УДК 519.6

Н.О. Возняк
Львівський національний університет ім. І.Франка,
кафедра прикладної математики

МОДЕЛЮВАННЯ ПОШИРЕННЯ КОМП'ЮТЕРНОГО ВІРУСУ В ІНТЕРНЕТІ

© *Возняк Н.О., 2005*

Наведено модель поширення комп'ютерного вірусу в інтернеті. Одержано умови початку поширення вірусу в зв'язаних підмережах. Проведено числовий аналіз моделі.

Model of computer virus spreading in internet is presented. Threshold conditions have been obtained for the model of computer virus spreading in connected networks. Numerical analysis of the model has been provided.

1. Постановка проблеми

Математичне дослідження комп'ютерних вірусів почалося значно пізніше ніж виник перший комп'ютерний вірус. Перші тривожні сигнали надійшли в період стрімкого розвитку глобальної мережі Інтернет. З того часу кількість шкідливих програм швидко зростає. І хоча протягом останніх семи років за кордоном з'явилися публікації за цією тематикою, в Україні проблема дослідження і аналізу поширення вірусів в інформаційних системах мало розглядається.

Відомо, що понад 95 % уражень хостів в глобальній мережі Інтернет здійснюється "хробаками" – шкідливими комп'ютерними програми, які мають здатність самостійно поширюватися каналами інформаційних систем, атакуючи та інфікуючи незахищені хости. Імовірність інфікації хоста та його подальшого поширення великою мірою залежать від того, наскільки інтенсивно цей хост взаємодіє з іншими вузлами мережі. Різнорідність підмереж, очевидно, впливає на швидкість поширення вірусу в кожній з них, тому велике значення має інтенсивність обміну інформацією між підмережами. На жаль, до цього часу недостатньо уваги приділяється дослідженням динаміки поширення комп'ютерних вірусів у великих мережах, які складаються з пов'язаних між собою підмереж.

2. Аналіз останніх досліджень

Одна з перших робіт [8], в якій було систематизовано й розглянуто основні питання комп'ютерної вірусології з точки зору математики, була написана в 1998 році. І хоча перший комп'ютерний вірус з'явився ще 1988 року, наслідки поширення його та інших вірусів залишилися недооціненими. Зокрема, в цій роботі наголошувалося, що важливим напрямком досліджень є експериментальне моделювання процесів поширення комп'ютерних вірусів і збирання статистичних даних. Було звернено увагу [3] на схожість і відмінність поширення вірусів у біологічних системах і комп'ютерних вірусів в інформаційних системах. Спроби застосувати досвід, одержаний в галузі математичної епідеміології, до інформаційних систем дали позитивні результати [3, 7]. Автори публікацій [4, 5, 8–10] виділяли закономірності, характерні для поширення комп'ютерних вірусів і модифікували відомі біологічні моделі для випадку комп'ютерних вірусів.

3. Цілі статті

Метою роботи є побудова моделі поширення комп'ютерного вірусу в складній мережі, яка складається з пов'язаних між собою підмереж. Іншою важливою проблемою, яка аналізується в статті, є визначення умов початку епідемії комп'ютерного вірусу в складних мережах.

4. Основний матеріал

4.1. Моделі поширення комп'ютерних вірусів

Для моделювання поширення біологічних та комп'ютерних вірусів зручно скористатися відомою SIR моделлю [3]. У випадку поширення комп'ютерних вірусів вона базується на поділі

множини хостів мережі на три підгрупи: незахищені (susceptible), інфіковані (infectious), імунізовані (removed). Нехай $S(t)$ – кількість незахищених хостів в момент часу t, $I(t)$ – кількість інфікованих хостів в момент часу t, $R(t)$ – кількість імунізованих хостів в момент часу t. SIR модель описують системою звичайних диференційних рівнянь:

$$\begin{cases} \dfrac{dS(t)}{dt} = -\beta S(t)I(t) \\ \dfrac{dI(t)}{dt} = \beta S(t)I(t) - \mu I(t), \\ \dfrac{dR(t)}{dt} = \mu I(t) \end{cases} \tag{1}$$

де $\beta > 0$ – ймовірність інфікування незахищеного хоста, $\mu > 0$ – ймовірність імунізації інфікованого хоста.

Для того, щоб одержати задачу Коші, систему рівнянь (1) доповнюють початковими умовами виду:

$$S(t_0) = S_0, I(t_0) = I_0, R(t_0) = R_0, t \geq t_0. \tag{2}$$

Одне з базових припущень, яке використовують для побудови моделі (1)–(2), полягає в тому, що хост одержує імунітет проти вірусного зараження лише тоді, коли він буде зараженим, проте не виключено, що деякі хости будуть імунізовані ще до того, коли на них потрапить вірус. Крім того, ускладнивши модель (1)–(2), можна врахувати в динаміці розвитку вірусу інтервал часу, коли ще відсутній ефективний антивірусний захист (вірус може вільно поширюватися в мережі) і інтервал часу, коли такий антивірус вже створено (тоді вірус розвивається за звичайною SIR моделлю). Такі складніші моделі та їхнє дослідження запропоновано в роботах [7, 9].

4.2. Поширення вірусу в інтернеті

Розглянемо складну комп'ютерну мережу (інтернет) як сукупність пов'язаних мереж. Мережі, що входять до складної мережі, назвемо підмережами.

Для узагальнення моделі (1)–(2) на випадок взаємопов'язаних K підмереж ми ввели коефіцієнти проникнення між підмережами δ_{ij}, які характеризують якість антивірусного бар'єра між підмережами, і отримали таку модель:

$$\begin{cases} \dfrac{dS_i(t)}{dt} = -\beta_i S_i(t)I_i(t) - \sum\limits_{\substack{j=1 \\ j\neq i}}^{K} \delta_{ji} S_i(t)I_j(t) \\ \dfrac{dI_i(t)}{dt} = \beta_i S_i(t)I_i(t) - \mu_i I_i(t) + \sum\limits_{\substack{j=1 \\ j\neq i}}^{K} \delta_{ji} S_i(t)I_j(t), i = 1,2,\ldots,K \\ \dfrac{dR_i(t)}{dt} = \mu_i I_i(t) \\ N_i = S_i(t) + I_i(t) + R_i(t) \end{cases} \tag{3}$$

де S_i – кількість незахищених хостів в i-й підмережі, I_i – кількість інфікованих хостів в i-й підмережі, R_i – кількість захищених хостів в i-й підмережі, β_i – швидкість поширення вірусу в i-й підмережі, μ_i – швидкість імунізації хостів в i-й підмережі відповідно. Модель (3) доповнюється початковими умовами (4):

$$\begin{cases} S_i(t_0) = S_{i0} \\ I_i(t_0) = I_{i0}, i = 1,2,\ldots,K. \\ R_i(t_0) = R_{i0} \end{cases} \tag{4}$$

Кожен з коефіцієнтів δ_{ij} задає ймовірності потрапляння вірусу з підмережі i в підмережу j.

Для практичного обчислення значень коефіцієнтів проникнення вірусу між підмережами можна здійснювати моніторинг комп'ютерів підмереж і обчислювати кількість вірусних атак між підмережами. Тоді ймовірність потрапляння вірусу до i підмережі за деякий час буде відповідно $\delta_{ji} = \frac{\int_{t_0}^{t_e} I_i(t)dt}{\int_{t_0}^{t_e} A_{ji}(t)dt}$, де I_i – кількість інфікованих комп'ютерів в момент часу t, A_{ji} – загальна кількість вірусних атак з підмережі j в підмережу i в момент часу t, t_0, t_e – час початку і кінця моніторингу відповідно.

Інший спосіб, який можна використати для того, щоб визначити параметри моделі (3)–(4), полягає в використанні процедури ідентифікації параметрів [6]. Розв'язок математичної моделі (3)–(4) є розв'язком задачі Коші для системи звичайних диференційних рівнянь, яка у загальному випадку може бути записана у вигляді:

$$\begin{aligned} y'(t) &= f(t, y(t), p), p > p_0 \\ y(t_0) &= y_0 \end{aligned}, \tag{5}$$

де $y(t,p) = (y_1(t,p), y_2(t,p), \ldots, y_n(t,p))^T$ – залежна змінна, $p = (p_1, \ldots, p_L)^T \in R^L$ – параметри моделі, які підлягають ідентифікації.

Процедура ідентифікації полягає у такому підборі значення параметрів p, щоб розв'язок $y(t_j,p)$ задачі (5) якнайменше відрізнявся від результатів вимірювань $y_j^{(e)} = \left(y_{1j}^{(e)}, y_{2j}^{(e)}, \ldots, y_{nj}^{(e)}\right)$ у вказаних точках t_j, $j = 1,m$.

Для розв'язання поставленої задачі необхідно мінімізувати функціонал:

$$\Phi(p) = \sum_{i=1}^{n}\sum_{j=1}^{m}\left[y_i(t_j, p) - y_{ij}^{(e)}\right]^2. \tag{6}$$

Вектор параметрів p може складатися як з параметрів δ_{ij}, так і з параметрів β_i, μ_i.

Отже, користуючись даними спостережень за розвитком епідемії, можна ідентифікувати параметри моделі $\delta_{ij}, \beta_i, \mu_i$, а одержаний результат буде тим точніший, чим більше результатів вимірювань ми задамо.

4.3. Аналіз поширення комп'ютерного вірусу в інтернеті

Визначимо для моделі (3)–(4) умови, за яких можливий розвиток епідемії в підмережі. Умовою початку епідемій в підмережах є додатне значення похідної від функції кількості інфікованих хостів в підмережі за часом:

$$\left.\frac{dI_i}{dt}\right|_{t=0} = \beta_i S_{i0} I_{i0} - \mu_i I_{i0} + \sum_{\substack{j=1 \\ i\neq j}}^{K} \delta_{ij} S_{i0} I_{j0} = S_{i0}\left(\beta_i I_{i0} + \sum_{\substack{j=1 \\ i\neq j}}^{K} \delta_{ij} I_{j0}\right) - \mu_i I_{i0} > 0 \tag{7}$$

звідки

$$S_{io} > \frac{\mu_i I_{io}}{\beta_i I_{io} + \sum_{\substack{j=1 \\ i\neq j}}^{K} \delta_{ij} I_{jo}}, \tag{8}$$

Одержаний результат означає, що в підмережі i не почнеться епідемія комп'ютерного вірусу, якщо початкова кількість незахищених хостів не перевищує певної величини, яка задається умовою (8).

Спробуємо оцінити вплив коефіцієнтів проникнення між підмережами δ_{ij} на умову (8). Для спрощення викладок будемо вважати, що кількість інфікованих хостів в кожній підмережі в початковий момент часу є однаковою. Розглянемо таку функцію:

$$S_{i0}(\delta_{i1},\ldots,\delta_{iK}) = \frac{\mu_i}{\beta_i + \sum\limits_{\substack{j=1 \\ i\neq j}}^{K} \delta_{ij}}, \tag{9}$$

Оскільки ми розглядаємо коефіцієнти проникнення вірусу між підмережами як ймовірності потрапляння вірусу з підмережі в підмережу, то $0 \le \delta_{ij} \le 1$. Максимальне значення $S_{i0}^{\max} = \frac{\mu_i}{\beta_i}$ функції (9) досягається в точці (0,...,0), а мінімальне – $S_{i0}^{\min} = \frac{\mu_i}{\beta_i + K}$ в точці (1,...,1).

Тобто, можна зробити висновок, що зі зростанням значень коефіцієнтів проникнення між підмережами зменшується порогове значення, за якого починається епідемія, а отже, чим більшими є коефіцієнти проникнення між підмережами, тим швидше розпочнеться епідемія в іншій підмережі.

З нерівності (8) одержуємо ще одну оцінку початку розвитку епідемії:

$$I_{i0} > \frac{S_{i0}\sum\limits_{\substack{j=1 \\ i\neq j}}^{K} \delta_{ij} I_{j0}}{\beta_i S_{i0} - \mu_i}, \tag{10}$$

Останній результат означає, що в підмережі i не почнеться епідемія комп'ютерного вірусу доти, поки початкова кількість хворих хостів в підмережі i не почне задовольняти нерівність (10).

4.4. Числові дослідження моделі поширення вірусу для випадку трьох підмереж

Розглянемо складну мережу, яка складається з трьох підмереж. Кожна підмережа містить 1000 хостів. У початковий момент часу всі хости є незахищеними, один хост в третій підмережі є інфікованим. У цьому випадку початкові умови задачі Коші є такими:

$$S_{10} = 1000, I_{10} = 0, R_{10} = 0, S_{20} = 1000, I_{20} = 0, R_{20} = 0, S_{30} = 999, I_{30} = 1, R_{30} = 0$$

Вхідні параметри $\beta_i, \mu_i, i = \{1,2,3\}$, які характеризують швидкості поширення вірусу і швидкості імунізації хостів для кожної з підмереж виберемо такими, які характеризують поширення "хробака" CodeRed в мережі Internet [9]: $\beta_i = 0.8 \cdot 10^{-5}, \mu_i = 0.01, i = \{1,2,3\}$. Поширенню цього "хробака" присвячено декілька досліджень [4,9], тому ми можемо порівняти одержані результати з результатами, одержаними в інших роботах.

Для випадку, коли підмережі не пов'язані між собою, тобто значення всіх коефіцієнтів проникності дорівнюють нулю, на рис. 1, а наведено графіки розподілу за часом кількості незахищених (лінія 1), інфікованих (лінія 2) та імунізованих (лінія 3) хостів в третій підмережі. У першій та другій підмережах жодних змін не відбувається: кількість незахищених комп'ютерів є сталою і дорівнює 1000, а кількість імунізованих і інфікованих дорівнює нулю.

Розглянемо випадок, коли підмережі пов'язані між собою з коефіцієнтами $\delta_A = \delta_{12} = \delta_{21} = 0, \delta_B = \delta_{23} = \delta_{32} = 10^{-3}, \delta_C = \delta_{13} = \delta_{31} = 10^{-6}$, що відповідає ситуації, коли підмережі 1 і 2 не зв'язані, підмережі 2 і 3 зв'язані з коефіцієнтом проникнення 10^{-3}, підмережі 1 і

3 зв'язані з коефіцієнтом проникнення 10^{-6}. На рис. 1, б наведено графіки розподілу за часом кількості незахищених (лінії 1, 4, 7), інфікованих (лінія 2, 5, 8) та імунізованих (лінія 3, 6, 9) хостів в першій, другій, третій підмережах відповідно. У першій та другій початково не заражених підмережах інфікування проходить з деяким запізненням.

Для випадку, коли підмережі пов'язані між собою з коефіцієнтами проникності $\delta_A = \delta_{12} = \delta_{21} = \delta_B = \delta_{23} = \delta_{32} = \delta_C = \delta_{13} = \delta_{31} = 10^{-6}$, що відповідає ситуації, коли всі підмережі є зв'язаними з однаковими коефіцієнтами проникнення 10^{-6}, на рис. 1, в наведено графіки розподілу за часом кількості незахищених (лінії 1, 4, 7), інфікованих (лінія 2, 5, 8) та імунізованих (лінія 3, 6, 9) хостів в першій, другій, третій підмережах відповідно. В першій та другій початково не заражених підмережах інфікування проходить з деяким запізненням, як і в попередньому випадку, однак в цьому випадку запізнення є сталим, тому лінії накладаються одна на одну.

Нами досліджено вплив значення коефіцієнта проникнення між підмережами на поширення вірусу. Результат було одержано для моделі (3)–(4) у випадку двох підмереж, зв'язаних з коефіцієнтами $\delta = \{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}\}$ для тих самих початкових умов. На рис.2 зображено розв'язки *I(t)* моделі (3)–(4) за різних значень параметрів δ (для спрощення вважатимемо, що в кожному окремо взятому випадку взаємовплив підмереж є однаковим). Одержаний розв'язок дає змогу стверджувати, що зі зменшенням зв'язності мережі збільшується час відставання між розвитком епідемії. Водночас спостерігаємо, що зменшення зв'язності мережі майже не призводить до зменшення кількості уражених хостів.

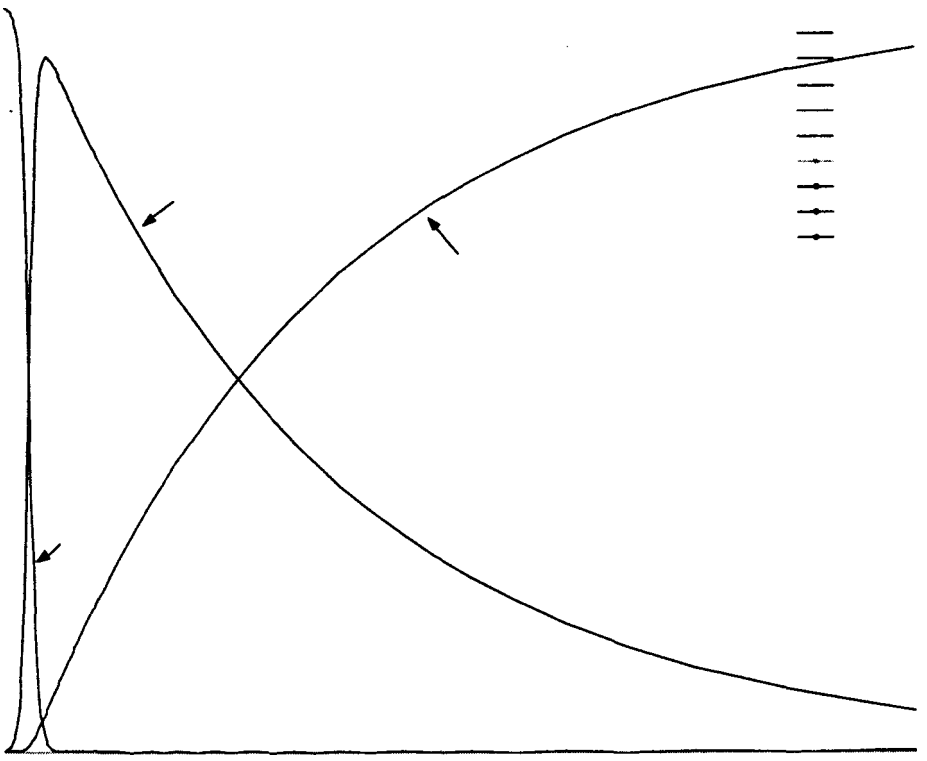

Рис. 1, а. Розв'язок моделі (3)–(4) за умов
$\delta_A = \delta_B = \delta_C = 0, N_1 = N_2 = N_3 = 1000, I_1(0) = I_2(0) = 0, I_3(0) = 1$

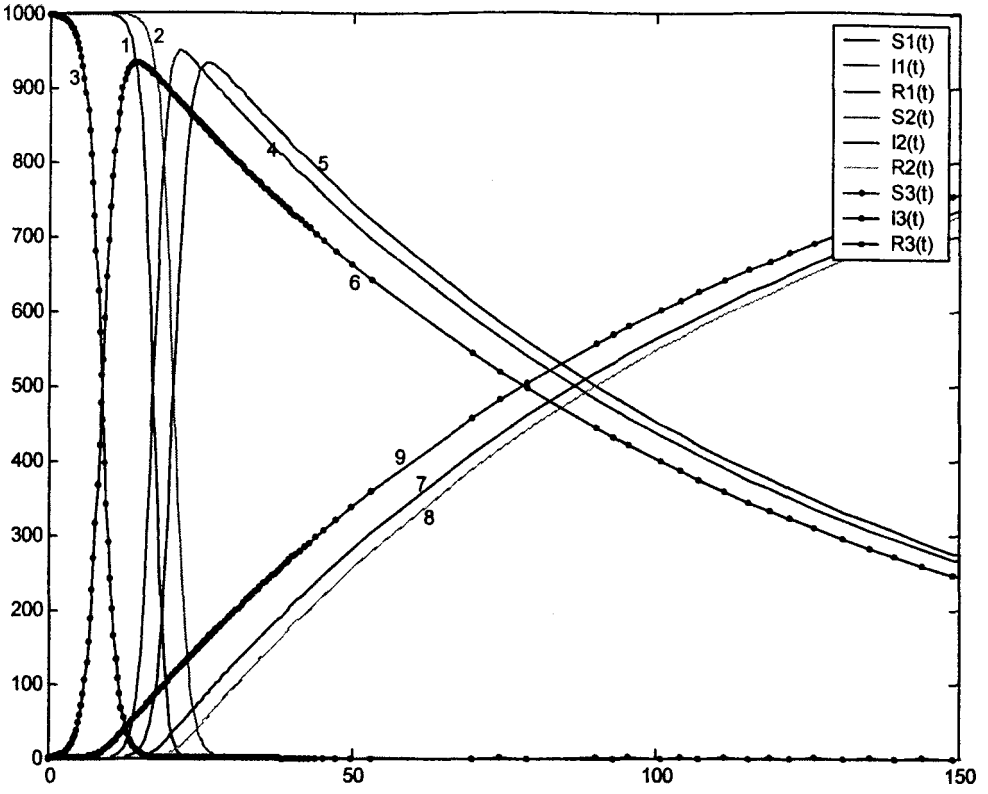

Рис. 1, б. Розв'язок моделі (3)–(4) за умов

$\delta_A = 0, \delta_B = 10^{-3}, \delta_C = 10^{-6}, N_1 = N_2 = N_3 = 1000, I_1(0) = I_2(0) = 0, I_3(0) = 1$

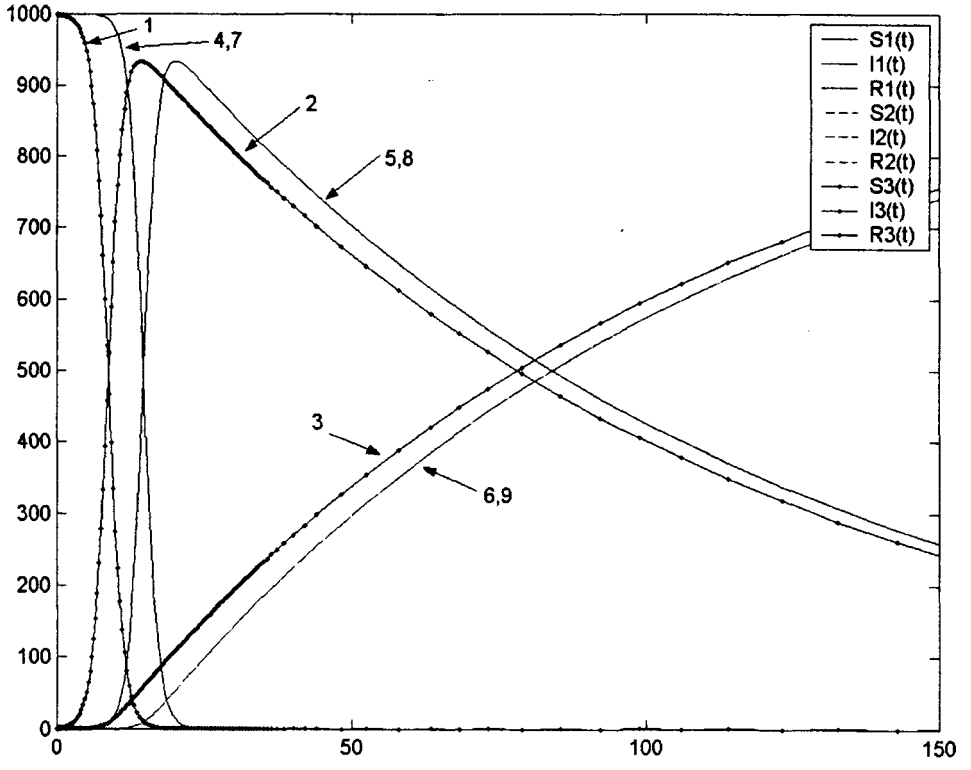

Рис. 1, в. Розв'язок моделі (3)–(4) за умов

$\delta_A = 10^{-6}, \delta_B = 10^{-6}, \delta_C = 10^{-6}, N_1 = N_2 = N_3 = 1000, I_1(0) = I_2(0) = 0, I_3(0) = 1$

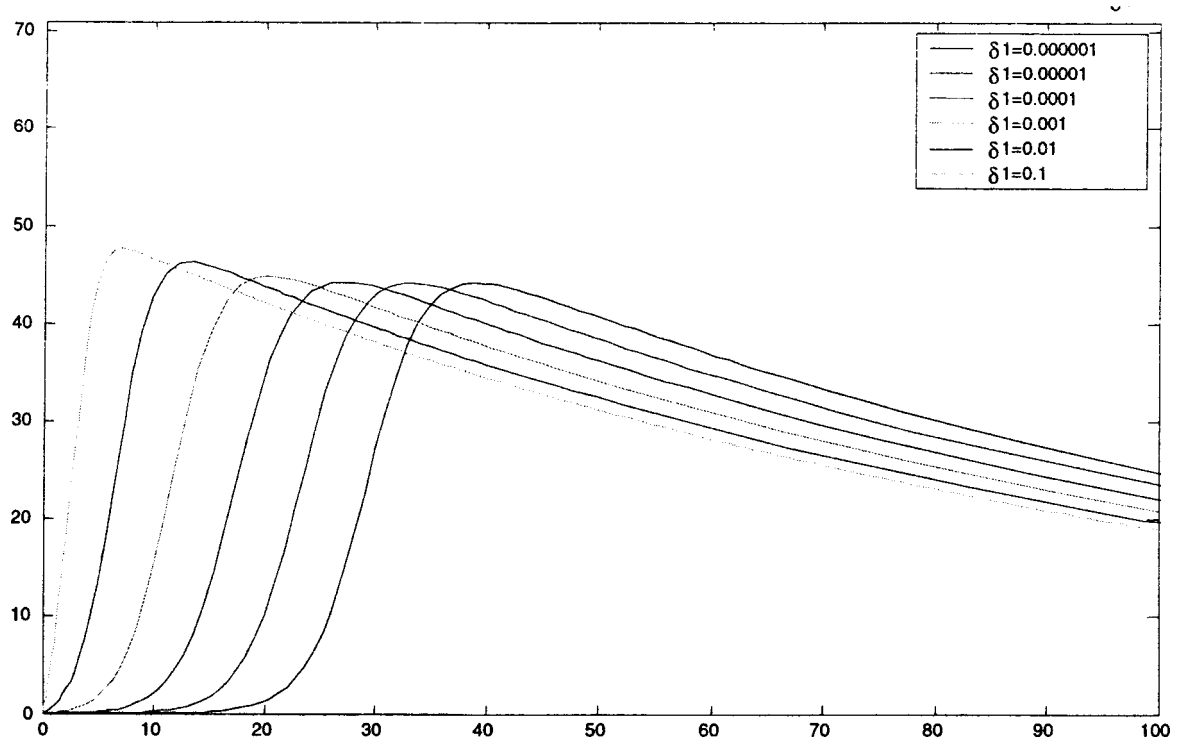

Рис. 2. Кількість інфікованих хостів I(t) у початково не інфікованій підмережі залежно від часу після потрапляння в неї вірусу для різних коефіцієнтів проникності між підмережами

Висновки

Запропонована нами математична модель описує поширення комп'ютерного вірусу у складній мережі, яка складається з K підмереж. Модель дає змогу спрогнозувати поширення вірусу в інтернеті і може бути корисною для практичного аналізу структури мережі і для теоретичних досліджень. Знайдено числовий розв'язок запропонованої моделі.

Проведені нами дослідження динаміки поширення вірусів в складній мережі показали, що значну роль у швидкості поширення комп'ютерних вірусів відіграє активність обміну інформацією між мережами та наявність антивірусного бар'єра у точках взаємопід'єднання мереж. Тому під час побудови сучасних комп'ютерних мереж необхідно враховувати величину епідеміологічного порогу, оскільки так можна істотно знизити ймовірність масового враження хостів у підмережі. Розробляючи стратегію захисту інтернету, треба пам'ятати не лише про захищеність окремо взятих підмереж, а й про вплив взаємозв'язків між ними на стійкість до вірусу системи загалом.

У роботі показано, що за рахунок зменшення взаємозв'язності підмереж вдається збільшити проміжок часу, необхідний для захисту хостів неінфікованої підмережі. Рекомендується, будуючи корпоративні комп'ютерні мережі, встановлювати у точках взаємопід'єднання мереж потужні антивірусні бар'єри – фаєрволи, антивірусні сканери, комплексні системи захисту, що дасть змогу уникнути глобального враження корпоративної мережі у випадку інфікування однієї із підмереж.

1. Возняк Н.О., Щербатий М.В. Поширення вірусів в інформаційних системах // Сучасні проблеми прикладної математики та інформатики: Тези доп. Десятої Всеукраїнської наукової конференції (Львів, 23–25 вересня 2003 р.). – Львів, 2003. – С. 37. 2. Chen Z., Gao L., Kwiat K., Modeling the Spread of Active Worms, Department of Electrical & Computer Engineering, University of Massachusetts, Amherst, MA 01002. 3. Leveille J., Epidemic Spreading in Technological Networks, 2002, HP, Technical Report: HPL-2002-287. 4. Moore D., Shannon C., Brown J., Code-Red: a case study on the spread and victims of an Internet worm, CAIDA, San Diego Supercomputer Center, University of California, San Diego. 5. Moore D., Shannon C., Voelker G. M., Savage S., Internet Quarantine:

Requirements for Containin7g Self-Propagating Code, San Diego Supercomputer Center, University of California, San Diego, INFOCOM 2003. 6. Shcherbatyy M., Ivankiv K., Modelling and Optimization of Infectious Disease Processes Based On Delay Differential Equations. // 22nd Annual Conference, ISCB. – Stockholm, Sweden, August 19-23, 2001. – P. 213. 7. Weissman G., Beubonica: Mitigating Duration and Peak Intensity with a New Model of Computer Virus Epidemics, Dartmouth Undergraduate Journal of Science, Spring 2001 Volume III, No. 2. 8. White S. R., Open Problems in Computer Virus Research // Virus Bulletin Conference, Oct 22, 1998 – Munich, Germany, 1998. 9. Zou C. C., Gong W., Towsley D., Code Red Worm Propagation Modeling and Analysis, Conference on Computer and Communications Security archive Proceedings of the 9th ACM conference on Computer and communications security. – Washington, DC, USA, 2002, pp. 138 – 147. 10. Zou C.C., Towsley D., Gong W., Email Virus Propagation Modeling and Analysis, University of Massachusetts, Amherst, Technical Report: TR-CSE-03-04.

УДК 004.9

В.К. Войчишин, О.С. Климишин, *Є.Я. Лещинський,
***Ю.В. Нікольський, Ю.М. Чорнобай**
Державний природознавчий музей НАН України,
*Національний університет "Львівська політехніка",
кафедра інформаційних систем та мереж

ІНФОРМАЦІЙНИЙ ПОРТАЛ ЛЬВІВСЬКОГО ДЕРЖАВНОГО ПРИРОДОЗНАВЧОГО МУЗЕЮ

© Войчишин В.К., Климишин О.С., Лещинський Є.Я., Нікольський Ю.В., Чорнобай Ю.М., 2005

Викладено принципи реалізації сучасної парадигми діяльності природничо-наукових музеїв, яка об'єднує традиційне колекціонування природних об'єктів як музейних предметів та накопичення інформації про природні об'єкти, процеси і явища у формі електронних баз даних. Запропоновано використати середовище розробки ZOPE для створення такої системи.

The article offers the approaches to realization the modern paradigm of the Natural History Museum activities, which unites traditional collecting of natural objects with accumulation the information in the form of electronic databases. Applying the system ZOPE to develop such system is proposed.

1. Вступ

Сучасні інформаційні технології мають тенденцію до інтеграції в єдине ціле елементів систем обробки інформації, які виконують різні функції. Прикладом цього є розвинутий web-сайт, який є одночасно пошуковою системою з багатокритеріальним пошуком, місцем збереження бази даних з необхідною інформацією, системою інтерактивного спілкування між інформаційним ресурсом та користувачем. Можна вказати інші функції, які виконують сучасні web-ресурси та виділити їхню спільну рису – інтеграцію в одному продукті засобів вирішення різних, але об'єднаних цим ресурсом задач.

У сучасних інформаційних технологіях такі ресурси отримали назву "портал", або "інформаційний портал". Портал на сучасному інформаційному ринку є потужною програмною та інформаційною структурою; створення його як специфічного інформаційного ресурсу є актуальною проблемою. Особливої важливості створення інформаційних порталів набуває у сферах людської діяльності або установах, основна функція яких полягає у накопиченні, збереженні, обробці великих обсягів інформації та забезпеченні доступу до цієї інформації широкій аудиторії. Типовим прикладом такої установи є музей.