Нейросетевая модель диагностирования компьютерных инцидентов на объектах критической информационной инфраструктуры

Neuro-network model of diagnosing computer security incidents on objects of critical information infrastructure

Саенко / Saenko I.
Игорь Борисович
(ibsaen@mail.ru)
доктор технических наук, профессор.
ФГБУН Санкт-Петербургский институт информатики и автоматизации Российской академии наук, ведущий научный сотрудник.
г. Санкт-Петербург

Авраменко / Avramenko V.
Владимир Семенович
(vsavr@mail.ru)
кандидат технических наук, доцент.
ВАС им. С. М. Буденного, профессор кафедры автоматизированных систем специального назначения.
г. Санкт-Петербург

Маликов / Malikov A.
Альберт Валерьянович
(mkv.vas@yandex.ru)
ФГКВОУ ВО «Военная академия связи им. Маршала Советского Союза С. М. Буденного» МО РФ (ВАС им. С. М. Буденного), адъюнкт кафедры автоматизированных систем специального назначения.
г. Санкт-Петербург

Ясинский / Yasinsky S.
Сергей Александрович
(yasinsky777@mail.ru)
доктор технических наук, доцент.
ЗАО «Институт телекоммуникаций», ведущий специалист.
г. Санкт-Петербург

Ключевые слова: компьютерный инцидент – computer incident; диагностический признак – diagnostic sign; искусственная нейронная сеть – artificial neural network; персептрон – perceptron; автоэнкодер – autoencoder.

Рассматривается модель диагностирования компьютерных инцидентов, фиксируемых на объектах критической информационной инфраструктуры. Определяется структура модели диагностирования. Приводится описание составных элементов и их назначение. Основой модели диагностирования является комбинированная искусственная нейронная сеть.

The paper considers a model of diagnosing computer incidents fixed on objects of critical information infrastructure. The structure of the diagnostic model is defined. Description of the constituent elements and their purpose is given. The basis of the diagnostic model is a combined artificial neural network.

Введение

Объекты критической информационной инфраструктуры представляют собой желаемую мишень для различного рода вредоносных воздействий, в том числе и компьютерных атак. Важность обеспечения безопасности таких объектов и реагирования на компьютерные инциденты подчеркивается принятием Федерального закона от 26.07.2017 № 187-ФЗ «О безопасности критической информационной инфраструктуры Российской Федерации». В соответствии с ним создается государственная система обнаружения, предупреждения и ликвидации последствий компьютерных атак на информационные ресурсы Российской Федерации (ГосСОПКА). Так, согласно приведенной в [1] статистике, в 2018 г. средствами ГосСОПКА выявлено более 4 млрд компьютерных воздействий на российские информационные ресурсы, из них более 17 тысяч наиболее опасных компьютерных атак. Для сравнения: за весь 2017 г. эти показатели составили 2,4 млрд воздействий и около 13 тысяч наиболее опасных атак.

Ситуацию с анализом компьютерных инцидентов на объектах критической информационной инфраструктуры усугубляет также тот факт, что совместно применяемые средства защиты информации имеют разнонаправленные функциональные возможности, вдобавок между ними существует слабая информационная и технологическая связь. В результате, компьютерный инцидент зачастую обрабатывается автономно, без оценки с позиции всей автоматизированной системы. Как следствие, решение, принимаемое на защиту, носит локальный характер и не учитывает возможное косвенное влияние нарушения на инфраструктуру атакованной системы.

В силу того, что средства защиты информации и средства автоматизации при обнаружении или детектировании компьютерного инцидента не предоставляют полную информацию о нарушении, для принятия решения на реагирование требуется дополнительная процедура по получению значении характеристик нарушения безопасности информации. Под нарушением безопасности информации понимается событие, заключающееся в появлении или реализации угрозы безопасности информации. Зачастую эта процедура осуществляется администратором путем изучения записеи в различных журналах событии, что влечет за собой значительные временные затраты. Частично задачу анализа нарушений безопасности информации в автоматическом режиме выполняют SIEM-системы [2]. Ранее проводились исследования по автоматизации диагностирования нарушении безопасности [3, 4]. Но в целом вопросы разработки методологических основ комплексного анализа нарушений безопасности информации требуют дальнейшего исследования. Противоречие между существующими подходами к анализу нарушения безопасности и требованиями к этому анализу обуславливают актуальность такого исследования.

Одним из этапов устранения данного противоречия является разработка модели диагностирования компьютерных инцидентов, позволяющей оперативно и достоверно определять значения характеристик нарушения безопасности (рис. 1). Особенностью даннои модели является использование искусственных неиронных сетеи (ИНС) в задаче диагностирования.

Определение перечня характеристик нарушения безопасности

Фиксируемые в автоматизированных системах компьютерные инциденты подлежат оперативному анализу с целью определения значений характеристик нарушения безопасности информации.

Множество характеристик нарушения безопасности представляет собой совокупность значений, по которым можно получить детальное представление о выявленном нарушении безопасности. Пример основных характеристик нарушения безопасности информации представлен в таблице 1.

Ряд характеристик нарушения безопасности может быть получен путем прямого измерения или расчета (таких как IP-адрес отправителя и получателя, идентификаторы пользователей, время и другие). Однако для ряда характеристик нарушения безопасности отсутствует возможность определения их функциональной зависимости от значений диагностических признаков. Ряд характеристик нарушения безопасности имеет взаимосвязи между собой. Так, например, результаты реализации компьютерного инцидента могут указывать на его цели.

На основании имеющегося набора значений характеристик нарушения безопасности информации принимается решение на реагирование на нарушение. В общем случае перечень характеристик нарушений безопасности должен соответствовать возможностям системы защиты по реагированию. В то же время получение значений характеристик нарушения является нетривиальной задачей, актуальной для автоматизированных систем различного назначения.

Обработка диагностических признаков

Применяемые в объектах критическои информационной инфраструктуры средства автоматизации и защиты информации генерируют и сохраняют в своих журналах большое количество служебной информации, обусловленнои в том числе и компьютерными инцидентами. Служебная информация содержит различного рода признаки нарушений безопасности информации, такие как типы событий, номера портов, адреса, идентификаторы процессов, время обнаружения и другие. Также могут быть использованы данные из других источников информации о состоянии автоматизированной системы в период реализации нарушения безопасности информации.

Многообразие признаков нарушения безопасности информации требует решения проблемы их эффективнои обработки. Существуют различные подходы к анализу и обработке больших массивов данных. К их числу относятся кластерныи анализ, факторныи анализ, искусственные неиронные сети, деревья решении, регрессионныи анализ, дискриминантныи анализ, корреляционныи анализ и другие [5–7].

Применительно к задаче диагностирования обученная ИНС способна осуществлять обобщение диагностических признаков с последующим нахождением значений характеристик нарушения безопасности, а также прогнозирование этих значений [8].

Учитывая большой объем выборки диагностических признаков, собираемых со всей автоматизированной системы в момент наступления факта нарушения безопасности, возникает необходимость произвести её уменьшение путем группирования однородных, достаточно схожих между собой признаков. Данный этап предварительной обработки данных из журналов событий заключается в отборе из всего множества событий $X = \{x_j\},\ j = \overline{1,n}$, где n – общее количество зарегистрированных событий во всех имеющихся журналах, такого подмножества информативных событий для диагностирования компьютерных инцидентов $X = \{x_i\}, i = \overline{1,m}$, где m – количество информативных событий, которое позволит определить значение характеристик нарушения безопасности. Предварительная обработка может осуществляться как экспертным методом, так и с применением одной из архитектур ИНС – автоэнкодера (рис. 2). Автоэнкодер позволяет уменьшить пространство входных признаков, экономно закодировав (сгруппировав) их в значительно меньшем числе нейронов [9]. Данный прием

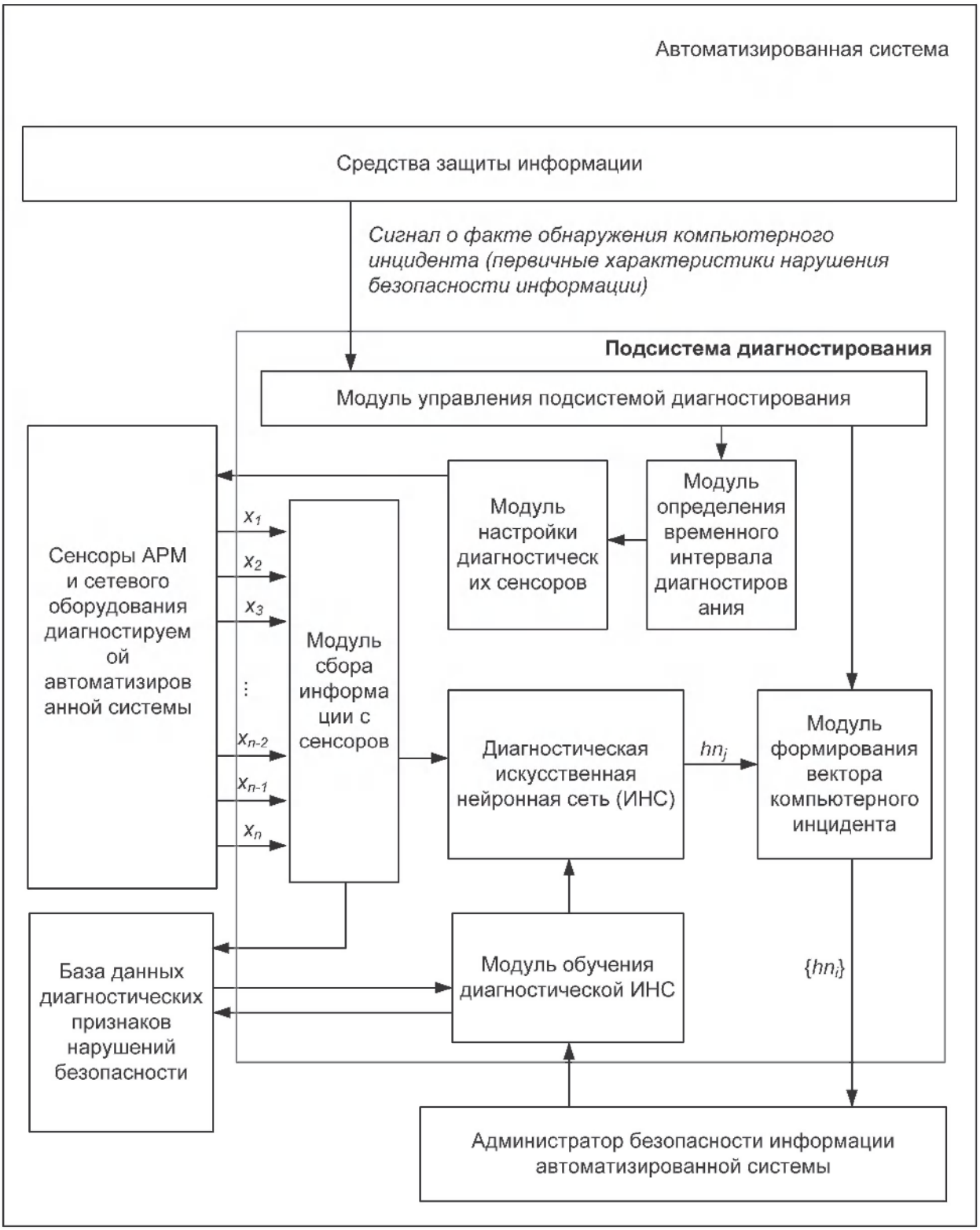

Рис. 1. Модель диагностирования компьютерного инцидента в автоматизированной системе

Таблица 1

Пример характеристик нарушения безопасности информации

№ п/п	Наименование характеристики	Значения характеристики
1.	Цель реализации угрозы	Нарушение конфиденциальности информации Нарушение целостности информации Нарушение доступности информации
2.	Объект воздействия	Автоматизированное рабочее место Серверное оборудование Сетевое оборудование
3.	Способ реализации	За счет эксплуатации известных уязвимостей За счет эксплуатации новых уязвимостей
4.	Источник нарушения	Внешний Внутренний
5.	Результаты реализации	Изменение Удаление Создание Блокировка и другие
6.	Стадия реализации	Реализована В процессе реализации
7.	Характер воздействия	Преднамеренное Непреднамеренное
8.	Последствия	Критический ущерб Некритический ущерб
9.	Время обнаружения	Момент времени обнаружения нарушения средствами зашиты информации
10.	Идентификатор (название) события	В соответствии с описанием журналов событий средств защиты, обнаруживших нарушение
11.	Идентификатор пользователя	В соответствии с настройками ИКС
12.	Адрес источника атаки	IP-адрес элемента инфокоммуникационной системы, где зафиксировано нарушение (при наличии)
13.	Адрес объекта атаки	IP-адрес, порт элемента инфокоммуникационной системы куда направлено нарушение (при наличии)

основан на статистическом методе – методе анализа главных компонент, определяющем линейное преобразование $X' = W \cdot X$. Описание состояния, характеризуемого вектором $\vec{x} \in R^n$, трансформируется в вектор $\vec{x}' \in R^m$ посредством матрицы $W \in R^{n+m}$, при $n > m$ с сохранением основной информации об исходном состоянии. Поскольку элементы исходного множества X могут иметь связи между собой, т.е. некоторые элементы коррелированны между собой, то автоэнкодер учтет эти корреляции.

Идентификация значения j-ой характеристики нарушения безопасности hn_j, осуществляется при помощи многослойного персептрона. Количество нейронов выходного слоя, с которых будут считываться эти значения, определяется в зависимости от числа возможных значений j-ой характеристики нарушения безопасности. Для идентификации бинарной характеристики нарушения безопасности используются два нейрона выходного слоя. На рис. 3 представлен трехслойный персептрон, состоящий из m нейронов входного слоя и p нейронов скрытого слоя, число которых определяется экспериментальным путем ($p<m$).

Задача классификации решается следующим образом. На этапе обучения на вход персептрона подаются тестовые векторы диагностических признаков, позволяющие однозначно судить о значении анализируемой характеристики нарушения (например, непреднамеренное нарушение).

Выходное значение нейронной сети сравнивается с требуемым значением. В случае расхождения значений происходит корректировка весовых коэффициентов. Корректировка производится методом обратного распространения ошибки. Функция активации нейронов – ReLU.

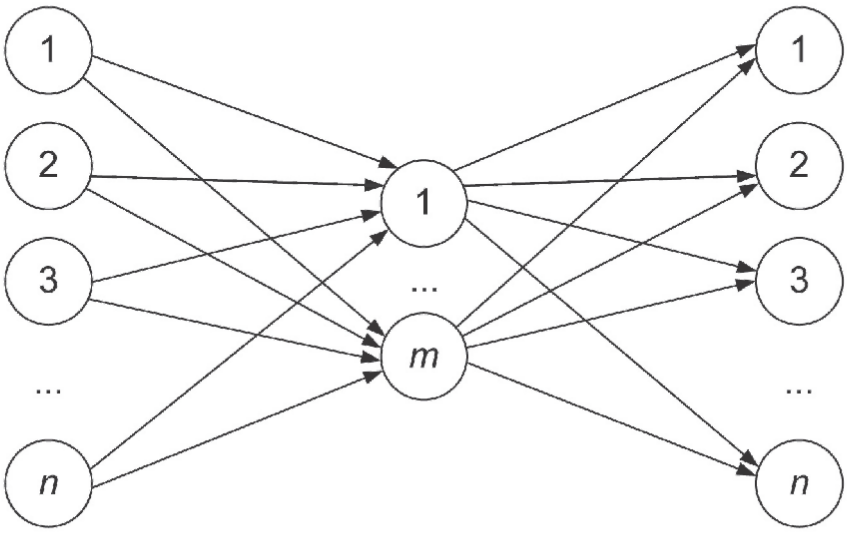

Рис. 2. Структура автоэнкодера для снижения размерности входных данных

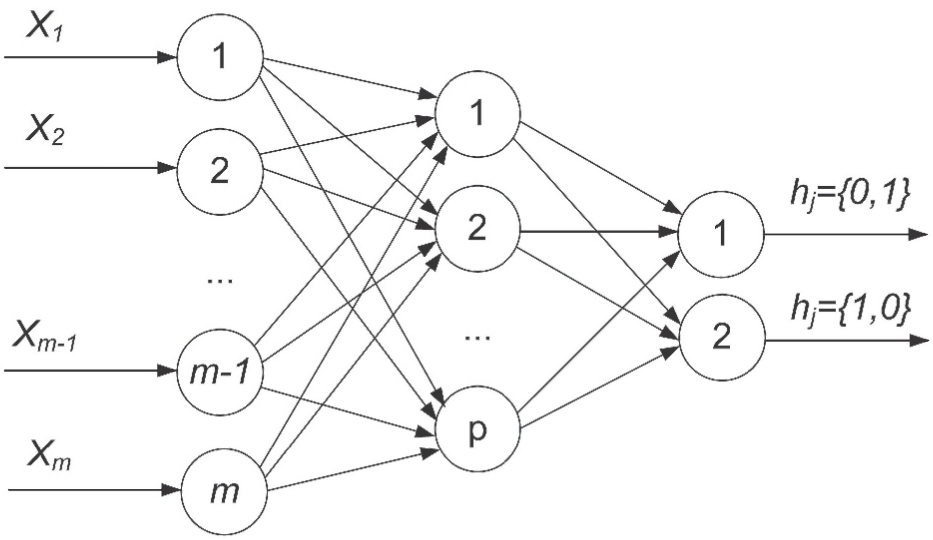

Рис. 3. Структура искусственной нейронной сети для идентификации значения бинарной характеристики нарушения безопасности

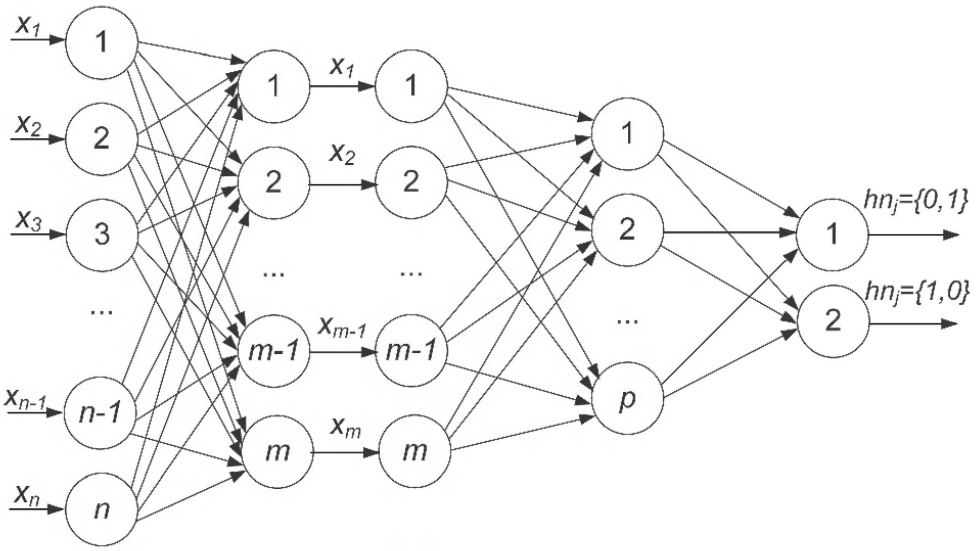

Рис. 4. Структура комбинированной искусственной нейронной сети для определения значения бинарной характеристики нарушения безопасности информации

Объединив указанные искусственные нейронные сети в единую последовательную цепочку, получаем комбинированную ИНС (рис. 4).

Автоэнкодер снижает размерность вектора исходных данных X и осуществляет первый этап по обобщению входных данных (выполняет их сжатие). В скрытом слое автоэнкодера формируются групповые диагностические признаки. Далее они поступают на вход трехслойного персептрона, который осуществляет процедуру обработки диагностических признаков и на выходе выдает значение бинарной характеристики нарушения безопасности.

Важной задачей, предшествующей обучению и работе комбинированной ИНС, является формирование множества входных данных [10, 11]. В работе [12] показано, что число ключевых слов в логах (важных для анализа журналов событий) не превосходит 170. В проведенном эксперименте было выбрано 128 слов и словосочетаний из журналов событий, появление которых является признаком компьютерного инцидент или предпосылки его появления. В зависимости от сигнала об инциденте от средств защиты выбирается размер временного интервала. Так, для DDoS-атаки интервал можно выбрать меньше, чем в случае нарушении правил разграничения доступа. Для каждого журнала событий автоматизированной системы строится вектор числа слов в выбранном временном интервале. На вход диагностической ИНС поступает суммарный вектор по всем журналам событий. Так, для автоматизированной системы, состоящей из 40 компьютеров с установленной операционной системой семейства Linux, могут быть эффективно использованы 400 журналов событий. Соответственно входной вектор диагностической ИНС должен содержать 51200 нейронов.

Обучение диагностической ИНС осуществляется на основе имеющейся базы данных о нарушениях безопасности информации, составленной экспертным путем. База данных состоит из 276 записей, которые включают вектора ключевых слов журналов событий для 20 компьютеров. Из 276 записей 113 соответствуют преднамеренному нарушению, а 163 – непреднамеренному нарушению.

Система диагностирования была реализована на языке программирования Python. Для настройки ИНС сети база данных была разделена на две части. Для обучения было взято 194 записи. Оставшиеся 82 записи использовались для тестирования. Оценка диагностической ИНС проводилась по общепринятым показателям точности Pr и полноты Rc:

$$Pr = \frac{TP}{TP + FP}, \quad Rc = \frac{TP}{TP + FN},$$

где TP – количество записей классифицируемых как преднамеренное нарушение, в то время как оно были преднамеренным, FP – количество записей классифицируемых как преднамеренное нарушение, в то время как оно фактически было непреднамеренным, FN – количество записей классифицируемых как непреднамеренное нарушение, в то время как оно на самом деле были преднамеренным нарушением.

Экспериментальная проверка модели

Для решения задачи оптимизации общего времени диагностирования возникает необходимость в преобразовании структуры ИНС путем уменьшения числа нейронов для сокращения времени обучения без существенной потери точности результата. Зависимость показателей точности и полноты от числа нейронов предпоследнего слоя комбинированной ИНС представлена в таблице 2.

Из таблицы 2 видно, что диагностическая ИНС показывает весьма высокую точность вычисления значения бинарной характеристики нарушения безопасности, на примере характера воздействия (преднамеренное/непреднамеренное). В случае если предпоследний слой содержит свыше 30 нейронов, показатель точности составляет более 75%, что говорит о большой доле правильно классифицированных преднамеренных нарушений среди всего объема нарушений, классифицированных как преднамеренные из тестовой выборки. Показатель полноты при этом составляет 85 % и указывает на долю правильно найденных значений преднамеренных нарушений из всего объема реально преднамеренных нарушений, имеющихся в тестовой выборке. При этом 30 нейронов предпоследнего слоя является оптимальным количеством для комбинированной ИНС для диагностирования системы из 20 компьютеров, поскольку увеличение числа нейронов не приводит к значительному приросту точности и полноты результатов. При этом время обучения ИНС возрастает с 330 секунд до 446 секунд.

Заключение

Предлагаемая модель диагностирования компьютерных инцидентов в автоматизированных системах, в отличие от известных подходов к диагностированию, позволяет повысить оперативность и достоверность получаемых значений характеристик нарушения безопасности. Определение значений характеристик нарушения безопасности осуществляется на основе ИНС. Способность обученного персептрона к решению задач классификации позволяет определять значения характеристики нарушения hn безопасности по имеющимся диагностическим признакам X в условиях отсутствия заданного в явном виде отношения $\boldsymbol{F : X \rightarrow hn}$.

Обученная ИНС позволяет достаточно быстро определить значение характеристики нарушения безопасности, что в свою очередь позволит обеспечить оперативное и обоснованное реагирование на выявленное нарушение.

Дальнейшие исследования связываются с совершенствованием диагностической ИНС, позволяющей определять несколько значений характеристик нарушения безопасности, исследование возможности применения технологии Transfer Learning на уже заранее обученных многослойных ИНС в целях оптимизации временных затрат на обучение, а также исследованим влияния размера временного интервала на точность диагностирования.

Работа выполнена при поддержке РФФИ (проект 18-07-01369) и бюджетной темы № АААА-А16-116033110102-5.

Таблица 2

Значения показателей точности *Pr*, полноты *Rc* в зависимости от числа нейронов скрытого слоя

Число нейронов	*TP*	*FP*	*TN*	*FN*	Точность, *Pr*	Полнота, *Rc*
5	19	14	34	15	0,58	0,56
10	20	13	35	14	0,61	0,59
15	23	11	37	11	0,68	0,68
20	24	10	38	10	0,71	0,71
25	26	9	39	8	0,74	0,76
30	29	9	39	5	0,76	0,85
35	29	8	40	5	0,78	0,85
40	29	8	40	5	0,78	0,85

Литература

1. Котенко, И. В. Создание новых систем мониторинга и управления кибербезопасностью / И.В. Котенко, И.Б. Саенко // Вестник Российской академии наук. – 2014. – Т. 84, № 11.– С. 993–1001.

2. Мурашов, Н. Н. Компьютерные атаки на информационные ресурсы Российской Федерации: факты и цифры [Электронный ресурс] / Н.Н. Мурашов // Информационная безопасность. – 2018. – № 6. – Режим доступа: http://lib.itsec.ru/articles2/infseccompsetey/kompyuternye-ataki-na-informatsionnye-resursy-rossiyskoy-federatsii-fakty-i-tsifry [Дата обращения: 30.05.2019], свободный. – Загл. с экрана.

3. Авраменко, В. С. Адаптивный контроль защищенности информации от несанкционированного доступа на основе информационных образов / В.С. Авраменко // Проблемы информационной безопасности. Компьютерные системы. – 2010. – № 2. – С. 45–49.

4. Авраменко, В. С. Распознавание пользователей на основе нечетких треугольных примитивов / В.С. Авраменко, С.Н. Бушуев // Известия ВУЗов. Приборостроение. – 1999. – Т. 42, №1. – С. 13–19.

5. Рубаков, С. В. Современные методы анализа данных / С.В. Рубаков // Наука. Инновации. Образование. – М.: ФГБУ «Российский научно-исследовательский институт экономики, политики и права в научно-технической сфере», 2008. – С. 165–176.

6. Cluster Analysis and Artificial Neural Networks. A Case Study in Credit Card Fraud Detection / E.M. Carneiro [et al.] // Proceedings of the 12th International Conference on Information Technology – New Generations, 2015. – P. 122–126.

7. Vaarandi R. A data clustering algorithm for mining patterns from event logs / R. Vaarandi // Proceedings of the 3rd IEEE Workshop on IP Operations and Management, 2003. – P. 119–126.

8. Kurd Z. Artificial Neural Networks in Safety-critical Applications / Z. Kurd. –University of York. Department of Computer Science, 2005.

9. Автоэнкодер: подход к снижению размерности векторного пространства с контролируемой потерей информации / Н.В. Акинина [и др.] // Известия ТулГУ. Технические науки. – 2016. – № 9. – С. 3–12.

10. Execution anomaly detection in distributed systems through unstructured log analysis / Q. Fu [et al.] // Proceedings of the 9th IEEE International Conference on Data Mining, 2009. – P. 149–158.

11. Detecting large-scale system problems by mining console logs / W. Xu [et al.] // Proceedings of the 22nd ACM Symposium on Operating Systems Principles, 2009. – P. 117–132.

12. Zou, D. Q. UiLog: Improving log-based fault diagnosis by log analysis / D.Q. Zou, H. Qin, H. Jin // Journal of computer science and technology. – 2016. – Vol. 31, No. 5. – P. 1038–1052.