На правах рукописи

КОСЕНКО Максим Юрьевич

МНОГОАГЕНТНАЯ СИСТЕМА ОБНАРУЖЕНИЯ И БЛОКИРОВАНИЯ БОТНЕТОВ ПУТЕМ ВЫЯВЛЕНИЯ УПРАВЛЯЮЩЕГО ТРАФИКА НА ОСНОВЕ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ

Специальность:
05.13.19 – Методы и системы защиты информации, информационная безопасность

АВТОРЕФЕРАТ
диссертации на соискание ученой степени кандидата технических наук

Челябинск – 2017

Работа выполнена на кафедре информационных технологий и экономической информатики ФГБОУ ВО «Челябинский государственный университет»

Научный руководитель: доктор технических наук, профессор
Мельников Андрей Витальевич

Официальные оппоненты: доктор технических наук, с. н. с.
Марков Алексей Сергеевич
Закрытое акционерное общество «Научно-производственное объединение «Эшелон»
президент

доктор технических наук, профессор
Соловьев Николай Алексеевич
ФГБОУ ВО «Оренбургский государственный университет», заведующий кафедрой программного обеспечения вычислительной техники и автоматизированных систем факультета математики и информационных технологий

Ведущая организация: ФГАОУ ВО «Тюменский государственный университет»,
г. Тюмень

Защита диссертации состоится 07 апреля 2017 г. в 10^{00} часов на заседании диссертационного совета Д 212.288.07 на базе ФГБОУ ВО «Уфимский государственный авиационный технический университет» по адресу: 450008, г. Уфа, ул. К. Маркса, 12.

С диссертацией можно ознакомиться в библиотеке ФГБОУ ВО «Уфимский государственный авиационный технический университет» и на сайте *www.ugatu.su*.

Автореферат разослан «___» _____ 20___ года.

Ученый секретарь
диссертационного совета,
доктор технических наук, доцент И. Л. Виноградова

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования

Большинство атак и мошеннических действий в Интернете осуществляется с помощью вредоносного программного обеспечения, которое включает в себя вирусы, трояны, черви, шпионские программы, ботнеты. Вредоносное программное обеспечение стало основным источником большинства зловредной активности в Интернете: целевые атаки, распределенные атаки типа «отказ в обслуживании», мошеннические действия, а также сканирование. Среди всех видов вредоносного программного обеспечения ботнеты являются основной платформой, которую злоумышленники используют как масштабный, согласованно действующий инструмент, используемый для поддержки постоянного роста преступной деятельности, такой, как DDoS, рассылка спама, фишинг и кража информации. Данные крупнейших мировых компаний, специализирующихся в области защиты информации, таких как Prolexic/Akamai, Incapsula, показывают, что ботнеты являются основной угрозой безопасности в Интернете и постоянно функционируют более 1200 ботнетов. Ботнетом называют сеть заражённых вредоносным программным обеспечением компьютеров, которые находятся под удаленным управлением злоумышленника. Традиционно обнаружение ботнетов осуществляется с помощью пассивного мониторинга и анализа сетевого трафика. Для обнаружения ботнетов выделяют подходы на основе поиска сигнатур либо аномалий в трафике. Менее популярны подходы, использующие анализ DNS трафика или применение узлов-ловушек. Основным недостатком существующих решений обнаружения ботнетов является то, что они не учитывают взаимосвязь многоагентной природы ботнетов и этапов их жизненного цикла. В результате обнаружение получается частичным, и блокировать деятельность ботнета не представляется возможным. В связи с этим, решаемая в диссертационной работе задача, заключающаяся в разработке многоагентной системы обнаружения и блокирования ботнетов путем выявления управляющего трафика на основе методов интеллектуального анализа сетевого трафика, является актуальной.

Степень разработанности темы

На сегодняшний день вопрос применения методов интеллектуального анализа данных в системах защиты информации широко освещен в работах отечественных и зарубежных авторов, таких как А.Е. Архипов, В.И. Васильев, Дж. Бинкли, Т. Йен, В.А. Камаев, А. Карасаридис, А.В. Козачок, А. Г. Корченко, И.В. Котенко, Н.Н. Куссуль, А.В. Лукацкий, К. Ливадас, Дж. Митчелл, А.Н. Назаров, А. Рамачандран, М. Рейтер, В.А. Сердюк, Д.С. Сильнов, С. Сингх, Е. Стинсон, В. П. Фраленко, И.А. Ходашинский и др.

В существующих исследованиях предложено множество подходов к обнаружению ботнетов, но эти решения имеют различные ограничения:

– отсутствие механизма автоматической генерации сигнатур ботов;

– обнаружение ботнетов с конкретной организационной структурой (централизованной или децентрализованной);

– обнаружение ботнетов, работающих по специфичному протоколу (IRC или HTTP и др.);

– обнаружение ботнетов с определенной вредоносной активностью (сканирование или рассылка спама);

– множество ложных срабатываний.

Таким образом, тема диссертационной работы, посвященная разработке алгоритмов и программного обеспечения системы, позволяющей автоматизировать процесс обнаружения и блокирования ботнетов с применением методов интеллектуального анализа данных, является актуальной.

Объектом исследования в данной работе являются системы защиты от ботнетов в открытых компьютерных сетях, включая Интернет.

Предметом исследования являются методы и алгоритмы обнаружения ботнетов с использованием интеллектуального анализа данных в рамках многоагентного подхода.

Целью диссертационной работы является повышение защищенности информационных систем от атак, использующих ботнеты, на основе разработки и применения многоагентной системы обнаружения и блокирования ботнетов с использованием алгоритмов интеллектуального анализа данных.

З**адачи исследования**

1. Исследовать распространенные атаки типа «отказ в обслуживании», процесс их реализации и механизмы защиты от них, проанализировать существующие подходы к выявлению ботнетов.
2. Разработать алгоритм обнаружения управляющего трафика ботнета в глобальных сетях с использованием технологий интеллектуального анализа данных.
3. Предложить архитектуру многоагентной системы обнаружения и блокирования ботнетов, проанализировать эффективность функционирования предложенных в диссертационном исследовании алгоритмов.
4. Разработать метод распределенного обнаружения управляющих компонент ботнета, позволяющий обнаруживать управляющие серверы и узлы сети, с которых осуществляется контроль атаки, основанный на сигнатуре управляющего трафика.
5. Разработать исследовательский прототип многоагентной системы обнаружения и блокирования ботнетов, дать рекомендации по практическому внедрению многоагентной системы обнаружения и блокирования ботнетов.

Научная новизна

– Предложен алгоритм обнаружения управляющего трафика ботнетов Botnet MultiAgent Recognition, основанный на интеллектуальном анализе данных с возможностью автоматического формирования сигнатуры управляющего трафика, что, в отличие от существующих методов, позволяет решать задачу обнаружения ботнетов в автоматическом режиме независимо от протокола их управления. При этом алгоритм позволяет обнаруживать ботнеты централизованной и децентрализованной организационных структур, а также не зависит от типа вредоносной деятельности ботов;

– Предложена архитектура многоагентной системы обнаружения ботнетов, соответствующая типовой архитектуре ботнета, что позволяет блокировать атаки на стороне её источника, тем самым разгрузив каналы передачи от вредоносного трафика, а также на основе разработанного алгоритма обнаружения управляющего трафика позволяет обнаруживать и блокировать пассивных участников ботнета;

– Предложен метод распределенного обнаружения управляющих компонент ботнета, основанный на сигнатуре управляющего трафика, который, в отличие от существующих алгоритмов, за счет использования многоагентного подхода позволяет обнаруживать управляющие серверы и узлы сети, с которых осуществляется контроль атаки.

Практическая значимость

Практическая значимость полученных результатов заключается в применении многоагентной системы для обнаружения и блокирования ботнетов путем обнаружения управляющего трафика ботнета на основе интеллектуального анализа данных, что обеспечивает повышение показателя F-меры обнаружения ботнета по сравнению с рядом известных систем от 9% до 26%, при этом доля ложных срабатываний не превышает 0,02. Разработанный прототип системы позволяет повысить эффективность предотвращения распределенных атак, совершаемых ботнетами, обеспечить централизованный мониторинг кибер-угроз в сети Интернет, обеспечить процесс проведения кибер-расследований благодаря возможности обработки накопленных данных.

Методы исследования

При решении поставленных в работе задач использовались теоретико-множественные, агентно-ориентированные методы представления моделей, объектно-ориентированные методологии проектирования и разработки программных систем, а также методы интеллектуального анализа данных. Для оценки эффективности предлагаемых решений использовались методы функционального и информационного моделирования.

Положения, выносимые на защиту

1. Результаты анализа состояния проблемы обнаружения ботнетов в открытых компьютерных сетях (включая Интернет), существующих методов защиты от распределенных атак типа «отказ в обслуживании» и методов обнаружения ботнетов.
2. Алгоритм обнаружения управляющего трафика ботнета Botnet MultiAgent Recognition на основе интеллектуального анализа данных.
3. Архитектура интеллектуальной многоагентной системы обнаружения и блокирования ботнетов.
4. Метод распределенного обнаружения управляющих компонент ботнета, позволяющий обнаруживать управляющие серверы и узлы сети, с которых осуществляется контроль атаки, основанный на сигнатуре управляющего трафика.
5. Исследовательский прототип многоагентной системы обнаружения и блокирования ботнетов.

Достоверность результатов

Полученные в диссертационной работе результаты не противоречат известным теоретическим положениям и подтверждаются результатами апробации и внедрения прототипа многоагентной системы, реализующего представленные алгоритм и метод обнаружения компонент ботнетов.

Апробация результатов

По теме диссертации опубликовано 12 научных статей и тезисов докладов, из них 3 статьи в изданиях, рекомендованных ВАК для публикации основных результатов диссертаций на соискание учёной степени кандидата наук. Имеется свидетельство о государственной регистрации программы для ЭВМ.

Основные положения диссертационной работы докладывались и обсуждались на следующих научных конференциях:

– XIV Международной научной конференции «Компьютерные науки и информационные технологии», Уфа – Гамбург – Норвежские Фьорды, 2012;

– II Международной конференции "Интеллектуальные технологии обработки информации и управления", Уфа, Россия, 2014.

– I Международной научно-практической конференции «Технологии цифровой обработки и хранения информации» (DSPTech’2015), г. Уфа, 2015 г.

– Международный семинар «Ситуационное управление, интеллектуальные, агентные вычисления и кибербезопасность в критических инфраструктурах» (CM/IAC/CS/CI-2016), г. Иркутск, 2016 г.;

– XVII Байкальской Всероссийской конференции с международным участием «Информационные и математические технологии в науке и управлении», г. Иркутск, 2012.

Разработанный прототип многоагентной системы обнаружения и блокирования ботнетов используется в компании предоставляющей услуги связи в Челябинской области ЗАО «Интерсвязь».

Результаты исследования используются в учебном процессе на кафедре «Информационные технологии и экономическая информатика» ФГБОУ ВО «Челябинский государственный университет» при проведении лекций и лабораторных работ по курсам «Система интеллектуального анализа данных» и «Защита информации» для студентов направления 02.04.02 «Фундаментальная информатика и информационные технологии» и 09.03.01 «Информатика и вычислительная техника».

Работа поддержана лабораторией квантовой топологии Челябинского государственного университета (грант Правительства РФ 14.Z50.31.0020).

Объём и структура работы

Диссертационная работа состоит из введения, четырех глав, заключения и списка литературы. Текст работы изложен на 149 страницах, содержит 43 рисунка и 33 таблицы. Список использованной литературы состоит из 149 наименований.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность темы диссертационного исследования, сформулированы цель и задачи работы, объект и предмет исследования, отмечается научная новизна, практическая значимость, положения, выносимые на защиту.

В первой главе проведен анализ современного состояния в области обнаружения ботнетов. В частности, проведен обзор ботнетов и их жизненного цикла. Рассмотрены существующие техники обнаружения вторжений и вредоносного программного обеспечения. Проведен обзор систем сбора информации о ботнетах и их отслеживания на основе метода приманки («honeypot'a»). Также проведен анализ существующих технологий и систем обнаружения и блокирования ботнетов. В результате проведенного анализа делается вывод о том, что существующие методы защиты от ботнетов и их атак не результативны. Это определяет необходимость разработки системы защиты с уровнем сложности не меньше, чем у самих ботнетов. Поэтому необходимо применять метод и систему, способную работать таким же распределенным способом, как и ботнет. Система должна обеспечивать возможность анализировать множество сетевых данных в разных сетях, обнаруживать сетевые атаки, влиять на фильтрацию трафика, выявлять сигнатуры вредоносного поведения и взаимодействовать между собой для эффективного выполнения перечисленных задач. Приведена классификация защитных механизмов от распределенной атаки типа «отказ в обслуживании», проведен обзор превентивных механизмов защиты от данного типа атаки. В заключении главы сформулированы цель и задачи исследования, поставленные в диссертационной работе.

Во второй главе разрабатываются архитектура многоагентной системы обнаружения и блокирования ботнетов, алгоритм обнаружения управляющего трафика ботнетов Botnet MultiAgent Recognition, метод отслеживания ip-адресов узлов сети, с которых осуществляется контроль атаки ботнетов.

Идея, вкладываемая в метод, основанный на многоагентном подходе, состоит в следующем. Для того чтобы обнаружить ботнет, в первую очередь необходимо обнаружить распределенную атаку типа «отказ в обслуживании», для осуществления которой чаще всего прибегают к использованию ботнетов. После обнаружения атаки необходимо блокировать её на стороне источника атаки, а атакующее средство взять под наблюдение для выявления характерных признаков работы бота. Далее нужно попытаться идентифицировать других участников ботнета путем поиска в различных сетях ранее обнаруженных признаков работы бота.

Для отображения структуры и функций системы обнаружения и блокирования была создана функциональная модель, представленная на рисунке 1. Модель создавалась с применением методологии функционального моделирования IDEF0. Функциональная модель представляет собой структурированное изображение функций системы, информации и объектов, связывающих эти функции.

Процессы в данной модели решают известные задачи обеспечения защиты информации, более того, в большинстве своем эти задачи имеют определенное решение: задача обнаружения атаки типа «распределенный отказ в обслуживании»; задача блокирования атаки; задача выявления характерных признаков работы бота (выявление управляющего трафика и формирование сигнатуры); задача обнаружения бота; задача координации агентов системы; задача контроля и мониторинга работы агентов; задача накопления информации о кибер-угрозах; задача визуализации кибер-атак и ботнетов.

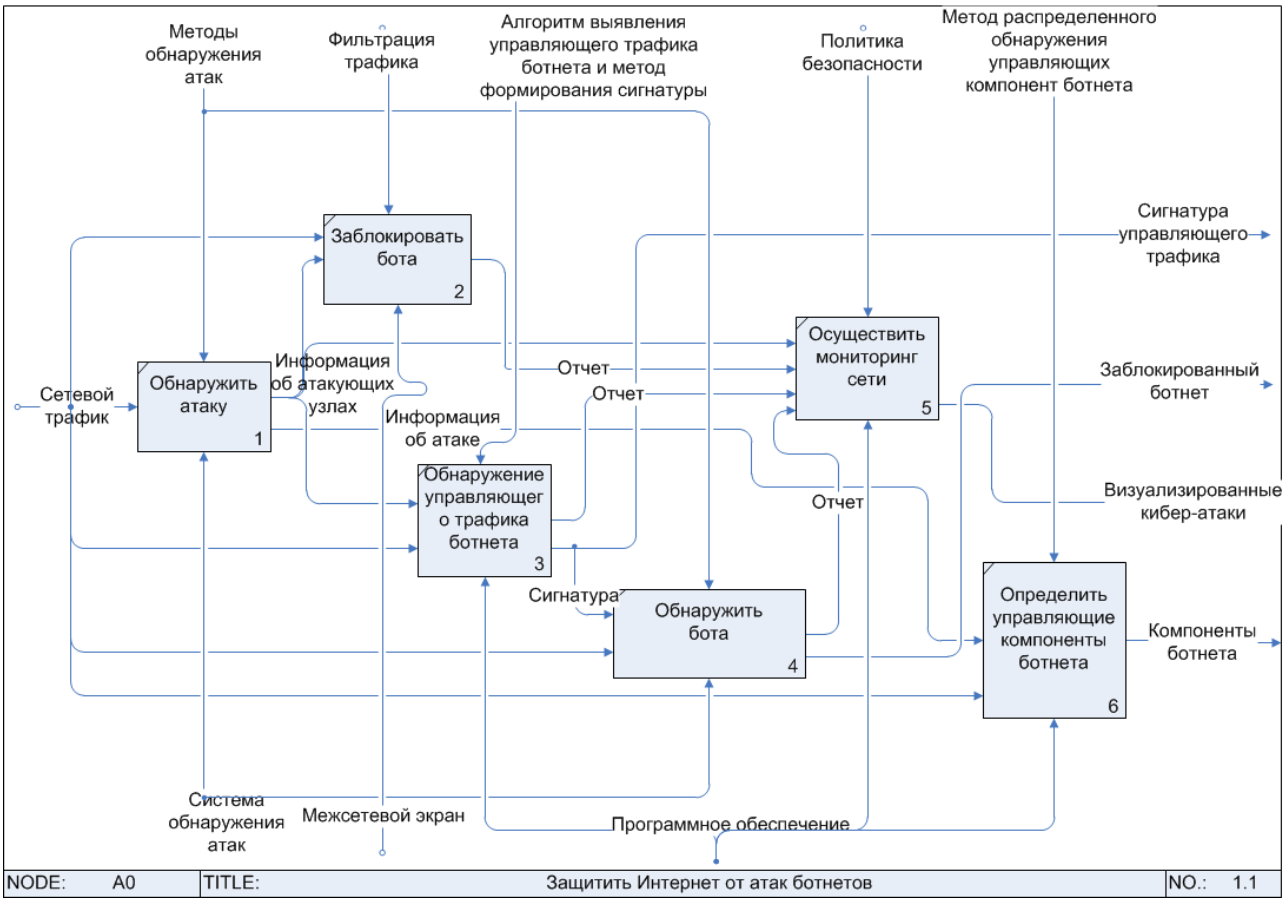

Рисунок 1 – Функциональная модель многоагентной системы обнаружения и блокирования ботнетов

На основе анализа функциональной модели сделан вывод о том, что эффективным средством противодействия ботнетам может стать система, аналогичная по архитектуре ботнетам. По своей сути ботнет является многоагентной системой, совместная работа агентов которой приводит к эффективным кибер-атакам. Таким образом, система защиты тоже должна быть многоагентной, что приведет к высоким результатам защиты как от атак, производимых ботнетами, так и как общее средство противодействия зловредным сетям. Многоагентный подход фактически избавляет от проблем масштабирования при росте системы обнаружения. Выявленные характерные признаки взаимодействия ботов с контролерами ботнетов используются для динамического формирования сигнатур ботов. Сигнатуры позволяют обнаружить

присутствие бота в других сетях. Такой подход помогает решить проблему автоматизации обнаружения ботов.

Для обнаружения ботнетов разработан алгоритм обнаружения управляющего трафика ботнета Botnet MultiAgent Recognition на основе методов интеллектуального анализа данных, по результату работы которого формируется сигнатура ботнета. Разработанный алгоритм обнаруживает трафик ботнета независимо от используемого протокола или организационной структуры ботнета.

На первом этапе осуществляется фильтрация ненужных потоков трафика. Это делается путем отброса сетевого трафика, который не направлен от внутренних узлов к внешним узлам. Таким образом, алгоритм игнорирует трафик, связанный с сообщениями между внутренними узлами, а также трафик, инициированный внешними узлами по отношению к внутренним узлам. Отфильтровываются также потоки, которые не находятся в состоянии «установлено, осуществляется передача данных». В дополнение осуществляется фильтрация по белым спискам, в рамках которой отфильтровываются все потоки, направленные к хорошо известным легитимным серверам. Предполагается, что данные серверы вряд ли могут являться управляющими серверами. Стоит отметить, что этап фильтрации не является критическим для нормального функционирования кластеризации. Этот этап полезен для снижения нагрузки трафика, повышения эффективности процесса кластеризации и фильтрации легитимного пользовательского трафика.

Следующим этапом является агрегация связных коммуникационных потоков с целью снижения нагрузки. В рамках временного интервала E (обычно несколько часов) все m TCP/UDP потоков, которые разделяют один и тот же протокол (TCP или UDP), адрес источника, адрес назначения и порт, объединяются в один коммуникационный поток $c_i = \{f_j\}_{j=1..m}$, где каждая f_j – это отдельный TCP/UDP поток. Множество $\{c_i\}_{i=1..n}$ объединяет все n коммуникационных потоков, наблюдаемых в интервале E, отражая сетевые взаимодействия наблюдаемого хоста.

Для того чтобы применить алгоритмы кластеризации для коммутационных потоков, сначала необходимо представить потоки в подходящем векторном представлении. Для этого в алгоритме извлекается ряд статистических признаков из каждого коммутационного потока c_i и переводится в d-мерный вектор $p_i < R^d$. Можно описать эту задачу в качестве функции F : C $\rightarrow$ R^d. Функция F определяется следующим образом: учитывая коммуникационный поток c_i, вычисляется дискретное распределение четырех случайных величин:

- количество потоков в час;
- количество пакетов в потоке;
- среднее число байт в пакетах;
- среднее количество байт в секунду.

С учетом дискретного распределения выборки каждой из этих четырех случайных величин вектора признаков получаются разной размерности. Для решения этой проблемы осуществляется нормализация векторов признаков путем техники биннинга данных.

Далее осуществляется кластеризация потоков. Цель заключается в поиске групп коммуникационных потоков, которые похожи друг на друга, тем самым выявляются потоки трафика к управляющим серверам. Кластеризация сетевых потоков осуществляется в два этапа, как показано на рисунке 2.

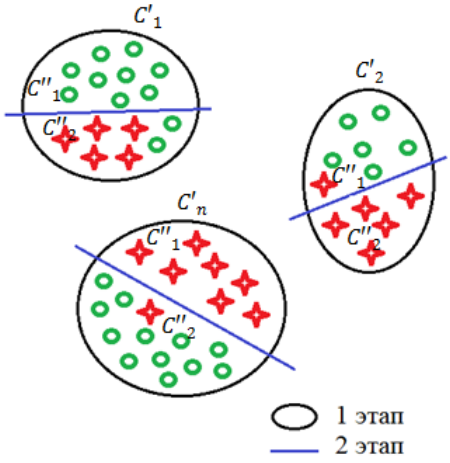

Рисунок 2 – Двухэтапная кластеризация

На первом этапе проводится первичная кластеризация в уменьшенном пространстве признаков $R^{d'}$, где d'<d. Уменьшение размерности пространства признаков с d=32 до d'=8 происходит путем вычисления среднего и дисперсии распределений fph, ppf, bpp и bps для каждого коммуникационного потока. Затем применяется алгоритм кластеризации XMeans на полученном представлении коммуникационного потока для нахождения крупных кластеров $\{C'_i\}_{i=1..y_1}$. Результатом первого шага кластеризации является множество $\{C'_i\}_{i=1..y_1}$, где y_1 – относительно большие кластеры. Таким образом, набор данных D разбивается на меньшие наборы данных (кластеры C'_i), содержащие набор точек, не очень отдаленных друг от друга. Поскольку кластеры $\{C'_i\}_{i=1..y_1}$ генерируются на первом шаге кластеризации, их количество не очень большое.

Для уточнения результата первого этапа кластеризации выполняется второй этап кластеризации на каждом отдельном наборе данных C'_i с помощью простого алгоритма кластеризации на полном описании коммуникационных потоков в R^d. Т.е. на втором шаге кластеризации используются все d=32 признака для представления коммуникационного потока. Второй шаг генерирует набор y_2 небольших и более точных кластеров $\{C''_i\}_{i=1..y_2}$.

После того, как получены результаты кластеризации, алгоритм проводит кросс-кластерную корреляцию. Идея состоит в том, чтобы выделить кластеры с максимальным пересечением среди всех кластеров, полученных на скомпрометированных узлах. Т.е. эти кластеры должны быть наиболее близки друг к другу. В кластерном анализе для количественной оценки близости вводится понятие метрики. Сходство и различие между объектами устанавливаются в зависимости от значения пересечения двух кластеров. В качестве меры близости для решения этой задачи используется следующее выражение:

$$sim(C''_i, C''_j) = \max_{\substack{i,j=1,y_2 \\ i \neq j}} \left(\left| C''_i \cap C''_j \right| \right) \quad (1)$$

В диссертационном исследовании не ставится задача определения личности и местонахождения злоумышленников. В работе предлагается метод, позволяющий предположить, что множество конкретных сетевых узлов являются компонентами ботнета, с которых осуществляется контроль атаки. Данный метод основан на анализе входящего трафика на атакуемую машину.

Метод распределенного обнаружения управляющих компонент ботнета основан на необходимости осуществления контроля выполнения атаки. При осуществлении атаки типа «отказ в обслуживании» организаторы атаки должны на протяжении всей атаки убеждаться, что атакуемый сервис находится в нерабочем состоянии. С этой целью они используют различное программное обеспечение, чаще основанное на протоколе ICMP. Таким образом, возникает возможность выявить злоумышленника путем анализа трафика атакуемой и атакующей машин. Функции отслеживания реализуются в модуле отслеживания узлов, контролирующих атаку. Данный модуль является частью агента обнаружения атаки. В результате работы модуля собирается следующая информация об узлах сети, с которых осуществляется контроль атаки ботнета: тип и версия операционной системы, тип устройства, запущенные службы, порты, маршрут следования данных, DNS информации, регистрационные данные whois, географическое местоположение узла, уязвимости узла.

В третьей главе проводится анализ эффективности разработанного алгоритма обнаружения управляющего трафика ботнета, основанного на применении технологий интеллектуального анализа данных, представлены результаты проведенных экспериментов.

Оценка эффективности разработанных алгоритмов проводилась по следующим критериям:

1. Точность алгоритма в пределах кластера – это доля трафика, действительно принадлежащая управляющему трафику ботнета, относительно всего трафика, отнесенного к этому кластеру:

$$PRECISION = \frac{TP}{TP+FP}, \quad (2)$$

где TP – истинно-положительные решения, т.е. решения, отнесенные к кластеру трафика ботнета и действительно им являющиеся;

FP – ложно-положительные решения, т.е. решения, отнесенные к кластеру трафика ботнета, но им не являющиеся.

2. Полнота алгоритма – это доля управляющего трафика ботнета в кластере относительно всего управляющего трафика ботнета, содержавшегося в собранных данных:

$$RECALL = \frac{TP}{TP+FN}, \quad (3)$$

где FN – ложно-отрицательные решения, т.е. решения, отнесенные к трафику, отличному от трафика ботнета, но являющиеся трафиком ботнета.

3. Метрика F-мера представляет собой гармоническое среднее между точностью и полнотой:

$$F = 2\frac{PRECISION \times RECALL}{PRECISION+RECALL} \quad (4)$$

Для проведения тестирования многоагентной системы обнаружения и блокирования ботнетов необходим работоспособный ботнет. С этой целью был

разработан метод автоматического формирования базы ботов и собран экспериментальный стенд в сети государственной организации. Автоматическое формирование базы ботов организовывалось путем разработки исследовательской honeypot-системы низкого взаимодействия, ориентированной на предоставление веб сервиса. Использование этой системы позволило получить экземпляр исходного кода бота, модификация которого дала возможность создать бота для экспериментального стенда.

Боты функционировали на реальных рабочих станциях, что обеспечило наличие разнородного пользовательского трафика. Рабочие станции работали под управлением операционной системы Windows 7, 8. Агенты располагались на виртуальных серверах с установленной операционной системой CentOS 7. Применялись два основных агента, участвующих в определении управляющего трафика бонета: агент исследования трафика скомпрометированных узлов и агент формирования сигнатуры. Информацию для обработки агенты исследования трафика получали по протоколу Netflow. Для этого коммутаторы в сети скомпрометированных узлов реализовывали функцию сенсора Netflow и передавали данные коллектору. Коллектор Netflow работает в качестве модуля агента исследования трафика. Структура сети эксперимента представлена на рисунке 3.

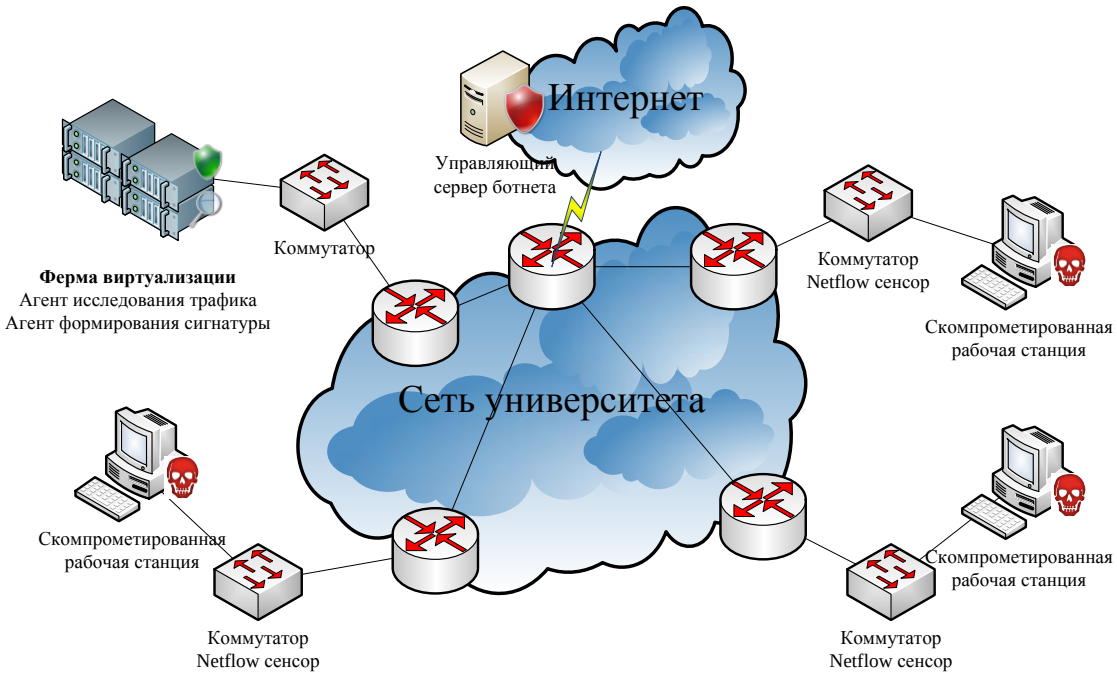

Рисунок 3 – Топология экспериментального стенда

Сбор данных для анализа проводился в течение 11 дней для всех трех наблюдаемых скомпрометированных узлов. В течение каждого дня первым, вторым и третьим узлом было передано порядка 900 тысяч, 700 тысяч и 400 тысяч пакетов (TCP, UDP) и 14000, 2400, 11000 сетевых потоков соответственно. По результатам процедур фильтрации и агрегирования объём данных уменьшился до 3900, 1500, 5400 агрегированных сетевых потоков в сутки. Естественно,

количество анализируемых потоков прямо зависит от активности пользователей скомпрометированных узлов.

Для реализации разработанного алгоритма обнаружения управляющего трафика проводилось сравнение трёх алгоритмов кластеризации: алгоритм кластеризации плотностным методом (DBSCAN), алгоритм, дополняющий метод кластеризации K-means эффективной оценкой числа кластеров (XMEANS), алгоритм максимального правдоподобия Expectation-maximization (EM). Данные алгоритмы самостоятельно определяют количество кластеров. Алгоритмы сравнивались с точки зрения способности генерировать кластеры таким образом, чтобы управляющий трафик ботнета распределялся в один или несколько кластеров, при этом данные кластеры содержали преимущественно управляющий трафик. Для оценки качества работы алгоритма обнаружения управляющего трафика использовалась метрика F-мера (F), являющаяся единой метрикой, объединяющей метрики точности (P) и полноты (R). В качестве меры близости двух векторов признаков x и y размерности n использовалось Евклидово расстояние:

$$dist(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}. \quad (5)$$

В результате эксперимента для каждого скомпрометированного узла был выделен набор кластеров. При этом эксперимент проводился для разных временных интервалов, используемых на этапе агрегирования потоков трафика. Результаты эксперимента представлены в таблице 1.

Таблица 1 – Показатели эффективности выделения управляющего трафика ботнета

Алгоритм	E	Узел 1			Узел 2			Узел 3		
		P	R	F	P	R	F	P	R	F
DBSCAN	2	0,0333	0,6154	0,0632	1,0000	0,7523	0,8586	0,0307	0,8511	0,0593
	4	0,0257	0,8400	0,0498	1,0000	0,7091	0,8298	0,0163	0,8571	0,0320
	6	0,0144	1,0000	0,0285	1,0000	0,6053	0,7541	0,0091	1,0000	0,0180
	8	0,0113	1,0000	0,0224	1,0000	0,6000	0,7500	0,0073	1,0000	0,0146
	10	0,0084	1,0000	0,0167	1,0000	0,5417	0,7027	0,0060	0,9444	0,0118
	24	0,0057	1,0000	0,0112	0,0194	0,6923	0,0378	0,0035	1,0000	0,0070
EM	2	0,0077	1,0000	0,0153	0,0635	1,0000	0,1196	0,0068	1,0000	0,0136
	4	0,0056	1,0000	0,0111	0,0380	1,0000	0,0731	0,0048	1,0000	0,0096
	6	0,1724	0,8824	0,2885	1,0000	0,7105	0,8308	0,1284	1,0000	0,2275
	8	0,1493	0,6250	0,2410	1,0000	0,6667	0,8000	0,1520	1,0000	0,2639
	10	0,1268	0,6429	0,2118	1,0000	0,6667	0,8000	0,1196	0,6111	0,2000
	24	0,1525	0,8182	0,2571	1,0000	0,4615	0,6316	0,0828	1,0000	0,1529
XMEANS	2	0,8788	0,7436	0,8056	0,9811	0,9541	0,9674	0,0476	0,7234	0,0892
	4	1,0000	0,4000	0,5714	1,0000	0,9455	0,9720	0,1875	0,4286	0,2609
	6	0,1282	0,5882	0,2105	1,0000	0,8947	0,9444	0,6667	0,9474	0,7826
	8	1,0000	0,3750	0,5455	1,0000	0,6667	0,8000	0,3830	0,9474	0,5455
	10	0,4167	0,3571	0,3846	1,0000	0,7083	0,8293	0,2414	0,3889	0,2979
	24	0,0408	0,5455	0,0759	1,0000	0,6154	0,7619	0,2917	0,5833	0,3889

Результаты показывают, что наиболее эффективным является использование алгоритма XMeans. В отличие от алгоритма XMeans, алгоритм DBSCAN определяет большую часть трафика бота, о чем говорят высокие показатели полноты, но не выделяет данный трафик в отдельный кластер, а совмещает его в один кластер с пользовательским трафиком. С точки зрения быстродействия алгоритмов, среднее время, затраченное на кластеризацию алгоритмом DBSCAN, сравнимо со средним временем, затрачиваемым при работе XMeans. При кластеризации с использованием алгоритма EM затрачивается существенно большее время. Таким образом, дополнительным преимуществом алгоритма XMeans является его быстродействие. Далее осуществлялся второй этап кластеризации, на котором в качестве входных данных используются кластеры, образованные на первом этапе кластеризации. На данном этапе использовался только один алгоритм кластеризации – XMeans. Полученные результаты отражены в таблице 2.

Таблица 2 – Показатели эффективности обнаружения управляющего трафика

Интервал	Узел 1			Узел 2			Узел 3		
	Precision	Recall	F-measure	Precision	Recall	F-measure	Precision	Recall	F-measure
2	0,87	0,65	0,74	0,85	0,96	0,9	0,79	0,79	0,79
4	0,63	1	0,77	0,25	0,5	0,33	0,93	0,93	0,93
6	0,25	0,75	0,38	0,2	0,5	0,29	0,75	0,75	0,75
8	0,5	0,75	0,6	0,17	1	0,29	0,88	0,58	0,7
10	0,83	1	0,91	0,89	0,89	0,89	1	1	1
24	0,67	0,67	0,67	0,75	1	0,86	0,33	0,5	0,4

Были рассчитаны доли ложно-положительных и ложно-отрицательных решений, представленных в таблице 3.

Таблица 3 – Доля ложно-положительных и ложно-отрицательных решений

Интервал	FPR				FNR			
	1 узел	2 узел	3 узел	Среднее	1 узел	2 узел	3 узел	Среднее
2	0,03	1	0,01	0,35	0,35	0,04	0,21	0,2
4	0,01	0,43	0,03	0,16	0	0,5	0,07	0,19
6	0,1	0,03	0,01	0,05	0,25	0,5	0,25	0,33
8	0,01	0,05	0,05	0,03	0,25	0	0,42	0,22
10	0,04	0,01	0	0,02	0	0,11	0	0,04
24	0,05	0,02	0,25	0,11	0,33	0	0,5	0,28

Проведённые тесты показали, что в качестве значения параметра временного интервала агрегирования сетевых потоков нужно использовать 10 часов. При этом алгоритм позволяет обнаружить управляющий трафик со средним значением F-меры 0,93. Частота ошибки первого рода, характеризующая частоту классификации легитимного трафика как трафик ботнета, равняется 0,02, а частота ошибок второго рода – 0,04.

Четвёртая глава посвящена разработке исследовательского прототипа многоагентной системы обнаружения и блокирования ботнетов, сформулированы требования к системе. Приведены результаты оценки эффективности разработанной системы, описана концепция практического внедрения многоагентной системы.

При разработке многоагентной системы использовалось множество зарекомендовавших себя средств защиты информации с открытым исходным кодом. Данные средства обеспечивают большинство функций агентов. В качестве системы обнаружения атак и системы обнаружения ботов была выбрана система Suricata. Базовой операционной системой при разработке исследовательского прототипа являлась CentOS. Поэтому в качестве межсетевого экрана для блокирования атак использовался Netfilter. В качестве связующего программного обеспечения использовалась платформа, реализующая систему обмена сообщениями между компонентами программной системы на основе стандарта AMQP (Advanced Message Queuing Protocol), RabbitMQ. Для приема файлов событий от агентов обнаружения атак и ботов использовался диспетчер бинарных файлов событий Barnyard2. Задача мониторинга и визуализации работы системы реализовывалась с помощью веб-системы мониторинга сетевой безопасности Snorby. Отдельно разрабатывались модуль кооперации, модуль отслеживания злоумышленника, модуль кластеризации трафика, модуль кросс-кластерной корреляции и модуль формирования сигнатуры ботнета. На рисунке 4 представлена диаграмма развертывания системы обнаружения ботнетов.

Подготовлена инфраструктура тестирования исследовательского прототипа. Проведено тестирование разработанной системы, построенной на базе предложенного алгоритма обнаружения управляющего трафика ботнетов. Прототип обнаружил 90% ботов используемого ботнета, в том числе ботов, не применяемых для проведения атаки. Был также обнаружен сетевой адрес, с которого проводился контроль выполнения распределенной атаки типа «отказ в обслуживании».

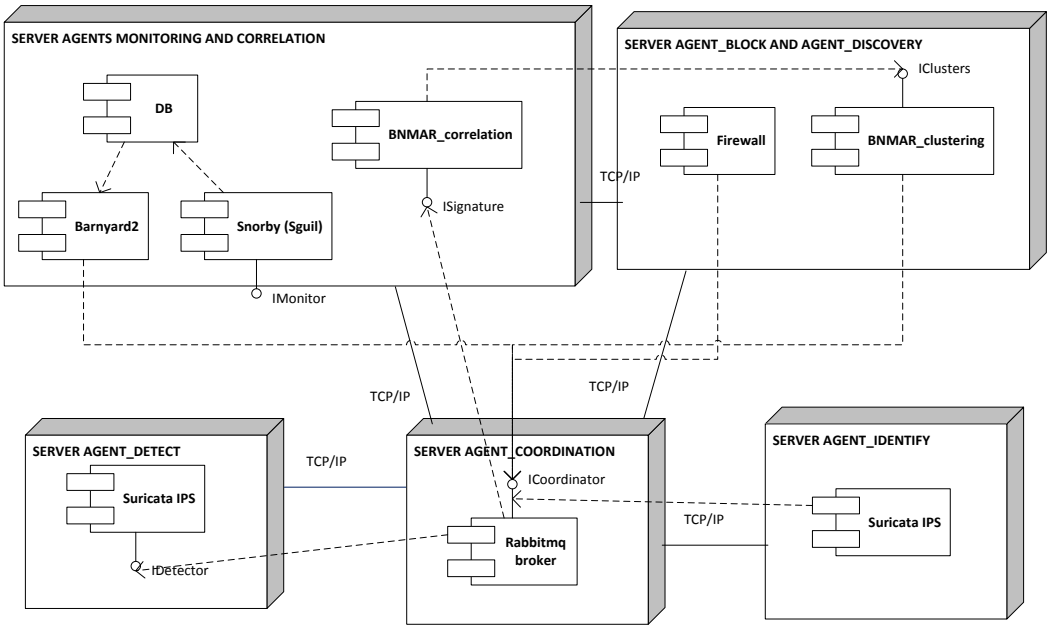

Рисунок 4 – Диаграмма развертывания системы обнаружения ботнетов

Сравнение эффективности разработанного прототипа проводилось со свободно распространяемыми защитными системами, близкими по предназначению: сетевая система предотвращения вторжений Snort и сетевая система обнаружения активности ботов BotHunter. В результате тестирования систем были получены следующие результаты. Показатель точности высокий у всех систем. У первых двух систем данный показатель максимальный в связи с использованием сигнатуры. Результирующим для общей оценки становится показатель полноты обнаружения, имеющий разброс от 15 до 35 %. В целом, показатель F-меры разработанного прототипа превышает аналогичные показатели Snort и BotHunter на 9 и 26 %. Разработанный прототип выявил сетевой адрес узла, с которого осуществлялся контроль атаки.

Предложена концепция практического внедрения многоагентной системы обнаружения и блокирования ботнетов. Концепция описывает возможность повсеместного распространения многоагентной системы для повышения эффективности её использования.

ОСНОВНЫЕ ВЫВОДЫ И РЕЗУЛЬТАТЫ РАБОТЫ

1. Проведен анализ состояния проблемы обнаружения ботнетов в открытых компьютерных сетях (включая Интернет), существующих методов защиты от распределенных атак типа «отказ в обслуживании» и методов обнаружения ботнетов. Анализ выявил отсутствие полноценной защиты от вредоносной активности ботнетов.

2. Предложен алгоритм обнаружения управляющего трафика ботнета Botnet MultiAgent Recognition на основе интеллектуального анализа данных с возможностью автоматического формирования сигнатуры управляющего трафика. В отличие от существующих методов, использование предложенного алгоритма позволяет решать задачу обнаружения ботнетов в автоматическом режиме, независимо от протокола их управления. При этом алгоритм позволяет обнаруживать ботнеты централизованной и децентрализованной организационных структур, а также не зависит от типа вредоносной деятельности ботов.

3. Предложена архитектура интеллектуальной многоагентной системы обнаружения и блокирования ботнетов, соответствующая типовой архитектуре ботнета, что позволяет блокировать атаки на стороне её источника, тем самым разгрузив каналы передачи от вредоносного трафика, а также на основе разработанного алгоритма обнаружения управляющего трафика позволяет обнаруживать и блокировать пассивных участников ботнета.

4. Предложен метод распределенного обнаружения управляющих компонент ботнета, основанный на сигнатуре управляющего трафика, который, в отличие от существующих алгоритмов, за счет использования многоагентного подхода позволяет обнаруживать управляющие серверы и узлы сети, с которых осуществляется контроль атаки.

5. Разработан исследовательский прототип многоагентной системы обнаружения и блокирования ботнетов. Эффективность разработанного прототипа подтверждена методом компьютерного моделирования ботнета и его вредоносной

деятельности. Прототип обеспечивает повышение показателя F-меры обнаружения ботнета по сравнению с рядом известных систем от 9% до 26%, при этом доля ложных срабатываний не превышает 0,02. Прототип обеспечивает обнаружение ботов, не используемых для проведения атаки, а также обнаружение сетевых адресов, с которых проводится контроль выполнения распределенной атаки типа «отказ в обслуживании».

6. Предложена концепция практического внедрения многоагентной системы обнаружения и блокирования ботнетов, описывающая возможность повсеместного распространения многоагентной системы обнаружения и блокирования ботнетов для повышения эффективности её использования.

Перспективы дальнейших исследований

Дальнейшим развитием диссертационной работы может быть исследование различных типов ботнетов и осуществляемых ими атак, совершенствование алгоритмов и архитектуры системы обнаружения и блокирования ботнетов, повышение показателей эффективности.

СПИСОК ПУБЛИКАЦИЙ ПО ТЕМЕ ДИССЕРТАЦИИ

В рецензируемых журналах из списка ВАК

1. Косенко М. Ю., Мельников А. В. Метод идентификации ботнетов на основе многоагентного подхода // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии, Воронеж: Воронежского государственного университета. – 2015. – С. 89–96.

2. Косенко М. Ю. Интеллектуальный анализ данных в задаче обнаружения ботнетов // Вестник УрФО. Безопасность в информационной сфере – № 1(19). – 2016. – С. 22–30.

3. Косенко М. Ю., Мельников А. В. Вопросы обеспечения защиты информационных систем от ботнет атак // Вопросы кибербезопасности. – № 4(17). – 2016. – С. 20–28.

Объекты интеллектуальной собственности

4. Свид. о гос. регистрации программы для ЭВМ № 2016661837. Система обнаружения управляющего трафика ботнета / М.Ю. Косенко, Д.Ю. Лавров. Зарег. 21.10.2016 г. – М.: Роспатент, 2016.

В трудах международных и всероссийских конференций

5. Косенко М.Ю., Лавров Д.Ю., Метод обнаружения трафика ботнета «Botnet MultiAgent Recognition» основанный на интеллектуальном анализе данных // Труды I Международной научно-практической конференции «Технологии цифровой обработки и хранения информации» (DSPTech’2015), Уфа, Россия. – 2015 – С. 153–157 (на англ. яз.).

6. Косенко М. Ю., Мельников А. В. Метод автоматического формирования базы ботов для классификации типов взаимодействия в ботнетах // Труды третьей международной конференции «Информационные технологии интеллектуальной поддержки принятия решений». – 2015. – С. 74–77.

7. Косенко М. Ю., Мельников А. В. Кооперация агентов многоагентной системы идентификации ботнетов // Труды Четвертой Международной научной конференции «Информационные технологии и системы», Челябинск: Челябинского государственного университета. – 2015. – С. 128–130.

8. Косенко М. Ю. Метод обнаружения ботнетов на основе многоагентного подхода // Труды второй международной конференции «Интеллектуальные технологии обработки информации и управления», 10-12 ноября, Уфа, Россия. – 2014. – С. 20–22.

9. Косенко М.Ю., Распределенное сетевое сканирование при проведении тестирования на проникновение // Компьютерные науки и информационные технологии CSIT’2013: труды 15-ой международной научной конференции, Уфа: Изд-во УГАТУ. – 2013. – С. 54–56 (на англ. яз.).

10. Косенко М. Ю. Модель выявления и блокирования распределенной атаки типа «отказ обслуживании» с источником атаки из облачной инфраструктуры // Информационные технологии и системы: труды Второй междунар. конф., Банное. – 2013. – С. 134–136.

11. Косенко М. Ю. Сбор информации при проведении тестирования на проникновение // Вестник УрФО. Безопасность в информационной сфере – № 3(9). – 2013. – С. 11–15.

12. Косенко М. Ю. Реализация распределенной атаки типа «отказ в обслуживании» на основе облачных технологий [Текст] // Информационные и математические технологии в науке и управлении. – Т. 2, ч. 2: Тр. XVII Байкал. Всерос. конф., Иркутск: ИСЭМ СО РАН. – 2012. – С. 174–180.

13. Косенко М. Ю. Злонамеренное использование облачных технологий // Информационные технологии и системы: материалы Первой междунар. конф., Банное, Россия. – 2012. – С. 67–70.

Диссертант М.Ю. Косенко

КОСЕНКО Максим Юрьевич

МНОГОАГЕНТНАЯ СИСТЕМА ОБНАРУЖЕНИЯ И БЛОКИРОВАНИЯ БОТНЕТОВ ПУТЕМ ВЫЯВЛЕНИЯ УПРАВЛЯЮЩЕГО ТРАФИКА НА ОСНОВЕ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ

Специальность 05.13.19 – Методы и системы защиты информации, информационная безопасность

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук