УДК 004. 942

А.П. Росенко

Методика обработки массива исходных данных, полученных экспертным путем

Предлагается методика обработки исходных данных, полученных экспертным путем, методами математической статистики. Основная идея метода заключается в необходимости разработки аппарата, позволяющего на основе методов математической статистики получить массивы исходных данных, адекватно отражающих последствия от воздействия различных угроз на безопасность КИ.

Ключевые слова: конфиденциальная информация, экспертные оценки, угрозы, внутренние угрозы, алгоритм, выборка, безопасность.

Актуальность проблемы. Экспертные оценки (ЭО) находят широкое применение при исследовании вопросов безопасности конфиденциальной информации [1]. Это связано с тем, что проблема безопасности КИ относится к трудно формализуемым проблемам. В связи с этим в настоящее время отсутствует достаточно разработанный научно-методический аппарат, позволяющий получить объективные исходные данные для исследования вопросов защиты информации, а применение существующих методов, разработанных в других областях знаний, затруднено в силу неадекватности реальных процессов, имеющих место при воздействии на АИС различных угроз. Для решения указанной проблемы в работе представлен метод обработки массива исходных данных, полученных экспертным путем.

Постановка задачи. Пусть требуется оценить влияние i-х ($i = 1, n$) угроз на безопасность КИ. Пусть также имеются j-е ($j = 1, k$) группы экспертов, осуществляющие такую оценку. В результате экспертной оценки получены частные выборки, представленные в виде матриц.

Требуется провести обработку массива полученных исходных данных методами математической статистики, установить степень однородности частных выборок, построить обобщенную статистическую модель в виде регрессионной модели, а также оценить адекватность регрессионной модели, доказать существенность коэффициентов регрессии и определить относительный вклад каждого оцениваемого фактора в значение безопасности КИ.

Алгоритм обработки массива исходных данных. Алгоритм обработки массива исходных данных, полученных с помощью экспертных оценок, представлен на рис. 1.

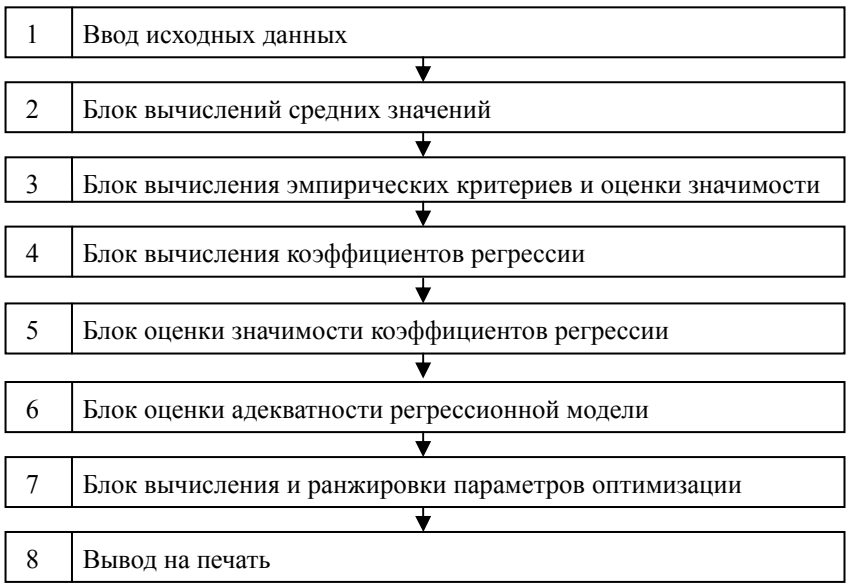

Рис. 1. Алгоритм обработки массива исходных данных

В соответствии с рис. 1 алгоритм состоит из восьми этапов. Первый этап предполагает формирование и ввод исходных данных, полученных в результате экспертного опроса. Особенностью этого этапа является то, что перед вводом исходных данных частная выборка проверяется на соответствие малой выборке по методике, изложенной в [2]. После указанной процедуры осуществляется ввод исходных данных. На втором этапе определяется суммарное среднее значение каждого *j*-го эксперимента для всех *i*-х оцениваемых внутренних угроз на *k*-м уровне, т.е.

$$\mathrm{A}^{(k)}{}_{ij} = \frac{1}{n}\sum_{i=1}^{n} a_{ij}\,, \tag{1}$$

и общее среднее значение по всем *j*-м экспериментам, оценивающим АИС по *i*-й внутренней угрозе:

$$\mathrm{A}^{(k)} = \frac{1}{m}\sum_{j=1}^{m} \mathrm{A}^{(k)}{}_{ij}\,, \tag{2}$$

или, подставляя (1) в (2), получим:

$$\mathrm{A}^{(k)} = \frac{1}{m}\frac{1}{n}\sum_{i=1}^{m}\sum_{i=1}^{n} a_{ij}\,, \tag{3}$$

где a_{ij} – балл, установленный *j*-экспертом *i*-й внутренней угрозе; *n* – число оцениваемых свойств; *m* – количество экспертов.

На третьем этапе устанавливается степень однородности средних арифметических оценок для различных групп экспертов по всем оцениваемым внутренним угрозам. Для этих целей определяется среднее арифметическое (M_j) для каждой группы экспертов, т.е.

$$\mathrm{M}_j = \frac{\sum_{j=1}^{n} X_j}{n}\,, \tag{4}$$

где χ_i – абсолютная величина оценки, данная *j*-м экспериментом; *n* – количество экспертов, оценивающих данную АИС.

В дальнейшем определяется общее среднее значение для всех групп экспертов по каждой АИС:

$$\mathrm{M} = \frac{1}{\sum_{j=1}^{m} n_j}\sum_{j=1}^{m} \mathrm{M}_J n\,,$$

где *m* – количество групп экспертов, а также дисперсия (D[M]) относительно среднего значения:

$$D[\mathrm{M}] = \frac{1}{\sum_{j=1}^{m} n_j}\sum_{j=1}^{m} (\mathrm{M}_J - \mathrm{M})^2 n_j\,. \tag{5}$$

Кроме того, определяются выборочные дисперсии (D_j), т.е. дисперсии по каждой группе экспертов:

$$D_j = \frac{\sum_{j=1}^{n} X^2 j}{n_j} - M^2{}_j\,, \tag{6}$$

и среднее арифметическое из этих дисперсий (D_0):

$$D_0 = \frac{1}{\sum_{i=1}^{m} n^{i=1}}\sum^{m} Djn\,. \tag{7}$$

Эмпирическое значение $F_Э$-критерия рассчитывается по формуле

$$F_Э = \frac{D[M](\sum_{j=1}^{m} n_I - m)}{D_0(m-1)}\,. \tag{8}$$

Используя табличные значения, например представленные в [3], определяется критическое значение $F_\alpha = \nu_1;\nu_2$, где α – вероятность отбрасывания истинной гипотезы;

$$\left.\begin{array}{l}\nu_1 = m-1 \\ \nu_2 = \sum_{i=1}^{m} n_j - m\end{array}\right\}$$ – число степеней свободы и проверяется условие:

если $F_Э \le F_{0,05=\nu_1;\nu_2}$ – средние арифметические отличаются случайным образом;

если $F_Э > F_{0,01=\nu_1;\nu_2}$ – отличия не случайны.

В дальнейшем устанавливается однородность дисперсий по критерию Бартлетта. Для расчета однородности дисперсий по критерию Бартлетта используются уже ранее полученные значения выборочных дисперсий (D_j) и общие средние арифметические (D_0) из этих дисперсий. Значение эмпирического критерия Бартлетта рассчитывается по формуле

$$B_Э = 2{,}3026[(\sum_{j=1}^{m} n_j - m)\lg D_0 - \sum_{j=1}^{m}(n-1)\lg D_J]. \quad (9)$$

Затем сопоставляется значение $B_Э$ с критическим значением $\chi^2_{\alpha;\nu}$. Если $B_Э \le \chi^2_{0,05;\nu}$ (где $\nu = m-1$), то гипотеза об однородности принимается, а если $B_Э > \chi^2_{0,01;\nu}$ – отклоняется.

На четвертом этапе производится оценка степени влияния различных угроз на безопасность КИ, а также определяются параметры статистических моделей, устанавливающих взаимосвязь указанных внутренних угроз и безопасности КИ. В этих целях выбрана полиномиальная модель не выше второго порядка вида

$$Y = \beta_0 + \sum_{i=1}^{k}\beta_i\chi_i + \sum_{i,j}\beta_{i,j}\chi_i\chi_j + \sum_{i=1}^{k}\beta_i\chi^2{}_i + \varepsilon, \quad (10)$$

где Y – значение, характеризующее безопасность КИ; k – количество исследуемых внутренних угроз; $\chi_i\chi_j$ – независимые переменные (единичные свойства), характеризующие линейные эффекты; $\chi_i\chi_j$ – эффекты взаимодействия; $\chi^2{}_i$ – эффекты второго порядка; β_0, β_i, $\beta_{i,j}$ – параметры модели, значения которых определяются по данным экспериментов; ε – остаток, характеризующий ошибку эксперимента и ошибку выбора модели.

Рассматриваемая модель является моделью второго порядка по независимым переменным и линейной по параметрам β. Как показано в [3], путем соответствующих замен нелинейные модели могут быть приведены к линейной типа

$$Y = \beta_0 + \sum_{i=1}^{k}\beta_i\chi_i + \varepsilon. \quad (11)$$

После определения оценок параметров β_i уравнение регрессии (статистическая модель) примет вид

$$\overline{Y} = \beta_0 + \sum_{i=1}^{k}\beta_i\chi_i, \quad (12)$$

где $\overline{Y}$ – предсказанное значение свойства Y.

В линейном случае в качестве показателей тесноты связи между угрозами используются парные коэффициенты корреляции, представляющие собой безразмерные статистические характеристики, определяемые из соотношений

$$\rho_{YX} = \frac{K_{TX}}{\sigma(X)\sigma(Y)}, \quad (13)$$

где K_{YX} – второй смешанный центральный момент (ковариация) случайных величин Y и X. Значение K_{YX} определяется по выражению

$$K_{YX} = M[(X-\overline{Y})(Y-\overline{Y})], \quad (14)$$

где $\sigma(X)$ и $\sigma(Y)$ – средние квадратические отклонения случайных величин (свойств) X и Y.

Статистические (выборочные) значения парных коэффициентов корреляции вычисляются по формулам:

$$r_{YX_i} = \frac{1}{nS(X)S(Y)} \sum_{u=1}^{n} (\chi_{iu} - \overline{X})(y_u - \overline{Y}), \quad (15)$$

между значениями X_i и Y:

$$r_{X_i X_j} = \frac{1}{nS(X_i)S(X_j)} \sum_{u=1}^{n} (\chi_{iu} - \overline{X}_i)(\chi_{ju} - \overline{X}_j), \quad (16)$$

между значениями X_i и X_j, где n – количество наблюдений; $\overline{Y}$, $\overline{X}_i$, $\overline{X}_j$ – средние значения соответствующих случайных величин (y_u, χ_{iu}, χ_{ju}) в рассматриваемой совокупности наблюдений; S(Y), S(X_i), S(X_j) – средние квадратические отклонения указанных случайных величин, определяемые по результатам наблюдений, например:

$$S^2(Y) = \frac{1}{n-1} \sum_{u=1}^{n} (y_u - \overline{Y})^2 . \quad (17)$$

На пятом этапе осуществляется оценка значимости коэффициентов регрессии. В этих целях устанавливается существенность коэффициентов регрессии по t-критерию Стьюдента для доверительной вероятности $1 - p = 0{,}95$.

На шестом этапе осуществляется оценка адекватности регрессионной модели по коэффициенту детерминации $R^2_{Y,x_1,\ldots,x_n}$, а также по коэффициенту множественной корреляции $R_{Y,x_1,\ldots,x_n}$.

На седьмом этапе вычисляются и ранжируются параметры оптимизации по коэффициенту K_{α_i}:

$$K_{\alpha_i} = \frac{t_{b_i}}{\sum_{i=1}^{n} t_{b_i}}, \quad (18)$$

где t_{b_i} – значимость коэффициентов корреляции по i-й угрозе; n – количество оцениваемых внутренних угроз.

После этого осуществляется ранжировка значений K_{α_i} и строятся диаграммы по убыванию (возрастанию) значения K_{α_i}.

На восьмом этапе производятся вывод результата на печать, обработка полученных результатов, построение диаграмм, формулировка выводов и практических рекомендаций.

Применение методики

Используя предложенную методику, проведена экспертная оценка с целью установления, какие последствия наиболее существенны при возможной реализации злоумышленником внутренних угроз. В качестве оцениваемых параметров принимались следующие последствия от воздействия внутренних угроз на КИ: кража КИ; подмена КИ; уничтожение КИ; нарушение штатной работы АИС; нарушение доступа к КИ; перехват КИ; ошибки сотрудников при обращении с КИ. Экспертам предлагалось определить степень последствий от воздействия внутренних угроз на КИ в процессе функционирования АИС путем назначения определенного балла. В качестве основного метода оценки использовался метод задания весовых коэффициентов (приписывание баллов). Выбор данного метода обусловлен тем, что оценку проводили сотрудники со стажем работы в должностях по информационной безопасности до трех лет, для которых экспертиза по указанному методу не представляла трудностей. Этим же обстоятельством обусловливается и выбор пятибалльной системы оценок (по шкале 1–5 баллов). Максимальное значение 5 баллов назначалось экспертом в том случае, если, по его мнению, последствия от воздействия внутренней угрозы оказывали на безопасность КИ существенное значение. Четыре балла назначалось экспертом, если последствия от внутренних угроз оказывали умеренное влияние на безопасность КИ. Оценка 3 балла назначалась в том случае, если указанное влияние проявлялось незначительно. Оценка 2 балла назначалась, если указанное выше влияние появлялось слабо. Минимальное значение 1 балл назначался при условии, что последствия от внутренней угрозы не оказывают влияния на безопасность КИ.

В экспертизе приняли участие 35 экспертов. В качестве объекта экспертизы рассматривались локальные ЭВМ, в которых циркулирует информация ограниченного распространения. Для обработки и анализа было отобрано 35 анкет. В связи с ограниченной выборкой вначале полученная статистическая выборка была проверена на соответствие каждой частной выборки малой выборке. По результатам такой проверки от каждой частной выборки к дальнейшей обработке было отобрано по 5 экспертных анкет. Таким образом, объем статистической выборки составил 30 реализаций.

Результаты первого этапа экспертной оценки дают основания выдвинуть гипотезу об однородности частных выборок, полученных от экспертов различных организаций. Для доказательства возможности принятия данного решения на втором и третьем этапах произведен более глубокий анализ экспертных оценок с использованием статистических критериев Бартлетта и Фишера. Используя указанные критерии, установлена однородность средних арифметических и дисперсии оценок по всем оцениваемым последствиям от внутренних угроз для всех частных выборок. На этом основании можно предположить, что рассматриваемые выборки также однородны, т.е. принадлежат одной и той же генеральной совокупности. Данное заключение позволило объединить частные выборки по оцениваемым последствиям от внутренних угроз в единую выборку и перейти к построению обобщенной статистической модели, устанавливающей связь между безопасностью КИ и оцениваемыми последствиями от воздействия внутренних угроз на КИ. В результате расчета получено следующее уравнение регрессии:

$$\mathbf{Y} = 0{,}044 + 0{,}383\mathbf{X_1} + 0{,}244\mathbf{X_2} + 0{,}208\mathbf{X_3} + 0{,}166\mathbf{X_7} + 0{,}012\mathbf{X_5} + 0{,}005\mathbf{X_6} + 0{,}006\mathbf{X_4}, \qquad (19)$$

где $\mathbf{Y}$ – безопасность КИ; $\mathbf{X_1} - \mathbf{X_7}$ – последствия от воздействия на КИ от внутренних угроз.

Оценка адекватности регрессионной модели осуществлялась следующим образом. Степень суммарного влияния семи факторов (последствий от воздействия внутренних угроз) на безопасность КИ выражается коэффициентом детерминации $R^2_{Y,x_1,\ldots,x_n} = 0{,}944$, который показывает, что 94,4% всех последствий оказывают определяющее значение на безопасность КИ. И только 5,6% последствий не вошли в круг исследований. Наибольшее влияние на безопасность КИ оказывают такие последствия от воздействия внутренних угроз, как кража КИ ($\mathbf{X_1}$), подмена КИ ($\mathbf{X_2}$), уничтожение КИ ($\mathbf{X_3}$), для которых установлена и наиболее тесная связь с безопасностью КИ ($\mathbf{Y}$), т.е. $r_{Y,X_1} = 0{,}879$; $r_{Y,X_2} = 0{,}892$; $r_{Y,X_3} = 0{,}787$. Теснота связи, характеризующаяся коэффициентом множественной корреляции $R_{Y,x_1,\ldots,x_n}$ от семи оцениваемых последствий, составляет 0,972 со среднеквадратическим отклонением от истинного значения $\sigma_R = 0{,}001$. Доказана существенность коэффициентов регрессии по *t*-критерию Стьюдента для доверительной вероятности $1 - p = 0{,}95$.

Определение относительного вклада каждого оцениваемого фактора x_i в значение безопасности КИ осуществлялось по коэффициенту K_{α_i}.

Ранжировка значений K_{α_i} по x_i представлена на рис. 2.

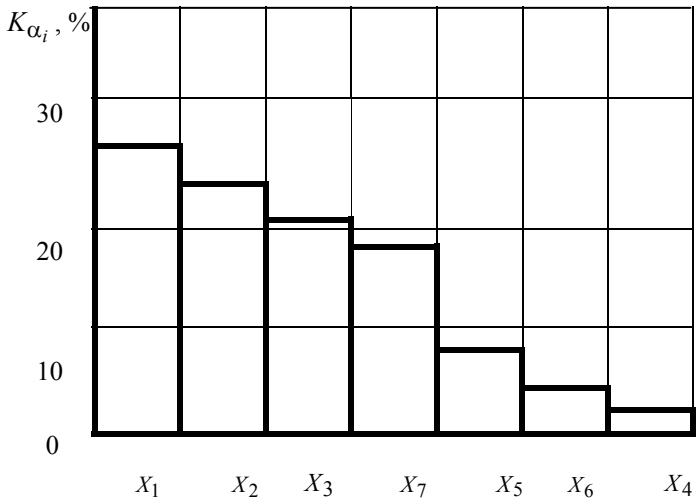

Рис. 2. Ранжировка значений K_{α_i} по x_i

Анализ рис. 2 свидетельствует о том, что наибольшее влияние на безопасность КИ оказывают такие последствия от воздействия внутренних угроз, как кража КИ, подмена КИ, уничтожение КИ и ошибки сотрудников при обращении с КИ.

Следует отметить, что полученные результаты в достаточной степени коррелируют с имеющимися статистическими оценками различных авторов по большинству оцениваемых возможных последствий от реализации внутренних угроз, а именно краже, подмене, уничтожении КИ, нарушении доступа и перехвата КИ. Исключение составляют результаты, полученные для таких последствий, которые связаны с ошибками сотрудников при обращении с КИ (их вес составляет 18%) и нарушением штатной работы АИС (их относительный вес составляет всего 2 %).

Выводы и практические рекомендации

Предложенная методика может быть использована для получения результатов и дальнейшего анализа процессов и явлений, возникающих в АИС при воздействии различных угроз, разработке организационно-профилактических мероприятий по локализации наиболее опасных угроз, а также в качестве исходных данных при проведении математического и имитационного моделирования.

Литература

1. Росенко А.П. Внутренние угрозы безопасности конфиденциальной информации: методология и теоретическое исследование. – М.: КРАСАНД, 2010. – 160 с.

2. Росенко А.П. О методе проверки данных экспертных оценок на принадлежность к малой (ограниченной) выборке // Современное состояние и приоритеты развития фундаментальных и прикладных исследований в области физики, математики и компьютерных наук: матер. 55-й Науч.-метод. конф. «Университетская наука – региону». – Ставрополь: Изд.-инф. центр «Фабула», 2010. – С. 156–157.

3. Елтаренко Е.А. Обработка экспертных решений / Е.А. Елтаренко, Е.К. Крупнова. – М.: Изд-во МИФИ, 1998. – 81 с.

Росенко Александр Петрович
Канд. техн. наук, доцент, зав. каф. компьютерной безопасности
Ставропольского государственного университета
Тел.: 8 (928) 011-78-77, 8 (865-2) 94-13-81
Эл. почта: rosenko@stavsu.ru

Rosenko A.P.

The method of processing the array of initial data obtained by an expert

The paper presents methods of processing raw data obtained by an expert, by the methods of mathematical statistics. The main idea consists in development of a structure that allows, to get arrays of input data, based on mathematical statistics methods that adequately reflect the effects from the in fluence of a variety of threats on confidential information.

Keywords: confidential information, expert opinions, threats, internal threats, the algorithm selection, and security.
