**На правах рукописи**

**КОТОВ Вадим Дмитриевич**

# ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА ОБНАРУЖЕНИЯ ВРЕДОНОСНЫХ ИНТЕРНЕТ-СТРАНИЦ НА ОСНОВЕ ТЕХНОЛОГИЙ МАШИННОГО ОБУЧЕНИЯ

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

**АВТОРЕФЕРАТ**
**диссертации на соискание ученой степени кандидата технических наук**

**Уфа – 2013**

Работа выполнена на кафедре вычислительной техники и защиты информации ФГБОУ ВПО «Уфимский государственный авиационный технический университет»

Научный руководитель: д-р техн. наук, профессор
**Васильев Владимир Иванович**

Официальные оппоненты: д-р техн. наук, профессор
**Мельников Андрей Витальевич**
ФГОБУ ВПО «Челябинский государственный университет», проректор по научной работе

канд. техн. наук, доцент
**Старцев Геннадий Владимирович**
ФГБОУ ВПО «Уфимский государственный авиационный технический университет», доцент кафедры автоматизированных систем управления

Ведущая организация: **ФГОБОУ ВПО «Самарский государственный университет»**

Защита диссертации состоится «28» июня 2013 г. в 10:00 часов на заседании диссертационного совета Д-212.288.07 при Уфимском государственном авиационном техническом университете в актовом зале корпуса № 1 по адресу: 450000, г. Уфа, ул. К. Маркса, 12.

С диссертацией можно ознакомиться в библиотеке Уфимского государственного авиационного технического университета.

Автореферат разослан «28» мая 2013 года.

Ученый секретарь
диссертационного совета
д-р. техн. наук, доцент **И.Л. Виноградова**

# ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность темы исследования.** В последнее время существенные изменения претерпели способы заражения компьютеров вредоносными программами. Это связано, в первую очередь, с увеличением роли сети Интернет в повседневной жизни. Интернет-браузеры и их расширения, подобно любым другим приложениям, обладают уязвимостями, которые могут быть эксплуатированы злоумышленниками. Особую опасность представляют те из них, которые позволяют выполнять произвольный код на системе жертвы. Именно это способствовало широкому распространению так называемых drive-by-download атак. Данный тип атаки основан на эксплуатации уязвимостей в браузере и/или его расширениях через вредоносные интернет-страницы. В результате атаки на компьютер жертвы устанавливается вредоносный исполняемый файл (например, троянская программа или руткит). Другими словами, вредоносные интернет-страницы являются переносчиками более крупных вредоносных программ.

Согласно обзору вредоносной активности в 2011 г., проведенному ЗАО "Лаборатория Касперского", количество атак через браузеры за год увеличилось с 580 млн. до 946 млн. При этом основным оружием злоумышленников для заражения компьютеров пользователей сети Интернет через браузер являются связки эксплойтов – серверные приложения, позволяющие своим владельцам осуществлять атаки drive-by-download. Особенно актуальна проблема для предприятий, сотрудники которых в процессе трудовой деятельности пользуются сетью Интернет. В этом случае под угрозой находятся конфиденциальные данные, которые могут быть скопированы и использованы злоумышленником против интересов компании.

Таким образом, существует необходимость защиты пользователей сети Интернет от подобного рода угроз. Современные подходы к обнаружению вредоносных программ сводятся либо к использованию сигнатур – последовательностей байт, идентифицирующих приложение, либо эвристических алгоритмов. Создание сигнатуры – это трудоемкий процесс, возможный только при наличии экземпляра вредоносной программы, поэтому он применим исключительно по отношению к известным угрозам и мало эффективен для обнаружения новых видов вредоносного содержимого.

К эвристическим методам обнаружения вредоносных программ относятся подходы, основанные на технологиях математической статистики и машинного обучения, которые при корректной формализации задачи позволяют на основе некоторого набора признаков отнести программу к классу нормальных или вредоносных приложений. В этом случае даже неизвестные вредоносные программы могут быть детектированы с определенной степенью вероятности. Поэтому технологии искусственного интеллекта в последнее время получили широкое распространение как в научном сообществе, так и в технических приложениях. Однако существует много непроработанных вопросов, связанных, в том числе с практической реализацией интеллектуальных систем обнаружения вредоносных интернет-страниц, поэтому данная тема является актуальной.

**Степень разработанности темы.** Проблема вредоносных страниц стала обсуждаться в научном сообществе вскоре после появления первых случаев массовых заражений пользователей сети Интернет через браузеры (2007 г.). Исследованиям в данной области посвящены работы таких ученых, как: M. Cova, C. Kruegel, G.Vigna, C. Charlie, B. Livshits и др.

Вместе с тем, в современных исследованиях существует ряд недостатков и непроработанных вопросов:

- не представлены анализ и предварительная категоризация обучающей выборки;
- рассмотрены только два из трех классов вредоносных страниц (при этом введение третьего класса значительно усложняет задачу классификации);
- используется большое количество динамических признаков, требующих выполнения вредоносного кода, что существенно замедляет процесс анализа страницы;
- не исследованы такие технологии как нейронные сети и иммунные вычисления;
- отсутствует сравнение систем обнаружения вредоносных интернет-страниц на основе машинного обучения с существующими антивирусными сканерами.

Таким образом, в рамках повышения эффективности обнаружения вредоносных интернет-страниц необходимо исследовать вышеперечисленные проблемы.

**Объект исследования –** Вредоносные интернет-страницы и серверные приложения, контролирующие drive-by-download атаки**.**

**Предмет исследования –** Применение технологий машинного обучения для решения задачи обнаружения вредоносных интернет-страниц.

**Цель работы**

Повышение эффективности обнаружения вредоносных интернет-страниц на основе технологий машинного обучения и их программной реализации.

**Задачи исследования**

1. Анализ существующих экземпляров вредоносных программ и изучение их свойств.
2. Построение формальной модели интернет-страницы.
3. Разработка методики получения выборки экземпляров вредоносных интернет-страниц.
4. Разработка исследовательского прототипа интеллектуальной системы обнаружения вредоносных интернет-страниц.
5. Проведение экспериментальной проверки разработанной системы, выбор наиболее эффективной технологии машинного обучения и сравнение с существующими антивирусными решениями.

**Научная новизна работы** заключается в следующем:

1. Исследованы вредоносные интернет-приложения как программные артефакты на основе вычисления статистических показателей их элементов, что позволило выявить основные характеристики рассматриваемых приложений.

2. Разработана модель интернет-страницы на основе формальных языков, отличающаяся от известных аналогов тем, что она обеспечивает математическое описание интернет-страницы как вектора в *n*-мерном пространстве признаков.

3. Предложена методика получения обучающей выборки для интеллектуальной системы обнаружения вредоносных интернет-страниц на основе технологий машинного обучения.

4. Разработаны и исследованы алгоритмы обнаружения вредоносных интернет-страниц на основе технологий машинного обучения, отличающиеся от известных аналогов тем, что они направлены на решение более сложной задачи классификации, и в них реализованы ранее не рассматриваемые (в контексте поставленной задачи) технологии машинного обучения, такие как, нейронные сети и иммунные вычисления.

5. Произведен экспериментальный анализ эффективности исследуемых технологий машинного обучения. Произведено исследование модели иммунной сети для решения задачи обнаружения вредоносных интернет-страниц. Выбрана наиболее эффективная модель машинного обучения, основанная на технологии нейронных сетей, демонстрирующая существенно более высокие показатели, чем большинство современных антивирусных сканеров.

**Теоретическая и практическая ценность** полученных результатов состоит в возможности использования разработанных алгоритмов обнаружения вредоносных интернет-страниц при построении систем защиты информации. Реализован исследовательский прототип интеллектуальной системы обнаружения вредоносных интернет-страниц. Предложена модульная архитектура системы, позволяющая добавлять новые алгоритмы машинного обучения и анализировать их эффективность. Разработанные алгоритмы построения интеллектуальной системы обнаружения вредоносных интернет-страниц позволяют с высокой эффективностью (более 87%) обнаруживать новые, ранее неизвестные экземпляры. При этом уровень ложных срабатываний в ходе экспериментов не превысил 11%.

**Методология и методы исследования.** В работе использовались методы теории формальных языков, теории множеств, теории вероятностей, теории машинного обучения. Для обработки результатов экспериментов использовались методы математической статистики.

**Положения, выносимые на защиту**

1. Результаты анализа вредоносных интернет-приложений как программных артефактов, позволившие выявить их основные характеристики (такие, как техники проведения атак, способы внедрения вредоносного кода, используемые методы защиты от обнаружения), необходимые при построении интеллектуальной системы обнаружения вредоносных интернет-страниц.

2. Модель интернет-страницы, основанная на теории формальных языков, позволяющая описать множество её статических признаков в терминах регулярных выражений и представить её в виде вектора в пространстве 73-х измерений.

3. Методика получения выборки вредоносных и безвредных интернет-страниц, основанная на методах математической статистики, позволяющая составить сбалансированный и репрезентативный набор данных для обучения интеллектуальной системы обнаружения вредоносных интернет-страниц.

4. Алгоритмы обнаружения вредоносных интернет-страниц на основе технологий машинного обучения, позволяющие повысить эффективность обнаружения вредоносных интернет-страниц при использовании их в дополнение к антивирусным сканерам.

5. Исследовательский прототип интеллектуальной системы обнаружения вредоносных интернет-страниц на основе технологий машинного обучения, реализующий разработанные алгоритмы обнаружения вредоносных интернет-страниц.

6. Результаты экспериментального исследования эффективности предложенных интеллектуальных алгоритмов обнаружения вредоносных интернет-страниц.

**Достоверность полученных результатов** основана на использовании в теоретических построениях законов и подходов, справедливость которых общепризнанна, а также известного и корректного математического аппарата; вводимые допущения мотивированы фактами, известными из практики. Достоверность и обоснованность научных положений подтверждена соответствием результатов теоретических и экспериментальных исследований.

**Апробация результатов.** Основные научные и практические результаты диссертационной работы докладывались и обсуждались на:

- 9-й и 10-й Международных научных конференциях "Security of Information and Networks" (Безопасность информации и сетей), г. Фамагуста, Северный Кипр, 2009 г.; г. Таганрог, 2010 г.;
- 2-м Всемирном конгрессе "Nature and Biologically Inspired Computing" (Вычисления, основанные на природных и биологических процессах), г. Китакюсю, Япония, 2010 г;
- Российско-Немецком семинаре "Innovation Information Technologies: Theory and Practice" (Инновационные информационные технологии: теория и практика), г. Уфа, 2011 г;
- Международном симпозиуме "Engineering Secure Software and Systems" (Проектирование безопасного программного обеспечения и систем), г. Париж, Франция, 2013 г.

**Публикации.** Результаты диссертационной работы отражены в 13 публикациях, в том числе в 4 статьях в рецензируемых журналах, рекомендованных ВАК, в 8 публикациях в сборниках материалов международных и всероссийских конференций и в 1 главе монографии.

**Структура и объем диссертации.** Диссертация состоит из введения, четырех глав, заключения, приложений и библиографического списка. Работа содержит 120 страниц машинописного текста, включая 18 рисунков и 25 таблиц. Библиографический список включает 78 наименований.

## СОДЕРЖАНИЕ ДИССЕРТАЦИИ

**Во введении** обоснована актуальность работы, сформулированы цель и основные задачи исследования, показаны научная новизна и практическая ценность работы, изложены основные положения, выносимые на защиту.

**В первой главе** проводится анализ современных интернет-угроз и существующих подходов к их обнаружению. На основе проведенного анализа сделан вывод о том, что современные методы борьбы с интернет-угрозами обладают существенными недостатками. Так, сигнатурные методы обладают низкой эффективностью при обнаружений новых экземпляров вредоносного кода; они нуждаются в регулярных обновлениях, а создание сигнатур является трудоемким процессом. Эвристические методы, основанные на принципах математической статистики и искусственного интеллекта, призваны восполнить недостатки сигнатурного подхода, однако на сегодняшний день их полномасштабное внедрение также затруднено в силу следующих проблем:

- вероятностный характер решений на основе эвристических подходов;
- высокий уровень ложных срабатываний;
- сложность создания обучающей выборки;
- сложность внедрения и настройки, поскольку стратегия обучения и характер данных для обучающей выборки зависят от специфики вычислительной среды.

Делается вывод о необходимости построения интеллектуальной системы обнаружения вредоносных интернет-страниц на основе технологий машинного обучения, создание которой позволяет во многом преодолеть вышеперечисленные недостатки.

В заключение главы формулируются цели и задачи исследования.

**Во второй главе** проводится анализ исходных текстов вредоносных интернет-приложений, исследуются эксплуатируемые ими уязвимости. Предложена модель интернет-страницы, основанная на теории формальных языков, разработана методика сбора получения репрезентативной обучающей выборки.

В ходе анализа исходных текстов вредоносных интернет-приложений было изучено 33 экземпляра, собрана статистика по реализованных в них функциям. Результаты анализа были использованы для выбора множества признаков интернет-страницы.

Было установлено, что имеется два типа вредоносных страниц: атакующие страницы, непосредственно эксплуатирующие уязвимости в браузерах и их расширениях и узлы перенаправления – инфицированные страницы, содержащие небольшой объем вредоносного кода, направляющего пользователя на атакующую страницу.

Формально задача обнаружения вредоносных страниц сводится к задаче классификации, т.е. разработке функции-гипотезы:

$$h_\theta^{(i)}(x) = \Pr[\, y = i \mid x, \theta], x \in P\,, \qquad (1)$$

где $h: P \rightarrow C$ является отображением множества признаков $P=\{p_j\}$ интернет-страниц на множество $C=\{'benign', 'redirect', 'exploit'\}$ классов, а $\theta$ – вектор параметров модели машинного обучения. При этом $p_i \in \mathbb{R}^n$ является вектором в $n$-мерном пространстве признаков интернет-страницы, $x \in P$ – вектором признаков, а $y \in \mathbb{N}$ – числовым кодом класса. Класс *benign* соответствует безвредным интернет-страницам, *redirect* – узлам перенаправления, *exploit* – атакующим страницам.

Для разработки модели интернет-страницы предлагается использовать теорию формальных языков, а именно – регулярные языки. Таблица 1 содержит базовые составляющие предлагаемой модели.

Таблица 1. Базовые элементы модели интернет-страницы

| **Элемент модели** | **Описание** |
|---|---|
| Алфавит $\Xi = \{0, 1, \dots, 255\}$ | Множество возможных числовых представлений символа размера 1 байт. |
| Документ $D \in L(\Xi^*)$ | Строка алфавита $\Xi$, представляющая собой содержимое интернет-страницы. |
| $\mathbb{T}$ | Множество строк, соответствующих всем допустимым тегам HTML. |
| $\mathbb{W}$ | Множество строк, соответствующих всем допустимым лексемам JavaScript. |
| Символ-пробел S | Регулярное выражение, соответствующее любому символу-пробелу (пробел, табуляция, новая строка и т.д.). |
| Любой символ □ | Регулярное выражение, соответствующее любому символу. |
| JavaScript-сценарий $W_D$ | L(‘<$S^*$*script*□$^*$>□$^*$</*script*>’) |

Таким образом, если представить интернет-страницу D как строку, то регулярные выражения реализуют удобный механизм описания различных элементов страницы, необходимых для вычисления значений признаков. Новые признаки могут вводиться в модель за счет создания регулярных выражений и определения различных функций над соответствующими им множествами строк. Модель интернет-страницы, предложенная в данной работе содержит 73 признака.

Для повышения эффективности алгоритмов классификации необходимо получить сбалансированную и репрезентативную обучающую выборку. Для этого была разработана и реализована методика сбора данных, сводящаяся к следующим шагам:

1. Разработка структуры базы данных для хранения информации об обучающей выборке.
2. Идентификация Интернет-ресурсов, содержащих списки доменных имен и IP адресов, содержащих вредоносное содержимое.

3. Загрузка вредоносных страниц. При этом в качестве параметра User-Agent HTTP-запроса подается строка, выбранная случайным образом из списка наиболее часто атакуемых браузеров. Загрузка безвредных страниц, полученных из списка Alexa.com Top 1'000'000.

4. Проверка полученных данных с помощью сервиса VirusTotal.com, включающего в себя 42 антивирусных сканера.

5. Анализ имен угроз, назначаемых вредоносным страницам антивирусами и приведение их к одному виду.

6. Осуществление случайной выборки (размером 500 экземпляров) из каждого класса вредоносных и безвредных страниц с целью формирования окончательного набора данных.

В качестве источников вредоносных страниц были выбраны:

- «черные» списки вредоносных URL: Malware Domain List (www. malwaredomainlist.com) и Clean-MX (support.clean-mx.de/clean-mx/viruses);
- средство тестирования на проникновение Metasploit, содержащее множество эксплойтов, реализующих drive-by-download атаки;
- вредоносные страницы, хранящиеся в базе данных VX Heavens (vx.netlux.org);
- вредоносные страницы, полученные из исходных кодов связок эксплойтов, проанализированных в данной работе.

**В третьей главе** определяются меры оценки эффективности моделей машинного обучения, проводится предварительный экспериментальный анализ предложенного подхода к решению упрощенной задачи обнаружения вредоносных интернет-страниц на основе технологий машинного обучения. Производится выбор признаков, характеризующих интернет-страницу, рассматриваются различные модели машинного обучения.

Перед реализацией исследовательского прототипа системы необходимо проверить, насколько эффективно машинное обучение при решении задачи обнаружения вредоносных интернет-страниц. Для этого был проведен предварительный эксперимент, в котором использовалась только часть обучающей выборки, содержащая атакующие и безвредные страницы. Соответственно, выходом функции-гипотезы $h(x)$ является только вероятность отнесения вектора признаков $x$ к классу вредоносных страниц. Каждая интернет-страница была представлена в виде вектора из 16 признаков, выбранных вручную. Был построен классификатор на основе персептрона с 20-ю нейронами скрытого слоя и одним нейроном выходного слоя. Эффективность обнаружения вредоносного содержимого для рассмотренной модели составила более 97%, а уровень ложных срабатываний менее 0,002%.

Для решения более сложной задачи, включающей в себя три класса интернет-страниц (безвредные страницы, узлы перенаправления и атакующие страницы), необходимо выбрать множество признаков, учитывающих свойства каждого из обозначенных классов. Сюда относятся признаки, характеризующие вредоносный код JavaScript; признаки HTML-содержимого страницы; признаки, характеризующие безвредный сценарий JavaScript.

В результате анализа было выбрано 73 признака. Часть из них получена на основе ручного анализа исходных текстов вредоносных интернет-приложений, другая часть – путем проведения статистического исследования различных свойств интернет-страниц.

Признаки 1-5 были получены в результате анализа техники эксплуатации уязвимостей повреждения памяти под названием «распыление кучи». Сюда входят такие признаки, как количество закодированных символов UNICODE, которые используются для размещения вредоносных инструкций в адресном пространстве браузера, количество циклов, количество шестнадцатиричных чисел и т.д.

Признаки 6-28 характеризуют обфусцированный (зашифрованный или упакованный) вредоносный код. Хорошими индикаторами обфусцированного кода являются число вхождений функций работы со строками в исходном коде сценария JavaScript, среднее значение и стандартное отклонение информационной энтропии страницы, полученной методом скользящего окна, частота символов-цифр и т.д.

Признаки 29-54 характеризуют HTML-структуру интернет-страницы. Сюда входят такие параметры как количество ссылок на странице, наличие мета-тегов и т.д. Кроме этого, была произведена случайная выборка из 1000 безвредных интернет-страниц, взятых из рейтинга Alexa.com Top 1’000’000, из которой было выделено 20 наиболее популярных HTML-тегов. Число вхождений каждого из этих тегов составляет еще 20 признаков интернет-страницы.

Признаки 55-73 были получены путем исследования частоты вхождений различных токенов (лексических единиц языка JavaScript) во вредоносных страницах. В результате было выбрано 19 токенов, удовлетворяющих условию $\mu_{i2}/\mu_{i1} < 0{,}001$, $\mu_{i3}/\mu_{i1} < 0{,}001$, где $\mu_{ij}$ – среднее количество вхождений i-го токена в страницах j-го класса (j=1 для безвредных страниц, j=2 для узлов перенаправления, j=3 для атакующих страниц).

В качестве примера в таблице 2 представлена небольшая часть предложенных признаков. Функция $size{:}L(\Xi^{+})\rightarrow\mathbb{N}$ вычисляет длину строки над алфавитом $\Xi$, функция $\mathrm{count}{:}\mathcal{R}\times L(\Xi^{+})\rightarrow\mathbb{N}$ определяет количество подстрок, соответствующих заданному регулярному выражению в заданной строке ($\mathcal{R}$ – множество регулярных выражений).

Таблица 2. Некоторые признаки интернет-страницы

| № | **Название признака** | **Функция вычисления признака** |
|---|---|---|
| 1 | Размер кода JavaScript в байтах | size(W) |
| 2 | Количество закодированных символов UNICODE | count(‘%u[A-F0-9]$^{4}$’, W) |
| 3 | Количество вызовов функции *unescape* | count(‘unescape’, W) |
| 4 | Количество шестнадцатиричных чисел | count(‘0x[A-F0-9]$^{+}$’, W) |
| 5 | Количество циклов | count(‘for’+’while’, W) |
| … | … | … |
| 72 | Число вхождений строки “cookie_path” | count(‘cookie_path’, W) |
| 73 | Число вхождений строки “overflow” | count(‘overflow’, W) |

На рисунке 1 представлена визуализация обучающей выборки в 2-х и 3-х измерениях, полученная методом анализа главных компонент. Из рисунка видно, что атакующие страницы и узлы перенаправления визуально разделены в пространстве, в то время как безвредные страницы и узлы перенаправления пересекаются, что значительно усложняет задачу классификации и является источником ложных срабатываний.

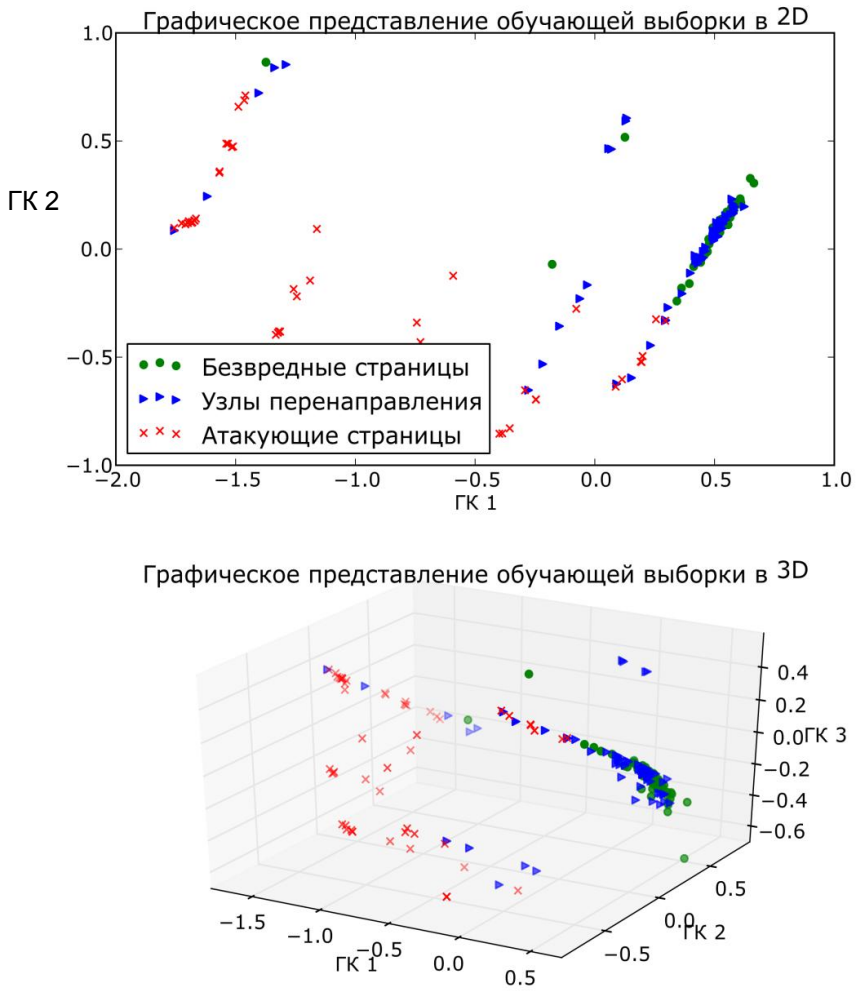


Рисунок 1 – Визуализация обучающих примеров в 2-х и 3-х измерениях

В разработанной системе использовались следующие технологии машинного обучения: логистическая регрессия, нейронная сеть (персептрон с различным количеством нейронов скрытого слоя), машина опорных векторов (модели без ядра и с радиально-базисным ядром), иммунные вычисления.

Поскольку иммунные вычисления являются сравнительно новой и мало исследованной технологией, рассмотрим её более подробно. Ключевой моделью иммунных вычислений является формальная иммунная сеть (ФИС). Иммунная сеть W представляет собой множество клеток $V_i$, $W=\{V_1 \ldots V_m\}$. Клетка V представляет собой дуплет V=(c, P), где $c \in \mathbb{N}$ – класс клетки, $P = (p_1 \ldots p_q) \in \mathbb{R}^q$ –

вектор в q-мерном Евклидовом пространстве, лежащий внутри единичного гиперкуба ( $P \leq 1$). Метрика $d(V_i, V_j)$ представляет собой расстояние между двумя клетками, такое, что $d(V_i,V_j)= P_i - P_j$ , где $P$ означает одну из возможных метрик (например, Евклидова норма, норма Чебышева и т.п.). Клетка $V_i$ «узнает» клетку $V_j$, если обе клетки относятся к одному и тому же классу и расстояние между ними меньше порогового значения h, $d(V_i, V_j) \leq h$. Существует два правила поведения формальной иммунной сети W:

- Апоптоз – если клетка $V_i$ «узнает» клетку $V_j$, то удалить $V_i$ из W;
- Иммунизация – если $V_i$ располагается ближе к $V_j$ чем все остальные клетки иммунной сети W, то добавить $V_i$ в множество W.

Пусть $A=[a_1,a_2,\ldots,a_N]^T$ – матрица, строки которой представляют собой набор обучающих векторов признаков, т.е. $a_i$ это один обучающий образ. Путем сингулярного разложения можно представить матрицу A в виде произведения A=USV, где U и V это матрицы правых и левых сингулярных векторов, а S – диагональная матрица сингулярных чисел. Тогда проекция входного образа Z в пространство ФИС может быть вычислена по формуле:

$$w_i = \frac{1}{s_i} Z^T v_i , \qquad (2)$$

где $w_i$ – i-я величина энергии связи, $s_i$ – i-е сингулярное значение матрицы A, $v_i$ – i-й правый сингулярный вектор матрицы A. Пространство ФИС бывает, как правило, одного, двух или трех измерений (i=1,2 или 3). Матрица U левых сингулярных векторов матрицы A представляет собой клетки формальной иммунной сети, к которым применяются вышеописанные правила.

Важным параметром иммунной сети является минимальное расстояние, на котором клетки узнают друг друга – $h$. Выбор наиболее подходящего порога активации происходит опытным путем, для этого формируется ФИС, после чего вычисляется его индекс неразделимости:

$$i = \ln(m_2) - \ln(m_1) - \ln(h) \qquad (3)$$

Параметр $m_1$ в формуле (2) соответствует начальному числу клеток ФИС, $m_2$ – числу клеток после апоптоза и иммунизации, а $h$ – порог узнавания клеток. Чем меньше индекс неразделимости, тем лучше качество распознавания иммунной сетью.

Алгоритм классификации на основе иммунокомпьютинга может быть представлен следующим образом:

1. Этап обучения:
    1) формирование обучающей матрицы;
    2) вычисление сингулярного разложения;
    3) апоптоз;
    4) иммунизация;
    5) вычисление индекса неразделимости;
    6) повторение шагов 3-5 с другими значениями h;
    7) выбор ФИС с наименьшим значением индекса неразделимости.

2. Этап классификации:
    1) отображение входного образа в пространство ФИС;
    2) определение ближайшей клетки ФИС;
    3) назначение класса ближайшей клетки входному образу.

В разработанной системе обнаружения вредоносных интернет-страниц используется модифицированный алгоритм классификации на основе иммунных вычислений, поскольку в исходном виде, в силу близости классов атакующих страниц и узлов перенаправления, ФИС показала низкую эффективность в ходе предварительных экспериментов. В новой версии рассматриваются 5 ближайших к входному образу клеток ФИС и назначается тот класс, который представлен наибольшим числом клеток.

**В четвертой главе** описывается реализация интеллектуальной системы обнаружения вредоносных интернет-страниц на языке Python, проводится экспериментальный анализ её эффективности и сравнение с существующими антивирусными сканерами.

Разработанная система обнаружения вредоносных интернет-страниц имеет модульную структуру, представленную на рисунке 2.

В состав разработанной системы входят следующие компоненты:

1. *База данных* – хранилище предварительно загруженных и категорированных вредоносных и безвредных интернет-страниц;
2. *Блок извлечения признаков* – подпрограмма, отвечающая за вычисление 73 параметров интернет-страницы;
3. *Блок нормализации признаков* – компонент, осуществляющий масштабирование и нормализацию признаков, а также кодирование номеров классов, соответствующих каждому обучающему примеру;
4. *Конфигурация системы* – множество описаний моделей, где каждая модель представляет собой программную реализацию технологии машинного обучения. Например, моделью может выступать нейронная сеть;
5. *Блок обучения* – модуль, ответственный за обучение нормализованных признаков. Кроме множества обучающих примеров, данный компонент получает на вход также множество описаний моделей, после чего выполняет процедуру обучения для каждой из полученных моделей;
6. *Множество моделей* – полученные в результате обучения программные реализации алгоритмов классификации. Получают на вход множество нормализованных новых примеров, отсутствующих в обучающей выборке;
7. *Модуль анализа данных антивирусами* – компонент, отвечающий за проверку каждого вновь поступающего экземпляра интернет-страницы антивирусами;
8. *Блок корреляции* – подпрограмма, отвечающая за выбор наиболее эффективной модели машинного обучения. Результатом работы блока корреляции является модель, обладающая наилучшими показателями, а корректно классифицированные примеры помещаются в базу данных.

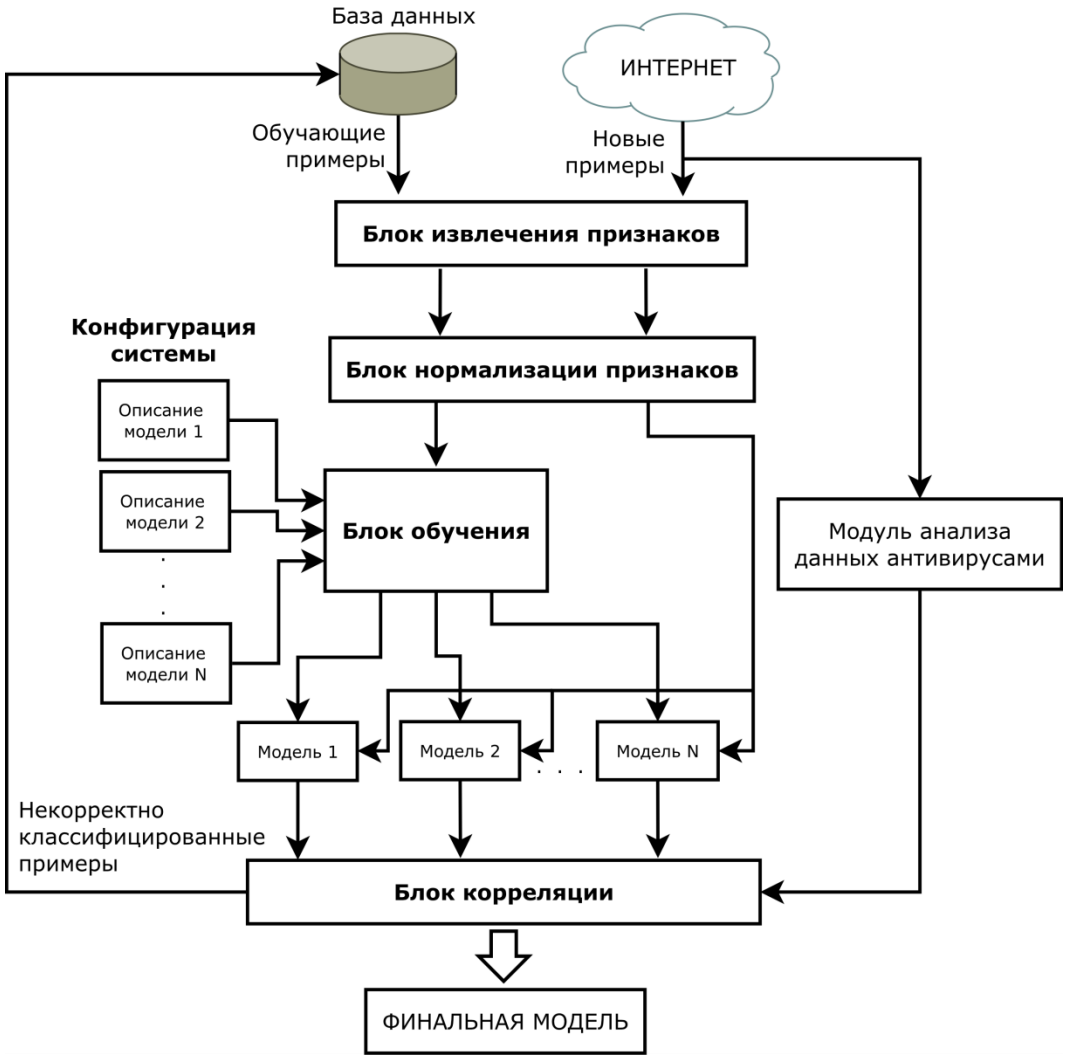


Рисунок 2 – Архитектура разработанной системы обнаружения вредоносных интернет-страниц

Разработанная система работает в двух режимах – обучения и оперативном режиме. В первом случае происходит обучение нескольких моделей на предварительно полученных и категорированных признаках, во втором – анализ новых страниц, загруженных из Интернета. Результатом работы системы является финальная модель машинного обучения, показавшая наилучшие результаты при анализе примеров, не вошедших в обучающую выборку, – эту задачу выполняет блок корреляции. Важным компонентом системы также является модуль анализа данных антивирусами, использующий в своей работе сервис VirusTotal.com. Цель этого компонента – обнаруживать экземпляры, которые система классифицирует некорректно.

Результаты анализа эффективности разработанной системы представлены в таблице 3. В число показателей эффективности машинного обучения входят следующие:

$C_0$, $C_1$, $C_2$ – процент корректно классифицированных безвредных страниц, узлов перенаправления и атакующих страниц соответственно;

$R_{TP}$ – уровень истинно положительных сигналов, показывающий сколько страниц, распознанных как вредоносные, оказались действительно вредоносными (атакующие страницы и узлы перенаправления выступают как один класс);

$R_{FP}$ – уровень ложных срабатываний, показывающий, сколько безвредных страниц было некорректно классифицировано (или как атакующая страница, или как узел перенаправления);

P – показатель точности, характеризующий, насколько «уверена» система, выдавая определенный сигнал;

F – мера, дающая общую оценку классификатора с точки зрения точности и полноты.

Таблица 3. Показатели эффективности моделей машинного обучения

| **Технология машинного обучения** | **$C_0$ (%)** | **$C_1$ (%)** | **$C_2$ (%)** | **$R_{TP}$ (%)** | **$R_{FP}$ (%)** | **P (%)** | **F (%)** |
|---|---|---|---|---|---|---|---|
| Логистическая регрессия | 80 | 66 | 91 | 77 | 20 | 90 | 83 |
| Машина опорных векторов | 85 | 78 | 93 | 83 | 14 | 92 | 87 |
| Нейронная сеть | 89 | 73 | 100 | 87 | 11 | 73 | 80 |
| Иммунная сеть | 84 | 75 | 89 | 81 | 16 | 90 | 85 |

Результаты, приведенные в таблице 3, были получены методом 10-ступенчатой кросс-валидации. Для каждой технологии машинного обучения строилось несколько моделей с различными параметрами (так, например, для нейронной сети – это число нейронов скрытого слоя), из которых выбиралась наиболее эффективная. Наиболее эффективной считается та модель, которая обладает лучшими показателями ($C_0$, $C_1$, $C_2$).

Среди представленных технологий, лучшей для предложенного пространства признаков является нейронная сеть. На рисунке 3 представлены кривые обучения, показывающие изменение процента некорректно классифицированных примеров в зависимости от итерации обучения. Из рисунка видно, что кривая, соответствующая тестовой выборке, практически совпадает с кривой, соответствующей обучающей выборке, что говорит о высокой обобщающей способности сети.

Сравнительный анализ предложенного подхода с современными антивирусными сканерами показал, что разработанная система обнаружения вредоносных интернет-страниц превосходит их по уровням обнаружения узлов перенаправления в среднем на 30% и атакующих страниц на 50%, при этом уступает по уровню ложных срабатываний (11% у разработанной системы и 0% у антивирусных сканеров).

**В заключении** изложены основные результаты работы.

**В приложениях** приведены исходные тексты программ, реализующих разработанные алгоритмы.

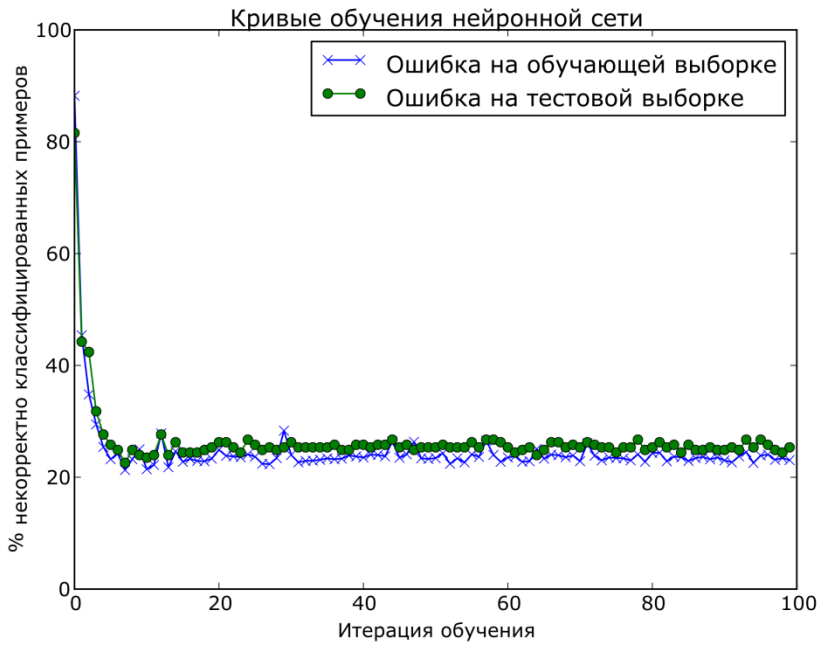


Рисунок 3 – Изменение процента некорректно классифицированных примеров в зависимости от итерации обучения

## ОСНОВНЫЕ ВЫВОДЫ И РЕЗУЛЬТАТЫ РАБОТЫ

1. Произведено исследование существующих вредоносных интернет-приложений как программных артефактов, включающее в себя статистический анализ их возможностей, а также ретроспективное исследование эксплуатируемых уязвимостей. На основе данного исследование обнаружились важные свойства интернет-атак (такие, как техники проведения атак, способы внедрения вредоносного кода, используемые методы защиты от обнаружения), играющие ключевую роль в разработке формальной модели интернет-страницы и выборе множества признаков.

2. Разработана модель интернет-страницы на основе регулярных выражений. Создано 7 базовых элементов (инвариантов) модели, на основе которых в модель было введено 73 признака. Признаки включают в себя характеристики HTML-структуры интернет-страницы (например, количество ссылок, наличие мета-тегов и т.д.), характеристики сценария javascript (например, количество циклов, количество шестнадцатиричных чисел и т.п.), а также общие признаки (такие, как среднее арифметическое и стандартное отклонение энтропии страницы, полученные при помощи скользящего временного окна).

3. Разработана и реализована методика получения выборки для машинного обучения, учитывающая вариабельность вредоносных интернет-документов и позволяющая получить сбалансированный набор данных. Было получено три группы данных: безвредные страницы, узлы перенаправления и атакующие страницы. Из каждой группы была произведена случайная выборка в количестве 500 экземпляров, что в общей сложности составляет 1500

обучающих примеров. Был проведен предварительный анализ обучающей выборки с использованием метода главных компонент, позволившего спроецировать обучающую выборку из пространства 73 измерений в 2-х и 3-х мерные пространства и визуализировать её.

4. Разработаны и исследованы алгоритмы обнаружения вредоносных интернет-страниц на основе технологий машинного обучения. Построены классификаторы на основе логистической регрессии, нейронной сети, машины опорных векторов и иммунных вычислений, предложен модифицированный алгоритм классификации на основе иммунной сети, позволяющий повысить эффективность классификации.

5. Реализован исследовательский прототип интеллектуальной системы обнаружения вредоносных интернет-страниц, реализующий разработанные алгоритмы. Предложена модульная архитектура системы, позволяющая добавлять новые алгоритмы машинного обучения и анализировать их эффективность.

6. Произведен экспериментальный анализ эффективности разработанной системы, в результате которого выбрана модель на основе персептрона, демонстрирующая эффективность обнаружения атакующих страниц, близкую к 100% при уровне ложных срабатываний не выше 11%. Сравнительный анализ предложенного подхода с современными антивирусными сканерами показал, что разработанная система обнаружения вредоносных интернет-страниц превосходит их по уровням обнаружения узлов перенаправления в среднем на 30% и атакующих страниц на 50%, при этом уступая по уровню ложных срабатываний (11% у разработанной системы и 0% у антивирусных сканеров). Достоверность результатов экспериментов достигается путем использования техники 10-ступенчатой кросс-валидации.

**Перспективы дальнейшей разработки темы.** В рамках дальнейших исследований планируется применение разработанных методики и алгоритмов для обнаружения вредоносных PDF-, SWF- и JAR-файлов. Реализация поискового робота, позволяющего идентифицировать новые экземпляры вредоносного содержимого в сети Интернет.

## ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

### В рецензируемых журналах из списка ВАК

1. Система обнаружения сетевых вторжений на основе механизмов иммунной модели / Котов В. Д., Васильев В. И. // Известия ЮФУ. Технические науки. 2011. № 12. С. 180–189.

2. Обнаружение атак на веб-сервер на основе иммунокомпьютинга / Котов В. Д., Васильев В. И. // Вестник компьютерных и информационных технологий. 2012. № 11. С. 45–49.

3. Современное состояние проблемы обнаружения сетевых вторжений / Котов В. Д., Васильев В. И. // Вестник УГАТУ: науч. журнал Уфимск. гос. авиац. техн. ун-та. 2012. Т. 16, № 3 (48). С. 198-204.

4. Обнаружение вредоносных интернет-страниц на основе технологии нейронных сетей / Котов В. Д. // Вестник УГАТУ: науч. журнал Уфимск. гос. авиац. техн. ун-та. 2012. Т. 16, № 8 (53). С. 73-79.

**В других изданиях**

5. Система обнаружения вторжений на основе технологий искусственных иммунных систем / Котов В. Д. // Материалы докладов Всероссийской научно-технической конференции студентов, аспирантов и молодых ученых ТУСУР-2009 – Ч3. – Томск: Изд-во «В-Спектр». 2009. С. 371–381.

6. Система обнаружения вторжений на основе искусственной иммунной системы / Котов В. Д., Васильев В. И. // Материалы докладов 2-й Международной конференции по безопасности информации и сетей. – Нью-Йорк: Изд-во ACM. 2009. С. 207-212 (статья на англ. языке).

7. Система обнаружения вторжений на основе технологий искусственных иммунных систем / Котов В. Д. // Интеллектуальные системы управления. – М: Машиностроение, 2010. С. 525-535.

8. Подход к обнаружению сетевых вторжений, основанный на иммунной модели / Котов В. Д., Васильев В. И. // Материалы докладов 3-й Международной конференции по безопасности информации и сетей. – Нью-Йорк: Изд-во ACM. 2010. С. 233-237 (статья на англ. языке).

9. Обнаружение атак на веб-сервер на основе принципов иммунных вычислений / Котов В. Д., Васильев В. И. // Материалы докладов 2-го всемирного конгресса "Вычисления, основанные на природных и биологических процессах". – Нью-Йорк: Изд-во IEEE. 2010. С. 25-30 (статья на англ. языке).

10. Иммунный подход к обнаружению и нейтрализации последствий инфекций сетевых червей / Котов В. Д., Васильев В. И. // Материалы докладов 4-го Российско-Немецкого семинара "Инновационные информационные технологии: теория и практика". – Уфа: УГАТУ. С. 15-19 (статья на англ. языке).

11. Обзор современных достижений в области обнаружения сетевых вторжений / Котов В. Д., Васильев В. И. // Материалы докладов 13-го международного семинара по информатике и информационным технологиям. Уфа: УГАТУ. 2011. С. 18-21 (статья на англ. языке).

12. Обнаружение вредоносных программ по поведенческим сигнатурам / Котов В. Д., Васильев В. И. // Материалы конференции ”Информационные технологии и системы”. – Челябинск: Изд-во Челяб. гос. ун-та. 2012. С. 70–72.

13. Анатомия связок эксплойтов. Предварительный анализ связок эксплойтов как программных артефактов / Котов В., Массаччи Ф. // Материалы докладов Международного симпозиума по проектированию безопасного программного обеспечения и систем. – Берлин: Изд-во Springer-Verlag. 2013. С. 181-196 (статья на англ. языке).

Диссертант В. Д. Котов