Министерство образования Республики Беларусь Учреждение образования

Белорусский государственный университет
информатики и радиоэлектроники

УДК004.032.26:004.056.5

Коваль
Ольга Сергеевна

Использование нейронных сетей для защиты информации в системе управления предприятия

АВТОРЕФЕРАТ

на соискание степени магистра технических наук
по специальности 1-98 80 01 «Методы и системы защиты информации, информационная безопасность»

_____

Научный руководитель
Вишняков Владимир Анатольевич
доктор технических наук, профессор

_____

Минск 2016

ВВЕДЕНИЕ

Компьютерные сети за несколько последних десятилетий из чисто технического решения превратились в глобальное явление, развитие которого оказывает влияние на большинство сфер экономической деятельности. В настоящее время проблема борьбы с компьютерной преступностью стала одной из первостепенных. Число компьютерных преступлений увеличивается ежегодно на 30 – 40 процентов [1]. Развитие глобальной сети Интернет способствует развитию компьютерных преступлений. С каждым годом киберпреступления охватывают все новые и новые сферы, связанные с компьютерной информацией: вредоносные программы – вирусы, кража конфиденциальной информации, взлом информационных ресурсов и т.д.

Современные исследования в области защиты информации направлены на создание таких методов и алгоритмов защиты, которые были бы способны обнаруживать и нейтрализовать неизвестные компьютерные вирусы, и таким образом не только повысить уровень компьютерной безопасности, но и избавить пользователя от постоянных обновлений антивирусного ПО или его модулей [2].

Предпосылкой для создания эффективных антивирусных систем является развитие нейросетевых технологий и эволюционного программирования, которые имеют биологические основы [4, 5]. Способность таких систем к обучению и обобщению результатов позволяет создавать на базе их интеллектуальные системы защиты информации, которые способны обнаруживать неизвестные компьютерные вирусы.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования. Компьютерные сети за несколько последних десятилетий из чисто технического решения превратились в глобальное явление, развитие которого оказывает влияние на большинство сфер экономической деятельности. В настоящее время одной из наиболее важных задач является защита информационных систем от несанкционированного доступа и минимизация рисков заражения компьютерных систем вирусами, которые также делают системы уязвимыми. В связи с этим тематика исследований является актуальной.

Связь работы с крупными научными программами и темами. Диссертационное исследование выполнялось в рамках НИР «Модели и средства

информационного управления и электронного маркетинга предприятия» № ГР 20115472 от 22.12.2011.

Цели и задачи исследования. *Целью* диссертационной работы является исследование методов и средств информационной безопасности на базе нейронных сетей в системах управления предприятием и обучение эффективной нейросетевой структуры для решения задачи защиты информации во внутрикорпоративных системах. Для достижения поставленной цели необходимо решить следующие *задачи*:

− провести анализ работы нейронных сетей, их архитектуру и основные свойства, правила построения обучающей выборки; ознакомиться с основными алгоритмами обучения нейронных сетей с целью дальнейшего применения для решения задачи защиты информации;

− исследовать модели вредоносного программного обеспечения, а также основные методики обнаружения вредоносных программ, выявить их достоинства и недостатки;

− выбрать и обучить нейросетевую структуру для классификации программного обеспечения во внутрикорпоративных системах на два класса: «чистая программа» и «зараженная программа»;

Объект и предмет исследования. *Объектом* исследования являются вредоносные программы в системе управления предприятия. *Предметом* исследования являются алгоритмы построения нейросетевыхструктур для обнаружения вредоносных программ.

Основные положения исследования, выносимые на защиту:

1. Анализ нейросетевых структур, применимых для классификации программного обеспечения и защиты информации во внутрикорпоративных системах.

2. Методики обнаружения вредоносных программ, их достоинства и недостатки. Методы защиты информации, использующие интеллектуальные технологии.

3. Нейросетевой подход для решения задач защиты информации, состоящий в последовательном объединении двух различных нейронных сетей: рециркуляционной нейронной сети и многослойного персептрона, которые соединяются последовательно. Выборка атрибутов и метаданных исполняемых файлов двух состояний: чистых и зараженных вирусом. Обученная нейронная сеть, исследование эффективности ее работы и возможности дальнейшего использования.

Апробация результатов исследования. Основные положения данной работы представлялись на XIII Белорусско-российской НТК «Технические средства защиты информации» в 2015 году.

Опубликованность результатов. По результатам выполненных исследований опубликована 2 научные работы: статья и тезисы докладов.

Структура и объем диссертации. Структурными частями диссертации являются: общая характеристика работы, введение, 3 главы, заключение, библиографический список, состоящий из 38 наименований, 3 приложения. Работа изложена на 90 страницах и включает в себя 14 рисунков и 20 таблиц.

СОДЕРЖАНИЕ РАБОТЫ

В первой главе рассматривается принцип работы нейронных сетей, их архитектура и основные свойства, правила построения обучающей выборки, а также процесс ее обучения.

Искусственная нейронная сеть – это вычислительная модель человеческого мозга[5,6]. Она представляет собой параллельно-распределенную систему процессорных элементов (нейронов), способных выполнять простейшую обработку данных, которая может настраивать свои параметры в ходе обучения на эмпирических данных[6]. Накопленные знания нейронной сети сосредоточены в весах межэлементный связей.

К основным свойсвам НС относятся такие свойства, как нелинейность, обучение на примерах, параллельная обработка данных, адаптивность, отказоустойчивость.

В математическом смысле искусственный нейрон – это абстрактная модель биологического нейрона[6]. Каждое значение x_i, поступающее по i-той синаптической связи, умножается на вес связи w_i. Тогда взвешенная сумма входов нейрона будет[6]:

$$S = w_1x_1 + w_2x_2 + ... + w_nx_n + b_0 = \sum_{i=1}^{n} w_ix_i + b_0 \qquad (1.1)$$

При реализации нейронной сети и определении ее параметров необходимо выбрать нейросетевую архитектуру и конфигурацию. При этом под архитектурой понимаются общие принципы построения нейронных сетей для определенного класса задач, а под конфигурацией – параметры конкретной нейросетевой модели.

Построение и обучение нейронных сетей часто называют искусством, поскольку выбор конфигурации сети и параметров ее обучения не всегда однозначен.

Наиболее распространенной нейросетевой структурой, применимой для решения широкого спектра задач, выступает многослойный персептрон.

В процессе работы нейронная сеть реализует преобразование данных, которое в общем виде может быть описано функцией многих переменных $Y = f(X)$. Это преобразование определяется конфигурацией сети (числом нейронов, способом их соединения, количеством связей, видом активационной функции и смещениями), а также весами межнейронных связей. Поэтому после того как конфигурация задана, необходимо выполнить соответствующую настройку весов, то есть произвести обучение нейронной сети, чтобы она могла выполнять требуемую обработку данных [6].

Выбор алгоритма зависит от особенностей решаемой задачи, в частности от количества данных, используемых для обучения, их качества и требования к уровню точности решения задачи.

Для современного этапа развития теории и практики обсепечения ЗИ характерно усиленное внимание к безопасности информационных объектов и, как следствие, к повышению требований по ЗИ, а также принятию международных стандартов в области ИБ [7]. Использование иителлектуальных технологий на различных этапах защиты приобретает все большее распространение в системах ЗИ [7].Одной из задач, которые позволяют решать интеллетуальные технологии, является обнаружение неизвестных вторжений/атак [9].Интеллектуальные системы защиты информации, обсепечивающие обнаружения атак, в качестве интеллектуального инструмента используют НС [7-9]. В данном случае НС представляется в виде отдельной системы обнаружения атак. При обработке трафика происходит анализ информации на наличие злоупотреблений, случаи с указанием на атаку перенаправляются к администратору безопасности.

Нейронные сети также используются в криптографии для восстановления достаточно длинных криптографических ключей, что приводит к решению задач повышенной размерности [11]. Решить задачи подобного плана с помощью обычных ЭВМ не представляется возможным. Метод разделения данных является одним из наиболее перспективных средств безопасного хранения криптографических ключей в распределенных вычислительных сетях [12].НС используются и при создании антивирусных систем [7]. Современные исследования в области защиты информации направлены на создание таких методов и алгоритмов защиты, которые были бы способны обнаружить и нейтрализовать неизвестные вредоносные программы, и, как следствие, не только повысить уровень компьютерной безопасности, но и избавить пользователя от постоянных обновлений антивирусного программного обеспечения. НС, а также искусственные иммунные системы [31], базирующиеся на основных принципах и

механизмах биологической иммунной системы, предоставляют новые возможности для создания современных систем защиты информации.

Во второй главе рассматриваются эволюция вредоносного ПО, методики обнаружения вредоносных программ. Также внимание уделяется нейронным сетям при обнаружении атак/аномалий инейросетевым структурам в целом, используемым в задачах защиты информации.

Термин «компьютерный вирус» был впервые введен Фредом Коэном в 1984 году, однако можно смело утверждать, что история компьютерных вирусов начинается с 1940 года, когдаДжон фон Нейман опубликовалсвои исследовательские работы по изучению самовоспроизводящихся математических автоматов.

Новым этапом в развитии вредоносных программ стало появление компьютерной сетевой инфраструктуры. В начале 1970-х годов появляется первая компьютерная сеть – прообраз Интернет – ARPANet[28]. Компьютерная сеть позволила связать большое количество компьютерных систем в единое целое, предоставляя удобные возможности для обмена информацией. Сетевая инфраструктура явилась идеальной средой для существования и размножения компьютерных вирусов.

Следующийэтап развития компьютерных вирусов связан с появлением персональных компьютеров. В результате значительно увеличивается число компьютеров, предназначенных для домашнего пользования - появляется потребность в разнообразном программном обеспечении, удовлетворяющим запросы новых пользователей. Такая ситуация на рынке программного обеспечения значительно упрощает задачу злоумышленников по заражению компьютерных систем.

Ситуация, прикоторой не применяются меры по защите компьютерной информации, стала причиной появления первых вирусных эпидемий, которые затрагивали обширные географические районы, в том числе, выходящие за пределы одного государства.

В 1992 году появляются первые вирусы, противоборствующие антивирусным программам. Часто в функции таких вирусов входили действия по уничтожению сигнатурных баз данных антивирусных продуктов, в результате чего антивирусная программа была не в состоянии корректно функционировать.

В настоящее время существует два основных метода обнаружения вредоносных программ, которые используются в современных антивирусных продуктах: сигнатурный и эвристический [1, 3].

К *сигнатурным методам* относят точные методы обнаружения вирусов и вредоносных программ,которые основываются на сравнении файла с образцами,

называемыми сигнатурами – уникальными фрагментами кода для каждого отдельного компьютерного вируса. Сигнатуры организовываются в антивирусные базы, чем обеспечивают точное обнаружение вредоносных программ.

При сканировании файла могут применяться различные методы сравнения сигнатур с кодом файла [3] :метод скользящего окна,метод контрольных точек,алгоритм Бойера-Мура. Сигнатурный метод обладает серьезным недостатком:он не способен определить новый, неизвестный вирус, а также модификации (применение полиморфных технологий, шифрование кода вируса,упаковка) уже известных вирусов.

Эвристические методы – это приблизительные, или вероятностныеметоды обнаружения вирусов, которые позволяют с определенной уверенностью предположить, что подозрительный файл на самом деле является вредоносной программой [14]. Суть эвристических методов заключается в том, что решение о заражении основывается на некоторых правдоподобных предположениях, а не на строгих выводах из имеющихся фактов и предпосылок.*Методы искусственного интеллекта* обнаружения вредоносныхпрограмм основаны на применении искусственных нейронных сетей и иммунных систем. Такие методы характеризуются вероятностным принципом обнаружения. На сегодняшний день существует несколько исследовательских работ, посвященных обнаружению вредоносных программ, основанных на применении как методов ИНС, так и методов ИИС [25-29],однако они имеют ряд ограничений. Современные исследования в области защиты информации направлены на создание таких методов и алгоритмов защиты, которые были бы способны обнаружить и нейтрализовать неизвестные вредоносные программы, и, как следствие, не только повысить уровень компьютерной безопасности, но и избавить пользователя от постоянных обновлений антивирусного программного обеспечения. Предпосылкой для создания таких систем безопасности стало развитие систем искусственного интеллекта [4, 5]. Искусственные нейронные сети, а также искусственные иммунные системы [31], базирующиеся на основных принципах и механизмах биологической иммунной системы, предоставляют новые возможности для создания современных систем защиты информации. Уникальность антивирусных программ, построенных на основе искусственного интеллекта, заключается в способности обнаружения неизвестных компьютерных вирусов.

В главе 3 происходит обучение нейросетевой структуры для классификации программного обеспечения во внутрикорпоративных системах на два класса: «чистая программа» и «зараженная программа».

Выборка для обучения НС строилась из атрибутов и метаданных PE-структуры исполняемых файлов двух состояний: чистых и зараженных вирусом.Исполняемые файлы, как чистые, так и зараженные вирусами, были предоставлены отделами информационной безопасности СП ЗАО «IBAGroup»и ОАО «Банк БелВЭБ».

Обучающая выборка была составлена из метаданных 100 файлов. На первом этапе все файлы поочередно загружались в *PEExplorer,* откуда их атрибуты и метаданные в шестнадцатеричном представлении сохранялись в Excel-таблицу.

Для обучения многослойного персептрона данные обучающей выборки должны быть приведены к десятичному представлению. Для автоматизации решения данной задачи на языке JavaScriptбыл написан конвертерчисел между различными системами счисления.

Для решения задачи защиты информации был выбран многослойный персептрон - наиболее распространенная нейросетевая структура, используемая для решения широчайшего спектра задач в различных областях деятельности человека.

Обучение проводилось на базе программного продукта IBM«SPSSStatistics».

Относительная погрешность классификации файлов с помощью обученной НС составила 5%, что является достаточно хорошим результатом. Следует, однако, отметить, что при обучении данной нейронной сети использовалась относительно небольшая выборка исполняемых файлов; для использования же нейронной сети при решении *реальных* задач защиты информации, например, во внутрикорпоративных системах, выборка файлов должна быть как можно больше, и вирусы, которыми часть файлов из выборки заражена, должны быть как можно более разнообразными.

ЗАКЛЮЧЕНИЕ

1. Компьютерные сети в настоящее время являются глобальным явлением, развитие которого оказывает влияние на большинство сфер экономической деятельности.В настоящее время одной из наиболее важных задач является защита информационных систем от несанкционированного доступа и минимизация рисков заражения компьютерных систем вирусами, которые также делают системы уязвимыми.

2. Основными методиками обнаружения вредоносных программ являются сигнатурный метод, метод скользящего окна, метод контрольных точек, алгоритм Бойера-Мура. Данные методики обладают существенным недостатком: они не позволяют определить новый, неизвестный вирус, а также модификации

уже известных вирусов. Для решения этой проблемы применяются эвристические (вероятностные) методы, а также поведенческий анализ.

3. Современные исследования в области защиты информации направлены на создание таких методов и алгоритмов защиты, которые были бы способны обнаруживать и нейтрализовать неизвестные компьютерные вирусы, и, таким образом, не только повысить уровень компьютерной безопасности, но и избавить пользователя от постоянных обновлений антивирусного ПО или его модулей.

4. Методы защиты информации, использующие искусственный интеллект, характеризуются вероятностным принципом обнаружения вредоносных программ и предоставляют возможность обнаружить и нейтрализовать неизвестные вирусы.

5. Наиболее распространенной нейросетевой структурой, используемой для решения задач широкого спектра деятельности человека, в том числе и задач защиты информации, является многослойный персептрон. В ходе работы было проведено обучение данной нейросетевой структуры при помощи специально построенной выборки исполняемых файлов двух состояний: чистых и зараженных вирусами. Обучение проводилось в SPSSStatistics – программе, произведенной компанией IBM и предназначенной для статистической обработки информации и данных.

6. Эффективность работы нейронной сети определялась с помощью контрольной выборки исполняемых файлов после ее обучения. Относительная погрешность классификации файлов составила 5%, однако следует отметить, что при обучении данной нейронной сети использовалась относительно небольшая выборка исполняемых файлов; для использования же нейронной сети при решении реальных задач защиты информации, например, во внутрикорпоративных системах, выборка файлов должна быть как можно больше, и вирусы, которыми часть файлов из выборки заражена, должны быть как можно более разнообразными.

7. Многослойный персептрон является базовой конфигурацией нейронной сети, на основе которой можно построить более сложные нейросетевые структуры. В работе был рассмотрен нейросетевой подход решения задач защиты информации, состоящий в последовательном объединении двух различных нейронных сетей: рециркуляционной нейронной сети и многослойного персептрона, которые соединялись последовательно. На первом этапе обработки входной информации в данной нейросетевой структуре происходит уменьшение размерности входного вектора входных данных при помощи нелинейной рециркуляционной нейронной сети. Это позволяет перейти от исходного

пространства данных к вспомогательному, которое характеризуется меньшей размерностью и большей информативностью исходного пространства. Рециркуляционная нейронная сеть позволяет автоматизировать анализ входных параметров, что, безусловно, упрощает процесс построения обучающей выборки. Второй этап состоит в обнаружении и распознавании атак. Для этого используется многослойный персептрон, который осуществляет обработку сжатого пространства входных образов с целью распознавания класса атаки.

8. Существуют так называемые генетические алгоритмы, которые используются при создании искусственной иммунной системы для защиты информации от вредоносных программ. Начальным этапом построения искусственной иммунной системы является генерация иммунных детекторов, в основе которых лежит многослойный персептрон. Пройдя стадии обучения и отбора, детекторы приобретают способность реагировать на вредоносные программы, сканируя их структуру, и, в то же время, игнорировать чистые файлы.

Список публикаций соискателя

1-А. О.С.Коваль, В.А. Вишняков «Обзор и анализ средств защиты информации с использованием искусственных нейронных сетей» // Тезисы докладов XII Белорусско-российской НТК «Технические средства защиты информации». Минск, БГУИР, 2015.

2-А. О.С.Коваль, А.И. Тавгень, Е.В.Яроц «Система сопровождения электронных учебных материалов» // Молодежный Научно-Технический Вестник, ФГБОУ ВПО «МГТУ им. Н.Э. Баумана», эл. NoФС77-51038.