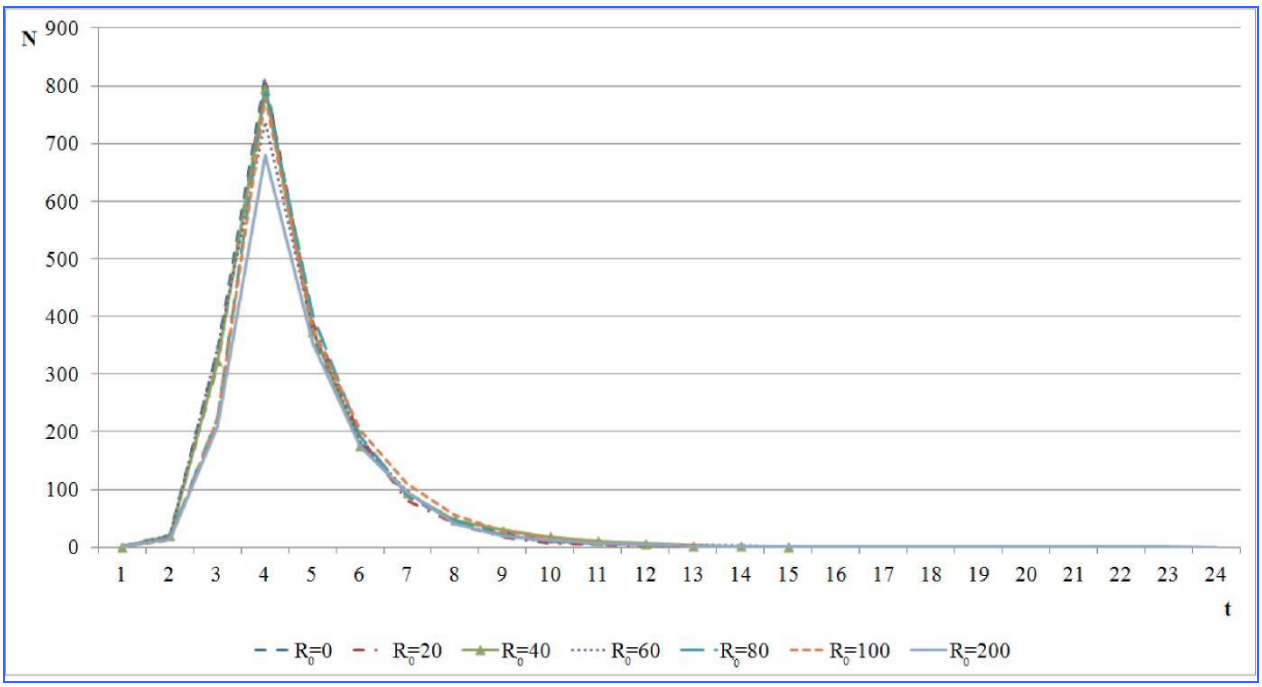

Рисунок 2.5 – Розрахунок впливу R_0 на процес атаки

За результатами можна зробити наступні висновки:

– введення підпроцесу захисту збільшує час всього процесу ЗПЗІ;

– при невеликих значеннях ймовірності захисту (γ<0,3) загроза реалізується практично на всіх вузлах у мережі (див. рисунок 2.4);

– при невеликих значеннях ймовірності захисту (γ<0,3) час процесу становить більше 50 умовних одиниць часу (див. рис. 2.4);

– при великій ймовірності захисту (≈0,9) процес триває ≈7 умовних одиниць часу й максимальна кількість атакуючих вузлів знижується залежно від ймовірності проведення атаки;

– при випадковому виборі споконвічно захищених вузлів картина процесу атаки практично не змінюється;

– при високій топологічній вразливості зростає тривалість процесу ЗПЗІ.

2.2 Розробка аналітичної моделі

Аналізуючи процес інформаційної взаємодії абонентів під час поширення забороненої інформації в ІТКМ, можна зробити наступні висновки. Маємо справу із трьома типами абонентів: атакуючі абоненти, які поширюють заборонену інформацію, захищені абоненти, що характеризуються тим, що не беруть участь у поширенні забороненої інформації й ніколи не будуть цим займатися, і потенційно вразливі абоненти, які можуть бути піддані негативному впливу з боку атакуючих вузлів і можуть почати поширювати заборонену інформацію. При цьому ми спостерігаємо два протиборчі підпроцеси атаки й захисту абонентів мережі. Для моделювання таких явищ часто застосовують епідеміологічні моделі [28], зокрема попереднім умовам опису точно відповідає SIR-модель Кермака-МакКендрика [29]. Характер графіків, отриманих у результаті імітаційного моделювання (див. рисунок 2.6), схожий з результатами, яка дає дана модель.

Дана модель є найбільш релевантною в даному дослідженні. Умови $N=1000$, $\varphi = 20$, $I_0=1$, $\beta=0,5$, $\gamma =0,5$, $R_0 =10$, $S(t)$ – кількість вузлів моделі SIR, на які була здійснена атака (Susceptibles–Infectives–Removed with immunity), – епідеміологічна модель, спрощено описує поширення захворювання, що передається від одного індивіда до іншого, яка розглядає суб'єктів з погляду трьох можливих станів: сприйнятливий, інфікований, імунізований.

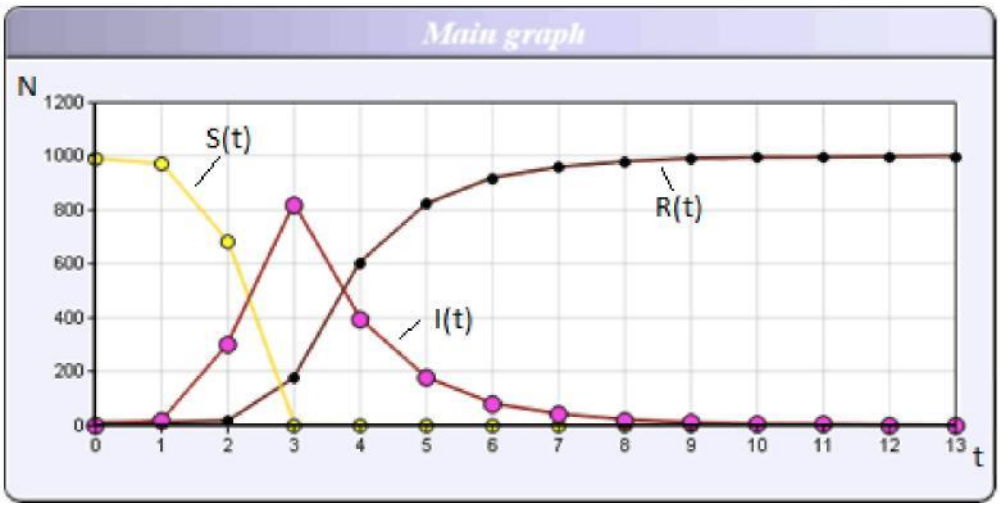

Рисунок 2.6 – Імітаційне моделювання

Система диференціальних рівнянь, що описують SIR-модель, має вигляд [30]:

$$\begin{cases} \frac{dI}{dt} = \beta \cdot \frac{S(t) \cdot I(t)}{N} - \gamma \cdot I(t) \\ \frac{dR}{dt} = \gamma \cdot I(t) \\ \frac{dS}{dt} = -\beta \cdot \frac{S(t) \cdot I(t)}{N} \end{cases}, \tag{2.1}$$

де $I(t)$ – кількість заражених (інфікованих) особин,

$S(t)$ – кількість сприйнятливих вузлів (особин),

$R(t)$ – кількість «виключених з імунізацією» (removed with immunity) вузлів,

$N=I(t)+S(t)+R(t)$ – кількість особин (вузлів мережі) в популяції,

γ – коефіцієнт відновлення/смерті,

β – швидкість зараження (інфікування),

t – час.

Дана система є надлишковою – будь-яке рівняння із трьох рівнянь можна виключити.

При використанні системи 2.1 для аналізу ЗПЗІ в ІТКМ одержано результати у вигляді графіків (рисунок 2.7), які хоча й правильно описують характер процесу, але не дають потрібної точності прогнозу.

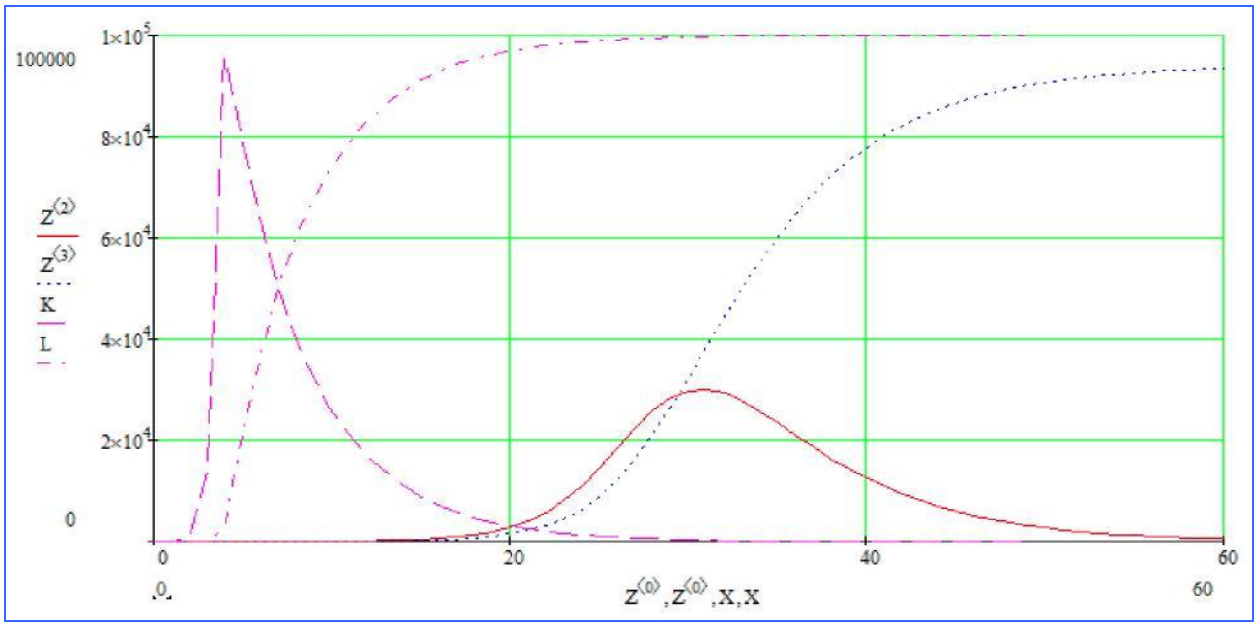

Рисунок 2.8 – Результати імітаційного моделювання

На рис. 2.7 N=100000, φ = 150, I_0=1, β=0,3, γ =0,2, R_0 =0) і аналітичного розв'язання ($Z^{<2>}$, $Z^{<3>}$ – аналітичний розв'язок для процесів атаки й захисту відповідно, K, L – результати імітаційного моделювання для процесів атаки і захисту відповідно)

Була висунута гіпотеза про те, що система 2.1 не дає потрібної точності у зв'язку з тим, що в моделі, яку вона описує, не враховуються топологічні особливості мережі. У зв'язку із цією гіпотезою було поставлено завдання адаптування системи 2.1 під прогнозування ЗПЗІ в ІТКМ шляхом інтегрування в неї параметра топологічної уразливості мережі φ.

Проаналізувавши графіки, отримані за результатами імітаційного моделювання й аналітичного розв'язання рівнянь (2.1), і простеживши фізичний вміст рівнянь у даній системі, можна прийти до наступного висновку.

Процес захисту не залежить від топології мережі, тому «змінювати» $R(t)$ не маємо права. А процес атаки залежить від структури зв'язків між абонентами в мережі. Параметр топологічної вразливості φ може впливати на $I(t)$ через коефіцієнт β.

У загальному виді адаптовану систему 2.1 для моделювання вразливостей мережі ІТКМ можна представити в наступному вигляді:

$$\begin{cases} \dfrac{dI}{dt} = C \cdot \beta \cdot \dfrac{S(t) \cdot I(t)}{N} - \gamma \cdot I(t) \\ \dfrac{dR}{dt} = \gamma \cdot I(t) \\ \dfrac{dS}{dt} = -C \cdot \beta \cdot \dfrac{S(t) \cdot I(t)}{N} \end{cases}, \tag{2.2}$$

де C – коефіцієнт, що залежить від параметра φ.

Аналіз топологій великомасштабних ІТКМ показав, що типові значення параметра φ для них перебувають у діапазоні від 100 до 600 [31].

3 АНАЛІЗ РЕЗУЛЬТАТІВ ДОСЛІДЖЕННЯ

3.1 Алгоритм збору даних про топологію доступної частини мережі

Під топологією будемо розуміти структуру інформаційних зв’язків між вузлами мережі. Топологічні характеристики (середній степінь зв’язності вузлів, розподіл степенів зв’язності вузлів, кластерний коефіцієнт мережі, середня довжина шляху мережі) у роботі розглядаються як основні технічні уразливості ІТКМ до реалізацій загроз. Інші уразливості: використання неліцензійного ПЗ у вузлах, некоректно налаштовані міжмережеві екрани й ін., у роботі не розглядаються.

Для моделювання ЗПЗІ необхідно мати топологію реального об’єкта. Пряме одержання цієї інформації ускладнене у зв’язку з наступним протиріччям. Для підвищення точності результатів моделювання необхідно мати топологію всієї мережі. Одержати таку інформацію без прав адміністратора не представляється можливим. Під час збору даних із правами абонента ІТКМ маємо справу із двома типами вузлів: відкритими й закритими. Якщо під час збору даних ми одержуємо ідентифікатори (id) вузла й суміжних з ним вузлів, то такий вузол називаємо відкритим. Якщо ж одержуємо тільки id вузла (абонент за допомогою налаштувань сховав інформацію про свої контакти), то такий вузол називаємо закритим. Також у мережі можуть існувати вузли, які з’єднані тільки із закритими вузлами. У такому випадку неможливо одержати навіть ідентифікатор вузла. Таких вузлів у мережі незначна частина. Емпірично показане [31], що закритих вузлів на порядок більше, ніж відкритих, тому під час збору даних втрачається значна частина даних.

Особливості практичної реалізації:

– частота запитів абонента про зв’язки вузла обмежена адміністраторськими заходами (наприклад, для обраної соціальної мережі це значення, наближено, становить 10 запитів на секунду). Це обмеження приводить до того, що, враховуючи масштабність ІТКМ (десятки мільйонів вузлів),

одержання інформації про топологію мережі перетворюється в тривалий процес. Враховуючи, що час сесії обмежений, дана особливість повинна враховуватися при практичній реалізації;

– відомі засоби (наприклад, Tictrac [30]) для вирішення завдання збору інформації про зв'язки вузлів в ІТКМ не ефективні, тому що прямо не призначені для досягнення цієї мети й мають множина недоліків;

– топологія реальної ІТКМ постійно змінюється (абоненти реєструються, додають зв'язки, видаляють зв'язки й облікові записи). У зв'язку із цим, необхідно постійно одержувати актуальну інформацію про ІТКМ для більш точного моделювання ЗПЗІ.

Топологія мережі представляється графом $G=\{V,E\}$, де V (множина вершин графа) – множина вузлів-абонентів, а E (множина ребер) – інформаційні зв'язки між вузлами.

Будемо вважати, що граф є неорієнтованим, тобто всі зв'язки – двоспрямовані. Будь-які дві вершини графа можуть бути зв'язані не більш ніж одним ребром. Для спрощення досліджень граф уважається не зваженим, тобто сила інформаційних зв'язків не відображається на ваги відповідних ребер.

Вузол являє собою людино-машинну систему, на одному комп'ютері не може перебувати кілька вузлів.

У запропонованій моделі вузол $v_i = \{id_i, flag_i\}$ зберігає унікальний ідентифікатор абонента мережі (id) і прапорець (flag). Змінна *flag* визначає статус вузла: відкритий (*flag*=1) або закритий (*flag*=0).

Алгоритм формування топології ІТКМ складається з послідовності кроків [31]:

– збір даних про топологію доступної частини мережі;

– формування повного графа мережі з урахуванням додавання недоступної частини на основі обчислених прогнозованих топологічних характеристик (розподіл степенів зв'язності, середня довжина шляху);

– формування вектора топологічної уразливості вузлів ІТКМ.

Граф доступної частини мережі – граф, що містить відкриті й закриті вузли й зв'язки між ними. Повний граф мережі – граф, що містить відкриті вузли й закриті вузли, що перейшли в стан відкритих, і зв'язки між ними. Сусідні вузли (суміжні вузли) – вузли, що мають зв'язки з даним вузлом.

Постановка завдання: потрібно скласти граф доступної частини мережі $G(V, E)$, де

V – множина вершин, що включає дві підмножини: $W=\{w_i\}$ – підмножина відкритих вершин; $U=\{u_i\}$ – підмножина закритих вершин;

E – множина зв'язків між вузлами ($e_{ij} = e_{ji}$ – зв'язок між i-*м* і *j*-м вузлами);

A – масив, що містить id пройдених вузлів (a_i – елементи масиву).

Схема алгоритму формування графа доступної частини мережі представлено на рисунку 3.1.

Змінні, використовувані в алгоритмі:

k – лічильник вузлів;

$Z=\{z_i\}$ – множина сусідніх вузлів k-го вузла;

flag – прапор, що визначає статус вузла (*flag*=1 – відкритий, *flag*=0 – закритий);

n – поточне значення довжини масиву A;

i – лічильник сусідніх вузлів;

X – тимчасова множина.

Алгоритм формування графа доступної частини мережі

Крок 1 (блок 2). Початкова установка. Обнулити множину вершин $V=0$ і зв'язків $E=0$. Ініціалізувати лічильник вузлів $(k = 1)$. Додати вершину v_1 у множину V $(V = V \cup v_1)$, зробити її поточною. Виконати $a_k = id(v_k)$.

Крок 2 (блоки 3,4). Виконати функцію $Get(a_k, Z, |Z|, flag)$ одержання множини Z сусідніх вузлів *k-го* вузла, де a_k – ідентифікатор *k-го* вузла, Z – множина, що повертається, $|Z|$ – його потужність, *flag* – прапор, що визначає статус вузла (відкритий/закритий). Якщо $flag=1$ (вузол відкритий), перейти до кроку 3, інакше $(flag=0)$ – до кроку 5.

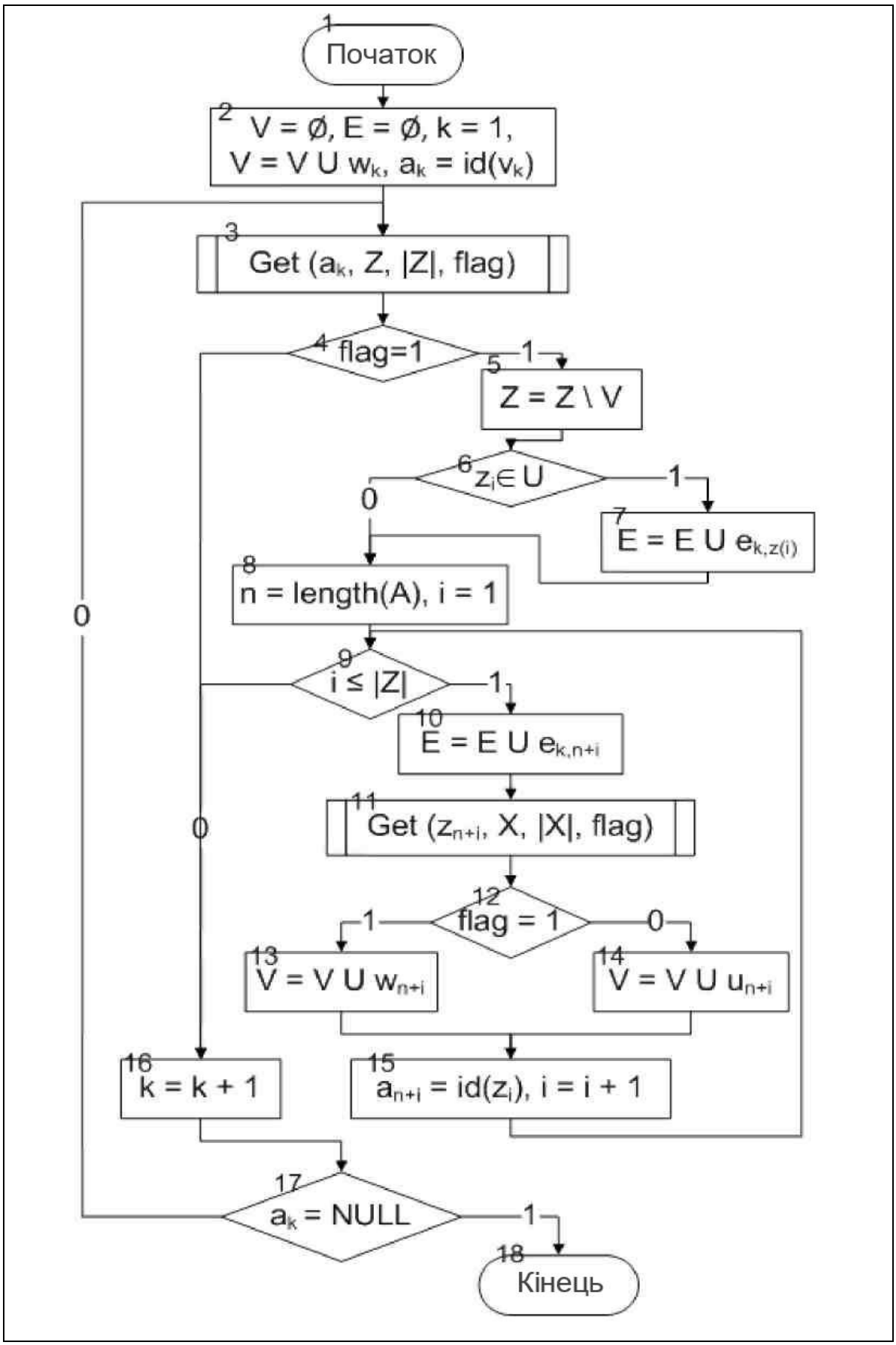

Рисунок 3.1 –Схема алгоритму формування графа доступної частини мережі

Крок 3 (блок 5-7). Для кожного z_i. $\in Z(i = 1,..,|Z|)$ якщо $z_i = v_k$, то $Z = Z \setminus z_i$ і якщо $z_i \in U$, то $E = E \cup e_{k, z(i)}$.

Крок 4 (блоки 8-15). Визначити довжину масиву *A (n = length(A))*. Для $z_{n+1} \in z(I = 1,\ldots,|Z|)$ додати ребро з *k-й* вершиною $E = E \cup e_{к,n+1}$. Виконати функцію *Get{*z_{n+i}*, X, |X|, flag). Якщо flag=1,* то $V = V \cup w_{n+i}$, інакше *(flag=Q)* $V = V \cup u_{n+i}$. Виконати a_{n+i}*=id(zi).*

Крок 5 (блоки 16,17). Перейти до наступного вузла $k = k + l$. Якщо $a_к$ = *NULL,* то кінець алгоритму, інакше перейти до кроку 2.

Розглянемо приклад поетапної реалізації алгоритму 3.1.

Етап 1. Виконується початкові налаштування згідно з першим кроком алгоритму: *k=1, V={*w_1*}, A[12].*

Етап 2. Виконується функція *Get(12, Z,* |Z|, *flag).* Одержується *Z*={43, 36, 39, 78}, |Z|=*4, flag=1.* Перехід до третього кроку алгоритму.

Етап 3. Перевіряється множина *Z* на наявність вузлів, уже доданих у множину *V,* і при наявності таких, видаляємо їх. Одержується *Z*={43, 36, 39, 78}, |Z|=4.

Етап 4. Визначається довжина масиву A(*n*=1). Додаються ребра, що зв'язують першу вершину з вузлами із множини *Z.* Одержується *E={*$e_{1,2}$, $e_{1,3}$, $e_{1,4}$, $e_{1,5}$*}.*

Виконується функція *Get* для всіх вузлів з множини *Z* і додаються вони у відповідні підмножини множини *V*. Одержується *W*={w_1, w_2}, *U*={u_3, u_4, u_5}. Записуємо ідентифікатори вузлів у масив *A*. Одержано *A* = [12, 43, 36, 39, 78].

Етап 5. Збільшується лічильник *k*=1+1=2. Другий елемент масиву *A* (a_2) існує, отже перехід до другого кроку алгоритму.

Після виконання перших п'яти етапів одержано граф, представлений на рисунку 3.2, на якому закриті вузли виділені сірим кольором, а відкриті – білим.

Етап 6. Виконується функція Get(43, *Z*, |Z|, *flag*). Одержано *Z*={12, 16, 25, 4}, |Z|=4, *flag*=1. Переходимо до третього кроку алгоритму.

Етап 7. Перевіряємо множину *Z* на наявність вузлів, уже доданих у множину *V*, і при наявності таких, видаляємо їх. Одержано *Z*={16, 25, 4}, |Z|=3.

Етап 8. Визначається довжина масиву A (n=5). Додаємо ребра, що зв'язують другу вершину з вузлами та множину Z. Одержуємо E={$e_{1,2}$, $e_{1,3}$, $e_{1,4}$, $e_{1,5}$, $e_{2,6}$, $e_{2,7}$, $e_{2,8}$}. Виконується функція *Get* для всіх вузлів з множини Z і додаються у відповідні підмножини множини V. Одержано W={w_1, w_2, w_8}, U={u_3, u_4, u_5, u_6, u_7}. Записуємо ідентифікатори вузлів у масив A, тоді A[12, 43, 36, 39, 78, 16, 25, 4].

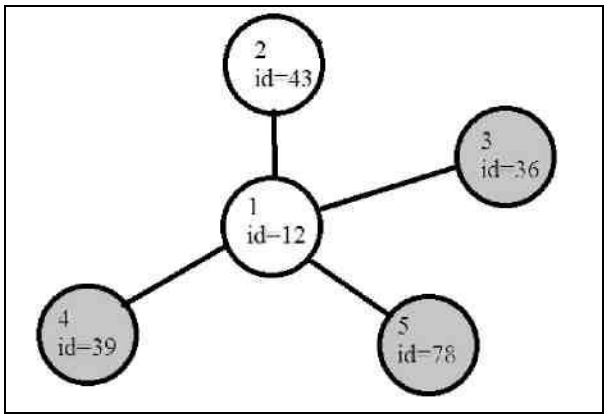

Рисунок 3.2 – Результат роботи алгоритму (1-5 етапи)

Етап 9. Збільшується лічильник k=2+1=3. Третій елемент масиву A (a_3) існує, отже перехід до другого кроку алгоритму.

Після виконання етапів 6-9 одержано граф, представлений на рисунку 3.3.

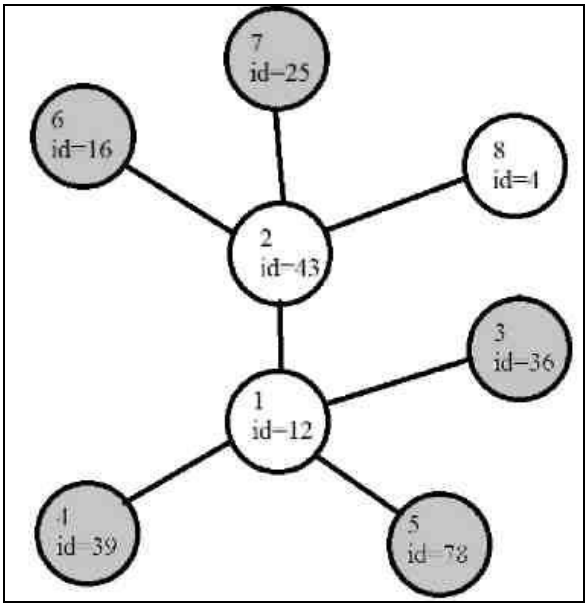

Рисунок 3.3 – Результат роботи алгоритму (1-9 етапи)

Етап 10. Виконуємо функцію *Get*(36, *Z*, |Z|, *flag*). Одержться Z= Ø, |Z|=0, *flag*=0. Переходимо до кроку 5 алгоритму.

Етап 11. Збільшується лічильник k=3+1=4. Четвертий елемент масиву A (a_4) існує, отже перехід до кроку 2 алгоритму.

Етап 12. Виконується функція *Get*(39, *Z*, |Z|, *flag*). Одержано *Z*= Ø, |Z|=0, *flag*=0. Перехід до кроку 5 алгоритму.

Етап 13. Збільшується лічильник k=4+1=5. П'ятий елемент масиву A (a_5) існує, отже перехід до кроку 2 алгоритму.

Етап 14. Виконується функція *Get*(78, *Z*, |Z|, *flag*). Одержано *Z*= Ø, |Z|=0, *flag*=0. Перехід до кроку 5 алгоритму.

Етап 15. Збільшується лічильник k=5+1=6. Шостий елемент масиву A (a_6) існує, отже перехвд до кроку 2 алгоритму.

Етап 16. Виконується функція *Get*(16, *Z*, |Z|, *flag*). Одержуємо *Z*= Ø, |Z|=0, *flag*=0. Переходимо до п'ятого кроку алгоритму.

Етап 17. Збільшується лічильник k=6+1=7. Сьомий елемент масиву A (a_7) існує, отже перехід до другого кроку алгоритму.

Етап 18. Виконується функція *Get*(25, *Z*, |Z|, *flag*). Одержуємо *Z*= Ø, |Z|=0, *flag*=0. Переходимо до п'ятого кроку алгоритму.

Етап 19. Збільшується лічильник k=7+1=8. Восьмий елемент масиву A (a_8) існує, отже перехід до другого кроку алгоритму.

Етап 20. Виконується функція *Get*(8, *Z*, |Z|, *flag*). Одержано *Z*={43, 36}, |Z|=2, *flag*=1. Перехід до кроку 3 алгоритму.

Етап 21. Перевіряється множина *Z* на наявність вузлів, вже доданих у множину *V*, і при наявності таких, вони видаляються. Одержано *Z* = Ø, |Z|=0, E={$e_{1,2}$, $e_{1,3}$, $e_{1,4}$, $e_{1,5}$, $e_{2,6}$, $e_{2,7}$, $e_{2,8}$, $e_{8,3}$}.

Етап 22. На даному етапі нічого не змінюється, бо *Z* = Ø.

Етап 23. Збільшується лічильник k=8+1=9. Дев'ятого елементу масиву A (a_9) не існує, отже робота алгоритму завершена.

Сформований у результаті алгоритму граф доступної ІТКМ представлено на рисунку 3.4.

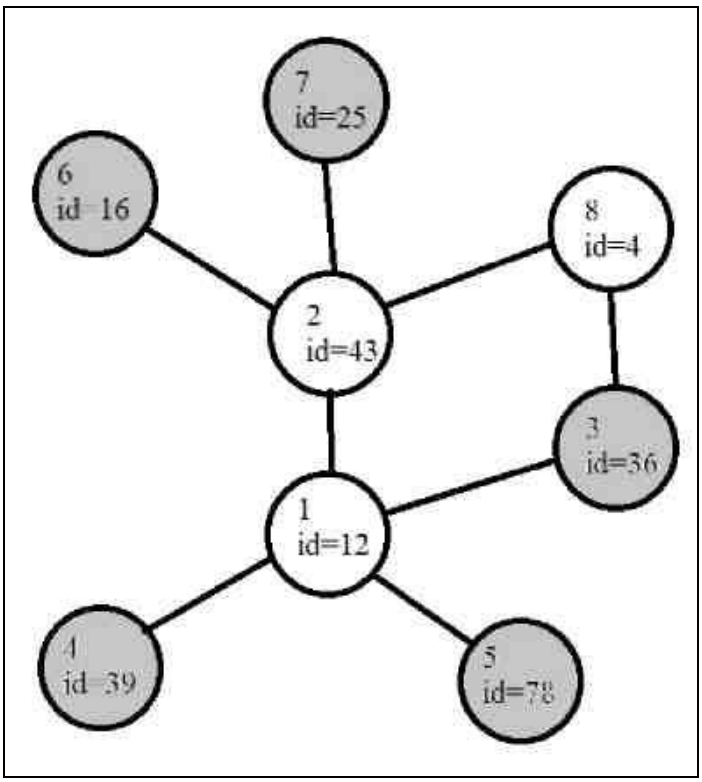

Рисунок 3.4 – Підсумковий результат роботи алгоритму

Результат роботи після кожного етапу відображено в таблиці 3.1.

Таблиця 3.1 – Поетапні результати роботи алгоритму

	V	***E***	***A***	***Z***		
	W1	0	12	0		
	W1	0	12	43,36,		
	W1	0	12	43,36,		
	Wl, W2,	Єi,2, Єi,3,	12,43,36,3	43,36,		
	Wl, W2,	Єi,2, Єi,3,	12,43,36,3	43,36,		
	Wl, W2,	Єi,2, Єi,3,	12,43,36,3	12,16,		
	Wl, W2,	Єi,2, Єi,3,	12,43,36,3	16,25,		
	Wl, W2,	Єi,2, Єi,3,	12,43,36,3	16,25,		
	Wl, W2,	Єi,2, Єi,3,	12,43,36,3	16,25,		
	Wl, W2,	Єi,2, Єi,3,	12,43,36,3	0		
	Wl, W2,	Єi,2, Єi,3,	12,43,36,3	0		
	Wl, W2,	Єi,2, Єi,3,	12,43,36,3	0		
	Wl, W2,	Єi,2, Єi,3,	12,43,36,3	0		
	Wl, W2,	Єi,2, Єi,3,	12,43,36,3	0		

Розглянутий приклад показує коректність алгоритму.

3.2 Розробка алгоритму формування повного графа

Розроблений алгоритм формування повного графа мережі, який враховує топологічні характеристики доступної частини мережі (розподіл ступенів зв’язності, середня довжина шляху).

Обчислення середнього степеня зв’язності мережі. Степінь зв’язності вузла (degree) – кількість суміжних з ним вузлів [31].

Середній ступінь зв’язності мережі (average degree) – середнє арифметичне степенів зв’язності по всій мережі. Використаний алгоритм обчислення середнього степеня зв’язності ґрунтується на обчисленні ступенів зв’язності у відкритих вузлів з урахуванням їхніх зв’язків із закритими. Середнє значення береться по відкритих вузлах.

Одержання розподілу степенів зв’язності вузлів у мережі. Розподіл степенів зв’язності вузлів – статистична характеристика, що показує кількість вузлів з кожним значенням зв’язності в мережі.

Облік відкритих і закритих вузлів при отриманні розподілу степенів зв’язності здійснюється аналогічним чином з обчисленням середнього степеня зв’язності.

Обчислення кластерного коефіцієнта мережі. Кластерній коефіцієнт вузла – характеристика, що показує «щільність» зв’язків навколо вузла. Кластерний коефіцієнт вузла обчислюється як відношення числа існуючих зв’язків між суміжними вузлами до значення загальної кількості можливих таких зв’язків:

$$C_i = \frac{2n_i}{k_i \cdot (k_i - 1)},$$

де k_i – ступінь зв’язності вузла, n_i – кількість зв’язків між суміжними вузлами.

Приклад обчислення кластерного коефіцієнта для вузла 1 (рисунок 3.5). Суцільними лініями показані існуючі зв’язки, пунктирними – потенційні. Степінь зв’язності k=4. Число можливих зв’язків між його суміжними вузлами дорівнює

$k(k-1)/2 = 4(4-1)/2=6$. Кількість існуючих зв'язків – 2. Кластерний коефіцієнт $C=2/6=1/3$.

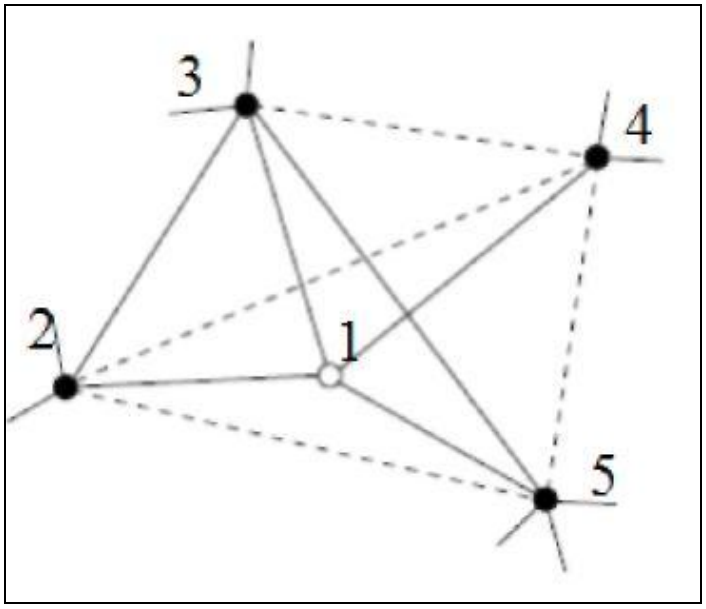

Рисунок 3.5 –Визначення кластерного коефіцієнту

Алгоритм обчислення коефіцієнта кластеризації мережі полягає в підрахунку кластерного коефіцієнта кожного вузла й знаходження середнього значення. Обчислення кластерних коефіцієнтів здійснюється тільки для відкритих вузлів з підрахунком клік, що утворені і відкритими і закритими вузлами. Середнє значення розраховується по відкритих вузлах.

Алгоритм обчислення середньої довжини шляху мережі. Середня довжина шляху вузла – середнє арифметичне найкоротших шляхів від заданого вузла до всіх інших. Середня довжина шляху мережі – середнє арифметичне середніх довжин шляхи всіх вузлів мережі.

Обчислення середньої довжини шляху в графі здійснюється тільки по відкритих вузлах. Закриті вузли при цьому «вилучалися» з мережі, тому що вони не несуть корисного інформаційного навантаження для даної топологічної характеристики. Даний алгоритм полягає в обчисленні суми середніх довжин шляху для кожного відкритого вузла, поділеної на їхню загальну кількість.

Схема алгоритму формування повного графа мережі з урахуванням недоступної частини представлено на рисунку 3.6.

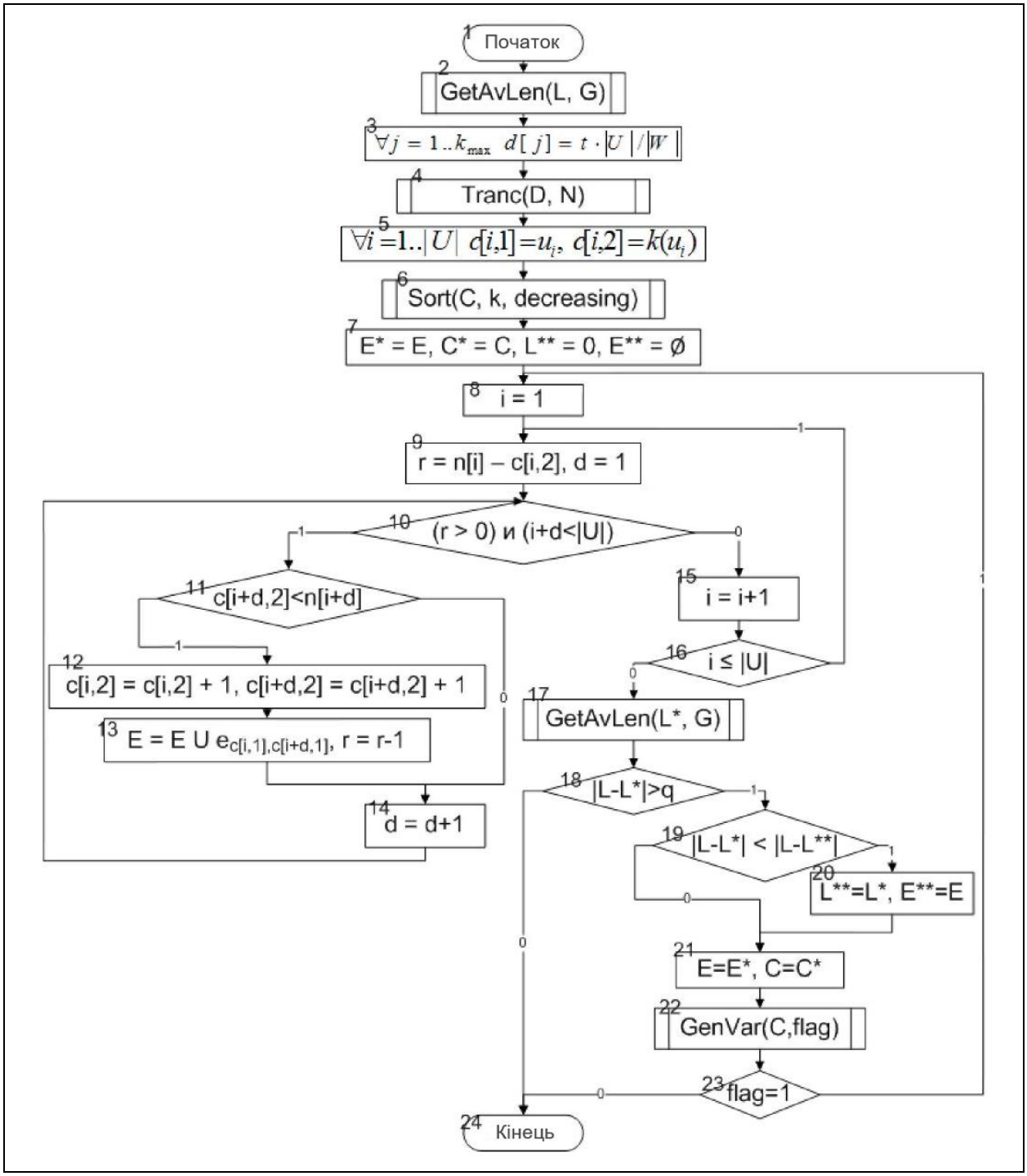

Рисунок 3.6 – Алгоритм генерації недоступної частини мережі

Топологічна уразливість ІТКМ – внутрішня властивість ІТКМ, заснована на характеристиках її топології, яке сприяє поширенню загрози забороненої інформації.

3.3 Формування вектора топологічної вразливості повного графа мережі

Топологічною уразливістю вузла мережі назвемо показник φ, який обчислюється за формулою:

$$\varphi_i = \frac{k_i \cdot (C_i + 1)}{L_i},$$

де k_i – ступінь зв'язності вузла,

Ci – кластерний коефіцієнт вузла,

L – середня довжина шляху вузла.

Дана характеристика показує, наскільки вразливий до атак, з погляду розташування в мережі, певний вузол.

Під час дослідження топологій реальних великомасштабних ІТКМ (105-108), можна виділити основні значущі положення:

– середній ступінь зв'язності вузлів у таких мережах становить 100-1000;

– середня довжина шляху визначається теорією шести рукостискань: у глобальних масштабах рівна 6, у реальних мережах становить значення 3-5;

– коефіцієнт кластеризації, як правило, варіюється в значеннях від 0,01 до 0,2.

Виходячи з перерахованого вище й отриманих експериментальних результатів, маємо типове значення коефіцієнта топологічної уразливості в діапазоні від 100 до 500.

Практичне застосування полягає в тому, що використовуючи коефіцієнт φ, можна оцінити топологічну уразливість конкретної реальної мережі.

При аналізі топологічних характеристик мережі можна підрахувати коефіцієнти уразливості для кожного вузла в мережі (вектор топологічної вразливості вузлів ІТКМ).

Вектор топологічної уразливості вузлів ІТКМ – вектор виду, що наведений в табл. 3.2

Таблиця 3.2 – Структура вектору топологічної вразливості вузлів

№ вузла	Значення φ
Вузол 1	φ_i
……..	…….
Вузол N	φ_N

Отриманий вектор можна використовувати під час прогнозування загрози поширення забороненої інформації. З одного боку, можна класифікувати за небезпекою атакуючі вузли, а з іншого – вибудувати найбільш ефективну стратегію протидії загрозі.

3.4 Алгоритм протидії загрозі поширення забороненої інформації

При наявності адміністративного ресурсу можна реалізувати автоматизовану систему протидії загрозі поширення забороненої інформації. Представлено узагальнений алгоритм роботи такої системи. Розглянуті функції реалізуються за допомогою типових засобів.

Крок 1. Введення даних – типове повідомлення, що містить інформацію, заборонену до поширення. База даних таких повідомлень формується з санкційного списку екстремістських матеріалів і єдиного реєстру доменних імен, покажчиків сторінок сайтів у мережі Інтернет і мережевих адрес, що дозволяють ідентифікувати сайти в мережі Інтернет, які містять інформацію, поширення якої в Україні заборонене.

Крок 2. Виявлення «маркерів», тобто слів і словосполучень, що мінімально змінюються в процесі переформулювання.

Крок 3. Синтез формального опису «маркерів» з використанням регулярних виразів або контекстно-вільної граматики.

Далі робота алгоритму розбивається на дві процедури, що паралельно виконуються, попередження й усунення наслідків загрози.

Процедура «Попередження загроз».

Початок → Типове повідомлення із забороненою інформацією → Виявлення «маркерів» → Синтез формального опису «маркерів» → Попередження.

Процедура 2 «Ліквідація наслідків»

Складання правил фільтрації повідомлень на основі формального опису → Конструювання ряду пошукових запитів за формальними правилами й підстроювання параметрів пошуку → Конфігурація технічних засобів фільтрації з використанням правил → Виконання запитів і аналіз результатів.

Процедура «Моделювання загрози поширення забороненої інформації».

Видалення знайдених сутностей зі збереженням зв’язності БД → Підвищення пріоритету процесу фільтрації відповідно до результатів моделювання → Відправлення повідомлення про проведені заходи в контролюючі органі.

Кінець.

Попередження.

Крок 4а. Складання правил фільтрації повідомлень на основі формального опису здійснюється шляхом компіляції регулярних виразів за допомогою засобів, призначених для фільтрації (див. крок 5а).

Крок 5а. Конфігурація технічних засобів фільтрації з використанням правил. Як правило, це антиспам-системи такі як Apache Spamassassin, FASTBL, DNSBl і ін.

Крок 6а. Моделювання загрози поширення забороненої інформації.

Крок 7а. Підвищення пріоритету процесу фільтрації відповідно до результатів моделювання загрози поширення забороненої інформації.

Ліквідація наслідків

Крок 4б. Конструювання послідовності пошукових запитів за формальними правилами, і підлаштовування параметрів пошуку (пріоритет, глибина й ін.)

Крок 5б. Виконання запитів і аналіз результатів. На даному етапі можливе уточнення запитів.

Крок 6б. Видалення знайдених сутностей зі збереженням зв’язності БД.

Крок 7б. Відправлення повідомлення про проведені заходи в контролюючі органи.

4 ОПИС РОЗРОБЛЕНОГО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

4.1 Особливості розробки програмного інструментарію

Розроблені алгоритми попередження загроз та формування топології ІТКМ реалізована у вигляді програмного комплексу.

Перша програма призначена для одержання доступної частини мережі. Хоча дане ПЗ орієнтоване на соціальну мережу «Fasebook», його легко можна переробити під іншу ІТКМ. Робота програми заснована на алгоритмі обходу завширшки. Програму написано мовою програмування Python. Для зберігання топології використовується об’єктно-орієнтована база даних *ZODB*. Одержання інформації здійснюється за допомогою API Fasebook. Програма збирає дані до настання однієї з наступних подій: отримана інформація про всі відкриті вузли в мережі, збір даних перерваний користувачем.

На початку роботи програми необхідно авторизуватися під акаунтом абонента, з якого почнеться збір інформації. Вихідні дані програми представляють собою текстовий файл, у якому в кожному рядку записаний ідентифікатор вузла, і через пробіл перераховані ідентифікатори суміжних з ним вузлів.

У результаті роботи програми була отримана частина топології соціальної мережі Fasebook, що містить 118834 відкритих вузлів й 16270504 закритих. Фрагмент вихідного файлу представлено на рисунку 4.1.

Друга програма призначена для формування повного графа ІТКМ на основі обчислених прогнозованих топологічних характеристик і формування його вектора топологічної уразливості. ПЗ створене для використання з використанням розподілених обчислювальних ресурсів. Програма написана в середовищі програмування Microsoft Visual Studio. Інтерфейсом взаємодії між процесами в додатку є MPI. Для представлення графа в пам’яті обчислювальної системи використовувалося два підходи: нерозподілений (локальний, використовувалася бібліотека Boost Graph Library) і розподілений (Parallel Boost Graph Library).

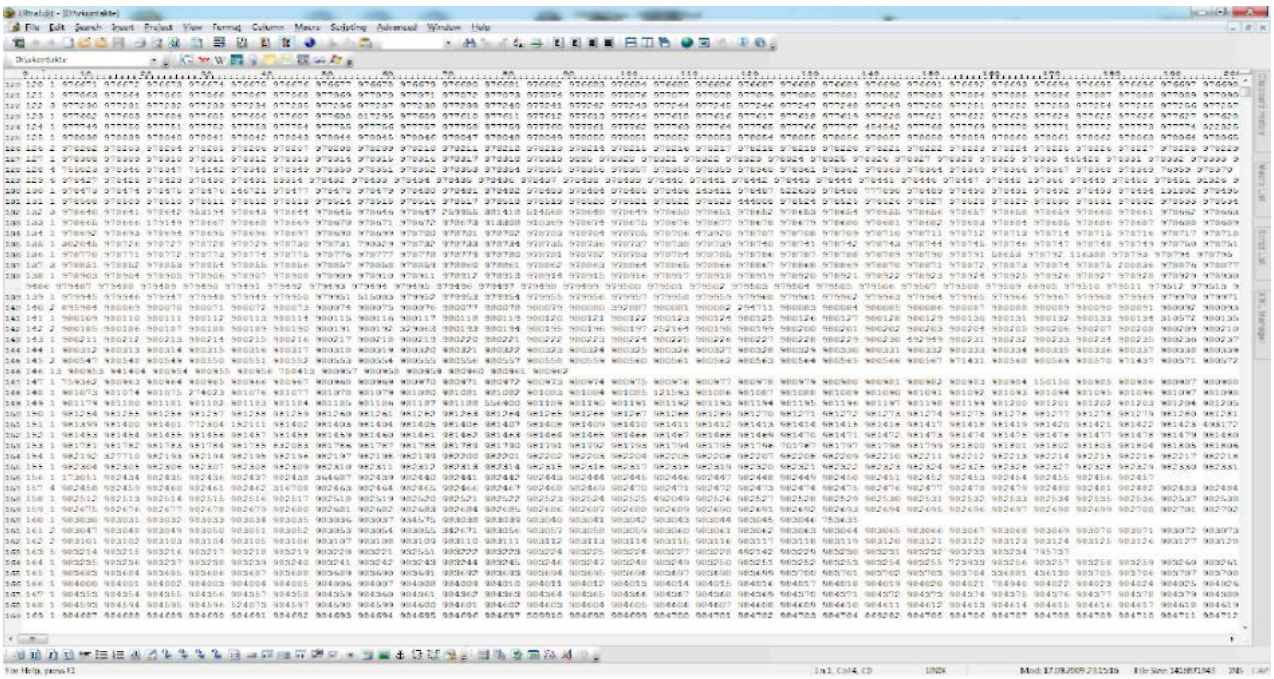

Рисунок 4.1 – Фрагмент вихідного файлу програми

На рисунку 4.2 показаний фрагмент (1000 вузлів) отриманої топології, побудований за допомогою ПЗ Pajek.

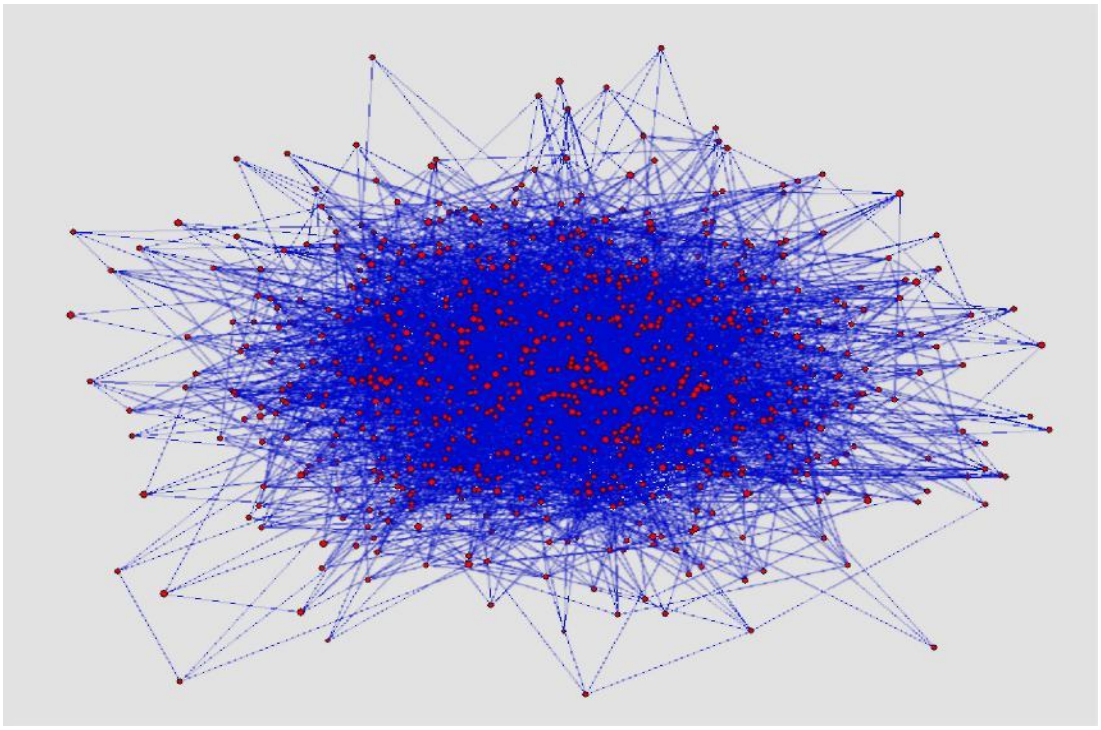

Рисунок 4.2 – Візуалізований фрагмент топології

Бібліотека «Parallel Boost Graph Library (PBGL)» надає гнучку й ефективну реалізацію концепції графів. Входить у збірку бібліотек boost, що розширюють функціональність C++, які вільно поширюються за ліцензією Boost Software License разом з вихідним кодом.

Бібліотека дозволяє обрати представлення графа, тип даних і алгоритм із великого набору алгоритмів, серед яких:

- пошук завширшки;
- пошук у глибину;
- алгоритм Беллмана-Форда;
- алгоритм Дейкстри;
- алгоритм Прима;
- алгоритм Краскала;
- знаходження компонентів зв’язності графа;
- завдання про максимальний потік;
- зворотний алгоритм Катхилла-Макки;
- алгоритм топологічного сортування й ін.

Формат вихідних даних аналогічний до даних, першої програми – текстовий файл, у якому в кожному рядку записаний ідентифікатор вузла, і через пробіл перераховані ідентифікатори суміжних з ним вузлів (топологія повного графа мережі). Другий вихідний файл – файл із вектором топологічної уразливості мережі.

4.2 Розподілене моделювання ЗПЗІ в ІТКМ

Моделювання ЗПЗІ на великомасштабній ІТКМ є трудомістким завданням. Його вирішення в прийнятний термін і одержання актуальних результатів можливо лише завдяки використанню розподілених обчислювальних ресурсів, тому розроблене ПЗ має характер моделювання обчислювального експерименту, дослідження проводилися на двох фрагментах ІТКМ. Перший (фрагмент соціальної мережі «Fasebook») отриманий у рамках даної наукової роботи, а другий (фрагмент із 16163521 вузли соціальної мережі «Facebook») отриманий незалежно з джерел й ін. [32, 33].

Експериментальне дослідження ЗПЗІ в ІТКМ здійснювалося на основі імітаційної моделі. Імітаційна модель реалізована у вигляді розробленого ПЗ під розподілену обчислювальну систему. Для реалізації паралельних обчислень на графі була використана бібліотека Parallel Boost Graph Library. Бібліотека є вільно розповсюджуваною й за своїми функціональними можливостями не має альтернатив.

Інтерфейсом взаємодії між процесами в додатку є MPI, для представлення графа в пам’яті обчислювальної системи використовувався розподілений підхід з використанням бібліотеки Parallel Boost Graph Library.

Формат вхідних даних – текстовий файл, у якому в кожному рядку записаний ідентифікатор вузла, і через пробіл перераховані ідентифікатори суміжних з ним вузлів (топологія повного графа мережі). У вихідному файлі фіксуються дані про динаміку ЗПЗІ, представлені списками атакуючих і захищених вузлів у кожний квант часу.

У додатку Г наведено частина коду програми.

5 АНАЛІЗ РЕЗУЛЬТАТІВ ЕКСПЕРИМЕНТАЛЬНИХ ДОСЛІДЖЕНЬ ТОПОЛОГІЇ ІТКМ

У ході експериментальних досліджень були отримані результати, що стосуються топології ІТКМ.

Після проведення експериментів можна зрівняти результати із представленими даними й зробити висновок про приналежність соціальних мереж до певного типу, виходячи з отриманих топологічних характеристик. Знаючи топологічні характеристики ІТКМ, можна генерувати на їхній основі мережі з такими ж параметрами будь-яких масштабів, що допоможе вивчати процеси, які відбуваються в них з використанням моделювання.

Кластерний коефіцієнт мережі обчислено за топологічним фрагментом мережі «Fasebook». Значення середнього кластерного коефіцієнта мережі вийшло рівним 0,048087. Докладні дані у вигляді діапазонів значення коефіцієнта й кількості вузлів, що потрапляють у відповідні інтервали, наведено в таблиці 5.1.

Таблиця 5.1 – Кластерний коефіцієнт («Fasebook»)

Кластерний коефіцієнт	Кількість вузлів
[0;0,1)	104810
[0,1;0,2)	9094
[0,2;0,3)	2423
[0,3;0,4)	1198
[0,4;0,5)	587
[0,5;0,6)	332
[0,6;0,7)	188
[0,7;0,8)	67
[0,8;0,9)	39
[0,9;1)	8
1	88

Аналізуючи отримані дані, можна сказати наступне. Більшість вузлів мають кластерний коефіцієнт в інтервалі [0; 0,1), що свідчить про низьку ступінь кластеризації розглянутого фрагмента мережі. Проте, є присутньою група вузлів з коефіцієнтом рівним одиниці, яка вибивається із залежності – чим більше

значення кластерного коефіцієнта, тим менше вузлів. Фізичний сенс цього явища можна пояснити в такий спосіб. У вибірці захопили групи користувачів, які підтримують тісні зв'язки між собою, наприклад, у зв'язку з родом діяльності. Захоплення, у свою чергу, таких груп визначається використаним методом вибірки – обходом завширшки.

Міланський університет і Facebook, проводячи спільне дослідження теорії шести рукостискань, одержали значення 4,74 [13].

Розбіжність у значеннях пояснюється кількістю вузлів у вибірці. Наведені дані дозволяють при дослідженні великомасштабних ІТКМ використовувати фіксоване значення середньої довжини шляху.

Використовуючи розроблену аналітичну модель, можна одержати прогноз за динамікою ЗПЗІ в ІТКМ за прийнятний час. Алгоритм одержання прогнозу складається з послідовності наступних кроків.

Крок 1. Визначити коефіцієнт топологічної уразливості розглянутої ІТКМ. Необхідно постійно проводити моніторинг значення даного параметра для самих великомасштабних і популярних мереж для використання його актуального значення.

Крок 2. З появою перших повідомлень із забороненою інформацією зібрати статистику таких повідомлень. Даний крок необхідно виконати на ранніх стадіях виникнення загрози. З одного боку, чим більше даних буде зібрано, тем точніше буде прогноз, з іншого боку, при затримці виконання даного кроку, актуальність прогнозу може бути втрачена.

Крок 3. Апроксимувати зібрані дані за допомогою системи диференціальних рівнянь, що описують модель, підібравши потрібні значення β і γ (імовірності атаки й захисту).

У результаті одержуємо прогноз на весь період поширення загрози забороненої інформації.

У якості рекомендацій запропоновано алгоритм роботи автоматизованої системи протидії загрозам поширенню забороненої інформації.

Створене програмне забезпечення призначене для моделювання автоматизації пошуку вузлів соціальної мережі, які є потенційними розповсюджувачами забороненої інформації. Розроблене програмне забезпечення підвищує ефективність за часом.

ВИСНОВКИ

Інформаційно-телекомунікаційні мережі є великомасштабними мережами з постійно зростаючим числом абонентів. З бурхливим ростом числа користувачів ІТКМ виникають проблеми інформаційної безпеки й захисту інформації в них.

Аналіз проблем інформаційної безпеки виявив, що окрім проблем, пов'язаних з використанням глобальної мережі Інтернет як розподіленої інформаційно-телекомунікаційної системи, які досить добре відомі й розв'язані, існує маловивчена проблема забороненого контенту.

Створення моделей і алгоритмів поширення загрози забороненої інформації – один із ключових підходів при вирішенні даного завдання. Проведений аналіз публікацій з даної тематики показує, що існуючі рішення малоефективні. Звичайно при моделюванні поширення загрози забороненої інформації не враховується топологія ІТКМ (модель мережі – повнозв'язний граф). А, якщо топологія враховується, то, як правило, використовується найпростіша SIS модель, а структура мережі відображається SF мережею. Під час моделювання ЗПЗІ важливо мати топологію, що відображає структуру зв'язків реальної мережі, а також використовувати адекватну модель інформаційної взаємодії вузлів. Ще однієї важливою проблемою є великомасштабність ІТКМ, яка заважає одержати дані з імітаційної моделі за прийнятний час. Розв'язання цього завдання полягає в створенні аналітичної моделі ЗПЗІ в ІТКМ.

Створено імітаційну модель ЗПЗІ в ІТКМ, що враховує топологічні характеристики мережі, а також особливості інформаційної взаємодії абонентів як людино-машинних систем. З її допомогою проведені експерименти, результати яких показали залежність реалізації ЗПЗІ від топологічної уразливості мережі.

Розроблена аналітична модель ЗПЗІ з урахуванням топологічної вразливості мережі. Релевантність результатів аналітичного розв'язання підтверджена серією експериментів на топології реальної мережі з використанням імітаційного моделювання.

Розроблено алгоритм формування вихідних даних за топологією мережі (множина вершин і зв’язків між ними доступної частини мережі), який враховує обмеження по збору даних і реалізований у вигляді розробленого програмного забезпечення.

Розроблено алгоритм формування повного графа мережі з урахуванням додавання недоступної частини на основі обчислених прогнозованих топологічних характеристик. Алгоритм реалізовано у вигляді розробленого програмного забезпечення.

Розроблено програмне забезпечення, яке дозволяє за прийнятний час одержати результати моделювання ЗПЗІ в ІТКМ за рахунок використання розподілених обчислювальних ресурсів.

За допомогою розробленого ПЗ було проведено моделювання, результат якого показав, що великомасштабні інформаційно-телекомунікаційні мережі не можна віднести до жодного з існуючих класів складних мереж.

Результати, отримані за значенням середньої довжини шляху (підтвердження теорії шести рукостискань), дозволяють при дослідженні великомасштабних ІТКМ використовувати фіксоване значення середньої довжини шляху.

Запропоновано алгоритм роботи автоматизованої системи протидії поширенню загрози забороненої інформації.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Бреер, В.В. Стохастические модели социальных сетей / В.В. Бреер; Управление большими системами, № 27. – 2013. - С. 169-204.

2. Анализатор Sniffer Pro LAN / Sniffer Technologies. URL: http://www.securitylab.ru/software/233623.php

3. Gjoka, M., Sirivianos, M., Markopoulou, A., Yang, X. Poking facebook: characterization of osn applications / M. Gjoka [et al.]; Proc. of WOSN - 2018.

4. Гусева, А.И. Технология межсетевых взаимодействий / А.И. Гусева; - М.: Бином, 2007. – 238 с.

5. Касперски, К. Компьютерные вирусы: изнутри и снаружи / К. Касперски; - Спб: "Питер", 2005. - 528 с.

6. Лукацкий, А. Обнаружение атак / А. Лукацкий; -: БХВ, 2013. - 624 с.

7. Собейкис, В.Г. Азбука хакера 3. Компьютерная вирусология / В.Г. Собейкис; - М.: Майор, 2006. - 512 с.

8. Drayer B., Brox T. Object Detection, Tracking, and Motion Segmentation for Object-level Video Segmentation // arxiv.org. 2016. – URL: https://arxiv.org/abs/1608.03066.

9. Hinton G. A practical guide to training restricted Boltzmann machines // Momentum. – 2010. – № 9(1).

10. Kim C., Li F. Multiple Hypothesis Tracking Revisited // Proceedings of the IEEE International Conference on Computer Vision. – 2019.

11. Konev A., Chigorin A., Krivovyaz G., Velizhev A., Konushin A. Traffic signs recognition on images with training on synthetic data // Technical vision in computer systems. – 2019. P. 65-66.

12. Russakovsky O., Deng J., Su H., Krause J., Satheesh S., Ma S., Huang Z., Kar- pathy A., Khosla A., Bernstein M., Berg A.C., Fei-Fei L. Imagenet large scale visual recognition challenge // IJCV. – 2015

13. Ruta A. A New Approach for In-Vehicle Camera Traffic Sign Detection and Recognition // IAPR Conference on Machine vision Applications (MVA). – 2009. – P. 509-513.

14. Аведьян Є.Д., Галушкин А.І., Селиванов С.А. Порівняльний аналіз структур полносвязных і сверточных нейронних мереж і їх алгоритмів навчання // Інформатизація й зв'язок. – 2017. – № 1.

15. Антощук С.Г. Відстеження об'єктів інтересу при побудові автоматизированных систем відеоспостереження за людьми // Електротехнічні й комп'ютерні системи. – 2018. – №8(84). – С. 151–156.

16. Zhai M., Roshtkhari M., Mori G. Deep Learning of Appearance Models for Online Object Tracking // arxiv.org . 2019. – URL: https://arxiv.org/abs/1607.02568 .

17. Zhang K., Liu Q. Robust Visual Tracking via Convolutional Networks // arxiv.org . 2019. – URL: https://arxiv.org/abs/1501.04505 .

18. Golbeck, J., Hendler, J. Inferring binary trust relationships in web-based social networks [Text] / J. Golbeck, J. Hendler; Transactions on Internet Technology - 2006. -Vol. 6, no. 4. - P. 497-529.

19. Xiang Y., Alahi A. Learning to Track: Online Multi-Object Tracking by Deci- sion Making // Proceedings of the IEEE International Conference on Computer Vision. – 2015.

20. Chetverikov G., Puzik O., Vechirska I. Multiple-valued structures of intellectual systems //Proceedings of the with Internations Computer Sciences and Information Technologies (CSIT). 2016, 7589907. -pp. 204-207

21. Granovetter, M. The strength of weak ties / M. Granovetter; American Journal of Sociology - 1973. - Vol. 78. - P. 1360-1380.

22. Granovetter, M. Threshold Models of Collective Behavior / M. Granovetter; American Journal of Sociology - 1978. - Vol. 83, no. 6. - P. 1420-1443.

23. Grimaldi, R. P. Discrete and Combinatorial Mathematics / R.P. Grimaldi; an applied introduction. - 4th edition. - New York, 1998.

24. Heberlein, L.T., Dias, G.V., Levitt, K.N, Mukherjee, B., Wood, J., Wolber, D.A.

25. Network security monitor / L.T. Heberlein [et al.]; Proc. of IEEE Symposium on Re-search in Security and Privacy. – Los Alamitos, CA, USA: IEEE Computer Society, 2020. - P. 296–304.

26. Hethcote, H.W. The Mathematics of Infectious Diseases / H.W. Hethcote; - 2015. - P. 599-653,

27. Hofmeyr, S.A., Forrest, S., Somayaji, A. Intrusion detection using sequences of system calls / S.A. Hofmeyr, S. Forrest, A. Somayaji; Journal of Computer Security. - Amsterdam: IOS Press, 2018. – Vol. 6, no 3. - P. 151-180.

28. Janky, B., Takacs, K. Social Control, Participation in Collective Action and Network Stability / B. Janky, K. Takacs; HUNNET Working Paper. - 2020.

29. Amaral, LAN, Scala, A., Barthelemy, M., Stanley HE (2000) Classes of small-world networks / Amaral LAN, A. Scala, M. Barthelemy, Stanley HE; Proceedings of the National Academy of Sciences of the United States of America. - 2017: 11149

30. Roberts, M.G., Heesterbeek, JAP Mathematical models in epidemiology / M.G. Roberts, Heesterbeek JAP; In JA. Filar (Ed.) Mathematical Models. Oxford: EOLSS Publishers Ltd, 2004.

31. Frauenthal, J.C. / J.C. Frauenthal; Mathematical Models in Epidemiology. – New York: Springer-Verlag, 2008. – 335 p.

32. Shostak I., Matyushenko I., Romanenkov Yu., Danova M., Kuznetsova Yu. Computer Support for Decision-Making on Defining the Strategy of Green IT Development at the State Level. In book: Green-IT Engineering: Social, Business and Industrial Applications, Vol. 171. Berlin, Heidelberg: Springer International Publishing, 533–559 (2018), https://doi.org/10.1007/978-3-030-00253-4

33. Shostak I., Kapitan R,. Volobuyeva L., and Danova M., Ontological Approach to the Construction of Multi-Agent Systems for the Maintenance Supporting Processes of Production Equipment. In Proc. : IEEE International Scientific and Practical Conference «Problems of Infocommunications. Science and Technology» (PICS&T-2018). Ukraine, Kharkiv, October 9-12, 2018. P. 209 – 214

34. Кукіер, К. Big Data: A Revolution That Will Transform How We Live, Work, and Think/К. Кукіер, В. Штойнберг, 2018. – 236 с.

35. Kasturirangan, R. Multiple Scales in Small-World Networks / R. Kasturirangan; Brain and Cognitive Science Department, MIT. - 20099.

36. Kenah, E., Robins, J. M. Network-based analysis of stochastic SIR epidemic models with random and proportionate mixing / E. Kenah, J. M. Robins; Departments of Epidemiology and Biostatistics Harvard School of Public Health. -2007.

37. Kephart, J.O., White, S.R. Directed-Graph Epidemiological Models of Computer Viruses / J.O. Kephart, S.R. White; Proceedings of the IEEE Computer Society Symposium on Research in Security and Privacy. -2019. P. 343 - 359.