АНАЛИТИЧЕСКИЕ МОДЕЛИ РАСПРОСТРАНЕНИЯ СЕТЕВЫХ ЧЕРВЕЙ

И. В. Котенко[1], В. В. Воронцов[2]

[1, 2]Санкт-Петербургский институт информатики и автоматизации РАН

[1, 2]СПИИРАН, 14-я линия ВО, д. 39, Санкт-Петербург, 199178

[1]<ivkote@comsec.spb.ru>, [2]<vorontsov@comsec.spb.ru>
[1, 2]<www.comsec.spb.ru>

УДК 004.94

Котенко И. В., Воронцов В. В. **Аналитические модели распространения сетевых червей** // Труды СПИИРАН. Вып. 4. — СПб.: Наука, 2007.

Аннотация. *В статье рассматриваются существующие аналитические модели распространения сетевых червей. Приводится описание как детерминированных моделей (SIS, SIR, SEIR, SAIR и PSIDR), так и стохастических.* — Библ. 11 назв.

UDC 004.94

Kotenko I. V., Vorontsov V. V. **Analytical models of network worm propagation** // SPIIRAS Proceedings. Issue 4. — SPb.: Nauka, 2007.

Abstract. *The paper considers existent analytical models of network worms spread dynamics. The deterministic (SIS, SIR, SEIR, SAIR and PSIDR) as well as stochastic models are described.* — Bibl. 11 items.

1. Введение

Сегодня *атаки сетевых червей* прочно удерживают пальму первенства во всех хит-парадах угроз сетевой безопасности. Этот вид вредоносного программного обеспечения наносит прямой финансовый ущерб, а также служит базисом для реализации многих других опасных угроз, среди которых несанкционированный доступ к данным, кража конфиденциальной и приватной личной информации и т.п.

В рамках данной работы под *сетевым червем* будем понимать программу, способную к самостоятельному поиску новых узлов для заражения и использующую для своего распространения коммуникационную сеть.

Существующие средства защиты не всегда оперативно справляются с эпидемиями сетевых червей, поэтому становится актуальной *задача создания систем обнаружения и защиты нового поколения, способных предотвратить или сдержать эпидемию на ранних стадиях*.

Для решения этой задачи необходимо уметь *моделировать эпидемии сетевых червей* с целью детального исследования этого феномена, анализа факторов, влияющих на распространение сетевых червей, и определения возможных механизмов обнаружения и противодействия. Одним из первых и важных этапов моделирования эпидемии сетевых червей является выбор модели эпидемии, адекватно описывающей процесс распространения сетевых червей.

Мощным подходом для анализа распространения компьютерных червей в масштабах сети Интернет является *аналитическое моделирование*. Преимуществом аналитического моделирования традиционно считается получение решения «*в общем виде*», а также высокая скорость моделирования конкретных сценариев для различных начальных условий. Кроме того, численные результаты, получаемые посредством таких моделей, позволяют анализировать поведение червей в сетях, состоящих из огромного числа элементов.

Таким образом, аналитическое моделирование может использоваться для исследования вирусных эпидемий в сети Интернет.

Сложность математического аппарата, используемого при аналитическом моделировании, связана со сложностью случайных процессов, происходящих при эпидемиях сетевых червей. К сложностям, с которыми приходится сталкиваться при аналитическом описании, относятся периоды бездействия в жизненном цикле червя или его адаптивное поведение.

Аналитический анализ распространения компьютерных червей уходит своими корнями к математическому описанию биологических процессов эпидемий. Математический аппарат для изучения динамики биологических эпидемий был разработан достаточно давно. Различные модели, заимствованные из эпидемиологии, уже применялись для моделирования динамики распространения Internet-червей.

Биологический подход к моделированию вирусной проблемы [1] по общему признанию начался с работ Д.Кефарта (J.Kephart) и С.Уайта (S.White) из IBM [2], однако модным это направление стало только в 2001 г., после вспышек эпидемий Code Red и Nimda [3].

Наибольший резонанс получили работы Н.Уивера (N.Weaver) и его коллег. Ими введена концепция Warhol-червей (или блицкриг-червей), а также исследованы различные концептуальные алгоритмы размножения самовоспроизводящихся кодов, кардинально повышающие эффективность их распространения [4, 5].

В настоящее время в СПИИРАН ведутся работы по созданию перспективных механизмов защиты от сетевых червей. Для решения этой задачи исследуются различные подходы к моделированию распространения сетевых червей. В данной статье описываются исследованные в рамках проводимых работ модели эпидемий. Рассматриваются модели двух классов: детерминированные (SIS-, SIR-, SEIR-, SAIR- и PSIDR-модели) и стохастические.

2. Детерминированные модели эпидемий

Большинство встречающихся в литературе моделей распространения червей представляют собой так называемые *детерминированные модели эпидемий.*

Они пригодны для моделирования эпидемии в той фазе, когда число инфицированных хостов достигло больших значений.

Рассмотрим несколько различных моделей, описывающих динамику распространения сетевых червей:

- SIS-модель (простая эпидемическая модель, или “Susceptible–Infected–Susceptible model”);
- SIR-модель (“Susceptible–Infected–Removed model”);
- SEIR-модель (“Susceptible–Exposed–Infected–Removed model”);
- SAIR-модель (“Susceptible–Antidotal–Infected–Removed model”);
- PSIDR-модель (“Progressive Susceptible–Infected–Detected–Removed model”).

Труды СПИИРАН. 2007. Вып. 4. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2007. Issue 4. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

2.1. Простая эпидемическая модель (SIS-модель)

В этой модели предполагается, что сетевой компьютер (хост) может перейти из неинфицированного состояния в инфицированное состояние [6]. Затем, например, благодаря удалению червя и установке заплаток на программное обеспечение (ПО) хост вновь возвращается в неинфицированное состояние, но впоследствии вновь может быть подвергнут заражению компьютерным червем.

Таким образом, процесс развития инфекции представляется как последовательность переходов между состояниями уязвимости к заражению (Susceptible), инфицированности (Infected) и снова уязвимости к заражению (Susceptible). Поэтому данная модель и обозначается как SIS-модель ("Susceptible-Infected-Susceptible").

Пусть произвольный хост сети, состоящей из постоянного количества N хостов, может находиться в двух состояниях – уязвимом (S) и инфицированном (I), т.е. $S + I = N$ [6].

Предположим, что при инфицировании хоста на нем может существовать только одна копия червя. Эта копия случайным образом выбирает в доступном адресном пространстве потенциальную жертву со средней постоянной скоростью β хостов в секунду. На поиск и заражение одного хоста в среднем тратится $1/\beta$ секунд.

В простейшем случае β определяется средней скоростью сканирования червем сети (V_s) и размером ее адресного пространства (N_{ip}) [6]:

$$\beta = V_S \cdot \frac{N}{N_{ip}} , \qquad (1)$$

где согласно спецификации протокола IP4 $N_{ip} = 2^{32}$.

Для описания динамики доли инфицированных хостов введем переменные $i = I/N$ и $s = S/N$. Уравнение динамики доли инфицированных хостов имеет следующий вид:

$$\frac{di}{dt} = \beta(1-i)i \ . \qquad (2)$$

Уравнение (2) имеет аналитическое решение. Предполагая, что в начальный момент времени $t_0 = 0$ доля инфицированных узлов составляет i_0, получим зависимость [6]:

$$i(t) = \frac{1}{1 + \left[\frac{1}{i_0} - 1\right]\exp(-\beta t)} \ . \qquad (3)$$

Из всего вышесказанного очевидным является то, что эпидемия сетевого червя в принятой модели полностью определяется двумя параметрами: скоро-

Труды СПИИРАН. 2007. Вып. 4. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2007. Issue 4. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

стью размножения червя (β) и начальной зараженностью рассматриваемой сети (i_0).

Функция (3) является логистической функцией **(**функция с S-образной (сигмоидной) формой графика**)** [6]. Она только по форме отличается от функции, представленной в [4], где постоянная интегрирования определена через константу, а не через начальную зараженность.

Динамика функции (3) характеризуется тремя четко различимыми этапами [6]:

- 1-й этап — *сравнительно медленное (но все же экспоненциальное) нарастание зараженности* до порогового уровня $i_{\text{пор}} \approx 0{,}05$. Скорость удвоения доли пораженных хостов равна $\ln(2)/\beta$;
- 2-й этап — *взрывная фаза* в диапазоне $0.05 < i < 0.95$. Ее продолжительность определяется только скоростью поиска β и приблизительно равна $5.89/\beta$;
- 3-й этап — *насыщение*, $i > 0.95$. На этом участке при случайном сканировании доступного адресного пространства инфицированные узлы контактируют преимущественно друг с другом, поэтому неинфицированные узлы могут оставаться «чистыми» неопределенно продолжительное время.

Для достижения порога насыщения $i = 0.95$ требуется время, определяемое следующим образом:

$$t = \frac{1}{\beta}\ln\left[19\left(\frac{1}{i_0} - 1\right)\right].$$

Следовательно, в пределах применимости рассматриваемой модели динамика эпидемии не зависит от масштабов сети. К примеру, сеть из миллиона хостов, в которой в начальный момент времени будет инфицирован один хост, будет практически полностью заражена за то же время, что и аналогичная сеть из ста миллионов хостов, где в начальный момент окажется 100 зараженных хостов.

Из полученных в работе [6] результатов важным для выполнения данного исследования является тот факт, что *существующие системы раннего оповещения о развитии эпидемии окажутся сколько-нибудь эффективными только на участке* $i \ll i_{\text{пор}}$ *и при крайне низких значениях* β.

Согласно данным [7] в случае Code Red это время составило бы не более 50 мин, а для Slammer — не превысило бы 15–20 с. Очевидно, что в обоих случаях эпидемии можно было бы только зафиксировать, но не предотвратить.

2.2. SIR–модель

Для оценки факторов, обеспечивающих затухание сетевых эпидемий, можно использовать SIR-модель (“Susceptible–Infected–Removed model”). В ней хосты существуют в трех состояниях: уязвимом (*S*), зараженном (*I*) и невосприимчивом (*R*). Таким образом, $S + I + R = N$.

Предположим, что узлы оказываются неуязвимыми только после излечения от инфекции.

Труды СПИИРАН. 2007. Вып. 4. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2007. Issue 4. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

Вводя *постоянную среднюю скорость «иммунизации»* в единицу времени γ и учитывая, что $i = \frac{I}{N}$, $r = \frac{R}{N}$ и $s = \frac{S}{N}$, получаем систему уравнений [6]:

$$\begin{cases} \frac{ds}{dt} = -\beta is, \\ \frac{di}{dt} = \beta is - \gamma i, \\ \frac{dr}{dt} = \gamma i. \end{cases} \quad (4)$$

В этой модели вводится понятие *порогового условия* для развития эпидемии. На участке возрастания $i(t)$ производная $\frac{di}{dt}$ должна быть больше 0.

Поскольку $s(t)$ непрерывно уменьшается за счет инфицированных машин, то получаем, что для начала эпидемии необходимо выполнение следующего условия:

$$s(0) > \frac{\gamma}{\beta} \equiv \rho . \quad (5)$$

Выполнение этого условия достигается легко: поскольку γ определяется запаздывающей человеческой реакцией и необходимостью загрузки громоздких «заплат», а β — постоянно улучшающимися техническими характеристиками сети и возможностями злоумышленника.

Например, злоумышленник может вставить небольшую паузу в цикл размножения, чтобы не создавать катастрофического трафика, и слегка понизить скорость инфицирования.

Доля же уязвимых узлов обычно очень велика. Поэтому практически всегда β значительно превосходит γ. Например, при моделировании эпидемии Code Red v2 в [8] соответствие с реальными данными достигнуто при соотношении $\frac{\beta}{\gamma} \approx 10^6$.

В реальных условиях «иммунитет» посредством установки антивирусного программного обеспечения, межсетевых экранов и «заплат» приобретают не только инфицированные узлы (*I*), но и уязвимые (*S*).

Допуская, что средняя скорость иммунизации примерно одинакова для узлов обоих типов и равна (столь же малой) величине γ, получим:

$$\begin{cases} \frac{di}{dt} = \beta i(1 - r - i) - \gamma i, \\ \frac{dr}{dt} = \gamma(1 - r), \end{cases} \quad (6)$$

Труды СПИИРАН. 2007. Вып. 4. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2007. Issue 4. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

и условие развития эпидемии (5) сохраняется.

Тогда

$$r(t) = 1 - \exp(-\gamma t),$$

из этого выражения следует, что при достаточно большом времени любую эпидемию теоретически можно преодолеть. Проблема заключается только в том, что это время может оказаться неприемлемо большим.

Как видно из рис. 1 и рис. 2 [6], при «вакцинации» уязвимых хостов для заметной эпидемической вспышки необходимо, чтобы скорость инфицирования превышала скорость иммунизации на два порядка и более. Это нужно для того, чтобы за время порядка $\frac{1}{\gamma}$ был пройден порог вспышки $i_{\text{пор}}$.

На практике же иммунизация незараженных узлов осуществляется гораздо медленнее. К тому же на часть хостов, на которых червь уничтожается, «заплаты» не устанавливаются, а при расширении Интернет появляются новые уязвимые узлы.

Поэтому повторные эпидемические вспышки могут происходить с завидной регулярностью.

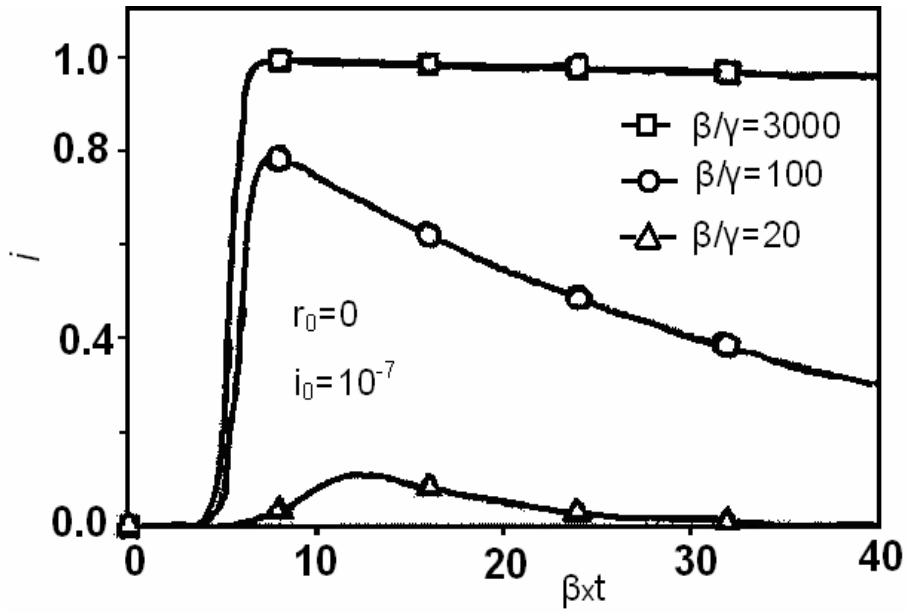

Рис. 1. Динамика развития эпидемии при модели эпидемии (6) и сравнительно малой начальной зараженности сети (10^{-7}). При $\frac{\beta}{\gamma} < 15$ вспышка не происходит [6].

Труды СПИИРАН. 2007. Вып. 4. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2007. Issue 4. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

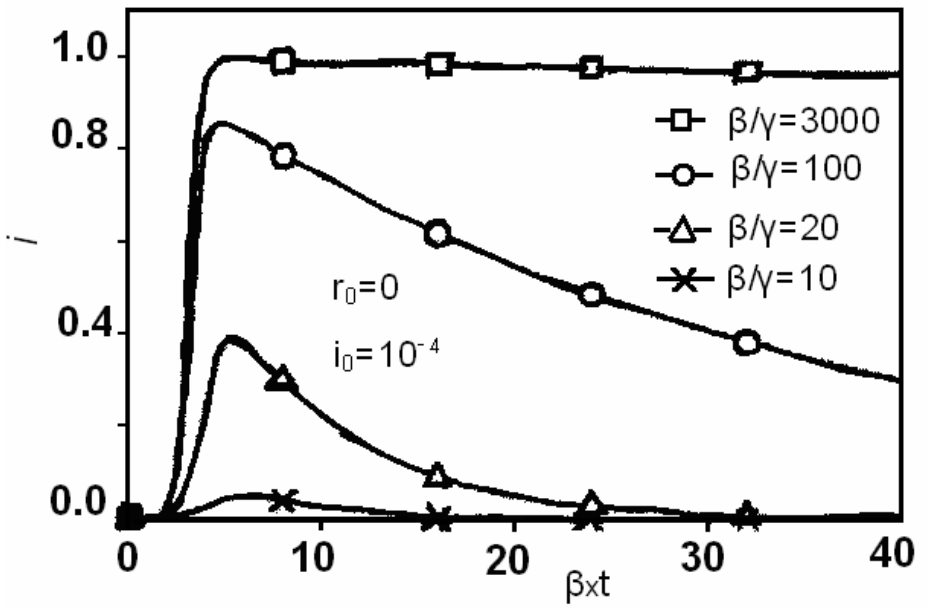

Рис. 2. Динамика развития эпидемии при модели эпидемии (6) и значительной начальной зараженности сети (10^{-4}). В этом случае пороговое значение $\frac{\beta}{\gamma} < 9$ [6].

Динамика системы с переменным числом узлов определяется скоростью прироста новых уязвимых (*S*) узлов α:

$$\begin{cases} \frac{ds}{dt} = -\beta is - (\gamma + \alpha)s + \alpha, \\ \frac{di}{dt} = \beta is - (\gamma + \alpha)i, \\ \frac{dr}{dt} = \gamma(1 - r) - \alpha r. \end{cases} \quad (7)$$

Условие развития эпидемии приобретает вид:

$$s > \frac{\gamma + \alpha}{\beta}. \quad (8)$$

2.3. SEIR-модель

В работе [9] рассматривается SEIR-модель (“Susceptible–Exposed–Infected–Removed model”), являющаяся модификацией SIR–модели.

В этой модели учитывается возможность того, что червь может иметь некий «инкубационный период», во время которого вирус не наносит какого-либо вреда инфицированному узлу.

Обычно червь заражает уязвимый узел (*S*) до входа в свою латентную стадию. В течение латентного периода (*E*, Exposed) узел считается зараженным, но не распространяет вирус.

Через некоторое время он становиться способным к заражению других хостов (I) и далее становиться «излеченным» (*R*).

Труды СПИИРАН. 2007. Вып. 4. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2007. Issue 4. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

В SEIR-модели требуется дополнительно вводить соотношение, характеризующее отношение между состояниями *E* и *I*.

Иногда используется другое обозначение модели — SLIR, где L — латентное состояние (от англ. latency).

2.4. SAIR-модель

В работе [10] авторами предлагается SAIR-модель (“Susceptible–Antidotal–Infected–Removed model”), основанная на SIR-модели.

Представление данной модели отображено на рис. 3.

Модель предусматривает использование на хостах сети антивирусного программного обеспечения.

Все моделируемые хосты в сети делятся на 4 группы:

S — неинфицированные хосты, восприимчивые к заражению;

A — неинфицированные хосты с установленным антивирусом;

I — инфицированные хосты;

R — «обезвреженные» хосты.

В приводимых ниже формулах *S, A, I u R* обозначают количество соответствующих хостов.

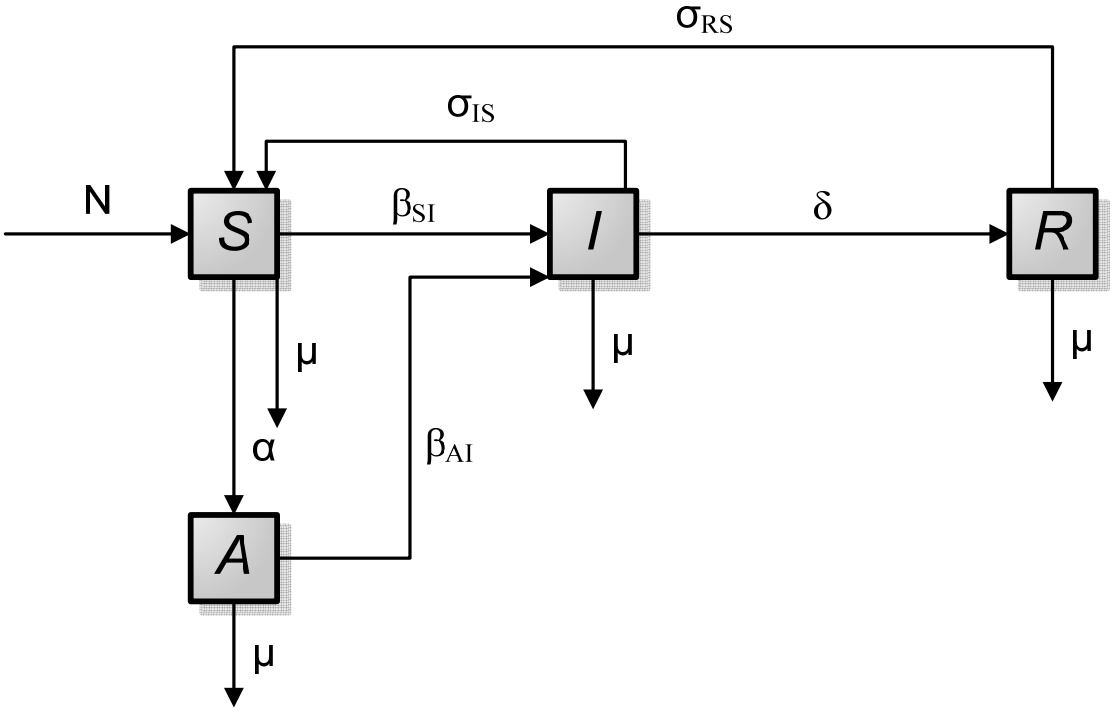

Рис. 3. Представление SAIR-модели.

Динамика модели описывается следующими уравнениями:

$$\frac{dS}{dt} = N - \alpha \cdot S \cdot A - \beta_{SI} \cdot S \cdot I - \mu \cdot S + \sigma_{IS} \cdot I + \sigma_{RS} \cdot R, \quad (9)$$

$$\frac{dI}{dt} = \beta_{SI} \cdot S \cdot I + \beta_{AI} \cdot A \cdot I - \sigma_{IS} \cdot I - \sigma \cdot I - \mu \cdot I, \quad (10)$$

Труды СПИИРАН. 2007. Вып. 4. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2007. Issue 4. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

$$\frac{dR}{dt} = \delta I - \sigma_{RS} R - \mu R, \quad (11)$$

$$\frac{dA}{dt} = \alpha SA - \mu A - \beta_{AI} AI. \quad (12)$$

Параметры, используемые в рассматриваемой модели, описываются следующим образом:

- N — частота добавления новых хостов в сети;
- μ — частота “смерти” хостов не из-за вируса;
- β_{SI} — частота заражения уязвимых хостов;
- $p_{SI} = \frac{I}{T-1}$ — вероятность установления связи с зараженным хостом;
- β_{AI} — частота заражения новым червем хостов с установленным антивирусом;
- $p_{AI} = \frac{I(1-\eta)}{T-1}$ — вероятность установления узлом, на котором установлен антивирус, связи с зараженным узлом;
- δ — частота удаления зараженных хостов;
- $\frac{k_i}{n_i}$ — вероятность выполнения зараженного файла (вероятность заражения хоста);
- n_i — количество исполняемых файлов на i-м хосте в предположении, что все файлы имеют равновероятную вероятность запуска на исполнение;
- k_i^i — количество инфицированных файлов на i-м хосте;
- $k_n{}^i$ — количество неинфицированных файлов на i-м хосте;
- σ_{IS} — частота восстановления зараженных хостов;
- $p_{IS} = \frac{A}{(T-1)\eta}$ — вероятность восстановления зараженного хоста;
- σ_{RS} — частота восстановления удаленных хостов при вмешательстве оператора;
- λ — преобразование уязвимого хоста в неуязвимый путем установки на нем антивирусного программного обеспечения;
- $p_{SA} = \frac{A}{T-1}$ — вероятность установки антивирусного программного обеспечения на инфицированном узле при установлении соединения между уязвимым узлом и узлом, на котором установлено антивирусное программное обеспечение.

Для простоты частоту добавления новых хостов полагают $N = 0$, подразумевая, что во время развития эпидемии не происходит добавления новых узлов. По этой же причине оправдывается и выбор $\mu = 0$. С учетом этих допущений моделируемая система описывается следующими уравнениями:

Труды СПИИРАН. 2007. Вып. 4. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2007. Issue 4. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

$$\frac{dS}{dt} = \alpha SA - \beta_{SI} SI + \sigma_{IS} I + \sigma_{RS} R\,, \tag{13}$$

$$\frac{dI}{dt} = \beta_{SI} SI + \beta_{AI} AI - \sigma_{IS} I - \sigma I\,, \tag{14}$$

$$\frac{dR}{dt} = \delta I - \sigma_{RS} R\,, \tag{15}$$

$$\frac{dA}{dt} = \alpha SA - \beta_{AI} AI\,. \tag{16}$$

Так как $\frac{dS}{dt} + \frac{dA}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0$, то $S + A + I + R = T = const$ для любых t.

Чтобы исследовать динамику модели введем *точки равновесия*.

Ситуация, при которой отсутствует заражение в системе ($I = 0$), характеризуется следующими условиями:

$$P_1^* = (S = 0, A = T, I = 0, R = 0)\,, \tag{17}$$

$$P_2^* = (S = T, A = 0, I = 0, R = 0)\,. \tag{18}$$

Таким образом, все узлы в сети либо уязвимы, либо неуязвимы при $I = 0$.

Выражения для *эндемического равновесия* выглядят следующим образом:

$$P_3^* = (S = \frac{\sigma_{IS} + \delta}{\beta_{SI}}, A = 0, I = \frac{\sigma_{RS} R}{\delta}, R = T - S - I)\,, \tag{19}$$

$$P_4^* = (S = (\beta_{AI})/\alpha, A = -\frac{\beta_{SI}\beta_{AI} I - \sigma_{IS}\alpha - \delta\alpha}{\alpha\beta_{AI}}, I = T - S - A, R = \frac{\delta I}{\sigma_{RS}})\,. \tag{20}$$

Распространение эпидемии определяется устойчивостью точек равновесия: если существует асимптотически устойчивая точка, характерная для нормального существования системы, то эпидемия может быть остановлена, в противном случае она развивается. Наличие таких точек можно получить, вычислив следующий якобиан:

$$J = \begin{bmatrix} -A\alpha - I\beta_{SI} & -\beta_{SI} S + \sigma_{IS} & \sigma_{RS} & -\alpha S \\ I\beta_{SI} & \beta_{SI} S + \beta_{AI} A - \sigma_{IS} - \delta & 0 & \beta_{AI} I \\ 0 & \delta & -\sigma_{RS} & 0 \\ A\alpha & -\beta_{AI} A & 0 & \alpha S - \beta_{AI} I \end{bmatrix}\,. \tag{21}$$

Устойчивость точки P_1^* достигается при

$$T < \frac{\sigma_{IS} + \delta}{\beta_{AI}}\,. \tag{22}$$

Устойчивость точки P_2^* в данной модели недостижима.

Труды СПИИРАН. 2007. Вып. 4. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2007. Issue 4. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

В эпидемиологической литературе известна концепция базовой частоты воспроизводства (R_0).

Этот параметр означает, что при $R_0 > 1$ происходит дальнейшее развитие эпидемии, а при $R_0 < 1$ возможно затухание эпидемии.

В рассматриваемой модели этот показатель может быть получен путем анализа устойчивости точки P_1^*.

Из (22) получаем:

$$R_{01} = \frac{\beta_{AI} T}{\sigma_{IS} + \delta} \ . \tag{23}$$

Таким образом, если $R_0 < 1$, то распространение червя по сети ограничивается. Тогда важным для нормального функционирования сети является следующее условие:

$$\beta_{AI} = \frac{\sigma_{IS} + \delta}{T}. \tag{24}$$

2.5. PSIDR-модель

В работах [9] и [11] предлагается PSIDR-модель распространения (“Progressive Susceptible–Infected–Detected–Removed model”), основанная на SIR-модели.

В PSIDR-модели предполагается, что эпидемические события разделены на *два периода* (см. рис. 4).

1. *Предварительный период (Pre-response).* Изначально червь инфицирует один хост в сети. В течение нескольких дней (часов) червь распространяется по сети, не будучи замеченным большинством пользователей. В терминах PSIDR-модели, эта фаза характеризуется положительной частотой заражения (рождение червя) β без попыток излечения. Уязвимые узлы становятся инфицированными с вероятностью β, если они контактировали с зараженным узлом.
2. *Период отклика (Respons).* Через некоторый период времени червь обнаруживается на некоторых хостах. Осуществляется выделение его сигнатур и внесение их в базы антивирусного программного обеспечения. Неинфицированные узлы становятся невосприимчивыми к данному червю, а инфицированные хосты «вылечиваются» с некоторой частотой, зависящей от скорости обновления антивирусных баз. Этот период в рассматриваемой модели характеризуется той же самой частотой рождения червя, но уязвимые узлы излечиваются с частотой μ, а уже инфицированные узлы обнаруживаются с частотой μ и излечиваются с частотой δ. Параметр μ, по сути, характеризует скорость распространения обновления антивирусных баз после первоначального обнаружения червя.

Труды СПИИРАН. 2007. Вып. 4. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2007. Issue 4. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

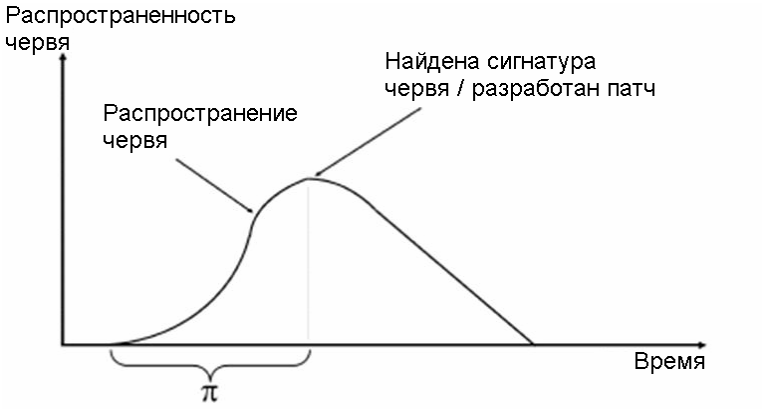

Рис. 4. Разделение эпидемических событий в PSIDR-модели.

В рассматриваемой модели интервал времени, разделяющий эти периоды, обозначается π. Это время характеризует задержку перед появлением готового «лекарства». Другими словами, разделение этих периодов можно наглядно изобразить, как показано на рис. 7. Используются следующие обозначения состояний: S — Susceptible, I — Infected, D — Detected, R — Removed.

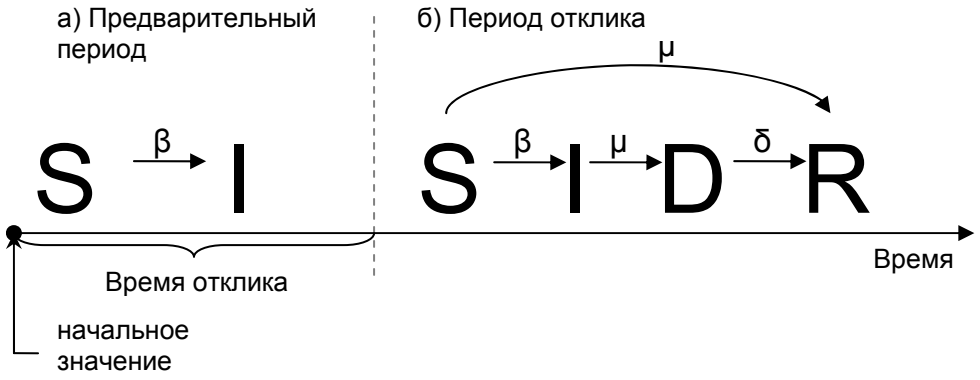

Рис. 5. Разделение эпидемических событий в PSIDR-модели.

Так же как и модели, рассмотренные выше, PSIDR-модель лучше всего описывается через последовательность состояний. Однако стоит рассмотреть следующие важные аспекты, возникающие при использовании для моделирования распространения червей данной модели:

1. *Изменчивость частоты «излечения»*. Подразумевается, что изначально инфицированный узел сразу не «излечивается». «Излечение» возможно только по истечению некоторого периода времени с начала эпидемии (это связано с обнаружением червя и разработкой «лекарства» для него). Поэтому в PSIDR-модели эпидемический период разделен на две части: 1) preresponse и 2) response. В течение первого периода вирус распространяется с частотой β и не излечивается **(**соответствующие параметры обнаружения (μ) и излечения (δ) равны нулю**)**. Далее через некоторый период времени π моделируемая система переходит во второй период, при котором инфицированные узлы могут быть излечены **(**соответствующие параметры обнаружения (μ) и излечения (δ) не равны нулю**)**. В ранее рассмотренных моделях эта ситуация не учитывалась.

Труды СПИИРАН. 2007. Вып. 4. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2007. Issue 4. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

2. *Возможность прямого перехода между состояниями S и R*. После появления «лекарства» уязвимые хосты могут перейти в невосприимчивое состояние, минуя стадию инфицирования в случае своевременного обновления антивирусного программного обеспечения. В рассматриваемой модели это описывается путем прямого перехода из состояния *S* в состояние *R* во время response-периода. Уязвимый хост становится невосприимчивым к атаке с частотой μ. Возможность такого прямого перехода из одного состояния в другое в ранних моделях не предусматривалась.
3. *Состояние обнаружения (detection).* Во время response-периода инфицированный хост (если он все еще функционирует) обнаруживается только при условии обновления антивирусного программного обеспечения. Если хост определен как инфицированный, то пользователь (или технический специалист) изолирует хост от внешней сети и через некоторое время «вылечивает» его. В рассматриваемой модели эта ситуация моделируется добавлением нового состояния, названного *D*-состоянием (от англ. detected), между состояниями *I* и *R*. Во время response-периода инфицированный хост определяется с частотой μ (фактически этот параметр зависит от скорости обновления антивирусного программного обеспечения) и далее «вылечивается» с частотой δ. Состояние *D* представляет собой период, во время которого инфицированный хост изымается пользователем из сети для дальнейшего «излечения». В предыдущих моделях данный фактор не учитывался. Однако стоит отметить, что некое подобие данного состояния присутствует в SEIR-модели (это состояние, когда узел инфицирован, но не является источником инфекции).

Таким образом, PSIDR-модель предполагает, что течение эпидемии можно разбить на два периода: в начале система может находиться в двух состояниях $S \to R$, а по истечении времени $t = \pi$ система переходит состояния $S \to I \to D \to R$ с возможностью прямого перехода между состояниями $S \to R$.

Опишем используемую модель.

Предполагается, что число узлов в сети (N) постоянно.

Предварительный период.

При $t < \pi$ должно выполняться следующее условие:

$$S(t) + I(t) = N. \tag{25}$$

Система описывается следующими дифференциальными уравнениями:

$$\frac{dS}{dt} = -\beta SI, \tag{26}$$

$$\frac{dI}{dt} = \beta SI. \tag{27}$$

Период отклика.

При $t \geq \pi$ должно выполняться следующее условие:

$$S(t) + I(t) + D(t) + R(t) = N. \tag{28}$$

Труды СПИИРАН. 2007. Вып. 4. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2007. Issue 4. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

Система описывается следующими дифференциальными уравнениями:

$$\frac{dS}{dt} = -\beta SI - \mu S, \tag{29}$$

$$\frac{dI}{dt} = \beta SI - \mu I, \tag{30}$$

$$\frac{dD}{dt} = \mu I - \delta D, \tag{31}$$

$$\frac{dR}{dt} = \delta D + \mu S. \tag{32}$$

Можно проверить, что $\frac{dS}{dt} + \frac{dI}{dt} + \frac{dD}{dt} + \frac{dR}{dt} = 0$ удовлетворяет условию (28).

Начальными условиями для системы являются:

$$S(0) > 0, \quad I(0) > 0, \quad D(0) = 0, \quad R(0) = 0.$$

Одним из преимуществ данной модели является наличие очевидных *стоимостных оценок*:

1. *Фиксированная оценка стоимости* (Fixing cost) характеризует время «лечения» хостов, и как много хостов инфицировано (т.е. соотношение «инфицированные хосты / хосты, находящиеся в *D*-состоянии»). Эта характеристика определяется как количество обнаруженных уязвимых хостов за каждый промежуток времени (площадь под кривой определяется через интеграл Римана):

$$\text{Fixing cost} = \int_{\pi}^{T} D(t)dt \approx \sum_{\pi}^{T} D(t). \tag{33}$$

2. *Оценка нанесенного урона* (Disruption cost) показывает, как много хостов инфицировано и как долго они будут находиться в этом состоянии:

$$\text{Disruption cost} = \int_{t_0}^{T} I(t)dt \approx \sum_{t_0}^{T} I(t). \tag{34}$$

3. *Максимальное число инфицированных узлов* позволяет характеризовать наихудшее состояние системы:

$$\text{Maximum number of infected nodes} = \max(I(t))\Big|_{t=t_0}^{T}. \tag{35}$$

4. *Время иммунизации.* В реальных сетях практически не встречается ситуация, при которой все хосты невосприимчивы к воздействию сетевых червей, но большинство из них может стать невосприимчивым к заражению. Поэтому в качестве этого параметра используется время, необходимое для иммунизации 95% узлов (хотя возможны комбинации в диапазоне 90...99%

Труды СПИИРАН. 2007. Вып. 4. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2007. Issue 4. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

в зависимости от потребностей). Таким образом, время, необходимое для приемлемой иммунизации сети, измеряется как функция от параметров конфигурации сети.

3. Стохастические модели эпидемий

Стохастические модели эпидемий являются другим широким классом аналитических подходов в рассматриваемой предметной области.

В качестве примера остановимся более подробно на модели из работы [4]. Авторы моделируют ранние этапы распространения червей, использующих полностью случайное сканирование всего пространства IP-адресов для поиска целей.

Основополагающее допущение, на котором основан проведенный анализ, состоит в том, что вероятность p, найти уязвимый хост за одно сканирование, есть величина неизменная во времени и равная $\frac{V}{2^{32}}$.

Здесь подразумевается равномерное распределение уязвимых хостов в Интернете. Это допущение интуитивно кажется правдоподобным, поскольку рассматривается начальная стадия заражения, когда число инфицированных хостов значительно ниже общего числа уязвимых хостов.

Если ввести ограничение на максимально допустимое число сканирований M произвольных хостов, становится возможным описать процесс эпидемии ветвящимся марковским процессом Гальтона–Ватсона. Действительно, каждый зараженный хост некоторого i-го поколения независимо от других инфицирует некоторое случайное число хостов, имеющих биномиальное распределение с параметрами M и p.

В рамках построенной модели авторы отвечают на два важных вопроса, а именно: закончится ли эпидемия гибелью червя и, если да, то сколько хостов он все-таки успеет поразить.

Формально гибель червя соответствует событию, состоящему в том, что для некоторого поколения число новых зараженных хостов, которые сформировали бы очередное поколение, стало равно 0. В работе показано, что необходимое и достаточное условие наступления такого события с вероятностью 1 есть выполнение неравенства $M \le \frac{1}{p}$.

Таким образом, имея оценку плотности уязвимости, можно ограничить число допустимых сканирований, обеспечив тем самым гарантированное вымирание популяции. На этом наблюдении основана предлагаемая в работе [4] система автоматического сдерживания распространения сетевых червей.

Теперь перейдем к определению оценки для общего числа инфицированных хостов в ходе эпидемии.

Можно показать, что поскольку биномиальное распределение с параметрами (M, p) приближается к распределению Пуассона с параметром $\lambda = Mp$, при выполняющихся на практике условиях, состоящих в том, что M достаточно большое, а вероятность p мала, общее число инфицированных хостов I имеет распределение следующего вида:

Труды СПИИРАН. 2007. Вып. 4. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2007. Issue 4. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

$$P\{I = k\} = \frac{I_0}{k(k - I_0)!}(k\lambda)^{(k-I_0)}e^{-k\lambda}, k = I_0, I_0 + 1, \ldots ,$$

где I_0 — начальное число инфицированных хостов.

Математическое ожидание и дисперсия случайной величины I определяются так:

$$\mathrm{E}[I] = \frac{I_0}{1-\lambda},\ \mathrm{D}[I] = \frac{I_0}{(1-\lambda)^3}\ .$$

Адекватность аналитической модели проверялась посредством имитационного моделирования, результаты которого показали высокую точность разработанного подхода [4].

4. Заключение

На основе анализа существующих работ в статье сделан обзор существующих детерминированных аналитических моделей распространения сетевых червей и дано представление о стохастических моделях. В частности, рассмотрены детерминированные модели SIS, SIR, SEIR, SAIR и PSIDR, а также некоторые результаты моделирования распространения сетевых червей с использованием данных моделей. Модели SIS, SIR и SEIR целесообразно использовать для высокоуровневого моделирования распространения сетевых червей, а модели SAIR и PSIDR — для более детального моделирования, наиболее приближенного к реальным условиям.

Моделирование и сравнение его результатов с данными наблюдений за развитием реальных эпидемий позволяет сделать очевидный и неутешительный вывод о неспособности существующих систем раннего оповещения обеспечить требуемый уровень безопасности.

Решением данной проблемы может стать применение качественно новых подходов к борьбе с распространением сетевых червей. Перспективной видится реализация механизмов сдерживания эпидемий на пограничном сетевом оборудовании с возможностью кооперативного взаимодействия между компонентами системы обнаружения и противодействия. Дальнейшее развитие работы связано с исследованием данного подхода.

Работа выполнена при финансовой поддержке РФФИ (проект №07–01–00547), программы фундаментальных исследований ОИТВС РАН (контракт №3.2/03), Фонда содействия отечественной науке и при частичной финансовой поддержке, осуществляемой в рамках проектов Евросоюза POSITIF (контракт IST–2002–002314) и RE–TRUST (контракт № 021186–2).

Литература

1. *Williamson M. M*. Biologically inspired approaches to computer security [Электронный ресурс] / HP Labs Technical Report HPL-2002-131, HP Labs Bristol, UK. 10 p. // <http://www.hpl.hp.com/techreports/2002/HPL-2002-131.html> (по состоянию на 12.03.2007).
2. *Kephart J. O., White S. R*. Directed-graph epidemiological models of computer viruses // Proceedings of the 1991 IEEE Computer Society Symposium on Research in Security and Privacy. Oakland, California, 1991. P. 343–359.

3. *Zou C. C., Gong W., Towsley D.* Code Red worm propagation modeling and analysis // 9th ACM Symposium on Computer and Communication Security. Washington DC, 2002. P. 138–147.
4. *Staniford S., Paxson V., Weaver N*. How to own the Internet in your spare time [Электронный ресурс] / 11th Usenix Security Symposium, San Francisco, August, 2002. 19 p. // <http://www.icir.org/vern/papers/cdc-usenix-sec02> (по состоянию на 12.03.2007).
5. *Weaver N*. Warhol Worms: The potential for very fast Internet plagues [Электронный ресурс] / 2001. 9 p. // <http://www.iwar.org.uk/comsec/resources/worms/ warhol-worm.htm> (по состоянию на 12.03.2007).
6. *Захарченко А*. Черводинамика: причины и следствия // Защита информации. Конфидент, 2004. № 2, С. 50–55.
7. *Zou C., Gao L., Gong W., Towsley D*. Monitoring and early warning for Internet worms // ACM Conference on Computer and Communications Security(CCS'03). Washington, DC, October 2003. P. 190–199.
8. *Chen Z., Gao L., Kwiat K*. Modeling the spread of active worms [Электронный ресурс] / IEEE INFOCOM 2003. P. 1890–1900. // <http://www.ieee-infocom.org/ 2003/papers/46_03.PDF> (по состоянию на 12.03.2007).
9. *Leveille J*. Epidemic spreading in technological networks [Электронный ресурс] / Technical Report HPL-2002-287, HP Laboratories Bristol, October 2002. 100 p. // <http://www.hpl.hp.com/techreports/2002/HPL-2002-287.pdf> (по состоянию на 12.03.2007).
10. *Piqueira J. R. C., Navarro B. F., Monteiro L. H. A*. Epidemiological models applied to viruses in computer networks // Journal of Computer Science. 2005. Vol. 1, No. 1. P. 31–34.
11. *Williamson M. M., Leveille J*. An epidemiological model of virus spread and cleanup [Электронный ресурс] / Proceeding of Virus Bulletin Conference, Toronto, Canada. 11 p. // <http://www.hpl.hp.com/techreports/2003/HPL-2003-39.html> (по состоянию на 12.03.2007).