Doi: 10.36724/2409-5419-2021-13-2-52-65

МОДЕЛИРОВАНИЕ УГРОЗ БЕЗОПАСНОСТИ ИНФОРМАЦИИ В ТЕРРИТОРИАЛЬНО-РАСПРЕДЕЛЕННЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ

МИНЯЕВ
Андрей Анатольевич

АННОТАЦИЯ

Введение: при проектировании систем защиты информации важным этапом является моделирование угроз безопасности, что подразумевает за собой определение перечня актуальных угроз для конкретной информационной системы, на основании которого принимаются решения по нейтрализации актуальных угроз. В настоящее время подавляющее большинство систем являются территориально-распределенными, что, потенциально, увеличивает количество актуальных угроз безопасности информации, ввиду сложности инфраструктуры, особенностей технологии обработки информации, незащищенных каналов связи. В этой связи, **целью исследования** является моделирование угроз безопасности информации в территориально-распределенных информационных системах. Проведенный анализ зарубежных и российских методологий моделирования угроз безопасности информации информационных систем показал, что основными проблемами являются большой объем данных для моделирования, а также недостатки экспертных методов. Установлено, что для решения поставленной задачи необходимо использовать **методы** машинного обучения, теорию адаптивных нечетких нейронных продукционных систем с алгоритмами нечеткого вывода и применение технологий Data Science при обработке большого объема данных. В работе были определены необходимые и достаточные показатели для моделирования угроз безопасности информации, на основе нечетких нейронных продукционных системах была предложена методика определения актуальных угроз безопасности информации. Методика автоматизирована и, гипотетически, исключает ошибки экспертов, увеличивает количество определяемых актуальных угроз безопасности информации, снижает финансовые затраты на закупку средств защиты информации, отличается от существующих тем, что процесс автоматизирован, обладает низкой вычислительной сложностью, отсутствует необходимость привлечения высококвалифицированных специалистов, позволяет определять перечень актуальных угроз в системах различных типов и классов, может быть адаптирована для работы с международными базами данных. **Практическая значимость** заключается в автоматизации процесса – разработке программы для ЭВМ, реализующей предложенную методику. **Обсуждение:** дальнейшие исследования целесообразно продолжить в определении наилучших параметров адаптивных нечетких нейронных продукционных систем и алгоритмов нечеткого вывода.

КЛЮЧЕВЫЕ СЛОВА: угрозы безопасности информации; территориально-распределенные информационные системы; методологии моделирования; Data Science; ANFIS.

Сведения об авторе:
старший преподаватель Санкт-Петербургского государственного университета телекоммуникаций им. проф. М.А. Бонч-Бруевича, г. Санкт-Петербург, Россия, minyaev.a@gmail.com

Для цитирования: *Миняев А.А.* Моделирование угроз безопасности информации в территориально-распределенных информационных системах // Наукоемкие технологии в космических исследованиях Земли. 2021. Т. 13. № 2. С. 52–65. Doi: 10.36724/2409-5419-2021-13-2-52-65

Введение

Моделирование угроз безопасности информации в информационных системах (ИС) является одним из основных этапов создания систем (подсистем) защиты информации (СЗИ) ИС. С точки зрения законодательства Российской Федерации (РФ) в области обеспечения безопасности информации этот этап жизненного цикла создания СЗИ является обязательным в соответствии с частью 2 статьи 19 закона № 152-ФЗ «О персональных данных», в соответствии с пунктом 4 приказа ФСТЭК России от 18 февраля 2013 г. № 21, а также в соответствии с приказом ФСТЭК России от 11 февраля 2013 г. № 17: «определение угроз безопасности информации, реализация которых может привести к нарушению безопасности информации в информационной системе, и разработку на их основе модели угроз безопасности информации».

В соответствии с методическими документами[1] (МД) регуляторов РФ этап моделирования УБИ необходимо проводить на этапе создания СЗИ, т.е. в том случае, когда еще на создана СЗИ, а иногда и сама ИС. В этом случае провести инструментальный анализ защищенности, необходимый в соответствии с МД не представляется возможным [1, 2]. В этой связи моделирование угроз безопасности информации (УБИ), как правило, проводится экспертным путем с привлечением специалистов в области информационной безопасности [3–9]. Такой подход является трудоемким и не исключает известные недостатки экспертного метода [10].

[1]Базовая модель угроз безопасности персональных данных при их обработке в информационных системах персональных данных (Выписка) (утв. ФСТЭК России 15.02.2008).
Методика определения актуальных угроз безопасности персональных данных при их обработке в информационных системах персональных данных (утв. ФСТЭК России 14.02.2008).
Методические рекомендации по разработке нормативных правовых актов, определяющих угрозы безопасности персональных данных, актуальные при обработке персональных данных в информационных системах персональных данных, эксплуатируемых при осуществлении соответствующих видов деятельности (утв. ФСБ России 31.03.2015).
Методика моделирования угроз безопасности информации (проект ФСТЭК России, 2020 г.).

Методологии моделирования угроз безопасности информации

В настоящее время в мире существуют множество методологий моделирования УБИ. Основными их них являются следующие [11–16]:

1. STRIDE and Associated Derivations.
2. PASTA (The Process for Attack Simulation and Threat Analysis).
3. LINDDUN (Linkability, Identifiability, Non-Repudiation, Detectability, Disclosure of Information, Unawareness, Non-Compliance).
4. CVSS (Common Vulnerability Scoring System).
5. Attack Trees.
6. Persona non Grata.
7. Security Cards.
8. hTMM.
9. Quantitative Threat Modeling Method.
10. Trike.
11. VAST Modeling.
12. OCTAVE.
13. Mitre.
14. БДУ ФСТЭК России.

В табл. 1 приведен анализ существующих методологий моделирования УБИ.

Таблица 1

Анализ существующих методологий моделирования УБИ

Метод моделирования УБИ	Преимущества	Недостатки
STRIDE	Является ли наиболее зрелым. Прост в использовании	Низкий уровень ложных срабатываний и умеренно высокий уровень ложных отрицательных результатов. Длительный по временным характеристикам
PASTA	Непосредственный вклад в управление рисками. Содержит встроенные приоритеты по смягчению угроз. Наличие исчерпывающей документации	Требует ввода данных по безопасности от операционного уровня, уровня управления, архитектуры и развития ИС. Содержит встроенные приоритеты по смягчению угроз. Трудоемкий

Продолжение таблицы 1

Метод моделирования УБИ	**Преимущества**	**Недостатки**
LINDDUN	Содержит встроенные приоритеты для минимизации УБИ	Количество УБИ может быстро возрастать по мере модификации, усовершенствования и усложнения ИС. На эффективность и результативность данного метода отрицательно влияет угрозы общего характера. Трудоемкий. Длительный по временным характеристикам
CVSS	Наличие автоматизированных компонентов. Содержит встроенную приоритизацию мер по минимизации УБИ. Имеет постоянные результаты при повторении	Непрозрачные расчеты баллов и возможные несоответствия, вызванные экспертными оценками. Сложен в использовании. Нет руководства по оценке подцелей, атак или рисков
Attack Trees	Имеет постоянные результаты при повторении. Прост в использовании, если есть полное представление об ИС	Полезен только тогда, когда ИС и проблемы безопасности в ней хорошо известны и поняты. Эксперты должны имеют высокий опыт в области кибербезопасности. Не приводится руководство по оценке подцелей, атак или рисков
Persona non Grata	Прост в применения. Вносит непосредственный вклад в управление рисками. Имеет постоянные результаты при повторении	Имеет тенденцию обнаруживать только определенное подмножество типов УБИ
Security Cards	Определяет сложные угрозы	Приводит ко многим ложным срабатываниям
hTMM	Отсутствие ложных срабатываний. Отсутствие пропущенных УБИ. Стабильный результат независимо от того, кто занимается моделированием УБИ. Экономическая эффективность	Применение Security Cards на основе предложений разработчиков. Обязательное использование инструментальных средств. Трудоемкий. Длительный
Quantitative TMM	Содержит встроенные приоритеты для минимизации УБИ. Имеет автоматизированные компоненты	Трудоемкий. Длительный
Trike	Непосредственный вклад в управление рисками. Содержит встроенные приоритеты для минимизации УБИ. Имеет автоматизированные компоненты	Система весов Trike кажется слишком расплывчатой. Плохо поддерживается и нет никакой документации
VAST Modeling	Подход позволяет интегрировать VAST в развитие организации и DevOps. Непосредственный вклад в управление рисками. Содержит встроенную приоритизацию мер по минимизации рисков. Имеет согласованные результаты при повторении. Имеет автоматизированные компоненты	Предполагает наличие высококвалифицированных специалистов в области ИБ. Трудоемкий. Длительный

Окончание таблицы 1

Метод моделирования УБИ	Преимущества	Недостатки
OCTAVE	Метод глубокий, но гибкий	Процесс моделирования УБИ требует значительного времени, большой объем документации
Mitre Att&ck	Своевременное реагирование на инциденты. Расследование инцидентов. Анализ деятельности нарушителей. Имеет автоматизированные компоненты	Процесс моделирования УБИ требует значительного времени. Большой объем документации. Требует высокой квалификации специалистов в области ИБ
БДУ ФСТЭК России	Достаточно полный объем перечня известных уязвимостей, УБИ, потенциальных нарушителей и возможных методов нейтрализации УБИ. Структурирован. Имеется документация регулятора для процесса моделирования УБИ	Нет автоматизированных средств для моделирования УБИ Перечень УБИ и уязвимостей не является исчерпывающим. УБИ не являются элементами иерархической классификационной системы угроз. Предполагает наличие высококвалифицированных специалистов в области ИБ. Трудоемкий. Длительный

По результатам проведенного исследования можно сказать о том, что существующие методологии в большинстве своем имеют существенные недостатки, а именно: большой объем данных, отсутствие документации, отсутствие автоматизированных средств моделирования УБИ, необходимость в высокой квалификации специалистов по информационной безопасности.

В связи с вышесказанным в настоящей работе были поставлены следующие задачи:

1. Подготовка набора данных для моделирования УБИ на основании известных баз данных УБИ и уязвимостей, а также разработанных ранее моделей угроз для территориально-распределенных ИС.
2. Анализ сформированного набора данных.
3. Форматирование набора данных для последующей автоматизированной обработки.
4. Выбор и сравнение качества работы нескольких моделей, определение наилучшей.
5. Оптимизация параметров в наилучшей модели.
6. Проверка модели.
7. Разработка программы для ЭВМ для моделирования УБИ «Модель угроз и нарушителя» (МУиН).
8. Итоговое представление результатов работы.

Подготовка набора данных для моделирования УБИ

Для преставления набора данных для автоматизированной обработки и разработки программного обеспечения использовался язык программирования Python 3 и технологии Data Science [16, 18].

В качестве объекта исследования была выбрана территориально-распределенная ИС, которая является информационной системой обработки персональных данных и государственной информационной системой одновременно [1].

ИС обеспечивают принцип централизованного хранения, накопления и многократного использования данных. Для экономии ресурсов и обеспечения информационной безопасности на автоматизированных рабочих местах (АРМ) пользователей хранение данных не осуществляется. Обработка осуществляется на стороне серверной части. Для этого могут использоваться технологии терминального доступа. Такая технология может быть реализована в IT-инфраструктуре путем развертывания терминальной фермы RDS (Romote Desktop Services). При такой технологии рабочие профили пользователей хранятся на серверах терминальной фермы. Это упрощает организацию доступа пользователей к информационным ресурсам ИС, а также наиболее эффективно обеспечивает процессы информационной безопасности. Также технология обеспечивает процессы при удаленной работе пользователей.

Для территориально-распределенных ИС характерно размещение серверных компонент, сетевого оборудования и АРМ пользователей на всей территории страны и, возможно, за ее пределами. В этом случае такие ИС имеют сложную архитектуру с точки зрения расположения своих компонентов и технологий обработки информации.

В состав типовой территориально-распределенной ИС входят следующие компоненты:

1. серверы, включающие:

– серверное оборудование;

– прикладные и специализированные программы, обеспечивающие обработку информации и ее представление в виде, необходимом для последующей автоматизированной обработки;

2. АРМ пользователей:

– типовые АРМ пользователя (стационарные персональные компьютеры);

– мобильное АРМ: планшеты, мобильные телефоны;

– планшеты.

Какивбольшинстветерриториально-распределенных системах присутствуют иные категории пользователей, не относящихся к работникам владельца ИС. К таким категориям пользователей относятся разработчики программного обеспечения и технических средств — лица, обеспечивающие разработку, поставку и внедрение программных, аппаратно-программных и аппаратных средств ИС. Обслуживающий персонал — лица, обеспечивающие штатное функционирование технических средств ИС (работники эксплуатационных подразделений владельца ИС, осуществляющие обслуживание и техническое сопровождение инженерных систем и помещений, в которых размещается оборудование ИС).

Сложностью моделирования УБИ для такой ИС является обработка большого объема данных [13], необходимого при моделировании УБИ, а именно: данные из БДУ ФСТЭК России[2] и/или зарубежных баз данных и знаний, сложность использования методических документов регуляторов РФ в области обеспечения безопасности информации, таких как:

Базовая модель угроз ПДн ФСТЭК России, 2008 г.

Методика определения актуальных угроз ПДн, 2008 г.

Методический документ ФСТЭК России. Меры защиты информации в государственных информационных системах (утв. ФСТЭК России 11.02.2014 г.)

Проект методического документа ФСТЭК России. Методика определения угроз безопасности информации в информационных системах (2020 г.)

Банк данных угроз ФСТЭК России.

«Методические рекомендации по разработке нормативных правовых актов, определяющих угрозы безопасности персональных данных, актуальные при обработке персональных данных в информационных системах персональных данных, эксплуатируемых при осуществлении соответствующих видов деятельности» (утв. ФСБ России 31.03.2015 N149/7/2/6–432).

В данной работе набор данных был сформирован на основании данных из БДУ ФСТЭК России и на основании ранее разработанных моделей угроз подобных ИС, описанной выше.

На рис. 1 и 2 представлены графики анализа данных БДУ ФСТЭК России.

Также был проведен анализ уязвимостей БДУ ФСТЭК России. На рис. 3 представлен график зависимости количества уязвимостей от производителя.

Как видно из графиков информация имеет большой объем, что является трудоемким в процессе моделирования УБИ [1, 17]. При экспертном подходе моделирования УБИ влечет за собой ошибки, связанные с таким подходом,

[2]https://bdu.fstec.ru/

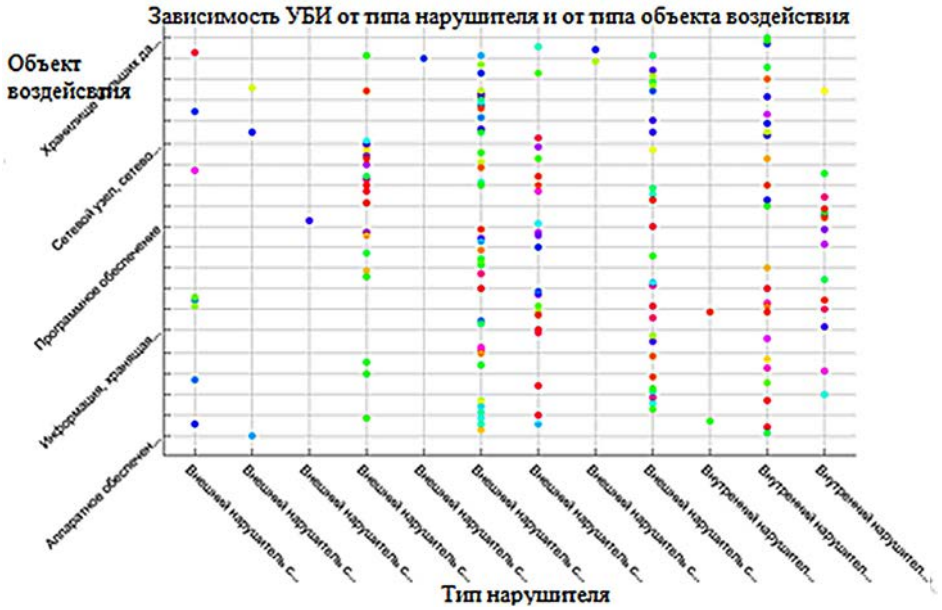


Рис. 1. Зависимость УБИ от типа нарушителя и от типа объекта воздействия

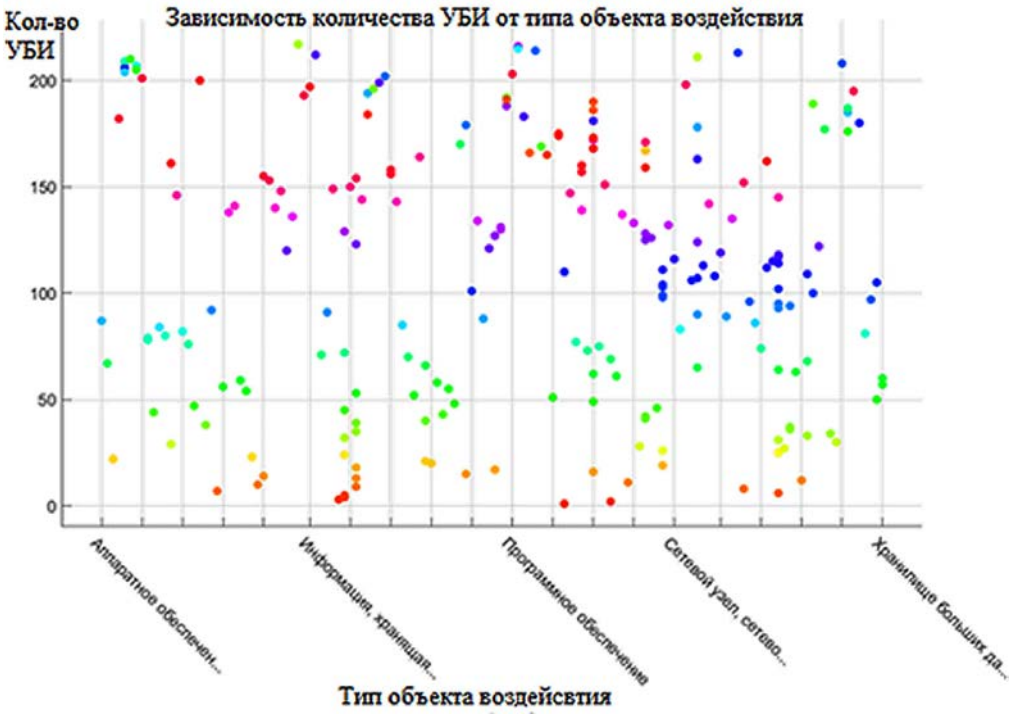


Рис. 2. Зависимость количества УБИ от типа объекта воздействия

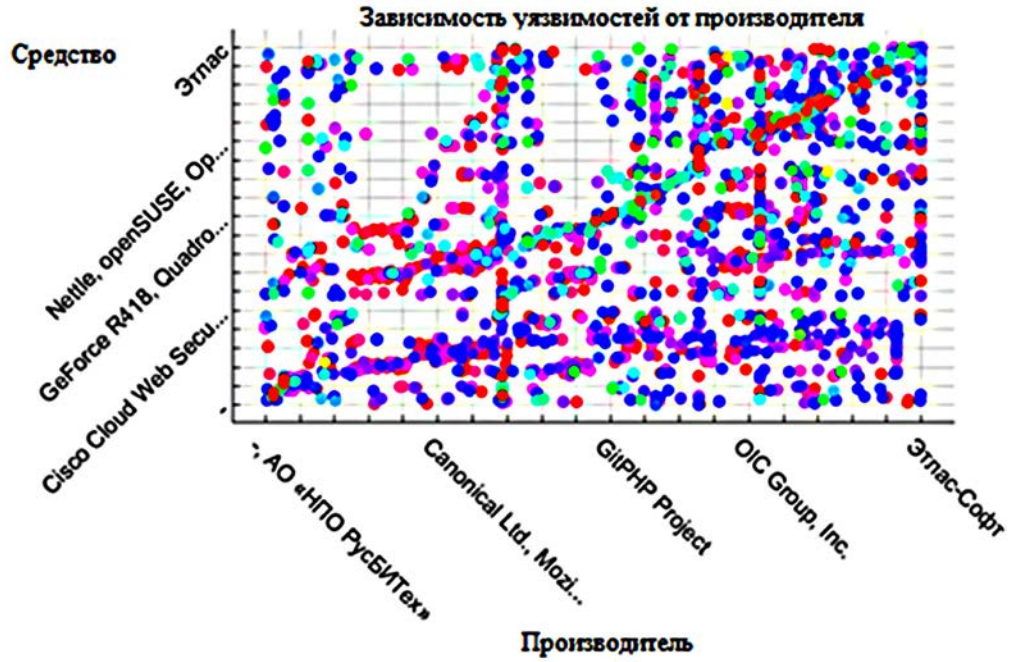


Рис. 3. График зависимости уязвимостей от производителя

такие как личное мнение эксперта, разрозненность и несогласованность мнений, трудоемкость. Методические документы регуляторов РФ определяют подход и этапы моделирования УБИ, при этом ошибки экспертов и трудоемкость не учитываются [1, 19–20].

В соответствии с МД регуляторов РФ актуальность УБИ определяется в зависимости от актуального нарушителя в ИС, перечнем потенциальных УБИ и уязвимостей в ИТ-инфраструктуре ИС, а также возможными последствиями от реализации УБИ. В этой связи набора данных

был сформирован из сведений базы данных угроз ФСТЭК России, моделей УБИ территориально-распределенных ИС и технических решений исследуемой ИС.

Форматирование набора данных

На рис. 4 и 5 изображены наборы данных dataframe БДУ ФСТЭК России до конвертации данных.

Для подготовки автоматизированной обработки данных необходимо выполнить конвертацию данных. Код на языке программирования Python 3 для преобразования данных dataframe (конвертирование данных) представлен ниже:

```
train = pd.read_csv('threats.csv', encoding='utf-8')
df = pd.get_dummies(train)
# Преобразование строковых данных
for col in list(df.columns):
    # Выбор колонок для преобразования
if ('ft²' in col or 'kBtu' in col or 'Metric Tons CO2e' in col or 'kWh' in
    col or 'therms' in col or 'gal' in col or 'Score' in col):
    # Конвертация
    df[col] = df[col].astype(float)
```

Выбор модели

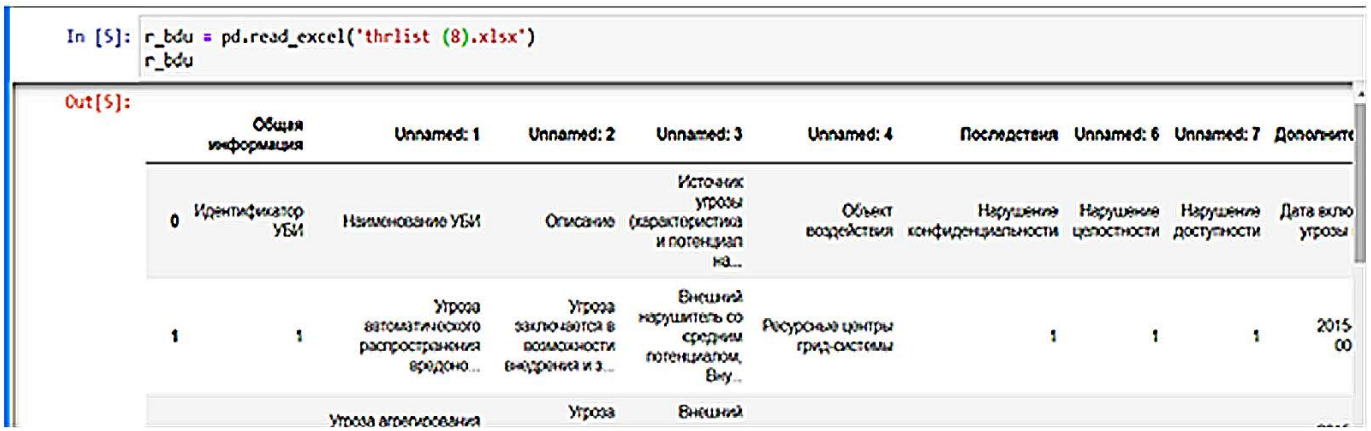
Было проведено исследование нечетких нейронных продуктивных сетей ANFIS (adaptive neuro-fuzzy inference system) с применением алгоритмов Сугено-Такаги, Такаги-Сугено-Канга (TSK), Ванга-Менделя и Мамдани. Отмечено, что зависимость погрешности от количества правил при проверке на тестовой выборке меньше у сети TSK. В этой связи для моделирования УБИ была выбрана нечеткая продуктивная сеть ANFIS, основанная на нечеткой системе вывода Такаги-Сугено-Канга (TSK).

Алгоритм ее работы заключается в реализации нечеткой продукционной модели, основанной на правилах типа:

$$R_i : IFx_i ISA_{i1} AND...ANDx_j ISA_{ij} AND...ANDx_{im} ISA_{im}, THEN$$

$$y = c_{i0} + \sum_{j=1}^{m} c_{ij} x_j), j = 1,...,n$$

Для этого была сформирована база правил для моделирования УБИ. Пример заполнения базы правил для моделирования УБИ на основании сформированного в данной работе набора данных приведен в табл. 2.

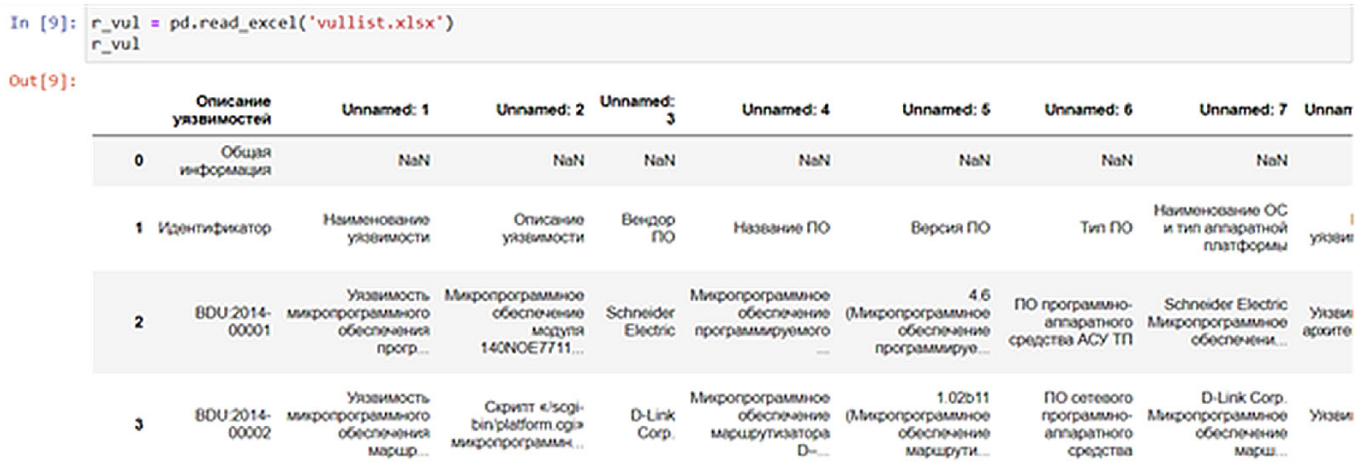


In [5]: r_bdu = pd.read_excel('thrlist (8).xlsx')
r_bdu

Out[5]:

	Общая информация	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Последствия	Unnamed: 6	Unnamed: 7	Дополнит
0	Идентификатор УБИ	Наименование УБИ	Описание	Источник угрозы (характеристика и потенциал на...	Объект воздействия	Нарушение конфиденциальности	Нарушение целостности	Нарушение доступности	Дата вклю угрозы
1	1	Угроза автоматического распространения вредоно...	Угроза заключается в возможности внедрения и з...	Внешний нарушитель со средним потенциалом, Вну...	Ресурсные центры грид-системы	1	1	1	2015-00
		Угроза агрегирования	Угроза	Внешний					

Рис. 4. Dataframe УБИ БДУ ФСТЭК России до конвертации



In [9]: r_vul = pd.read_excel('vullist.xlsx')
r_vul

Out[9]:

	Описание уязвимостей	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6	Unnamed: 7	Unnam
0	Общая информация	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
1	Идентификатор	Наименование уязвимости	Описание уязвимости	Вендор ПО	Название ПО	Версия ПО	Тип ПО	Наименование ОС и тип аппаратной платформы	уязви
2	BDU:2014-00001	Уязвимость микропрограммного обеспечения прогр...	Микропрограммное обеспечение модуля 140NOE7711...	Schneider Electric	Микропрограммное обеспечение программируемого ...	4.6 (Микропрограммное обеспечение программируе...	ПО программно-аппаратного средства АСУ ТП	Schneider Electric Микропрограммное обеспечени...	Уязви архите
3	BDU:2014-00002	Уязвимость микропрограммного обеспечения маршр...	Скрипт «/scgi-bin/platform.cgi» микропрограмм...	D-Link Corp.	Микропрограммное обеспечение маршрутизатора D-...	1.02b11 (Микропрограммное обеспечение маршрути...	ПО сетевого программно-аппаратного средства	D-Link Corp. Микропрограммное обеспечение марш...	Уязви

Рис. 5. Dataframe уязвимостей БДУ ФСТЭК России до конвертации

Таблица 2

База правил метода моделирования УБИ

№	ЕСЛИ (IF)			ТО (THEN)
	Тип нарушителя (источник воздействия)	ИТ – инфраструктура (объект воздействия)	Версия ПО	
1	Внешний нарушитель с низким потенциалом, Внутренний нарушитель с низким потенциалом	Виртуальная машина VMWare	6.5 (VMWare Workstation), от 7.0.0 до 7.1.4 включительно (VMWare Workstation)	Угроза несанкционированного доступа к защищаемым виртуальным машинам со стороны других виртуальных машин
2	Внешний нарушитель с высоким потенциалом	Мобильное устройство (аппаратное устройство) на базе iOS	(Android), до 10.3.3 включительно (iOS)	Угроза контроля вредоносной программой списка приложений, запущенных на мобильном устройстве
…				
n	Внешний нарушитель со средним потенциалом, Внутренний нарушитель со средним потенциалом	Средство защиты информации	12.4 (Cisco IOS), 12.4 (Cisco IOS), 15.0 (Cisco IOS), 15.0 (Cisco IOS), 15.1 (Cisco IOS), 15.1 (Cisco IOS), 12.2 (Cisco IOS), 12.2 (Cisco IOS), 15.2 (Cisco IOS), 15.2 (Cisco IOS)	Угроза несанкционированного воздействия на средство защиты информации

Количество правил *n* для случая моделирования УБИ на основе БДУ ФСТЭК России равно 222. Для определения уязвимостей на основе БДУ ФСТЭК России оно будет составлять *n* = 30321. Правила представлены в табл. 3 как единое, фактически оно представляет множество правил, состоящих отдельно по типу нарушителя, типу СрЗИ (например, SecretNet, Dallas Lock и т.д.) и по воздействию. Набор данных был сформирован уже с учетом этих нюансов. Объект воздействия — совокупность данных из БДУ ФСТЭК России и данных об ИТ-инфраструктуре ИС, взятых из моделей УБИ и проектных решений по СЗИ.

В предлагаемой методике определения актуальных УБИ ANFIS базируется на следующих положениях[1]:

– входные переменные являются четкими;

– функции принадлежности (ФП) определены функцией Гаусса:

$$\mu_{A_{ij}}(x_j) = \exp(-\frac{1}{2}(\frac{x_j - a_{ij}}{b_{ij}})^2)$$

где *x* — входные сети a_{ij}, b_{ij} — настраиваемые параметры ФП.

— нечеткая импликация Ларсена — нечеткое произведение;

– Т-норма — нечеткое произведение;

– композиция не производится;

– метод дефаззификации — метод центроида.

Функциональная зависимость после дефаззификации для получения выходной переменно принимает следующий вид [18]:

$$y' = \frac{\Sigma_i^n((c_{i0} + \Sigma_{j=1}^m c_{ij}x_j)\Pi_j^m \mu_{A_{ij}}(x_j'))}{\Sigma_{i=1}^n \Pi_j^m \mu_{A_{ij}}(x_j')} = \frac{\Sigma_i^n((c_{i0} + \Sigma_{j=1}^m c_{ij}x_j)\Pi_j^m \exp\left[-(\frac{x_j' - a_{ij}}{b_{ij}})^2\right]}{\Sigma_{i=1}^n \Pi_j^m \exp\left[-(\frac{x_j' - a_{ij}}{b_{ij}})^2\right]} \quad (1)$$

Выражение (1) лежит в основе сети ANFIS с применением алгоритма TSK, которая включает в себя пять слоев:

1) Слой состоит из элементов, которые выполняют фаззификацию входных четких переменных:

$$x_j'(j = 1, ..., n).$$

Элементы данного слоя вычисляют значения степеней ФП $\mu_{A_{ij}}\left[x_j'\right]$, заданных функциями Гаусса с параметрами a_{ij}, b_{ij} .

2) Количество элементов этого слоя равно количеству правил в базе (база правил, приведенная в данной работе выше), выполняет нечеткую импликацию степеней принадлежности правил.

[1] Хижняков Ю.Н. Алгоритмы нечеткого, нейронного, нейро-нечеткого управления в системах реального времени. Пермь: ПНИПУ, 2013. 160 с.

3) Данный слой генерирует значения функций $(c_{j0} + \Sigma_{j=1}^{m} c_{ij} x_j^{'})$, которые перемножаются на результаты вычислений элементами второго слоя.

4) Первый элемент слоя 4 необходим для активизации заключений правил в соответствии со значениями, агрегированных в 3 слое, степеней принадлежности предпосылок правил. Второй элемент четвертого слоя производит дополнительные вычисления для последующей дефаззификации результата работы сети ANFIS.

5) Данный слой состоит из одного нормализующего элемента и производит дефаззификацию результатов работы сети ANFIS.

Исходя из вышесказанного следует, что сеть ANFIS TSK содержит два параметрических слоя (слой 1 и 3). Настраиваемыми в процессе обучения сети ANFIS параметрами являются:

– в 1 слое — нелинейные параметры a_{ij}, b_{ij} ФП фаззификатора;

– в 3 слое — параметры c_{i0} и c_{ij} линейных функций $(c_{j0} + \Sigma_{j=1}^{m} c_{ij} x_j^{'})$ из заключений базы правил.

При наличии n правил и m-входных переменных число параметров 1 слоя равно $2nm$, а 2 — $n(m+1)$. Суммарное общее число настраиваемых параметров равно $n(3m+1)$.

На следующем шаге рассчитываются параметры c_{i0} и c_{ij} с линейных функций при условии фиксированных значений параметров a_{ij}, b_{ij}. Параметры c_{i0} и c_{ij} находятся путем решения системы линейных уравнений.

Выходную переменную из выражения (1) представляем в следующем виде:

$$y^{'} = \Sigma \sum_{i=1}^{n} w_i^{'} (c_{i0} + \Sigma \sum_{j=1}^{m} c_{ij} x_j),$$

где

$$w_i^{'} = \frac{\Pi_{j=1}^{m} \mu_{A_{ij}}(x_j^{'})}{\Sigma_{i=1}^{n} \Pi_j^{m} \mu_{A_{ij}}(x_j^{'})} = \frac{\Pi_j^{m} \exp\left[-(\frac{x_{ij}^{'} - a_{ij}}{b_{ij}})^2\right]}{\Sigma_{i=1}^{n} \Pi_j^{m} \exp\left[-(\frac{x_{ij}^{'} - a_{ij}}{b_{ij}})^2\right]} = const$$

Алгоритм обучения сети ANFIS с применением алгоритма TSK.

При k обучающих примерах $x_1^{(k)}, x_2^{(k)}, ..., x_m^{(k)}, y^{(k)}$, где $k = 1,\ldots, K$ и замене значений выходных переменных $y^{'(k)}$ значениями эталонных переменных $y^{(k)}$, получим систему из k линейных уравнений вида:

$$\begin{bmatrix} w_1^{'(1)} & w_1^{'(1)} x_1^{(1)} & \ldots & w_1^{'(1)} x_m^{(1)} & \ldots & w_n^{'(1)} & w_n^{'(1)} x_1^{(1)} & \ldots & w_n^{'(1)} x_m^{(1)} \\ w_1^{'(2)} & w_1^{'(2)} x_1^{(2)} & \ldots & w_1^{'(2)} x_m^{(2)} & \ldots & w_n^{'(2)} & w_n^{'(2)} x_1^{(2)} & \ldots & w_n^{'(2)} x_m^{(2)} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ w_1^{'(k)} & w_1^{'(k)} x_1^{(k)} & \ldots & w_1^{'(k)} x_m^{(k)} & \ldots & w_n^{'(k)} & w_n^{'(k)} x_1^{(k)} & \ldots & w_n^{'(k)} x_m^{(k)} \end{bmatrix} x \begin{bmatrix} c_{10} \\ \ldots \\ c_{1m} \\ \ldots \\ c_{n0} \\ \ldots \\ c_{nm} \end{bmatrix} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \ldots \\ y^{(k)} \end{bmatrix} \quad (3)$$

где $w_i^{'(k)}$ агрегированная степень истинности предпосылок по i-му правилу при предъявлении k-го входного вектора $(x_1^{(k)}, x_2^{(k)}, ..., x_m^{(k)})$.

Таким образом, (3) в сокращенном виде:

$$Wc = y$$

Размерность матрицы W равна $k(m+1)n$, при этом, как правило, количество строк k значительно больше количества столбцов: $k(m+1)n$. Решение этой системы уравнений можно провести за один шаг при помощи псевдоинверсии матрицы W:

$$c = W^{+} y = (W^T W)^{-1} W^T y$$

После определения линейных параметров ij фиксируем и рассчитываем фактические выходные сигналы сети для всех примеров, для чего используем линейную зависимость:

$$y^{'} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \ldots \\ y^{(k)} \end{bmatrix} = Wc$$

определяем вектор ошибок:

$$e = y^{'} - y$$

производим уточнение параметры:

$$a_{ij}^{(k)}(t+1) = a_{ij}^{(k)}(t) - C \frac{dE^{(k)}(t)}{da_{ij}^{(k)}}$$

$$b_{ij}^{(k)}(t+1) = b_{ij}^{(k)}(t) - C \frac{dE^{(k)}(t)}{db_{ij}^{(k)}}$$

После уточнения нелинейных параметров процесс адаптации параметров запускаем до тех пор, пока не наступит повторяемость результатов. Данный алгоритм является гибридным. Его особенность заключается в разделении этапов процесса обучения. Данный алгоритм более эффективен, чем аналогичные методы, например, метод Уидроу-Хоффа, у которого уточнение всех параметров выполняется параллельно и одновременно.

Структура нечеткой нейронной продукционной сети ANFIS с применением алгоритма TSK представлена на рис. 6.

Для проведения вычислений и определения актуальных УБИ в данной работе была разработана программа для ЭВМ «МУиН» на языке программирования Python 3, а также расчеты были проведены в среде MATLAB для сравнения и иллюстрации исследований.

Оптимизация параметров в наилучшей модели, проверка модели

При первоначальных исходных данных и параметров сети ANFIS ошибка обучения сети составляла 3,6–3,7. В ходе проведения экспериментов было установлено, что при определенных параметрах сети ANSIS и оптимизированного набора исходных данных ошибка обучения уменьшается.

На рис. 7–9 представлены настройки сети ANFIS в среде MATLAB.

В результате проведения обучения сети ANFIS с параметрами, указанными на рис. 8–10, при сформированном наборе данных ошибка обучения сети была равной 0,014, что является достаточно хорошим результатом работы.

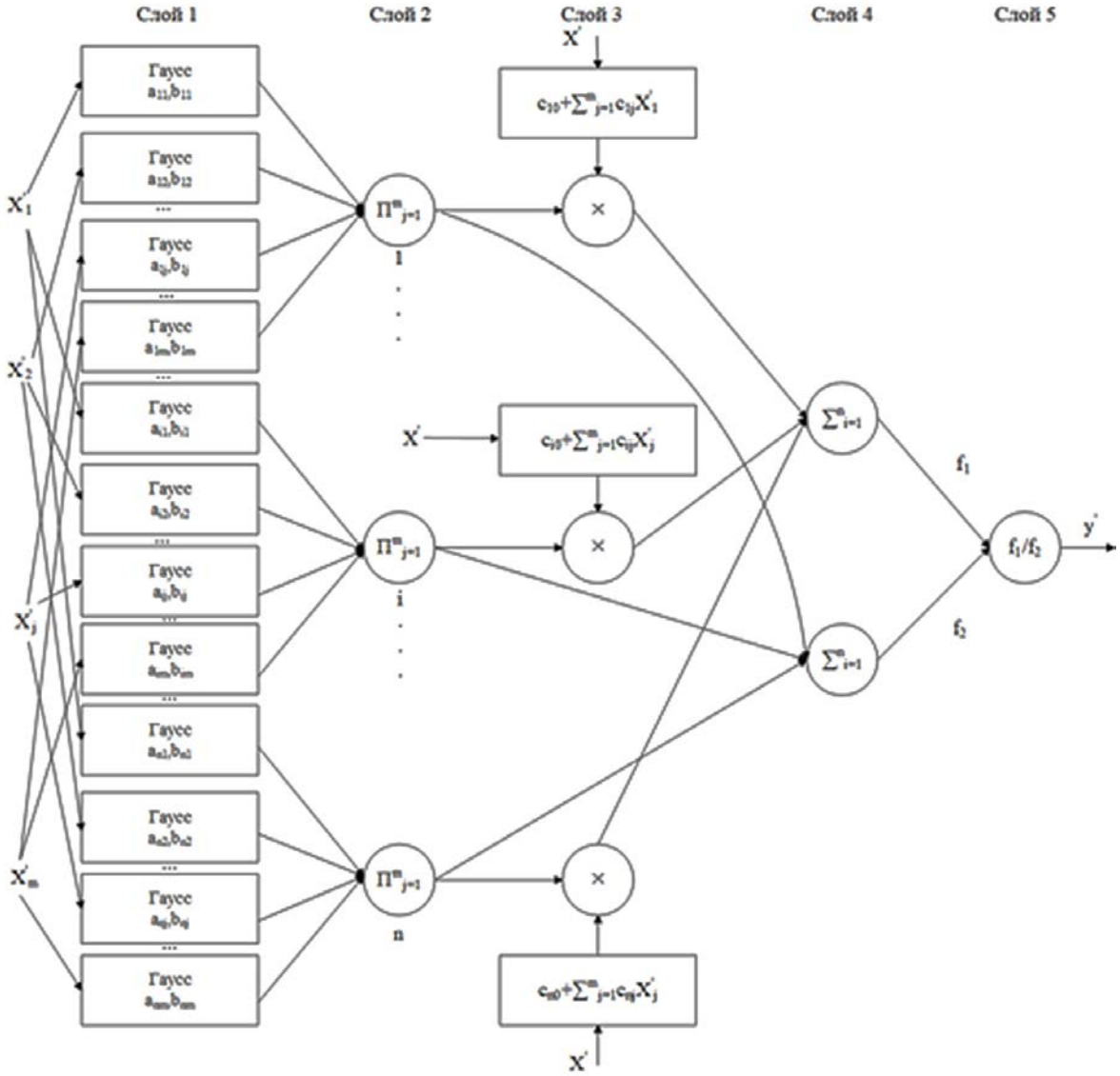


Рис. 6. Сеть ANFIS с применением алгоритма TSK

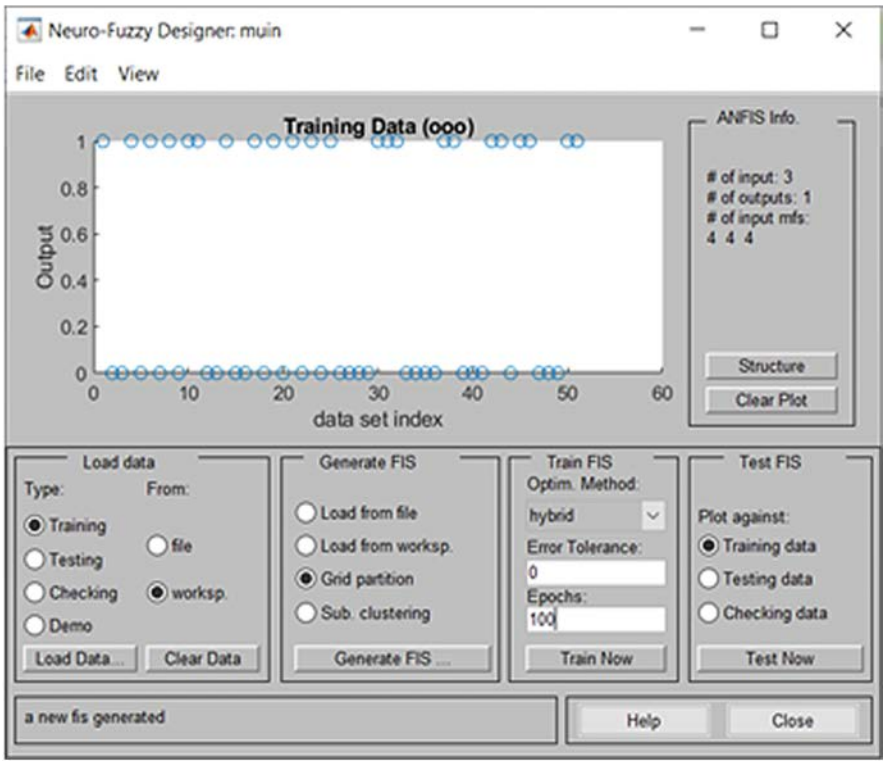


Рис. 7. Настройки ANFIS в среде MATLAB и распределение обучающих данных

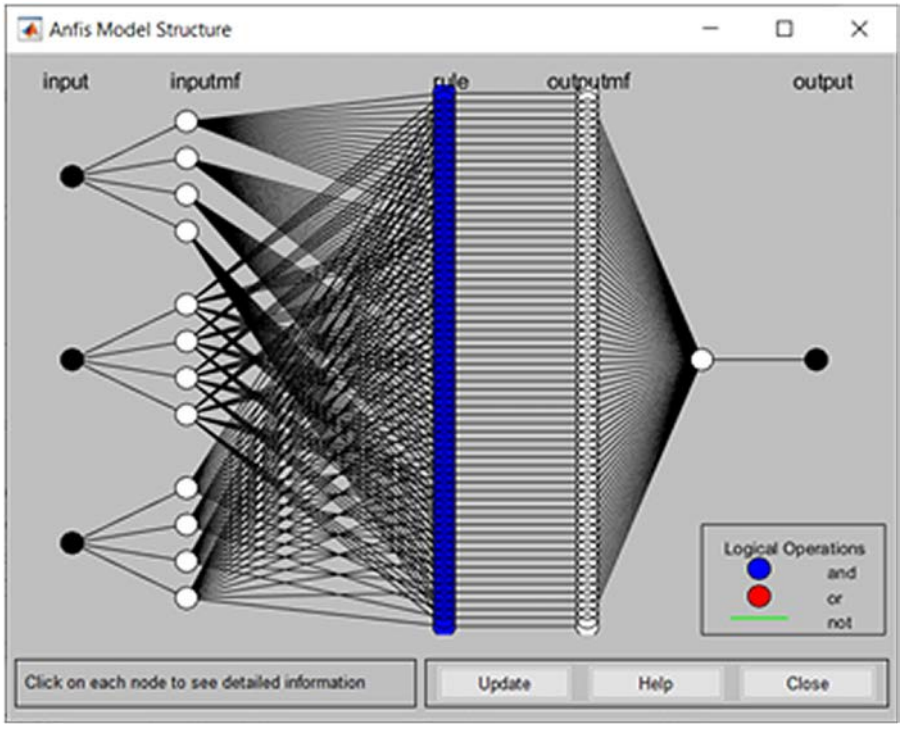


Рис. 8. Структура сети ANFIS

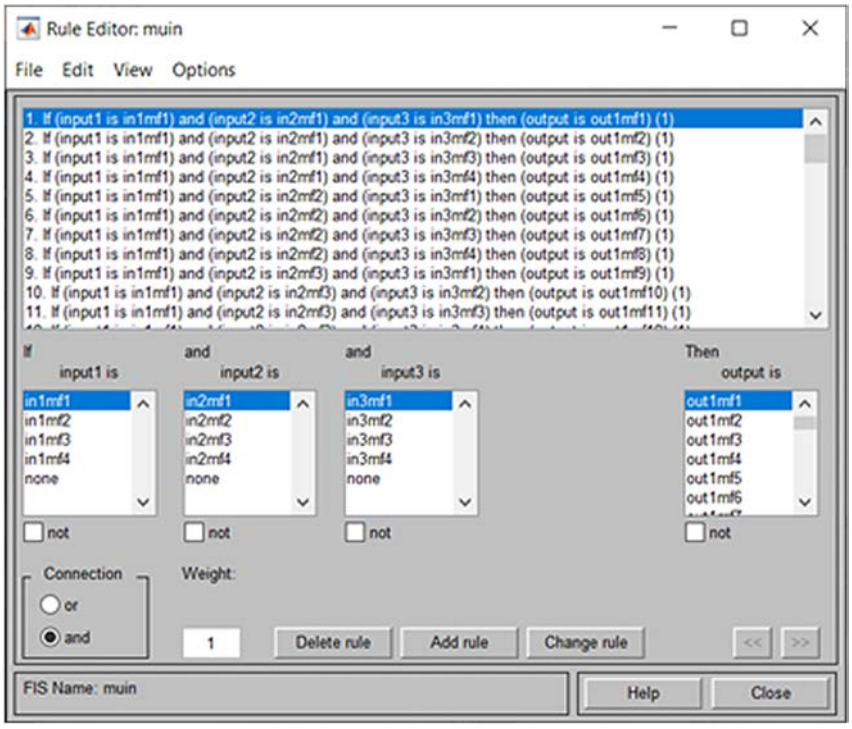


Рис. 9. Правила сети ANFIS

Заключение

Была решена задача подготовки данных для реализации метода моделирования УБИ. Проведен анализ известных зарубежных и российских баз данных и знаний угроз безопасности информации, проведен анализ моделей угроз ряда территориально-распределенных ИС. На основе проведенного анализа выявлены недостатки и преимущества существующих методологий. Установлены ключевые особенности архитектуры территориально-распределенных ИС, определены сложности при моделировании УБИ для таких ИС. Подготовлен набор данных и предложена методика определения актуальных УБИ, проведены необходимые работы для автоматизированной обработки набора данных, связанные с конвертацией данных. Были проанализированы нечеткие нейронные сети ANFIS, алгоритмы их работы. Выбрана наилучшая, основанная на алгоритме нечеткого вывода Такаги-Сугено-Канга. Проведены эксперименты, по результатам которых определены наилучшие параметры системы ANFIS, при которых RMSE достигает значения в диапазоне 0,012–0,023, что является локальным минимумом на заданном интервале и позволяет доказать выполнение поставленной в настоящей работе задачи. Разработана программа для ЭВМ, реализующую предложенную методику определения актуальных УБИ на языке программирования Python 3. Полученные результаты можно использовать для определения актуальных УБИ для территориально-распределенных ИС. В качестве продолжения работы по данной тематике могут являться разработка автоматизированной методики определения актуального нарушителя безопасности информации, а также уменьшения ошибки обучения сети.

Литература

1. *Миняев А.А., Будько М.Ю.* Метод оценки эффективности системы защиты информации территориально распределенных информационных систем // Информатизация и связь. 2017. № 3. С. 119–121.

2. *Будько М.Ю., Миняев А.А.* Методика оценки эффективности системы защиты персональных данных информационной системы // Проблема комплексного обеспечения информационной безопасности и совершенствование образовательных технологий подготовки специалистов силовых структур: Межвузовский сборник трудов VI Всероссийской научно-технической конференции (ИКВО НИУ ИТМО, 10 декабря 2015 г.). 2016. С. 43–45.

3. *Osako T.* Proactive Defense model based on Cyber threat analysis // FUJITSU. Sci. Tech. J. 2016. No. 52 (3). Pp. 72–77.

4. *Noor U.* A machine learning-based FinTech cyber threat attribution framework using high-level indicators of compromise // Future Generation Computer Systems. 2019. Pp. 227–242.

5. *Trifonov R.* Artificial intelligence methods for cyber threats intelligence // Int. J. Comput. 2017. Pp. 129–135.

6. *Huang L.* Adaptive strategic cyber defense for advanced persistent threats in critical infrastructure networks // Perform. Eval. Rev. 2018. No. 46 (2). Pp. 52–56.

7. *Ушаков И.А., Котенко И.В., Овраменко А.Ю., Преображенский А.И., Пелёвин Д.В.* Комбинированный подход к обнаружению инсайдеров в компьютерных сетях // Вестник Санкт-Петербургского государственного университета технологии и дизайна. Серия 1: Естественные и технические науки. 2020. № 4. С. 66–71.

8. *Ушаков И. А.* Обнаружение инсайдеров в корпоративной компьютерной сети на основе технологий анализа больших данных // Вестник Санкт-Петербургского государственного университета технологии и дизайна. Серия 1: Естественные и технические науки. 2019. № 4. С. 38–43.

9. *Савинов Н. В., Токарева К. А., Ушаков И. А., Красов А. В., Сахаров Д. В.* Исследование модели сети ЦОД на основе политик Cisco ACI // Защита информации. Инсайд. 2019. № 4 (88). С. 32–43.

10. *Livshitz I. I., Yurkin D. V., Minyaev A. A.* Formation of the Instantaneous Information Security Audit Concept // Communications in Computer and Information Science. 2016. Vol. 678. Pp. 314–324.

11. *Agrawal A., Ahmed C. M., Chang E. Poster.* Physics-Based Attack Detection for an Insider Threat Model in a Cyber-Physical System // In Proceedings of the 2018 on Asia Conference on Computer and Communications Security, 2018. Pp. 821–823.

12. *Mead N., Shull F., Vemuru K., Villadsen O. A.* Hybrid Threat Modeling Method // Software Engineering Institute, Carnegie Mellon University. 2018. URL: http://resources.sei.cmu.edu/library/asset-view.cfm? AssetID=516617 (дата обращения 15.06.2020).

13. *Khan R., McLaughlin K., Laverty D., Sezer Sakir.* STRIDE-based Threat Modeling for Cyber-Physical Systems // In Proceedings of the 2017 IEEE PES Innovative Smart Grid Technologies Conference Europe.

14. *Shevchenko N., Chick T.A., O'Riordan P., Scanlon T.P., Woody C.* Threat modelling: A summary of available methods // Carnegie Mellon University Software Engineering Institute. 2018. Pp. 1–24.

15. *Yue Li, Teng Zhang, Xue Li, Ting Li*. A Model of APT attack Defense Based On Cyber Threat Detection // Communications in Computer and Information Science, Cyber Security,15th International Annual Conference, CNCERT 2018. Pp. 122–134.

16. *Cao X., Gong N.Z.* Mitigating Evasion Attacks to Deep Neural Networks via Region-based Classification // Proceedings of the 2017 Annual Computer Security Applications Conference (ACSAC). ACM, 2017. Pp. 278–287.

17. *Котенко И. В., Ушаков И. А., Пелёвин Д. В., Преображенский А. И., Овраменко А. Ю.* Выявление инсайдеров в корпоративной сети: подход на базе UBA и UEBA // Защита информации. Инсайд. 2019. № 5 (89). С. 26–35.

18. *Mohassel P., Zhang Y.* SecureML: A System for Scalable PrivacyPreserving Machine Learning // Proceedings of the 2017 IEEE Symposium on Security and Privacy (S&P). IEEE, 2017. Pp. 19–38.

19. *Кузнецов И. А., Липатников В. А., Сахаров Д. В.* Управление АСМК организации интегрированной структуры с прогнозированием состояния информационной безопасности // Электросвязь. 2016. № 3. С. 28–36.

20. *Проноза А. А., Виткова Л. А., Чечулин А. А., Котенко И. В., Сахаров Д. В.* Методика выявления каналов распространения информации в социальных сетях // Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления. 2018. Т. 14. Вып. 4. С. 362–377.

MODELING INFORMATION SECURITY THREATS IN TERRITORIAL-DISTRIBUTED INFORMATION SYSTEMS

ANDREY A. MINYAEV,
St. Petersburg, Russia, minyaev.a@gmail.com

KEYWORDS: information security threats; distributed information systems; modeling methodology; data science; ANFIS.

ABSTRACT

Introduction: in the design of information security systems importance is the modeling of security threats, which implies the definition of a list of software threats to the information system, on the basis of which decisions are made to neutralize actual threats. Today the number of urgent threats to information security is increasing, due to the complexity of the infrastructure, information processing technologies, and unprotected communication channels. In this regard, the goal is to simulate security threats in geographically distributed information systems. The analysis of foreign and Russian methodologies for threat modeling showed problems associated with a large amount of data for modeling, as well as expert methods. It has been established that to solve the problem posed use machine learning methods, the theory of adaptive fuzzy neural production systems with fuzzy inference algorithms and the use of Data Science technologies when processing large amounts of data. The paper uses such data protection methods as data protection tools, based on fuzzy neural security systems, it is proposed to determine the actual threats to information security. The proposed methodology is automated and hypothetically eliminates expert errors, increases the number of frequently used topical threats to information security, reduces financial costs for the purchase of information security tools, differs in that the process is automated, has low computational complexity, there is no need attracting highly qualified specialists, allows you to determine the list of actual threats in systems of various types and classes, can be adapted to work with databases. The practical significance lies in the automation of the process – the development

of a computer program that implements the proposed methodology. **Discussion:** further research is advisable to continue determining the best parameters of adaptive fuzzy neural production systems and fuzzy inference algorithms.

REFERENCES

1. Minyaev A.A., Budko M. Yu. Method for assessing the effectiveness of the information protection system of geographically distributed information systems. I*nformatization and communication*. 2017. No. 3. Pp. 119–121.
2. Budko M. Yu., Minyaev A.A. Methodology for assessing the effectiveness of the personal data protection system of the information system. *Problema kompleksnogo obespecheniya informacionnoj bezopasnosti i sovershenstvovanie obrazovatel'nyh tehnologij podgotovki specialistov silovyh struktur: Mezhvuzovskij sbornik trudov VI Vserossijskoj nauchno-tehnicheskoj konferencii* [The problem of comprehensive information security and the improvement of educational technologies for training specialists from law enforcement agencies: Interuniversity collection of works of the VI All-Russian scientific and technical conference, IKVO NRU ITMO, December 10, 2015]. 2016. Pp. 43–45.
3. Osako T. Proactive Defense model based on Cyber threat analysis. *FUJITSU. Sci. Tech. J.* 2016. No. 52 (3). Pp. 72–77.
4. Noor U. A machine learning-based FinTech cyber threat attribution framework using high-level indicators of compromise. *Future Generation Computer Systems*. 2019. Pp. 227–242.
5. Trifonov R. Artificial intelligence methods for cyber threats intelligence. *J. Comput*. 2017. Pp. 129–135.
6. Huang L. Adaptive strategic cyber defense for advanced persistent threats in critical infrastructure networks. *Perform. Eval. Rev.* 2018. No. 46 (2). Pp. 52–56.
7. Ushakov I.A., Kotenko I.V., Ovramenko A. Yu., Preobrazhensky A.I., Pelevin D.V. Combined approach to detecting insiders in computer networks. *Bulletin of the St. Petersburg State University of Technology and Design. Series 1: Natural and technical sciences.* 2020. No. 4. Pp. 66–71.
8. Ushakov I.A. Detection of insiders in a corporate computer network based on big data analysis technologies. *Bulletin of the Saint Petersburg State University of Technology and Design. Series 1: Natural and technical sciences*. 2019. No. 4. Pp. 38–43.
9. Savinov N.V., Tokareva K.A., Ushakov I.A., Krasov A.V., Sakharov D.V. Investigation of a data center network model based on Cisco ACI policies. *Information Security. Inside*. 2019. No. 4 (88). Pp. 32–43.
10. Livshitz I.I., Yurkin D.V., Minyaev A.A. Formation of the Instantaneous Information Security Audit Concept. *Communications in Computer and Information Science*. 2016. Vol. 678. Pp. 314–324.
11. Agrawal A., Ahmed C.M., Chang E. Poster. Physics-Based Attack Detection for an Insider Threat Model in a Cyber-Physical System. *In Proceedings of the 2018 on Asia Conference on Computer and Communications Security*, 2018. Pp. 821–823.
12. Mead N., Shull F., Vemuru K., Villadsen O.A. *Hybrid Threat Modeling Method.* Software Engineering Institute, Carnegie Mellon University. 2018. URL: http://resources.sei.cmu.edu/library/asset-view.cfm? AssetID=516617 (date of access 15.06.2020).
13. Khan R., McLaughlin K., Laverty D., Sezer Sakir. STRIDE-based Threat Modeling for Cyber-Physical Systems. *In Proceedings of the 2017 IEEE PES Innovative Smart Grid Technologies Conference Europe*.
14. Shevchenko N., Chick T.A., O'Riordan P., Scanlon T.P., Woody C.. Threat modelling: A summary of available methods. *Carnegie Mellon University Software Engineering Institute*. 2018. Pp. 1–24.
15. Yue Li, Teng Zhang, Xue Li, Ting Li. A Model of APT attack Defense Based On Cyber Threat Detection. *Communications in Computer and Information Science, Cyber Security,15th International Annual Conference*, CNCERT 2018. Pp. 122–134.
16. Cao X., Gong N.Z. Mitigating Evasion Attacks to Deep Neural Networks via Region-based Classification. *Proceedings of the 2017 Annual Computer Security Applications Conference (ACSAC)*. ACM, 2017. Pp. 278–287.
17. Kotenko I.V., Ushakov I.A., Pelevin D.V., Preobrazhensky A.I., Ovramenko A. Yu. Identifying Insiders on the Corporate Network: AUBA and UEBA Approach. *Protection of information. Inside*. 2019. No. 5 (89). Pp. 26–35.
18. Mohassel P., Zhang Y. SecureML: A System for Scalable Privacy Preserving Machine Learning. *Proceedings of the 2017 IEEE Symposium on Security and Privacy (S&P).* IEEE, 2017. Pp. 19–38.
19. Kuznetsov I.A., Lipatnikov V.A., Sakharov D.V. Management of ASMK organization of an integrated structure with forecasting the state of information security. *Electrosvyaz.* 2016. No. 3. Pp. 28–36.
20. Pronoza A.A., Vitkova L.A., Chechulin A.A., Kotenko I.V., Sakharov D.V. Methods for identifying channels of information dissemination in social networks. *Bulletin of St. Petersburg University. Applied math. Informatics. Management processes.* 2018. Vol. 14. Iss. 4. Pp. 362–377.

INFORMATION ABOUT AUTHOR:

Minyaev A.A., Senior Lecturer, The Bonch-Bruevich Saint-Petersburg State University of Telecommunications.

For citation: Minyaev A.A. Modeling information security threats in territorial-distributed information systems. *H&ES Research*. 2021. Vol. 13. No. 2. Pp. 52–65. Doi: 10.36724/2409-5419-2021-13-2-52-65 (In Rus)