УДК 004.49.5

С.Г. СЕМЕНОВ, канд. техн. наук, доц., НТУ "ХПИ", Харьков,
В.В. ДАВЫДОВ, асп., НТУ "ХПИ", Харьков

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ РАСПРОСТРАНЕНИЯ КОМПЬЮТЕРНЫХ ВИРУСОВ В ГЕТЕРОГЕННЫХ КОМПЬЮТЕРНЫХ СЕТЯХ АВТОМАТИЗИРОВАННЫХ СИСТЕМ УПРАВЛЕНИЯ ТЕХНОЛОГИЧЕСКИМ ПРОЦЕССОМ

Усовершенствованы математические модели распространения компьютерных вирусов в гетерогенной компьютерной сети, учитывающие ее топологические и архитектурные особенности. Проведены сравнительные исследования разработанных математических моделей и построены сравнительные графики зависимости количества зараженных узлов от времени функционирования компьютерной системы при распространении эпидемии. Ил.: 4. Библиогр.: 8 назв.

Ключевые слова: математическая модель, компьютерные вирусы, топологические особенности, гетерогенная компьютерная сеть.

Постановка задачи. На современном этапе развития средств и систем управления в различных отраслях производства, одним из основных факторов, влияющих на эффективность их функционирования, является степень защищенности существующих средств автоматизации и компьютеризации в автоматизированных системах управления технологическими процессами (АСУТП). Как показали исследования, существующие компьютерные системы зачастую подвержены различного рода атакам злоумышленного программного обеспечения (ПО). Особым опасностям подвергаются компьютерные системы, используемые для контроля и управления технологическими процессами в системах критического применения.

Исследование систем атак и защиты информации показали высокую степень их взаимозависимости, при этом уровень технического и программного обеспечения злоумышленников в большинстве практических случаев выше. В связи с этим существующие средства защиты АСУТП не всегда своевременно справляются с идентификацией злоумышленного ПО и его локализацией в случае распространения эпидемии. Именно поэтому вопросы разработки и внедрения новых методов, моделей, алгоритмов и систем защиты информации от злоумышленного ПО остается актуальным и в настоящее время. Особо острой в указанном перечне представляется проблематика разработки и внедрения адекватных математических моделей распространения

компьютерных вирусов, учитывающих специфику их злоумышленного воздействия в гетерогенных компьютерных сетях.

Анализ литературы показал, что в работах [1 – 5] ряд авторов уделяют особое внимание биологическим подходам математического моделирования. Так, в работах [3, 6 – 8] были проведены анализ и сравнительные исследования существующих биологических моделей распространения компьютерных вирусов. Это позволило выявить ряд характерных особенностей распространения компьютерных вирусов в компьютерных сетях и получить математическую оценку скорости заражения отдельных объектов в различных режимах функционирования.

Однако анализ существующих математических моделей распространения компьютерных вирусов позволил выявить ряд их недостатков, связанных, в первую очередь, с пренебрежением учета топологических и архитектурных особенностей компьютерной сети. Поэтому разработка математической модели распространения компьютерных вирусов в АСУТП с учетом топологических и архитектурных особенностей компьютерной сети является актуальной научной задачей.

Целью статьи является разработка математической модели распространения компьютерных вирусов в гетерогенных компьютерных сетях с учетом топологических особенностей компьютерной сети.

В работе [8] были выделены следующие математические модели компьютерных вирусов: *SI* (Suspected-Infected), *SIR*(Suspected-Infected-Recovered), *SEIQR* (Suspected-Exposed-Infected-Quarantined-Recovered), *PSIDR* (Progressive Suspected-Infected-Detected-Recovered). Анализ их достоинств и недостатков позволил определить модели *SI*, *SIR*, *PSIDR* как наиболее адекватно описывающие процесс распространения компьютерных вирусов.

1. Модель SIT. Исследования показали, что модель *SI* [1, 2] характеризуется наличием двух типов объектов управления: зараженные (I) и не зараженные (S). Характерной особенностью данной модели является пренебрежение антивирусным ПО (лечащего фактора), что приводит к тому, что эпидемия не может угаснуть.

Обобщенная структура компьютерной системы на основе модели *SI* может быть представлена с помощью выражения:

$$N = S(t) + I(t), \qquad (1)$$

где: N – общее количество объектов в системе; $S(t)$ – количество уязвимых объектов; $I(t)$ – количество зараженных объектов.

В данной математической модели отсутствует учет топологических характеристик. Такой учет возможен путем введения функции связности $f(c_i)$, где $c_i \in C$, C – множество топологий в системе.

В этом случае, с учетом топологических особенностей компьютерной сети, модель *SI* претерпевает ряд изменений и трансформируется в модель *SIT* (*SI-T*opology), а динамическое изменение характеристик данной модели описывается с помощью системы:

$$\begin{cases} \dfrac{dS_i(t)}{dt} = -\dfrac{\beta f(c_i) I(t)}{N} S_i(t), \\ \dfrac{dI(t)}{dt} = \dfrac{\beta f(c_i) I(t)}{N} S(t), \\ \dfrac{dS(t)}{dt} = \sum\limits_{i=1}^{K} \dfrac{dS_i(t)}{dt} / k, \\ \dfrac{dI(t)}{dt} + \dfrac{dS(t)}{dt} = 0, \end{cases} \qquad (2)$$

где: β – частота заражения; c_i – топология конкретной подсети; k – количество топологий в исследуемой сети.

Исследуем временные характеристики поведения компьютерных вирусов в соответствии с моделью *SIT*.

Графики зависимостей изменения количества зараженных узлов от времени функционирования компьютерной системы в условиях распространения эпидемии по моделям *SI* и *SIT* представлены на рис. 1. Предполагается, что в рассматриваемой АСУТП архитектура компьютерной сети представлена определенной, заранее известной, топологией (тип "звезда"), где $f(c_i) = 0.6$.

Анализ графиков на рис. 1 показал, что в соответствии с моделью *SIT* время распространения эпидемии меньше до 2 раз по сравнению с моделью *SI*, при этом скорость заражения в модели *SI* выше в 1.1 раз.

Среди достоинств приведенной модели является простота реализации. Однако, пренебрежение наличием антивирусного ПО ограничивает применение математической модели *SIT* стадией заражения.

2. Модель *SIRT*. Исследования показали, что модель *SIR* [1, 2] характеризуется наличием трех типов объектов управления: зараженные (*I*), не зараженные (*S*), вылеченные объекты, обладающие иммунитетом (*R*).

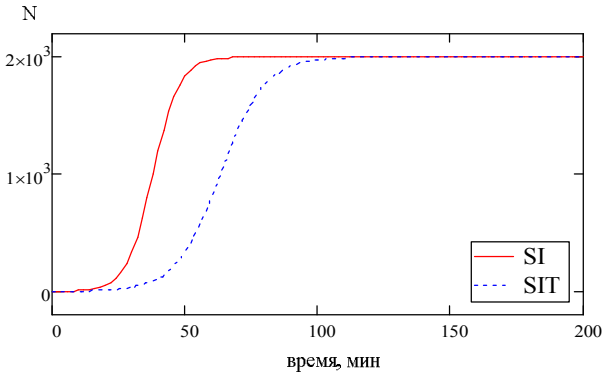

Рис. 1. Графики зависимостей изменения количества зараженных узлов от времени функционирования компьютерной системы в условиях распространения эпидемии по моделям *SI* и *SIT*, при коэффициенте заражения β = 0.2

Обобщенная структура компьютерной системы на основе данной модели может быть представлена с помощью выражения:

$$N = S(t) + I(t) + R(t), \tag{3}$$

где: N – общее количество объектов в системе; $S(t)$ – количество уязвимых объектов; $I(t)$ – количество зараженных объектов; $R(t)$ – количество вылеченных объектов, обладающих иммунитетом.

С учетом топологических особенностей компьютерной сети (фукции связности $f(c_i)$), модель *SIR* претерпевает ряд изменений и трансформируется в модель *SIRT* (*SIR-T*opology), а динамическое изменение характеристик данной модели описывается с помощью системы:

$$\begin{cases} \dfrac{dS_i(t)}{dt} = -\dfrac{\beta f(c_i) I(t)}{N} S_i(t), \\ \dfrac{dI(t)}{dt} = \dfrac{\beta f(c_i) I(t)}{N} S(t) - \delta I(t), \\ \dfrac{dR(t)}{dt} = \delta I(t), \\ \dfrac{dS(t)}{dt} = \sum\limits_{i=1}^{K} \dfrac{dS_i(t)}{dt} / k, \\ \dfrac{dS(t)}{dt} + \dfrac{dI(t)}{dt} + \dfrac{dR(t)}{dt} = 0, \end{cases} \tag{4}$$

где: β – частота заражения; c_i – топология конкретной подсети; δ – частота лечения, "скорость иммунизации"; k – количество топологий в исследуемой сети.

Данная модель подразумевает, что эпидемия возможна лишь при $\beta > \delta$.

Исследуем временные характеристики поведения компьютерных вирусов в соответствии с моделью *SIRT*.

Графики зависимостей изменения количества зараженных узлов от времени функционирования компьютерной системы в условиях распространения эпидемии по моделям *SIR* и *SIRT* представлены на рис. 2. Топологические допущения условий моделирования аналогичны модели *SIT*.

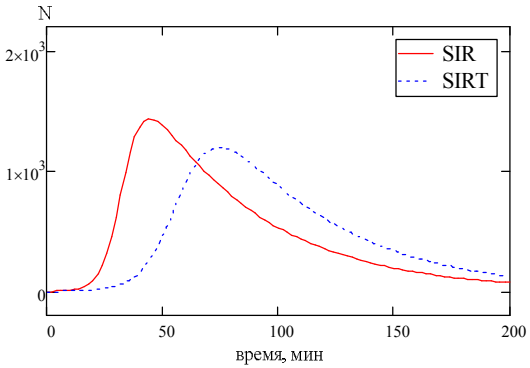

Рис. 2. Графики зависимостей изменения количества зараженных, незараженных и вылеченных узлов от времени функционирования компьютерной системы в условиях распространения эпидемии по моделям *SIR* и *SIRT* при коэффициенте заражения $\beta = 0.25$ и частоте лечения $\delta = 0.02$

Проведенные исследования модели *SIRT* показали, что введение дополнительного типа объекта управления и учет возможного фактора лечения, позволило повысить точность конечного результата в условиях наличия обновляемого антивирусного ПО.

В реальных условиях для лечения компьютерных систем существует необходимость идентификации и локализации злоумышленного ПО. Данная процедура требует определенных (от доли секунды до десятка часов) временных затрат. Данный фактор в модели *SIRT* не учитывается, что также снижает область применения данной модели [8].

3. Модель PSIDRT. Модель *PSIDR* [5] характеризуется наличием четырех типов объектов управления: зараженные (*I*), не зараженные (*S*),

вылеченные объекты, обладающие иммунитетом (R) и найденные зараженные объекты (D).

Эта модель, описывающая поведение системы в условиях воздействия злоумышленного ПО, выделяет 2 этапа:

1) Только заражение объектов (модель идентична модели *SI*);

2) Добавление фактора лечения, при этом вылеченные узлы не заражаются повторно.

Обобщенная структура компьютерной системы на основе модели *PSIDR* может быть представлена с помощью выражения:

$$N = S(t) + I(t) + D(t) + R(t), \tag{5}$$

где: N – общее количество объектов в системе; $S(t)$ – количество уязвимых объектов; $I(t)$ – количество зараженных объектов; $R(t)$ – количество вылеченных объектов, обладающие иммунитетом; $D(t)$ – количество объектов, в которых обнаружен вирус.

С учетом топологических особенностей компьютерной сети (функции связности $f(c_i)$), модель *PSIDR* претерпевает ряд изменений и трансформируется в модель *PSIDRT* (*PSIDR-T*opology), а динамическое изменение характеристик данной модели описывается с помощью системы:

$$\begin{cases} \dfrac{dS_i(t)}{dt} = -\beta S_i(t) I(t) - \mu f(c_i) S_i(t), \\ \dfrac{dI(t)}{dt} = \beta S(t) I(t) - \mu I(t), \\ \dfrac{dR(t)}{dt} = \delta D(t) + \mu f(c_i) S(t), \\ \dfrac{dD(t)}{dt} = \mu I(t) - \delta D(t), \\ \dfrac{dS(t)}{dt} = \sum\limits_{i=1}^{K} \dfrac{dS_i(t)}{dt} / k, \\ \dfrac{dS(t)}{dt} + \dfrac{dI(t)}{dt} + \dfrac{dR(t)}{dt} + \dfrac{dD(t)}{dt} = 0, \end{cases} \tag{6}$$

где: β – частота заражения; $I(t)$ – количество зараженных объектов; μ – вероятность вылечивания (переход в состояние R); c_i – топология конкретной подсети; $S(t)$ – количество уязвимых объектов; $R(t)$ – количество вылеченных (с иммунитетом) объектов; $D(t)$ – количество

обнаруженных зараженных объектов (на первой стадии $D(t) = 0$); k – количество топологий в исследуемой сети; δ – частота лечения.

Исследуем временные характеристики поведения компьютерных вирусов в соответствии с моделью *PSIDRT*.

Графики зависимостей изменения количества обнаруженных зараженных узлов от времени функционирования компьютерной системы в условиях угасания эпидемии (второй этап) по моделям *PSIDR* и *PSIDRT* представлены на рис. 3. Графики зависимостей изменения количества зараженных узлов от времени функционирования компьютерной системы в условиях угасания эпидемии по моделям *PSIDR* и *PSIDRT* представлены на рис. 4. Топологические допущения условий моделирования аналогичны модели *SIT*.

Анализ графиков рис. 3 и рис. 4 показал, что в соответствии с моделью *PSIDRT* происходит замедление процесса обнаружения вируса в 1.65 раз и уменьшение максимального количества обнаруженных объектов в 1.1 раз. Это приводит к замедлению процесса лечения компьютерной системы до 1.01 раз.

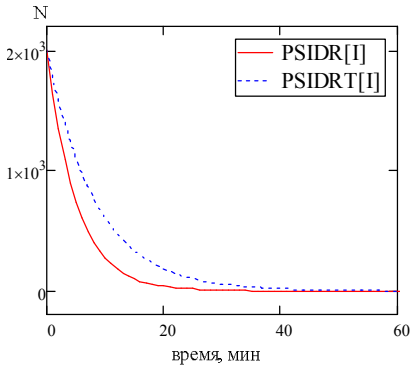

Рис. 3. Графики зависимостей изменения количества обнаруженных зараженных узлов от времени функционирования компьютерной системы в условиях угасания эпидемии (второй этап) по модели *PSIDR* и *PSIDRT*, при коэффициенте заражения $\beta = 0.25$, частоте лечения $\delta = 0.02$, вероятности вылечивания $\mu = 0.2$

Проведенный анализ модели *PSIDRT* показал, что разбиение модели распространения компьютерных угроз на два этапа дает возможность независимого анализа процесса заражения и лечения. Введение задержки между началами этих двух этапов, идентификации, локализация и лечение злоумышленного ПО, позволило устранить один из недостатков модели *SIRT*. Введение коэффициента связности сети не повлияло на время вылечивания системы. Максимальное количество обнаруженных

зараженных объектов уменьшилось в 1.1 раз, что негативно влияет на процесс лечения.

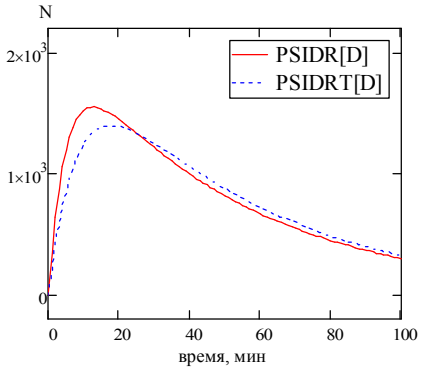

Рис. 4. Графики зависимостей изменения количества зараженных узлов от времени функционирования компьютерной системы в условиях угасания эпидемии (второй этап) по моделям *PSIDR* и *PSIDRT*, при коэффициенте заражения $\beta = 0.25$, частоте лечения $\delta = 0.02$, вероятности вылечивания $\mu = 0.2$

Выводы. В результате проведенной работы были разработаны математические модели распространения компьютерных вирусов в гетерогенных компьютерных сетях с учетом топологических особенностей компьютерной сети. Анализ и сравнительные исследования разработанных математических моделей позволили сделать вывод о повышении их точности по сравнению с известными математическими моделями до двух раз, что позволило сделать вывод о целесообразности применения разработанных моделей в компьютерных сетях АСУТП.

Список литературы: **1.** *Котенко И.В.* Аналитические модели распространения сетевых червей / *И.В. Котенко, В.В. Воронцов* // Труды СПИИРАН. Вып. 4. – СПб.: Наука, 2007. **2.** *Rohloff K.* Stochastic Behavior of Random Constant Scanning Worms / *K. Rohloff , T. Basar* // Computer Communications and Networks, 2005. ICCCN 2005. – Proceedings. 14th International Conference on 17-19 Oct. 2005, – P. 339 – 344. **3.** *Cohen F.* Computer viruses, theory and experiments, Computers & Security. – 1987. – Vol. 6. – P. 22 – 35. **4.** *Jeffrey O.* Directed-Graph Epidemiological Model of Computer Viruses / *O. Jeffrey Kephart, R. Steve White* // IEEE Symposium on Security and Privacy, 1991. – P. 343. ***5.*** *Williamson M.* Epidemiological model of virus spread and cleanup / *M. Williamson, J. Leveille* // HP Laboratories Bristol (February 27th, 2003) [Електроний ресурс]. ***6.*** *Zesheng Chen, Lixin Gao, Kevin Kwait.* Modeling the spread of active worms. INFOCOM 2003. [Електроний ресурс] **7.** *Williamson M.M., Leveille J.* Epidemiological model of virus spread and cleanup. HPL-2003-39 [Електроний ресурс] ***8.*** *Давыдов В.В.* Сравнительный анализ моделей распространения компьютерных вирусов в автоматизированных системах управления технологическим процессом / *В.В. Давыдов* // Системи обробки інформації / Харків: ХУПС, 2012. – Вип. 3 (101). – Том 2. – С. 147 – 151.

Статью представил д.т.н., проф. НТУ "ХПИ" Серков А.А.

УДК 004.49.5

Математична модель розповсюдження комп'ютерних вірусів в гетерогенних мережах автоматизованих систем керування технологічним процесом / Семенов С.Г., Давидов В.В. // НТУ "ХПІ". Серія: Інформатика та моделювання. – Харків: НТУ "ХПІ". – 2012. – № 38. – С. 163 – 171.

Удосконалені математичні моделі розповсюдження комп'ютрених вірусів в гетерогенній комп'терній мережі, що враховує її топологічні та архітектурні особливості. Було проведено порівняльні дослідження розроблених математичних моделей та побудовані порівняльні графіки залежності кількості заражених вузлів від часу функціонування комп'ютерної мережі при розповсюдженні епідемії. Іл.: 4. Бібліогр.: 8 назв.

Ключові слова: математична модель, комп'ютерні віруси, топологічні особливості, гетерогенна комп'ютерна мережа.

UDC 004.49.5

Mathematic model of computer virus spread in heterogeneous computer networks of SCADA systems / Semenov S.G., Davydov V.V. // Herald of the National Technical University "KhPI". Subject issue: Information Science and Modeling. – Kharkov: NTU "KhPI". – 2012. – №. 38. – P. 163 – 171.

Improved mathematical models of computer viruses spread in heterogeneous computer network, taking into account its topological and architectural features. Comparative analyses of mathematical models developed and built by comparative graphs of the number of infected nodes on a computer system operating time during propagation of the epidemic. Figs.: 4. Refs.: 8 titles.

Keywords: mathematical model, computer viruses, topological specifics, heterogeneous computer network.

Поступила в редакцию 18.06.2012