М. И. Тенетко, О. Ю. Пескова
Россия, г. Таганрог, Технологический институт ЮФУ

ПРИМЕНЕНИЕ ЛИНГВИСТИЧЕСКИХ ПЕРЕМЕННЫХ ПРИ ОЦЕНКЕ РИСКОВ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

В настоящее время одним из наиболее перспективных методов проведения аудита информационной безопасности предприятия стал метод оценивания угроз и уязвимостей безопасности, направленных на информационные активы предприятия. Для проведения аудита по этому методу необходимо выполнить следующие этапы оценки [1]:

− инвентаризация информационных ресурсов и оценка степени их значимости для успешного выполнения бизнес-процессов предприятия;

− описание архитектуры информационной системы предприятия;

− описание информационных потоков, возникающих в процессах обработки информации.

На основании данных, полученных при выполнении этих этапов, формируется следующая информация:

− перечень угроз, способных нарушить безопасность данных информационных активов;

− перечень уязвимостей в системе обеспечения информационной безопасности; способных вызвать реализацию перечисленных угроз;

− взаимосвязь между отдельной угрозой и уязвимостью (риск);

− ранжированный перечень рисков, сопоставленных с активами.

Проведённое в работе [2] исследование методов анализа и оценки угроз и уязвимостей показало, что эти методы имеют значительные ограничения. Во-первых, они требуют предоставления исчерпывающей исходной информации об объекте оценки, что практически невозможно. Во-вторых, оценки даются относительно дискретной шкалы и являются, как правило, численными. Такая схема представления данных не учитывает особенностей представления сложных знаний в человеческом мозге. В результате математическая модель объекта оценки, служащая основой для принятия решений, обедняется, упрощается и, возможно, даже искажается.

Действительно, лишь два этапа оценки позволяют рассматривать факты в приемлемые сроки и без искажений. Это описание архитектуры информационной системы, которое по определению предполагает однозначность, и описание информационных потоков, недвусмысленность которых, как правило, является важным требованием политики информационной безопасности.

Такие понятия, как степень значимости (ценности, важности) актива, угрозы и уязвимости, действительные для данной информационной системы, а также степень взаимосвязи между отдельной угрозой, уязвимостью и активом, являются в известной степени субъективными. Мы заинтересованы в таком описании перечисленных фактов, которое позволило бы оперировать субъективностью и принимать решения с эффективностью, наиболее близкой к эффективности человека-эксперта.

Для этих целей, в частности, подходит лингвистический подход, предложенный Л. Заде в работе [3]. Этот подход основывается на нахождении приближённого лингвистического описания явлений и процессов. В его основе лежит понятие лингвистической переменной. Подход не является целиком качественным; он опирается и на математические вычисления, но эти вычисления совершаются «за кулисами».

Описание лингвистической переменной

Приведём описание лингвистической переменной [3]. Лингвистическая переменная – это переменная, значениями которой являются слова или предложения естественного или искусственного языка. Например, *«Возраст»* – лингвистическая переменная, если она принимает лингвистические, а не числовые значения, то есть значения *«очень молодой»*, *«молодой»*, *«старый»*, *«не молодой и не старый»* и т. п.

Формально лингвистическая переменная описывается набором (*X*, *T* (*X*), *U*, *G*, *M*), в котором *X* – название этой переменной, *T* (*X*) – терм-множество *X*, то есть совокупность её лингвистических значений, *U* – универсальное множество, *G* — синтаксическое правило, порождающее термы множества *T* (*X*), *M* – семантическое правило, которое каждому лингвистическому значению *X* ставит в соответствие его смысл *M* (*X*), причём *M* (*X*) является нечётким подмножеством множества *U*. Напомним, что нечёткое множество представляет собой множество, в котором степень принадлежности элемента не ограничивается значениями {0, 1}, а может принимать любые значения из диапазона [0, 1].

Конкретное название *X*, порождённое синтаксическим правилом *G*, называется термом. Терм, состоящий из одного слова или нескольких слов, всегда фигурирующих вместе, называется атомарным термом. Терм, состоящий из одного или более атомарных термов, называется составным термом.

Смысл лингвистического значения *X* характеризуется функцией принадлежности □□U □ □ ı□, которая каждому элементу *u* □*U* ставит в соответствие значение принадлежности этого элемента к *X*.

Назначение семантического правила – связать принадлежности так называемых первичных термов в составном лингвистическом значении с принадлежностью составного значения. Неопределённости, такие как *«очень»*, *«вполне»*, *«чрезвычайно»* и т. п., а также союзы *«и»* и *«или»* понимаются при этом как нелинейные операторы, или модификаторы, преобразующие смысл соответствующих термов.

Совокупность значений лингвистической переменной составляет терм-множество этой переменной. Это множество может иметь, вообще говоря, бесконечное число элементов.

Однако для практических задач достаточно фиксированного конечного терм-множества; такое ограничение с точки зрения практического использования не слишком обедняет возможностей аппарата.

Считается, что количество элементов терм-множества должно быть не более семи (не считая модификаторов и их производных); это связано с особенностями восприятия человеком исследуемых объектов.

Для того чтобы лучше понять, что такое лингвистическая переменная, можно обратиться к рис.1, на котором изображена иерархическая структура связи между лингвистической переменной *«Возраст»*, нечёткими ограничениями, представляющими смысл её значений, и значениями базовой переменной *«возраст»*.

Оценивание значимости информационных активов

Рассмотрим, каким образом можно решить задачу оценивания значимости информационных активов предприятия при помощи лингвистических переменных. Допустим, мы имеем множество активов $A = \{a_1, a_2, ..., a_n\}$. Множество вполне чёткое; предприятие должно иметь чёткое представление о своих тайнах. Перед нами стоит задача определить для каждого элемента-актива a_i □*A* лингвистическую оценку его значимости.

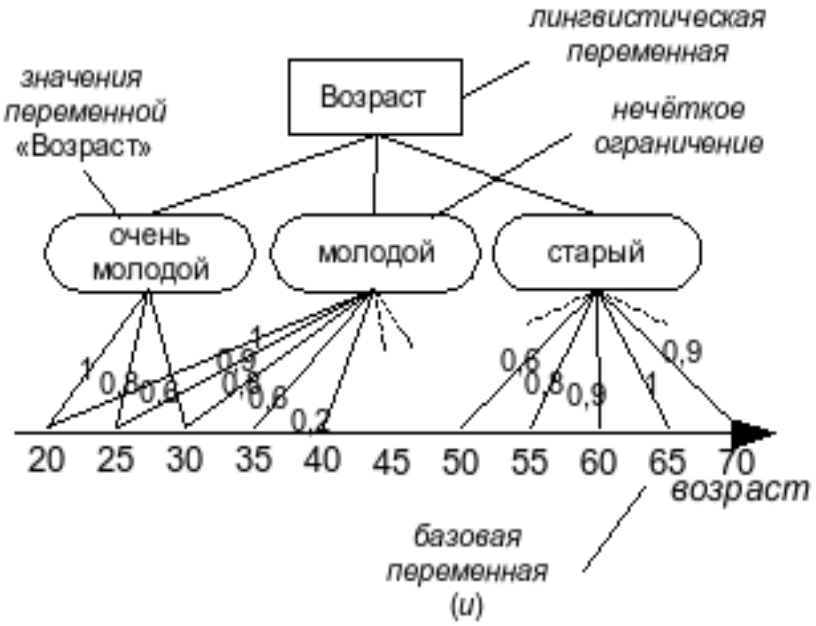

Рис. 1. Иерархическая структура лингвистической переменной

Определим лингвистическую переменную «*Значимость*» (*X*, *T* (*X*), *U*, *G*, *M*), где *T* (*X*) = {*«незначительная», «малая», «средняя», «высокая», «критическая»*}. Затем нам необходимо определить универсальное множество *U*.

Основная сложность заключается в том, что мы в данном случае не можем назвать чёткое множество *U*, элементы которого могли бы количественно охарактеризовать значимость, ценность, критичность информационных активов для бизнес-процессов компании, то есть мы не знаем, как выразить степень ценности актива в форме функции тех или иных точно измеренных величин. Шкала стоимости активов, финансовых потерь в случае нарушения безопасности активов и т. п. не может быть использована, поскольку довольно часто нет возможности определить точную сумму стоимости либо потерь.

В этом случае можно определить класс ценных активов и приписывать каждому терму *X* степень принадлежности к этому классу. Разумеется, полученные таким образом значения функции принадлежности основаны лишь на впечатлениях эксперта, которые он не в состоянии точно формализовать.

Другими словами, эксперт определяет функцию принадлежности не на множестве математически точно определённых объектов, а на множестве обозначенных некими символами впечатлений.

Такие определения имеют смысл для человека, но не для ЭВМ, однако, тем не менее, могут обрабатываться при помощи ЭВМ.

Результатом определения ценности активов является множество $\tilde{A} = \{(a_i, X_j)\}$ — множество упорядоченных пар, первым компонентом которых являются элементы $a_i \in A$, а вторым компонентом — элементы $X_j \in T(\tilde{X})$. Каждая упорядоченная пара устанавливает соответствие между активом a_i и степенью его ценности X_j, выраженной лингвистически.

Представление рисков

Рассмотрим возможность представления рисков. Риск, как правило, представляет собой определённую взаимосвязь между угрозой и уязвимостью. Определим множество угроз — $T = \{t_1, t_2, ..., t_k\}$ и множество уязвимостей — $V = \{v_1, v_2, ..., v_m\}$.

Традиционно взаимосвязь между парой объектов $x \in X$, $y \in Y$ выражается в форме упорядоченной пары (x, y). Множество, элементами которого являются упорядоченные пары, называется отношением между множеством X и множеством Y и обозначается $X \times Y$. Как правило, отношение определяется прямым (декартовым) произведением множеств X и Y: $X \times Y$.

Учитывая это, определим множество всех возможных сочетаний угрозы и уязвимости (множество рисков R), являющееся декартовым произведением множеств T и V и состоящее из упорядоченных пар (t_k, v_m): $R = T \times V = \{(t_1, v_1), (t_1, v_2), ..., (t_k, v_m)\}$. Затем определим нечёткое множество оценённых рисков:

$$\tilde{R} = \{(t_1, v_1)/\mu_{11}, (t_1, v_2)/\mu_{12}, ..., (t_k, v_m)/\mu_{km}\}.$$

Как видно, множество $\tilde{R}$ есть множество кортежей, первым компонентом которых являются элементы $t_i \in T$, вторым — элементы $v_j \in V$, а третьим — элементы $\mu_{ij} \in M$. Множество R является универсальным множеством множества $\tilde{R}$, а множество M является множеством значений функции принадлежности из диапазона [0, 1].

Поясним смысл нечёткого множества $\tilde{R}$. Значение принадлежности упорядоченной пары из множества R к множеству $\tilde{R}$ — это субъективное выражение уверенности эксперта в том, что для данной информационной системы угроза t_i реализуется через уязвимость v_j.

Речь в данном случае не идёт о вероятности реализации угрозы через уязвимость. Вероятность характеризует долю реализаций данной угрозы при функционировании информационной системы в заданный период времени. Принадлежность характеризует субъективную меру того, насколько упорядоченная пара (t_i, v_j) соответствует в представлении эксперта высказыванию *«данная угроза реализуется через данную уязвимость»*.

Иными словами, независимо от значений вероятности возникновения угроз из множества T мы оцениваем, насколько правомерно высказывание о том, что угроза t_i реализуется через уязвимость v_j. Экспертное определение и назначение вероятности возникновения угрозы связано со сбором статистических данных; экспертное определение и назначение принадлежности связано со сложившимися представлениями эксперта о характере исследуемых угроз и уязвимостей, т. е. с его персональным опытом.

Разумеется, на накопление персонального опыта могут влиять в том числе и выявленные экспертом в своё время статистические закономерности. Однако человек-эксперт практически никогда не занимается непосредственным выявлением и сопоставлением статистических закономерностей, вычислением значений вероятности и т. п. Знания, соответствующие его опыту, формируются, вообще говоря, в виде образов, не имеющих чётких границ.

В определённой степени можно говорить о том, что эти образы — своего рода метаданные; образы на качественном уровне представляют абсолютно все количественные результаты экспериментов, служащие основой для формирования персонального опыта, в том числе и статистические наблюдения. В дальнейшем человек-эксперт при принятии решения обращается именно к образам, таким как *«ве-*

роятность возникновения этого события очень велика» (лингвистически заданное значение вероятности), *«скорее всего, между этими двумя объектами существует взаимосвязь»* (лингвистически заданное бинарное отношение), *«из этих двух объектов первый существенно значимее второго»* (лингвистически заданное отношение предпочтения) и т. п.

По этой причине описанная модель представления взаимосвязи между угрозой и уязвимостью может оказаться эффективнее модели, основанной на вероятностях событий. Модель позволяет более точно описывать процессы, нарушающие безопасность активов, без потери времени, связанного со сбором фактов.

Можно слегка видоизменить модель для того, чтобы избежать использования численных оценок принадлежности.

Зададим лингвистическую переменную *«Взаимосвязь между угрозой и уязвимостью»* с терм-множеством $T(Y) = \{$*«незначительное», «низкое», «среднее», «значительное», «высокое»*$\}$ и универсальным множеством U, определённым как класс однозначно взаимосвязанных угроз и уязвимостей. Кортеж $\{(t_k, v_m)/\square_{km}\}$, лежащий в основе множества $\tilde{R}$ (см. формулу №1), в таком случае заменяется на кортеж $\{(t_k, v_m)/Y_j\}$, где Y_j □□ (Y) — значение лингвистической переменной *«Взаимосвязь между угрозой и уязвимостью»*.

Однако, с другой стороны, численные оценки принадлежности позволяют наглядно показать взаимосвязь между угрозами и уязвимостями. Любое множество, элементами которого являются отношения, в том числе и нечёткое, может быть представлено в виде двудольного графа [4]. Нечёткое отношение (t_k, v_m) легко сводится к матрице инциденций следующего вида (табл.1).

Таблица 1

Матрица инциденций нечёткого отношения (t_k, v_m)

	v_1	v_2	...	v_m
t_1	□₁₁	□₁□	□̃₁	□₁m
t_2	□□₁	□□□	□□̃	□2m
...	□̃₁	□̃□	□̃	□̃m
t_k	□k1	□k2	□̃k	□km

Двудольный граф нечёткого отношения приведён на рис. 2.

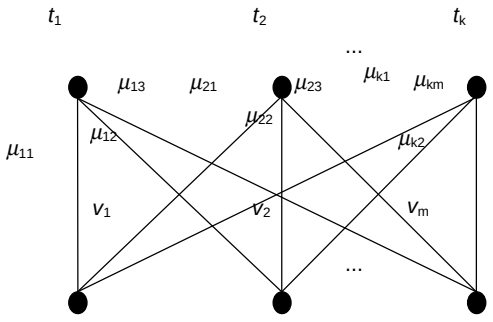

Рис. 2. Двудольный граф нечёткого отношения (t_k, v_m)

Известно, что □ij принимает значения из непрерывного диапазона [0, 1]. Это значит, что для наглядного графического изображения связи между угрозами и уязвимостями можно использовать следующие приёмы:

− раскраска рёбер графа цветом, интенсивность которого линейно связана со значением $\square_{ij}$;

− обозначение рёбер графа линиями, толщина которых линейно связана со значением $\square_{ij}$;

− задание непрерывного диапазона цветов (например, диапазон цветов радуги: от красного до фиолетового) и линейное сопоставление с ним значений $\square_{ij}$.

Эти приёмы не играют особой роли в процессе принятия решения, однако могут облегчить понимание неспециалистами процесса аудита.

Сопоставление риска и актива

Как правило, риски не рассматриваются сами по себе; они имеют значение, поскольку направлены на защищаемые активы. Чтобы сопоставить риск и актив, расширим кортеж $\{(t_k, v_m)/\square_{km}\}$, описанный в формуле №1, до вида $\{((t_k, v_m)/\square_{km}), (a_n, X_i)\}$. Как видно, полученный кортеж представляет собой бинарное отношение двух других кортежей.

Первый из них в свою очередь является нечётким бинарным отношением между угрозой и уязвимостью. Второй является обычным бинарным отношением между активом и степенью его ценности.

Такое представление учитывает одновременно ценность актива и степень связанности угрозы с уязвимостью, что даёт возможность впоследствии оценить величину риска нарушения безопасности данного актива.

Заключение

Предложенная модель оценивания информационных активов и описания рисков позволяет максимально эффективно использовать персональный опыт эксперта при проведении аудита информационной безопасности.

Она опирается на концепцию лингвистических метаданных, нечётких образов, описывающих в разуме человека персональный опыт. Вместе с тем модель позволяет обрабатывать данные с помощью ЭВМ.

Основное достоинство такого подхода — сведение к минимуму затрат на исследование информационной системы с целью выявления статистических закономерностей и обнаружения рисков, а также адекватная оценка информационных активов предприятия.

В заключение следует сказать, что вопросы формирования терм-множества лингвистической переменной, сопоставления термов с функциями принадлежности и применения лингвистических модификаторов заслуживают отдельного исследования.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Петренко С. А., Петренко А. А.* Аудит безопасности Intranet. – М.: ДМК Пресс, 2002. – 416 с.

2. *Тенетко, М. И., Пескова, О. Ю.* Обзор методов анализа и оценки информационных рисков. — http://www.security.ase.md/publ/ru/pubru101/21.pdf .

3. *Заде Л.* Понятие лингвистической переменной и его применение к принятию приближённых решений. – М.: Мир, 1976. – 163 с.

4. *Кофман А.* Введение в теорию нечётких множеств. – М. Радио и связь, 1982. – 432 с.