На правах рукописи

Большев Александр Константинович

Алгоритмы преобразования и классификации трафика для обнаружения вторжений в компьютерные сети

05.13.11 – Математическое обеспечение вычислительных машин, комплексов и компьютерных сетей
05.13.19 – Методы и системы защиты информации, информационная безопасность

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук

Санкт-Петербург – 2011

Работа выполнена в *Санкт-Петербургском государственном электротехническом университете «ЛЭТИ» им. В.И. Ульянова (Ленина) (СПбГЭТУ) на кафедре «Математическое обеспечение ЭВМ».*

Научный руководитель: *доктор технических наук, профессор, Лисс Александр Рудольфович*

Официальные оппоненты: *профессор, доктор технических наук, профессор кафедры «Вычислительная техника» СПбГЭТУ, Водяхо Александр Иванович*

профессор, доктор технических наук, профессор Санкт-Петербургского государственного университета, Карпов Андрей Геннадьевич

Ведущая организация: *Санкт-Петербургский государственный университет телекоммуникаций им. проф. М. А. Бонч-Бруевича*

Защита состоится «21» декабря 2011 г. в 15 часов на заседании совета по защите докторских и кандидатских диссертаций *Д212.238.01* при *СПбГЭТУ «ЛЭТИ»*, расположенном по адресу: *197376, Россия, Санкт-Петербург, улица Профессора Попова, дом 5*

С диссертацией можно ознакомиться в библиотеке *СПбГЭТУ «ЛЭТИ»*.

Автореферат разослан «11» ноября 2011 г.

Ученый секретарь
совета по защите докторских и
кандидатских диссертаций *Д212.238.01*,
к.т.н. *Щеголева Н.Л.*

Общая характеристика работы

Актуальность работы. Обнаружение сетевых атак является в данный момент одной из наиболее острых проблем сетевых технологий. По данным DARPA, незащищенный компьютер, подключенный к сети Интернет, будет взломан не позднее, чем через 2-3 часа. Масштабные эпидемии сетевых червей, DDoS атаки с бот-сетей размером более 10000 компьютеров, автоматизированные средства поиска уязвимостей в сетях – все это делает обеспечение безопасности локальных сетей весьма трудоемким делом. Сейчас трудно найти сеть, в которой отсутствуют такие активные средства предупреждения атак как антивирус, брандмауэр, системы предупреждения вторжений уровня хоста и так далее. К сожалению, одних активных средств отражения атак недостаточно. Поэтому, в дополнение к ним применяют пассивные средства борьбы с атаками – сетевые системы обнаружения вторжений.

Сетевые системы обнаружения вторжений (ССОВ) просматривают весь сетевой трафик (или трафик определенного участка сети) и при обнаружении каких-либо отклонений в нем сигнализируют об этом. Формальные ССОВ работают по принципу антивирусной программы – пакеты, попадающие на сенсоры, сравниваются с БД сигнатур и, в случае обнаружения совпадения, объявляется тревога. К сожалению, даже формальных ССОВ становится недостаточно для надежной защиты сети. По данным CERT, количество известных новых методов вторжения только за 2010 год превысило 25000. Это значит, что в среднем, каждый день появляется порядка 70 новых атак. Физически невозможно обновлять БД сигнатур формальных ССОВ за такие промежутки времени. Кроме того, увеличение объема сигнатур отрицательно сказывается на производительности систем. Решением этой проблемы является применение систем обнаружения вторжений на основе выявления аномальной активности или эвристических ССОВ.

На данный момент существует достаточно большое количество эвристических ССОВ, работающих на прикладном уровне OSI. В области обнаружения вторжений на сетевом/транспортном уровнях до сих пор не предложено ни одной системы, способной работать в реальном времени.

Цель и задачи исследования – разработка модели обнаружения вторжений на сетевом/транспортном уровнях и построение эвристической сетевой системы обнаружения вторжений на основе полученной модели.

Для достижения поставленных целей были решены следующие задачи:

1. классификация и анализ архитектуры современных систем обнаружения вторжений;

2. исследование и анализ существующих моделей, методов и систем эвристического обнаружения вторжений, выбор основных критериев оценки эвристических методов обнаружения вторжений, оценка существующих методов и систем;

3. разработка представления трафика в виде пространства векторов; разработка набора быстрых алгоритмов, реализующего такое преобразование;

4. исследование и выбор численных методов сокращения размерности полученного пространства;

5. выбор и настройка метода извлечения знаний и формирования базы знаний о трафике целевой сети;

6. разработка модели эвристической системы обнаружения вторжений на основе полученных алгоритмов и выбранных методов;

7. исследование и оценка полученной модели;

8. разработка комплекса программ реализующих модель и исследование его работоспособности в реальных сетях.

Объектом исследования диссертационной работы является процесс обнаружения вредоносных действий и аномальных явлений на основе анализа трафика в компьютерной сети.

Предметом изучения является набор моделей, эвристических методов и алгоритмов, обучающихся на положительном и/или смешанном сетевом трафике и предназначенных для обнаружения вторжений и аномальных явлений в компьютерных сетях.

Методы исследования. Теоретическая часть работы выполнена на основе методов математической статистики и функционального анализа, методологии извлечения знаний и искусственного интеллекта. В экспериментальной части работы применяются численные методы, алгоритмы и методы извлечения знаний. Для выполнения экспериментальной части созданы программные комплексы IceIDS2 и IceIDS2s.

Научная новизна полученных результатов заключается в следующем:

1. систематизированы существующие методы обнаружения вторжений в сеть, обучающиеся только на положительном или смешанном трафике;

2. разработан набор алгоритмов быстрого преобразования трафика в множество векторов, пригодных для извлечения признаков вторжений;

3. разработана модель системы обнаружения вторжений на основе быстрого алгоритма преобразования трафика, одноклассового классификатора на базе искусственных нейронных сетей (ИНС) и методов сжатия пространства данных;

4. разработана методика определения аномалий на основе метода главных компонент (МГК) и ИНС, позволяющая обнаруживать нехарактерные явления в динамических системах, методика может применяться и вне области обнаружения вторжений в сеть;

5. разработан комплекс программ для обнаружения вторжений в компьютерные сети, основанный на предложенной модели и алгоритмах.

Практическая значимость. Практическую значимость имеют полученные автором следующие результаты:

- Набор алгоритмов быстрого преобразования трафика в множество векторов.

- Методика выявления аномалий в динамических системах.

- Программные комплексы обнаружения вторжений IceIDS2 и IceIDS2s.

Результаты работы внедрены в СПбГЭТУ «ЛЭТИ» на кафедре «МО ЭВМ», в сетях компаний ООО «Торговый дом «Китай» и ООО «Некки».

На защиту выносятся следующие основные результаты и положения:

1. способ представления трафика в виде набора векторов, пригодных для последующего использования в методах классификации;

2. набор алгоритмов преобразования трафика в множество векторов;

3. модель эвристической системы обнаружения вторжений на основе разработанных алгоритмов;

4. комплекс программ, позволяющий обнаруживать вторжения и аномальные явления в реальных компьютерных сетях.

Апробация работы. Основные результаты, полученные в ходе работы над диссертацией представлены на:

- Научно-технических конференциях профессорско-преподавательского состава СПбГЭТУ, Санкт-Петербург, 2009 и 2011 гг.

- XV Международной конференции «Современное образование: содержание, технологии, качество.», СПбГЭТУ «ЛЭТИ» 2009 г.

- Всероссийской Конференции «ИНФОС-2009», Вологодский Государственный Технический Университет, 2009 г.

- Конференции «Технологии Microsoft в теории и практике программирования», Санкт-Петербург, 2010 г.

Публикации. Основные теоретические и практические результаты диссертации опубликованы в 10 статьях и докладах, среди которых 4 публикации в ведущих рецензируемых изданиях, рекомендованных в действующем перечне ВАК. Доклады доложены и получили одобрение на 5 международных, всероссийских и межвузовских научно-практических конференциях, перечисленных в конце автореферата. Система IceIDS2 зарегистрирована в качестве программного средства (свидетельство о регистрации ПС IceIDS2 №2009614325).

Структура и объем диссертации. Диссертация состоит из введения, 4 глав, заключения и библиографии. Общий объем диссертации 155 страниц, из них 146 страниц текста, включая 43 рисунка. Библиография включает 88 наименований на 9 страницах.

Содержание работы

Во Введении обоснована актуальность диссертационной работы, сформулирована цель и аргументирована научная новизна исследований, показана практическая значимость полученных результатов, представлены выносимые на защиту научные положения.

Первая глава посвящена анализу современного состояния области обнаружения вторжений и в частности эвристического обнаружения вторжений в сеть, выработке критериев оценки эвристических ССОВ и исследованию существующих систем.

Процесс осуществления угроз информационной системе получил название «атака» или «вторжение». Атака – любое действие нарушителя, направленное на нарушение заданной функциональности вычислительной системы или получение несанкционированного доступа к информационным, вычислительным или сетевым ресурсам. Атака, как и любое действие, имеет свой жизненный цикл, разделяющий ее на этапы подготовки, вторжения, атакующего воздействия и развития атаки.

Система обнаружения вторжений (Intrusion Detection System, IDS) - программное или аппаратное средство, предназначенное для выявления фактов неавторизованного доступа в компьютерную систему или сеть, либо несанкционированного управления ими.

Современные методы обнаружения вторжений базируются на двух принципах: сигнатурный (формальный, описывающий каждую атаку особой моделью или сигнатурой) и эвристический (обнаружение аномалий, базирующийся не на моделях информационных атак, а на моделях штатного функционирования наблюдаемой информационной системы).

Сетевая СОВ (Network-based IDS, NIDS) отслеживает вторжения, проверяя сетевой трафик, захватываемый на её сенсорах. ССОВ условно разделяют на два класса систем: работающие на уровне приложений модели OSI (обнаруживают вторжения на основе анализа поля данных пакета) и работающие на

сетевом/транспортном уровне модели OSI (обнаруживают вторжения анализируя управляющие пакеты протоколов IP, TCP и др., а также заголовки пакетов с данными).

Можно выделить следующие наиболее важные критерии оценки эффективности работы эвристических ССОВ (HNIDS):

1. CR – количество корректно распознанных аномальных и нормальных пакетов; здесь также будет корректно предположить, что любые атаки обычно не входят в нормальный трафик сети и классифицируются как аномальные;

2. FP (False Positive, ложная тревога) – количество нормальных пакетов принятых за аномальные;

3. PPs (Packet per second, пакетооборот) – максимальное количество пакетов, которое система может обработать за 1 секунду на этапе тестирования;

4. η (устойчивость системы) – процент отрицательных векторов в обучающей выборке при котором система начинает работать нестабильно;

5. FN (False Negative) – количество аномальных пакетов, принятых за нормальные.

В условиях реальных современных сетей на применение HNIDS накладываются особые требования, связанные с высокими уровнями трафика (большими и сверхбольшими показателями пакетооборота в сети). Во-первых, скорость этапа тестирования имеет наивысший приоритет. Во-вторых, при проверке большого количества пакетов в секунду, любая FP генерит сообщение в журналах аномальностей. Если значение FP системы достаточно велико, то журналы системы очень быстро заполнятся ошибками распознавания и восприятие человеком настоящих аномалий в этом шуме будет сильно затруднено.

Другой важный момент заключается в том, что мы изначально не сможем разделить тренировочные данные на нормальные и аномальные (далее – положительные и отрицательные). То есть, тренировочная выборка либо может состоять целиком из данных, которые мы считаем положительными (или отрицательными), либо считается, что тренировочная выборка – смешанная.

В таблице 1 указаны существующие HNIDS, которые удовлетворяют указанным выше условиям.

На основании проведенного обзора, можно сделать следующие выводы:

- ни одна из рассмотренных систем или моделей не способна работать в сетях с высоким уровнем трафика из-за больших значений FP;

- за исключением fpMAFIA, OTAD и SPADE рассмотренные методы не предназначены для работы с большими объемами обучающих выборок

Таблица 1

Существующие HNIDS

Метод	Данные	Основа	CR(%)	FP(%)
SPADE	реальные, $\oplus \cup \ominus$	временные закономерности	∼70	∼0.02
OTAD	реальные, $\oplus$	одноклассовый SVM	59.1	3.1
Геометрический подход Арнольда и Эскина	модельные, $\oplus$	K-ближайших соседей	89	10
fpMAFIA	реальные, $\oplus$	адаптивная решетка	90.2	5.4
DT-SVM	реальные, $\oplus$	деревья решений, SVM	94.5	0.3
Кластеризация по Эскину	модельные, $\oplus \cup \ominus$	single linkage clustering	65.7	0.178

(например, в небольшой корпоративной сети показатель в 11 гигабайт трафика в день (или порядка 2-3 сотен тысяч пакетов) является нормой, но большинство методов не готовы работать с выборками такого размера);

- время обучения вляется слишком большим и не подходит для реальных условий;
- большинство моделей рассчитаны на обнаружение только узкого класса атак и не способны к обнаружению других аномалий.

Следовательно, разработка новых методов обнаружения на основе обучения только на положительном или смешанном трафике является актуальной и важной задачей исследования.

Вторая глава описывает выбор признаков вторжений из трафика, разработку алгоритмов извлечения этих признаков из «сырых» пакетов с дальнейшим преобразованием в набор векторов и методику сокращения размерности полученного множества.

Самыми распространенными протоколами на сетевом и транспортном уровнях модели TCP/IP являются: TCP, UDP, ICMP, IP; именно они и будут рассмотрены. Подавляющее большинство атак осуществляется с использованием только этих пакетов. Другие протоколы (например, IGMP) будут отображаться на вектор как IP-пакет.

Если рассмотреть различные методы вторжений в сеть на сетевом или транспортном уровнях, то можно легко заметить, что в подавляющем большинстве случаев оно «растянуто» и включает в себя несколько TCP-сегментов между двумя оконечными точками, т.е. является сессией. Поэтому, не имеет смысла выделять признаки из отдельных пакетов TCP, необходимо выделять признаки из TCP-сессии. Стоит отметить, что даже для дейтаграммы UDP и сессии TCP некоторые поля IP-пакета, содержащие сегмент или дейтаграмму, будут отображены на вектор совместно со свойствами сессии или дейтаграммы.

В результате классификации вторжений, а также исследования современных атак на транспортном и/или межсетевом уровне сети, были выбраны следующие признаки, которые будут извлекаться из трафика для пакетов или сессий:

- Общие: протокол, характеристики фрагментации, TTL, ToS, количество отправленных/полученных байт, является ли широковещательным, IP-опции, корректность CRC и др.
- ICMP-пакетов: код и тип ICMP-сообщения.
- UDP-пакетов и TCP-сессий: сервис, land (равен ли порт клиента порту серверу).
- TCP-сессий: продолжительность, количество флагов в сессии, менялся ли размер окна, встречался ли нулевой sequence, количество пакетов, количество сессий у того же сервиса, отношения отправленных/полученных к общему количеству байт сессии, количество байт под опции TCP, тип и версия ОС инициатора сессии, MSS и др.

Всего было выделено 45 признаков вторжений, которые составили вектор признаков для эвристического обнаружения вредоносной или аномальной активности.

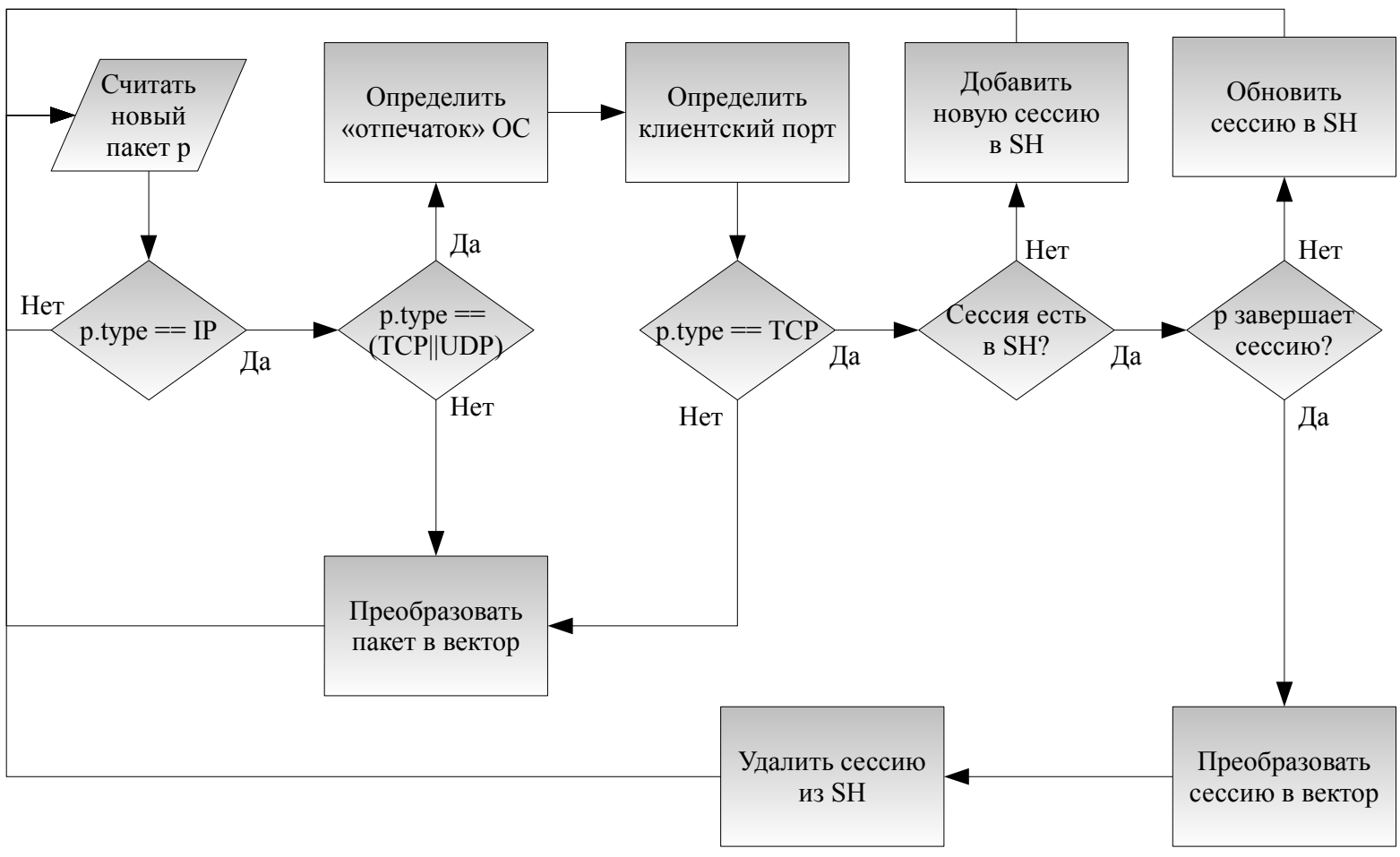

Рисунок 1 – Блок-схема алгоритма преобразования TEA1

Одна из основных проблем состоит в том, что если HNIDS «не успела» захватить момент начала соединения между сервером и клиентом, то весьма

трудно определить какая из сторон сессии или UDP-дейтаграммы является сервером, а какая – клиентом. Механизм выделения клиентских портов в стеке TCP/IP устроен таким образом, что при соединении с сервером клиентский порт выбирается случайным образом из довольно большого интервала (или интервалов). Так как в самом пакете не содержится какой-либо специфичной для транспортного уровня информации, которая могла бы точно указать, какая из сторон является сервером, необходимы некоторые эвристики для такого определения. Во-первых можно определять максимальный серверный порт в конкретной сети. Во-вторых можно использовать предопределенный список общепринятых или частных для данной организации серверных портов в сети. Помимо этого, т.к. интервал клиентских портов характерен для различных видов стеков TCP/IP, определив тип и версию операционной системы, можно упростить определение стороны сервера в сессии.

Для решения задачи определения типа и версии операционной системы (ОС) удаленного хоста разработан алгоритм 1-POSD, который определяет тип сетевого стека основываясь на единичном пакете. Сложность алгоритма $O(1)$.

Обобщенная блок-схема алгоритма преобразования трафика в набор векторов (TEA1) представлена на рисунке 1. Его особенности:

- сессии хранятся в специальной структуре данных Session Data (SD), представляющую из себя несколько хэш-таблиц с перекрестными ссылками друг на друга и хэшем int32 (используется модифицированная автором функция Роберта Дженкинса);
- сложность преобразования одного UDP/IP/ICMP-пакета в вектор или обновления TCP-сессии равна $O(1)$; сложность преобразования одного TCP-пакета в вектор (или добавления сессии) равна $O(log_2^2(n))$, где n – число TCP-сессий в сети в текущий момент времени.

В результате работы алгоритма, мы имеем набор векторов размерности $M = 45$. При больших выборках, такая размерность может оказать решающее влияние на производительность системы. Кроме того, очень вероятно, что в полученной выборке содержится лишняя информация. Следовательно, необходимо найти способ «свернуть» пространство выборки и сократить её размерность.

Метод главных компонент (МГК, англ. Principal component analysis, PCA) – один из основных способов уменьшить размерность данных, потеряв наименьшее количество информации. Вычисление главных компонент сводится к вычислению собственных векторов и собственных значений ковариационной матрицы исходных данных.

Пусть имеется матрица переменных X размерностью $N \times M$, где N – число образцов (строк), а M – это число независимых переменных (столбцов), которых, как правило, много ($M >> 1$). В методе главных компонент используются новые, формальные переменные $v_a (a = 1, \ldots, A)$, являющиеся линейной комби-

нацией исходных переменных $x_m (m = 1, \ldots, M)$:

$$v_a = p_{a1} x_1 + \cdots + p_{aM} x_M$$

С помощью этих новых переменных матрица X разлагается в произведение двух матриц V и P:

$$X = V P^* + E = \sum_{a=1}^{M} v_a p_a + E, \tag{1}$$

матрица V называется матрицей счетов, ее размерность – $N \times A$, матрица P называется матрицей нагрузок, ее размерность $M \times A$, E – это матрица остатков, размерностью $N \times M$.

Новые переменные v_a называются главными компонентами. Число столбцов – t_a в матрице T, и p_a в матрице P – равно A, которое называется числом главных компонент PC. Эта величина заведомо меньше числа переменных M и числа образцов N.

Важным свойством МГК является ортогональность (независимость) главных компонент. Поэтому матрица счетов V не перестраивается при увеличении числа компонент, а к ней просто прибавляется еще один столбец – соответствующий новому направлению. Тоже происходит и с матрицей нагрузок P.

Преобразование новых данных X_{new} в пространство новой размерности выполняется произведением:

$$V_{new} = X_{new} P^*$$

Таким образом, мы получаем пространство новой размерности $M' < M$. Имея набор главных компонент, мы можем «свернуть» пространство любого вектора, полученного на выходе алгоритма преобразования, как на этапе обучения, так и при тестировании.

По сравнению с независимым анализом компонент (ICA) или нелинейными методами сокращения размерности пространства, МГК обеспечивает более высокую скорость работы и возможность распараллеливания (например с применением GPGPU).

Третья глава посвящена разработке модели обнаружения аномалий поведении динамического объекта и модели эвристической системы обнаружения вторжений.

Компьютерную сеть, в который установлены сенсоры системы обнаружения вторжений, можно представить в виде динамического объекта. В каждый момент времени объект генерирует сигналы (пакеты). Когда сеть работает в нормальном режиме, сигналы поддаются описанию с помощью неких закономерностей. В случае, когда в сети происходит аномальное событие (связанное со вторжением или неполадками в работе аппаратных/программных систем), характер генерации пакетов изменяется. Предположим, что за некоторый промежуток времени (период обучения) мы можем накопить информацию (в виде набора векторов) которая представляет из себя весь (или почти весь) спектр

сигналов поведения объекта. Тогда, выработав закономерности на основании имеющегося набора, можно на этапе тестирования сравнивать новые пакеты с выработанными закономерностями и в случае обнаружения несоответствий сообщать об обнаружении аномалии.

Стоит отметить, что мы можем гарантировать лишь то, что в обучающей выборке основная часть векторов будет положительной (неаномальной). Возможно, там будут присутствовать некоторые шумы, но их количество не будет значительным. Накопить две одинаковые выборки положительного (характерного для сети, неаномального) и отрицательного (аномального) трафика в реальных условиях невозможно. Для выработки закономерностей в таких случаях уместно применять методы кластеризации или классификации с обучением на одном классе.

Задача классификации с обучением на одном классе (one-class classification, одноклассовой классификации) формулируется следующим образом: пусть A — полное множество объектов, а X — некоторое подмножество A. Существует отображение $y^* : X^m \to 1, P(X^m \in X) \simeq 1$. Требуется построить алгоритм $a : X \to 1$ и $a : (A \backslash X) \to 0$. В терминах предметной области обнаружения вторжений это выглядит следующим образом: имея обучающую выборку пакетов положительного трафика X^m, построить алгоритм, который переведет попадающий на сенсоры пакет в 1, если он неаномальный и в 0, если он аномальный (см. рисунок 2).

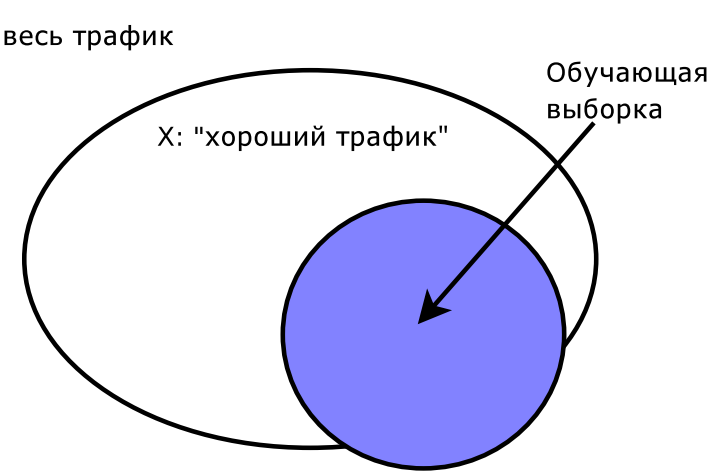

Рисунок 2 – Классификация с обучением на одном классе

Исследование применимости классификации с обучением на одном классе к обнаружению аномалий в сети (с использованием методов одного ближайшего соседа (1-Nearest neighbour method) и одноклассового SVM) на эталонных данных KDD Cup '99 Data показало, что даже без предварительной настройки можно получить вполне конкурентоспособные (с методами из таблицы 1) результаты по критериям CR и FP. Тем не менее, эти методы весьма чувствительны к шумам в обучающей выборке, требуют больших ресурсов на этапе обучения и их использование затруднено на больших выборках (>100000 векторов).

Одним из способов реализации одноклассового классификатора является

использование многослойного перцептрона в качестве адаптивного фильтра. Многослойный перцептрон – это искусственная нейронная сеть (ИНС) прямого распространения. Ниже указан несколько модифицированный автором метод классификации на основе многослойного перцептрона. Пусть существует набор векторов V_n, каждый из которых является положительным и имеет размерность m. Выберем некую метрику D которая описывает расстояние между векторами (в качестве расстояния можно взять Евклидово или Чебышева). Построим ИНС прямого распространения с m входными нейронами, h нейронами скрытого слоя и m выходными нейронами, при этом скрытый слой имеет сигмоидальную функцию активации, выходной – линейную. Для его обучения по методу обратного распространения будем использовать обучающую выборку X^n вида $(x_1, x_1), ..., (x_n, x_n)$. То есть, построенная ИНС будет работать как *адаптивный фильтр*, который, получив на вход сигнал (вектор) должен без искажений передавать его на выход. Таким образом для векторов, *похожих* на вектора обучающей выборки расстояние $D(x_i, y_i) \to 0$, где y_i – выходной вектор ИНС при получении на вход вектора x_i. После обучения, если выбрана нестандартная метрика D нужно пропустить всю обучающую выборку через ИНС и получить пороговое значение $thres = (1 + \theta) \max_{\forall i \in (1,n)} D(x_i, y_i)$, где $\theta \in (-1, 1)$ – пороговый коэффициент. Далее, чтобы узнать, является ли произвольный вектор x положительным, достаточно получить посредством ИНС вектор y и проверить – не превышает ли порог $thres$ значение $D(x, y)$. Видно, что после обучения, ИНС будет эмулировать *тождественное отображение* (с неким максимальным пороговым искажением $thres$) для векторов из положительного множества; вектора из отрицательного множества будут искажаться гораздо больше.

Три основных параметра, которые напрямую влияют на эффективность работы такого одноклассового классификатора, это:

- h – количество нейронов скрытого слоя напрямую влияет на объем т.н. «памяти» нейросети; если h будет больше или равно m, то возникнет опасность «переобучения», то есть сеть напрямую приблизиться к тождественному отображению и будет любые данные поступившие на вход передавать с минимальным изменением на выход; если же h будет слишком маленьким, то возможно «недообучение» сети, то есть положительный класс будет не до конца определен;

- η – коэффициент скорости обучения (LR, learning rate), влияет как на скорость обучения, так и на эффективность «запоминания»; при больших η сеть может никогда не обучиться, при малых – обучаться слишком долго или «переобучиться»;

- μ – коэффициент инерционности (LM, Learning Momentum), влияющий как на скорость обучения, так и на качество (performance) обучения.

Окончательная схема модели эвристической ССОВ представлена на рисунке 3. Модель имеет два состояния: режим обучения и режим тестирования (работы, основной режим). При обучении, на вход модели попадает некое количество «сырого» эталонного трафика (трафика, характерного для данной сети, захваченного в ней на протяжении достаточно длительного промежутка времени); при помощи алгоритма TEA1 из трафика получается набор векторов X^M. После чего, необходимо найти разложение на главные компоненты с помощью МГК по формуле (1). На наборе V новой размерности $M' < M$ уже можно проводить обучение искусственной нейронной сети. Обучение проводится при помощи АОР таким образом, чтобы сигнал на входе сети соответствовал сигналу на выходе («адаптивный фильтр»)

$$W_{\text{ИНС}} = \text{АОР}(\text{ИНС}, V, V)$$

, где $W_{\text{ИНС}}$ – матрица весов обученной ИНС. После обучения, можно найти значение порога $thres$ (максимального расстояния между сигналами на входе и выходе нейронной сети для всей обучающей выборки):

$$thres = \max_i D(\text{ИНС}(v_i), v_i)$$

, где v_i – i-й вектор обучающей выборки, D – метрика расстояния. В данном случае, в качестве метрики расстояния используется евклидово. Весь этот процесс отображен на верхней части рисунка 3.

На этапе тестирования, сенсор захватывает из сети пакет, преобразует его в вектор, отображает в новое пространство сокращенной размерности и определяет при помощи ИНС его аномальность. Для нового вектора x_{new}, полученного из пакета при помощи TEA1, этот процесс будет выглядеть как

$$d = D(\text{ИНС}(x_{new}P^*), x_{new}P^*), d \leq thres$$

, здесь в случае $d \geq thres$ вектор x_{new} (и соответствующий ему пакет) – аномален.

Полученная модель может быть адаптирована к использованию в других областях, например в обнаружении аномалий сигналов от сложных динамических объектов; в работе проведено исследование, показавшее достаточную эффективность подобного подхода.

Четвертая глава посвящена программной реализации эвристической системы обнаружения вторжений, исследованию работоспособности и эффективности предложенной системы.

Для исследования эффективности предложенных методов было исследовано две модели – модель на данных KDD и модель на данных DARPA. Обе модели тестировались в двух вариациях – с использованием сокращения размерности и без него. На основе моделей разработано два варианта сетевой системы обнаружения вторжений.

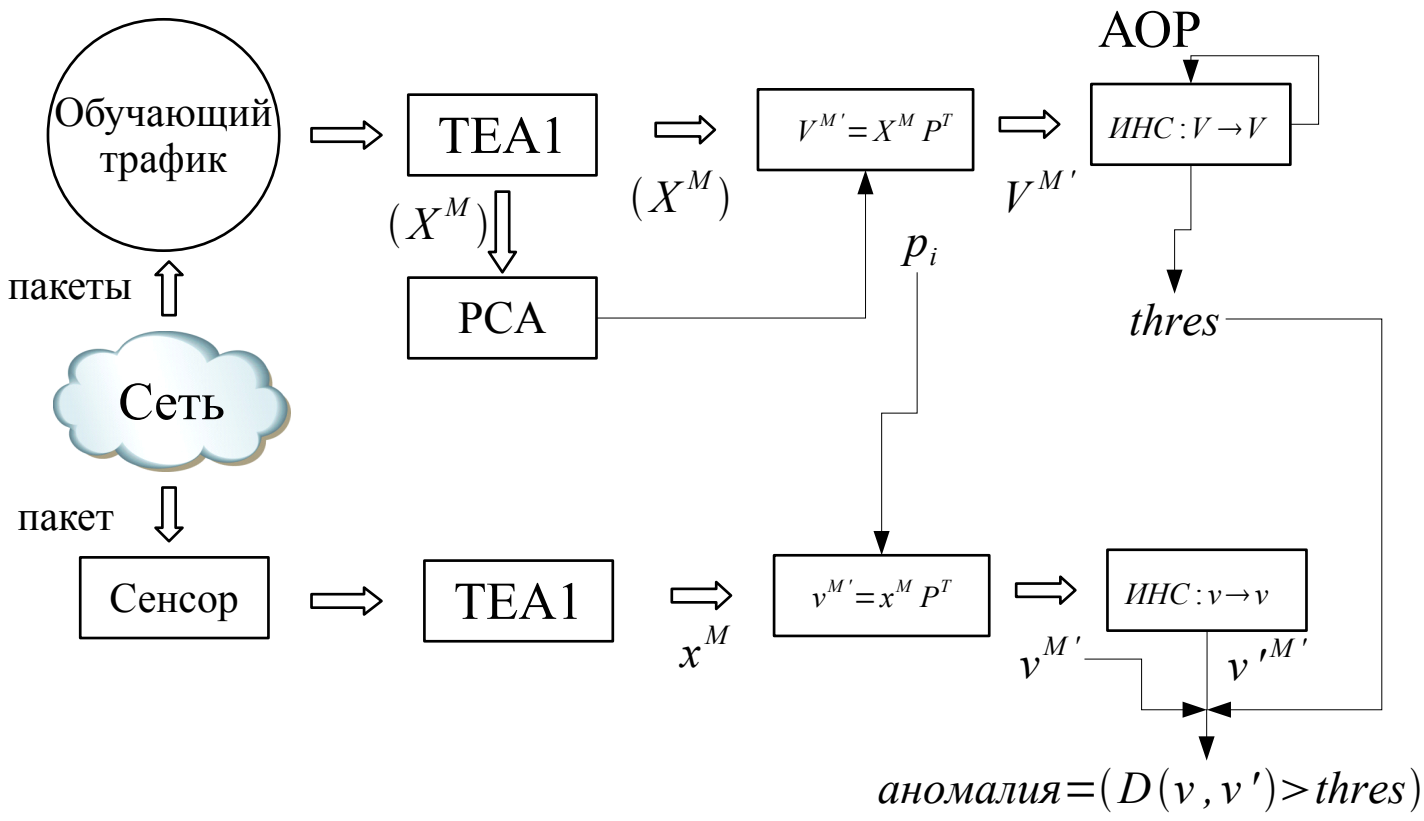

Рисунок 3 – Схема работы модели ССОВ на базе TEA1 и ИНС

Модель на данных KDD без сокращения размерности данных показала наилучший результат: CR = 0.967 и $FR = 1.33 * 10^{-5}$. При этом коэффициент инерционности слабо влияет на результат обучения и чем большей «памятью» (количеством нейронов внутреннего слоя) обладает ИНС, тем меньшим будет порог и тем лучшим будет качество обнаружения аномалий. При внесении в модель модуля сокращения размерности на базе МГК было показано, что наилучшие результаты по обнаружению вторжений происходят при 40-50% компонент в пространстве меньшей размерности. Основные зависимости относительно η, μ, h_l, выявленные при исследовании модели без МГК, сохраняются и в модели с МГК.

Сравнение полученных на данных KDD Cup '99 результатов с результатами моделей других авторов на тех же данных показывает (см. табл. 2), что практически всегда при фиксации одного из критериев, модель ИНС/МГК по другим критериям оказывается лучше аналогов. Единственное исключение – метод fpMAFIA, но он проигрывает модели по критериям производительности и устойчивости к шумам в обучающей выборке.

На данных DARPA IDS Evaluation модель без использования МГК показала 17(из 18 всего) распознанных атак при 4 ложных срабатываниях; модель на базе алгоритма TEA1 и МГК – 18(из 18 всего) распознанных атак при 10 ложных срабатываниях.

Система IceIDS2 не использует МГК и разработана на языке Java, с использованием библиотек FANN (реализация ИНС) и jPCAP (низкоуровневаня работа с сетью). IceIDS2 использует упрощенную версию алгоритма TEA1, работающую с векторами состоящими из 27 компонент. Разработанная система

Таблица 2

Результаты исследования

Модель(h_l, η, μ, p_c)	CR	FP	FN
ИНС(0.7,0.07,0.9,0.4)	0.96142	0.00003	0.03855
Геометрический подход	0.98	0.1	-
ИНС(0.8,0.05,0.1,0.4)	0.98	0.01	0.0008
Геометрический подход	0.28	0.01	-
ИНС(0.5,0.01,0.01,0.4)	0.95	0.01	0.03
Кластеризация	0.667	0.021	-
ИНС(0.7,0.005,0.1,0.4)	0.668	0.001	0.329
ИНС(0.5,0.03,0.01,0.5)	0.95	0.021	0.028
Zhang & Zulkernine	0.67	0.04	-
ИНС(0.9,0.05,0.9,0.4)	0.672	0.001	0.327
ИНС(0.7,0.001,0.9,0.7)	0.93	0.04	0.02
fpMAFIA	0.902	0.054	-
ИНС(0.9,0.02,0.9,0.8)	0.902	0.077	0.019
ИНС(0.4,0.085,0.1,0.4)	0.729	0.054	0.217

зарегистрирована как программное средство (Свидетельство о регистрации ПС IceIDS2 №2009614325). В процессе испытаний в реальной сети система обнаружила 15 попыток вторжения при 12 ложных срабатываниях. Скорость обработки пакетов в некоторые моменты времени достигала 252 килопакетов в секунду на машине Xeon 2.8GHz, 2048Mb RAM. На обучение IceIDS2 затратила более 51 часа реального времени.

Вторая система (IceIDS2s) реализована на языке C# и включает в себя блок сокращения размерности пространства данных, а также полную версию алгоритма TEA1 В качестве сенсора выступает утилита tcpdump и/или библиотека sharpPcap. Алгоритм преобразования реализован с использованием структур данных из библиотеки .Net System.Collections. Метод главных компонент и сжатие пространства обучающей выборки реализованы на основе библиотек alglib и Microsoft Research Accelerator (для быстрых матричных операций при помощи GPGPU). Библиотека NeuronDotNet используется для операций с ИНС. Тестирование системы в трех реальных сетях показали высокую эффектиность обнаружения вторжений(по сравнению с IceIDS2 и Snort). Скорость обработки при этом достигала 425 кпакетов/с. Кроме того, по сравнению с IceIDS2, время обучения IceIDS2s значительно меньше.

В Заключении подводятся итоги работы, делаются выводы об эффективности и применимости полученных результатов, описываются дальнейшние пути развития работы.

Основные результаты работы

При решении поставленных в диссертационной работе задач получены следующие основные научные и практические результаты:

1. Разработано представление сетевого трафика в виде набора векторов, содержащих признаки вторжений и/или аномальных явлений в целевой сети. Разработан набор алгоритмов преобразования трафика в этот набор векторов; алгоритм имеет сложность $O(log_2^2 n)$ и может работать в режиме реального времени.

2. Разработана модель эвристической сетевой системы обнаружения вторжений на базе предложенного алгоритма, метода главных компонент и одноклассового классификатора на базе искусственных нейронных сетей. Модель, после обучения, способна обнаруживать вторжения и нехарактерные явления в трафике целевой сети. Модель может быть адаптирована для обнаружения аномалий в других сложных динамических объектах.

3. Проведено экспериментальное исследование эффективности обнаружения вторжений и аномалий при помощи полученной модели. Результаты исследования показали, что модель является конкурентоспособной в области методов обнаружения вторжений обучающихся только на положительном или смешанном трафике.

4. Разработаны комплексы программ IceIDS2 и IceIDS2s представляющие из себя сетевые эвристические системы обнаружения вторжений. Проведены испытания и внедрения программных комплексов в реальных сетях. Результаты испытаний показали эффективность работы систем по критериям CR, PPs и FP.

Список публикаций

В изданиях рекомендованных ВАК РФ:

1. Большев А.К. Яновский В.В. Подход к обнаружению аномального трафика в компьютерных сетях с использованием методов кластерного анализа. // Известия Государственного Электротехнического Университета, серия «Информатика, управление и компьютерные технологии», выпуск 3/2006, Изд-во СПбЭТУ, СПб, 2006, С. 38-45.

2. Большев А.К. Яновский В.В. Применение нейронных сетей для обнаружения вторжений в компьютерные сети. // Вестник Санкт-Петербургского университета, Серия 10 ПМПУ вып.1/2010, СПб, 2010, С. 129-136.

3. Большев А.К. Лисс А.Р. Применение искусственных нейронных сетей и методов сокращения размерности данных при решении задач обнаружения сетевых вторжений. // Известия СпбГЭТУ «ЛЭТИ», выпуск 5/2010, Изд-во СПбЭТУ, СПб, 2010, С. 51-56.

4. Большев А.К. Лисс А.Р. Прототип эвристической системы обнаружения вторжений в компьютерные сети на основе метода главных компонент. // Научно-технические ведомости Санкт-Петербургского государственного политехнического университета, Серия «Информатика, Телекоммуникации, Управление» вып.4(103)/2010, Изд-во Политехнического университета, Санкт-Петербург, 2010, с. 200-205.

Статьи и материалы конференций:

5. Большев А.К. Яновский В.В. Модель системы обнаружения атак отказа в обслуживании на основе метода опорных векторов. // Известия Государственного Электротехнического Университета, выпуск 4/2008, Изд-во СПбГЭТУ, СПб, 2008, С. 25-31.

6. Большев А.К. Обнаружение вторжений в сеть при помощи одноклассовой классификации методом опорных векторов. // Сб. тр. 62-й науч.-техн. конф. проф.-преп. состава СПбГЭТУ «ЛЭТИ» 2009. 27 января - 8 февраля 2009, С. 121-127.

7. Большев А.К. Применение нейронных сетей для обнаружения вторжений в образовательных компьютерных системах // XV Международная конференция «Современное образование: содержание, технологии, качество.», СПбГЭТУ «ЛЭТИ» 22.04.2009, т1, С. 115-117

8. Большев А.К. Яновский В.В. Применение алгоритмов одноклассовой классификации при построении сетевых систем обнаружения вторжений. // ИНФОС-2009, Вологодский Государственный Технический Университет, 26-27.06.2009. С. 28-31.

9. Большев А.К. Лисс А.Р. Использование PCA и ИНС для обнаружения вторжений в сеть. // Технологии Microsoft в теории и практике программирования, Изд-во Политехнического университета, Санкт-Петербург, 2010, С. 12-13.

10. Большев А.К. Лавров А.А. Метод идентификации ОС удаленного хоста на основе анализа временных характеристик стека TCP/IP в задачах сетевого мониторинга. // Сб. тр. 64-й науч.-техн. конф. проф.-преп. состава СПбГЭТУ «ЛЭТИ». СПб., 2011. С. 104-110.