Зубков Евгений Валерьевич. Алгоритмы и методики интеллектуального анализа событий информационной безопасности в сетях и системах телекоммуникаций: диссертация ... кандидата : 05.12.13 / Зубков Евгений Валерьевич;[Место защиты: ФГБОУ ВО Сибирский государственный университет телекоммуникаций и информатики], 2016

Федеральное государственное бюджетное образовательное учреждение высшего образования «Сибирский государственный университет телекоммуникаций и информатики»

На правах рукописи

Зубков Евгений Валерьевич

АЛГОРИТМЫ И МЕТОДИКИ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА СОБЫТИЙ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ В СЕТЯХ И СИСТЕМАХ ТЕЛЕКОММУНИКАЦИЙ

05.12.13 - Системы, сети и устройства телекоммуникаций

(технические науки)

Диссертация

на соискание ученой степени кандидата технических наук

ТОМ 1

Научный руководитель: доктор технических наук, профессор Белов Виктор Матвеевич

Новосибирск-2016

ОГЛАВЛЕНИЕ

ТОМ 1

ВВЕДЕНИЕ 6

1. МЕТОДИЧЕСКИЕ ОСНОВЫ ПРИМЕНЕНИЯ ИНТЕЛЕКТУАЛЬНОГО

АНАЛИЗА ДАННЫХ В ЗАДАЧАХ ОБНАРУЖЕНИЯ ВТОРЖЕНИЙ 12

1.1 Элементы интеллектуального анализа данных 12

1.2 Данные как материал интеллектуального анализа 13

1.3 Предпосылки к использованию МИАД в задачах сетевой ИБ 15

1.4 Обзор методов интеллектуального анализа данных 21

1.4.1 Наивный байесовский подход (Naive Bayes Approach) 21

1.4.2 Метод опорных векторов 24

1.4.3 Метод ближайших соседей (Nearest Neighbor) 26

1.4.4 Метод деревьев решений (Decision Trees) 27

1.4.5 Искусственные нейронные сети 29

1.4.6 Нечеткая логика 32

1.4.7 Г енетические алгоритмы 34

1.5 Исследования в области интеллектуального анализа данных для выявления

сетевых вторжений 37

1.6 Оценка эффективности методик обнаружения сетевых вторжений 41

1.7 Выводы 43

2. РАЗРАБОТКА КОМПЛЕКСНОЙ МЕТОДИКИ ИССЛЕДОВАНИЯ

СТАТИСТИКИ СИБ В СЕТЯХ ТЕЛЕКОММУНИКАЦИЙ 45

2.1 Трудности, связанные с исследованием статистики СИБ, и пути их

преодоления 45

2.2 Анализ структуры данных СИБ и возможности применения существующих

решений для их кластеризации 47

2.3 Элементы теории информации и энтропиный подход 50

2.4 Общие принципы выбранного для исследований подхода 54

2.5 Вычисление наиболее информативного признака (НИП) 57

2.6 Вычисление наиболее информативного значения (НИЗ) 60

2.7 Контроль однородности кластеров 66

2.8 Вычисление динамической зависимости СИБ 67

2.9 Оценка степени угрозы ИБ 71

2.10 Выводы 78

3. ВЫЧИСЛИТЕЛЬНЫЙ ЭКСПЕРИМЕНТ 81

3.1 Общие положения 81

3.2 Исходные данные 81

3.3 Кластеризация сетевых СИБ 82

3.3.1 Итерация 1: исследование множества А0 82

3.3.2 Итерация 2: исследование множества А0-1 91

3.3.3 Итерация 3: исследование множества А0-0 91

3.3.4 Итерация 4: исследование множества А0(2)-1 95

3.3.5 Итерация 5: исследование множества А0(2)-0 95

3.3.6 Итерация 6: исследование множества А0(3)-1 98

3.3.7 Итерация 7: исследование множества А0(3)-0 99

3.3.8 Итерация 8: исследование множества А0(4)-1 102

3.3.9 Итерация 9: исследование множества А0(4)-0 103

3.3.10 Итерация 10: исследование множества А0(5)-1 106

3.3.11 Итерация 11: исследование множества А0(5)-0 106

3.3.12 Итерация 12: исследование множества А0(6)-1 109

3.3.13 Итерация 13: исследование множества А0(6)-0 110

3.3.14 Итерация 14: исследование множества А0(7)-1 113

3.3.15 Итерация 15: исследование множества А0(7)-1-1 115

3.3.16 Итерация 16: исследование множества А0(7)-1-0 116

3.3.17 Итерация 17: исследование множества А0(7)-0 116

3.3.18 Итерация 18: исследование множества А0(8)-1 119

3.3.19 Итерация 19: исследование множества А0(8)-0 119

3.3.20 Итерация 20: исследование множества А0(9)-1 122

3.3.21 Итерация 21: исследование множества А0(9)-0 122

3.3.22 Итерация 22: исследование множества А0(10)-1 124

3.3.23 Итерация 23: исследование множества А0(10)-0 125

3.3.24 Итерация 24: исследование множества А0(11)-1 127

3.3.25 Итерация 25: исследование множества А0(11)-0 128

3.3.26 Результаты кластеризации СИБ 128

3.4 Вычисление динамической зависимости СИБ 130

3.5 Оценка угрозы зафиксированных СИБ 132

3.6 Выводы 137

4. ПРИКЛАДНОЕ ПРИМЕНЕНИЕ РАЗРАБОТАННЫХ АЛГОРИТМОВ И

МЕТОДИК 138

4.1 Значение прикладного применения МИАД при исследовании сетевых

СИБ 138

4.2 Структура ПАК 139

4.3 Вычислительная подсистема 140

4.4 Подсистема хранения данных 141

4.5 Подсистема визуализации 141

4.5.1 Главное окно

4.5.2 Окно детализации 146

4.6 Результаты опытно-экспериментальных испытаний 148

4.7 Перспективы дальнейшего развития ПАК 153

4.8 Выводы 155

ЗАКЛЮЧЕНИЕ 157

СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ 159

СЛОВАРЬ ТЕРМИНОВ 160

СПИСОК ЛИТЕРАТУРЫ 161

СПИСОК ИЛЛЮСТРАТИВНОГО МАТЕРИАЛА 178

ТОМ 2

ПРИЛОЖЕНИЕ 1 3

ПРИЛОЖЕНИЕ 2 10

ПРИЛОЖЕНИЕ 3 11

ПРИЛОЖЕНИЕ 4 18

ПРИЛОЖЕНИЕ 5 124

ПРИЛОЖЕНИЕ 6 179

ПРИЛОЖЕНИЕ 7

ЗАКЛЮЧЕНИЕ

В результате проведенных исследований были получены следующие основные результаты:

1. Впервые выбраны критерии МИАД для анализа статистики сетевых СИБ, заключающиеся в способности анализировать большие массивы данных, характеризующиеся малым количеством признаков СИБ и высокой вариативностью их значений с использованием номинальных шкал.
2. Впервые разработаны алгоритм и методика кластеризации данных о СИБ, зафиксированных СОВ, на основе энтропийного подхода. Результатом их работы является представление данных в виде однородных групп (кластеров), каждая из которых соответствует предварительно заданным параметрам.
3. Впервые разработаны алгоритм и методика оценки динамической зависимости между кластерами. Их применение позволяет получить числовую величину, характеризующую степень согласованности по времени СИБ различных кластеров. Важно отметить, что данные приобретают новый признак, измеренный в шкале отношений, который может рассматриваться в качестве эквивалента расстояния между кластерами и потенциально расширяет спектр доступных МИАД для проведения дальнейших исследований.
4. Впервые разработаны алгоритм и методика оценки степени текущей угрозы на основе данных, прошедших предварительную экспертную оценку. Для этого введены две характеристики: степень изученности СИБ кластера и показатель непротиворечивости экспертной оценки. Дано математическое обоснование соответствия этих величин решаемой задаче. Согласованность результата и исходных параметров подтверждена вычислительным экспериментом.
5. Впервые разработана структура ПАК для исследования статистики СИБ, основанная на предложенных МИАД, и выполнена его программная реализация. ПАК состоит из трех подсистем (вычислительной, хранения данных и визуализации) и решает задачи кластеризации данных о СИБ, вычисления динамической зависимости между полученными кластерами и взаимодействия с БД. Для работы с результатами вычислений предусмотрен кроссплатформенный графический интерфейс, позволяющий работать с данными в интерактивном режиме.
6. Проведены вычислительный эксперимент и серия опытно - экспериментальных испытаний ПАК. Согласованность полученных результатов с теоретическими положениями подтверждает правильность и достоверность алгоритмов и методик, положенных в основу ПАК.

Диссертационная работа обладает научным и практическим потенциалом. Кластеризованные данные, представленные в форме самостоятельных статистических сущностей и обладающие набором дополнительных признаков, являются материалом для дальнейших исследований, в том числе с помощью МИАД, использующих расстояние между сущностями в качестве основной метрики. Существует возможность добавления прогностических функций за счет реализации контура обратной связи, что в целом приведет к повышению уровня автоматизации ПАК. Кроме того, дополнительно проведенные испытания показали потенциальную возможность использования разработанного ПАК для исследования трафика