УДК 004.4

Г. Э. Яхъяева, **О. В. Ясинская**

Новосибирский государственный университет
ул. Пирогова, 2, Новосибирск, 630090, Россия

E-mail: gulnara@math.nsc.ru; yasinskaya.olga@gmail.com

МЕТОДЫ СОГЛАСОВАНИЯ ЗНАНИЙ ПО КОМПЬЮТЕРНОЙ БЕЗОПАСНОСТИ, ИЗВЛЕЧЕННЫХ ИЗ РАЗЛИЧНЫХ ДОКУМЕНТОВ *

Рассматривается проблема согласования знаний по компьютерной безопасности, извлеченных из разных текстов на естественном языке. Дается описание поставленной задачи с помощью теоретико-модельного формализма. Знание о конкретной компьютерной атаке формализуется в виде недоопределенной алгебраической системы (названной обобщенным прецедентом). База знаний представляет собой множество обобщенных прецедентов. Согласованное значение истинности предложения вычисляется в виде интервала, определенного на отрезке $[0,1]$. Приводятся алгоритмы вычисления согласованного значения истинности, описывается программная реализация разработанных методов.

Ключевые слова: информационная безопасность, компьютерная атака, прецедент компьютерной атаки, прецедентная модель, обобщенная нечеткая модель, обобщенный прецедент.

Введение

Современный этап развития общества характеризуется возрастающей ролью информационной сферы. Безопасность рядовых пользователей, промышленных предприятий и корпораций и всего государства в целом существенным образом зависит от обеспечения информационной безопасности. В дальнейшем, в ходе технического прогресса важность информационной безопасности будет только возрастать [1].

Сейчас основная задача специалистов, обеспечивающих компьютерную безопасность, – это оперативная реакция на изменения текущего статуса защищенности всех компонент систем и своевременное обнаружение изменения этого статуса. Для этой цели удобно иметь программную систему, позволяющую без необходимости приобретения особых навыков оперативно определить тип компьютерной атаки, узнать самую свежую информацию о возможных последствиях компьютерных атак и способах их предотвращения.

При проектировании интеллектуальных систем первоочередной проблемой является задача представления знаний и дальнейшая их обработка [2]. Компьютерная программа, чтобы быть эффективной для заданной предметной области, должна располагать знанием о данной предметной области, представленном в используемом этой программой формализме. Задача представления знаний состоит главным образом в выявлении того, каковы наиболее адекват-

* Исследование выполнено при поддержке Министерства образования и науки Российской Федерации (соглашение № 14.B37.21.0400 «Методы извлечения и порождения знаний для обеспечения информационной безопасности»).

Яхъяева Г. Э., Ясинская О. В. Методы согласования знаний по компьютерной безопасности, извлеченных из различных документов // Вестн. Новосиб. гос. ун-та. Серия: Информационные технологии. 2013. Т. 11, вып. 3. С. 63–73.

ISSN 1818-7900. Вестник НГУ. Серия: Информационные технологии. 2013. Том 11, выпуск 3

ные формализмы для представления знаний и каковы наиболее эффективные методы манипулирования этими знаниями.

Особенно остро эта проблема стоит, когда речь идет о знаниях, относящихся к области информационной безопасности и компьютерных угроз. Это связано с тем, что в этой области ценность информации в гораздо большей степени зависит от ее новизны, чем в большинстве других областей науки и технологии. Для эффективной защиты от компьютерной угрозы необходимо как можно раньше узнать о ее появлении. Один из наиболее актуальных источников такой информации – тексты на естественном языке, представленные в сети Интернет.

Одной из методологий обработки знаний, извлеченных из текстов на естественном языке, является теоретико-модельный подход представления знаний. Он основан на разработанном теоретико-модельном подходе к формализации онтологий предметных областей [3; 4]. В рамках предлагаемого подхода знания, извлекаемые из текстов, написанных на естественном языке, представляются в виде алгебраических систем (*прецедентов* предметной области). На основе прецедентов строится *прецедентная модель* предметной области. Значением истинности предложения на прецедентной модели является набор тех прецедентов, для которых это предложение является истинным в точном смысле. В результате фазификации прецедентной модели получается *нечеткая модель*, в которой значениями истинности предложений являются числа из интервала [0, 1]. Фазифицируя некоторое множество прецедентных моделей как единое целое мы получаем *обобщенную нечеткую модель*. Формальное (теоретико-модельное) описание всех этих моделей можно найти в работах [5; 6].

Одним из путей порождения новых знаний при помощи текстов естественного языка является сравнение и интеграция знаний, содержащихся в разных текстах [7]. В процессе извлечения знаний из текстов на естественном языке строятся различные обобщенные нечеткие модели, формализующие извлеченное знание. И, таким образом, возникает необходимость в *согласовании* различных алгебраических систем. Данная работа посвящена решению вопроса о согласовании знаний, полученных из разных источников для предметной области информационной безопасности.

Принцип согласования обобщенных нечетких моделей

Одной из интерпретаций обобщенной нечеткой модели может быть следующая. Пусть есть некоторый эксперт по предметной области, описываемой языком σ_A. Например, таким экспертом может быть системный администратор предприятия, предметной областью – компьютерная безопасность. Эксперт имеет дело с некоторым множеством ситуаций – прецедентов данной предметной области (например, с некоторым множеством кибератак). Это множество прецедентов можно рассматривать как вероятностное пространство. Элементарными исходами этого вероятностного пространства будут прецеденты. Естественно, эксперт не знает полного описания каждой из указанных прецедентов и тем более не знает истинностного значения всех предложений сигнатуры σ_A на каждом из этих прецедентов. Тем не менее эксперт, исходя из известной ему информации, может давать вероятностные оценки истинности интересующих нас предложений. Например, эксперт может утверждать «70 % компьютерных атак используют DoS-атаки» или «не менее 60 % компьютерных атак направлено на кражу информации». В работах [6; 8] показано, как такие вероятностные знания эксперта можно формализовать с помощью обобщенной нечеткой модели.

А теперь предположим, что у нас имеется не один, а несколько экспертов в данной предметной области. Каждый из экспертов обладает своими уникальными знаниями о предметной области. Очевидно, что эти знания, в общем случае, не будут совпадать (так как разные эксперты могли получать свои знания из разных источников, могли принадлежать различным научным школам и т. п.). Однако, принимая решение, мы бы хотели учитывать мнения всех экспертов, т. е. находить компромиссное решение.

На формальном языке эту проблему можно описать следующим образом. Пусть рассматриваемая предметная область описывается сигнатурой σ_A, где A – множество индивидуу-

мов (основное множество) данной предметной области. Для описания предметной области построено конечное число обобщенных нечетких моделей

$$\left\{\mathfrak{A}_{K_i} = \left\langle A, \sigma_A, \xi_i \right\rangle \mid i = 1, ..., n\right\},$$

где K_i – множество прецедентных моделей, порождающих модель $\mathfrak{A}_{K_i}$, и ξ_i – означивание всех предложений сигнатуры σ_A на модел $\mathfrak{A}_{K_i}$. Заметим, что значениями истинности предложений на обобщенной нечеткой модели являются различные подмножества интервала [0, 1].

Тогда *проблему согласования* конечного числа моделей $\mathfrak{A}_{K_1}, ..., \mathfrak{A}_{K_n}$ можно сформулировать следующим образом: описание процедуры (алгоритма), позволяющего для любого $\varphi \in S(\sigma_A)$, исходя из значений истинности $\xi_1(\varphi), ..., \xi_n(\varphi)$ этого предложения на моделях $\mathfrak{A}_{K_1}, ..., \mathfrak{A}_{K_n}$, построить согласованное значение истинности $Tr(\varphi) \subseteq [0, 1]$.

Данную проблему можно было бы формализовать как построение *n*-местной функции $f : (\rho([0,1]))^n \to \rho([0,1])$. Однако при построении этой функции нужно учитывать, что согласованные значения истинности для разных предложений не должны противоречить друг другу. Например, странно бы выглядело, если бы наш принцип согласования выдавал бы $Tr(\varphi) = Tr(\neg\varphi) = 1$.

Таким образом, более разумно сформулировать принцип согласования *n* обобщенных нечетких моделей как *n*-местную функцию f, определенную на множестве всех обобщенных нечетких моделей, т. е.

$$f : \left\langle \mathfrak{A}_{K_1}, ..., \mathfrak{A}_{K_n} \right\rangle \mapsto \mathfrak{A}_K.$$

Более того, хотелось бы, чтобы данный принцип согласования работал на любом конечном множестве моделей и не зависел от порядка рассмотрения моделей. А эти свойства, как известно, достигаются при помощи свойств ассоциативности и коммутативности.

Пересечение и объединение обобщенных нечетких моделей. Как было отмечено выше, каждая обобщенная нечеткая модель $\mathfrak{A}_K$ однозначно определяется множеством прецедентных моделей K. Таким образом, определяя функцию согласования мы, по существу, должны задавать операции на множествах прецедентных моделей. Для начала рассмотрим теоретико-множественные операции пересечения и объединения.

Определение 1. *Пусть* $\mathfrak{A}_{K_1}$ *и* $\mathfrak{A}_{K_2}$ *– обобщенные нечеткие модели, порожденные классами прецедентных моделей* K_1 и K_2 *соответственно. Модель* $\mathfrak{A}_{K_1 \cap K_2}$ *называется* ***пересечением*** *моделей* $\mathfrak{A}_{K_1}$ *и* $\mathfrak{A}_{K_2}$.

Замечание 1. $\mathfrak{A}_{K_1 \cap ... \cap K_n} = \mathfrak{A}_{(K_1 \cap ... \cap K_{n-1}) \cap K_n}$.

Отметим, что операция пересечения является частичной, т. е. не все модели согласуются при помощи данной операции.

Предложение 1. *Пусть* $\mathfrak{A}_{K_1}, ..., \mathfrak{A}_{K_n}$ *– обобщенные нечеткие модели, порожденные классами прецедентных моделей* $K_1, ..., K_n$ *соответственно. Тогда для любого* $\varphi \in S(\sigma_A)$

$$\bigcap_{i=1}^{n} K_i \neq \varnothing \Rightarrow \xi_{K_1 \cap ... \cap K_n}(\varphi) \subseteq \bigcap_{i=1}^{n} \xi_{K_i}(\varphi) \neq \varnothing.$$

Доказательство. Пусть $\bigcap_{i=1}^{n} K_i \neq \varnothing$. Тогда для любого $q \in \xi_{K_1 \cap ... \cap K_n}(\varphi)$ найдется такая прецедентная модель $\mathfrak{A}_E$, что $\mathfrak{A}_E \in K_1 \bigcap ... \bigcap K_n$ и $Fuz(\mathfrak{A}_E) \models_q \varphi$. Очевидно, что $\mathfrak{A}_E \in K_i$ $(i = 1, ..., n)$. Следовательно, $q \in \xi_{K_i}(\varphi)$ $(i = 1, ..., n)$, т. е. $q \in \bigcap_{i=1}^{n} \xi_{K_i}(\varphi)$.

■

Определение 2. *Пусть* $\mathfrak{A}_{K_1}$ *и* $\mathfrak{A}_{K_2}$ *– обобщенные нечеткие модели, порожденные классами прецедентных моделей* K_1 и K_2 *соответственно. Модель* $\mathfrak{A}_{K_1 \cup K_2}$ *называется* ***объединением*** *моделей* $\mathfrak{A}_{K_1}$ *и* $\mathfrak{A}_{K_2}$.

Замечание 2. $\mathfrak{A}_{K_1 \cup \ldots \cup K_n} = \mathfrak{A}_{(K_1 \cup \ldots \cup K_{n-1}) \cup K_n}$.

Предложение 2. *Пусть* $\mathfrak{A}_{K_1}, \ldots, \mathfrak{A}_{K_n}$ – *обобщенные нечеткие модели, порожденные классами прецедентных моделей* $K_1, \ldots, K_n$ *соответственно. Тогда для любого* $\varphi \in S(\sigma_A)$

$$\xi_{K_1 \cup \ldots \cup K_n}(\varphi) = \bigcup_{i=1}^{n} \xi_{K_i}(\varphi).$$

Доказательство. Рассмотрим $q \in \xi_{K_1 \cup \ldots \cup K_n}(\varphi)$. Найдется такая прецедентная модель $\mathfrak{A}_E$, что $\mathfrak{A}_E \in K_1 \cup \ldots \cup K_n$ и $Fuz(\mathfrak{A}_E) \models_q \varphi$. Тогда найдется такое i $(i \in \{1, \ldots, n\})$, что $\mathfrak{A}_E \in K_i$, т. е. $q \in \xi_{K_i}(\varphi)$. Следовательно, $q \in \bigcup_{i=1}^{n} \xi_{K_i}(\varphi)$.

Возьмем теперь $q \in \bigcup_{i=1}^{n} \xi_{K_i}(\varphi)$. Тогда найдется такое i $(i \in \{1, \ldots, n\})$, что $q \in \xi_{K_i}(\varphi)$. Следовательно, существует такая прецедентная модель $\mathfrak{A}_E$, что $\mathfrak{A}_E \in K_i$, и $Fuz(\mathfrak{A}_E) \models_q \varphi$. Таким образом, получим, что $\mathfrak{A}_E \in K_1 \cup \ldots \cup K_n$ и $q \in \xi_{K_1 \cup \ldots \cup K_n}(\varphi)$.

■

На практике обычно используются вероятностные оценки событий, являющиеся либо числами, либо интервалами из множества $[0,1] \cap \mathbb{Q}$. Таким образом, интересно рассматривать не произвольные обобщенные модели, а интервальные модели.

Определение 3. *Обобщенная нечеткая модель* $\mathfrak{A}_K$ *называется* ***интервальной моделью****, если для любого* $\varphi \in S(\sigma_A)$ *значение истинности* $\xi_K(\varphi)$ *является интервалом на множестве* $[0,1] \cap \mathbb{Q}$.

Заметим, что класс интервальных моделей не замкнут относительно операции объединения. Таким образом, операция объединения не является подходящим формализмом для описания поставленной задачи.

С другой стороны, хотя класс интервальных моделей замкнут относительно операции пересечения, данная операция является частичной. И, следовательно, тоже не подходит для формализации согласования моделей.

Произведение обобщенных нечетких моделей.

Определение 4. *Рассмотрим множества прецедентов* E_1 *и* E_2. *Будем говорить, что* E_1 *и* E_2 ***изоморфны*** $(E_1 \cong E_2)$, *если существует такое взаимно однозначное отображение* $f : E_1 \to E_2$, *что* $\mathfrak{A} \cong f(\mathfrak{A})$ *для любого* $\mathfrak{A} \in E_1$.

Определение 5. *Рассмотрим прецедентные модели* $\mathfrak{A}_{E_1}$ *и* $\mathfrak{A}_{E_2}$. *Модель* $\mathfrak{A}_E$ *назовем* ***произведением*** *моделей* $\mathfrak{A}_{E_1}$ *и* $\mathfrak{A}_{E_2}$ *и обозначим* $\mathfrak{A}_E = \mathfrak{A}_{E_1} * \mathfrak{A}_{E_2}$, *если*:

1) $E = E_1 \cup E_2^*$, *где* $E_2^* \cong E_2$ *и* $E_2^* \cap E_1 = \varnothing$;

2) *Для любого* $\varphi \in S(\sigma_A)$ *имеем* $\tau_E(\varphi) = \tau_{E_1}(\varphi) \cup \tau_{E_2}(\varphi)$.

В работе [6] было доказано, что операция * является ассоциативной, коммутативной и замкнутой на множестве всех прецедентных моделей.

Предложение 3. *Рассмотрим прецедентные модели* $\mathfrak{A}_{E_1}$, $\mathfrak{A}_{E_2}$ *и* $\mathfrak{A}_E$ *такие, что* $\mathfrak{A}_E = \mathfrak{A}_{E_1} * \mathfrak{A}_{E_2}$. *Пусть нечеткие модели* $\mathfrak{A}_{\mu_1}$, $\mathfrak{A}_{\mu_2}$ *и* $\mathfrak{A}_\mu$ *являются фазификациями данных моделей. Тогда для любого предложения* $\varphi \in S(\sigma_A)$ *имеем*

$$\mu(\varphi) = \frac{\mu_1(\varphi) \cdot \|E_1\| + \mu_2(\varphi) \cdot \|E_2\|}{\|E_1\| + \|E_2\|}.$$

Доказательство данного предложения также можно найти в работе [6].

Следствие 1. *Рассмотрим прецедентные модели* $\mathfrak{A}_{E_1}$, $\mathfrak{A}_{E_2}$ *и* $\mathfrak{A}_E$ *такие, что* $\mathfrak{A}_E = \mathfrak{A}_{E_1} * \mathfrak{A}_{E_2}$. *Пусть нечеткие модели* $\mathfrak{A}_{\mu_1}$, $\mathfrak{A}_{\mu_2}$ *и* $\mathfrak{A}_\mu$ *являются фазификациями данных*

моделей. Тогда для любого предложения $\varphi \in S(\sigma_A)$ *имеем*

$$\min\{\mu_1(\varphi), \mu_2(\varphi)\} \le \mu(\varphi) \le \max\{\mu_1(\varphi), \mu_2(\varphi)\}.$$

Определение 6. *Пусть* $\mathfrak{A}_{K_1}$ *и* $\mathfrak{A}_{K_2}$ *– обобщенные нечеткие модели, порожденные классами прецедентных моделей* K_1, K_2 *соответственно. Модель* $\mathfrak{A}_{K_1 * K_2}$ *называется* ***произведением*** *моделей* $\mathfrak{A}_{K_1}$ *и* $\mathfrak{A}_{K_2}$, *если*

$$K_1 * K_2 = \{\mathfrak{A}_{E_1} * \mathfrak{A}_{E_2} \mid \mathfrak{A}_{E_1} \in K_1 \text{ и } \mathfrak{A}_{E_2} \in K_2\}.$$

Так как произведение прецедентных моделей является коммутативной и ассоциативной операцией, то и произведение обобщенных нечетких моделей также будет коммутативно и ассоциативно.

Теорема 1. *Пусть* $\mathfrak{A}_{K_1}$, $\mathfrak{A}_{K_2}$ *– интервальные модели. Тогда для любого предложения* $\varphi \in S(\sigma_A)$ *имеем*

$$\left.\begin{array}{l} \xi_{K_1}(\varphi) = [\alpha_1, \alpha_2] \\ \xi_{K_2}(\varphi) = [\beta_1, \beta_2] \end{array}\right\} \Rightarrow \xi_{K_1 * K_2}(\varphi) \subseteq [\min(\alpha_1, \beta_1)]; \max(\alpha_2, \beta_2)].$$

Доказательство. Следует из следствия 1.

База знаний по компьютерной безопасности

На основе методологии обобщенных нечетких моделей в Новосибирском государственном университете была разработана программная система RiskPanel [3], по существу являющаяся рабочим местом специалиста (группы специалистов) по обеспечению корпоративной информационной безопасности [9].

Ядром данной системы является база знаний по информационной безопасности. Для организации базы знаний и работы с ней используется технология OntoBox [10] – система представления и хранения данных в формате онтологий, обладающая мощными и гибкими инструментами обработки. Ее использование позволяет обеспечить большую степень модульности и мобильности баз знаний, что является преимуществом при разработке сложных информационных систем.

Для описания прецедентов в базе знаний OntoBox создано семь категорий признаков (классов) – симптомы, угрозы, уязвимости, последствия, потери, контрмеры и конфигурация. Каждая из этих категорий признаков представлена в виде древовидной структуры. Прецеденты в базе характеризуются обладанием определенных признаков из каждой категории. Каждый прецедент формируется исходя из некоторого текста на естественном языке, найденного в сети Интернет.

В ходе анализа текстов, предоставляемых для формирования прецедентов, было обнаружено, что их подавляющее большинство обладает четкой, но не полной информацией, т. е. для каждого конкретного прецедента мы не имеем полной информации об обладании / необладании всеми описанными в базе знаний признаками. Для разрешения этой проблемы было предложено использовать методологию семантики открытого мира, широко применяемую в системах логики описаний (Description Logic) [11]. Основная идея данного подхода заключается в рассмотрении всех возможных интерпретаций неизвестной информации. Таким образом, для математического описания прецедента компьютерной атаки мы будем рассматривать обобщенную нечеткую модель, обладающую определенными свойствами, которую будем называть неполным прецедентом.

Определение 7. *Рассмотрим множество* $U \subseteq S(\sigma_A)$ *и означивание* $\nu : U \to \{0,1\}$. *Будем говорить, что прецедент* $\mathfrak{A}$ ***согласуется*** *с означиванием* ν (*и обозначать* $\mathfrak{A} \uparrow \nu$), *если для любого предложения* $\varphi \in U$ *имеем*

$$\mathfrak{A} \models \varphi \Leftrightarrow \nu(\varphi) = 1.$$

[3] Свидетельство о государственной регистрации программ для ЭВМ № 2011617412 от 23.09.2011 г.

Определение 8. *Рассмотрим множество* $U \subseteq S(\sigma_A)$ *и означивание* $\nu : U \to \{0,1\}$. *Обобщенная нечеткая модель* $\mathfrak{A}_K$ *называется* ***обобщенным прецедентом****, порожденным означиванием* ν, *если*

$$K = \{\mathfrak{A} \mid \mathfrak{A} - \text{прецендент и } \mathfrak{A} \uparrow \nu\}.$$

В данном формализме всю базу знаний системы RiskPanel мы можем рассматривать как конечное множество обобщенных прецедентов. И, делая выводы из данной базы знаний, мы должны согласовывать данные модели.

Для базы знаний, формализованной в виде множества обобщенных прецедентов, наиболее адекватным является принцип согласования, основанный на произведении обобщенных нечетких моделей, так как именно он согласуется с семантикой открытого мира.

Заметим, что каждый обобщенный прецедент $\mathfrak{A}_K$ не является интервальной моделью. Более того, для любого предложения $\varphi \in S(\sigma_A)$ значение истинности $\xi_K(\varphi)$ принадлежит множеству $\{\{0\}, \{1\}, \{0,1\}\}$.

Таким образом, алгоритм подсчета значений истинности в модели согласования, предложенный в теореме 1, нам не подходит. Сформулируем алгоритм подсчета значений истинности в модели согласования обобщенных прецедентов.

Теорема 2. *Пусть* $\mathfrak{A}_{K_1}, ..., \mathfrak{A}_{K_n}$ *– обобщенные прецеденты. Тогда для любого предложения* $\varphi \in S(\sigma_A)$ *имеем*

$$\xi_{K_1 * ... * K_n}(\varphi) = \left\{ \frac{\alpha}{n}; \frac{\alpha+1}{n}; ...; \frac{\alpha+\beta}{n} \right\},$$

где $\alpha = \left\| \left\{ \mathfrak{A}_{K_i} \mid \xi_{K_i}(\varphi) = \{1\} \right\} \right\|$ *и* $\beta = \left\| \left\{ \mathfrak{A}_{K_i} \mid \xi_{K_i}(\varphi) = \{0,1\} \right\} \right\|$.

Доказательство. Пусть $q \in \xi_{K_1 * ... * K_n}(\varphi)$. Тогда найдутся такие прецеденты $\mathfrak{A}_1 \in K_1, ..., \mathfrak{A}_n \in K_n$, что, в силу предложения 3,

$$q = \frac{\varepsilon_1(\varphi) + ... + \varepsilon_n(\varphi)}{n},$$

где для любого $i = 1, ..., n$ имеем $\varepsilon_i(\varphi) = 1$, если $\mathfrak{A}_i \models \varphi$, и $\varepsilon_i(\varphi) = 0$, если $\mathfrak{A}_i \not\models \varphi$. Очевидно, что

$$\alpha \le \varepsilon_1(\varphi) + ... + \varepsilon_n(\varphi) \le \alpha + \beta.$$

Таким образом, получим, что $q \in \left\{ \frac{\alpha}{n}; \frac{\alpha+1}{n}; ...; \frac{\alpha+\beta}{n} \right\}$.

Возьмем теперь $q \in \left\{ \frac{\alpha}{n}; \frac{\alpha+1}{n}; ...; \frac{\alpha+\beta}{n} \right\}$. Допустим, $q = \frac{\gamma}{n}$. Очевидно, что $\alpha \le \gamma \le \alpha + \beta$. Пусть

$$\left\{ \mathfrak{A}_{K_i} \mid \xi_{K_i}(\varphi) = \{1\} \right\} = \left\{ \mathfrak{A}_{K_{i_1}}, ..., \mathfrak{A}_{K_{i_\alpha}} \right\};$$

$$\left\{ \mathfrak{A}_{K_j} \mid \xi_{K_j}(\varphi) = \{0,1\} \right\} = \left\{ \mathfrak{A}_{K_{j_1}}, ..., \mathfrak{A}_{K_{j_\beta}} \right\}.$$

Из каждого обобщенного прецедента первого множества выберем произвольно по одному прецеденту, т. е. $\mathfrak{A}_{i_s} \in K_{i_s}$ $(s = 1, ..., \alpha)$. Из $\gamma - \alpha$ обобщенных прецедентов второго множества выберем по одному прецеденту, на котором предложение φ истинно, а из оставшихся обобщенных прецедентов выберем по одному прецеденту, на котором предложение φ ложно, т. е. выберем такие $\mathfrak{A}_{j_s} \in K_{j_s}$ $(s = 1, ..., \beta)$, что $\mathfrak{A}_{j_s} \models \varphi$, если $s \le \gamma - \alpha$, и $\mathfrak{A}_{j_s} \not\models \varphi$, если $s \ge \gamma - \alpha$.

Очевидно, что

$$\mathfrak{A}_{i_1} * ... * \mathfrak{A}_{i_\alpha} * \mathfrak{A}_{j_1} * ... * \mathfrak{A}_{j_\beta} \models_q \varphi.$$

Таким образом, получим, что $q \in \xi_{K_1 * ... * K_n}(\varphi)$. ■

Заметим, что, согласовывая конечное множество обобщенных прецедентов, мы не будем получать интервальную модель. Но при $n \to \infty$ значения истинности предложений на согласованной модели будут стремиться к интервалам на множестве $[0, ..., 1] \cap \mathbb{Q}$. Таким образом, на практике, имея дело с достаточно большим множеством прецедентов, мы можем воспринимать значения истинности на согласованной модели как интервалы.

Атомарно-обобщенные прецеденты

Определение 9. *Обобщенный прецедент называется атомарно-обобщенным, если он порожден означиванием подмножества множества всех атомарных предложений.*

Рассмотрим бесквантороное предложение $\varphi(A_1, ..., A_n)$ от n атомарных предложений. Приведем это предложение к виду СДНФ, т. е. $\varphi(A_1, ..., A_n) = \omega_1 \vee ... \vee \omega_k$, где ω_i $(i \in \{1, ..., k\})$ – элементарные конъюнкции, состоящие из атомарных предложений $A_1, ..., A_n$.

Введем следующие обозначения:

$$Con(\varphi) = \{\omega_1, ..., \omega_k\};$$
$$Con(\varphi, \{0\}) = \{\omega \in Con(\varphi) \mid \xi_K(\omega) = \{0\}\};$$
$$Con(\varphi, \{1\}) = \{\omega \in Con(\varphi) \mid \xi_K(\omega) = \{1\}\};$$
$$Con(\varphi, \{0, 1\}) = \{\omega \in Con(\varphi) \mid \xi_K(\omega) = \{0, 1\}\}.$$

Теорема 3. *Пусть $\mathfrak{A}_K$ – атомарно-обобщенный прецедент и φ – бесквантороное предложение сигнатуры σ_A. Тогда*

$$\xi_K(\varphi) = \begin{cases} \{0\}, & Con(\varphi) = Con(\varphi, \{0\}); \\ \{1\}, & (Con(\varphi, \{1\}) \neq \varnothing) \text{ или} \\ & (\|Con(\varphi, \{0, 1\})\| = 2^{\|(A_i \mid \xi_K(A_i) = \{0, 1\})\|}); \\ \{0, 1\}, & \text{в противном случае.} \end{cases}$$

Доказательство. Очевидно, что

$$\xi_K(\varphi) = \{0\} \Leftrightarrow \forall \mathfrak{A} \in K\ (\mathfrak{A} \models \varphi) \Leftrightarrow \forall \mathfrak{A} \in K\ (\mathfrak{A} \models \omega_1, ... \mathfrak{A} \models \omega_n) \Leftrightarrow$$
$$\Leftrightarrow \xi_K(\omega_1) = ... = \xi_K(\omega_n) = \{0\} \Leftrightarrow Con(\varphi) = Con(\varphi, \{0\}).$$

С другой стороны,

$$Con(\varphi, \{1\}) \neq \varnothing \Leftrightarrow \exists \omega_i \forall \mathfrak{A} \in K\ (\mathfrak{A} \models \omega_i) \Rightarrow$$
$$\Rightarrow \forall \mathfrak{A} \in K\ (\mathfrak{A} \models \varphi) \Leftrightarrow \xi_K(\varphi) = \{1\}.$$

Пусть $A_1, ..., A_n$ – множество атомарных предложений, входящих в формулу φ. Рассмотрим множество элементарных конъюнкций

$$V = \{A_1^{\varepsilon_1}\ \&\ ...\ \&\ A_n^{\varepsilon_n} \mid \exists \mathfrak{A} \in K : \mathfrak{A} \models A_1^{\varepsilon_1}\ \&\ ...\ \&\ A_n^{\varepsilon_n}\}.$$

Пусть $\alpha = \|(A_i \mid \xi_K(A_i) = \{0, 1\})\|$. Очевидно, что $\alpha \neq 0$, так как иначе бы мы имели $Con(\varphi, \{0, 1\}) \neq \varnothing$. Следовательно, $\|V\| = 2^\alpha$.

Допустим теперь, что $Con(\varphi) \neq Con(\varphi, \{0\})$ и $Con(\varphi, \{1\}) = \varnothing$. Следовательно, $Con(\varphi, \{0, 1\}) \neq \varnothing$. Более того, $Con(\varphi, \{0, 1\}) \subseteq V$.

Рассмотрим два случая: $Con(\varphi, \{0, 1\}) = V$ и $Con(\varphi, \{0, 1\}) \neq V$.

Пусть $Con(\varphi, \{0, 1\}) = V$. Тогда для любого прецедента $\mathfrak{A} \in K$ найдется такой конъюнкт $\omega_i \in Con(\varphi, \{0, 1\})$, что $\mathfrak{A} \models \omega_i$. Следовательно, на любом прецеденте $\mathfrak{A} \in K$ предложение φ будет истинно, т. е. $\xi_K(\varphi) = \{1\}$.

Предположим теперь, что $Con(\varphi, \{0, 1\}) \subset V$. Тогда найдется такой прецедент $\mathfrak{A}' \in K$, на котором все конъюнкты из $Con(\varphi, \{0, 1\})$ будут ложны. А так как мы предположили, что

$Con(\varphi, \{1\}) = \varnothing$, то на прецеденте $\mathfrak{A}'$ будет ложно и все предложение φ. С другой стороны, так как $Con(\varphi, \{0,1\}) \neq \varnothing$, то найдется такой прецедент $\mathfrak{A}''$, что $\mathfrak{A}'' \models \varphi$. Следовательно, мы получим $\xi_K(\varphi) = \{0,1\}$. ■

Модуль согласования знаний

В рамках программной системы обеспечения корпоративной информационной безопасности RiskPanel был разработан модуль согласования знаний, извлеченных из различных прецедентов компьютерных атак. На данный момент интерфейс модуля позволяет вычислять значение истинности в виде интервала для формулы, представленной в виде СДНФ [12].

Рассмотрим интерфейс модуля более подробно (рис. 1). Для того чтобы предоставить данные, подаваемые на вход главному алгоритму, пользователю необходимо ввести параметры формулы с помощью предоставляемых средств. Сначала выбираются признаки, входящие во все конъюнкции интересующей СДНФ. В выпадающем списке «Категория признака в СДНФ» можно выбрать категорию признака – симптомы, угрозы, уязвимости, последствия, потери, контрмеры и конфигурация. При выборе категории в окне ниже отображается дерево возможных значений признака данной категории, напоминающее дерево папок в файловой системе. Информация о категориях признаков и возможных значениях хранится в файле базы данных OntoBox. Пользователь выбирает значение признака в дереве и нажимает кнопку «Добавить в конъюнкции», после чего выбранное значение признака и соответствующая ему кнопка «Удалить» добавляются в окно с результирующей СДНФ. Далее задается количество конъюнкций в формуле, после чего в окно с результирующей СДНФ добавляется необходимое количество выпадающих списков с двумя значениями – «+» и «–», где «–» символизирует отрицание аргумента. Данные из этого окна с СДНФ будут переданы на вход основному алгоритму при нажатии на кнопку «Получить значение формулы».

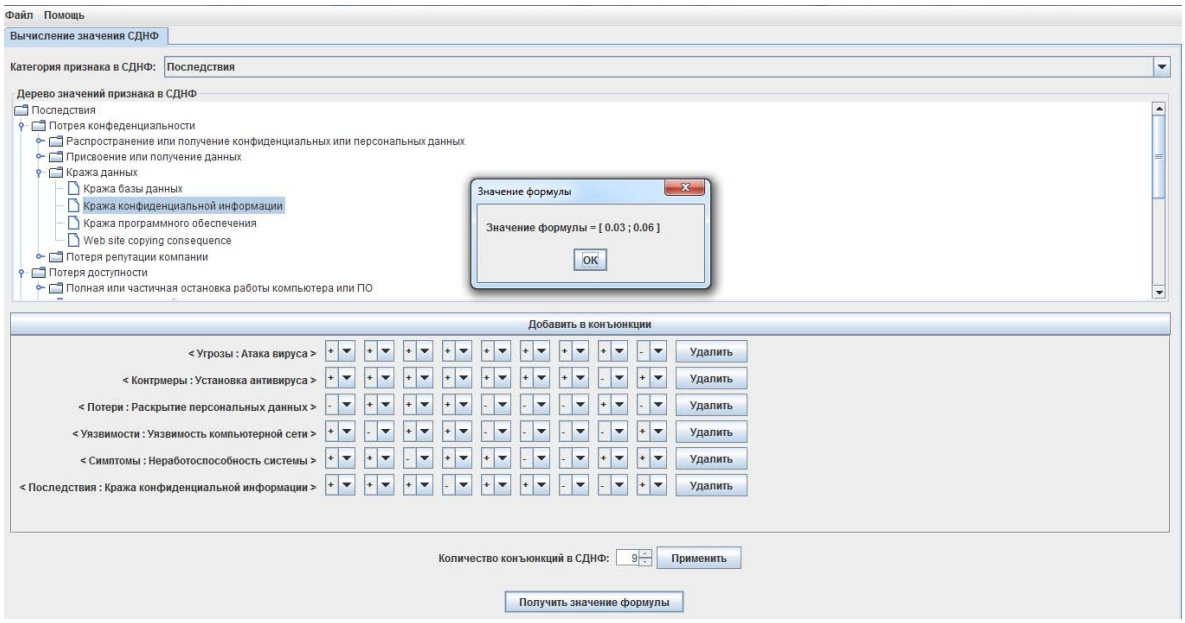

Рис. 1. Общий вид модуля интеграции знаний

Значение формулы вычисляется в виде интервала (см. теорему 2). Значение начала интервала равно отношению количества прецедентов, на которых формула истинна, к количеству всех имеющихся прецедентов. Значение конца интервала равно отношению количества прецедентов, на которых формула истинна, плюс количество прецедентов, на которых истинность формулы не определена, к количеству всех имеющихся прецедентов.

Алгоритм определения истинности формулы на обобщенном прецеденте разработан на основе теоремы 3 и представлен на рис. 2.

```
алг getSDNFVerityOnPrecedent(арг Precedent precedent, арг список
SDNFFormulaAttributes, арг матрица SDNFBoolMatrix)
нач двоич rightValue, цел unknownAttributesCount, список removedIndexes
|     нц для Attribute attr из SDNFFormulaAttributes
|     |     цел attrValueOnPrec := checkIfPrecHasAttr(attr, precedent)
|     |     если (attrValueOnPrec = UNKNOWN_ATTR)
|     |     |     unknownAttrsCount := unknownAttrsCount + 1
|     |     иначе
|     |     |     если (attrValueOnPrec = HAS_ATTR)
|     |     |     |     rightValue := true
|     |     |     иначе
|     |     |     |     rightValue := false
|     |     |     список boolRow := SDNFBoolMatrix.get
|     |     |                 (SDNFFormulaAttributes.indexOf(attr))
|     |     |     нц для цел i от 0 до boolRow.size()
|     |     |     |     если (removedIndexes не содержит i)
|     |     |     |     |     если (boolRow.get(i) != rightValue)
|     |     |     |     |     |     removedIndexes.add(i)
|     |     |     кц
|     кц
|     цел remainingConjCount := SDNFBoolMatrix.get(0).size() -
|                                                 removedIndexes.size()
|     если (remainingConjCount = 0)
|     |     возврат SDNF_FALSE
|     если (unknownAttributesCount = 0)
|     |     возврат SDNF_TRUE
|     если (remainingConjCount < 2^unknownAttributesCount)
|     |     возврат SDNF_UNKNOWN
|     иначе
|     |     возврат SDNF_TRUE
кон
```

Рис. 2. Алгоритм определения истинности формулы на обобщенном прецеденте

Сначала из формулы исключаются ложные конъюнкции, которые противоречат имеющейся информации о прецеденте. Если конъюнкций в формуле не осталось, то формула ложна на прецеденте. Далее если конъюнкции остались, то при отсутствии в конъюнкциях значений признаков, для которых неизвестно, обладает ими прецедент или нет, формула считается истинной на прецеденте. Если конъюнкции остались, и при этом они содержат значения признаков, для которых неизвестно, обладает ими прецедент или нет, то алгоритм действует следующим образом. Если оставшихся конъюнкций меньше, чем 2^n, где n – количество значений признаков в конъюнкциях, для которых неизвестно, обладает ими прецедент или нет, то истинность формулы на прецеденте не определена, иначе формула истинна на прецеденте.

Для определения принадлежности прецеденту каждого из значений признаков, входящих в конъюнкции, требуется $O(n)$ операций, где n – количество всех значений признаков всех категорий, хранящихся в базе знаний OntoBox. Для исключения ложных на прецеденте конъюнкций на основе полученной информации о принадлежности для каждого из значений признаков требуется $O(k)$ операций, где k – количество конъюнкций в СДНФ. Всего количество значений признаков, участвующих в конъюнкциях, не может превышать n. В результате

итоговая алгоритмическая сложность разработанного подхода к определению истинности СДНФ на прецеденте равна $O(n(n+k))$.

Далее, если в базе знаний OntoBox хранится m прецедентов компьютерных атак, то на вычисление значения истинности формулы в виде интервала потребуется $O(mn(n+k))$ операций.

Заключение

Данная работа посвящена описанию математического аппарата и программной реализации одного из модулей системы RiskPanel, направленного на согласование знаний по компьютерной безопасности, полученных из различных интернет-источников.

Алгоритмы, реализованные в этом модуле, разработаны на основе методологии обобщенных нечетких моделей. Знания, полученные из одного источника, формализуются в виде алгебраической системы и хранятся в базе знаний системы RiskPanel. Для реализации согласования знаний строится обобщенная нечеткая модель, являющаяся произведением всех хранящихся в базе алгебраических систем.

Интерфейс системы позволяет вычислять истинностное значение любого бесквантороного предложения. На вход системы подается предложение, представленное в виде СДНФ. Значение истинности высчитывается как вероятностный интервал.

Разработанный алгоритм имеет полиномиальную сложность.

Список литературы

1. *Васенин В. А.* К созданию международной системы мониторинга и анализа информационного пространства для предотвращения и прекращения военно-политических киберконфликтов // Информационные технологии. 2012. № 9. С. 2–10.

2. *Тейз А., Грибомон П., Юлен Г., Пирот А., Ролан Д., Снайерс Д., Воклер М., Гоше П., Вольпер П., Грегуар Э., Дельсарт Ф.* Логический подход к искусственному интеллекту: От модальной логики к логике баз данных: Пер. с фр. М.: Мир, 1998. 494 с.

3. *Пальчунов Д. Е.* Решение задач поиска информации на основе онтологий // Бизнес-информатика. 2008. Т. 1. С. 3–13.

4. *Пальчунов Д. Е.* Моделирование мышления и формализация рефлексии: Ч. 2. Онтологии и формализации понятий // Философия науки. 2008. № 2. С. 62–99.

5. *Palchunov D. E., Yakhyaeva G. E.* Interval Fuzzy Algebraic Systems // Proceedings of the Asian Logic Conference 2005. World Scientific Publishers, 2006. P. 23–37.

6. *Пальчунов Д. Е., Яхъяева Г. Э.* Нечеткие алгебраические системы // Вестн. Новосиб. гос. ун-та. Серия: Математика, механика, информатика. 2010. Т.10, вып. 3. С. 75–92.

7. *Пальчунов Д. Е.* Поиск и извлечение знаний: порождение новых знаний на основе анализа текстов естественного языка // Философия науки. 2009. № 4 (43). С. 70–90.

8. *Яхъяева Г. Э., Ясинская О. В.* Применение методологии прецедентных моделей в системе риск-менеджмента, направленного на раннюю диагностику компьютерного нападения // Вестн. Новосиб. гос. ун-та. Серия: Информационные технологии. 2012. Т. 10, вып. 2. С. 106–115.

9. *Пальчунов Д. Е., Яхъяева Г. Э., Хамутская А. А.* Программная система управления информационными рисками RiskPanel // Программная инженерия. 2011. № 7. С. 29–36.

10. *Малых А. А., Манцивода А. В.* Онтобокс: онтологии для объектов // Изв. Иркут. гос. ун-та. 2009. Т. 2, № 2. С. 94–104.

11. The Description Logic Handbook / Ed. by F. Baader. N. Y.: Cambridge Univ. Press, 2003. 555 p.

12. *Пальчунов Д. Е., Ульянова Е. А.* Методы автоматического порождения поисковых эвристик // Вестн. Новосиб. гос. ун-та. Серия: Информационные технологии. 2010. Т. 8, вып. 3. С. 5–12.

Материал поступил в редколлегию 02.08.2013

G. E. Yakhyaeva, O. V. Yasinskaya

MATCHING METHODS IN COMPUTER SECURITY KNOWLEDGE LEARNED FROM MULTIPLE DOCUMENTS

This paper considers the problem of matching the knowledge of computer security learned from different texts in natural language. A description of the problem with the model-theoretic formalism is presented. Knowledge of the particular computer attack is formalized as underdetermined algebraic system (named generalized precedent). The knowledge base is a set of generalized precedents. The matched value of the truth of a sentence is calculated as an interval defined on [0, 1]. The paper presents the algorithms for calculating the matched value of the truth, the software implementation of the developed methods is described.

Keywords: information security, cyber-attack, cyber-attack precedent, precedent model, generalized fuzzy model, generalized precedent.

References

1. *Vasenin V. A.* K sozdaniu mezhdunarodnoy sistemy monitoringa i analiza informacionnogo prostranstva dlya predotvrascheniya i prekrascheniya voenno-politicheskih kiberkonfliktov // Informacionnye tehnologii. 2012. № 9. S. 2–10.

2. *Teyz A., Gribomon P., Ulen G., Pirot A., Rolan D., Snayers D., Vokler M., Goshe P., Volper P., Greguar E., Delsart F.* Logicheskiy podhod k iskusstvennomu intellektu: Ot modalnoy logiki k logike baz dannyh: Per. s fr. M.: Mir, 1998. 494 s.

3. *Palchunov D. E.* Reshenie zadach poiska informacii na osnove ontologiy // Biznes-informatika. 2008. T. 1. S. 3–13.

4. *Palchunov D. E.* Modelirovanie myshleniya i formalizaciya refleksii: Ch. 2. Ontologii i formalizacii ponyatiy // Filosofiya nauki. 2008. № 2. S. 62–99.

5. *Palchunov D. E., Yakhyaeva G. E.* Interval Fuzzy Algebraic Systems // Proceedings of the Asian Logic Conference 2005. World Scientific Publishers, 2006. P. 23–37.

6. *Palchunov D. E., Yahyaeva G. E.* Nechetkie algebraicheskie sistemy // Vestn. Novosib. gos. un-ta. Seriya: Matematika, mehanika, informatika. 2010. T.10, vyp. 3. S. 75–92.

7. *Palchunov D. E.* Poisk i izvlechenie znaniy: porozhdenie novyh znaniy na osnove analiza tekstov estestvennogo yazyka // Filosofiya nauki. 2009. № 4 (43). S. 70–90.

8. *Yahyaeva G. E., Yasinskaya O. V.* Primenenie metodologii precedentnyh modeley v sisteme risk-menedzhmenta, napravlennogo na rannuu diagnostiku komputernogo napadeniya // Vestn. Novosib. gos. un-ta. Seriya: Informacionnye tehnologii. 2012. T. 10, vyp. 2. S. 106–115.

9. *Palchunov D. E., Yahyaeva G. E., Hamutskaya A. A.* Programmnaya sistema upravleniya informacionnymi riskami RiskPanel // Programmnaya inzheneriya. 2011. № 7. S. 29–36.

10. *Malyh A. A., Mancivoda A. V.* Ontoboks: ontologii dlya obektov // Izv. Irkut. gos. un-ta. 2009. T. 2, № 2. S. 94–104.

11. The Description Logic Handbook / Ed. by F. Baader. N. Y.: Cambridge Univ. Press, 2003. 555 p.

12. *Palchunov D. E., Ulyanova E. A.* Metody avtomaticheskogo porozhdeniya poiskovyh evristik // Vestn. Novosib. gos. un-ta. Seriya: Informacionnye tehnologii. 2010. T. 8, vyp. 3. S. 5–12.