М. В. Тимонин, В. С. Лаврентьев

ИСПОЛЬЗОВАНИЕ ТЕОРИИ НЕЧЕТКОЙ МЕРЫ ДЛЯ АГРЕГАЦИИ СОСТАВЛЯЮЩИХ РИСКА ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

Риск, связанный с информационной безопасностью организации, является многомерным сложным понятием, включающим множество связанных друг с другом переменных. Основой построения модели риска является его декомпозиция на логические составляющие, представляющие более мелкие области проблемы, такие как, например, «безопасность рабочих станций» или «безопасность данных в системе резервного копирования», которые, в свою очередь, разделяются на еще более мелкие компоненты до тех пор, пока оценка элемента не сведется к тривиальному вопросу. Следующим шагом является ответ на эти вопросы, т. е. оценка составляющих на основе информации об исследуемой проблеме, распространение информации снизу вверх и расчет интересующего нас кумулятивного значения, т. е. величины риска.

Несмотря на то что традиционным определением риска является произведение вероятности негативного события на ущерб, в области информационной безопасности (ИБ) на данный момент такой подход имеет ряд проблем, по крайней мере, если рассматривать вероятности в классическом, частотном смысле. Существует достаточно много проблем, препятствующих проведению точных количественных оценок. Прежде всего, недостаточность данных — статистики по взломам и атакам практически нет, особенно такой, которая ответила бы на вопрос: насколько сильно мои данные подвержены опасности?

Проблема усиливается тем, что потенциальный источник атак — не стохастический генератор, подчиняющийся только случайному распределению, а зачастую интеллектуальный агент, т. е. человек, действующий рационально и, главное, направленно. Таким образом, даже имея некоторую частотную характеристику распределения типа атак, использовать лишь ее для оценки риска ИБ не имеет большого смысла, потому что обеспечение защиты от наиболее часто встречающихся атак не гарантирует безопасности данных.

Подобные размышления приводят к тому, что оценивать следует не вероятность потенциальных происшествий, а их реализуемость с учетом введенных мер, иными словами, уровень защищенности организации. Такой подход позволяет максимально утилизировать использование информации: организация, как правило, располагает данными об устройстве собственной системы ИБ и целях осуществления защиты, существуют стандарты, предоставляющие рекомендации по ее построению (ГОСТ, ISO/BS, NIST), в редких случаях присутствуют даже некоторые данные по инцидентам, произошедшим в организации в прошлые годы.

Таким образом, проблема из категории «расчет вероятности» может быть переведена в категорию «агрегация данных». Однако возникает вопрос о том, какой же оператор наиболее уместно применять для агрегации значений. Для иллюстрации механизмов процесса рассмотрим элемент разложения риска, изображенный на рис. 1. Какими данными располагает аудитор в данном примере? Имеется список компонентов, входящих в более общее понятие, существует некое представление об относительной важности компонентов, о том, как они взаимодействуют — дополняют ли друг друга или являются взаимозаменяемыми (как, например, два последних компонента в примере), какие компоненты необходимы, какие достаточны и т. д. Наконец, возможно провести оценку того, насколько полно компоненты реализованы, правильно работают.

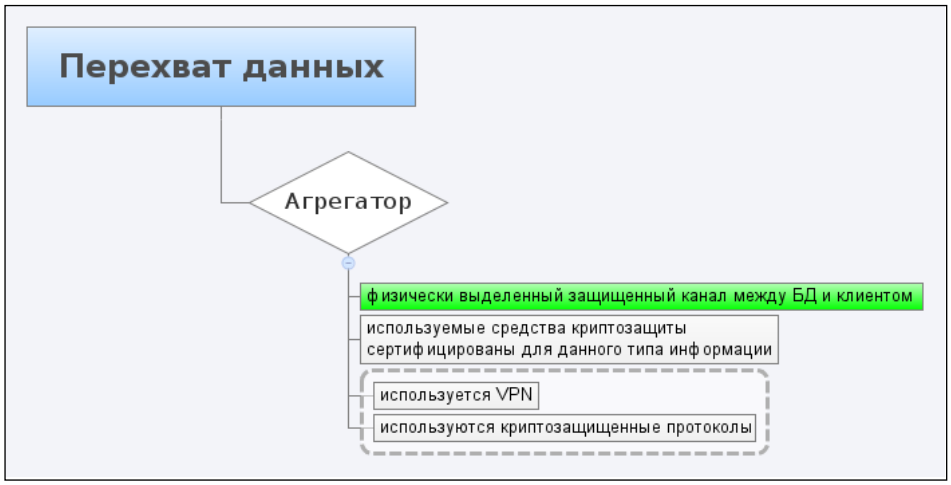

Рис. 1. Элемент диаграммы риска

Необходимо привести все критерии на общую шкалу, объединить и потом произвести обратное преобразование. Т. е. в общем виде:

$\rho = F^{-1}(H(F(a_1, a_2, .., a_n)))$,

где a1,....,an — компоненты, а ρ — агрегированное значение.

Таким образом, необходим аксиоматический подход — сформировать список требований к оператору и использовать соответствующий математический аппарат.

Оператор, применимый для агрегации данных в проблеме исследования риска информационной безопасности, должен, несомненно, обладать рядом свойств, определяемых решаемой задачей. Список свойств приведен ниже. Вслед за Грабичем [1] свойства разделены на математические и семантические, хотя первые являются лишь выражением некоторых других семантических свойств, которые более удобно описывать в математическом виде. Свойства описывают спектр применимости оператора при моделировании риска.

Семантические свойства

1. Возможность задания значимости агрегируемых компонентов при наличии такой необходимости.

2. Возможность выражения характера агрегации, в частности:

2a. Конъюнктивно или дизъюнктивно направленная (экстремумы min и max);

2b. Какие критерии являются необходимыми (ветирование);

2c. Какие критерии являются достаточными.

3. Возможность выражения взаимодействия между критериями:

3a. Взаимозаменяемость;

3b. Поддержка.

4. Возможность ясной семантической интерпретации оператора.

Математические свойства

5. Идемпотентность $H(a, a, ..., a) = a, \forall a$.

6. Непрерывность H является непрерывной функцией $a_1, .., a_n$.

7. Монотонность (неубываемость) $a'_i > a_i \Rightarrow H(a_1, a'_i, ..., a_n) \geq H(a_1, a_i, ..., a_n) \forall a'_i, a_i \in A$.

8. Компенсируемость $min(a_i) \leq H(a_1, .., a_n) \leq max(a_i)$.

Проанализировав вышеуказанные свойства и различные операторы, встречающиеся в литературе, в частности сочетания AND/OR, взвешенное среднее, OWA и порождаемые им xOWA[2], xOWG, взвешенные минимум/максимум [3], авторами было принято решение использовать для построения модели интеграл Шоке [4].

Обозначим как $X=\{x_1, x_{2,}.., x_n\}$ множество критериев, а как 2^X — множество всех подмножеств множества X.

Определение 1.1: Нечеткой мерой [5] (или емкостью [4]) на множестве X называется функция $м: 2^X \to [0,1]$, удовлетворяющая следующим свойствам:

1. $m(\varnothing)=0$, $m(X)=1$,

2. $A \subset B \Rightarrow m(A) \le m(B), \forall A, B \in 2^X$.

$m(A)$ может рассматриваться как значимость критерия A. Таким образом, в дополнение к обычным весам мы получаем возможность определять значимость групп критериев.

Нечеткая мера называется *аддитивной*, если $м(A \cup B)= м(A)+ м(B), A \cap B = \varnothing$, *субаддитивной*, если $м(A \cup B)\le м(A)+ м(B), A \cap B = \varnothing$ и *супераддитивной*, если $м(A \cup B)\ge м(A)+ м(B), A \cap B = \varnothing$.

Определение 1.2: Дискретным интегралом Шоке элементов $a_{1,}.., a_n$ для меры *м* называется

$$C_м(a_{1,}.,a_n)= \sum_{i=1}^{n}(a_{(i)} - a_{(i-1)})м(A_{(i)}),$$ где A(i)= $\{x_{(i)},..., x_{(n)}\}$

$a_{(1)},..., a_{(n)}$ перестановка $a_{1,}.., a_n$, такая что $a_{(1)} \le a_{(2)} \le ... \le a_{(n)}$ и $a_{(0)}= 0$.

В случае, когда мера *м* аддитивна, интеграл сокращается до взвешенного среднего

$$C_м(a_{1,}.,a_n)= \sum_{i=1}^{n} a_i м(\{i\}).$$

Теперь проанализируем требования к оператору на примере и покажем, как реализовать различные сочетания критериев с помощью интеграла Шоке.

1. Возможность задания значимости агрегируемых компонентов при наличии такой необходимости — операция идентична той, которая производится при использовании взвешенного среднего. На данный момент при построении модели веса компонентов будут выставляться экспертом на основе знаний о системе.

2. Возможность выражения характера агрегации.

2a. Конъюнктивно или дизъюнктивно направленная (экстремумы min и max).

Строго конъюнктивная агрегация (AND) характеризуется следующим образом

$$\begin{matrix} м(a_1,..., a_n)=1 \\ м(B) = 0, \forall B \ne \{a_1,..., a_n\} \end{matrix}.$$

Соответственно, строго дизъюнктивная (OR)

$$\begin{matrix} м(\varnothing) = 0 \\ м(B) = 1, \forall B \subset 2^X \end{matrix}.$$

Существует также индекс, показывающий склонность нечеткой меры в сторону дизъюнктивной или конъюнктивной агрегации [6]

$$orness(C_м)= \sum_{T \subseteq N\, t \in (0\ n)} \frac{1}{(n-1)\binom{n}{t}} м(T)= \sum_{T \subseteq N} \frac{n-t}{(n-1)(t+1)} m^м(T).$$

2b. Какие критерии являются необходимыми (ветирование).

Необходимость критерия x_j фактически означает следующую декомпозицию агрегации [9]:

$$H(a_{1,}.,a_j,...,a_n)= a_j \wedge G(a_{1,}.,a_{j-1},a_{j+1},...,a_n).$$

С помощью интеграла Шоке это моделируется следующим образом:

$$м(A) = 0, \forall A \subset X \setminus \{x_j\}.$$

2c. Какие критерии являются достаточными.

Критерий x_j является достаточным, если агрегацию можно представить в следующем виде [9]:

$$H(a_{1,}.,a_j,...,a_n)= a_j \vee G(a_{1,}.,a_{j-1},a_{j+1},...,a_n).$$

Или с помощью интеграла Шоке: $м(A) = 1, \forall A$ э $\{x_j\}$.

3. Веса групп критериев — более сильная с семантической точки зрения операция позволяет выражать комплиментарность или заменяемость, иными словами, свойства критериев усиливать

значимость друг друга или быть взаимозаменяемыми. Выражение подобных взаимодействий становится возможным благодаря отсутствию свойства аддитивности.

С помощью нечеткой меры комплиментарность моделируется следующим образом:

$$м(ij) > м(i) + м(j),$$

что означает, что совместный вес критериев i и j больше, чем сумма индивидуальных весов, т. е. компоненты усиливают друг друга. Соответственно, взаимозаменяемость моделируется как

$$м(ij) < м(i) + м(j),$$

что означает, что информация, содержащаяся в критериях, частично перекрывается.

В общем случае для описания неаддитивной нечеткой меры $м$ требуется задание 2^n-2 коэффициентов, что существенно усложняет процесс моделирования. Однако, как будет показано далее, существует несколько путей снижения данного числа. Стоит упомянуть, что каждый достаточный или необходимый критерий уменьшает количество определяемых коэффициентов в 2 раза.

Кроме того, интеграл Шоке идемпотентен, непрерывен, монотонен (при условии монотонности нечеткой меры $м$) и является компенсирующим.

Неаддитивность меры приводит к семантической гибкости, однако в то же время затрудняет определение величины вклада каждой переменной в общее значение, поскольку оно будет зависеть не только от непосредственного веса переменной (т. е. значения нечеткой меры), но и от всех комбинаций, в которые данная мера входит. Для определения относительной меры каждого компонента в общей сумме используется индекс Шепли [7].

Определение 1.3 [8]: Индексом взаимодействия $м$, обозначаемым $I^м$, называется следующее значение:

$$I^м(A) = \sum_{B \subseteq X \setminus A} \frac{(n-b-a)!b!}{(n-a+1)!} \sum_{K \subseteq A} (-1)^{|A \setminus K|} м(B \cup K), \forall A \subseteq X .$$

Индекс, рассчитанный для любых двух величин, демонстрирует характер взаимодействия между ними. Если значение положительное, критерии усиливают друг друга, а если отрицательное — являются взаимозаменяемыми. Индекс взаимодействия для единичных элементов называется величиной Шепли (или индексом Шепли).

$$u_i(м) = I^м(i) = \sum_{A \subseteq N \setminus i} \frac{(n-a-1)!a!}{n!} [м(A \cup i) - м(A)] .$$

Индекс Шепли выражает значимость критерия в агрегированном значении. Сумма значений индекса Шепли для всех критериев всегда равняется единице.

Выводы

Использование теории нечетких множеств и интеграла Шоке в качестве оператора агрегации, несомненно, имеет большой потенциал при решении проблемы моделирования риска информационной безопасности. Сочетание семантической выразительности с математическими свойствами, демонстрируемыми интегралом Шоке, приводят к возможности более точного решения задач данного класса. В дальнейшем будет продемонстрирован вариант практического применения и проведено сравнение с вероятностным подходом.

СПИСОК ЛИТЕРАТУРЫ:

1. *Grabisch M., Orlovski S. A. and Yager R. R.* Fuzzy aggregation of numerical preferences // The Handbook of Fuzzy Sets Series. Vol. 4: Fuzzy Sets in Decision Analysis, Operations Research and Statistics. R. Slowinski (ed.). Kluwer Academic, 1998. P. 31–68.
2. *Yager R. R. and Filev D. P.* Induced ordered weighted averaging operators // IEEE Transaction on Systems. Man and Cybernetics 29. 1999. P. 141–150.
3. *Dubois D. and Prade H.* Weighted minimum and maximum operations in fuzzy set theory // Information Sciences. 1986. № 39. P. 205–210.

4. *Choquet* G. Theory of capacities // Annales de l'Institut Fourier. 1953. № 5. P. 131–295.
5. Sugeno M. Theory of fuzzy integrals and its applications. PhD thesis. Tokyo Institute of Technology, 1974.
6. *Marichal J.-L.* Tolerant or intolerant character of interacting criteria in aggregation by the Choquet integral // Eur. Journal of Operational Research. 2004. № 155(3). P. 771–791.
7. *Murofushi T.* A technique for reading fuzzy measures (I): the Shapley value with respect to a fuzzy measure // 2nd Fuzzy Workshop. Nagaoka, Japan. October 1992. In Japanese. P. 39–48.
8. *Grabisch M.* k-order additive fuzzy measures // The 6th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU). Granada, Spain, 1996. P. 1345–1350.
9. *Grabisch M.* Alternative representations of discrete fuzzy measures for decision making // The 4th Int. Conf. on Soft Computing. Iizuka, Japan.