УДК 004.056.57

А. Н. Гусаров, Д. О. Жуков,
А. В. Косарева

ОПИСАНИЕ ДИНАМИКИ РАСПРОСТРАНЕНИЯ КОМПЬЮТЕРНЫХ УГРОЗ В ИНФОРМАЦИОННО-ВЫЧИСЛИТЕЛЬНЫХ СЕТЯХ С ЗАПАЗДЫВАНИЕМ ДЕЙСТВИЯ АНТИВИРУСОВ

Разработана модель динамики развития цепной вирусной эпидемии в компьютерной сети с запаздыванием действия антивируса. Проведен анализ полученной модели, показано ее отличие от существующих аналогов и рассмотрены возможности ее применения.
E-mail: e-mail:nigo@bk.ru

Ключевые слова: *информационно-вычислительные сети, компьютерные угрозы, вирусы, антивирусы.*

Исследование и математическое моделирование динамики компьютерных угроз — сравнительно молодая область информатики. Следует упомянуть, что первые работы, посвященные исследованию динамики распространения компьютерных вирусов, появились еще до создания глобальной сети Интернет.

Однако основной прорыв в развитии исследований динамики вирусных эпидемий произошел после эпидемии таких вирусов, как Code Red I и II, Nimda и Slammer [1–4] и др.

Несмотря на актуальность исследований динамики информационных угроз, в настоящее время существует только две основные модели [5–9], описывающие развитие вирусных эпидемий, берущие свое начало из биологии и не учитывающие многих особенностей компьютерных технологий, например, таких как топология компьютерных сетей.

Анализ сетевых эпидемий последнего времени не позволяет быть уверенным в эффективности применяемых мер защиты [4, 10–12], на что, в частности, указывают результаты работы [4], в которой (на основе анализа наблюдений и моделирования эпидемий) была разработана концепция червя Warhol (мгновенный вирус) и исследованы различные алгоритмы размножения компьютерных вирусов, кардинально повышающие эффективность их распространения [4].

Для результативных мер по обеспечению безопасности необходимо продолжить исследование новых теоретических моделей описания динамики информационных угроз и на их основе провести анализ возможных уязвимостей.

Модели развития вирусных эпидемий в компьютерных сетях. Наиболее часто используемыми моделями, описывающими распространение вирусов, являются так называемые SI- и SIR-модели; первая, SI-модель, описывает развитие эпидемии в сетях без защиты, а вторая, SIR-модель, — с учетом действия антивирусов.

Модель SI [5–9] исходит из того, что любой из входящих в атакуемую сеть компьютеров может находиться в одном из двух состояний: уязвимом (S) и инфицированном (I), причем общее число компьютеров в сети $N = I + S$. Данная модель основана на уравнении

$$\frac{di}{dt} = \beta(1 - i)i, \tag{1}$$

где i — доля зараженных компьютеров; β — константа скорости размножения вирусов.

В SIR-модели факторы, обеспечивающие затухание сетевых эпидемий, оцениваются исходя из того, что сетевые узлы могут находиться в трех состояниях: уязвимом (S), зараженном (I) и невосприимчивом (R). Отметим, что узлы оказываются неуязвимыми только после излечения от инфекции, а общее число узлов сети составляет $N = S + I + R$. Вводя постоянную среднюю скорость иммунизации в единицу времени γ для описания динамики развития эпидемий [5–9], можно получить следующую систему уравнений:

$$\begin{aligned} \frac{di}{dt} &= \beta(1 - r - i) - \gamma; \\ \frac{dr}{dt} &= \gamma(1 - r). \end{aligned} \tag{2}$$

В одной из модификаций SIR-модели кроме этого, рассматривается еще возможность прироста сети с постоянной скоростью за счет появления новых узлов. При распространении эпидемии быстрота заражения сети в значительной степени определяется стратегией поведения вирусов. Если происходит рассылка копий вирусов по случайно выбранным адресам, то скорость заражения сети снижается по сравнению с вариантом, когда адресное пространство сети было первоначально разделено между вирусами.

Распространение червей происходит, как правило, путем поиска уязвимых узлов в сети, затем они ищут и используют уязвимые сервисы, работающие на этих узлах, для проникания и заражения. Например, для обнаружения уязвимого узла некоторые черви используют случайное сканирование адресов, при котором червь случайно генерирует IP-адреса. Затем червь пытается запустить уязвимые сервисы на узлах с этими IP-адресами.

Можно привести несколько основных механизмов сканирования сетей, используемых червями: случайное сканирование адресов; последовательное сканирование адресов, когда червь сканирует IP-адреса последовательно; преимущественное сканирование локальных адресов, когда червь генерирует IP-адреса, которые с высокой вероятностью могут оказаться действующими адресами хост-компьютеров;

метод декомпозиции, когда червь разделяет диапазон IP-адресов после проникания в другой компьютер.

В SI-модели случайное распространение или целенаправленное распространение вирусов учитывается только эмпирическим образом, за счет уменьшения или увеличения константы скорости β их размножения.

Построение математической модели разветвленной эпидемии в сети с запаздыванием действия антивируса. Часто используемая для описания вирусных атак SIR-модель носит эмпирический характер, и многие ее параметры являются искусственными, поэтому для практических целей было бы уместно использовать более строго обоснованную модель. Действия угроз в компьютерной сети и, в частности, распространение вирусов имеют определенные аналогии с цепными процессами, известными в химии и физике. Использование таких подходов для анализа динамики вирусных эпидемий может позволить отойти от принятых в настоящее время биологических аналогий и получить новые результаты, необходимые для моделирования стратегии защиты информационных систем от вирусных угроз.

Для построения такой модели рассмотрим сеть, состоящую из L компьютеров, в которой возможен процесс распространения вирусов, имеющих коэффициент размножения ξ. Процесс распространения вирусов начинается раньше, чем они будут обнаружены и появится эффективный антивирус, способный их необратимо уничтожить. Антивирусы появляются только на некотором шаге процесса распространения вирусов, отстающем от начала на h_0 шагов, т.е. на шаге $k = h - h_0$ (происходит запаздывание).

Число антивирусов, появляющихся для вирусов на $(h+1)$-м шаге или на $(k+1)$-м шаге, обозначим как N_{k+1}, а появившихся на k-м шаге (шаге h для вирусов) — как N_k.

Число зараженных на $(h+1)$-м шаге компьютеров можно обозначить как P_{h+1}, а зараженных на шаге h — как P_h. Изменение числа инфицированных машин равно разности числа заражений и числа вирусов, уничтоженных на $(h+1)$-м шаге.

Имеются следующие случайные события, образующие полную систему:

— компьютер заражен вирусом с вероятностью $\dfrac{P_h}{L}$ (P_h — число вирусов на шаге h, L — число компьютеров в сети);

— на компьютере с вероятностью $\dfrac{N_k}{L}$ имеется антивирус;

— на компьютере с вероятностью $\left(1 - \dfrac{P_h}{L} - \dfrac{N_k}{L}\right)$ нет ни вируса, ни антивируса.

Число заражений на $(h+1)$-м шаге составляет $\xi P_h\Big(1-\dfrac{P_h}{L}-\dfrac{N_k}{L}\Big)$, так как заражение уже зараженного компьютера мы рассматривать не будем, а компьютер, на котором есть антивирус, заразиться не может.

Число вирусов, уничтоженных на $(h+1)$-м шаге, должно составить $P_h\dfrac{N_k}{L}$, где $\dfrac{N_k}{L}$ — вероятность того, что на $(h+1)$-м шаге любой из P_h вирусов, существовавших на шаге h, может встретить антивирус. Таким образом,

$$P_{h+1}-P_h=\xi P_h\Big(1-\frac{P_h}{L}-\frac{N_k}{L}\Big)-P_h\frac{N_k}{L}. \tag{3}$$

Изменение на $(k+1)$-м шаге числа компьютеров, на которых установлен антивирус, определяется разностью $N_{k+1}-N_k$, где N_{k+1} — число машин с антивирусом на $(k+1)$-м шаге, N_k — число машин с антивирусом на шаге k:

$$N_{k+1}-N_k=\xi P_h\Big(1-\frac{N_k}{L}\Big). \tag{4}$$

Здесь ξP_h учитывает, что антивирус устанавливается на $(h+1)$-м шаге на тех машинах, на которых на шаге h был обнаружен вирус, а член $\Big(1-\dfrac{N_k}{L}\Big)$ учитывает, что антивирус устанавливается только там, где его нет.

Поскольку длительность каждого шага равна τ, то все время процесса t и число шагов h связаны между собой следующим соотношением: $t=h\tau$, а $t_0=h_0\tau$ $(k=h-h_0)$.

Переходя от числа шагов h и k к времени процесса, получаем

$$P(t+\tau)-P(t)=$$
$$=\xi P(t)\left\{1-\frac{P(t)}{L}-\frac{N(t-t_0)}{L}\right\}-\frac{P(t)N(t-t_0)}{L}; \tag{5}$$
$$N(t-t_0+\tau)-N(t-t_0)=\xi P(t)\left\{1-\frac{N(t-t_0)}{L}\right\}. \tag{6}$$

Обозначив $t-t_0=y$ и разложив уравнения (5) и (6) в ряд Тейлора, получим

$$P(t)+\tau\frac{dP(t)}{dt}+\frac{\tau^2}{2}\frac{d^2P(t)}{dy^2}+\ldots-P(t)=$$
$$=\xi P(t)\left\{1-\frac{P(t)}{L}-\frac{N(y)}{L}\right\}-\frac{P(t)N(y)}{L}; \tag{7}$$
$$N(y)+\tau\frac{dN(y)}{dy}+\frac{\tau^2}{2}\frac{d^2N(y)}{dy^2}+\ldots-N(y)=\xi P(t)\left\{1-\frac{N(y)}{L}\right\}. \tag{8}$$

Ограничиваясь в левых частях уравнений (7) и (8) не более чем первыми производными, получаем уравнение

$$\tau \frac{dP(t)}{dt} = \xi P(t)\left\{1 - \frac{P(t)}{L} - \frac{N(y)}{L}\right\} - \frac{P(t)N(y)}{L} \tag{9}$$

и

$$\tau \frac{dN(y)}{dy} = \xi P(t)\left\{1 - \frac{N(y)}{L}\right\} \tag{10}$$

с начальными условиями $P(t = 0) = P_0$, $N(y = 0) = P(t_0)$, где $y = t - t_0$ соответственно.

Анализ модели разветвленной эпидемии в сети с запаздыванием действия антивируса. Численное решение системы уравнений (9) и (10) позволяет получить зависимости, моделирующие цепной процесс распространения вирусов, когда в момент времени $t = t_0$ начинает действовать антивирус.

Как отмечалось ранее, результативность распространения вирусов определяется в основном коэффициентом размножения вируса ξ и средней длительностью τ одного шага развития вирусной эпидемии.

На рис. 1 приведены результаты численного решения системы уравнений (9) и (10) для сети, состоящей из 100 000 машин, с начальным числом вирусов $P_0 = 5$ и длительностью $\tau = 25$ условных единиц для различных ξ при условии начала действия антивируса в момент времени $t_0 = 15$ условных единиц времени после начала размножения вирусов.

Кривая *1* показывает изменение с течением времени числа машин, на которых установлен антивирус $N(t)$, а кривая *1′* — число инфицированных компьютеров $P(t)$; обе кривые получены для $\xi = 10$. Кривые

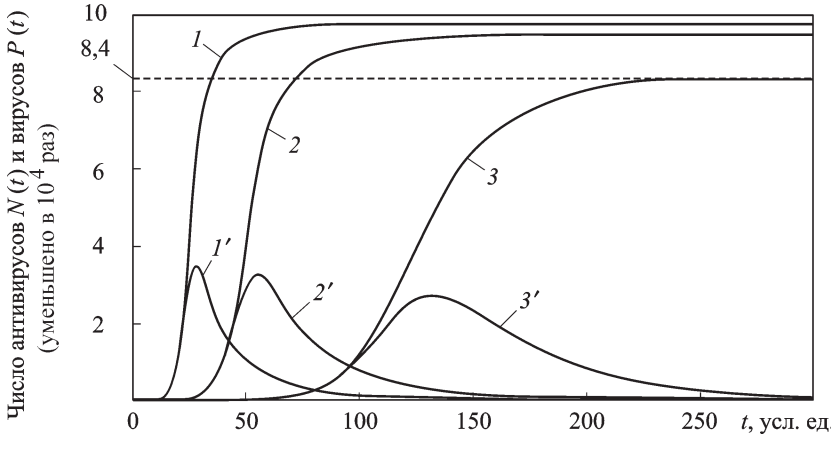

Рис. 1. Результаты численного решения системы уравнений (9) и (10) для сети, состоящей из 100 000 машин ($P_0 = 5$ и $\tau = 25$ условных единиц):
1 и *1′* — изменение $N(t)$ и $P(t)$ для $\xi = 10$; *2* и *2′* — то же для $\xi = 5$; *3* и *3′* — то же для $\xi = 2$

2 и *2′* характеризуют изменение с течением времени числа антивирусов и вирусов в сети, состоящей из 100 000 компьютеров, для случая, когда $\xi = 5$. Кривые *3* и *3′* имеют аналогичный смысл при $\xi = 2$.

Как следует из рис. 1, кривая роста числа антивирусов по своему характеру близка к логистической кривой.

Следует обратить внимание на то, что предельное значение антивирусов не достигает значения 100 000 — числа компьютеров в сети после ликвидации вирусной атаки и на предельные значения $N(t)$ (см. рис. 1, кривая *1*). Величина $N(t)$ достигает значения $8{,}4 \cdot 10^4$, следовательно, в рассматриваемой сети остаются $1{,}6 \cdot 10^4$ компьютеров, на которых нет антивирусного ПО, и может возникнуть новая вспышка эпидемии. Подобная картина часто наблюдается на практике, когда пользователи не устанавливают антивирусное ПО на свои компьютеры после предотвращения эпидемии, если во время эпидемии они не были инфицированы. По всей видимости, это и может служить причиной повторных всплесков активности червей и вирусов, как это наблюдалось в 2001 г. с червем Code Red I, Code Red II и Nimda [1–4].

Состояние сети, состоящей из 100 000 машин, для различных значений τ показано на рис. 2, из которого следует, что увеличение значения τ приводит к общему увеличению времени всего процесса распространения и гибели вирусов.

При прогнозировании вирусных атак и применение комплекса защитных мероприятий особый интерес представляет время запаздывания действия антивирусного ПО. На рис. 3 показано изменение $N(t)$ — кривые *1* и $P(t)$ — кривые *1′* для сети, состоящей из 100 000 машин, с $\xi = 3$, $P_0 = 5$ и $\tau = 25$ условных единиц для $t_0 = 15, 100, 150$ соответственно.

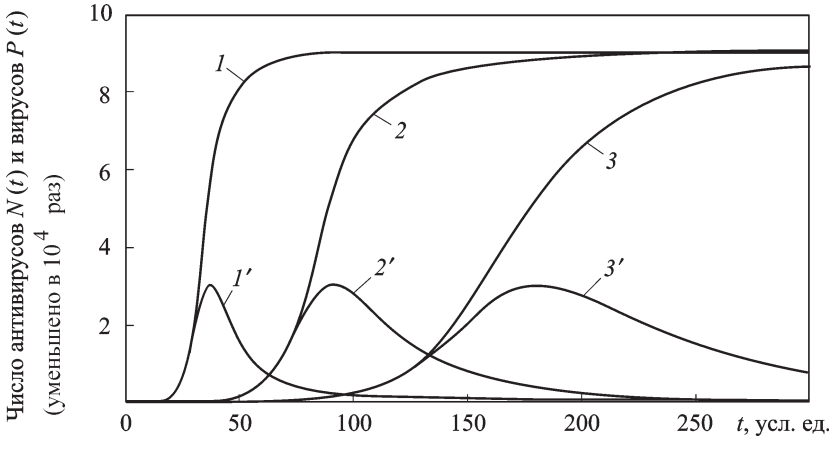

Рис. 2. Результаты численного решения системы уравнений (9) и (10) для сети, состоящей из 100 000 машин, с $P_0 = 5, \xi = 3$ и $t_0 = 15$:
1, 2, 3 и *1′*, *2′*, *3′* — изменение соответственно $N(t)$ и $P(t)$ для $\tau = 10, 25, 50$

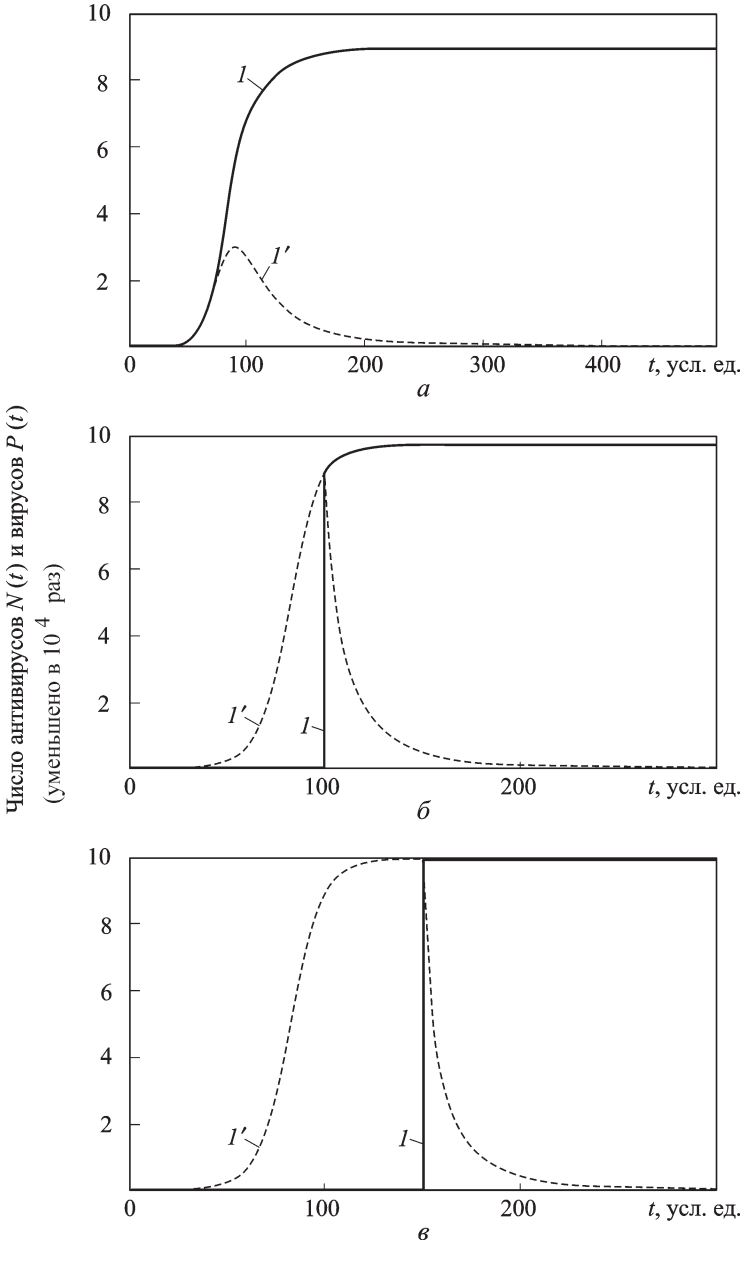

Рис. 3. Результаты численного решения системы уравнений (9) и (10) для сети, состоящей из 100 000 машин, при $P_0 = 5$, $\xi = 3$, $\tau = 25$ условных единиц и $t_0 = 15$ (*а*); 100 (*б*) и 150 (*в*):
1 и *1′* — изменение $N(t)$ и $P(t)$

Приведенные данные показывают вполне очевидный результат моделирования: чем позже начинает действовать антивирусное ПО, тем сильнее оказывается заражена сеть.

Выводы. Рассмотренная в настоящей работе модель цепного развития эпидемии компьютерных вирусов в сети с запаздыванием действия антивирусов в отличие от существующей эмпирической SIR-модели является более обоснованной. Уравнения (9) и (10) образуют систему уравнений, которая существенным образом отличается от системы уравнений (2), используемой в модели SIR. В частности, в уравнении (9) убыль вирусов в правой части определяется произведением числа вирусов на вероятность их встречи с антивирусом, в то время как SIR-модель указывает на убыль вирусов, происходящую с постоянной средней скоростью "иммунизации" в единицу времени γ, т.е. убыль в SIR-модели является постоянной; в предлагаемой модели скорость изменения числа антивирусов связана с числом вирусов, уже существующих в данный момент в сети (уравнение (10)). В SIR-модели принято, что скорость появления антивирусов не зависит от числа имеющихся вирусов. Таким образом, изменение их числа не связано с эпидемией и имеет постоянную скорость.

Кроме того, полученная модель в отличие от SIR-модели позволяет предсказывать результат, при котором в рассматриваемой сети может возникнуть новая вспышка эпидемии, так как часть компьютеров остается незащищенной антивирусом из-за того, что первоначальная эпидемия закончилась раньше, чем антивирусное ПО было установлено на все компьютеры.

СПИСОК ЛИТЕРАТУРЫ

1. Z o u C. C., G o n g W., T o w s l e y D. Code red worm propagation modeling and analysis.// In 9th ACM Symposium on Computer and Communication Security. – Washington DC, USA. – 2002. – P. 138–147.
2. G a u d i n S h. "2003 'Worst year ever' for viruses, worms" (http://www.esecurityplanet.com/trends/article.php/3292461).
3. S t a n i f o r d S., P a x s o n V., W e a v e r N. How to own the Internet in your spare time // 11th Usenix Security Symposium. – San Francisco, USA. – August 2002. – P. 149–167. – ISBN 1-931971-00-5.
4. W e a v e r N. "Warhol worms: The potential for very Fast Internet plagues" (http://www.cs.berkeley.edu/∼nweaver/warhol.html).
5. T h e W o r k s h o p on rapid malcode (WORM). – Washington DC, USA. – ACM PRESS. – October 27, 2003. – ISBN 1-58113-785-0.
6. J a s m i n L e v e i l l e. Epidemic spreading in technological networks (http://www.hpl.hp.com/techreports/2002/HPL-2002-287.pdf).
7. S e n t h i l k u m a r C. G. Worms: how to stop them? (http://www.csif.cs.ucdavis.edu/∼cheetanc/worms/proposal.ps).
8. G a r e t t o M., G o n g W., T o w s l e y D. Modeling malware spreading dynamics" IEEE INFOCOM 2003 (http://www.ieee-infocom.org/2003/papers/46_01.PDF).

9. Захарченко А. А. Бой с тенью: компьютерные вирусы и причины сетевого хаоса // Защита информации. Конфидент. – 2003. – № 6. – С. 49–52.
10. John Leyden. The trouble with anti-virus (http://www.the-register.co.uk/content/56/32680.html).
11. Duncan Graham-Rowe. Computer antivirus strategies in crisis (http://www.newscientist.com/news/news.jsp?id=ns99994119).
12. Zou C. C., Gao L., Gong W., Towsley D. Monitoring and early warning for Internet worms // Proceedings of the 10th ACM Conference on Computer and Communications Security, CCS 2003. – Washington DC, USA. – October 27–30, 2003. – ACM PRESS 2003. – ISBN 1-58113-738-9.

Статья поступила в редакцию 2008

Алексей Николаевич Гусаров родился в 1981 г., окончил Московский государственный университет приборостроения и информатики в 2005 г. Автор 38 научных работ в области информационных технологий.

A.N. Gusarov (b. 1981) graduated from the Moscow State University of Instrument Engineering and Information Technology in 2005. Author of 38 publications in the field of information technologies.

Дмитрий Олегович Жуков родился в 1965 г., окончил МХТИ им. Д.И. Менделеева в 1988 г. и МГУ им. М.В. Ломоносова в 1995 г. Д-р техн. наук, профессор, директор центра новых информационных технологий Московского государственного университета приборостроения и информатики. Автор 122 научных работ в области информационных технологий и повышения качества образования.

D.O. Zhukov (b. 1965) graduated from the Mendeleev Moscow Chemical and Technological Institute in 1988 and the Lomonosov Moscow State University in 1995. D. Sc. (Eng.), professor, director of center of new technologies of the Moscow State University for Instrument Engineering and Information Technology. Author of 122 publications in the field of information technologies and imprivement of education quality.

Анастасия Владимировна Косарева родилась в 1988 г. Студентка факультета информатики Московского государственного университета приборостроения и информатики. Автор 24 научных работ в области нанотехнологии и информационных технологий.

A.V. Kosareva (b. 1988) — student of the Moscow State University for Instrument Engineering and Information Technology. Author of 24 publications in the field of nanotechnology and information technologies.