УДК 004.056, 004.942

ОПТИМИЗАЦИЯ СОСТАВА СРЕДСТВ ЗАЩИТЫ ИНФОРМАЦИИ МЕТОДОМ ПОСЛЕДОВАТЕЛЬНОГО АНАЛИЗА ВАРИАНТОВ

Е. Ф. Кустов, А. А. Магазев
Омский государственный технический университет, г. Омск, Россия

Аннотация. **В представленной работе рассматривается задача оптимизации состава средств защиты информации, формулируемая в рамках одной марковской модели безопасности. В данной задаче в качестве целевой функции выступает функционал стоимости состава средств защиты, а в качестве допустимого множества решений – наборы средств защиты, обеспечивающие заданное бесперебойное время функционирования информационной системы. В работе обсуждается проблема решения указанной оптимизационной задачи; в частности, анализируется возможность ее решения методом последовательного анализа вариантов. Для сравнения эффективности этого метода был произведен ряд численных экспериментов, результаты которых сравнивались с результатами, полученными прямым перебором. В заключение работы обсуждаются дальнейшие перспективы и направления исследований.**

Ключевые слова: информационная система, угроза безопасности, марковская модель, оптимизация, целочисленное программирование, метод последовательного анализа вариантов.

DOI: 10.25206/2310-4597-2019-1-197-202

I. Введение

С каждым годом проблема оптимизации выбора эффективных средств защиты информации становится всё острее, и как следствие, возникает потребность в нахождении оптимальных путей, а также наиболее выгодных средств защиты информации. Скорость развития информационных технологий с каждым годом становится всё стремительнее, вследствие чего в информационных системах появляется всё больше уязвимостей, которые могут повлечь за собой нанесение ущерба предприятию. Поэтому на рынке появляются различные способы защиты информации. Следовательно, остро встаёт вопрос об оптимальном выборе средств защиты информации. Эту задачу эффективно можно решить с помощью математического моделирования.

Математическое моделирование играет важную роль в защите информации. Использование различных математических моделей, позволяет дать теоретическое обоснование методов и механизмов защиты информации. Кроме того, математические модели позволяют дать количественную оценку *эффективности* и *надёжности,* которая является важнейшей величиной, для построения надёжной системы защиты информации.

Наиболее интересными являются модели, сформулированные в рамках терминов случайных марковских процессов. Данные модели широко применяются в различных областях информационной безопасности: построение модели угроз безопасности [1], оптимизация выбора средств защиты информации [2], обнаружение кибер-атак на информационную систему [3] и т.д.

В статье [2] рассматривается задача оптимизации комплекса средств защиты информации, сформулированная в рамках одной марковской модели угроз информационной системы. Данная работа является продолжением работы [2]. В частности, мы исследуем способы решения указанной оптимизационной задачи методом последовательного анализа вариантов [4].

II. Постановка задачи

Рассмотрим информационную систему (в дальнейшем просто систему), на которую воздействует n независимых внешних угроз с вероятностями $q_1, q_2, \ldots, q_n$. Предположим, что одновременная реализация более двух угроз невозможна и, кроме того, очередная угроза может проявиться только после успешного отражения предыдущей. Следовательно, в каждый момент времени $t = 0,1,2,\ldots$ система находится в одном из $n + 1$ возможных состояний: $s_0, s_1, \ldots, s_n, s_{n+1}$. Состояние s_0 называется безопасным, так как на систему не воздействует ни одна из угроз. На систему в состоянии s_i произвела воздействие i-я угроза, где $i = 1,\ldots,n$. При этом на следующем шаге возможно два исхода события:

1) Данная угроза будет успешно отражена с вероятностью r_i и система вернется в состояние s_0;
2) Данная угроза с вероятностью $\bar{r}_i = 1 - r_i$ приведет к выводу системы из строя.

В данном случае мы будем считать, что система переходит в состояние s_{n+1}.

Граф состояний системы приведен на рис. 1.

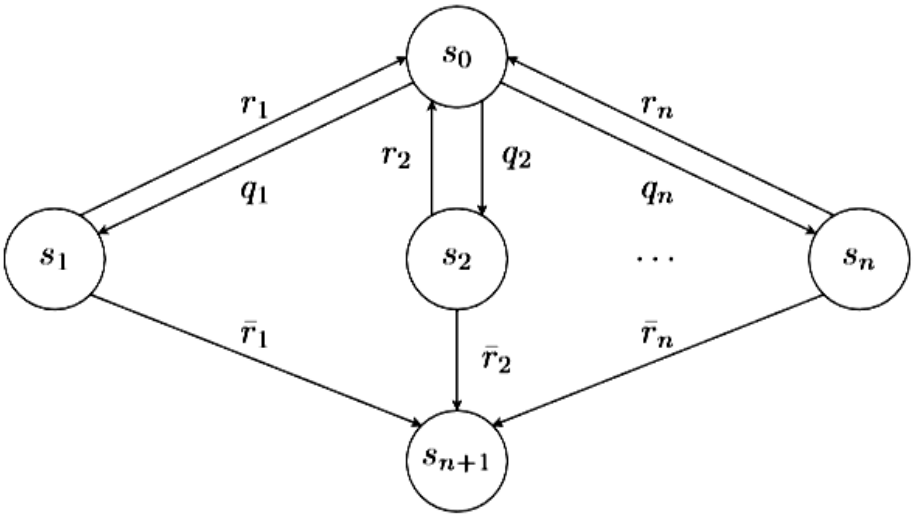

Рис. 1. Граф состояний системы

В динамике система представляет собой марковскую цепь с матрицей переходных вероятностей

$$\Pi = \begin{pmatrix} q_0 & q_1 & q_2 & \dots & q_n & 0 \\ r_1 & 0 & 0 & \dots & 0 & \bar{r}_1 \\ r_2 & 0 & 0 & \dots & 0 & \bar{r}_2 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ r_n & 0 & 0 & \dots & 0 & \bar{r}_n \\ 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix},$$

где $q_0 = 1 - \sum_{i=1}^{n} q_i$.

Вероятность $p_i(t)$ означает, что система находится в момент времени t в состоянии s_i. Эта вероятность определяется через вероятности состояний системы в момент времени $t-1$ согласно формуле

$$p_i(t) = \sum_{j=0}^{n+1} p_j(t-1)\Pi_{ji}.$$

В начальный момент времени $t = 0$ система находится в безопасном состоянии s_0, то есть $p(0) = (1,0,...,0)$.

Как было показано в работе [2], вероятность безопасного состояния системы в произвольный момент времени t выражается следующей формулой:

$$p_0(t) = \frac{1}{w}\left(\frac{q_0 + w}{2}\right)^{t+1} - \frac{1}{w}\left(\frac{q_0 - w}{2}\right)^{t+1}.$$

Здесь положительная величина w определяется как $w^2 = q_0^2 + 4\sum_{i=1}^{n} q_i r_i$.

Временем релаксации τ назовем время, за которое вероятность безопасного состояния системы уменьшается в два раза (по сравнению с моментом времени $t = 0$). Из равенства $p_0(0)/p_0(\tau) = 2$ получаем

$$\tau = \log_{\frac{q_0+w}{2}} \frac{w}{2} - 1. \qquad (1)$$

Пусть T – некоторый фиксированный момент времени. Нас будут интересовать условия, при которых вероятность нахождения системы в безопасном состоянии на временах, меньших или равных T, будет относительно большой. Перепишем неравенство $\tau \geq T$, используя формулу (1), и решим его относительно w. Получаем

$$w \geq 2x^* - q_0, \tag{2}$$

где x^* – вещественный корень уравнения

$$x^{T+1} - x + \frac{q_0}{2} = 0, \tag{3}$$

принадлежащий отрезку $[q_0,1]$. Подставляя в (2) выражение для параметра w, получаем ограничение на значения параметров защиты r_i:

$$\sum_{i=1}^{n} q_i r_i \geq x^*(x^* - q_0). \tag{3}$$

Предположим, что имеется m различных средств защиты, предотвращающих угрозы информационной безопасности. Каждое средство защиты связанно с соответствующей булевой переменной, принимающей значения 1 или 0. Следовательно, все средства защиты можно представить в виде m – мерного булевого вектора $\mathbf{x} = (x_1, \ldots, x_m) \in \{0,1\}^m$.

Обозначим вероятность успешного парирования α-ым средством защиты i-ой угрозы через $r_{i,\alpha}$, где $1 \leq \alpha \leq m$. В один и тот же момент времени сразу несколько средств защиты могут блокировать данную угрозу, поэтому вероятность отражения всеми средствами защиты представим в виде суммы m совместных событий:

$$r_i(\mathbf{x}) = \sum_{n=1}^{m} (-1)^{k-1} \sum\nolimits_{\alpha_1 < \alpha_2 < \cdots < \alpha_n} (r_{i,\alpha_1} x_{\alpha_1})(r_{i,\alpha 2} x_{\alpha_2}) \ldots (r_{i,\alpha_n} x_{\alpha_n}). \tag{4}$$

Пусть c_α – стоимость α- го средства защиты. Рассмотрим целевую функцию C: $\{0,1\}^m \rightarrow \mathbf{R}$:

$$C(\mathbf{x}) = \sum_{\alpha=1}^{m} c_\alpha x_\alpha. \tag{5}$$

Значение данного функционала на векторе $\mathbf{x}$ даёт стоимость соответствующей конфигурации системы защиты.

Используя фиксированный момент времени $T > 0$ можно ограничить значения параметров $r_i(\mathbf{x})$ областью допустимых значений $R_T(q_1,\ldots, q_n) \subset \mathbf{R}^n$, при которых время релаксации будет больше, либо равно T. Также естественным ограничением будет условие сведения стоимости конфигурации к минимуму, то есть $C(\mathbf{x}) \rightarrow \min$. Принимая во внимания указанные ограничения, получаем следующую оптимизационную задачу:

$$C(\mathbf{x}) = \sum_{\alpha=1}^{m} c_\alpha x_\alpha \rightarrow \min,\ \mathbf{x} \in X, \tag{6}$$

где

$$X = \{x \in \{0,1\}^m : \sum_{i=1}^{n} q_i r_i(x) \geq x^*(x^* - q_0)\}. \tag{7}$$

Здесь $q_0 = 1 - \sum_{i=1}^{n} q_0$, а x^* представляет собой вещественный корень уравнения (3), принадлежащий отрезку $[q_0,1]$.

Наиболее прямой способ решить данную оптимизационную задачу – прямой перебор. Очевидно, что при этом необходимо выполнить 2^m итераций, где m – число рассматриваемых средств защиты. Чтобы снизить количество итераций и ускорить вычисления, применим *метод последовательного анализа вариантов* [4].

III. Теория

Метод последовательного анализа вариантов основан на пошаговом конструировании решений и отсеве в процессе такого конструирования тех, которые не могут быть достроены до оптимальных.

Напомним некоторую терминологию. Всякий вектор $\mathbf{x} = (x_1, \ldots, x_m) \in \{0,1\}^m$ будем называть *решением*. Множество всех решений обозначим Ω. Решение будем называть *допустимым*, если оно удовлетворяет неравенству (7). Множество всех допустимых решений обозначим Ω_f. Рассмотрим вектор $\mathbf{x}_{(p)} = (x_1, \ldots, x_p)$, $p < m$. Если он может быть достроен до допустимого решения $(x_1,\ldots, x_p, x_{p+1}, \ldots, x_m)$, то будем называть $\mathbf{x}_{(p)}$ *допустимым частным решением*.

Для решения оптимизационной задачи (6) – (7) необходимо построить набор σ *элиминирующих тестов* ξ_i, которые производят отсев частичных условий, которые не могут быть достроены до допустимых. В такой набор обязательно входят тест ξ_0 – анализ допустимости решений, который сводится к проверке неравенства (7), и тест ξ_1 – сравнение допустимых решений по значению целевой функции (6). *Оценкой* каждого частичного решения $\mathbf{x}_{(p)}$ назовём число $\gamma_C(\mathbf{x}_{(p)})$:

$$\gamma_C(x_{(p)} \le \min\{C(\mathbf{x}) \mid \mathbf{x} \in \Omega\}. \tag{8}$$

Пусть C^* – верхняя граница (рекорд) для минимума задачи (6) – (7). На первых шагах допустим, что рекорд равен бесконечности, а в дальнейшем положим, что он равен значению целевой функции на наилучшим найденном допустимом решении. Тогда дополнительный элиминирующий тест ξ_2 для всякого множества частичных решений h будет задаваться соотношением:

$$\xi_2(h) = \{\mathbf{x}_{(p)} \mid \gamma_C(\mathbf{x}_{(p)}) > C^*\}. \tag{9}$$

Рассмотрим еще один элиминирующий тест ξ_3, который основан на анализе допустимых частичных решений. Перепишем неравенство (7) в виде $g(\mathbf{x}) \le 0$. Очевидно, что для функции $g(\mathbf{x})$ существует нижняя граница $\gamma_g(\mathbf{x}^0_{(p)})$ её значений на множестве $\Omega_f(\mathbf{x}^0_{(p)})$, где $\mathbf{x}^0_{(p)}$ – частичное решение, полученное из $\mathbf{x}_{(p)}$. Следовательно, если имеет место неравенство $\gamma_g(\mathbf{x}^0_{(p)}) > 0$, то частичное решение $\mathbf{x}^0_{(p)}$ *не может быть достроено* до допустимого решения. Исходя из этого, мы можем определить тест ξ_3 следующим условием:

$$\xi_3(h) = \{\mathbf{x}_{(p)} \mid \mathbf{x}_{(p)} \in h, \max \gamma_g(\mathbf{x}_{(p)} > 0)\}. \tag{10}$$

В нашем случае функцию $g(\mathbf{x})$ можно представить в виде (см. формулу (4)):

$$g(\mathbf{x}) = g_1(\mathbf{x}) - g_2(\mathbf{x}), \tag{11}$$

где все $g_1(\mathbf{x})$ и $g_2(\mathbf{x})$ являются неубывающими функциями. Тогда, если для частичного решения $\mathbf{x}^0_{(p)}$ выполняется:

$$\gamma_g(\mathbf{x}_{(p)}) = g_1(x_1^0, \ldots, x_p^0, 0, \ldots, 0) - g_2(x_1^0, \ldots, x_p^0, 1, \ldots, 1) > 0,$$

то его следует исключить из списка согласно тесту ξ_3, так как его невозможно достроить до допустимого.

Решение задачи можно интерпретировать, как продвижение по *дереву решений*, где узлами будут являться частичные решения, причём висячими узлами являются полные решения (см. [4]).

IV. Результаты экспериментов

Основываясь на описанной выше теории, нами была написана программа на языке C++, решающая оптимизационную задачу (6)–(7). На вход программы подавались следующие данные:

- m – исходное число имеющихся средств защиты;
- n – количество угроз безопасности информационной системы;
- q_i – вероятности реализации угроз;
- T – верхняя граница времени релаксации системы;
- r_{i,α_n} – вероятности отражения угроз средствами защиты;
- c_α – стоимости средств защиты.

На выходе программа выдает вектор **x** – набор средств защиты, который удовлетворяет набору элиминирующих тестов σ, а также C_{min} – минимальную стоимость набора средств защиты.

С помощью разработанной нами программы был проведен ряд численных экспериментов, оценивающих эффективность метода последовательного анализа вариантов по сравнению с методом полного перебора. Для этого программа запускалась со значениями параметра m из диапазона от 5 до 16. Остальные переменные задавались случайно с помощью стандартной функции **rand** библиотеки **cstdlib**. В целях получения более объективных результатов, для каждого m было проведено 5 независимых измерений, при каждом из которых остальные входные параметры задачи задавались случайно. После проведения каждой серии экспериментов для всякого фиксированного значения m было вычислено среднее значение числа требуемых для решения задачи итераций и его дисперсия. Полученные результаты представлены в табл. 1.

ТАБЛИЦА 1
РЕЗУЛЬТАТЫ ЧИСЛЕННОГО ЭКСПЕРИМЕНТА

Кол-во средств защиты *m*	**Кол-во итераций при прямом переборе**	**Среднее число итераций при методе ПАВ**	**Дисперсия числа итераций при методе ПАВ**
5	32	38	87
6	64	59	676
7	128	123	2071
8	256	224	19315
9	512	372	45042
10	1024	717	397303
11	2048	1258	2145988
12	4096	2791	1660622
13	8192	4303	11149851
14	16384	2819	2379965
15	32768	8793	35255705
16	65536	11064	50953893

Для иллюстрации мы также приводим графики числа итераций при полном переборе и среднего числа итераций при методе последовательного анализа вариантов.

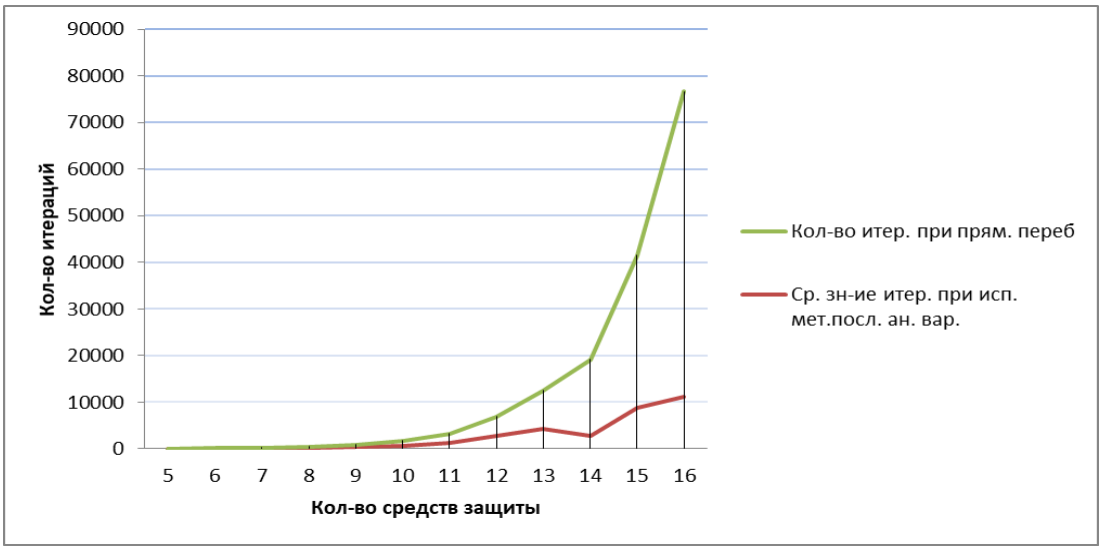

Рис. 2. Сравнение числа итераций при прямом переборе и с использованием метода последовательного анализа вариантов

Как видно из приведённого выше графика число итераций, требуемое для решения оптимизационной задачи (6)–(7), при методе последовательного анализа вариантов меньше чем при методе полного перебора, причем эта разница с ростом m становится более выраженной. Исходя из этого, можно предположить, что метод последовательного анализа вариантов более эффективен для поиска оптимальной конфигурации СЗИ, в особенности, при больших значениях m.

V. Обсуждение результатов

Исходя из представленных выше результатов эксперимента, можно сделать вывод, что метод последовательного анализа вариантов даёт количественный выигрыш, и на практике использование данного метода целесообразно. Тем не менее, для более точных оценок необходимо провести дальнейшие теоретические исследования с целью более строгого анализа зависимости числа итераций от m при применении описанного нами метода.

VI. Выводы и заключение

В представленной работе была рассмотрена оптимизационная задача (6)–(7). Оптимизационная задача формулировалась в рамках марковской модели угроз, исследованной в работе [2]. Произведён численный эксперимент для оценки эффективности метода последовательного анализа вариантов. Эффективность данного метода заключается в выигрыше количества итераций, относительно метода прямого перебора вариантов. Было вычислено средние значение количества итераций и дисперсия выборки. На основе дисперсии, можно предположить, что результаты выполнения программы сильно зависят от совокупности случайных переменных, поданных на вход.

Список литературы

1. Щеглов К. А., Щеглов А. Ю. Марковские модели угрозы безопасности информационной системы // Известия высших учебных заведений. Приборостроение. Автоматика. Вычислительная техника. 2015. № 12 (58). С. 957–965.

2. Magazev A., Tsyrulnik V. Optimizing the selection of information security remedies in terms of a Markov security model // IOP Conf. Series: Journal of Physics: Conf. 2019. Vol. 1096. 012160. DOI:10.1088/1742-6596/1096/1/012160.

3. Гаврилова Е. А. Исследование методов обнаружения сетевых атак // Научные записки молодых исследователей. Автоматика. Вычислительная техника. 2017. № 4 (56). С. 55–58.

4. Ковалёв М. М. Дискретная оптимизация (целочисленное программирование). Минск: Изд-во БГУ им. В.И. Ленина, 1977. 193 с.

УДК 004.056, 004.942

ОБ ОЦЕНКЕ СРЕДНЕГО ВРЕМЕНИ ЖИЗНИ ИНФОРМАЦИОННОЙ СИСТЕМЫ, ПОДВЕРГАЮЩЕЙСЯ УГРОЗАМ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

А. А. Магазев, А. А. Касенов
Омский государственный технический университет, г. Омск, Россия

Аннотация. **В работе изучается модель угроз, воздействующих на информационную систему. Модель представлена на языке случайных марковских процессов и описывает функционирование информационной системы как последовательность воздействий и отражений угроз. В результате исследования были получены явные аналитические формулы для среднего времени жизни информационной системы. Информационная система в свою очередь представлена структурной схемой с последовательным и параллельным соединением её элементов. Рассмотрим более подробно каждое соединение в частности. В первом случае, в случае последовательного соединения, при выводе из строя одного элемента - вся система выходит из строя. Во втором же случае для вывода из строя всей информационной системы необходимо чтобы все элементы такой системы вышли из строя. В дальнейшем такую систему можно сколько угодно усложнять.**

Ключевые слова: информационная система, угроза безопасности, марковский процесс.

DOI: 10.25206/2310-4597-2019-1-202-207

I. Введение

Моделирование в рамках информационной безопасности является одним из базовых методов в рамках информационной безопасности. Ключевым недостатком альтернативного метода – метода натурных испытаний – является его высокая стоимость, а также трудоемкость. Также метод натурных испытаний не даёт теоретического обоснования методов защиты информации.

В настоящей работе мы продолжаем исследование одной марковской модели угроз информационной безопасности, предложенной в статьях [1, 2], и частично проанализированной в работе [3].