УДК 51-74:004

МОДЕЛЮВАННЯ ПРОЦЕСІВ РОЗПОВСЮДЖЕННЯ ШКІДЛИВОГО ПЗ В МЕРЕЖІ ІНТЕРНЕТ

М. С. Дякуненко[1], І. В. Стьопочкіна[1]

[1] *Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Фізико-технічний інститут*

Анотація

В роботі обґрунтовано застосування математичних моделей для моделювання розповсюдження шкідливого програмного забезпечення: моделі SIR та дискретної математичної моделі — просторово розподіленого клітинкового автомату. Одержано функції переходу між станами для моделі клітинкового автомату. Моделювання на основі клітинкового автомату здійснено з урахуванням різних початкових умов та мережевих топологій. Проаналізовано та порівняно результати на основі обраних математичних моделей, які свідчать про успішне розв'язання поставленої задачі.

Ключові слова: SIR модель, клітинковий автомат, моделювання поширення шкідливого ПЗ

Вступ

На сьогоднішній день шкідливе програмне забезпечення є одним з найбільших ризиків для безпеки комп'ютерних мереж будь-якого розміру. Шкідливе програмне забезпечення (надалі шкідливе ПЗ) — це загальний термін, який використовується для позначення будь-якої комп'ютерної програми, створеної навмисно для здійснення несанкціонованої діяльності, яка в багатьох випадках наносить шкоду системі, в якій вона розміщується [1]. Боротьба з поширенням шкідливого ПЗ ведеться за допомогою антивірусного програмного забезпечення, а також шляхом введення політик та правил користування комп'ютерними мережами.

Саме тому важливою є розробка математичних моделей, що слугують для імітації поширення шкідливого ПЗ в комп'ютерних мережах. Подібна задача розв'язувалась в [2], однак принципи функціонування клітинкового автомату суттєво залежали від особливостей поведінки шкідливого ПЗ, які із часом стають неактуальними. Такі математичні моделі, що широко розповсюджені в інших галузях, наприклад, в епідеміології, були б дуже корисними для спеціалістів з інформаційної безпеки, оскільки дозволяли б моделювати розповсюдження шкідливого програмного забезпечення в мережі, та, таким чином, допомагали б у перевірці ефективності прийнятих контрзаходів, а також у прийнятті відповідних рішень щодо стримування поширення або, принаймні, мінімізації шкідливих наслідків. З часу перших спроб вивчення способів боротьби з комп'ютерними вірусами були проведені аналогії з біологічними вірусами, оскільки поведінка шкідливого ПЗ подібна до поведінки вірусів в людській популяції [1]. Здатність будь-якої математичної моделі імітувати поведінку інфекції значною мірою залежить від припущень, зроблених під час процесу моделювання. Наразі більшість із існуючих математичних моделей базуються на парадигмі, успадкованій від математичної епідеміології (зокрема, на моделі SIR — Кермака-Маккендріка).

Хоча дифузійні математичні моделі, що успадковані від математичної епідеміології, мають сильну математичну основу вони не враховують індивідуальні властивості елементів, що формують комп'ютерну мережу, оскільки використовують лише параметри що мають загальний характер, такі як швидкість зараження та швидкість одужання. Таким чином, у моделях, заснованих на диференціальних рівняннях, можна отримати хороші результати щодо глобальної поведінки, але вони не застосовні для моделювання в локальних та корпоративних мережах.

Окрім моделей, заснованих на диференціальних рівняннях, існує певна кількість моделей, заснованих на дуже різних парадигмах. З них можна виділити дискретну стохастичну модель, засновану на використанні клітинкового автомату. Клітинкові автомати — це прості обчислювальні моделі (особливий тип скінченних автоматів, здатних ефективно моделювати складні системи). Вони складаються з скінченної кількості одиниць, які називаються клітинами, які зв'язані між собою за допомогою певної топології. В кожен момент часу кожна клітина знаходиться в одному стані з числа кінцевої кількості можливих станів. Цей стан змінюється з дискретним плином часу відповідно до правил переходу, змінні яких є станами в попередній момент часу клітини та її сусідів [2]. За допомогою клітинкового автомату можна врахувати індивідуальні особливості елементів мережі, а також задавати будь-яку топологію мережі. Тому дискретні моделі, засновані на клітинкових автоматах, можуть подолати недоліки, що мають дифузійні математичні моделі, що описуються системою диференціальних рівнянь.

1. Постановка задач

Основною метою роботи є моделювання поширення шкідливого ПЗ в комп'ютерній мережі за допомогою моделей Кермака–Маккендріка SIR та клітинкового автомату, а також співставлення одержаних результатів. Цікавим питанням є порівняння відповідності цих результатів, одержаних для реальної мережі.

Виходячи з мети роботи слід розв'язати наступний перелік задач:

- Для розв'язування системи диференціальних рівнянь моделі SIR використати чисельні методи, зокрема, для розв'язання системи диференціальних рівнянь використати методи Рунге-Кутта 4-5 порядку.
- Підібрати параметри для моделі SIR, що якнайкраще апроксимують результат імітаційного моделювання клітинкового автомату.
- Побудувати дискретну математичну модель — просторово розподілений клітинковий автомат, а також запропонувати функції переходу між станами для моделі клітинкового автомату. Моделювання на клітинковому автоматі здійснюватиметься з урахуванням різних початкових умов та на графі реальної мережі.
- Порівняти результати моделювання, а також зробити висновки.

2. Обгрунтування вибору моделей для дослідження

Для моделювання поширення шкідливого ПЗ була використана SIR-подібна модель, оскільки, як вже зазначалося, поведінка шкідливого ПЗ подібна до поведінки вірусів в людській популяції, а також здійснено імітаційне моделювання за допомогою клітинкового автомату на реальній мережі.

2.1. Модель SIR

В моделі SIR чисельність елементів залишається незмінною [3]. Сприйнятливі елементи — комп'ютери, сервери в мережі (надалі елементи мережі), що можуть бути заражені шкідливим ПЗ, інфіковані — елементи мережі, що заражені шкідливим ПЗ і здатні його передавати, і, нарешті, вилучені — елементи мережі, що були інфіковані, але з різних причин в даний час є не зараженими (на карантині, виведені з ладу, відновлені антивірусним ПЗ). Відповідно розглядаються три залежні від часу змінні: кількість сприйнятливих елементів мережі $S(t)$, кількість заражених $I(t)$ та кількість вилучених, $R(t)$. Оскільки вважається, що кількість елементів мережі залишається постійною, маємо, що $S(t) + I(t) + R(t) = N$ для будь-якого моменту часу t, де N являє собою загальну кількість елементи мережі. Крім того, беруться до уваги два параметри, а саме: швидкість передачі, β і швидкість відновлення, α, (швидкість, з якою заражений елемент мережі відновлюється за одиницю часу). Динаміка моделі регулюється наступним набором диференціальних рівнянь [3]:

$$\begin{cases} \dfrac{dS}{dt} = -\beta \cdot S \cdot I \\ \dfrac{dI}{dt} = \beta \cdot S \cdot I \ - \ \alpha \cdot I \\ \dfrac{dR}{dt} = \alpha \cdot I \end{cases} \tag{1}$$

Позначення $S'(t)$, $I'(t)$ і $R'(t)$ позначає відповідні похідні $S(t)$, $I(t)$ і $R(t)$ відносно часу t, тобто варіацію в часі цих величин. Перше рівняння системи, $S'(t) = -\beta \cdot S(t) \cdot I(t)$, встановлює, що зміна кількості сприйнятливих елементів мережі із часом залежить від кількості контактів між сприйнятливими та зараженими елементами мережі, а також від швидкості передачі. Друге рівняння системи, $I'(t) = \beta \cdot S(t) \cdot I(t) - \alpha \cdot I(t)$, вказує на те, що зміна кількості заражених елементів мережі є різницею між новими зараженими елементами, та тими, які вилучені з мережі. Нарешті, третє рівняння, $R'(t) = \alpha \cdot I(t)$, показує, що приріст вилучених елементів пропорційний кількості інфікованих, де коефіцієнт відновлення є константою пропорційності. В [1] встановлено, що коли $\beta \leq \alpha$, $(R_0 \leq 1)$, де $R_0 = \beta/\alpha$ – це усереднена кількість додатково заражених, які виникають при введенні одного зараженого елемента в абсолютно сприйнятливу популяцію, то система досягає стану рівноваги без зараження. Тоді як якщо $\beta > \alpha$, $(R_0 > 1)$, то спостерігається збільшення кількості заражених. Модель SIR моделює розповсюдження шкідливого ПЗ виходячи із початкової кількості вузлів у різних станах, та зі значень коефіцієнтів швидкості передачі та швидкості відновлення.

2.2. Клітинковий автомат

Клітинковий автомат (надалі КА) — це проста модель обчислень, складена з одиниць пам'яті, що називаються клітинками: $k_1, k_2, ..., k_n$, які розташовуються за певною топологією, визначеною за допомогою графу комплексної мережі $G = (V, E)$. Як наслідок, кожен елемент клітинкового автомату означає вершину G: $k_i \in V$ з $1 \leq i \leq n$, де $n - |V|$ і локальні взаємодії між цими клітинами моделюються за допомогою ребер: існує зв'язок між k_i і k_j, якщо $(k_i, k_j) \in E$ [4].

Кожна клітина k_i наділена станом на кожному кроці часу t: $s_i^t \in \mathcal{S}$, де $\mathcal{S}$ є кінцевим набором станів. Цей стан змінюється за допомогою локальної функції переходу, змінними якої є стани основного елементу та сусідніх елементів на попередньому кроці часу.

Сприйнятливі пристрої — це ті, які не заражені шкідливим програмним забезпеченням (пристрій не містить шкідливого коду); інфіковані пристрої характеризуються тим, що на них потрапило шкідливе програмне забезпечення і вони здатні поширити його на пов'язані з ними елементи мережі, і, нарешті, вилучені пристрої — це ті пристрої, які по різним причинам вилучені з мережі і не здатні передавати шкідливе ПЗ. Коефіцієнти КА такі: параметр зара-

женння $0 \leq h \leq 1$, це ймовірність того що пристрій стане зараженим, та матиме можливість заразити інші поєднані з ним елементи, параметр $0 \leq a \leq 1$ означає ймовірність вилучення елементу з мережі після зараження.

Модель розповсюдження базується на клітинковому автоматі, основними характеристиками якого є наступні:

- Кожен вузол $k_i \in V$ графу G являє собою клітинку КА.
- Набір станів формується трьома станами: сприйнятливим (S), інфікованим (I) та вилученим (R): $\mathcal{S} = \{S, I, R\}$.
- Клітинка КА, стан якої $s_i^t = I$, може заразити лише пов'язані з нею елементи $\{k_j | (k_i, k_j) \in E \wedge s_j^t = S\}$.
- Локальними перехідними функціями між станами є наступні:
  - Перехід від сприйнятливого до зараженого: якщо $s_j^t = S$ то $s_j^{t+1} = I$ з ймовірністю h. Очевидно, що вузол залишиться сприйнятливим ($s_j^{t+1} = S$) з ймовірністю $1 - h$.
  - Перехід від зараженого до вилученого: якщо $s_i^t = I$ то $s_i^{t+1} = R$ з ймовірністю a. Вузол може залишитися інфікованим ($s_i^{t+1} = I$) з ймовірністю $1 - a$.

3. Постановка експерименту та результати

3.1. Вихідні дані

Для дослідження поширення шкідливого ПЗ було використано датасет з сайту `https://snap.stanford.edu/data/Oregon-1.html`. Це граф маршрутизаторів, що організовані у підграфи, які називаються автономними системами (АС). Кожна АС обмінюється потоками трафіку з деякими сусідами. Виведений з огляду маршрутизаторів в Орегоні між 31 березня 2001 року та 26 травня 2001 року.

Характеристики графу вказані в Таблиці 1. Граф також є неорієнтованим, ненаправленим та без кратних ребер.

Табл. 1. Характеристики графу

Кількість вузлів	11174
Кількість зв'язків	23409
Середній коефіцієнт кластеризації	0.2964
Кількість трикутників	19894
Діаметр графу	9

3.2. Програмна реалізація моделей

Перш за все для клітинкового автомату були обрані ймовірності переходу між станами, від сприйнятливого до зараженого і від зараженого до вилученого. В даному випадку ймовірність переходу від сприятливого до зараженого буде 0.85, відповідно ймовірність того що елемент залишиться сприйнятливим — 0.15. Ймовірність того, що елемент з зараженого перейде в вилучений — 0.25, відповідно ймовірність того, що елемент залишиться зараженим — 0.75. На нульовый ітерації були заражені вузли $\{1, 22, 12294, 9920, 10875\}$. На Рис. 1 наведено динаміку зміни станів елементів мережі за ітераціями.

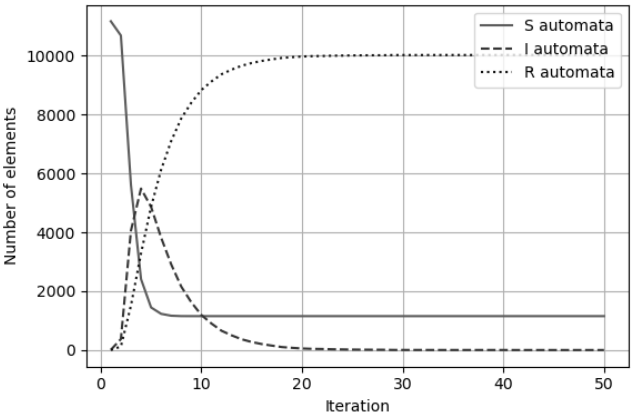

Рис. 1. Результат імітації на клітинковому автоматі

В Табл. 2 наведені отримані дані після завершення моделювання поширення ПЗ в мережі.

Табл. 2. Результат імітації на клітинковому автоматі

Максимальна кількість заражених вузлів	5450
Кількість вилучених на останній ітерації	10017
Кількість сприйнятливих на останній ітерації	1158

На Рис. 2 зображено графік моделі SIR. Параметри SIR: $\beta = 0.00022,\ \alpha = 1.0$. Параметри підібрані так, щоб наблизити кількість вилучених і сприйнятливих до результату імітаційної моделі. В Табл. 3 наведено результат моделі SIR. Як можна побачити, при імітації на клітинковому автоматі видно значний сплеск заражених елементів у порівнянні з моделлю SIR.

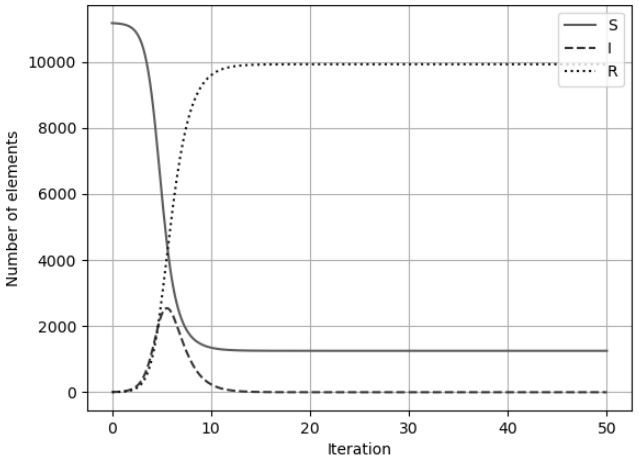

Рис. 2. Результат моделі SIR

Табл. 3. Результат моделі SIR

Максимальна кількість заражених вузлів	2545.899
Кількість вилучених на останній ітерації	9924.801
Кількість сприйнятливих на останній ітерації	1253.199

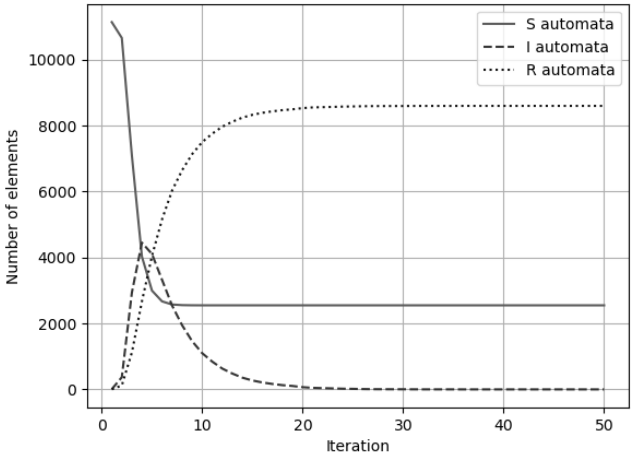

Рис. 3. Результат імітації на клітинковому автоматі при видаленні 3 вершин з найбільшою кількістю зв'язків

За допомогою програми `Gephi 0.92` визначено, що перші три вузли з найбільшою кількістю зв'зків — $\{701, 1239, 7018\}$. При їх видаленні з мережі кількість заражених значно зменшується у порівнянні з результатами в Табл. 2. Графік імітаційного моделювання наведений на Рис. 3, результат моделювання наведений в Табл. 4.

Табл. 4. Результат імітації на клітинковому автоматі при видаленні 3 вершин з найбільшою кількістю зв'язків

Максимальна кількість заражених вузлів	4453
Кількість вилучених на останній ітерації	8625
Кількість сприйнятливих на останній ітерації	2550

Висновки

При застосуванні моделі Кермака-Маккендрика використовуються такі параметри, як швидкість зараження та швидкість одужання, які мають загальний характер. Значення параметра є постійним для всіх елементів мережі, або в деяких випадках воно відповідає заданому розподілу ймовірностей. Відповідно, використання параметрів, індивідуалізованих для кожного з елементів мережі, не розглядається. В тому числі в моделі не враховуються зв'язки між вузлами. Оскільки,наприклад, якщо буде заражений вузол з великою кількістю зв'язків це може спричинити можливе зараження великої кількості пов'язаних з ним вузлів і значне зростання кількості заражених елементів в мережі та пришвидшення передачі шкідливого ПЗ. Коли поширення шкідливого програмного забезпечення аналізується макроскопічно, отримані результати забезпечують досить хорошу апроксимацію того, що насправді відбувається. Однак, якщо аналізувати таке поширення в локальних мережах, отримані результати погано апоксимують реальні дані, оскільки в мікроскопічному масштабі динаміка дуже чутлива до локальних зв'язків. Таким чином, у моделях що базуються на диференціальних рівняннях, можна отримати хороші результати щодо глобальної поведінки, але нам не вистачить інформації про індивідуальну поведінку кожного з елементів у мережі.

Щоб подолати недоліки моделей, що базуються на диференціальних рівняннях, необхідно розглянути моделі, що враховують індивідуальні характеристики кожного елемента системи, та зв'язки між елементами. Основні недоліки, притаманні цим моделям, можна виправити застосуванням іншого типу моделі, заснованої на клітинкових автоматах. При цьому можна було б врахувати індивідуальні характеристики кожного з комп'ютерів або пристроїв, які підключені до мережі. Моделі на основі клітинкових автоматів дають змогу розглянути різні топології мережі, а також здійснити імітаційне моделювання для реальної мережі. На основі результатів, отриманих з моделювання поширення шкідливого ПЗ за допомогою клітинкового автомату, можна простежити динаміку розповсюдження шкідливого коду і зробити висновки про принципи протидії поширенню шкідливого ПЗ. Зокрема, вузли, які відіграють критичну роль у поширенні шкідливого ПЗ добре виділяються серед інших на основі цього підходу; і саме щодо них треба у першу чергу застосовувати заходи антивірусного захисту та лікування. Також вони можуть бути центрами поширення "ліків" по всіх дочірніх вузлах, що дасть змогу природно розповсюдити відповідні складові антивірусного захисту саме у ті вузли, де вони потрібні.

Перелік використаних джерел

1. Rey, Ángel Mathematical modeling of the propagation of malware: A review. *Security and Communication Networks.*— 2015. — P.2561–2579.
2. І.В Стьопочкіна, М.В. Грайворонський Моделювання розпосюдження комп'ютерних вірусів на основі імовірнісного клітинкового автомату. *//Захист інформації.* — 2015. — №4, С.1-9.
3. W. O. Kermack and A. G. McKendrick A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character.* — 1927. — P.700-721. — `http://www.jstor.org/stable/94815`
4. Rey, Ángel and Queiruga-Dios, Araceli and Hernández, Guillermo and Tabernero, A. Modeling the Spread of Malware on Complex Networks. — 2020. — P.109-116