На правах рукописи

ВАГАНОВ МИХАИЛ ЮРЬЕВИЧ

ГИБРИДНАЯ ИСКУССТВЕННАЯ ИММУННАЯ СИСТЕМА ЗАЩИТЫ КОМПЬЮТЕРА ОТ ПРОЦЕССОВ С АНОМАЛЬНОЙ АКТИВНОСТЬЮ

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

АВТОРЕФЕРАТ

диссертации на соискание ученой степени

кандидата технических наук

Омск – 2012

Работа выполнена в Омском государственном университете им. Ф.М. Достоевского

Научный руководитель:
Доктор физико-математических наук, доцент
Белим Сергей Викторович

Официальные оппоненты:
Доктор технических наук, доцент
Лебедев Илья Сергеевич
Доктор технических наук, профессор
Суханов Андрей Вячеславович

Ведущая организация:
ФГБОУ ВПО «БибАДИ»

Защита состоится «15» мая 2012 г. в 15 часов 50 мин. на заседании диссертационного совета Д при национальном исследовательском университете информационных технологий, механики и оптики по адресу:

С диссертацией можно ознакомиться в библиотеке НИУ ИТМО.

Автореферат разослан «14» апреля 2012 г.

Ученый секретарь
диссертационного совета Д
к.т.н., доцент Поляков В.И.

Общая характеристика работы

Актуальность темы

Вместе с ростом распространенности вычислительных систем возрастает количество вредоносных программ, распространяющихся по компьютерным сетям. Основным средством борьбы с вредоносными программами на сегодняшний день является сигнатурный анализ, то есть выявление последовательности команд, наносящих вред компьютеру. Построение баз данных сигнатур компьютерных вирусов является предметом деятельности большого количества компаний, занимающихся выпуском соответствующего программного обеспечения. Этот подход на сегодняшний день приводит к фильтрации вредоносного кода с достаточно высокой эффективностью. Однако в связи с резким ростом объема данных, который наблюдается в последнее время, данный подход наталкивается на ряд сложностей, связанных с необходимостью анализа всех обрабатываемых компьютером данных, что сказывается на производительности системы. Также резко возросла скорость появления новых вирусов и соответственно появилась необходимость увеличения частоты появления новых баз данных сигнатур вредоносного кода. Эти проблемы приводят к необходимости выработки новых альтернативных методов обнаружения и подавления вредоносных процессов.

Одним из новых направлений борьбы с компьютерными вирусами является разработка искусственных иммунных систем. Искусственные иммунные системы используют подходы для борьбы с вредоносным влиянием аналогичные механизмам наблюдаемым у живых организмов. А именно, обнаружение вирусов и выработка иммунного ответа – антител. Такой подход позволяет компьютерным системам дообучаться в процессе функционирования, самостоятельно выявляя компьютерные вирусы по их активности и самостоятельно вырабатывая средства борьбы с вредоносным кодом.

На сегодняшний день разработано несколько систем защиты информации, построенных по аналогии с иммунными системами живых организмов. Следует отметить, что все эти системы носят специализированный характер. Задача построения универсальной системы остается не разрешенной. Однако построение даже специализированных систем имеет высокую ценность, так как существуют большие классы однотипных компьютерных вирусов, нейтрализация которых остается актуальной задачей.

Вследствие способности к дообучению, искусственные иммунные системы могут быть отнесены к системам искусственного интеллекта, которые получают все большее распространение, в том числе и в системах защиты информации.

Целью работы является совершенствование методов борьбы с вредоносным кодом в компьютерных системах общего назначения.

Для достижения поставленной цели были решены следующие **задачи**:

1. Разработка и реализация системы слежения за аномальной активностью процессов.

2. Разработка алгоритма автоматического выделения участков вредоносного кода в легальных процессах.

3. Разработка алгоритма подавления вредоносного кода.

4. Реализация и апробация программного комплекса, реализующего разработанные алгоритмы.

Объектом исследования является вредоносный код, его детектирование и нейтрализация. .

Методы исследования. Для решения поставленных в диссертационной работе задач были использованы методы интеллектуального анализа данных, искусственных иммунных систем и теории нейронных сетей.

Научная новизна

В диссертационной работе разработаны алгоритмы детектирования процессов с аномальной активностью и автоматической выработки механизмов подавления таких процессов. При этом получены следующие результаты, обладающие научной новизной:

1. Разработка и реализация системы слежения за аномальной активностью процессов. Новизна подхода состоит в динамическом отслеживании объема ресурсов, потребляемых процессом на основе системы детекторов в реальном времени.

2. Разработка алгоритма автоматического выделения участков вредоносного кода в легальных процессах. Новизна подхода состоит в поиске участков дизассемблированного кода, отвечающих за аномальное поведение с использованием методов искусственного интеллекта.

3. Разработка алгоритма подавления вредоносного кода. Новизна подхода состоит в автоматическом формировании сервисов, отвечающих за подавление каждого вида аномальной активности – антител.

4. Реализация и апробация программного комплекса, реализующего разработанные алгоритмы. Реализовано антивирусное программное обеспечение с использованием искусственных иммунных систем.

Практическая и научная значимость результатов

1. Разработанная система слежения за аномальной активностью процессов может быть использована не только в рамках искусственной иммунной системы, но и как самостоятельная система обнаружения вторжений.

2. Разработанный алгоритм автоматического выделения участков вредоносного кода позволяет автоматически формировать базы сигнатур компьютерных вирусов.

3. Разработанный алгоритм подавления вредоносного кода, позволяет автоматически нейтрализовывать новые компьютерные вирусы.

4. Реализованный программный комплекс представляет собой закон-

ченный антивирусный пакет с возможностями самостоятельного автоматического дообучения, без регулярного обновления баз вирусов извне.

Основные научные результаты выносимые на защиту

1. Система слежения за аномальной активностью процессов.
2. Алгоритм автоматического выделения участков вредоносного кода.
3. Алгоритм подавления вредоносного кода.
4. Прикладное программное обеспечение, реализующее предложенные алгоритмы.

Апробация работы Основные результаты диссертации докладывались и обсуждались на следующих конференциях: «Информационные технологии и автоматизация управления» (2009 г., г.Омск), IV Международная научно - практическая конференция « Актуальные проблемы безопасности информационных технологий»,(Красноярск, 2010), международная конференция «Автоматизация управления и интеллектуальные системы и среды» (г. Нальчик,2010), XVIII Всероссийская научно-практическая конференция «Проблемы информационной безопасности в системе высшей школы» (г. Москва, 2011).

Публикации Материалы диссертации опубликованы в 6 печатных работах, из них 2 статьи в журналах из списка, рекомендованного ВАК.

Структура и объем диссертации

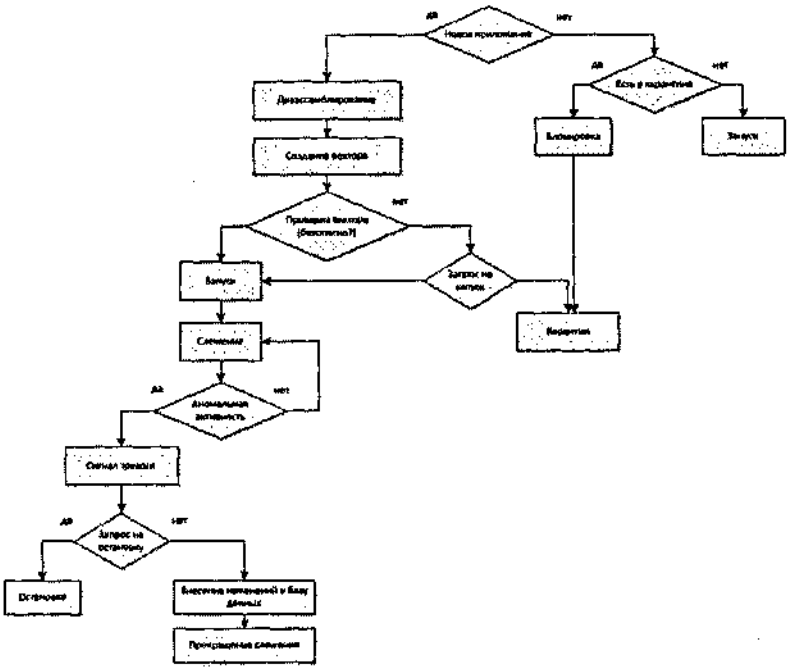
Диссертация состоит из введения, 4 глав, заключения и библиографии. Общий объем диссертации 92 страницы, включая 14 рисунков и 10 таблиц. Библиография включает 103 наименования.

Краткое содержание работы

Во введении обосновывается актуальность темы диссертации, формулируются цель и задачи исследования, обсуждается новизна и практическая ценность выносимых на защиту результатов, даётся краткая характеристика содержания работы.

Первая глава носит обзорный характер и посвящена описанию существующих подходов к созданию искусственных иммунных систем. Также в первой главе приведены основные алгоритмы построения систем обнаружения вторжений.

Вторая глава посвящена описанию архитектуры и алгоритмов работы искусственной иммунной системы. Общая схема функционирования искусственной иммунной системы в штатном режиме приведена на рисунке .



При запуске приложения в системе происходит проверка является ли оно новым. Для выявления новых ведется список обработанных приложений. Если приложение не является новым, то производится проверка, не помещено ли приложение в карантин. Если приложение в карантине, то оно блокируется. Снять блокировку может только пользователь в «ручном» режиме. Если приложение не помещено в карантин, то оно

запускается на выполнение средствами операционной системы.

В том случае, когда происходит запуск не исполнявшегося ранее приложения, запускается алгоритм проверки наличия вредоносных участков кода. В случае присутствия последовательности команд присущих вредоносным программам происходит блокировка приложения и выдается запрос пользователю с предупреждением о возможной «зараженности» приложения. Если пользователь разрешает запуск приложения, то оно переходит в режим выполнения, а искусственная иммунная система в режим слежения. В случае запрета пользователем на выполнение приложение помещается в карантин. Если в результате проверки на наличие вредоносного кода таковой не обнаружен, происходит запуск приложения и переход искусственной иммунной системы в режим слежения.

В режиме слежения система периодически сканирует основные параметры процесса детектируя аномальное поведение. В случае обнаружения завышенного потребления ресурсов система генерирует сигнал тревоги и выдает сообщение пользователю. Если пользователь подтверждает аномальное поведение, происходит анализ исполняемого кода, соответствующие данные добавляются к обучающему множеству, а приложение помещается в карантин. Если пользователь определяет поведение процесса как нормальное, то обновляется профиль нормального поведения и продолжается процесс слежения за процессом.

Режим слежения реализуется с помощью набора детекторов, фиксирующих параметры активности процесса. Реализованные в данной работе детекторы можно разделить на следующие семь групп, отслеживающих интенсивность различных вызовов:

1. Работа с файлами (39 детекторов)

Например: CreateFile, CopyFile, DeleteFile, GetFileType, ReadFile, OpenFile, WriteFile

2. Работа с сетью (33 детектора)

Например: bind, accept, connect, getaddrinfo, listen, recv, send

3. Работа с реестром (19 детекторов)

Например: NtCreateKey, NtDeleteKey, NtEnumerateKey, NtFlushKey, NtOpenKey, NtSetValueKey

4. Работа с процессом авторизации (9 детекторов)

Например: CredUICmdLinePromptForCredentials, CredEnumerate

5. Управление сервисами (17 детекторов)

Например: ChangeServiceConfig, ControlService, CreateService, DeleteService, StartService

6. Управление установкой приложений (13 детекторов)

Например: FindActCtxSectionGuid, CreateActCtx, QueryActCtx

Общее число детекторов составило 130.

Для определения является ли поведение аномальным использовался нейросетевой подход. В качестве нейронной сети был выбран трехслойный персептрон с сигмоидальной функцией отклика. Входной слой содержит 12 нейронов, выходной – один нейрон.

Для определения вредоносных участков кода был разработан алгоритм выделения сигнатур последовательных команд:

1. Дизассебмлирование процесса (посредством запуска консольной версии дизассемблера IDA-32)
2. Поиск стоп-точек (начало и конец процедур)
3. Выделение блоков команд
4. Отбрасывание операндов
5. Запись последовательностей инуструций (сигнатур)
6. Присваивание номера каждой уникальной последовательности.
7. Формирование вектора наличия сигнатур для каждого процесса (каждой последовательности инструкций соответствует координата вектора, значение координат: 1 – присутствует последовательность, 0 – отсутствует).

Таким образом, каждый исполняемый файл можно представить в виде вектора в многомерном Евклидовом пространстве, осями координат которого являются уникальные последовательности инструкций.

$$P_i = p_i^1, p_i^2, ..., p_i^N \tag{0.1}$$

$$p_i^k = \begin{cases} 0, n_{i,j} = 0 \\ 1, n_{i,j} > 0 \end{cases}$$

где P - вектор, характеризующий программу i, p_i^k – булево значение, определяющее, встречается ли последовательность инструкций k в программе i, $n_{i,j}$ - количество нахождений инструкции k в программе i.

Полученная сигнатура используется в двух случаях:

1. Определение наличия вредоносного кода в новых приложениях.
2. При обнаружении аномальной активности.

Наличие вредоносного кода в приложении определяется с помощью еще одной нейросети (трехслойного персептрона), на вход которой подается сигнатурный вектор приложения. Для обучения нейросети используется набор вредоносных и не вредоносных приложений. Более подробно процесс обучения может быть описан следующим образом:

1. Производится сбор большого количества вредоносных программ и гарантированно не вредоносных программ.
2. Производится дизассемблирование и выделения последовательностей команд для всех исполняемых кодов, как вредоносных, так и не вредоносных.
3. Отбрасываем уникальные последовательности команд, встречающиеся менее чем в 10% приложений.
4. Отбрасываем стандартные инструкции, встречающиеся более чем в 85% приложений.
5. Для исключения зависимых последовательностей команд используем

критерий Пирсона χ^2.

6. Каждому приложению, как вредоносному так и не вредоносному сопоставляем сигнатурный вектор.

Нейросеть изначально обучается на некотором наборе приложений. Периодически производится переобучение нейросети с учетом сигнатурных векторов приложений выявленных по аномальной активности системой слежения. Таким образом, система является интеллектуальной и дообучается в процессе функционирования, самостоятельно подстраиваясь под вновь появляющиеся вредоносные программы.

Рассмотренная система является искусственной иммунной, так как обладает основными характеристиками, предъявляющимися к таким системам:

1. Обнаружение вторжений новых вирусов – производится нейросетью на основе аномальной активности.

2. Выработка антител – вычисление сигнатурных векторов и дообучение соответствующей нейросети.

3. Реакция на повторное вторжение (обезвреживание) – проверка сигнатур и помещение в карантин новых приложений, содержащих известные вирусы.

В третьей главе рассмотрена программная реализация искусственной иммунной системы. Подробно описаны реализация детекторов и алгоритм выделения сигнатур.

Для возможности получения данных о текущей деятельности процесса был реализован WDM драйвер режима ядра, включающий в себя функции сбора различных проявления активности процессов, а именно: работу с файловой системой и устройствами, сетевую и межпроцессную активность. Данный драйвер так же отвечает за создание и обновление таблицы процессов, которая требуется для выявления скрытых и замаскированных процессов. В виду достаточно большого количества вари-

антов сокрытия процесса, в данной работе одновременно используется несколько различных способов получения списка процессов (в частности, используя функции ToolHelp API, Native API, ZwQuerySystemInformation, а так же анализ структуры EPROCESS).

Для регистрации всех обращений к файловой системе, в драйвер мониторинга был включен функционал драйвера фильтра файловой системы, основной задачей которого является перехват $IRP-$пакетов с командами IRP_MJ_CREATE. Тот факт, что драйвер фильтра занимает в иерархии более высокий уровень, нежели драйвер файловой системы, позволяет ему модифицировать поток между приложениями и драйвером файловой системы.

Четвертая глава посвящена описанию компьютерного эксперимента. Эксперимент проводился в три стадии. Первая стадия заключалась в обучении подсистемы анализа активности процессов. Во второй стадии формировалась база данных приложения и происходило обучение подсистемы программных профилей. При обучении в качестве входных данных использовались наборы известных вредоносных и нормальных программ. В третьей стадии было произведено тестирование совместной работы подсистем профилей и анализа активности процессов. В поставленном эксперименте были задействованы 8 персональных компьютеров с установленной Windows Seven, 32bit. Каждый из компьютеров являлся рабочим местом одного из трех типов сотрудников (менеджер, оператор или бухгалтер) организации, согласившейся принять участие в эксперименте. Для проверки данных перед началом обучения использовался Norton Antivirus (с последними, на момент тестирования, базами). Выбор платформы (x32) обусловлен возможностью использования бесплатной версии дизассемблера IDA Pro.

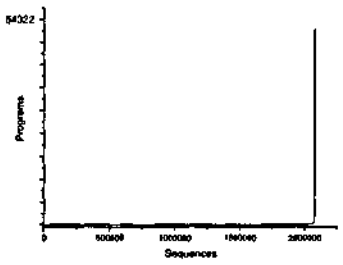
Для составления актуального набора вредоносных программ, были развернуты спам-ловушки на доменах gorodomsk.org, nobrain.ru и email.ws.

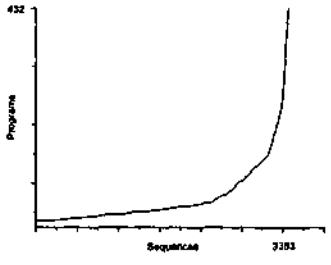
Так же были реализованы скрипты, позволяющие переходить по ссылкам в теле письма и сохранять вредоносное программное обеспечение. Общее количество извлеченных исполняемых вложений (включая архивы) составило 4000. Полученные исполняемые файлы были разделены на две группы – одна использовалась на стадии обучения, вторая – в ходе последующего тестирования.

В процессе создания базового набора детекторов использовались данные о нормальной активности процессов, собранные за 480 часов (60 рабочих дней), а так же данные о суточной активности вредоносных программ из первого набора.

Для подготовки данных ко второй стадии эксперимента, были реализованные скрипты на языке python, копирующие исполняемые файлы с расширениями exe и dll на сетевой диск. Далее было произведено удаление повторяющихся файлов, после чего количество файлов для анализа составило 65 000 файлов. Непосредственно перед началом дизассемблирования, было проведено антивирусное сканирование, что позволило исключить возможность попадания вредоносного программного обеспечения.

В процессе дизассемблирования вредоносных и нормальных программ было получено 65 000 файлов .asm. Из каждого файла были извлечены последовательности инструкций. Общее число уникальных последовательностей составило 1974179 .

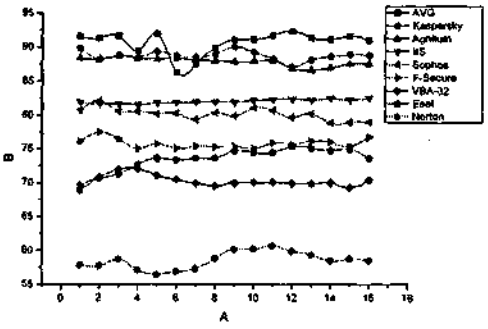


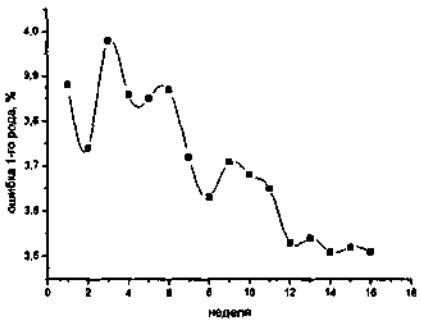


После применения метода последовательного сокращения, количество последовательностей сократилось до 3383 . После применения метода Пирсона, число уникальных последовательностей составило 623.

Для каждой из программ был создан профайл, который, в последствии использовался для обучения нейронной сети. В качестве модели нейронной сети был выбран трехслойный персептрон. Количество нейронов входного слоя равно 632, количество нейронов скрытого и третьего слоев составляло 211 и 2 соответственно. В качестве метода обучения использовался метод коррекции ошибки (обучение с учителем). Для реализации выбранной модели использовалась библиотека FANN и соответствующий интерфейс для языка python.

Для тестирования обнаружения вредоносных программ использовалась вторая группа вредоносных программ, не использованных при обучении нейронной сети. Процент верных обнаружений составил 82%. Значение ошибок первого рода не превысило 4%. В таблице представлены результатов работы искусственной иммунной системы существующими на рынке антивирусными решениями.





Как хорошо видно из графиков при выбранном обучающем наборе построенная искусственная иммунная система проигрывает в эффективности наилучшим антивирусным программам порядка 6%. Однако данный факт легко объясняется маленьким размером использованного обучающего множества по сравнению с антивирусными базами коммерческих программных продуктов. Из рисунка 3 хорошо виден монотонный рост эффективности разработанной системы с течением времени вследствие дообучения. Поведение графика на рисунке 4 демонстрирует существенное снижение числа ошибок первого рода с течением времени.

	Базы Август 2011		Базы Февраль 2011	
	Обнаружено	Ошибка 1-го рода	Обнаружено	Ошибка 1-го рода
AVG	91.60%	0.00%	64.90%	0.02%
Kaspersky	89.80%	2.97%	48.10%	0.01%
Agnitum	88.42%	0.03%	32.60%	0.07%
BitDefender	88.40%	2.16%	54.10%	0.04%
Avira	86.80%	0.84%	64.70%	0.13%
Avast	85.96%	0.13%	41.00%	0.03%
Panda Security	84.90%	0.05%	34.60%	0.02%
Trend Micro	84.70%	0.18%	43.40%	0.04%
DrWeb	84.20%	0.69%	37.70%	0.20%
ИИС	81.87%	3.94%	-	-
Sophos	80.70%	0.01%	64.20%	2.24%
F-Secure	76.10%	0.08%	68.50%	0.04%
VBA32	69.60%	0.55%	35.10%	0.07%
Eset	68.90%	0.04%	38.70%	0.02%
Norton	57.90%	0.11%	39.10%	0.02%

Из таблицы видно, что результаты построенной искусственной системы сравнимы, а иногда и превосходят результаты большинства антивирусных программ. При этом эффективность всех антивирусных программ существенно снижается при отсутствии внешнего обновления баз данных. Тогда как эффективность разработанной искусственной иммунной системы системы монотонно повышается вследствие самостоятельного обучения и не требует обновления из внешних источников.

В Заключении приведены результаты работы.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЫВОДЫ

В диссертационной работе содержится решение научной задачи разработки алгоритмов автоматического детектирования процессов с аномальным поведением, а также автоматическое формирование системы подавления аномальных процессов.

В ходе исследования получены следующие результаты:

1. Разработана и реализованная искусственная иммунная система слежения за аномальной активностью процессов. В рамках данной системы выработаны механизмы слежения за аномальной активностью процессов. Система строится на основе технологий искусственного интеллекта и обладает способностью к обучению.

2. Предложен алгоритм автоматического выделения участков вредоносного кода. Для каждого процесса строится профиль, характеризующий последовательности команд присущие данному процессу. После устранения не характерных инструкций и связанных последовательностей, удается вычленить сигнатуры приемлемой длины для обнаружения вредоносного кода.

3. Предложен алгоритм подавления вредоносного кода. Данный алгоритм использует нейросетевой подход на основе сигнатур последовательностей команд и обладает высокой эффективностью.

4. Разработано прикладное программное обеспечение на базе языка C++ для детектирования процессов с аномальной активностью и выработки системы подавления подобных процессов в дальнейшем. Разработанное программное обеспечение обладает достаточно высокой эффективностью, сравнимой с большинством коммерческих решений. При этом система способна к самообучению, не требует постоянных обновлений баз данных и характеризуется высокой стабильностью.

СПИСОК ПУБЛИКАЦИЙ

Журналы из списка, рекомендованного ВАК

1. Ваганов М.Ю. Разработка искусственной иммунной системы, предназначенной для обнаружения заражений компьютерной системы // Безопасность информационных технологий. 2011. №1. С. 80–81.

2. Ваганов М.Ю. Обнаружение вредоносных программ на основе анализа активности рабочей станции //Вестник омского университета. 2010. №4. С. 134–136.

Другие издания

3. Ваганов М.Ю. Реализация системы обнаружения вторжений как части искусственной иммунной системы // Математические структуры и моделирование, 2010, вып. 21, С. 104-112.

4. Ваганов М.Ю. Программная реализация модели искусственной иммунной систем // Материалы первой международной конференции «Автоматизация управления и интеллектуальные системы и среды» (АУИСС-2010). Россия, Терскол. - Нальчик: Изд-во КБНЦ РАН, 2010. Т.4. С. 125-126

5. Ваганов М.Ю. Белим С.В. Сетевые искусственные иммунные системы // *Информационные технологии и автоматизация управления: материалы науч.-практ. конф.* ОмГТУ, 20-24 апреля 2009 года – Омск: Изд-во ОмГТУ, 2009.

6. Ваганов М.Ю. Система обнаружения вредоносных программ на основе анализа активности рабочей станции. // Сборник материалов IV Международной научно-практической конференции « Актуальные проблемы безопасности информационных технологий», Красноярск, 2010, С.44-45.

Подписано в печать 29.02.12
Формат 60x84x16, бумага писчая.
Оперативный способ печати.
Усл. печ. л. 1,5. Тираж 100 экз., заказ № 0112

Отпечатано в «Полиграфическом центре КАН»
тел. (3812) 24-70-79, 8-904-585-98-84.
E-mail: pc_kan@mail.ru
644050, г. Омск, ул. Красный Путь, 30
Лицензия ПЛД № 58-47 от 21.04.97