УДК 004.891

д.т.н., доц. **Гунченко Ю.О.** (ОНУ)
к.т.н., доц. **Бойчук В.О.** (ХмНУ)
Юмашов В.С. (ХмНУ)

МЕТОД ВИЯВЛЕННЯ ЗАГРОЗИ ПОШИРЕННЯ ЗАБОРОНЕНОЇ ДО РОЗПОВСЮДЖЕННЯ ІНФОРМАЦІЇ В КОМП'ЮТЕРНИХ МЕРЕЖАХ

У статті розглядаються комп'ютерні мережі (КМ), що знаходяться під впливом загрози поширення забороненої інформації. Досліджуються моделі загрози поширення цієї інформації і особливості комп'ютерних мереж. Наведено основні проблеми інформаційної безпеки, які актуальні для даного дослідження. Розробляється і досліджується імітаційна модель, що враховує топологічні характеристики мережі, а також особливості інформаційної взаємодії абонентів як людино-машинних систем. і аналітична модель загрози поширення забороненої до розповсюдження інформації в КМ з урахуванням топологічної вразливості мережі. Релевантність результатів аналітичного рішення підтверджена серією експериментів на топології реальної мережі з використанням імітаційного моделювання. Також розробляється методика формування топології великомасштабної КМ з проведенням досліджень, яка враховує основні топологічні характеристики доступної частини мережі і працює в умовах недостатньої репрезентативності вибірки вихідних даних. Запропоновано систему протидії загрозі поширення забороненої до розповсюдження інформації.

Ключові слова: безпека комп'ютерних мереж, заборонена інформація, моделювання, формування топології.

Вступ. Комп'ютерні мережі (КМ) забезпечують практично повний спектр можливостей для обміну інформацією між користувачами - мережевими абонентами. Комп'ютерна мережа надає різні сервіси для організації соціальних взаємовідносин 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 існують кілька найбільш популярних представників. У США це «Facebook», «MySpace», «Twitter» і «LinkedIn»; «Nexopia» - в Канаді, «Bebo» - в Великобританії, «Facebook», «dol2day» - в Німеччині. В Україні на сьогоднішній день найпопулярними є «ВКонтакте» і «Одноклассники». Сучасною проблемою таких систем є їх низький рівень інформаційної безпеки. Для забезпечення захисту інформації в телекомунікаційних мережах, включаючи Інтернет, розроблено безліч методів і засобів, запропонованих в працях П.Д. Зегжди, С.П. Расторгуєва, Р. Бретта, В.І. Завгороднього, В.А. Герасименко, А.А. Малюка, В. Столінгса, К. Касперски, С. Норкатта, В.В. Домарева. Тим не менш, 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 безлічі функцій захисту принциповою відносно цих систем є функція попередження прояву забороненої інформації. Вона реалізується за рахунок механізмів прогнозування загрози поширення і розсилки повідомлень з попередженнями про наслідки дій із забороненим контентом. Використання інших функцій (попередження, виявлення, локалізації та ліквідації загрози) припускає наявність повного контролю над системою, що в справжніх умовах неможливо.

Постановка задачі. Одним з підходів до прогнозування загрози поширення забороненої інформації (ЗгЗІ) є моделювання, наприклад, з використанням моделей впливу, моделей просочування і зараження (Д.А. Новиков, Д.А. Губанов і А.Г. Чхартішвілі, J. Leveille, D . Watts і S. Strogatz, R. Albert та A. Barabasi, J. Leskovec, M. Gjoka, SN Dorogovtsev, MEJ Newman і RM Ziff, JO Kephart і SR White і ін.). Дані моделі, як правило, не враховують топологічні особливості мережі (розподіл ступенів зв'язності, кластерний коефіцієнт, середня

довжина шляху). Взаємодія між абонентами в рамках цих математичних моделей описується переважно гомогенним графом, що при моделюванні великомасштабних мереж (більше 10 млн. вузлів) може дати похибку прогнозування ЗгЗІ більше 30%. Крім того, дані підходи носять в основному теоретичний характер, практика їх використання не виходить за рамки експериментів. Таким чином, дослідження, спрямовані на створення моделей та алгоритмів ЗгЗІ, актуальні і мають теоретичне і практичне значення у вирішенні проблеми забезпечення інформаційної безпеки в системах і мережах телекомунікацій.

Мета роботи полягає в підвищенні точності прогнозування загрози поширення забороненої інформації в комп'ютерних мережах.

Для досягнення мети роботи необхідно вирішити наступні завдання:

1. Провести інформаційний огляд і експерименти для виявлення істотних характеристик об'єкта і зовнішніх факторів, що впливають на процес реалізації ЗгЗІ. Виконати аналіз основних підходів до моделювання ЗгЗІ.

2. Розробити імітаційну модель ЗгЗІ в КМ.

3. Синтезувати і показати адекватність аналітичної моделі ЗгЗІ в КМ.

4. Розробити методику формування топології КМ.

5. Змоделювати процес реалізації ЗгЗІ на топології реальної великомасштабної КМ. Провести експериментальне дослідження з отриманими результатами.

Виклад основного матеріалу досліджень. За результатами огляду предметної області було поставлено завдання створення імітаційної і аналітичної моделей поширення загрози забороненої інформації в КМ. Імітаційна модель необхідна для отримання експериментальних результатів для синтезування аналітичної моделі. Необхідність створення аналітичної моделі обґрунтовується тим, що для імітаційного моделювання на топології існуючих комп'ютерних мереж (десятки мільйонів вузлів) необхідні великі витрати часу. Не враховуючи час на збір інформації про топологію мережі, який може становити близько тижня, безпосередньо моделювання загрози поширення забороненої інформації займає кілька годин навіть при використанні розподілених обчислювальних ресурсів. Аналітична модель може дати прогноз ЗгЗІ майже миттєво. З її допомогою можна отримати актуальні дані (до того моменту, коли кількість атакуючих абонентів буде максимальним) по динаміці ЗгЗІ.

Процес ЗгЗІ характеризується наступними особливостями. У мережі існують вузли трьох типів. Перший тип - атакуючі вузли, це вузли, які розповсюджують заборонену до поширення інформацію. Другий тип - захищені вузли, які характеризуються тим, що не беруть участі в поширенні забороненої інформації і ніколи не будуть цим займатися. Третій тип - потенційно вразливі. Вузли такого типу не беруть участі в процесі поширення загрози, але можуть бути схильні до негативного впливу з боку атакуючих вузлів і можуть почати поширювати заборонену інформацію.

Імітаційне моделювання:

Наведемо алгоритм реалізації ЗгЗІ, грунтуючись на описі процесів, що протікають в реальних КМ. Схема реалізації загрози представлена на рис. 1.

Крок 1. Поширення інформації забороненої до поширення (ЗІ) (далі процес «атаки») ініціює будь-який абонент-зловмисник (на рисунку - вузол 1), поширюючи повідомлення з ЗІ (реалізує загрозу) по його списку контактів. Атаку може починати один зловмисник або група.

Крок 2. Абоненти-одержувачі (вузли 2,3,4), прийнявши повідомлення з ЗІ, читають його і включаються в процес атаки, поширюючи його далі по своїх контактах (вузол 3), або ігнорують або взагалі видаляють повідомлення (вузол 2), тобто в атаці не беруть участь. Процес атаки зазвичай йде лавиноподібно. Атакуючі абоненти не закінчують атаку, одного разу передавши повідомлення із забороненою інформацією. Вікно атаки, як правило, триває протягом досить значного проміжку часу і залежить від типу подачі ЗІ в повідомленні, зацікавленості абонента і тд.

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

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

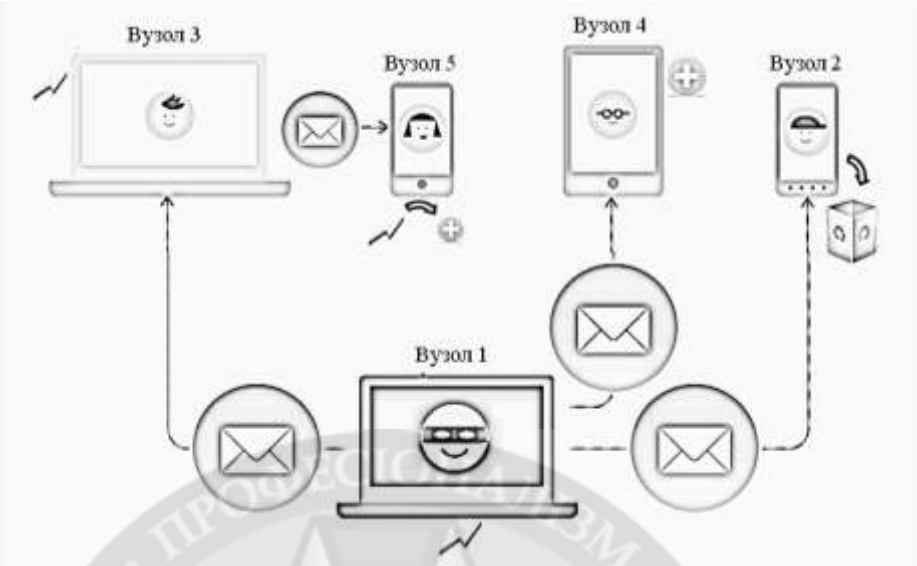

Рис. 1. Схема реалізації ЗгЗІ

Таким чином, ЗгЗІ в КМ являє собою складний динамічний процес, що складається з двох протидіючих підпроцесів атаки і захисту вузлів мережі. Для моделювання таких явищ часто застосовують епідеміологічні моделі, зокрема нашому опису точно відповідає SIR-модель Кермака-Маккендріка. Характер графіків, отриманих в результаті імітаційного моделювання (рис. 2 і 3), схожий з результатами, які дає дана модель. За наведеними вище причинами дана модель була взята за основу в цьому дослідженні. При використанні системи диференціальних рівнянь SIR-моделі для аналізу ЗгЗІ в КМ отримали результати у вигляді графіків, які хоча і правильно описують характер процесу, але не дають потрібної точності прогнозу.

На основі описаного алгоритму була побудована імітаційна модель ЗгЗІ в КМ.

Імітаційна модель ЗгЗІ:

Вхідні дані: N - кількість вузлів, що дорівнює кількості абонентів мережі, k - середній ступінь зв'язності вузлів, α - параметр, що відображає середню довжину шляху і рівень мережевої кластеризації, β - параметр, що відображає силу загрози, ймовірність здійснення атаки, γ - параметр відображає ступінь протидії загрозі, ймовірність захисту абонента (в моделі вважається, що β і γ однакові для кожного абонента), I_0 (абоненти-зловмисники - початкові джерела загрози) R_0 (абоненти, початково-несприйнятливі до атакуючих дій). Вихідні дані: $I(t)$, $R(t)$, $S(t)$ - чисельні масиви даних, що описують динамічний процес реалізації ЗгЗІ (кількості атакуючих, захищених і потенційно вразливих вузлів в кожну умовну одиницю часу відповідно).

Крок 1. Створення топології КМ – графа $G_{sw} =< V, E >$, 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 під це завдання, за рахунок заданих топологічних параметрів N, k, α .

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. $\forall v_i^I$ якщо $\exists e_{ij}$ і $v_j \in V^S$ $j = 1..N$, то з ймовірністю β виконати: $V^S \backslash v_j$ і $V^I \cup v_j$; з ймовірністю γ виконати: $V^I \backslash v_i$, $V^R \cup v_i$.

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

Після вибору типу мережі та введення параметрів відбувається імітаційне моделювання за алгоритмом реалізації ЗгЗІ. Потім результати відправляються в функцію побудови графіків для виведення результатів в графічному вигляді. На рис. 2 і 3 наведені результати імітаційного моделювання ЗгЗІ на мережі з параметрами $(N = 10000, \beta = 0{,}2, \gamma = 0{,}1, I_0 = 1, R_0 = 2000)$.

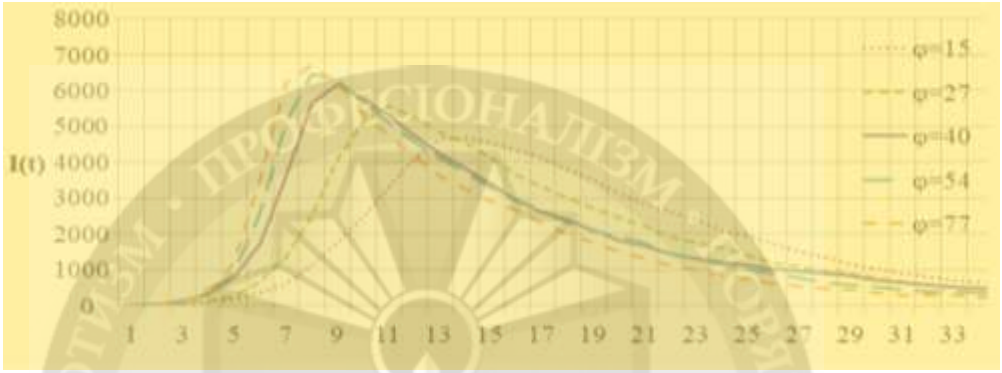

Рис. 2. Результати імітаційного моделювання ЗгЗІ

Рис. 3. Результати імітаційного моделювання ЗгЗІ

SIR (від англ. Susceptibles - Infectives - Removed with immunity) - епідеміологічна модель, спрощено описує поширення захворювання, що передається від одного індивіда до іншого, яка розглядає суб'єктів з точки зору трьох можливих станів: сприйнятливий, інфікований, імунізований.

Система диференціальних рівнянь, що описують SIR-модель, має вигляд [6,7]:

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

де $I(t)$ - кількість заражених (інфікованих) особин, $S(t)$ – кількість сприйнятливих особин, $R(t)$ - кількість «виключених з імунізацією» (removed with immunity) особин, $N = I(t) + S(t) + R(t)$ - кількість особин в популяції, γ - коефіцієнт відновлення / смерті, β - швидкість зараження (інфікування), t - час. Дана система є надлишковою – будь-яке рівняння з трьох рівнянь можна виключити.

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

В результаті отримали аналітичну модель, яка описується системою диференційних рівнянь:

$$\begin{cases} \dfrac{dI}{dt} = 2 \cdot \ln \varphi \cdot \beta \cdot \dfrac{S(t) \cdot I(t)}{N} - \gamma \cdot I(t) \\ \dfrac{dR}{dt} = \gamma \cdot I(t) \\ \dfrac{dS}{dt} = -2 \cdot \ln \varphi \cdot \beta \cdot \dfrac{S(t) \cdot I(t)}{N} \end{cases},$$

де φ - коефіцієнт топологічної вразливості мережі - показник, який обчислюється за формулою:

$$\varphi = \frac{k \cdot (C+1)}{L},$$

k - середній ступінь зв'язності вузлів мережі, C - коефіцієнт кластеризації мережі, L - середня довжина шляху мережі.

Релевантність результатів аналітичного рішення підтверджена серією експериментів на топологіях реальних мереж з використанням імітаційного моделювання. При цьому похибка для процесу захисту склала не більше 10%, для процесу атаки - не більше 15%.

Для моделювання ЗгЗІ необхідно мати топологію реального об'єкта. Пряме отримання цієї інформації утруднено у зв'язку з тим, що для підвищення точності результатів моделювання необхідно мати топологію всієї мережі. Отримати таку інформацію без прав адміністратора не представляється можливим. При зборі даних з правами абонента КМ маємо справу з двома типами вузлів: відкритими і закритими. Якщо в ході збору даних ми отримуємо ідентифікатори (id) вузла і суміжних з ним вузлів, то такий вузол називаємо відкритим. Якщо ж отримуємо тільки id вузла (абонент за допомогою налаштувань приховав інформацію про свої контакти), то такий вузол називаємо закритим. Також в мережі можуть існувати вузли, які з'єднані тільки з закритими вузлами. В такому випадку неможливо отримати навіть ідентифікатор вузла. Таких вузлів в мережі незначна частина. Емпірично показано, що закритих вузлів на порядок більше, ніж відкритих, тому при зборі даних втрачається значна частина інформації.

При наявності адміністративного ресурсу можна реалізувати автоматизовану систему протидії загрозі поширення забороненої інформації. Узагальнений алгоритм роботи такої

системи представлений на рис. 4. Розглянуті функції реалізуються за допомогою типових засобів.

Крок 1. Введення даних - типове повідомлення, що містить інформацію, заборонену до поширення. База даних таких повідомлень формується зі списку екстремістських матеріалів і реєстру доменних імен, покажчиків сторінок сайтів в мережі «Інтернет» і мережевих адрес, що дозволяють ідентифікувати сайти в мережі «Інтернет», що містять інформацію, поширення якої в Україні заборонено.

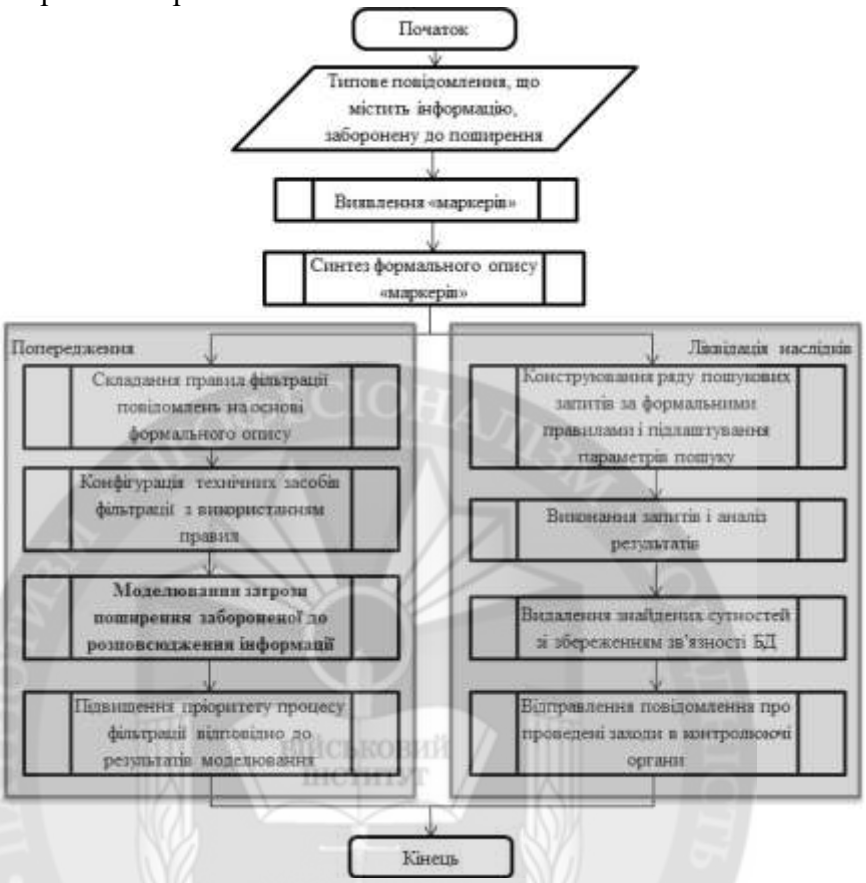

Рис. 4. Узагальнений алгоритм протидії поширенню загрози забороненої до розповсюдження інформації

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

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

Далі робота алгоритму розбивається на дві частини, паралельно виконуються процедури попередження і усунення наслідків загрози.

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 FASTBL, Apache Spamassassin, Kaspersky Antispam, Yandex Spamooborona, dnsbl і ін.

Крок 6а. Моделювання загрози поширення забороненої до розповсюдження інформації.

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

Ліквідація наслідків

Крок 4б. Конструювання ряду пошукових запитів за формальними правилами і підлаштування параметрів пошуку (пріоритет, глибина і тд.)

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.

Висновки. В результаті інформаційного огляду і проведених експериментів було виявлено, що на ЗгЗІ в комп'ютерних мережах істотно впливає топологія інформаційних зв'язків між абонентами і модель інформаційної взаємодії між ними. Виконаний аналіз основних підходів до моделювання ЗгЗІ показав, що найбільш адекватними моделями для цього завдання є моделі впливу, просочування і зараження. Створена імітаційна модель ЗгЗІ в КМ. З її допомогою проведені експерименти, результати яких показали залежність реалізації ЗгЗІ від топологічної вразливості мережі. Розроблено аналітичну модель. Похибка для процесу захисту склала не більше 10%, для процесу атаки - не більше 15%. Розроблено методику формування топології КМ. Створено алгоритм формування вихідних даних про топологію мережі (велика кількість вершин і зв'язків між ними доступної частини мережі), який враховує обмеження по збору даних. Розроблено алгоритм формування повного графа мережі з урахуванням додавання недоступної частини на основі обчислених прогнозованих топологічних характеристик. В ході експериментальних досліджень були отримані результати, що стосуються топології розглянутих КМ. Отримане значення середньої довжини шляху сходиться з результатами незалежних досліджень та дає можливість використовувати його в аналітичній моделі як фіксований параметр. В якості рекомендацій запропонований алгоритм роботи системи протидії поширенню загрози забороненої до розповсюдження інформації.

Приклади ефективного застосування механізмів прогнозування ЗгЗІ в КМ дають підставу констатувати адекватність і функціональність основних теоретичних побудов і розроблених на їх основі алгоритмічних і інструментальних засобів.

ЛІТЕРАТУРА:

1. Ferrara E. Topological features of Online Social Networks. Communications on Applied and Industrial Mathematics / E. Ferrara, G. Fiumara, 2011.

2. Easley D. Networks, Crowds and Markets Reasoning About a Highly Connected World / D. Easley, J. Kleinberg, 2010.

3. Albert R. Statistical mechanics of complex networks / R. Albert, A. Barabasi; Rewiews of Modern Physics, 2002.

4. Губанов Д.А. Соціальні мережі: моделі інформаційного впливу, керування і протистояння / Д.А. Губанов, Д.А. Новиков, А.Г. Чхартишвили; под. ред. чл..-корр. РАН Д.А. Новикова – М.: Видав. Фізико-математичної літератури, 2010. – 228 с.

5. Закон України про захист суспільної моралі [Електронний ресурс] – Режим доступу: http://zakon2.rada.gov.ua/laws/show/1296-15

6. Frauenthal, J.C. Mathematical Models in Epidemiology / J.C. Frauenthal; New York: Springer-Verlag, 1980. – 335 s.

7. Hethcote, H.W. The Mathematics of Infectious Diseases / H.W. Hethcote; - 2000.

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

REFERENCES:

1. Ferrara E. Topological features of Online Social Networks. Communications on Applied and Industrial Mathematics / E. Ferrara, G. Fiumara, 2011.

2. Easley D. Networks, Crowds and Markets Reasoning About a Highly Connected World / D. Easley, J. Kleinberg, 2010.

3. Albert R. Statistical mechanics of complex networks / R. Albert, A. Barabasi; Rewiews of Modern Physics, 2002.
4. Hubanov D.A. Sotsialni merezhi: modeli informatsiinoho vplyvu, keruvannia i protystoiannia / D.A. Hubanov, D.A. Novykov, A.H. Chkhartyshvyly; pod. red. chl..-korr. RAN D.A. Novykova – M.: Vydav. Fizyko-matematychnoi literatury, 2010. – 228 s.
5. Zakon Ukrainy pro zakhyst suspilnoi morali [Elektronnyi resurs] – Rezhym dostupu: http://zakon2.rada.gov.ua/laws/show/1296-15
6. Frauenthal, J.C. Mathematical Models in Epidemiology / J.C. Frauenthal; New York: Springer-Verlag, 1980. – 335 s.
7. Hethcote, H.W. The Mathematics of Infectious Diseases / H.W. Hethcote; - 2000.
8. Roberts, M.G., Heesterbeek, JAPMathematical models in epidemiology / M.G.Roberts, Heesterbeek JAP;In JA. Filar (Ed.) Mathematical Models.Oxford: EOLSS Publishers Ltd, 2004.

Рецензент: д.т.н., проф. Мясіщев О.А., завідувач кафедри кібербезпеки та комп'ютерних систем і мереж Хмельницького національного університету

д.т.н., доц. Гунченко Ю.А., к.т.н. Бойчук В.О., Юмашов В.С.

МЕТОД ВИЯВЛЕННЯ ЗАГРОЗИ ПОШИРЕННЯ ЗАБОРОНЕНОЇ ДО РОЗПОВСЮДЖЕННЯ ІНФОРМАЦІЇ В КОМП'ЮТЕРНИХ МЕРЕЖАХ

В статье рассматриваются компьютерные сети (КС), которые находяться под влиянием угрозы распространения запрещенной информации. Исследуются модели угрозы распространения этой информации и особенности компьютерных сетей. Приведены основные проблемы информационной безопасности, которые актуальны для даного исследования. Разрабатывается и исследуется имитационная модель, учитывающая топологические характеристики сети, а также особенности информационного взаимодействия абонентов как человеко-машинных систем. и аналитическая модель угрозы распространения запрещенной к распространению информации в КС с учетом топологической уязвимости сети. Релевантность результатов аналитического решения подтверждена серией экспериментов на топологии реальной сети с использованием имитационного моделирования. Также разрабатывается методика формирования топологии крупномасштабной КС с проведением исследований, которая учитывает основные топологические характеристики доступной части сети и работает в условиях недостаточной репрезентативности выборки исходных данных. Предложена система противодействия угрозе распространения запрещенной к распространению информации.

Ключевые слова: безопасность компютерных сетей, запрещенная информация, моделирования, формирование топологии.

Ph.D. Gynchenko Y.A., Ph.D. Boychuk V.O., Yumashov V.S.

METHOD FOR DETECTION OF THREATS DISTRIBUTE ILLEGAL TO SPREAD INFORMATION IN COMPUTER NETWORKS

The article deals with computer networks (CN), which are influenced by the threat of banned information. Investigate the threat model disseminate this information and especially computer networks. The basic problem of information security, which are relevant to our study.

Developed and researched simulation model that takes into account the topological characteristics of the network, and also features information interaction subscribers of human-machine systems. and the threat of analytical model forbidden to disseminate information to the CM based topological network vulnerability. The relevance of the results of analytical decision confirmed a series of experiments on real network topology using simulation. Also, the technique of forming a large-scale topology of KM research, which takes into account the basic topological characteristics of the available network up and running in low representativeness of the sample source data. The system to counter the threat propagation prohibited to disseminate information.

Keywords: security computer networks, prohibited information, modeling, forming topology.