УДК 004.056

С.Г. Самохвалова, А.В. Дмитриева

ПРОЕКТИРОВАНИЕ И РЕАЛИЗАЦИЯ МОДУЛЯ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ ДЛЯ ПРОВЕДЕНИЯ АУДИТА ИНФОРМАЦИОННЫХ СИСТЕМ ПЕРСОНАЛЬНЫХ ДАННЫХ

В статье рассматривается подход к построению модуля интеллектуального анализа данных, который выполняет одну из функций процесса аудита информационных систем персональных данных – оценивает состояние защищенности конкретной информационной системы.

Ключевые слова: аудит, информационная система персональных данных, нечеткая логика, гибридная сеть.

DESIGN AND IMPLEMENTATION OF THE MODULE OF INTELLECTUAL DATA ANALYSIS FOR AUDIT INFORMATION SYSTEM OF PERSONAL DATA

The article considers the approach to building the data mining module, which performs the same functions: the system evaluates the state of the information system security.

Key words: audit, information system of personal data, fuzzy logic, hybrid network.

В современном мире проблема обеспечения информационной безопасности персональных данных становится одной из наиболее актуальных и злободневных. Персональные данные представляют собой информационные ресурсы, которые широко применяются во всех сферах жизни общества и без которых было бы невозможно большинство информационных процессов. Персональные данные обрабатываются специализированными информационными системами, которые в целях обеспечения информационной безопасности необходимо периодически анализировать. Именно данную задачу и обязан выполнять аудит – механизм обеспечения информационной безопасности.

Аудит – процесс получения качественных и количественных оценок о текущем состоянии информационной безопасности организации, компании в соответствии с определенными критериями и показателями безопасности [1]. Аудит предназначен для оценки состояния информационной системы и разработки рекомендаций по применению комплекса организационных мер и программно-технических средств, направленных на обеспечение защиты ресурсов информационной системы от угроз ее безопасности [2].

Можно выделить следующие цели проведения аудита:

1) анализ возможных рисков, связанных с реализацией угроз в отношении информационных ресурсов;

2) оценка текущего уровня защищенности информационной системы;

3) оценка соответствия текущего состояния защищенности требуемому по нормативно-правовым актам;

4) поиск слабых мест в системе защиты;

5) выработка соответствующих рекомендаций по повышению уровня безопасности системы.

Аудит информационных систем персональных данных – один из механизмов обеспечения информационной безопасности. Основной целью данного мероприятия является оценка уровня защищенности системы и выработка конкретных рекомендаций для повышения информационной безопасности и, соответственно, для дальнейшего совершенствования функционала и обеспечивающих подсистем. Аудит информационных систем персональных данных подразумевает исследование информационной системы на основании документов, регулирующих отношения в области обработки персональных данных.

Основным нормативно-правовым документом, регулирующим отношения в области обработки персональных данных, является Федеральный закон №152 «О персональных данных». Этим законом регулируются отношения, связанные с обработкой персональных данных, осуществляемой федеральными органами государственной власти, органами государственной власти субъектов Российской Федерации, иными государственными органами (далее – государственные органы), органами местного самоуправления, иными муниципальными органами (далее – муниципальные органы), юридическими и физическими лицами с использованием средств автоматизации, в том числе в информационно-телекоммуникационных сетях, или без использования таких средств [3].

Следующий документ – Постановление Правительства Российской Федерации № 1119 от 1 ноября 2012 г. «Об утверждении требований к защите персональных данных при их обработке в информационных системах персональных данных». Настоящий документ устанавливает требования к защите персональных данных при их обработке в информационных системах персональных данных (далее – информационные системы) и уровню защищенности таких данных [4]. Определяются типы актуальных угроз, характерные для той или иной ИСПДн:

1) угрозы 1-го типа актуальны для информационной системы, если для нее в том числе актуальны угрозы, связанные с наличием недокументированных (недекларированных) возможностей в системном программном обеспечении;

2) угрозы 2-го типа актуальны для информационной системы, если для нее в том числе актуальны угрозы, связанные с наличием недокументированных (недекларированных) возможностей в прикладном программном обеспечении;

3) угрозы 3-го типа актуальны для информационной системы, если для нее актуальны угрозы, не связанные с наличием недокументированных (недекларированных) возможностей в системном и прикладном программном обеспечении.

Постановление также устанавливает четыре уровня защищенности для ИСПДн, в соответствии с категорией обрабатываемых ПДн, актуальных угроз и количества субъектов ПДн (в зависимости от формы отношений между организацией и субъектами). Выделяются следующие категории обрабатываемых ПДн:

1) биометрические персональные данные – сведения, которые характеризуют физиологические и биологические особенности человека, на основании которых можно установить его личность и которые используются оператором для установления личности субъекта персональных данных;

2) специальные категории персональных данных – персональные данные, касающиеся расовой, национальной принадлежности, политических взглядов, религиозных или философских убеждений, состояния здоровья, интимной жизни субъектов персональных данных;

3) общедоступные персональные данные – персональные данные субъектов персональных данных, полученные только из общедоступных источников персональных данных, созданных в соответствии со статьей 8 Федерального закона «О персональных данных»;

4) иные категории персональных данных – персональные данные, не относящиеся к первым трем пунктам.

Помимо этого, ИСПДн разделяют по количеству субъектов ПДн: 1) менее 100000 субъектов; 2) более 100000 субъектов.

По перечисленным признакам проводится классификация ИСПДн по уровням защищенности. В Постановлении рассматриваются также требования для обеспечения того или иного уровня защищенности.

Модуль интеллектуального анализа данных базируется на теории нечеткой логики. Нечеткая логика предназначена для формализации человеческих способностей к неточным или приближенным рассуждениям, которые позволяют более адекватно описывать ситуации с неопределенностью [5]. Математический раздел нечеткой логики является в своем роде обобщением классической логики и теории множеств. Нечеткая логика основывается на понятии нечеткого множества, которое впервые было предложено Лютфи Заде в 1965 г.

Теория нечеткой логики позволяет отойти от привычных понятий обычной классической логики, вводя в высказывания степень неопределенности. Именно поэтому высказывания в нечеткой логике могут принимать не только значения «Истина» или «Ложь» («0» или «1»), но и любые значения в интервале [0, 1].

Модуль интеллектуального анализа данных представляет собой модульную нечеткую нейронную сеть. Нечеткие нейронные сети, или гибридные сети, объединяют в себе достоинства нейронных сетей и систем нечеткого вывода. Поэтому они наиболее удобны и наименее трудоемки при решении поставленных задач.

В среде Matlab механизм гибридных сетей реализован в модуле ANFIS – Adaptive Neuro-Fuzzy Inference System (адаптивная система нейро-нечеткого вывода). Гибридная сеть, построенная с использованием данного инструмента, представляет собой нейронную сеть с единственным выходом и несколькими входами, являющимися лингвистическими переменными.

Термы входных переменных в этом случае описываются стандартными функциями принадлежности, а термы выходной переменной – константой или линейной функциями принадлежности.

ANFIS реализует алгоритм Сугено. При этом все весовые коэффициенты равны единице.

Для реализации модуля интеллектуального анализа данных необходимо определить входные и выходные показатели. Ими будут требования к уровням защищенности, а также итоговая оценка уровня защищенности ИСПДн.

Эти требования могут определяться однозначно и неоднозначно.

Однозначно определяемые требования: 1) перечень лиц, допущенных к ПДн (X3); 2) должностное лицо, ответственное за обеспечение безопасности ПДн (X5); 3) автоматическая регистрация в электронном журнале безопасности изменения полномочий сотрудника оператора по доступу к ПДн (X7); 4) структурное подразделение, ответственное за обеспечение безопасности ПДн (X8).

И неоднозначно определяемые требования: 1) режим обеспечения безопасности помещений, где обрабатываются ПДн (X1); 2) сохранность носителей (X2); 3) средства защиты информации, прошедшие процедуру оценки соответствия требованиям законодательства РФ (X4); 4) ограничение доступа к содержанию электронного журнала сообщений (X6).

Такое разделение свидетельствует, что однозначно определяемым показателям можно дать оценку 1 или 0 соответственно при выполнении или невыполнении данного требования в системе. Оценка неоднозначно определяемых показателей может принимать значения в интервале [0, 1].

Необходимо на основании имеющихся сведений определить лингвистические переменные.

Для однозначно определяемых лингвистических переменных определены два лингвистических терма: 1) S – требование не выполняется; 2) L – требование выполняется.

Соответственно числовые значения для данных термов: $X_j = 0$ для терма S и $X_j = 1$ для терма L.

Для неоднозначно определяемых лингвистических переменных вводятся три лингвистических терма, описанных следующим образом: 1) S – низкий уровень выполнения требования; 2) M – средний уровень выполнения требования; 3) L – высокий уровень выполнения требования.

Соответственно числовые значения для данных термов принадлежат интервалам: $(X_i) \in [0;0,3]$ для S, $(X_i) \in [0,3;0,9]$ –для M и $(X_i) \in [0,9;1]$ – для L.

Для выходной, результирующей оценки Y_i определяются пять лингвистических термов следующего вида: 1) S – низкая оценка уровня защищенности; 2) SM – оценка уровня защищенности ниже среднего; 3) M – средняя оценка уровня защищенности; 4) ML – оценка уровня защищенности выше среднего; 5) L – высокая оценка уровня защищенности.

Соответственно числовые значения для данных термов принадлежат интервалам: $(Y_i) \in [0;0,1]$ – для S; $(Y_i) \in [0;0,3]$ – для SM; $(Y_i) \in [0,3;0,7]$ – для M; $(Y_i) \in [0,7;0,9]$ – для ML и $(Y_i) \in [0,9;1]$ – для L.

Построение нечеткой нейронной сети начинается с определения базы правил для входных показателей.

Пусть исследуемая информационная система персональных данных классифицируется по 4-му уровню защищенности: категория обрабатываемых данных – общедоступные данные; количество субъектов персональных данных не превышает 100000; третий тип актуальных угроз, связанный с отсутствием недекларированных возможностей в ОС и ПО. Тогда входными параметрами, определяемыми экспертом, будут:

1) режим обеспечения безопасности помещений, где обрабатываются ПДн (X1);

2) сохранность носителей (X2);

3) перечень лиц, допущенных к ПДн (X3);

4) средства защиты информации, прошедшие процедуру оценки соответствия требованиям законодательства РФ (X4).

Сумма правил составляет 54, так как в данном случае три входные лингвистические переменные определяются тремя лингвистическими термами (S, M, L), а четвертая – двумя (S, L): $2 \times 3^3 = 54$.

Правила описываются в форме:

$$R_j : ЕСЛИ\ X1\ есть\ A_1^j\ и\ X2\ есть\ A_2^j\ \ и\ X3\ есть\ A_3^j\ \ и\ X4\ есть\ A_4^j, ТО\ Y_j = B_k^j,$$

где R_j – номер j-го правила; *X1, X2, X3, X4* – входные показатели; Y_j – значение выхода j-го правила; $A_1^j, A_2^j, A_3^j, A_4^j, B_k^j$ – нечеткие подмножества.

Фрагмент правил представлен на рис. 1:

	A	B	C	D	E	F
1		Входные факторы				
2	№	X1	X2	X3	X4	Y
3	1	S	S	S	S	S
4						
5	2	S	S	L	M	M
6						
7	3	S	S	S	L	SM
8						
9	4	S	S	L	S	SM
10						
11	5	S	S	S	M	SM

Рис. 1. Фрагмент базы правил для нечеткой нейронной сети.

Далее на основе данных правил необходимо создать обучающую выборку. Для этого в каждом правиле для каждой входной переменной случайно выбирается значение из интервалов, соответствующее тому или иному лингвистическому терму. Для этого можно использовать численный метод Монте-Карло. Фрагмент обучающей выборки изображен на рис. 2:

№	X1	X2	X3	X4	Y
1	0.289	0.049	0	0.208	0.013
2	0.162	0.073	1	0.301	0.43
3	0.044	0.147	0	0.97	0.14
4	0.189	0.254	1	0.26	0.27
5	0.17	0.149	0	0.82	0.205
6	0.28	0.146	1	0.983	0.37
7	0.245	0.891	0	0.116	0.208
8	0.042	0.573	1	0.831	0.555
9	0.15	0.673	0	0.953	0.62
10	0.115	0.585	1	0.093	0.611

Рис. 2. Фрагмент обучающей выборки.

Редактор ANFIS предусматривает различные инструменты для построения, генерирования и редактирования структуры нечеткой нейронной сети. Команда anfisedit позволяет загружать входные параметры для построения сети, настраивать ее элементы и демонстрирует внешний вид сети. Команда mfedit позволяет настраивать функции принадлежности, менять методы нечетких логических «ИЛИ» и «И» и способ дефаззификации (приведение к четкости). Помимо функций принадлежности, можно также редактировать базу правил.

В окне просмотра правил редактора ANFIS осуществляется проверка построенной нечеткой нейронной сети.

В строке входных данных задаются значения параметров, для которых необходимо произвести оценку защищенности. Допустим, эксперт оценивает требования к уровням защищенности для конкретной ИСПДн следующим образом:

1) режим обеспечения безопасности помещений, где обрабатываются ПДн (X1), – 1 (L);

2) сохранность носителей (X2), – 0,7 (M);

3) перечень лиц, допущенных к ПДн (X3), – 1 (L);

4) средства защиты информации, прошедшие процедуру оценки соответствия требованиям законодательства РФ (X4), – 0,6 (M).

Тогда итоговый выходной показатель будет равен 0,8. Это значит, что данная информационная система персональных данных соответствует требованиям нормативно-правовой документации на 80%. Такая оценка свидетельствует, что уровень защищенности системы выше среднего. Чтобы система выдавала наиболее высокий результат, следует увеличить значение одного из показателей (X2 или X4) с уровня M на уровень L. Результат работы нечеткой нейронной сети представлен на рис. 3.

В ходе работы был рассмотрен и построен модуль интеллектуального анализа данных, позволяющий оценить состояние защищенности информационной системы персональных данных. В дальнейшем будет рассматриваться доработка модуля до системы поддержки принятия решений, которая позволит осуществить основные функции процесса аудита информационных систем персональных данных: описание информационной системы, построение моделей угроз и злоумышленников, определение уровня защищенности, оценка этого уровня, а также рекомендации по улучшению состояния защищенности.

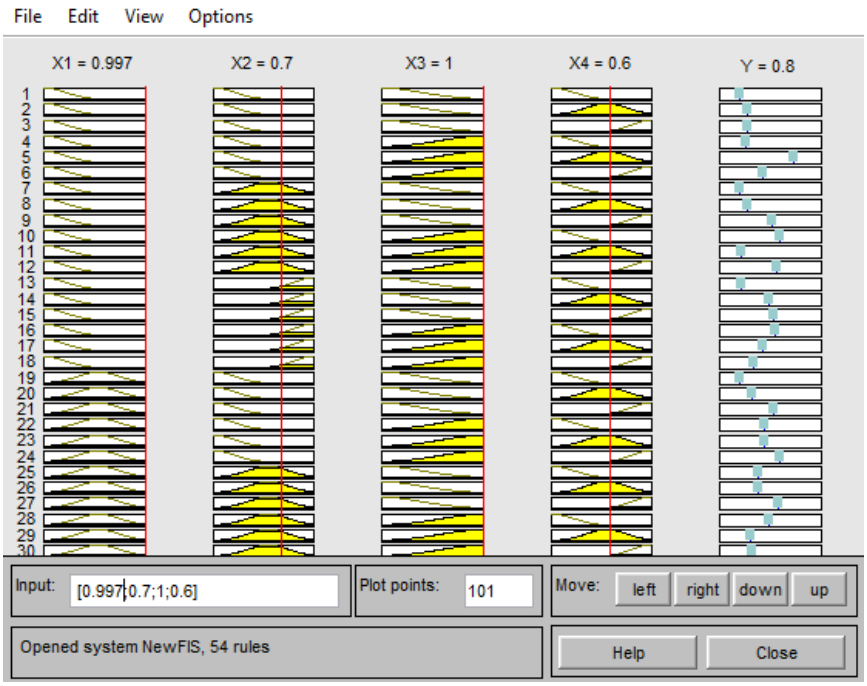

Рис. 3. Результат работы нечеткой нейронной сети.

1. Петренко, С.А., Аудит безопасности Intranet: учебное пособие / С. А. Петренко, А. А. Петренко. – М.: ДМК Пресс, 2002. – 406 с.

2. Аверченков, В.И. Аудит информационной безопасности органов исполнительной власти [Электронный ресурс]: учебное пособие / В. И. Аверченков [и др.]. – Электрон. текстовые данные. – Брянск: Брянский гос. техн. ун-т, 2012. – 100 с. – Режим доступа: http://www.iprbookshop.ru/6992.html. – ЭБС «IPRbooks».

3. Федеральный закон от 27 июля 2006 г. №152-ФЗ «О персональных данных»: офиц. текст – М.: Кремль, 2006. – 22 с.

4. Постановление Правительства «Об утверждении требований к защите персональных данных при их обработке в информационных системах персональных данных» №1119, утвержденное 1 ноября 2012 г: офиц. текст. – М.: Кремль, 2012. – 4 с.

5. Леоненков, А.В. Нечеткое моделирование в среде MATLAB и fuzzyTECH: учебное пособие. – СПб. БХВ Петербург, 2005. – 736 с.