Оптимизация выбора средств защиты информации в рамках одной марковской модели безопасности

А.А. Магазев[1], **В.Ф. Цырульник**[1]

[1]Омский государственный технический университет, пр. Мира 11, Омск, Россия, 644050

Аннотация. В работе рассматривается модель информационной безопасности, формулируемая на языке марковских процессов. В рамках этой модели функционирование информационной системы описывается как последовательность отказов и восстановлений, которые возникают вследствие воздействий на систему угроз безопасности. Мы проводим детальное исследование модели и вводим ее важную характеристику, называемую *временем релаксации*, с помощью которой конструируем допустимую область параметров безопасности модели. В заключение мы формулируем и обсуждаем проблему выбора средств защиты информации как проблему нелинейной оптимизации с булевыми переменными.

1. Введение

Роль моделирования в области информационной безопасности и защиты информации трудно переоценить. Разработка и использование различных математических моделей в этой области необходимы для надлежащего теоретического обоснования механизмов и методов защиты, используемых при проектировании и эксплуатации реальных объектов. Кроме того, применение математических моделей позволяет более строго формулировать такие важные характеристики систем защиты информации как *эффективность* и *надежность*, которые являются количественными мерами качества функциональной реализации выполняемых системой задач.

Особое место среди использующихся в настоящее время моделей информационной безопасности занимают модели, формулируемые в терминах случайных марковских процессов. Спектр прикладных задач, решаемых с применением подобных моделей, необычайно широк: обнаружение кибер-атак в компьютерных сетях [1], моделирование процессов распространения компьютерных вирусов [2], обнаружение вторжений в компьютерных системах и вычислительных сетях [3, 4], оптимизация и повышение надежности защищенных информационных систем [5].

В работах [6, 7, 8, 9] исследуется класс марковских моделей, в которых компьютерные системы, подвергающиеся угрозам информационной безопасности, рассматриваются как системы с отказами и восстановлениями. Подобный подход к исследованию защищенных информационных систем позволил привлечь развитый математический аппарат теории надежности, хорошо зарекомендовавший себя при расчетах и проектировании сложных технических систем. В частности, в статьях [6, 7] была предложена марковская модель с конечным числом состояний, характеризующих степень влияния угроз на информационную систему. В статье авторов [10] указанная модель была изучена более углубленно; в частности, было исследовано асимптотическое поведение модели при больших временах,

а также была введена ее важная характеристика, называемая *временем релаксации*. Кроме того, в работе [10] также был предложен алгоритм построения области значений параметров защиты системы, при которых время релаксации является не меньшим некоторой заданной величины.

Настоящая работа является продолжением исследования [10]. Ниже мы приводим подробное описание рассматриваемой модели, даем определение времени релаксации и даем описание алгоритма построения допустимой области параметров защиты системы. Кроме того, в терминах исследуемой модели мы формулируем и обсуждаем проблему оптимального выбора средств защиты информации как проблему нелинейной оптимизации с булевыми переменными.

2. Описание модели

Рассмотрим компьютерную систему (в дальнейшем просто *систему*), на которую действует n независимых угроз с вероятностями q_1, q_2, …, q_n. Будем считать, что одновременное появление двух и более угроз невозможно и, кроме того, очередная угроза может проявиться только после успешного парирования предыдущей. В соответствие с этим, в каждый момент времени $t = 0, 1, 2, \ldots$ система находится в одном из $n+1$ возможных состояний: s_0, s_1, …, s_n, s_{n+1}. В состоянии s_0, называемом *безопасным*, ни одна из угроз не реализуется. Состояние s_i, $1 \le i \le n$, характеризуется действием i-ой угрозы. При этом в последующий момент времени имеется две альтернативы: либо данная угроза будет успешно отражена с вероятностью r_i и система вернется в состояние s_0, либо с вероятностью $\bar{r}_i = 1 - r_i$ эта угроза приведет к выводу системы из строя. В последнем случае мы будем считать, что система переходит в состояние s_{n+1}. Граф состояний системы приведен на рис. 1.

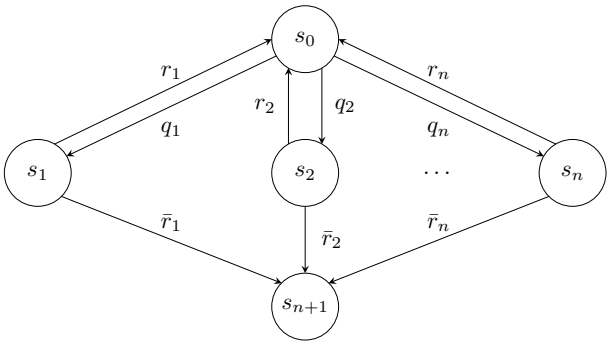

Рисунок 1. Граф состояний модели

Динамика системы представляет собой простую марковскую цепь с матрицей переходных вероятностей

$$\Pi = \begin{pmatrix} q_0 & q_1 & q_2 & \ldots & q_n & 0 \\ r_1 & 0 & 0 & \ldots & 0 & \bar{r}_1 \\ r_2 & 0 & 0 & \ldots & 0 & \bar{r}_2 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ r_n & 0 & 0 & \ldots & 0 & \bar{r}_n \\ 0 & 0 & 0 & \ldots & 0 & 1 \end{pmatrix}. \qquad (1)$$

Здесь мы ввели обозначение $q_0 = 1 - \sum_{i=1}^{n} q_i$.

Обозначим через $p_i(t)$ вероятность состояния s_i в момент времени t. Эта величина определяется через вероятности состояний системы в предыдущий момент времени $t-1$ согласно формуле

$$p_i(t) = \sum_{j=0}^{n+1} p_j(t-1)\,\Pi_{ji}, \tag{2}$$

или в матричной форме

$$\mathbf{p}(t) = \mathbf{p}(t-1)\cdot\Pi, \tag{3}$$

где $\mathbf{p}(t) = (p_0(t), p_1(t), \ldots, p_{n+1}(t))$ — вектор вероятностей состояний системы в момент t. Используя (3) можно записать

$$\mathbf{p}(t) = \mathbf{p}(0)\cdot\Pi^t, \quad t = 0, 1, 2, \ldots, \tag{4}$$

где через Π^t обозначена t-ая степень матрицы Π.

Будем считать, что в начальный момент времени $t = 0$ система находилась в безопасном состоянии s_0, то есть $\mathbf{p}(0) = (1, 0, \ldots, 0)$. Простыми вычислениями можно показать, что в этом случае для вероятностей $p_i(t)$ имеем выражения [10]:

$$p_0(t) = w^{-1}\left[\left(\frac{q_0+w}{2}\right)^{t+1} - \left(\frac{q_0-w}{2}\right)^{t+1}\right]; \tag{5}$$

$$p_i(t) = p_0(t-1)q_i, \quad 1 \le i \le n; \tag{6}$$

$$p_{n+1}(t) = 1 - p_0(t) - p_0(t-1)\sum_{i=1}^{n} q_i. \tag{7}$$

Здесь положительная величина w, которую мы далее будем называть *w-параметром* модели, определяется как

$$w^2 = q_0^2 + 4\sum_{i=1}^{n} r_i q_i > 0. \tag{8}$$

Отсутствие защиты в системе характеризуется условием $r_i = 0$, $1 \le i \le n$. Согласно (8) тогда имеем $w = q_0$, то есть w-параметр в этом случае совпадает с вероятностью отсутствия реализации любой из угроз. Отсюда для вероятностей состояний системы будем иметь:

$$p_0(t) = q_0^t, \quad p_i(t) = q_0^{t-1}q_i, \quad 1 \le i \le n; \quad p_{n+1}(t) = 1 - q_0^{t-1}. \tag{9}$$

Из этих равенств легко следуют предельные выражения для вероятностей состояний системы при $t \to \infty$:

$$\lim_{t\to\infty} p_0(t) = \cdots = \lim_{t\to\infty} p_n(t) = 0, \quad \lim_{t\to\infty} p_{n+1}(t) = 1.$$

Эти же предельные соотношения будут иметь место и в общем случае, когда не все параметры r_i равны нулю. Чтобы показать это, достаточно убедиться, что величины $(q_0 \pm w)/2$, стоящие в круглых скобках правой части равенства (5), по модулю всегда меньше единицы. Указанный факт, в свою очередь, является следствием того, что $(q_0 \pm w)/2$ представляют собой вещественные корни квадратного уравнения

$$f(x) = x^2 - q_0 x - \sum_{i=1}^{n} q_i r_i = 0,$$

которые в силу неравенств $f(\pm 1) > 0$ и $f(0) < 0$ принадлежат интервалу $(-1, 1)$.

Из формул (5) – (7) следует, что вероятности $p_i(t)$, $1 \leq i \leq n+1$ полностью определены, если известна вероятность $p_0(t)$. Таким образом, далее мы можем сосредоточиться только на исследовании функции $p_0(t)$.

Из сказанного выше следует, что $p_0(t)$ будет уменьшаться с ростом t, однако в общем случае данная зависимость является не монотонной. Это легко видно из формулы (5): величина $p_0(t)$ представляется как разность двух стоящих в квадратных скобках слагаемых, первое из которых является монотонно убывающей функцией от t, а второе имеет осцилляционный характер (ввиду отрицательности величины $(q_0 - w)/2$). Скорость убывания $p_0(t)$ и выраженность ее осцилляций зависят, вообще говоря, от конкретных значений параметров модели q_i и r_i. Например, на рис. 2 приведены графики функции $p_0(t)$ при двух различных значениях параметров q_1 и r_1 в случае одной угрозы.

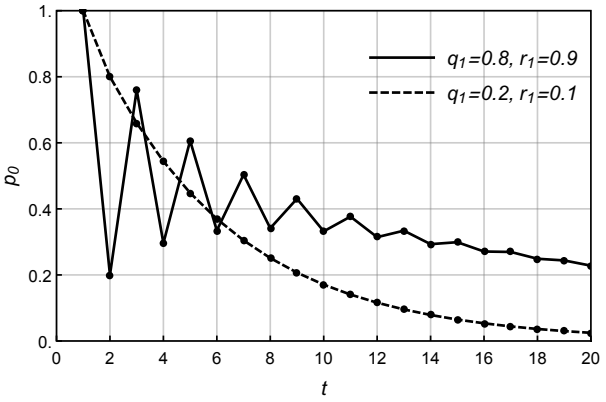

Рисунок 2. График $p_0(t)$ для различных значений q_1 и r_1 в случае одной угрозы

В силу того, что

$$\lim_{t\to\infty} \left|\frac{q_0 - w}{q_0 + w}\right|^t = 0,$$

второе слагаемое, стоящее в квадратных скобках в выражении (5), является величиной бесконечно малой более высокого порядка, чем соответствующее первое слагаемое. Это означает, что «амплитуда» осцилляций функции $p_0(t)$ быстро убывает с ростом t, и на больших временах мы можем приближенно считать

$$p_0(t) \approx p_0^*(t) = w^{-1} \left(\frac{q_0 + w}{2}\right)^{t+1}. \tag{10}$$

Нетрудно оценить условие применимости приближения (10). Пусть $\varepsilon > 0$. Тогда требование $|p_0(t) - p_0^*(t)| < \varepsilon$ эквивалентно неравенству

$$t > \log_{\frac{w-q_0}{2}} \varepsilon w - 1. \tag{11}$$

Таким образом, приближенное выражение (10) отличается от истинной вероятности $p_0(t)$ на величину, не большую ε, если система рассматривается на временах, удовлетворяющих условию (11).

3. Время релаксации и построение допустимой области параметров защиты

Будем рассматривать динамику системы на временах, удовлетворяющих неравенству (11) для некоторого малого $\varepsilon > 0$. Тогда с точностью до величин порядка ε мы можем приближенно считать

$$p_0(t) \approx w^{-1} \left(\frac{q_0 + w}{2} \right)^{t+1}. \tag{12}$$

В дальнейшем для нас будет важным то, что в рамках рассматриваемого приближения величина $p_0(t)$ является монотонно убывающей функцией времени.

Временем релаксации τ модели назовем время, за которое вероятность безопасного состояния системы уменьшается в *два раза* (по сравнению с начальным моментом времени). С использованием (12) из условия $p_0(0)/p_0(\tau) = 2$ немедленно получаем явное выражение для τ:

$$\tau = \log_{\frac{q_0+w}{2}} \left(\frac{w}{2} \right) - 1. \tag{13}$$

Пусть $T > 0$ — некоторый фиксированный момент времени. Наша ближайшая задача будет заключаться в нахождении значений параметров защиты $r_1, \ldots, r_n$, при которых $\tau \geq T$. Другими словами, нас будут интересовать условия, при которых время релаксации модели не меньше некоторого заранее заданного значения.

Используя формулу (13), перепишем неравенство $\tau \geq T$:

$$\frac{w}{2} \leq \left(\frac{q_0 + w}{2} \right)^{T+1}. \tag{14}$$

Будем рассматривать данное неравенство, как ограничение на параметры защиты системы $r_1, \ldots, r_n$; так как последние входят только в выражение для w-параметра, нам необходимо решать неравенство (14) относительно переменной w. Нетрудно видеть, что соответствующее решение имеет вид

$$w \geq 2x^* - q_0, \tag{15}$$

где x^* — вещественный корень полиномиального уравнения

$$x^{T+1} - x + \frac{q_0}{2} = 0, \tag{16}$$

принадлежащий отрезку $[q_0, 1]$.

Подставляя в неравенство (15) явное выражение для w-параметра (8), получаем ограничение на значения параметров защиты r_i:

$$\sum_{i=1}^{n} q_i r_i \geq x^*(x^* - q_0). \tag{17}$$

Вместе с неравенствами

$$r_i \leq 1, \quad i = 1, \ldots, n; \tag{18}$$

требование (17) определяет в пространстве значений параметров защиты выпуклую область $R_T(q_1, \ldots, q_n) \subset \mathbb{R}^n_+$, которую мы будем называть *допустимой*. Таким образом, только для значений параметров $r_1, \ldots, r_n$ из допустимой области $R_T(q_1, \ldots, q_n)$ время релаксации τ системы будет не меньшим, чем фиксированное значение T.

4. Задача выбора оптимального набора средств защиты

Допустим, что имеется m различных средств защиты, отражающих угрозы информационной безопасности. С каждым средством защиты удобно связать соответствующую булеву переменную, принимающую значение 1, если данное средство функционирует, и 0 — в обратном случае. Таким образом, вся совокупность средств защиты информации характеризуется m-мерным булевым вектором $\mathbf{x} = (x_1, \ldots, x_m) \in \{0,1\}^m$.

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

$$r_i(\mathbf{x}) = \sum_{\gamma=1}^{m} (-1)^{\gamma-1} \sum_{\alpha_1<\alpha_2<\cdots<\alpha_\gamma} (r_{i,\alpha_1} x_{\alpha_1})(r_{i,\alpha_2} x_{\alpha_2}) \ldots (r_{i,\alpha_\gamma} x_{\alpha_\gamma}). \tag{19}$$

Отметим, что в общем случае величина $r_i(\mathbf{x})$ является полиномом m-го порядка от булевых переменных x_α.

В рамках описанной нами выше модели, i-ая угроза безопасности будет успешно парирована системой защиты с вероятностью $r_i(\mathbf{x})$, либо не парирована с вероятностью $1 - r_i(\mathbf{x})$. Отсюда следует, что величины $r_i(\mathbf{x})$, $1 \leq i \leq n$, являются важными параметрами системы защиты, характеризующими качество ее функционирования. В связи с этим, естественной представляется задача о выборе такой конфигурации системы защиты, при которой значения параметров $r_i(\mathbf{x})$ будут удовлетворять определенным условиям. Сформулируем в терминах нашей модели одну из таких задач, часто возникающую на практике.

Обозначим c_α стоимость α-го средства защиты (в условных денежных единицах). Рассмотрим линейный функционал $C : \{0,1\}^m \to \mathbb{R}$:

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

Очевидно, что значение этого функционала на векторе $\mathbf{x}$ дает общую стоимость соответствующей конфигурации системы защиты (в которой присутствуют или отсутствуют конкретные средства защиты, в зависимости от значений соответствующих булевых переменных).

Зафиксируем некоторый момент времени $T > 0$. Следуя предыдущему разделу, мы можем ограничить значения параметров $r_i(\mathbf{x})$ областью допустимых значений $R_T(q_1, \ldots, q_n) \subset \mathbb{R}^n$, при которых время релаксации модели будет не меньшим, чем T. При этом ограничении естественно потребовать условие минимальности стоимости конфигурации системы защиты, то есть минимума функционала $C(\mathbf{x})$. Таким образом, мы приходим к следующей оптимизационной задаче:

$$C(\mathbf{x}) = \sum_{\alpha=1}^{m} c_\alpha x_\alpha \to \min, \quad \mathbf{x} \in X, \tag{20}$$

где

$$X = \{\mathbf{x} \in \{0,1\}^m : \sum_{i=1}^{n} q_i r_i(\mathbf{x}) \geq x^*(x^* - q_0)\}. \tag{21}$$

Напомним, что здесь $q_0 = 1 - \sum_{i=1}^{n} q_i$, а x^* представляет собой вещественный корень уравнения (16), принадлежащий отрезку $[q_0, 1]$.

В силу того, что величины $r_i(\mathbf{x})$ являются полиномами m-ой степени от булевых переменных x_α, оптимизационная задача (20), (21) относится к классу задач нелинейного дискретного программирования. Как известно, общих эффективных методов решения таких задач на сегодняшний день не существует [11, 12]. Тем не менее, при не очень больших значениях m (что, как правило, и имеет место на практике) задача (20), (21) может быть решена простым методом перебора. При этом данная задача может допускать одно или несколько решений, либо не иметь его вовсе. Проиллюстрируем это на простом примере.

Рассмотрим систему, на которую действуют две угрозы с вероятностями $q_1 = 0.59$ и $q_2 = 0.18$, откуда $q_0 = 0.23$ Допустим, что у нас имеется $m = 5$ различных средств защиты, стоимости которых равны $c_1 = 600$ у.е., $c_2 = 700$ у.е., $c_3 = c_5 = 400$ у.е., $c_4 = 200$ у.е. Таким образом, функционал стоимости системы защиты в данном примере имеет вид

$$C(\mathbf{x}) = 600x_1 + 700x_2 + 400(x_3 + x_5) + 200x_4.$$

Вероятности $r_{i,\alpha}$ отражения угроз данными средствами защиты заданы следующей матрицей:

$$\|r_{i,\alpha}\| = \begin{pmatrix} 0.96 & 0.50 & 0.47 & 0.50 & 0.07 \\ 0.66 & 0.43 & 0.60 & 0.69 & 0.92 \end{pmatrix}.$$

Найдем решения оптимизационной задачи (20), (21) для заданных значений параметров при $T = 5, 10, 100$.

1) Случай $T = 5$. Корень x^* уравнения (16), принадлежащий отрезку $[q_0, 1]$, равен $x^* = 0.975$. Область X согласно (21) имеет вид

$$X = \{\mathbf{x} \in \{0,1\}^5 \colon 0.59\, r_1(\mathbf{x}) + 0.18\, r_2(\mathbf{x}) \geq 0.726\},$$

где $r_1(\mathbf{x})$ и $r_2(\mathbf{x})$ определяются формулой (19) (мы не выписываем здесь их явные выражения ввиду громоздкости последних). Соответствующее решение $\bar{\mathbf{x}}$ оптимизационной задачи (20), (21) единственно и имеет вид:

$$\bar{\mathbf{x}} = (1,0,0,1,0), \quad C(\bar{\mathbf{x}}) = 800 \text{ у.е.}$$

2) Случай $T = 10$. В этом случае $x^* = 0.988$, так что область X имеет вид

$$X = \{\mathbf{x} \in \{0,1\}^5 \colon 0.59\, r_1(\mathbf{x}) + 0.18\, r_2(\mathbf{x}) \geq 0.749\}.$$

Оптимизационная задача (20), (21) допускает здесь два решения $\bar{\mathbf{x}}_1$ и $\bar{\mathbf{x}}_2$:

$$\bar{\mathbf{x}}_1 = (1,0,0,1,1), \quad \bar{\mathbf{x}}_2 = (1,0,1,1,0), \quad C(\bar{\mathbf{x}}_1) = C(\bar{\mathbf{x}}_2) = 1200 \text{ у.е.}$$

3) Случай $T = 100$. Здесь $x^* = 0.999$, откуда

$$X = \{\mathbf{x} \in \{0,1\}^5 \colon 0.59\, r_1(\mathbf{x}) + 0.18\, r_2(\mathbf{x}) \geq 0.772\}. \tag{22}$$

В этом случае оптимизационная задача (20), (21) не имеет решений, так как не существует конфигураций $\mathbf{x} \in \{0,1\}^5$, принадлежащих области (22).

5. Заключение

В работе рассмотрена модель угроз информационной безопасности, описываемая в терминах марковских процессов. Динамика модели представляется последовательностью сбоев и восстановлений компьютерной системы, происходящих в результате воздействия случайных внешних угроз. Приведены явные аналитические формулы для вероятностей состояний системы, обсуждаются их некоторые предельные случаи и анализируется поведение системы на больших временах. В терминах времени релаксации модели приводится алгоритм построения допустимой области значений параметров защиты системы. Кроме того, в терминах рассмотренной модели сформулирована задача выбора средств защиты информации как задача нелинейного программирования с булевыми переменными.

6. Литература

[1] Piétre-Cambacédès, L. Beyond Attack Trees: Dynamic Security Modeling with Boolean Logic Driven Markov Processes (BDMP) / L. Piétre-Cambacédès, M. Bouissou // Proceedings of the 2010 European Dependable Computing Conference. — IEEE Computer Society, 2010. — P. 199–208. DOI: 10.1109/EDCC.2010.32.

[2] Далингер, Я. Математические модели распространения вирусов в компьютерных сетях различной структуры / Я.М. Далингер, Д.В. Бабанин, С. М. Бурков // Информатика и системы управления. — 2011. — № 4. — С. 3–11.

[3] Ye, N. A Markov Chain Model of Temporal Behavior for Anomaly Detection / N. Ye // Proceeding on the 2000 IEEE Systems, Man, and Cybern. Information Assurance and Security Workshop. — IEEE Computer Society, 2000. — P. 171–174.

[4] Kovalev, S.M. Anomaly detection based on Markov chain model with production rules / S.M. Kovalev, A.V. Sukhanov // Программные продукты и системы. — 2014. — Т. 107, № 3. — С. 40–43. DOI:10.15827/0236-235X.107.040-043.

[5] Щеглов, К.А. Математические модели эксплуатационной информационной безопасности / К.А. Щеглов, А.Ю. Щеглов // Вопросы защиты информации. — 2014. — № 3. — С. 52–65.

[6] Клименко, Е.С. Марковская модель оценки влияния внутренних угроз на безопасность конфиденциальной информации / Е.С. Клименко, А.П. Росенко // Известия ЮФУ. Технические науки. — 2007. — Т. 76, №4. — С. 123–126.

[7] Росенко, А.П. Математическое моделирование влияния внутренних угроз на безопасность конфиденциальной информации, циркулирующей в автоматизированной информационной системе / А.П. Росенко // Известия ЮФУ. Технические науки. — 2008. — Т. 85, № 8. — С. 71–81.

[8] Щеглов, К.А. Марковские модели угрозы безопасности информационной системы / К.А. Щеглов, А.Ю. Щеглов // Известия высших учебных заведений. Приборостроение. — 2015. — Т. 58, № 12. — С. 957–965. DOI: 10.17586/0021-3454-2015-58-12-957-965.

[9] Щеглов, К.А. Моделирование угрозы безопасности информационной системы с использованием аппроксимирующих функций / К.А. Щеглов, А.Ю. Щеглов // Известия высших учебных заведений. Приборостроение. — 2016. — Т. 59, № 1. — С. 50–59. DOI: 10.17586/0021-3454-2016-59-1-50-59.

[10] Магазев, А.А. Исследование одной марковской модели угроз безопасности компьютерных систем / А.А. Магазев, В.Ф. Цырульник // Моделирование и анализ информационных систем. — 2017. — Т. 24, № 4. — С. 445–458. DOI: 10.18255/1818-1015-2017-4-445-458.

[11] Onn, S. Nonlinear discrete optimization / S. Onn. — European Mathematical Society, 2010. — 137 p.

[12] Ковалев, М.М. Дискретная оптимизация (целочисленное программирование) / М.М. Ковалев. — М.: Едиториал, 2003. — 192 с.

Optimizing the selection of information security remedies in terms of one Markov security model

A.A. Magazev[1], V.F. Tsyrulnik[1]

[1]Omsk State Technical University, Mira ave. 11, Omsk, Russia, 644050

Abstract. In this work, an information security model formulated in terms of Markov processes is considered. In the framework of this model the functioning of an information system is described as a sequence of failures and recovery actions which appear as results of security threats acting on the system. We provide a detailed investigation of the model and introduce its important characteristic, called the relaxation time, by means of which we construct the permitting domain of the security parameters of the model. Finally, we formulate and discuss the problem of the selection for information security remedies as a problem of nonlinear optimization with Boolean variables.

Keywords: Information system, Security threat, Markov process, Nonlinear discrete optimization.