УДК 004.056.57

Д.В. Бабанин

МОДЕЛИ РАСПРОСТРАНЕНИЯ КОМПЬЮТЕРНЫХ ВИРУСОВ НА ОСНОВЕ ЦЕПЕЙ МАРКОВА

Московский государственный институт электроники и математики (технический университет), г. Москва

Две модели распространения компьютерных вирусов, основанных на цепях Маркова и описание их использования

Компьютерные вирусы на сегодняшний день являются постоянной угрозой, представляющей опасность как для отдельных пользователей ПК, так и для предприятий. Актуальной является задача проектирования локальных сетей таким образом, чтобы сеть была максимально защищена от вирусов своей структурой. Для точной оценки защищенности сети от вирусов необходимо иметь возможность смоделировать развитие вирусной эпидемии на выбранной конфигурации сети.

Существующие модели распространения вирусов в компьютерных сетях (SI, SIR[1], AAWP[2], PSIDR[3]) не учитывают структуру сети, основываясь на предположении, что сеть является полносвязным графом.

Рассмотрим локальную сеть, состоящую из N компьютеров. Каждый компьютер может находиться в одном из двух состояний – незараженный или зараженный.

Сеть можно представить в виде графа, узлами которого являются компьютеры, а дугами – каналы связи между ними, по которым могут распространяться вирусы. Вес связи w_{ij} означает вероятность перехода вируса по каналу связи между компьютерами i и j за единицу времени.

Модель на основе цепи Маркова для всей сети

Общее состояние сети является совокупностью состояний всех компьютеров сети, которое можно описать вектором из N элементов, где значение i-го элемента соответствует состоянию i-го компьютера: I (infected), если компьютер заражен, и S (suspected), если компьютер не заражен.

Состояние сети в следующий момент времени зависит только от текущего состояния сети, и не зависит от предыдущих. Поэтому процесс распространения вируса в сети можно представить как цепь Маркова.

Построение матрицы переходов

Переходные вероятности вычисляются по формуле

$$P_{ij} = P[f^t = s^j \mid f^{t-1} = s^i] \qquad (1)$$

Сеть перейдет из состояния s_i в состояние s_j при условии, если каждый компьютер в сети перейдет из состояния s_k^i в состояние s_k^j, где k – номер компьютера в сети. Вероятность этого события описывается следующей формулой:

$$P_{ij} = P[f^t = s^j \mid f^{t-1} = s^i]$$
$$= P[f_1^t = s_1^j \cap f_2^t = s_2^j \cap \ldots \cap f_N^t = s_N^j \mid$$
$$f_1^{t-1} = s_1^i \cap f_2^{t-1} = s_2^i \cap \ldots \cap f_N^{t-1} = s_N^i] \quad (2)$$
$$= \prod_{k=1}^{N} P[f_k^t = s_k^j \mid f_k^{t-1} = s_k^i].$$

Определение вероятности изменения состояния компьютера

Вероятность $P[f_k^t = s_k^j \mid f_k^{t-1} = s_k^i]$ перехода k-го компьютера из состояния s_k^i в состояние s_k^j можно вычислить следующим образом. Необходимо рассмотреть четыре варианта для различных состояний компьютера на предыдущем и следующем шаге: переход из состояния S в состояние I, из S в S, I в S, и I в I.

S → I. Пусть $P_{зар}(k, s^i)$ – вероятность заражения незараженного k-го компьютера из состояния сети s^i.

S → S. Поскольку событие перехода компьютера в зараженное состояние и событие, при котором незараженный компьютер останется незараженным, образуют полную группу событий, то вероятность $P[f_k^t = S \mid f_k^{t-1} = S]$ будет равна $1 - P_{зар}(k, s^i)$.

I → S. Так как модель не учитывает излечения компьютера от вирусов, то переход из состояние 1 в состояние 0 невозможен, т.е. имеет нулевую вероятность.

I → I. Так как модель не учитывает излечения компьютера от вирусов, то вероятность перехода из состояния 1 в состояние 1 равна единице.

В результате получается следующая формула:

$$P[f_k^t = s_k^j \mid f_k^{t-1} = s_k^i] = \begin{cases} P_{зар}(k, s^i), & если\ s_k^i = S, s_k^j = I \\ 1 - P_{зар}(k, s^i), & если\ s_k^i = S, s_k^j = S \\ 0, & если\ s_k^i = I, s_k^j = S \\ 1, & если\ s_k^i = I, s_k^j = I \end{cases} \quad (3)$$

Определение вероятности заражения компьютера

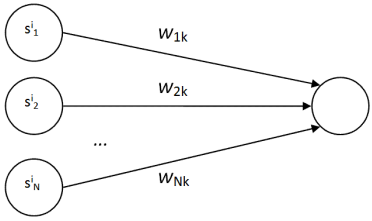

Рис. 1. Схема заражения узла

Узел k перейдет из незараженного состояния в зараженное за единицу времени в случае, если вирус к нему поступит хотя бы с одного другого узла.

Поскольку события заражения k-го компьютера от различных узлов являются независимыми, то вероятность заражения незараженного k-го узла будет равна:

$$P_{зар}(k, s^i) = 1 - \prod_{m-1}^{N} (1 - P_{передачи}(m, k, s_m^i)) \quad (4)$$

Вероятность передачи вируса от узла m узлу k при состоянии сети s_m^i можно вычислить следующим образом:

- если компьютер m заражен, вероятность равна w_{mk}, (как это было определено в начале статьи),
- если компьютер m не заражен, то вероятность передачи вируса с него равна нулю.

$$P_{передачи}(m, k, s_m^i) = \begin{cases} w_{mk}, \text{если } s_m^i = \text{I} \\ 0, \text{если } s_m^i = \text{S} \end{cases} \quad (5)$$

Модель на основе цепи Маркова для отдельных узлов

Если рассматривать в виде марковской цепи не процесс распространения вируса по всей сети, а строить отдельную марковскую цепь для каждого узла, можно значительно сократить объем вычислений.

В каждый момент времени каждый компьютер с определенной вероятностью может быть незараженным (S), либо зараженным (I). Вектор состояния в данном случае состоит из двух элементов – вероятности того, что компьютер не заражен и вероятности того, что компьютер заражен: $p_k^t = \{P_k^t(S), P_k^t(I)\}$, где k – номер компьютера, t – номер шага.

Матрица переходных вероятностей для данного узла будет иметь следующий вид:

$$P = \begin{pmatrix} P(f^t = S \mid f^{t-1} = S) & P(f^t = I \mid f^{t-1} = S) \\ P(f^t = S \mid f^{t-1} = I) & P(f^t = I \mid f^{t-1} = I) \end{pmatrix} \quad (6)$$

Поскольку данная модель не учитывает возможность излечения, то переход из состояния I в состояние S невозможен, а из состояния I возможно попасть только обратно в состояние I, то

$$P[f^t = S \mid f^{t-1} = I] = 0$$
$$P[f^t = I \mid f^{t-1} = I] = 1$$

Так как сумма элементов строки матрицы переходов всегда равна единице, то

$$P[f^t = S \mid f^{t-1} = S] = 1 - P[f^t = I \mid f^{t-1} = S]$$

Обозначим $P_{зар}(k) = P[f^t = I \mid f^{t-1} = S]$, где k – это номер узла, для которого составляется модель. В результате матрица переходов выглядит следующим образом:

$$\mathbf{P} = \begin{pmatrix} 1 - P_{зар}(k) & P_{зар}(k) \\ 0 & 1 \end{pmatrix} \quad (7)$$

где $P_{зар}(k)$ - это вероятность заражения k-го узла, которая, как было показано выше, вычисляется по формуле

$$P_{зар}(k) = 1 - \prod_{m=1}^{N} (1 - P_{передачи}(m,k)) \quad (8)$$

Передача вируса от узла m узлу k произойдет при одновременном наступлении следующих событий

- если компьютер m заражен на предыдущем шаге, вероятность этого события равна $P_k^{t-1}(I)$;
- если вирус пройдет по связи m-k, вероятность этого события равна w_{mk} (как это было определено в начале статьи).

Следовательно, вероятность передачи вируса от узла m узлу k равна произведению вероятности зараженности компьютера m на предыдущем шаге, умноженной на вероятность перехода вируса по связи m-k:

$$P_{передачи}(m,k) = P_m^{t-1}(I) \cdot w_{mk} \quad (9)$$

В результате подстановки получаем следующую матрицу переходов для отдельного узла:

$$\mathbf{P} = \begin{pmatrix} \prod_{m=1}^{N} (1 - P_m^{t-1}(I) \cdot w_{mk}) & 1 - \prod_{m=1}^{N} (1 - P_m^{t-1}(I) \cdot w_{mk}) \\ 0 & 1 \end{pmatrix} \quad (10)$$

Как можно заметить, матрица переходов, зависит от состояния узлов сети на предыдущем шаге, следовательно цепь Маркова для каждого узла является неоднородной.

Использование модели на основе цепи Маркова для всей сети

Для использования модели на основе цепи Маркова необходимо задать начальное распределение π_0 – вектор вероятностей нахождения сети в том или ином состоянии в начальный момент времени. Выбирается единственное начальное состояние сети f_0, для которого вероятность принимается равной единице, для остальных – нулю: $\pi_0 = \{p_j^0\}$, где $p_j^0 = P[f_0 = s_j] = \begin{cases} 1, f_0 = s_j \\ 0, f_0 \neq s_j \end{cases}$

Исходя из теории марковских цепей, распределение на шаге t будет равно $\pi_t = \pi_{t-1}\mathbf{P}$.

Математическое ожидание среднего числа зараженных компьютеров на шаге n можно вычислить следующим образом:

Для каждого состояния сети s_j легко определить количество зараженных компьютеров: поскольку s_j представляет собой вектор, состоящий из N элементов, то количество зараженных компьютеров для состояния s_j определяется как

$$N_I(s_j) = \sum_{i=1}^{N} \begin{cases} 1, s_{ij} = I \\ 0, s_{ij} \neq I \end{cases}$$

Поскольку сумма всех элементов вектора состояния на шаге t всегда равно единице, то математическое ожидание количества зараженных компьютеров будет равно

$$M[N_I^t] = \sum_{j=1}^{2^N} p_j^{(t)} \cdot N_I(s_j)$$

Проводя вычисления для t = 0...t_{max}, получим зависимость среднего количества зараженных компьютеров от номера шага t.

Использования модели марковских цепей для отдельных узлов

Использование модели для отдельных узлов значительно проще.

Начальное распределение задается для каждого узла сети: $\pi_j^0 = \{P_j^0(S), P_j^0(I)\}$, где j – номер узла. Удобнее всего выбрать в сети зараженные компьютеры, для которых вероятность нахождения в зараженном состоянии равна 1, т.е. $\pi_j^0 = \{0; 1\}$, а для остальных – наоборот $\pi_j^0 = \{1; 0\}$.

Далее для каждого шага t производятся следующие действия:

- Для каждого j-го узла по формуле (10) строится матрица переходов.
- Вектор начального на предыдущем шаге умножается на полученную матрицу переходов, в результате получается вектор распределения для первого шага: $\pi_j^t = \pi_j^{t-1}\mathbf{P}$.
- Имея множество векторов распределения на шаге t равно $\{\pi_1^t, \pi_2^t, ..., \pi_N^t\}$, а следовательно, зная вероятность нахождения каждого узла в зараженном состоянии в момент времени t ($P_j^t(I)$), можно вычислить математическое ожидание количества зараженных компьютеров на этом шаге:

$$M[N_I^t] = \sum_{j=1}^{N} P_j^t(I)$$

Проводя вычисления для t = 0...t_{max}, получим зависимость среднего количества зараженных компьютеров от номера шага t.

Заключение

Были предложены модели, позволяющие рассчитывать распространение компьютерных вирусов в вычислительных сетях различной топологии. Был описан механизм их использования для получения информации о характере распространения вирусной эпидемии в сети. Использование данных моделей позволит оценить защищенность сетей различных топологий от вирусных атак.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. А.А. Захарченко. «Черводинамика: причины и следствия». Защита информации. Конфидент, 2004, № 2, с. 50–55.

2. Zesheng Chen, Lixin Gao, and Kevin Kwait. Modeling the spread of active worms. INFOCOM 2003. [Электронный ресурс] http://www.ieee-infocom.org/2003/papers/46_03.PDF

3. M. M. Williamson, J. Leveille "An epidemiological model of virus spread and cleanup" HPL-2003-39 [Электронный ресурс] http://www.hpl.hp.com/techreports/2003/HPL-2003-39.pdf