ИНФОРМАТИКА
ИНФОРМАЦИОННЫЕ СИСТЕМЫ

УДК 519.684

Математическая модель поиска коэффициента памяти компьютера, инфицированного вирусом, и собственных параметров компьютерных вирусов в аспекте теории вычислителей с неабсолютной памятью

В. В. Бахтин
Пермский государственный национальный исследовательский университет
Россия, 614990, г. Пермь, ул. Букирева, 15
bakhtin_94@bk.ru; 89125851599

Предложена математическая модель поиска коэффициента памяти компьютера, зараженного вирусом, и введены определения константы вируса, относительной константы вируса и коэффициента уничтожения вируса. Показан алгоритм расчета параметров зараженного компьютера и компьютерного вируса, что позволяет провести классификацию компьютерных вирусов в соответствии с их коэффициентами уничтожения и константами вирусов. В качестве примера проведен расчет данных параметров для внедряемого COM-вируса.

Ключевые слова: *робот; компьютерный вирус; классификация вирусов; константа вируса; относительная константа вируса; коэффициент уничтожения вируса; моделирование памяти.*

DOI: 10.17072/1993-0550-2017-4-79-85

Введение

В настоящее время программное обеспечение роботов создается непосредственно для каждой серии роботов, однако, существует возможность использования компьютерных вирусов для создания и моделирования роботов с абсолютной и неабсолютной памятью. Для этого необходимо заражать операционную систему робота соответствующим вирусом, с помощью которого будет моделироваться некоторый коэффициент памяти λ для данного робота (или компьютера). Если этот коэффициент удовлетворяет двойному неравенству $0<\lambda<1$, то робот будет забывать часть полученной информации, ее будет уничтожать вирус, за счет этого эффекта забывания компьютер или робот, зараженный вирусом, будет неким психологическим аналогом человека.

В предлагаемом способе создания психологических аналогов человека необходимо знать коэффициент памяти λ для конкретного вируса, поэтому необходимо разработать математическую модель функционирования вируса, с помощью которой можно вычислить значение этого коэффициента.

Рассмотрим базовые положения теории роботов с неабсолютной памятью.

1. Базовые определения

Элементарным обучением s_i будем называть количество информации, полученное вычислителем из внешних источников, массив информации, который обрабатывает робот за один такт.

В работах [1, 2] приведено соотношение, позволяющее вычислять количество информации робота или иного вычислителя, получаемое им в результате непрерывного получения роботом информации в течение определенного количества тактов:

$$S_i = s_i + \lambda S_{i-1},$$

где – i – порядковый номер массива информации, воздействующего на робота и порождающего у него элементарное обучение s_i, S_i – суммарное обучение робота.

Суммарным обучением робота назовем обучение, полученное роботом в результате воздействия на него каждого из массивов информации, обработанных на предыдущих i-тактах, λ – коэффициент памяти, характеризующий долю предыдущего суммарного обучения, которую помнит робот к моменту воздействия на него массива информации с порядковым номером i, $\lambda \in [0,1]$.

Тактом заражения назовем одну итерацию работы компьютерного вируса, за которую компьютерный вирус успевает заразить один файл. В контексте теории роботов с неабсолютной памятью, тактом будет называться воздействие на робота нового массива информации, который создает у компьютера элементарное обучение s_i.

Предположим, что $s_i = s_i^{vir} + s_i^{us}$, где s_i^{vir} – информация, которую компьютер получил от вируса, s_i^{us} – информация, полученная компьютером от пользователя. Предположим, что информацию компьютер получает только от вируса, т. е. s_i^{us}=0. То есть, формула расчета суммарного обучения принимает следующий вид:

$$S_i = s_i^{vir} + \lambda S_{i-1}, \tag{1}$$

где, $s_i = \underline{s^{vir}} = const$ для каждого конкретного неполиморфного вируса, то есть вируса, который будет записывать один и тот же текст во все заражаемые файлы.

Введем определение константы вируса $\underline{s^{vir}}$ – константой вируса называется количество байт информации, на которое увеличится объем зараженного файла за один такт заражения. Для расчета константы вируса будем использовать следующую формулу:

$$\underline{s^{vir}} = S_1^f - S_0^f .$$

Вместе с определением константы вируса стоит определить понятие относительной константы вируса – относительной константой вируса назовем отношение константы вируса $\underline{s^{vir}}$ к первоначальному объему незараженного файла S_0^f. То есть расчет относительной константы вируса будем производить по формуле

$$\varphi_1^{vir} = \frac{\underline{s^{vir}}}{S_0^{f1}} .$$

Значение относительной константы вируса зависит от размера файла до заражения, поэтому расчет данной величины будет производиться для каждого файла в отдельности.

Чтобы получить некоторый обобщенный показатель, введем понятие средней относительной константы вируса, которая будет рассчитываться как отношение суммы константы вируса для n файлов и суммы изначальных объемов этих файлов. Ниже приведена соответствующая формула:

$$\overline{\varphi^{vir}} = \frac{n \cdot \underline{s^{vir}}}{\sum_{i=1}^{n} S_0^{fi}} .$$

Коэффициент уничтожения – коэффициентом уничтожения компьютерного вируса назовем величину, равную разности единицы и коэффициента памяти робота (или компьютера) λ, зараженного данным вирусом. Рассчитываться он будет по следующей формуле:

$$\omega = 1 - \lambda .$$

2. Внедряемый вирус

Компьютерный вирус – это вредоносная программа, способная создавать свои копии и (или) другие вредоносные программы [3]. А вредоносная программа, в свою очередь, это программа, используемая для осуществления несанкционированного доступа к информации и (или) воздействия на информацию или ресурсы автоматизированной информационной

системы [3]. Из этих определений становится понятно, что главным отличительным свойством компьютерного вируса является саморепликация, то есть создание собственных копий в других исполняемых файлах или файлах данных. Отметим, что именно это мы называем заражением файла компьютерным вирусом.

Одним тактом будем называть заражение нового файла исследуемым вирусом. Опишем то, как происходит процесс заражения очередного файла. Примером, на котором мы будем это рассматривать, послужит внедряемый в конец файла *.COM вирус. COM-файл – исполняемый файл, содержащий инструкции на языке ассемблер. Сокращение от английского command – командный файл, содержащий команды для исполнения. Внедрение в другой исполняемый файл происходит по следующему алгоритму:

1. Вирус находит непросмотренный COM-файл (другой исполняемый файл) в той же директории, где непосредственно находится, если файл найден, то переходит к пункту 2. Если после просмотра всех файлов в директории не найден пригодный для заражения, то вирус останавливает свою работу, заражения не произойдет.

2. Вирус проверяет возможность чтения файла, если ошибок чтения не происходит, то переходит к пункту 3 (иначе – возвращается к пункту 1).

3. Осуществляется проверка длины файла, если она меньше 64 000 байт, то переход на пункт 4, иначе – пункт 1.

4. Вирус переходит на последний байт файла и читает его, то есть считывает его потенциальную сигнатуру – фрагмент, который присутствует в его собственном коде, программную подпись, которая оставляется вирусом для идентификации уже зараженных файлов.

5. Проверяется зараженность данным вирусом (сигнатурная проверка), если считанный байт является сигнатурой, то переход к пункту 1, в противном случае данный файл еще не заражен и следует переходить к следующему пункту алгоритма.

6. Вирус приступает к заражению; в первые три байта заражаемого файла записывает команду перехода на тело вируса, т. е. на первый байт после окончания тела самого заражаемого файла. То есть на первые три байта записывается команда jmp "метка" (таким образом, при запуске этого исполняемого файла первой же командой будет осуществляться передача управления вирусу, который начнет свою деятельность с этого момента).

7. Тело вируса записывается в конец заражаемого файла.

8. Записать в конец тела вируса, которое только что было записано в конец заражаемого файла, инструкцию jmp "метка" (данная метка указывает на начало тела заражаемой программы). Выполнение данной инструкции передает управление обратно зараженному файлу, после этого он продолжит работать в штатном режиме, выполнит все необходимые операции и благополучно завершится, возвращая код 0.

3. Математическая модель

Для построения модели выберем следующую целевую функцию[4]:

$$\min_{\lambda} Y(\lambda) = \sum_{i=1}^{n} \left(S_i^э - S_i^m \right)^2,$$

где $S_i^э$ – полученное экспериментально значение количества информации, которая находится в исследуемой директории к концу такта с номером i, S_i^m – количество информации в исследуемой директории к концу такта с номером i, рассчитанное теоретически.

В данном случае в качестве минимизируемой функции была использована функция, построенная с помощью метода наименьших квадратов (МНК). Минимизация целевой функции позволит вывести формулу для расчета значений количества информации на определенном такте заражения для конкретного вируса. Такая формула требуется для построения математической модели функционирования робота (компьютера), зараженного определенным вирусом.

Используем формулу (1) для расчета S_i^m, подставим ее в целевую функцию и получим следующее соотношение:

$$\min_{\lambda} Y(\lambda) = \sum_{i=1}^{n} \left(S_i^э - s_i^{vir} - \lambda S_{i-1}^э \right)^2.$$

Возведем правую часть выражения в квадрат и сгруппируем по слагаемым с λ, сохраняя знаки суммы у соответствующих слагаемых.

Получим следующую формулу:

$$Y(\lambda)=\sum_{i=1}^{n}S_{i-1}^{э\,2}\lambda^2+\sum_{i=1}^{n}\left(2s_i^{vir}S_{i-1}^{э}-2S_i^{э}S_{i-1}^{э}\right)\lambda+$$

$$+\sum_{i=1}^{n}\left(S_i^{э2}-2S_i^{э}s_i^{vir}+s_i^{vir2}\right).$$

Определим точки экстремума целевой функции, для этого потребуется продифференцировать полученное выражение. В результате дифференцирования получим следующую производную:

$$y(\lambda)=2\sum_{i=1}^{n}S_{i-1}^{э\,2}\lambda+2\sum_{i=1}^{n}\left(s_i^{vir}S_{i-1}^{э}-S_i^{э}S_{i-1}^{э}\right).$$

Приравняем полученную производную к 0 и найдем точку экстремума, она будет определяться однозначно, поскольку степень производной по λ равняется единице.

Получим следующее равенство:

$$\sum_{i=1}^{n}S_{i-1}^{э\,2}\lambda=\sum_{i=1}^{n}\left(S_i^{э}S_{i-1}^{э}-s_i^{vir}S_{i-1}^{э}\right).$$

Исходя из вида полученной функции $Y(\lambda)$, можно утверждать, точка экстремума, которая сейчас будет найдена, является точкой минимума, так как построенная функция по переменной λ является параболой. Это доказывается тем, что коэффициент, расположенный в формуле перед λ, всегда является положительным, так как представляет собой произведение константы, которая больше 0 и суммы $\sum_{i=1}^{n}S_{i-1}^{э\,2}$. Первое слагаемое $S_0^{э}$ равно сумме размерностей файлов, содержащихся в исследуемой директории на момент начала моделирования. Так как для заражения необходимы хотя бы файл вируса и первый заражаемый файл, то это слагаемое обязательно будет положительным. То есть, справедливо неравенство $2\sum_{i=1}^{n}S_{i-1}^{э\,2}>0$. Следовательно, исследуемая функция представляет собой параболу, ветви которой направлены вверх. Из этого следует, что вершина параболы представляет собой точку минимума. Проведем оставшиеся преобразования и получим следующую формулу для расчета точки минимума:

$$\lambda_{min}=\frac{\sum_{i=1}^{n}S_i^{э}}{\sum_{i=1}^{n}S_{i-1}^{э}}-\frac{\sum_{i=1}^{n}s_i^{vir}}{\sum_{i=1}^{n}S_{i-1}^{э}}.$$

Для типа внедряемых вирусов, которые не уничтожают содержимое заражаемого файла, константа вируса удовлетворяет соотношению:

$$\lambda_{min}=\frac{\sum_{i=1}^{n}S_i^{э}}{\sum_{i=1}^{n}S_{i-1}^{э}}-n\frac{\underline{s^{vir}}}{\sum_{i=1}^{n}S_{i-1}^{э}}. \qquad (2)$$

Формулу (2) можно использовать для расчета коэффициента памяти любого неполиморфного внедряемого вируса.

4. Пример расчета параметров вируса

Перейдем к построению математической модели функционирования конкретного вируса, который работает по описанному выше алгоритму и внедряется в конец исполняемого файла.

Определим начальное значение количества информации в исследуемой директории $S_0^{э}$ = 3 528 байт, в данной директории находятся незараженные исполняемые файлы, число файлов n = 15, заражение каждого следующего незараженного файла будем считать новым тактом и увеличивать счетчик i на единицу.

Константа вируса для данной вредоносной программы, которая показывает, сколько байт содержит тело вируса, равна:

$$\underline{s^{vir}}=S_1^{f}-S_0^{f}=358-64=294 \text{ байта}.$$

Применим полученные математические выражения для расчета конкретного коэффициента уничтожения вируса ω. Проведем экспериментальные замеры количества информации в исследуемой директории после каждого такта i после заражения каждого следующего COM-файла.

Результаты проведенных измерений представлены в табл. 1. Полученные значения будут использованы для расчета коэффициента памяти компьютера, зараженного исследуемым вирусом. С помощью этого коэффициента несложно получить коэффициент уничтожения вируса. Считаем, что данный вирус не является полиморфным, то есть не изменяет своего программного кода от одного заражения к другому. Это хорошо видно из значений, приведенных в табл. 1, так как количество информации на каждом такте увеличивается на фиксированную величину, то есть на константу вируса $\underline{s^{vir}}$.

Таблица 1. *Результаты потактового экспериментального заражения файлов в директории COM-вирусом*

n	$S_i^э$, байт
0	3528
1	3822
2	4116
3	4410
4	4704
5	4998
6	5292
7	5586
8	5880
9	6174
10	6468
11	6762
12	7056
13	7350
14	7644
15	7938

Подставим полученные экспериментальные значения из табл. 1 в формулу (2) и получим коэффициент памяти λ для вычислителя, зараженного исследуемым вирусом:

$$\lambda_{min} = \frac{88200}{83790} - 15\frac{294}{83790} = 1.$$

Полученное значение коэффициента памяти λ означает, что робот или компьютер, зараженный исследованным вирусом, сохраняет абсолютную память, то есть не будет происходить забывания полученной информации. Коэффициент уничтожения рассматриваемого вируса равен

$$\omega = 1 - \lambda_{min} = 0.$$

Таким образом, для моделирования психологического аналога человека данный вирус не подходит, так как его коэффициент уничтожения ω равен 0, а значит необходимо найти или создать вирус, коэффициент уничтожения которого лежит между 0 и 1: $\omega \in (0,1)$.

5. Алгоритм определения коэффициента памяти компьютера, зараженного вирусом

В вышестоящих пунктах были рассмотрены формулы, которые потребуются для нахождения коэффициента памяти вычислителя, зараженного определенным внедряемым вирусом.

Построим алгоритм определения коэффициента памяти робота, зараженного COM-вирусом. Для решения данной задачи предлагаем следовать следующему алгоритму:

1. Вычислить значение функции $Y(\lambda)$ с граничным значением интервала $\lambda = 0$, т. е. значение $Y(0)$.

2. Вычислить значение функции $Y(\lambda)$ с граничным значением интервала $\lambda = 1$, т. е. значение $Y(1)$.

3. Вычислить значение точки минимума λ_{min}, используя полученные в пункте 3 формулы.

4. Вычислить значение функции $Y(\lambda)$ в полученной на предыдущем шаге точке минимума λ_{min}, т. е. значение $Y(\lambda_{min})$.

5. Выбрать минимальное значение из полученных значений функции $Y(\lambda)$, т. е. выбрать $min[Y(0), Y(1), Y(\lambda_{min})]$.

6. За решение поставленной задачи примем λ, которое соответствует минимальному значению, полученному в сравнении в пункте 5.

Теперь получим выражения для расчета значений функции на границах отрезка, для точки $\lambda = 0$, подставим соответствующее значение в функцию $Y(\lambda)$ и получим следующие выражения:

$$Y(0) = \sum_{i=1}^{n} \left(S_i^э - s_i^{vir} - 0 S_{i-1}^э \right)^2 =$$

$$= \sum_{i=1}^{n} \left(S_i^э - s_i^{vir} \right)^2 = \sum_{i=1}^{n} \left(S_i^{э2} - 2 S_i^э s_i^{vir} + s_i^{vir2} \right).$$

Также нам потребуется выражение для расчета значения функции в точке $\lambda = 1$, подставим соответствующее значение в функцию $Y(\lambda)$ и получим следующую формулу:

$$Y(1) = \sum_{i=1}^{n} \left(S_i^э - s_i^{vir} - 1 S_{i-1}^э \right)^2 =$$

$$= \sum_{i=1}^{n} (S_{i-1}^{э\ 2} + 2 s_i^{vir} S_{i-1}^э - 2 S_i^э S_{i-1}^э +$$

$$+ S_i^{э2} - 2 S_i^э s_i^{vir} + s_i^{vir2}).$$

Полученный алгоритм необходимо проверить на практике, для этого можно использовать исследованный ранее вирус, поскольку часть необходимых параметров для него уже рассчитана. В качестве необходимых экспериментальных значений будем использовать полученные ранее значения из табл. 1. Значение константы вируса остается прежним: $\underline{s^{vir}} = 294$ байта. Перейдем непосредственно к расчетам в рамках алгоритма, для начала найдем значение целевой функции в точке $\lambda = 0$:

$$Y(0) = 542818080 - 2 \cdot 294 \cdot 88200 +$$
$$+15 \cdot 294 \cdot 294 = 489659940.$$

Теперь найдем значение целевой функции в точке $\lambda = 1$, это потребует от нас несколько больше вычислений:

$$Y(1) = 542818080 - 2 \cdot 294 \cdot 88200 -$$
$$-2 \cdot 516887280 + 1296540 +$$
$$+49268520 + 492253020 = 5186160.$$

Точка минимума для данной целевой функции уже была нами найдена ранее, значение ее равно $\lambda_{min} = 1$. Таким образом, количество необходимых вычислений значительно снижается. Значение целевой функции в точке минимума совпадает с вычисленным нами только что значением в точке $\lambda = 1$. Перейдем непосредственно к сравнению полученных значений:

$$min[489659940, 5186160, 5186160] =$$
$$= 5186160.$$

Таким образом, за решение задачи мы принимаем $\lambda_{min} = 1$, так как полученное минимальное значение целевой функции находится подстановкой именно этого значения:

$$Y(1) = Y(\lambda_{min}) = 5186160.$$

Это подтверждает данные предварительного расчета, более того, в процессе решения задачи для данного внедряемого вируса было обнаружено следующее интересное, в плане дальнейшего исследования, соотношение:

$$Y(\lambda_{min}) = 4 \cdot n \cdot (\underline{s^{vir}})^2.$$

Продолжая исследование, попробуем установить, чем вызвано такое соотношение, является ли это случайным совпадением для данного вируса, или же данное равенство не что иное, как закономерность для всех вирусов данного класса.

6. Относительная константа вируса

Продолжая исследовать данный внедряемый вирус, рассчитаем значения относительной константы вируса φ_i^{vir} для каждого из исполняемых файлов в исследуемой директории, после чего найдем среднюю относительную константу вируса $\overline{\varphi^{vir}}$. Изначальные размерности исполняемых файлов и значение относительной константы вируса для каждого из представленных в директории файлов представлены в табл. 2.

Таблица 2. *Значения относительной константы вируса* φ_i^{vir} *для исследуемых* n *исполняемых файлов*

n	S_0^{fi}, байт	φ_i^{vir}
1	470	0.63
2	300	0.98
3	35	8.40
4	503	0.58
5	146	2.01
6	142	2.07
7	57	5.16
8	66	4.45
9	68	4.32
10	52	5.65
11	124	2.37
12	46	6.39
13	68	4.32
14	66	4.45
15	501	0.59

Теперь найдем среднюю относительную константу вируса $\overline{\varphi^{vir}}$, используя значения из табл. 2, получим следующее значение:

$$\overline{\varphi^{vir}} = \frac{n \cdot \underline{s^{vir}}}{\sum_{i=1}^{n} S_0^{fi}} = \frac{15 \cdot 294}{2644} = 1.67.$$

Выводы

Моделирование поведения компьютера, зараженного внедряемым вирусом, показывает, что вредоносное программное обеспечение можно использовать для создания вычислителей с неабсолютной памятью, для этого необходимо подбирать вирусы с соответствующими условиям задачи коэффициентами уничтожения ω.

На основе этого коэффициента и относительной константы вируса φ_i^{vir} возможно ввести новую классификацию компьютерных вирусов в аспекте теории вычислителей с неабсолютной памятью.

Представленный алгоритм поиска параметров может быть применен для нахождения и построения вирусов, необходимых для моделирования обучения группы роботов, которые обучаются совместно, подробнее об этом в работе [5].

Список литературы

1. *Пенский О.Г., Черников К.В.* Основы математической теории эмоциональных роботов. Пермь: изд-во Перм. гос. ун-та. 2010. 256 с.
   URL: https://arxiv.org/find/cs/1/au:+Pensky_O/0/1/0/all/0/1 (дата обращения: 22.09.2017).
2. *Черников К.В.* Математические модели роботов с неабсолютной памятью: автореф. дис. на соиск. учен. степ. канд. физ.-мат. н. (05.13.18). ПНИПУ. Пермь, 2013. 16 с.
3. *Терминология* ГОСТ Р 51275-2006: Защита информации. Объект информатизации. Факторы, воздействующие на информацию. Общие положения.
4. *Виноградов И.М.* Математическая энциклопедия. М.: Советская энциклопедия, 1977. Т. 5.
5. *Пенский О.Г., Ощепкова Н.В., Бахтин В.В.* Математические модели воспитания группы роботов с неабсолютной памятью: в сб.: Искусственный интеллект в решении актуальных социальных и экономических проблем XXI века. Пермь, 2017. С. 192–195.

A mathematical model of searching for the memory coefficient of a computer, infected with a virus, and computer viruses' settings in the context of the theory of solvers with non-absolute memory

V. V. Bakhtin

Perm State University; 15, Bukireva st., Perm, 614990, Russia
bakhtin_94@bk.ru; 89125851599

A mathematical model for the search for the memory coefficient of a computer infected with a virus is proposed, and the definitions of the virus constant, the relative constant of a virus and the destruction rate of a virus are introduced. An algorithm for calculating the parameters of an infected computer and a computer virus is shown, which makes it possible to classify computer viruses in accordance with their destruction coefficients and virus constants. As an example, the calculation of these parameters for an introduced COM virus was performed.

Keywords: *robot; a computer virus; classification of viruses; the constant virus; virus relative constant; coefficient of destruction of the virus; modeling memory.*