МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ М.В.ЛОМОНОСОВА
(МГУ)

№ госрегистрации 114100140107

УТВЕРЖДАЮ
Заместитель проректора МГУ
_____ А.Э.Сазонов
«30» июня 2016 г.

ОТЧЕТ
О ПРИКЛАДНЫХ НАУЧНЫХ ИССЛЕДОВАНИЯХ

Исследование и разработка инновационной технологии построения программных средств обеспечения компьютерной безопасности, основанных на использовании методов машинного обучения и математической статистики для анализа данных поведенческой биометрии пользователей при работе в рамках стандартного человеко-машинного интерфейса, для решения задач активной аутентификации и идентификации пользователей, обнаружения внутренних вторжений и предотвращения попыток хищения конфиденциальной информации

по теме:
ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ ПОСТАВЛЕННЫХ ПЕРЕД ПНИ ЗАДАЧ

(промежуточный)

Этап 4

ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2014-2020 годы»

Соглашение о предоставлении субсидии от 27.06.2014 № 14.604.21.0056

Руководитель ПНИ,
д.ф.-м.н., профессор _____ 30.06.2016 _____ И.В.Машечкин
подпись, дата

Москва 2016

СПИСОК ИСПОЛНИТЕЛЕЙ

Руководитель проекта д.ф.-м.н., профессор	30.06.16 подпись, дата	И.В. Машечкин (введение, заключение, реферат)
Исполнители темы		
к.ф.м.н., доцент	30.06.16 подпись, дата	М.И. Петровский (подразделы 1.1, 1.3)
м.н.с.	30.06.16 подпись, дата	И.С. Попов (подраздел 1.2)
к.ф.м.н., доцент	30/06/16 подпись, дата	А.Н. Терехин (подраздел 1.3, приложение А)
математик 1-й кат.	30.06.16 подпись, дата	Д.В. Царёв (подраздел 1.2, раздел 3)
математик 1-й кат.	30.06.16 подпись, дата	П.М.Саликов (разделы 2, 3, приложение А)
математик 1-й кат.	30.06.16 подпись, дата	О.Е.Горохов (разделы 2, 3,приложение А)
программист	30.06.16 подпись, дата	Д.А.Никифоров (разделы 2, 3, приложение А)
нормоконтролер к.ф.м.н., ассистент	30.06.16 подпись, дата	В.В. Глазкова (приложение Б, В)

РЕФЕРАТ

Отчет 200 с., 61 рис., 15 табл, 18 источника, 3 прил.

КОМПЬЮТЕРНАЯ БЕЗОПАСНОСТЬ, ПОВЕДЕНЧЕСКАЯ БИОМЕТРИЯ, АКТИВНАЯ АУТЕНТИФИКАЦИЯ, ИДЕНТИФИКАЦИЯ ПОЛЬЗОВАТЕЛЕЙ, ДИНАМИКА РАБОТЫ С КЛАВИАТУРОЙ И МЫШЬЮ, ОБНАРУЖЕНИЕ ВНУТРЕННИХ ВТОРЖЕНИЙ, ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ, МАШИННОЕ ОБУЧЕНИЕ, МОДЕЛИРОВАНИЕ ПОВЕДЕНИЯ ПОЛЬЗОВАТЕЛЯ.

Объектом исследования являются методы анализа данных поведенческой биометрии пользователей при работе в рамках стандартного человеко-машинного интерфейса, для решения задач компьютерной безопасности, включая задачи активной аутентификации и идентификации пользователей, обнаружения внутренних вторжений и предотвращения попыток хищения конфиденциальной информации.

Отчет содержит описание исследований, которые были проведены на 4 этапе работ настоящего ПНИ. В результате исследований, были получены следующие основные результаты:

- в соответствие с п. 3.13 Технического задания проведены экспериментальные исследования ЭО ПК для обеспечения компьютерной безопасности на основе анализа поведенческой информации работы пользователей компьютерных систем по разработанной Программе и методикам экспериментальных исследований, продемонстрировано соответствие полученных теоретических и практических исследований требованиям технического задания;
- проведен анализ и оценка РИД, обоснована актуальность и перспективы использования данных РИД для построения инновационных программных систем перспективного класса UEBA (User and Entity Behavior Analytics – анализ поведения пользователей и систем), основанных на анализе характеристик работы пользователя с текстовой информацией (содержимое файлов, электронных писем, коротких сообщений, интернет информации).

СОДЕРЖАНИЕ

ОПРЕДЕЛЕНИЯ, ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

В настоящем отчете о НИР применяют следующие термины с соответствующими определениями.

ТЗ	Техническое задание на выполнение прикладных научных исследований (Приложение 1 к Соглашению № 14.604.21.0056 о предоставлении субсидии от 27.06.2014).
Контент документа	Содержимое документа.
ИБ	Информационная безопасность.
ИАД	Интеллектуальный анализ данных.
ППК	Прикладной программный комплекс.
ЭО ПК	Экспериментальный образец программного комплекса.

ВВЕДЕНИЕ

Цель работы в целом — исследование и разработка комплекса научных решений, направленных на создание программных средств анализа индивидуальных особенностей поведения пользователей компьютерных систем (поведенческой биометрии) при работе в рамках стандартного человеко-машинного интерфейса, с целью создания инновационной технологии построения систем компьютерной безопасности.

В рамках настоящих ПНИ проводились работы, соответствующие четвертому этапу «Экспериментальные исследования поставленных перед ПНИ задач». В рамках отчетного этапа решались следующие задачи:

- проведение экспериментальных исследований ЭО ПК в соответствие с Программной и методикой экспериментальных исследований;
- обеспечение работоспособности рабочих станций, общего системного и специального программного обеспечения для проведения работ по проекту, проведение регламентных работ по обеспечению сохранности данных и программ, относящихся к проводимым работам по проекту;
- проведение оценки РИД, полученных при выполнении ПНИ, с целью их вовлечения в хозяйственный оборот.

Результатом исследований настоящего этапа является обоснование характеристик и свойств разработанного ЭО ПК и подтверждение соответствия результатов ПНИ требованиям Технического задания. В отчете содержится информация о проведенных экспериментальных исследованиях по разработанной Программе и методикам экспериментальных исследований ЭО ПК по следующим направлениям:

- оценка скорости работы методов сбора и предобработки поведенческих биометрических данных, а также объемов собираемых биометрических данных в условиях реальной работы;
- оценка точности и скорости работы методов предобработки текстовых данных, в том числе методов рубрикации, группировки, автоматического аннотирования и выявления ключевых слов на экспериментальных и/или эталонных тестовых данных;
- оценка точности и скорости работы методов машинного обучения и математической статистики для построения и применения поведенческих моделей.

В отчете содержится информация о проведении оценки РИД, полученных при выполнении ПНИ, с целью их вовлечения в хозяйственный оборот. Основным результатом анализа РИД является обоснование перспективности использования результатов для построения инновационных систем IT безопасности. В частности, одна из ведущих мировых компаний, занимающихся анализом рынка информационных технологий Gartner (http://www.gartner.com) сделала вывод о перспективности технологий UEBA (User and Entity Behavior Analytics – анализ поведения пользователей и систем). И отсутствии в настоящее время решений, класса систем UEBA, позволяющих использовать для решения задач информационной безопасности анализ характеристик работы пользователей с текстовой информацией (в частности утверждается, что в настоящее время таких систем в мире нет и они могут появиться только через несколько лет).

В отчете содержится информация о выполнении обеспечения работоспособности рабочих станций, общего системного и специального программного обеспечения для проведения работ по проекту и проведении оценки РИД, полученных при выполнении ПНИ, с целью их вовлечения в хозяйственный оборот.

По результатам полученным на 3-м этапе настоящих ПНИ были поданы на регистрацию РИД две заявки в Федеральный институт промышленной собственности (ФИПС) о регистрации программы для ЭВМ «Система мониторинга, теневого копирования и автоматического аннотирования текстовых данных при работе пользователя с электронными документами» и «Система двухфакторной аутентификации на основе анализа поведенческой биометрической информации об особенностях работы пользователя с компьютерной мышью». Было проведено дополнительное патентное исследование. Проведен поиск патентных документов, включая заявки и патенты на изобретения, полезные модели, технические решения которых касаются методов анализа данных поведенческой биометрии пользователей при работе в рамках стандартного человеко-машинного интерфейса, для решения задач компьютерной безопасности, включая задачи активной аутентификации и идентификации пользователей, обнаружения внутренних вторжений и предотвращения попыток хищения конфиденциальной информации. По результатам патентного анализа сделан вывод об отсутствии конфликтов с существующими патентами и о возможности получения правовой охраны планируемых РИД настоящих ПНИ.

1 Проведение экспериментальных исследований по разработанной Программе и методикам экспериментальных исследований ЭО ПК

С целью демонстрации соответствия результатов теоретических исследований настоящих ПНИ требованиям ТЗ и в соответствие с пунктом 4.1 плана графика исполнения обязательств при выполнении ПНИ, были проведены экспериментальные исследования по разработанной Программе и методикам экспериментальных исследований ЭО ПК по следующим направлениям:

- оценка скорости работы методов сбора и предобработки поведенческих биометрических данных, а также объемов собираемых биометрических данных в условиях реальной работы;
- оценка точности и скорости работы методов предобработки текстовых данных, в том числе методов рубрикации, группировки, автоматического аннотирования и выявления ключевых слов на экспериментальных и/или эталонных тестовых данных;
- оценка точности и скорости работы методов машинного обучения и математической статистики для построения и применения поведенческих моделей.

1.1 Оценка скорости работы методов сбора и предобработки поведенческих биометрических данных, а также объемов собираемых биометрических данных в условиях реальной работы

1.1.1 Оценка скорости работы методов сбора и предобработки биометрической информации об особенностях динамики работы пользователя с устройствами ввода-вывода в рамках стандартного человеко-машинного интерфейса

1.1.1.1 Оценка скорости работы методов сбора и предобработки биометрической информации для решения задачи фоновой идентификации пользователей на основе событий ввода данных с помощью клавиатуры и мыши

Общая схема постановки экспериментов

Проводимые эксперименты состояли из следующих шагов: сбор данных на рабочих местах пользователей, предобработка данных (очистка, выделение составных событий,

разбиение данных на окна на основе порогов на минимальное и максимальное количество событий в окне, расчет векторов признаков на основе выбранной модели представления), постобработка рассчитанных векторов (дискретизация по квантилям, выбор наиболее стабильных или наиболее значимых признаков), формирование из полученных векторов обучающих и тестовых наборов, построение модели на основе различных одноклассовых методов классификации на обучающем наборе данных, применение построенных моделей для классификации тестового набора, оценка точности полученных результатов.

Объем собираемых биометрических данных динамики работы с клавиатурой и мышью определяется суммарным размером данных, используемых для:

- хранения исходных событий о работе пользователя с клавиатурой и мышью; события хранятся в виде текстовых файлов в целевых каталогах;
- хранения предобработанных данных, в виде промежуточных текстовых файлов в тех же каталогах;
- хранения рассчитанных векторов признаков, соответствующих временным окнам, по которым разделяется пользовательская активность динамики работы с клавиатурой и мышью; вектора хранятся в виде текстовых файлов.

Таким образом, размер собираемых данных соответствует суммарному размеру объектов целевого каталога, используемого компонентами сбора в качестве корневого для текущей пользовательской сессии. Оценка объема производилась в следующих режимах работы пользователя:

- объём пиковой нагрузки генерации данных, производимой при активной работе пользователя с устройствами ввода непрерывно в течении одного часа;
- объем дневной генерации данных, производимых пользователем в течение одной сессии типичного восьмичасового рабочего дня;
- недельный объём данных – суммарный объем ежедневных сессий пользователя, в течение одной недели.

При проведении указанных экспериментальных исследований одним из основных критериев оценки скорости работы методов сбора является "незаметность" работы компонент сбора для пользователя – задержка реакции на действия пользователя с устройствами ввода во всех используемых приложениях не должна визуально отличаться от поведения приложений при неактивной компоненте сбора.

Также при проведении сбора данных оценивалось количество выполненных событий, которые успевал обработать перехватчик за отведенное время. Данная проверка связана со

спецификой реализации режима перехвата пользовательских событий ввода на целевой платформе: ОС Windows принудительно отключает сторонние перехватчики событий ввода, в случае если время обработки потока событий ими превышает системный порог. Данная проверка применялась к каждому выполненному событию и сопоставлялась с информацией логов программной компоненты сбора данных. Проверка статуса активности перехватчика показала, что перехватчик успевает обработать все выполненные события за отведённое время.

Оценка скорости работы методов сбора и предобработки поведенческих биометрических данных динамики работы с клавиатурой, а также объемов собираемых биометрических данных в условиях реальной работы

Настоящие экспериментальные исследования проводились в соответствие с пунктом 4.3.1.1 Программы и методики экспериментальных исследований. Целью проводимых экспериментальных исследований является оценка скорости работы методов сбора и предобработки поведенческих биометрических данных динамики работы с клавиатурой, а также объемов собираемых биометрических данных для решения задач фоновой идентификации пользователей.

В качестве набора данных, участвующих в эксперименте, использовался набор данных динамики работы с клавиатурой, сформированный в процессе локального сбора при фоновой работе пользователей за компьютерами. Набор содержит данные работы девяти пользователей, собранные в течение трех-пяти дней в фоновом режиме при работе за компьютером. Производилась оценка скорости сбора данных как количество событий, совершенных пользователем с клавиатурой в единицу времени. Рассматриваемые наборы содержат по несколько тысяч событий динамики работы с клавиатурой на каждого пользователя.

Экспериментальные исследования по оценке скорости и объема собираемых данных проводились на рабочей станции WST_WIN7 (программные характеристики - ОС Windows 7 (32bit); аппаратные характеристики: процессор Core2 Duo, частота 2.4 ГГц, 2 ядра; оперативная память 1Гбайт; дисковый накопитель HDD, 80 Гбайт; сетевой адаптер 100Мбит/сек).

Результаты по скорости сбора и объему собираемых данных динамики работы с клавиатурой приведены в протоколе №1 по пункту № 4.3.1.1 Программы и методики экспериментальных исследований «Подсистемы 1» ЭО ПК для сбора и анализа

биометрической информации об особенностях динамики работы пользователя с устройствами ввода-вывода в рамках стандартного человеко-машинного интерфейса (см. Приложение А: таблица 1 Протокола испытаний №1).

Для проведения экспериментов по проверке «незаметности» работы компонента сбора данных динамики работы с клавиатурой были привлечены десять независимых экспертов, на рабочие станции которых были установлены компоненты сбора; в процессе их работы был собран объем 15Мб данных динамики работы с клавиатурой и задержка реакции на действия пользователя с устройствами ввода визуально не отличалась от работы при неактивной компоненте сбора.

Результаты по скорости работы методов предобработки для рассматриваемого набора данных динамики работы с клавиатурой приведены в протоколе №2 по пункту № 4.3.1.1 Программы и методики экспериментальных исследований «Подсистемы 1» ЭО ПК для сбора и анализа биометрической информации об особенностях динамики работы пользователя с устройствами ввода-вывода в рамках стандартного человеко-машинного интерфейса (см. Приложение А: таблица 3 Протокола испытаний №3).

Оценка скорости работы методов сбора и предобработки поведенческих биометрических данных динамики работы с мышью, а также объемов собираемых биометрических данных в условиях реальной работы

Настоящие экспериментальные исследования проводились в соответствие с пунктом 4.3.1.1 Программы и методики экспериментальных исследований. Целью проводимых экспериментальных исследований является оценка скорости работы методов сбора и предобработки поведенческих биометрических данных динамики работы с мышью, а также объемов собираемых биометрических данных для решения задач фоновой идентификации пользователей.

В качестве набора данных, участвующих в эксперименте, использовался набор данных динамики работы с мышью, сформированный в процессе локального сбора при фоновой работе пользователей за компьютерами. Набор содержит данные работы девяти пользователей, собранные в течение трех-пяти дней в фоновом режиме при работе за компьютером. Производилась оценка скорости сбора данных как количество событий, совершенных пользователем с мышью в единицу времени. Рассматриваемые наборы содержат по несколько сотен тысяч событий динамики работы с мышью на каждого пользователя.

Экспериментальные исследования по оценке скорости и объема собираемых данных проводились на рабочей станции WST_WIN7 (программные характеристики - ОС Windows 7

(32bit); аппаратные характеристики: процессор Core2 Duo, частота 2.4 ГГц, 2 ядра; оперативная память 1Гбайт; дисковый накопитель HDD, 80 Гбайт; сетевой адаптер 100Мбит/сек).

Результаты по скорости сбора и объему собираемых данных динамики работы с мышью приведены в протоколе №1 по пункту № 4.3.1.1 Программы и методики экспериментальных исследований «Подсистемы 1» ЭО ПК для сбора и анализа биометрической информации об особенностях динамики работы пользователя с устройствами ввода-вывода в рамках стандартного человеко-машинного интерфейса (см. Приложение А: таблица 1 Протокола испытаний №1).

Для проведения экспериментов по проверке «незаметности» работы компонента сбора данных динамики работы с мышью были привлечены десять независимых экспертов, на рабочие станции которых были установлены компоненты сбора; в процессе их работы был собран объем 234 Мб данных динамики работы с мышью и задержка реакции на действия пользователя с устройствами ввода визуально не отличалась от работы при неактивной компоненте сбора.

Результаты по скорости работы методов предобработки для рассматриваемого набора данных динамики работы с мышью приведены в протоколе №2 по пункту № 4.3.1.1 Программы и методики экспериментальных исследований «Подсистемы 1» ЭО ПК для сбора и анализа биометрической информации об особенностях динамики работы пользователя с устройствами ввода-вывода в рамках стандартного человеко-машинного интерфейса (см. Приложение А: таблица 3 Протокола испытаний №3).

1.1.1.2 Оценка скорости работы методов сбора и предобработки биометрической информации для решения задачи активной аутентификации пользователей на основе событий ввода данных с помощью клавиатуры и мыши

Экспериментальные исследования проводились в соответствие с пунктом 4.3.1.1 Программы и методики экспериментальных исследований. Целью проводимых экспериментальных исследований является оценка скорости работы методов сбора и предобработки биометрической информации, а также объемов собираемых данных для решения задачи активной аутентификации пользователей на основе событий ввода данных с помощью клавиатуры и мыши.

При проведении указанных экспериментальных исследований, одним из основных критериев оценки скорости работы методов сбора является время отклика элементов

пользовательского интерфейса при нажатиях пользователем клавиш клавиатуры и(или) осуществлении действий с мышью – при проведении серии действий по вводу данных для осуществления обучения модели, либо непосредственно при вводе данных для совершения попытки авторизации. Время отклика считается приемлемым, в случае отсутствия явно видимых пользователю задержек от момента совершения действий по вводу данных и до момента получения ответной реакции от системы. Для проведения экспериментов по этой проверке были привлечены десять независимых экспертов, которые осуществляли ввод парольных слов и графических жестов; явно видимых пользователю задержек от момента совершения действий по вводу данных и до момента получения ответной реакции от системы обнаружено не было.

Объем собираемых данных определяется суммарным размером временных файлов, используемых для промежуточного хранения вычисляемых биометрических характеристик, и зависит от числа учитываемых попыток в модели обучения пользователя. Для проведения экспериментов оценивался суммарный объем данных при обычной схеме обучения (до 15 удачных попыток ввода ключевого слова при использовании клавиатуры и до 10 попыток повторения заданного шаблона при использовании мыши), и при увеличенном/максимальном числе попыток (до 50 повторений).

Средний объём файла, получаемого после сбора данных, необходимых для построения модели использования мыши пользователем, составил 155 Кб на 10 повторений шаблона. В силу линейности увеличения размера файла при увеличении количества подписей, средний размер модели из 50 подписей составил полтора мегабайта. При этом объём, занимаемый одной подписью, линейно зависит от длительности времени её ввода (считается только время с нажатой кнопкой мыши - то есть суммарное время рисования всех фрагментов подписи). В проводимых экспериментах среднее время ввода тестовой подписи составляло 5 секунд. В случае изменения длительности ввода подписи размер файла будет изменяться в такой же пропорции.

При аутентификации на основе динамики работы с клавиатурой средний объем файла, содержащего информацию о 15 попытках ввода пароля, состоящего из 10 символов, составил 20 Кб; размер модели из 50 вводов составил 80 Кб. При проведении экспериментов также было установлено, что размер файла линейно зависит от длины пароля.

Экспериментальные исследования по оценке скорости и объема собираемых данных проводились на рабочей станции WST_WIN7 (программные характеристики - ОС Windows 7 (32bit); аппаратные характеристики: процессор Core2 Duo, частота 2.4 ГГц, 2 ядра;

оперативная память 1Гбайт; дисковый накопитель HDD, 80 Гбайт; сетевой адаптер 100Мбит/сек).

Результаты по скорости сбора и объему собираемых данных приведены в протоколе №3 по пункту № 4.3.1.1 Программы и методики экспериментальных исследований «Подсистемы 1» ЭО ПК для сбора и анализа биометрической информации об особенностях динамики работы пользователя с устройствами ввода-вывода в рамках стандартного человеко-машинного интерфейса (см. Приложение А: таблица 5 Протокола испытаний №5).

Результаты по скорости работы методов предобработки данных приведены в протоколе №4 по пункту № 4.3.1.1 Программы и методики экспериментальных исследований «Подсистемы 1» ЭО ПК для сбора и анализа биометрической информации об особенностях динамики работы пользователя с устройствами ввода-вывода в рамках стандартного человеко-машинного интерфейса (см. Приложение А: таблица 7 Протокола испытаний №7).

1.1.2 Оценка скорости работы методов сбора и предобработки данных о работе пользователя с информационными и вычислительными ресурсами компьютерной системы

Экспериментальные исследования проводились в соответствие с пунктами 4.3.2.1, 4.3.2.2, 4.3.2.3 Программы и методики экспериментальных исследований.

Целью проводимых экспериментальных исследований является оценка скорости работы методов сбора и предобработки поведенческих биометрических данных, объемов собираемых биометрических данных в условиях реальной работы «Подсистемы 2»; оценка точности и скорости работы методов машинного обучения и математической статистики для построения и применения поведенческих моделей «Подсистемы 2»; а также проверка функциональности «Подсистемы 2» ЭО ПК.

Агенты сбора информации выполняются в фоновом режиме, чтобы осуществлять сбор с АРМ пользователей следующей информации:

- время начала и завершения сеанса работы АРМ;
- применяемое пользователем в процессе работы АРМ программное обеспечение (с учетом фильтрации информации о работе системных и служебных процессов ОС);
- факты работы пользователей с внешними запоминающими устройствами, включая время чтения/записи файлов с указанием их имен;

- факты обращений пользователей к файлам и папкам на АРМ других пользователей или серверах;
- характеристики работы АРМ пользователя в ЛВС (время и длительность сетевых соединений, объем передаваемой информации, используемые для передачи приложения и протоколы, адреса удаленных серверов);
- факты самостоятельной установки и удаления пользователями программного обеспечения и/или технических средств на АРМ пользователя.

Для выявления перечисленных фактов осуществляется сбор следующей информации:

- время начала/завершения сеанса работы пользователя с АРМ в ЛВС;
- информацию о запускаемых пользователем программах и приложениях;
- параметры работы пользователя с внешними запоминающими устройствами (запись/чтение информации);
- регистрация времени записи файлов на внешние носители;
- параметры обращения пользователя к файлам и папкам, в том числе на удаленных АРМ;
- характеристики работы пользователя в ЛВС (время и длительность сетевых соединений, объем передаваемой информации, используемые для передачи приложения и протоколы, адреса удаленных серверов и служб);
- интенсивность работы пользователя с клавиатурой и манипулятором "мышь" в каждом из активных положений;
- информацию об установке или удалении пользователем программных и технических средств на основе изменения состава соответствующих записей системного реестра.

При отправке информации на сервер консолидации поддерживаются следующие стратегии ее передачи (задание расписания загрузки агентом собранной информации на сервер консолидации):

- фиксированный объем информации - агент накапливает определенный объем информации или фиксированное количество записей журналов и затем передает их на сервер консолидации;
- через равные промежутки времени - агент через равные промежутки времени передает всю имеющуюся у него в локальном хранилище информацию независимо от ее объема;
- в режиме реального времени - агент немедленно передает информацию о каждой вновь прочитанной записи в журнале регистрации.

Вся информация, передаваемая агентом на сервер консолидации, защищена с использованием технологии SSL.

Экспериментальные исследования по пункту № 4.3.2.1 Программы и методики экспериментальных исследований, проводимые с целью оценки скорости работы методов сбора и предобработки поведенческих биометрических данных, а также объемов собираемых биометрических данных в условиях реальной работы, выполняются по следующему сценарию.

1. Выделяется 5-10 рабочих мест пользователей, на которых устанавливаются агенты сбора для всех указанных выше типов фактов и событий.
2. Осуществляется сбор данных с указанных рабочих мест в течении рабочей недели (не менее 5 рабочих дней).
3. В процессе сбора данных фиксируется объем собираемых данных и время сбора на всех основных этапах:
   - сбор и сохранение событий в локальном хранилище для каждого из типов собираемых фактов (перечисленных выше);
   - время и объем передачи информации на сервер консолидации (в различных режимах) для каждого рабочего места;
   - объем и время сохранения событий на сервере консолидации со всех рабочих мест, на которых осуществляется мониторинг.

   По результатам экспериментов рассчитываются следующие характеристики:

- минимальное, максимальное и среднее число событий каждого типа в сутки, а также минимальный, максимальный и средний объем собираемой информации в локальном хранилище и на сервере консолидации (с учетом заполняющихся словарей);
- минимальное, максимальное и среднее время обработки событий каждого типа с учетом выбранной стратегии передачи информации.

Экспериментальные исследования по оценке скорости и объема собираемых данных, а также методов предобработки данных проводились на рабочей станции WST_WIN7_MON: программные характеристики - ОС Windows 7 (32bit), Python 2.6, библиотека OpenSSL; аппаратные характеристики: процессор Core2 Duo, частота 2.4 ГГц, 2 ядра; оперативная память 1Гбайт; дисковый накопитель HDD, 80 Гбайт; сетевой адаптер 100Мбит/сек.

Результаты по скорости сбора и объему собираемых данных приведены в протоколе №1 по пункту № 4.3.2.1 Программы и методики экспериментальных исследований «Подсистемы 2» ЭО ПК для сбора и анализа данных о работе пользователя с информационными и вычислительными ресурсами компьютерной системы (см. Приложение А: таблица 9 Протокола испытаний №9).

Результаты по скорости работы методов предобработки данных приведены в протоколе №2 по пункту № 4.3.2.1 Программы и методики экспериментальных исследований «Подсистемы 2» ЭО ПК для сбора и анализа данных о работе пользователя с информационными и вычислительными ресурсами компьютерной системы (см. Приложение А: таблица 11 Протокола испытаний №11).

1.1.3 Оценка скорости работы методов сбора и предобработки данных об особенностях работы пользователя с потоками текстовой информации

Данный пункт посвящён проведению экспериментальных исследований по пункту 4.3.3.1 Программы экспериментальных исследований, представленной в отчёте за предыдущий этап настоящих ПНИ (Раздел 4 Программы и методик экспериментальных исследований ЭО ПК).

Целью проводимых экспериментальных исследований является оценка скорости работы методов сбора и предобработки поведенческих биометрических данных, а также объемов собираемых биометрических данных в условиях реальной работы «Подсистемы 3» ЭО ПК, которая служит для сбора и анализа информации об особенностях работы пользователя с потоками текстовой информации.

Согласно пункту 6.5.1 Программы и методик экспериментальных исследований ЭО ПК поведенческие биометрические данные считаются собранными и предобработанными, когда выполнится заполнение структур представления поведенческой информации соответствующими данными. Для оценки скорости работы методов сбора и предобработки поведенческих биометрических данных выполняется следующая последовательность действий:

- на одном из компьютеров, с установленным агентом мониторинга, выполнить процедуру копирования папки, содержащей большое количество (порядка 10000) текстовых документов;
- замерить интервал времени от завершения процедуры копирования папки до заполнения структур представления поведенческой информации всеми текстовыми документами из рассматриваемой папки;
- среднее время работы методов сбора и предобработки будет равно длине рассчитанного временного интервала, поделённой на число текстовых документов папки.

Функции сбора и предобработки поведенческой информации выполняет агент мониторинга «Подсистемы 3» ЭО ПК. С точки зрения пользователя, на чьё рабочее место устанавливается агент мониторинга, критичным является то, как на производительность его компьютера влияет работа агента. При сборе поведенческой информации основная нагрузка на агент мониторинга приходится в том случае, когда требуется сохранять данные о наблюдаемых операциях и выполнять теневое копирование соответствующих документов. Примером описанного случая является выполнение пользователем операции создания документа. Поэтому для оценки скорости работы методов сбора и предобработки поведенческих биометрических данных была проведена серия экспериментов, заключающаяся в копировании (т.е. создании новых документов) большого числа документов на наблюдаемом компьютере пользователя.

В качестве наблюдаемого компьютера для проведения тестов агента мониторинга использовалась *рабочая станция поддержки агентов мониторинга, сбора и предобработки* (см. Приложение А Программы и методик экспериментальных исследований ЭО ПК):

- программное обеспечение: ОС Windows 7 (32bit); Python 2.6; библиотека OpenSSL;
- аппаратные характеристики: процессор Core2 Duo (частота 2.4 ГГц, 2 ядра); оперативная память 1Гбайт; дисковый накопитель HDD, 80 Гбайт; сетевой адаптер 100Мбит/сек.

Для проведения всестороннего исследования нагрузки на вычислительные ресурсы наблюдаемой машины также был подключён один физический жёсткий диск и один внешний жёсткий диск. Двум физическим жёстким дискам тестовой машины соответствуют логические диски *C* и *D*, причём диск *C* является системным, а используемый внешний жёсткий диск соответствует логическому диску *T*.

Тестовый стенд, на котором проводились экспериментальные исследования, состоит из:

- агента мониторинга, установленного на диске *C* наблюдаемого компьютера;
- директории EnronFolder, содержащей все тестовые документы из набора экспериментальных данных НТЭД2 (10228 текстовых документов), сформированного на втором этапе настоящих ПНИ;
- директории ObservedFolder — директория, наблюдаемая агентом мониторинга, т.е. все операции, происходящие с данной директорией, не фильтруются агентом мониторинга;
- директории TestFolder — вспомогательная директория для проведения тестов.

Для оценки производительности агента мониторинга рассматривались следующие тесты:

– копирование всех документов из директории EnronFolder в директорию TestFolder с незапущенным агентом мониторинга;
– копирование всех документов из директории EnronFolder в директорию TestFolder с запущенным агентом мониторинга;
– копирование всех документов из директории EnronFolder в директорию ObservedFolder с запущенным агентом мониторинга.

Первоначально были проведены две серии экспериментов, заключающиеся в выполнении описанных тестов с различными конфигурациями стенда:

– 1-ая конфигурация: директории EnronFolder располагается на диске *D*, а директории ObservedFolder и TestFolder — на диске *C*;
– 2-ая конфигурация: директории EnronFolder располагается на внешнем жёстком диске *T*, а директории ObservedFolder и TestFolder — на диске *D*.

При проведении 3-его теста, с использованием 1-ой конфигурации стенда, получается двойная нагрузка на диск *C*, т.к. выполняется создание новых документов непосредственно в процессе их копирования и при создании теневых копий агентом мониторинга. Поэтому рассматривается также и 2-ая конфигурация стенда, при которой копии документов будут создаваться на диске D, а их теневые копии на диске C.

Для оценки скорости работы методов сбора и предобработки поведенческих биометрических данных замерялось время копирования всех документов набора НТЭД2 для каждого теста. Полученные показатели для двух рассмотренных серий экспериментов с двумя рассмотренными конфигурациями тестового стенда приведены в протоколе №1 по пункту № 4.3.3.1 Программы экспериментальных исследований «Подсистемы 3» ЭО ПК для сбора и анализа информации об особенностях работы пользователя с потоками текстовой информации (см. Приложение А: таблица 13 Протокола испытаний №13).

Создание теневой копий документа требуется от агента мониторинга в случаях операции создания, изменения или любой операции с документом, который ранее не был зарегистрирован агентом. Проведённые эксперименты показали, что при одновременном выполнении операций, требующих создание теневых копий, с большим количеством документов, скорость их выполнения в худшем случае уменьшается в 1.6 раза.

Однако, в повседневной работе пользователя такие массовые операции с документами встречаются редко. Поэтому была проведена 3-ья серия экспериментов с 1-ой конфигурацией стенда, с той лишь разницей что выполнялось последовательное копирование файлов пользователя «vkaminski» из набора НТЭД2 (616 текстовых документов) с задержкой в

1 секунду. Полученные тестовые показатели для рассматриваемой серии экспериментов также приведены в протоколе №1 по пункту № 4.3.3.1 Программы экспериментальных исследований (см. Приложение А: таблица 13 Протокола испытаний №13).

Из полученных данных для 3-ей серии экспериментов следует, что каких-либо задержек (с точностью до секунды) при сборе и предобработки поведенческой информации об особенностях работы пользователя с потоками текстовой информации зафиксировано не было. Это объясняется тем, что теневое копирование выполнятся только после выполнения пользовательской операции и успевает завершиться менее чем за 1 секунду. Таким образом, если пользователь не выполняет частые операции, требующие от агента создания теневых копий, с большим числом документов, то он даже не заметит каких-либо изменений в характеристиках работы наблюдаемого компьютера.

Согласно пункту 6.5.1 Программы и методик экспериментальных исследований ЭО ПК оценка объема собираемых биометрических данных в условиях реальной работы осуществляется путём расчёта среднего объёма собранной поведенческой информации за сутки, полученной от агентов мониторинга. В ходе экспериментальных исследований были произведены замеры среднего объёма текстовых документов, с которыми пользователь работал за сутки, и соответствующий объём поведенческой информации пользователя при работе с данными документами. Поведенческая информация пользователя включает данные об операциях пользователя с электронными документами и текстовое содержимое соответствующих электронных документов. Также отметим, что вся собираемая поведенческая информация сжимается архиватором *gzip*. Результаты описанных замеров приведены в протоколе №1 по пункту № 4.3.3.1 Программы и методик экспериментальных исследований «Подсистемы 3» ЭО ПК для сбора и анализа информации об особенностях работы пользователя с потоками текстовой информации (см. Приложение А: таблица 13 Протокола испытаний №13).

1.2 Оценка точности и скорости работы методов предобработки текстовых данных, в том числе методов рубрикации, группировки, автоматического аннотирования и выявления ключевых слов на экспериментальных и/или эталонных тестовых данных

Данный подраздел посвящён проведению экспериментальных исследований по пунктам 4.3.3.2 и 4.3.3.4 Программы экспериментальных исследований, представленной в отчёте за предыдущий этап настоящих ПНИ (Раздел 4 Программы и методик экспериментальных исследований ЭО ПК), описания которых приводятся ниже в пунктах 1.2.1 и 1.2.2 соответственно.

1.2.1 Оценка работы методов предобработки текстовых данных

Целью проводимых экспериментальных исследований является оценка точности и скорости работы методов предобработки текстовых данных, в том числе методов рубрикации, группировки, автоматического аннотирования и выявления ключевых слов на экспериментальных и/или эталонных тестовых данных.

Согласно пункту 6.5.2 Программы и методик экспериментальных исследований ЭО ПК для каждого из разработанных методов рубрикации, группировки, автоматического аннотирования и выявления ключевых слов процедура оценки точности и скорости работы следующая:

- загрузить экспериментальные данные в систему, используя процедуру сбора поведенческой информации с текстовыми данными;
- с помощью автоматизированного рабочего места (АРМ) аналитика инициировать процесс (рубрикация, группировка, автоматическое аннотирование), реализующий рассматриваемый метод, для выборки, сформированной из загруженных экспериментальных данных, и замерить время его работы;
- сформировать отчёт, содержащий результаты применения рассматриваемого метода к выборки экспериментальных данных;
- выгрузить результирующие данные сформированного отчёта для расчёта соответствующих оценок точности. В качестве оценок качества алгоритмов автоматического аннотирования будут использоваться стандартные метрики *ROUGE* (аббревиатура для *Recall-Oriented Understudy for Gisting Evaluation*). Для оценки качества алгоритмов многотемной классификации (рубрикации «с учителем») будет использоваться

традиционный критерий *Hamming Loss*. Оценка качества алгоритмов кластеризации (рубрикации «без учителя») производится с помощью расчёта следующих базовых критериев, таких как чистота получаемых кластеров (англ. *Purity*), нормализованная взаимная информация (англ. *Normalized Mutual Information, NMI*) и Rand-индекс (англ. *Rand Index, RI*).

Для проведения экспериментальных исследований использовался *вычислительный сервер (Подсистема2/Подсистема3)* (см. Приложение А Программы и методик экспериментальных исследований ЭО ПК):

- программное обеспечение: ОС Microsoft Windows 2008R2 (64bit); Microsoft Office 2007; Microsoft SQL Server 2012; Microsoft SQL Server 2005; Microsoft SQL Server 2005 Analysis Services; Microsoft SQL Server 2005 Management Studio; интерпретатор Python 2.7.3; модуль Python pywin 219; Библиотека Python Win32 Extensions; Библиотека Natural Language Toolkit (NLTK);
- аппаратные характеристики: два процессора x64 с тактовой частотой 2 Ггц, 4 ядра; оперативная память объемом 48 Гбайт; два дисковый накопителя HDD, объемом 2Тбайт; два сетевых адаптера с пропускной способностью 1 Гбит.

1.2.1.1 Оценка метода автоматического аннотирования

Самым распространенным средством для оценки качества алгоритмов аннотирования является полностью автоматизированная утилита ROUGE (аббревиатура для Recall-Oriented Understudy for Gisting Evaluation). На вход подается множество построенных алгоритмом аннотаций и на каждую такую аннотацию одна (или более) модельная аннотация. Далее с помощью специальных метрик аннотации сравниваются, и алгоритму присваивается некоторая оценка (средняя по всем аннотациям). Впоследствии для оценки качества различных алгоритмов автоматического аннотирования утилита ROUGE стала использоваться на крупнейшей конференции, посвящённой задачам автоматического аннотирования Document Understandng Conference (DUC).

Эталонные наборы данных включают в себя текстовые документы и модельные аннотации к ним, которые обычно строятся человеком. Причем для одного документа может быть построено несколько модельных аннотаций, т.к., вообще говоря, одна аннотация может быть хорошей для одного человека и, в тоже время, быть плохой для другого. Для оценки методов однодокументного аннотирования, в которых аннотация строится к каждому

документу из коллекции, принято использовать набор данных DUC 2002 (533 документа, 1112 модельных аннотаций), который состоит из новостных статей.

В работе Lin C. Y. (Looking for a few good metrics: Automatic summarization evaluation— how many samples are enough // Proceedings of the NTCIR Workshop. – 2004. – Т. 4.) проводится анализ применимости различных метрик для различных задач аннотирования, в частности, для оценки алгоритмов однодокументного аннотирования на наборе DUC 2002 рекомендуется использовать метрики ROUGE-2, ROUGE-L, ROUGE-S, ROUGE-W. Также одним из результатов данной работы является вывод, что наборы DUC 2002 обладают достаточным количеством модельных аннотаций для объективной оценки качества алгоритмов аннотирования.

В «Подсистеме 3» автоматическое аннотирование было реализовано с помощью разработанного метода (представленного в отчёте за первый этап настоящих ПНИ), заключающегося в оценке значимости (релевантности) отдельных фрагментов (предложений) текста и последующем составлении аннотации документа из наиболее значимых фрагментов. Вычисление релевантности фрагментов текста основано на выделение основных тематик документа, расчёте их значимости (глобального веса тематики) и оценке их представленности (весе) в каждом фрагменте документа. Для выделения основных тематик текста использовалась неотрицательная матричная факторизация (NMF). Далее релевантность фрагмента текста рассчитывается как норма вектора, являющегося результатом поэлементного умножение вектора глобальных весов тематик и вектора весов тематик в рассматриваемом фрагменте.

Для сравнения реализации разработанного метода автоматического аннотирования на наборе DUC 2002 результирующие аннотации должны состоять из 100 слов, поэтому при формированияи аннотации алгоритм последовательно выбирал предложения текста для аннотации в порядке убывания их релевантности до тех пор, пока суммарное количество слов в выбранных предложениях не превысит 100. Затем выбранные предложения упорядочивались в порядке появления их в тексте, и получаемая таким образом аннотация подавалась на вход утилите *ROUGE*[1].

Полученные показатели метрик ROUGE и скорости работы метода автоматического аннотирования приведены в протоколе №1 по пункту № 4.3.3.2 Программы экспериментальных исследований «Подсистемы 3» ЭО ПК для сбора и анализа информации

[1] Параметры *ROUGE*: ROUGE-1.5.5.pl -e /duc2002 -m **-l 100** -2 4 -n 2 -w 1.2 -s -a duc2002.xml

об особенностях работы пользователя с потоками текстовой информации (см. Приложение А: таблица 14 Протокола испытаний №14).

1.2.1.2 Оценка метода рубрикации

Задача классификации многотемных (англ. multi-label) документов заключается в определении принадлежности документа к одному или нескольким классам (из предопределённого набора классов) на основании анализа совокупности признаков, характеризующих данный документ. В задаче классификации многотемных документов в обучающей выборке $Z = \{x_i, y_i\}_{i=1}^{m}$ для каждого примера $x_i \in X$ задан не единственный класс, а множество релевантных классов $y_i \subseteq \{1,...,q\}$, и целью алгоритма машинного обучения является построение на основе обучающей выборки классификатора $f_Z : X \to 2^q$, предсказывающего для заданного примера все релевантные классы (X — исходное пространство признаков, q — число классов).

Рубрикация текстовых данных была реализована с помощью разработанного на первом этапе настоящих ПНИ метода классификации многотемных документов, который использует подход «попарных сравнений» (бинарная декомпозиция типа «каждый-против-каждого»). В этом методе каждая пара возможно пересекающихся классов разделяется двумя бинарными классификаторами, которые отделяют перекрывающиеся и неперекрывающиеся области. Затем отдельные вероятности, предсказанные бинарными классификаторами, объединяются вместе для оценки результирующих вероятностей принадлежности документа классам на основе расширенной модели Бредли-Терри «с ничьёй».

Метрики оценки качества методов многотемной (multi-label) классификации существенно отличаются от метрик оценки для обычной классификации. Для оценки качества многотемной классификации традиционно используется критерий Hamming Loss, который измеряет симметрическое различие между предсказанным и истинным наборами классов (рубрик) документов:

$$HammingLoss = \frac{1}{k}\sum_{i=1}^{k}\frac{1}{|Y|}\left|(f_Z(x_i))\nabla(y_i)\right|,$$

где $a\nabla b = (a \cup b) \setminus (a \cap b)$, $a \subseteq Y$, $b \subseteq Y$, k – размер тестового набора. Другими словами, критерий $HammingLoss$ оценивает, сколько раз пара «пример-метка» классифицируется неправильно, т.е. метка, принадлежащая примеру, не предсказывается или метка, не принадлежащая примеру, предсказывается (см. рисунок 1). Чем меньше значение

$HammingLoss$, тем лучше качество многотемного классификатора. Качество классификации совершенно, когда $HammingLoss = 0$. Заметим, что когда $|y_i| = 1$ для всех примеров, то задача многотемной классификации вырождается в традиционную классификацию и значение $HammingLoss$ равно обычной ошибке классификации, домноженной на $2/|Y|$.

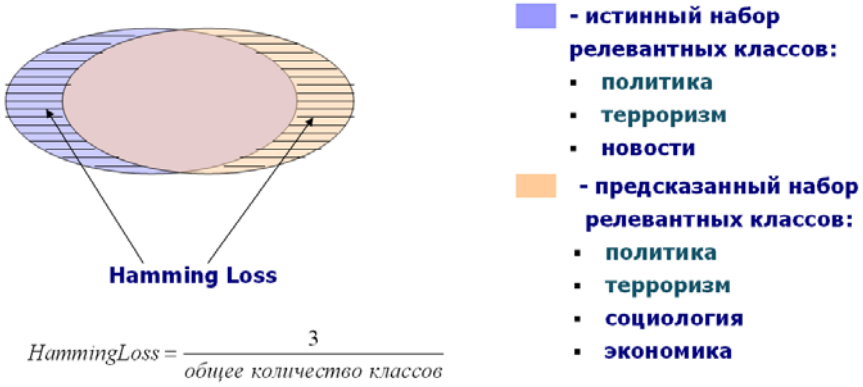

Рисунок 1 — Иллюстрация критерия Hamming Loss для оценки качества многотемной классификации.

Для оценки точности и скорости работы реализации многотемной классификации использовался популярный эталонный набор данных Reuters-21578. Документы данного набора разделены на тренировочные и тестовые (7068 тренировочных документов, 2745 тестовых документов), каждый документ соотнесен с несколькими тематиками, всего в наборе 120 различных тематик.

Полученные показатели критерия Hamming Loss и скорости работы метода многотемной классификации (рубрикации) приведены в протоколе №1 по пункту № 4.3.3.2 Программы экспериментальных исследований «Подсистемы 3» ЭО ПК для сбора и анализа информации об особенностях работы пользователя с потоками текстовой информации (см. Приложение А: таблица 14 Протокола испытаний №14).

1.2.1.3 Оценка метода группировки

Кластеризация (группировка или рубрикации «без учителя») коллекции документа была реализована с помощью разработанного на первом этапе настоящих ПНИ метода, основанного на использовании неотрицательной матричной факторизации (NMF) для выделения основных тематик коллекции документов и представления текстов документов в пространстве тематик. Использование неотрицательной матричной факторизации позволяет установить, какие термы наибольшим образом характеризуют каждую из выделенных тематик

(т.е. решается задача выявления ключевых слов), и определить принадлежность документа к тому или иному кластеру, который соответствует одной из выделенных тематик.

Представлением документа в пространстве тематик является вектор с неотрицательными элементами, показывающими веса каждой тематики в данном документе. Реализованный алгоритм кластеризации использует информацию о весах тематик в документе для определения кластера документа — номер кластера для документа с номером *j* определяется как $n = \arg\max_i H_{i,j}$, где H_{ij} — вес *i*-ой тематики в тексте *j*-ого документа. Помимо кластеризации на основе неотрицательной матричной факторизации были также реализованы автоматическое аннотирование, выявление ключевых тематик и ключевых слов.

Эксперименты по оценке качества кластеризации (группировки) были проведены на эталонном тестовом наборе 20 Newsgroups. Используемый набор содержит 18828 документов, документы набора предварительно кластеризованы (разделены) на 20 тематических групп. Результаты вычисления основных критериев оценки качества алгоритмов кластеризации (Manning C. D. et al. Introduction to information retrieval. – Cambridge: Cambridge university press, 2008. – Т. 1. – С. 496.), таких как чистота получаемых кластеров (англ. Purity), нормализованная взаимная информация (англ. Normalized Mutual Information, NMI) и Rand-индекс (англ. Rand Index, RI), приведены в протоколе №1 по пункту № 4.3.3.2 Программы экспериментальных исследований «Подсистемы 3» ЭО ПК для сбора и анализа информации об особенностях работы пользователя с потоками текстовой информации (см. Приложение А: таблица 14 Протокола испытаний №14).

Оценка реализации функциональности «Подсистемы 3» ЭО ПК, которая служит для сбора и анализа информации об особенностях работы пользователя с потоками текстовой информации. «Подсистема 3» ЭО ПК считается выдержавшим проверку, если полученные результаты соответствуют ожидаемым результатам контрольных примеров №№ 3-5 Приложение Б Программы и методик экспериментальных исследований ЭО ПК и/или отражены в комплекте документации.

Проверка построения и применения модели рубрикации «с учителем» (многотемная классификация). Контрольный пример №3.

1. Запустить АРМ аналитика «Подсистемы 3». В результате откроется АРМ аналитика в виде ММС консоли, которая показана на рисунке 2.

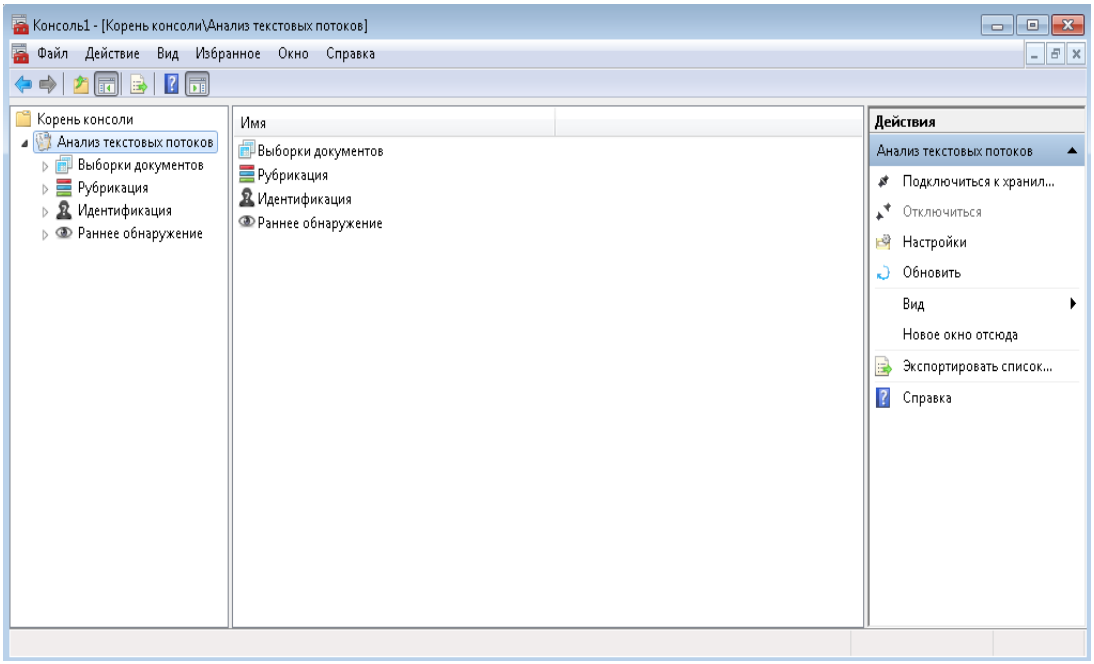

Рисунок 2 — Главное окно MMC консоли АРМ.

2. Выполнить подключение к единому хранилищу модуля консолидации. Главное окно консоли состоит из 3 частей (см. рисунок 2). Справа отображается меню для выбранного корневого узла («Анализ текстовых потоков»). В данном меню есть раздел «Подключиться к хранилищу», который позволяет подключиться к базе данных единого хранилища, предварительно открыв окно для ввода параметров подключения (см. рисунок 3).

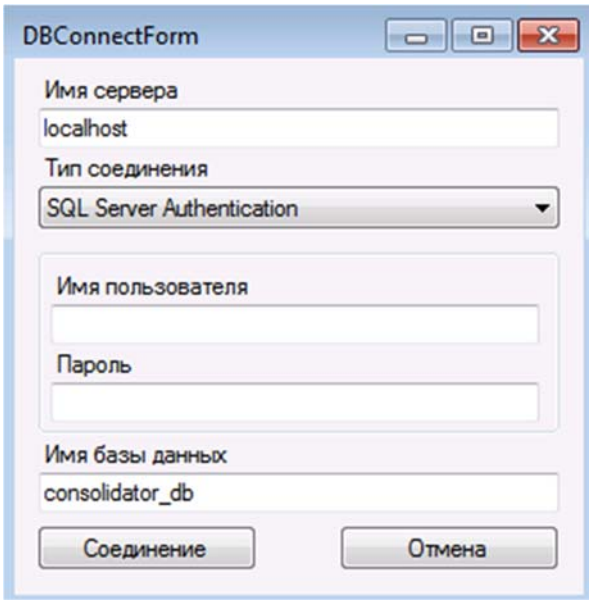

Рисунок 3 — Окно ввода параметров для подключения к хранилищу.

3. Задать список тем для рубрикации. Нажав на раздел «Настройки» в правом меню корневого узла «Анализ текстовых потоков» (см. рисунок 2), откроется окно настроек, в котором присутствует раздел «Категории» (см. рисунок 4). Этот раздел меню задает настройки построения моделей для определения тематик документов. В этом разделе можно просматривать список существующих тем, добавлять новую тематику (вводя её название в соответствующую область «Введите название темы…» и нажав кнопку «Добавить»), а также удалять тематики (для этого нужно выбрать нужную тематику в соответствующем списке и нажать кнопку «Удалить тему»). Кроме того, кнопка «Сброс» позволяет восстановить список тематик, заданный по умолчанию.

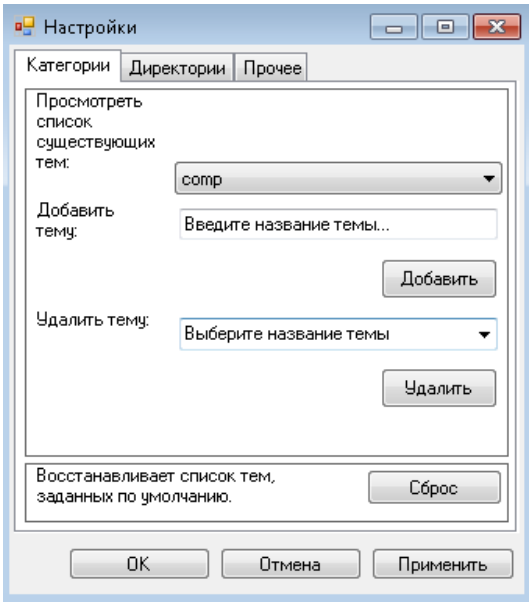

Рисунок 4 — Окно настройки категорий (тем) рубрикации.

4. Создать выборку документов для последующего построения модели. При активации меню «Создать выборку» узла «Выборки документов», находящегося в левой части меню главного окна консоли АРМ (см. рисунок 2), откроется форма создания выборок, представленная на рисунке 5.

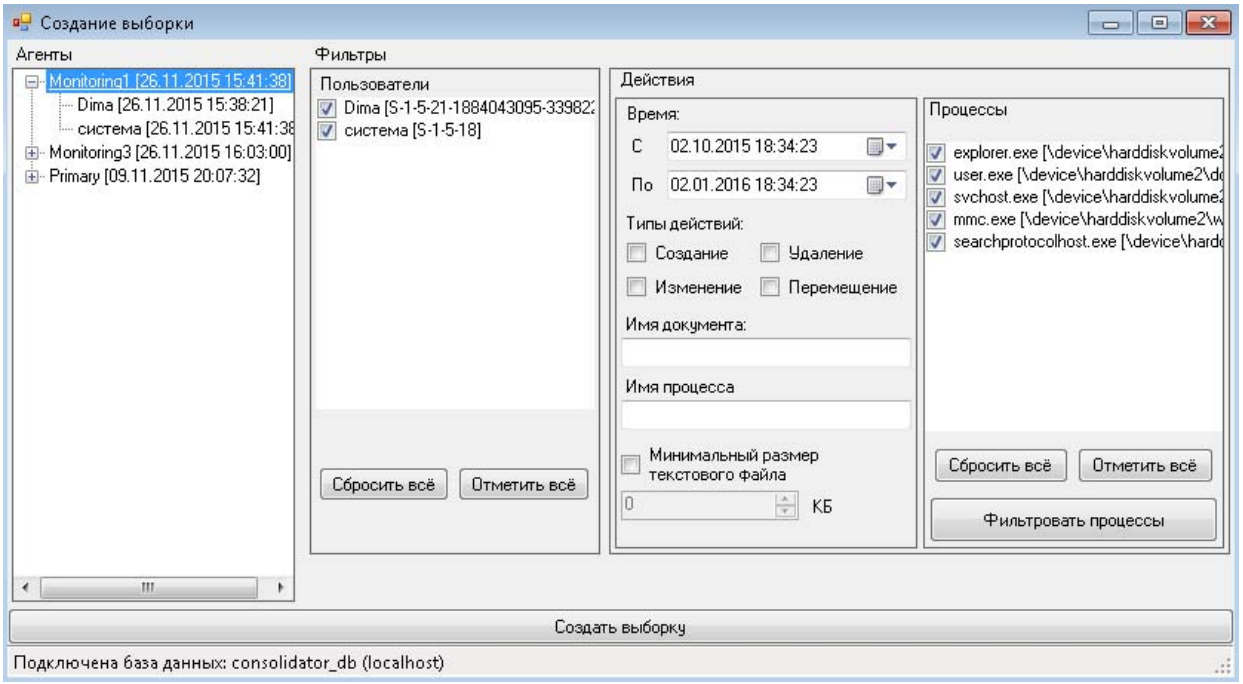

Рисунок 5 — Окно формы создания выборок поведенческой информации.

5. Просмотреть выборку. При активации меню «Просмотреть» для выборки происходит вывод событий данной выборки в отдельной форме в виде таблицы (см. рисунок 6).

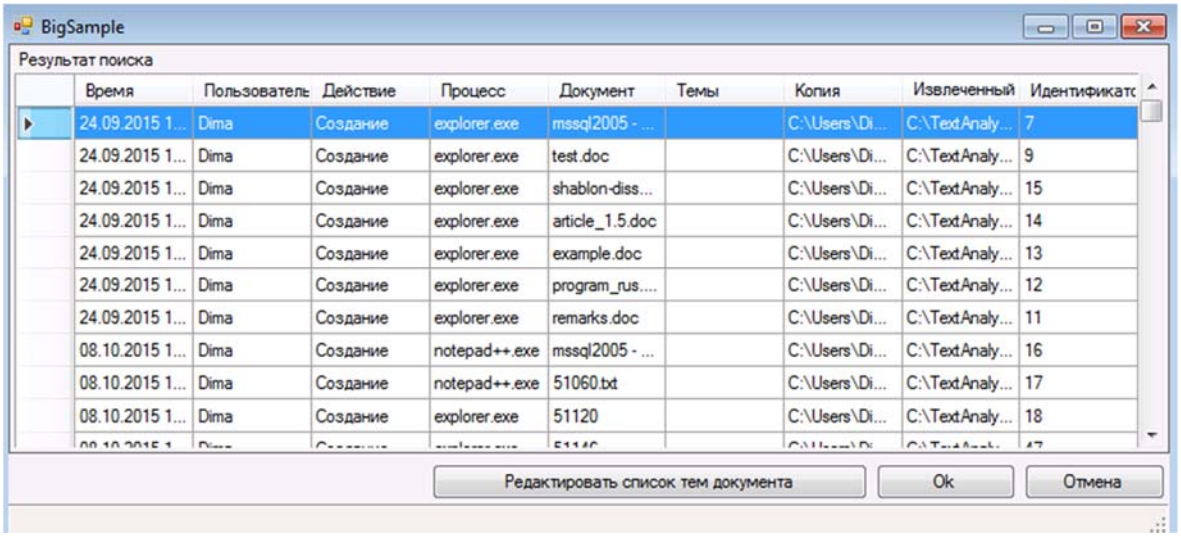

Рисунок 6 — Окно формы просмотра выборки поведенческой информации.

6. Построить модель рубрикации «с учителем». Открытие формы интерфейса построения моделей рубрикации (см. рисунок 7) происходит при выборе пункта меню «Создать модель» в разделе «Модели рубрикации», который входит в узел «Рубрикация», находящийся в левой части меню главного окна консоли АРМ (см. рисунок 2). В разделе «Тип рубрикации» пользователь должен выбрать тип строящейся модели, т.е.

«многотемная» (тип «группировка (кластеризация)» используется при рубрикации «без учителя»).

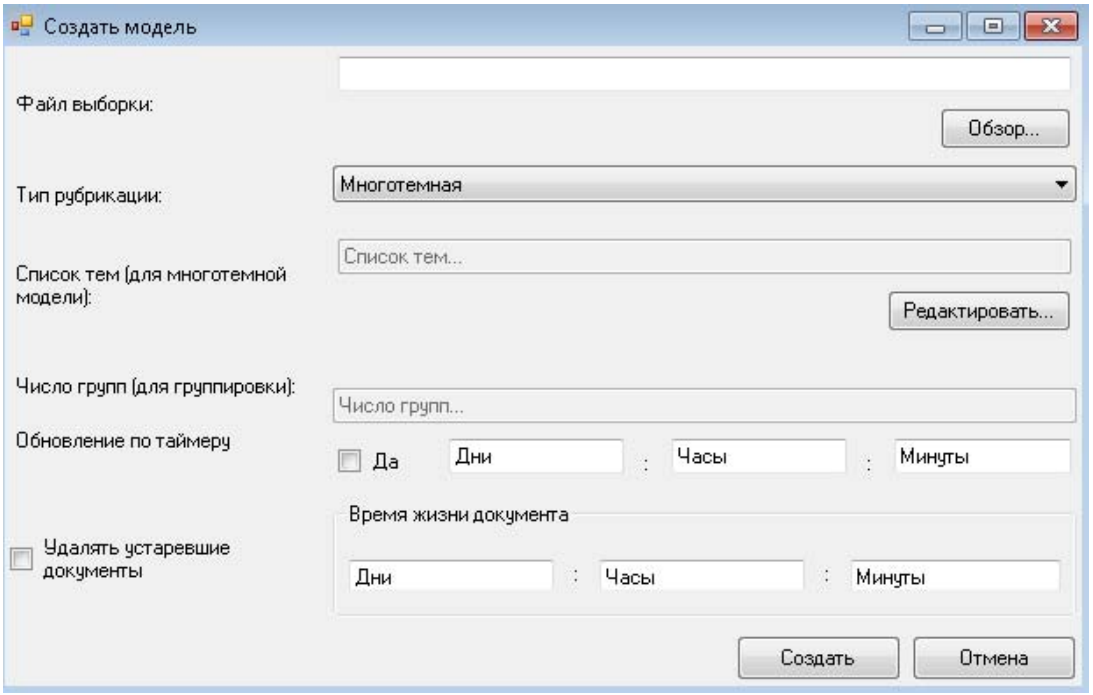

Рисунок 7 — Окно формы построения модели рубрикации «с учителем».

7. Построить отчет по созданной модели рубрикации «с учителем». Форма интерфейса построения отчетов по применению модели рубрикации «с учителем» (см. рисунок 8) открывается при активации пункта меню «Создать отчет» узла «Отчеты рубрикации», а также при выборе пункта меню «Создать отчет» для модели, по которой нужно сформировать отчет.

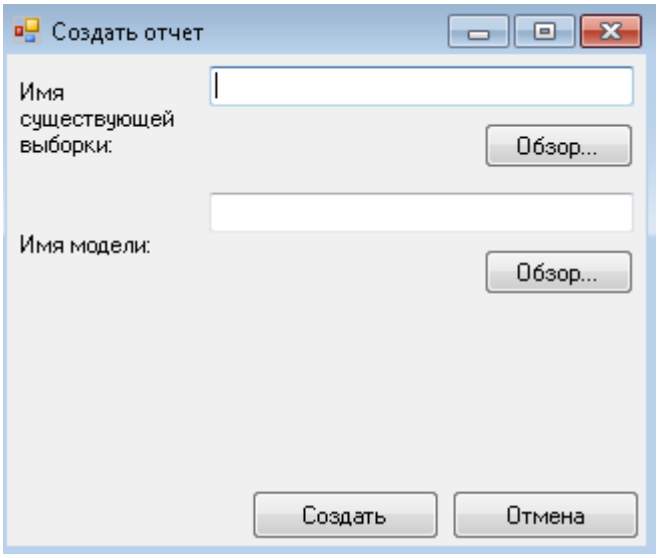

Рисунок 8 — Окно формы построения отчетов по применению модели рубрикации «с учителем».

8. Просмотреть отчёт рубрикации «с учителем». Форма отчёта рубрикации «с учителем» представляет таблицу событий поведенческой информации, которая открывается с помощью меню «Просмотреть» для выбранного отчёта. В случае отчета по рубрикации «с учителем», строки содержат список тем (колонка «Темы» в таблице), присвоенных соответствующим документам (см. рисунок 9).

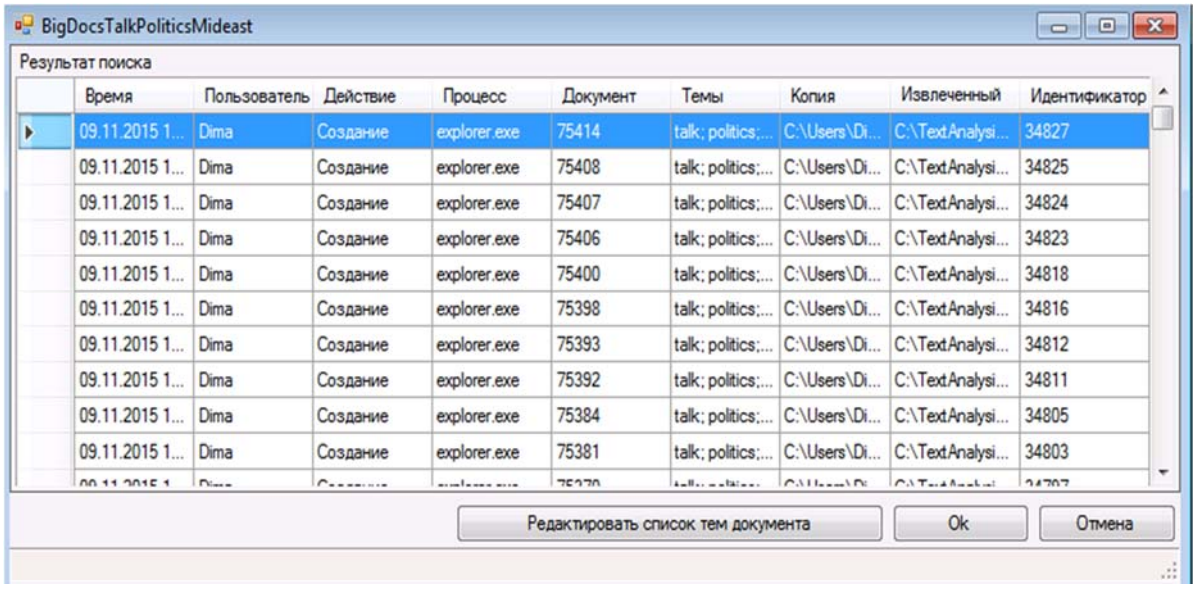

Рисунок 9 — Окно формы просмотра отчёта рубрикации «с учителем».

Проверка построения модели рубрикации «без учителя» (кластеризация). Контрольный пример №4.

1. Запустить АРМ аналитика «Подсистемы 3». В результате откроется АРМ аналитика в виде MMC консоли, которая показана на рисунке 2.
2. Выполнить подключение к единому хранилищу модуля консолидации. Главное окно консоли состоит из 3 частей (см. рисунок 2). Справа отображается меню для выбранного корневого узла («Анализ текстовых потоков»). В данном меню есть раздел «Подключиться к хранилищу», который позволяет подключиться к базе данных единого хранилища, предварительно открыв окно для ввода параметров подключения (см. рисунок 3).
3. Создать выборку документов для последующего построения модели. При активации меню «Создать выборку» узла «Выборки документов», находящегося в левой части меню главного окна консоли АРМ (см. рисунок 2), откроется форма создания выборок, представленная на рисунке 5.
4. Просмотреть выборку. При активации меню «Просмотреть» для выборки происходит вывод событий данной выборки в отдельной форме в виде таблицы (см. рисунок 6).
5. Построить модель рубрикации «без учителя». Открытие формы интерфейса построения моделей рубрикации (см. рисунок 10) происходит при выборе пункта меню «Создать модель» в разделе «Модели рубрикации», который входит в узел «Рубрикация», находящийся в левой части меню главного окна консоли АРМ (см. рисунок 2). В разделе «Тип рубрикации» пользователь должен выбрать тип строящейся модели, т.е. «группировка (кластеризация)» (тип «многотемная» используется при рубрикации «с учителем»).

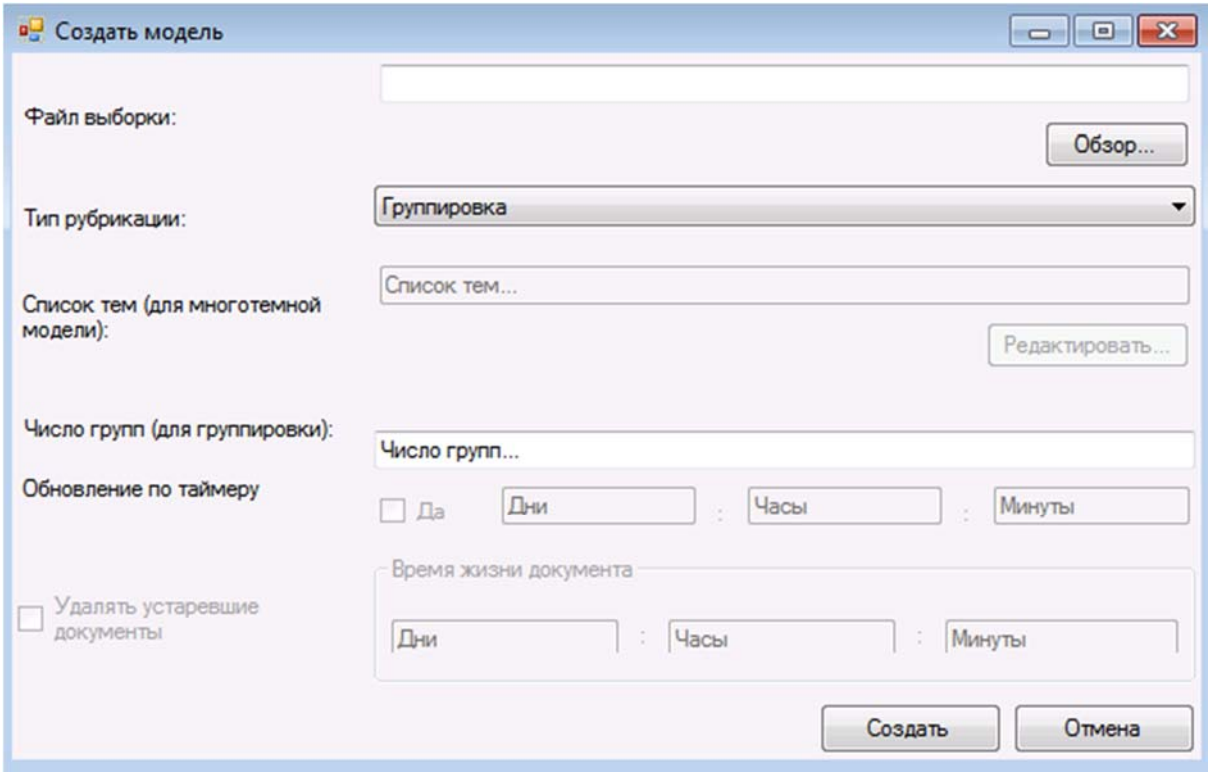

Рисунок 10 — Окно формы построения модели рубрикации «без учителя».

6. Просмотреть модель рубрикации «без учителя». Форма модели рубрикации «без учителя» представляет таблицу событий поведенческой информации, которая открывается с помощью меню «Просмотреть» для выбранной модели. Для модели рубрикации в случае группировки появляется столбец, показывающий номер кластера, наиболее характерного для данного документа. В зависимости от номера кластера ячейка, содержащая его номер, окрашивается в соответствующий цвет (см. рисунок 11).

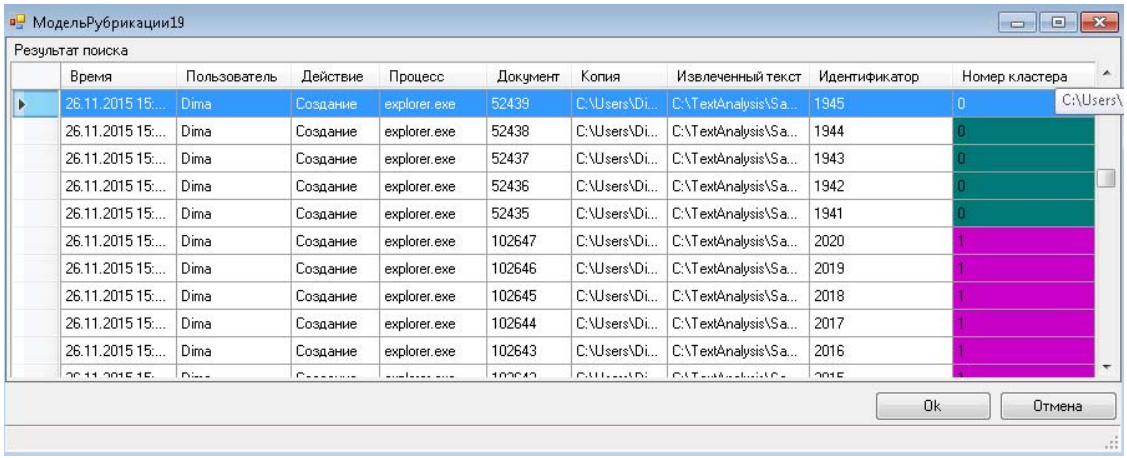

Рисунок 11 — Окно формы просмотра модели рубрикации «без учителя».

Проверка построения аннотаций. Контрольный пример №5.

1. Запустить АРМ аналитика «Подсистемы 3». В результате откроется АРМ аналитика в виде MMC консоли, которая показана на рисунке 2.
2. Выполнить подключение к единому хранилищу модуля консолидации. Главное окно консоли состоит из 3 частей (см. рисунок 2). Справа отображается меню для выбранного корневого узла («Анализ текстовых потоков»). В данном меню есть раздел «Подключиться к хранилищу», который позволяет подключиться к базе данных единого хранилища, предварительно открыв окно для ввода параметров подключения (см. рисунок 3).
3. Создать выборку документов для последующего построения модели. При активации меню «Создать выборку» узла «Выборки документов», находящегося в левой части меню главного окна консоли АРМ (см. рисунок 2), откроется форма создания выборок, представленная на рисунке 5.
4. Просмотреть аннотацию к теневой копии документа из сформированной выборки. При активации меню «Просмотреть» для выборки происходит вывод событий данной выборки в отдельной форме в виде таблицы (см. рисунок 6). События в таблице представляют собой последовательность операций с документами, отсортированных по времени. Каждая строка таблицы соответствует одному документы, а столбцы соответствуют параметрам операции с документом. При двойном клике на строчке таблицы возникает окно с просмотром истории действий с соответствующим документом (см. рисунок 12). В данном окне для каждого действия создания теневой копии при двойном клике открывается окно с выделенным текстом копии и аннотацией к данному тексту (см. рисунок 13).

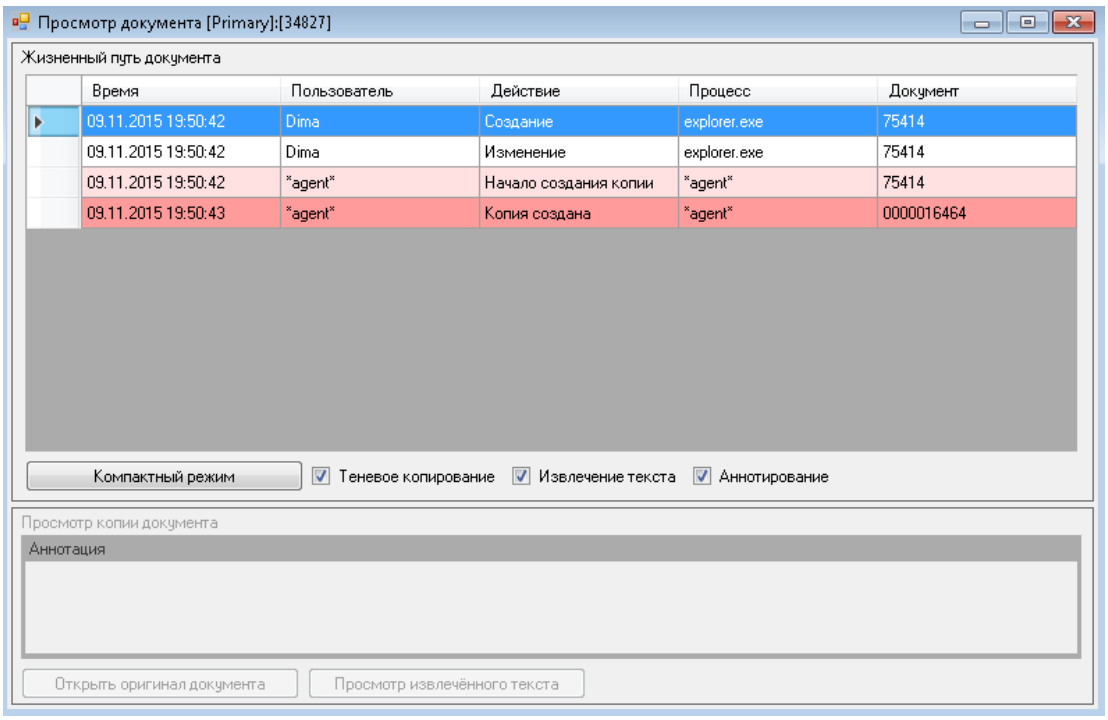

Рисунок 12 — Окно формы просмотра истории действий с документом.

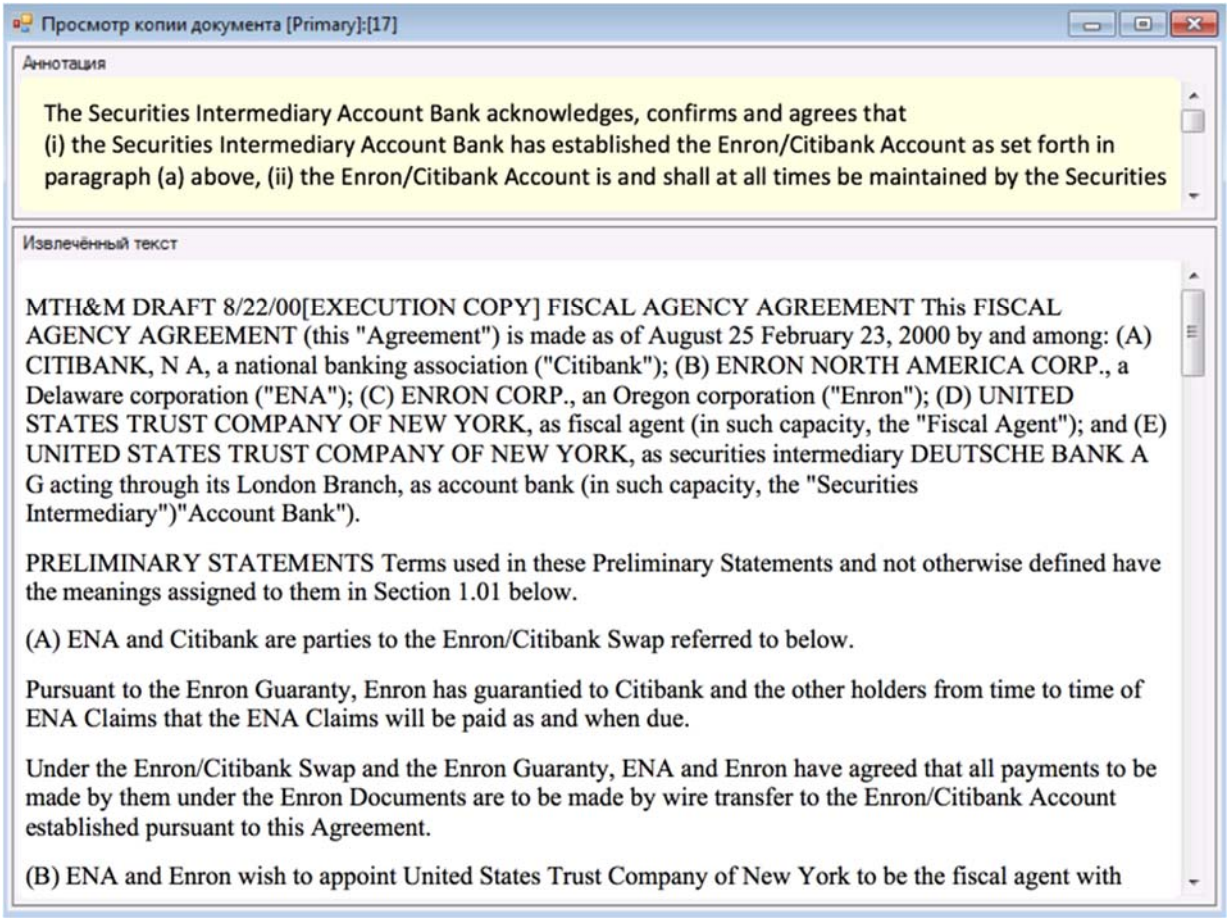

Рисунок 13 — Окно формы просмотра аннотации к теневой копии документа.

1.3 Оценка точности и скорости работы методов машинного обучения и математической статистики для построения и применения поведенческих моделей

1.3.1 Оценка точности и скорости работы методов машинного обучения и математической статистики для построения и применения поведенческих моделей для анализа биометрической информации об особенностях динамики работы пользователя с устройствами ввода-вывода в рамках стандартного человеко-машинного интерфейса

1.3.1.1 Оценка точности и скорости работы методов машинного обучения и математической статистики для построения и применения поведенческих моделей для анализа биометрической информации для решения задачи фоновой

идентификации пользователей на основе событий ввода данных с помощью клавиатуры и мыши

Общая схема постановки экспериментов

Настоящие экспериментальные исследования проводились в соответствие с пунктами 4.3.1.2, 4.3.1.3 Программы и методики экспериментальных исследований. Целью проводимых экспериментальных исследований является сравнительный анализ точности и скорости фоновой идентификации пользователей на основе динамики их работы с клавиатурой с применением различных методов машинного обучения (метод опорных векторов (SVM), нечеткий метод поиска исключений на основе потенциальных функций (Fuzzy), метод k ближайших соседей (kNN), метод репликторных нейронных сетей (RNN)) для построения и применения поведенческих моделей, используемых в «Подсистеме 1» ЭО ПК, выбор наилучших параметров метода, а также выбор методов и параметров предобработки и постобработки данных при построении модели представления.

При идентификации пользователя критериями качества работы методов являются количество действий пользователя, которые система ошибочно распознала как чужие, и количество действий другого пользователя, которые системы распознала как свои. В качестве количественной характеристики качества работы классификатора будет использоваться показатель AUC (площадь под ROC-кривой) — площадь, ограниченная ROC-кривой и осью доли ложных положительных классификаций. Чем выше показатель AUC, тем качественнее классификатор. При применении моделей также фиксируется время, затрачиваемое на распознавание для каждого временного окна, при этом оценивается максимальное, минимальное и среднее время применения моделей.

Оценка точности и скорости работы методов машинного обучения и математической статистики для построения и применения поведенческих моделей динамики работы пользователей с клавиатурой

Для решения фоновой идентификации на основе динамики их работы с клавиатурой предлагается использовать следующий сценарий распознавания пользователей:

- выбирается контрольная группа пользователей;
- каждый из пользователей работает в обычном режиме на своем рабочем месте, на котором установлена система сбора информации в течение нескольких рабочих дней непрерывно;

– собранные данные используются для построения моделей идентификации пользователей по особенностям работы с клавиатурой. Фиксируется время, затрачиваемое на построение моделей: вычисляются максимальное, минимальное и среднее время построения моделей;
– далее каждый из пользователей работает в обычном режиме на своем рабочем месте, на котором установлена система сбора информации в течение нескольких рабочих дней непрерывно и собираемые данные используются для проверки полученных моделей идентификации;
– применение построенных моделей к собранным данным и оценка качества классификации. Модели классификации применяются как к собственным новым данным пользователя, для которого была построена модель так и перекрестно по принципу «каждый против каждого».

Перебирались следующие параметры выделения окон на основе порогов на минимальное и максимальное количество событий в окне: верхняя граница окна (400, 500, 600); нижняя граница окна (100, 200, 300); пауза прерывания окна (30 сек, 40 сек, 50 сек); процент перекрытия между окнами (0, 0.25, 0.5).

Перебирались следующие параметры при построении модели представления при формировании векторов признаков: количество наиболее популярных одиночных нажатий (50, 75, 100); количество наиболее популярных диграфов (50, 75, 100); использование статистики клавиш по группам (да, нет); использование статистики сочетаний из двух клавиш по группам (да, нет).

Перебирались следующие параметры и методы постобработки признаков при построении модели представления: количество квантилей при дискретизации признаков (3, 5, 7, 10); расчет межквартильных расстояний (да, нет); дискретизация на основе гистограмм (да, нет); применение метода главных компонент (да, нет); количество главных компонент (20, 50, 70); выбор наиболее стабильных признаков (да, нет); выбор признаков по уровню значимости (да, нет), количество отбираемых признаков (250, 200, 150, 100); уровень значимости отбираемых признаков (0.3, 0.5).

Перебирались следующие параметры метода опорных векторов при обучении модели классификации: параметр соотношения между долей выбросов и долей опорных векторов (0.1, 0.3, 0.5, 0.7), функция ядра (“rbf”, “poly”, “linear”), ширина ядра (0.01, 0.001, 0.002, 0.005, 0.008).

Перебирались следующие параметры нечеткого метода поиска исключений на основе потенциальных функций при обучении модели классификации: функция ядра (“rbf”, “poly”, “linear”), скорость убывания степени принадлежности в зависимости от расстояния до центра

кластера (1.5, 1.7, 1.9), доля ожидаемых исключений (0.1, 0.3, 0.5), ширина ядра (0.005, 0.01, 0.1).

Перебирались следующие параметры метода ближайших соседей при обучении модели классификации: количество ближайших соседей (3, 5, 10).

Перебирались следующие параметры метода репликаторных нейронных сетей при обучении модели классификации: конфигурация сети (16-4-16, 256-16-256, 32-8-4-8-32).

Экспериментальные исследования проводились на вычислительном сервере SRV_SYS1: программные характеристики (ОС Microsoft Windows 2012R2, 64bit; интерпретатор Python3; модуль Python pandas 0.16.2; модуль Python numpy 1.9.3; модуль user-agent 1.0.1), аппаратные характеристики (два процессора x64 с тактовой частотой 2.4 Ггц, 4 ядра; оперативная память объемом 64 Гбайт; два дисковый накопителя HDD, объемом 2Тбайт в режиме RAID1; два дисковый накопителя HDD, объемом 1Тбайт; два сетевых адаптера с пропускной способностью 1 Гбит).

Были получены следующие результаты экспериментов. Лучшие результаты по точности классификации показал нечеткий метод поиска исключений на основе потенциальных функций. Оптимальными оказались следующие параметры: функция ядра – “rbf”, скорость убывания степени принадлежности в зависимости от расстояния до центра кластера -1.5; доля ожидаемых исключений - 0.1; количество квантилей – 10; выбор наиболее стабильных стабильных признаков - да, количество отбираемых признаков – 200; верхняя граница окна – 500 событий; нижняя граница окна – 300 событий; пауза прерывания окна – 40 секунд; процент перекрытия между окнами – 0; количество наиболее популярных одиночных нажатий – 50; количество наиболее популярных диграфов – 100; использование статистики клавиш по группам - да; использование статистики сочетаний из двух клавиш по группам – нет.

Результаты по скорости построения и применения моделей классификации приведены в протоколе №1 по пункту № 4.3.1.2 Программы и методики экспериментальных исследований «Подсистемы 1» ЭО ПК для сбора и анализа биометрической информации об особенностях динамики работы пользователя с устройствами ввода-вывода в рамках стандартного человеко-машинного интерфейса (см. Приложение А: таблица 2 Протокола испытаний №2).

Результаты по точности классификации приведены в протоколе №3 по пункту № 4.3.1.2 Программы и методики экспериментальных исследований «Подсистемы 1» ЭО ПК для сбора и анализа биометрической информации об особенностях динамики работы пользователя с

устройствами ввода-вывода в рамках стандартного человеко-машинного интерфейса (см. Приложение А: таблица 6 Протокола испытаний №6).

Проверка по пункту № 4.3.1.3 Программы и методики экспериментальных исследований осуществлялась на основе оценки соответствия функциональности компонента фоновой идентификации аутентификации с целью сопоставления полученных результатов и ожидаемых результатов контрольных примеров №№21-22 Приложение Б Программы и методики экспериментальных исследований и/или проверки того, что они отражены в комплекте программной документации.

Проверка режима сбора данных об активности пользователя при работе с устройствами ввода состояла из следующих шагов:

– осуществление запуска приложения компонента сбора для начала контрольного сеанса сбора данных произведено без каких-либо сообщений об ошибках;
– выполнение сеанса обычной работы с использованием устройств ввода длительностью в 15 минут прошло без сообщений об ошибках и визуально не отличалось от работы при неактивной компоненте сбора;
– нажатие CTRL+SHIFT+S с использованием служебных клавиш в левой половине клавиатуры привело к открытию окна управления сборщиком данных (см. рисунок 14);

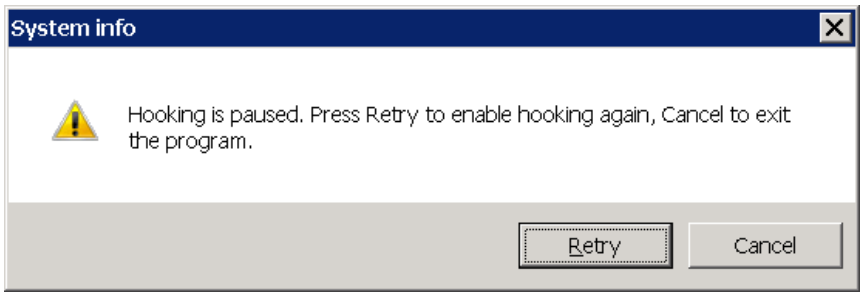

Рисунок 14 — Окно управления сборщиком данных.

– завершение контрольного сеанса сбора данных нажатием на кнопку «Cancel». В файловой системе сформирована структура каталогов с текстовыми файлами, содержащими логи программы (program.log и _internal.log), описание целевой платформы (systeminfo.log) и собранные в течении сессии данные об использовании клавиатуры и мыши (keybard.csv и mouse.csv).

Проверка режима фоновой идентификации пользователя при работе с устройствами ввода состояла из следующих шагов.

1. Активирование рабочего место аналитика (модуль идентификации) путем запуска соответствующего Python модуля на сервере прошло без каких-либо ошибок.

2. Запуск приложения компонента сбора в режиме идентификации прошел без каких-либо ошибок.
3. Выполнение сеанса обычной работы с использованием устройств ввода продолжительностью, достаточной для формирования временного окна. В результате в дополнение к сформированной в результате сбора данных структуре каталогов и файлов в целевом каталоге создаются файлы features_keyboard.csv, structure.csv (см. рисунок 15).

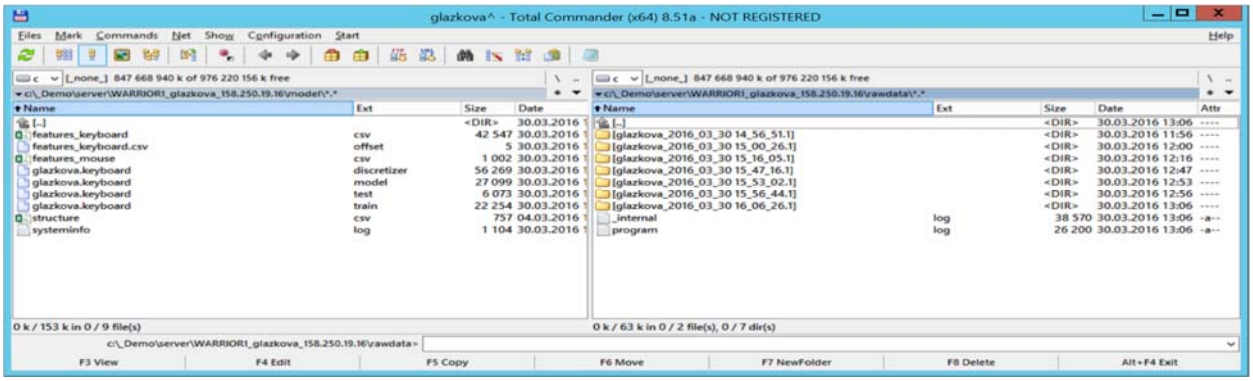

Рисунок 15 — Структура файлов в целевом каталоге модуля идентификации динамики работы с клавиатурой.

4. В логах сервера (модуля идентификации) появляется запись о результате классификации очередного вектора признаков, полученных в результате обработки сформированного временного окна пользователя (см. рисунок 16).

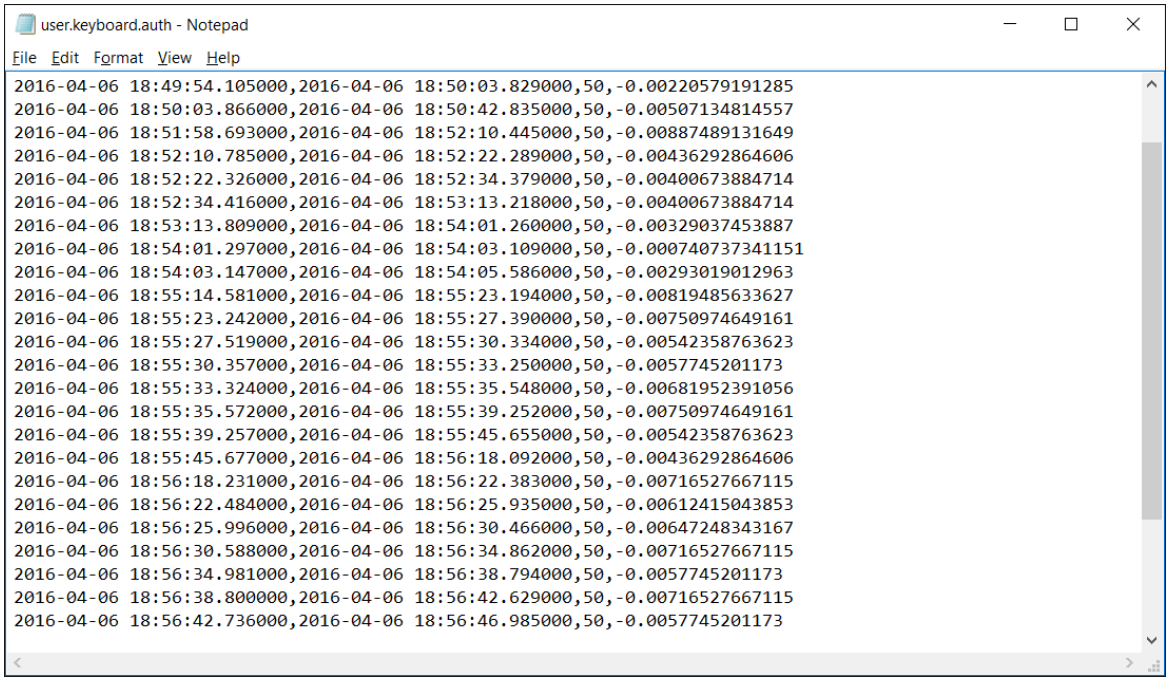

Рисунок 16 — Результаты классификации данных динамики работы с клавиатурой в логах сервера идентификации.

5. В компоненте визуализации модуля идентификации отображается информация о результатах аномальности текущего анализируемого временного окна (см. рисунок 17).

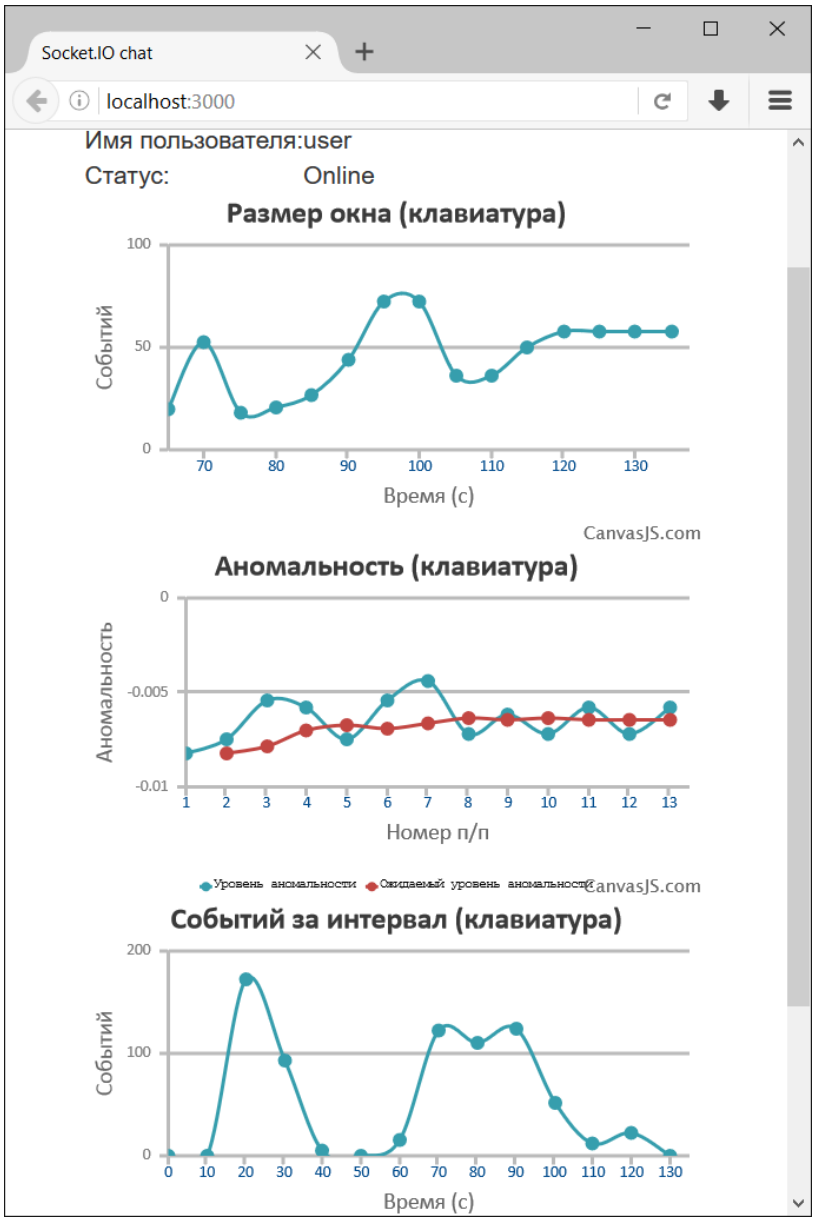

Рисунок 17 — Визуализация результатов идентификации пользователя на основе динамики работы с клавиатурой.

Оценка точности и скорости работы методов машинного обучения и математической статистики для построения и применения поведенческих моделей динамики работы пользователей с мышью

Для решения фоновой идентификации на основе динамики работы с мышью использовался следующий сценарий распознавания пользователей:

- выбирается контрольная группа пользователей;
- каждый из пользователей работает в обычном режиме на своем рабочем месте, на котором установлена система сбора информации в течение нескольких рабочих дней непрерывно;
- собранные данные используются для построения моделей идентификации пользователей по особенностям работы с мышью. Фиксируется время, затрачиваемое на построение моделей: вычисляются максимальное, минимальное и среднее время построения моделей;
- далее каждый из пользователей работает в обычном режиме на своем рабочем месте, на котором установлена система сбора информации в течение нескольких рабочих дней непрерывно и собираемые данные используются для проверки полученных моделей идентификации;
- применение построенных моделей к собранным данным и оценка качества классификации. Модели классификации применяются как к собственным новым данным пользователя, для которого была построена модель так и перекрестно по принципу «каждый против каждого».

Перебирались следующие параметры выделения окон на основе порогов на минимальное и максимальное количество событий в окне: верхняя граница окна (1000, 2000, 3000); нижняя граница окна (500, 1000, 1500); пауза прерывания окна (1 сек, 2 сек, 3 сек); процент перекрытия между окнами (0, 0.25, 0.5).

Перебирались следующие параметры при построении модели представления при формировании векторов признаков: количество областей экрана (4, 9, 16), количество интервалов расстояний (3, 20), количество направлений перемещения мыши (2, 4, 8).

Перебирались следующие параметры и методы постобработки признаков при построении модели представления: количество квантилей при дискретизации признаков (2, 8, 10); расчет межквартильных расстояний (да, нет); дискретизация на основе гистограмм (да, нет); применение метода главных компонент (да, нет); количество главных компонент (20, 50, 70); выбор наиболее стабильных признаков (да, нет); выбор признаков по уровню значимости (да, нет), количество отбираемых признаков (250, 200, 150, 100); уровень значимости отбираемых признаков (0.3, 0.5).

Перебирались следующие параметры метода опорных векторов при обучении модели классификации: параметр соотношения между долей выбросов и долей опорных векторов (0.1, 0.3, 0.5, 0.7), функция ядра (“rbf”, “poly”, “linear”), ширина ядра (0.01, 0.001, 0.002, 0.005, 0.008).

Перебирались следующие параметры нечеткого метода поиска исключений на основе потенциальных функций при обучении модели классификации: функция ядра (“rbf”, “poly”, “linear”), скорость убывания степени принадлежности в зависимости от расстояния до центра кластера (1.5, 1.7, 1.9), доля ожидаемых исключений (0.1, 0.3, 0.5), ширина ядра (0.005, 0.01, 0.1).

Перебирались следующие параметры метода ближайших соседей при обучении модели классификации: количество ближайших соседей (3, 5, 10).

Перебирались следующие параметры метода репликаторных нейронных сетей при обучении модели классификации: конфигурация сети (16-4-16, 256-16-256, 32-8-4-8-32).

Экспериментальные исследования проводились на вычислительном сервере SRV_SYS1: программные характеристики (ОС Microsoft Windows 2012R2, 64bit; интерпретатор Python3; модуль Python pandas 0.16.2; модуль Python numpy 1.9.3; модуль user-agent 1.0.1), аппаратные характеристики (два процессора x64 с тактовой частотой 2.4 Ггц, 4 ядра; оперативная память объемом 64 Гбайт; два дисковый накопителя HDD, объемом 2Тбайт в режиме RAID1; два дисковый накопителя HDD, объемом 1Тбайт; два сетевых адаптера с пропускной способностью 1 Гбит).

Были получены следующие результаты экспериментов. Лучшие результаты по точности классификации показал метод опорных векторов. Оптимальными оказались следующие параметры: функция ядра “rbf”, количество областей экрана – 9, количество направлений – 4, количество интервалов расстояний – 20, максимальный размер окна – 2000 событий, минимальный размер окна – 1000 событий, пауза – 1 сек, процент перекрытия окон – 0.5, количество квантилей – 2, расчет межквартильных расстояний (нет); дискретизация на основе гистограмм (нет); применение метода главных компонент (нет); выбор наиболее стабильных признаков (нет).

Результаты по скорости построения и применения моделей классификации приведены в протоколе №3 по пункту № 4.3.1.2 Программы и методики экспериментальных исследований «Подсистемы 1» ЭО ПК для сбора и анализа биометрической информации об особенностях динамики работы пользователя с устройствами ввода-вывода в рамках стандартного человеко-машинного интерфейса (см. Приложение А: таблица 6 Протокола испытаний №6).

Результаты по точности классификации приведены в протоколе №4 по пункту № 4.3.1.2 Программы и методики экспериментальных исследований «Подсистемы 1» ЭО ПК для сбора и анализа биометрической информации об особенностях динамики работы пользователя с устройствами ввода-вывода в рамках стандартного человеко-машинного интерфейса (см. Приложение А: таблица 4 Протокола испытаний №4).

Проверка по пункту № 4.3.1.3 Программы и методики экспериментальных исследований осуществлялась на основе оценки соответствия функциональности компонента фоновой идентификации аутентификации с целью сопоставления полученных результатов и ожидаемых результатов контрольных примеров №№21-22 Приложение Б Программы и методики экспериментальных исследований и/или проверки того, что они отражены в комплекте программной документации.

Проверка режима сбора данных об активности пользователя при работе с устройствами ввода состояла из следующих шагов:

- осуществление запуска приложения компонента сбора для начала контрольного сеанса сбора данных произведено без каких-либо сообщений об ошибках;
- выполнение сеанса обычной работы с использованием устройств ввода длительностью в 15 минут прошло без сообщений об ошибках и визуально не отличалось от работы при неактивной компоненте сбора;
- нажатие CTRL+SHIFT+S с использованием служебных клавиш в левой половине клавиатуры привело к открытию окна управления сборщиком данных (см. рисунок 14);
- завершение контрольного сеанса сбора данных нажатием на кнопку «Cancel». В файловой системе сформирована структура каталогов с текстовыми файлами, содержащими логи программы (program.log и _internal.log), описание целевой платформы (systeminfo.log) и собранные в течении сессии данные об использовании клавиатуры и мыши (keybard.csv и mouse.csv).

Проверка режима фоновой идентификации пользователя при работе с устройствами ввода состояла из следующих шагов.

1. Активирование рабочего место аналитика (модуль идентификации) путем запуска соответствующего Python модуля на сервере прошло без каких-либо ошибок.
2. Запуск приложения компонента сбора в режиме идентификации прошел без каких-либо ошибок.
3. Выполнение сеанса обычной работы с использованием устройств ввода продолжительностью, достаточной для формирования временного окна. В результате в

дополнение к сформированной в результате сбора данных структуре каталогов и файлов в целевом каталоге создаются файлы features_mouse.csv, structure.csv (см. рисунок 15).

4. В логах сервера (модуля идентификации) появляется запись о результате классификации очередного вектора признаков, полученных в результате обработки сформированного временного окна пользователя (см. рисунок 18).

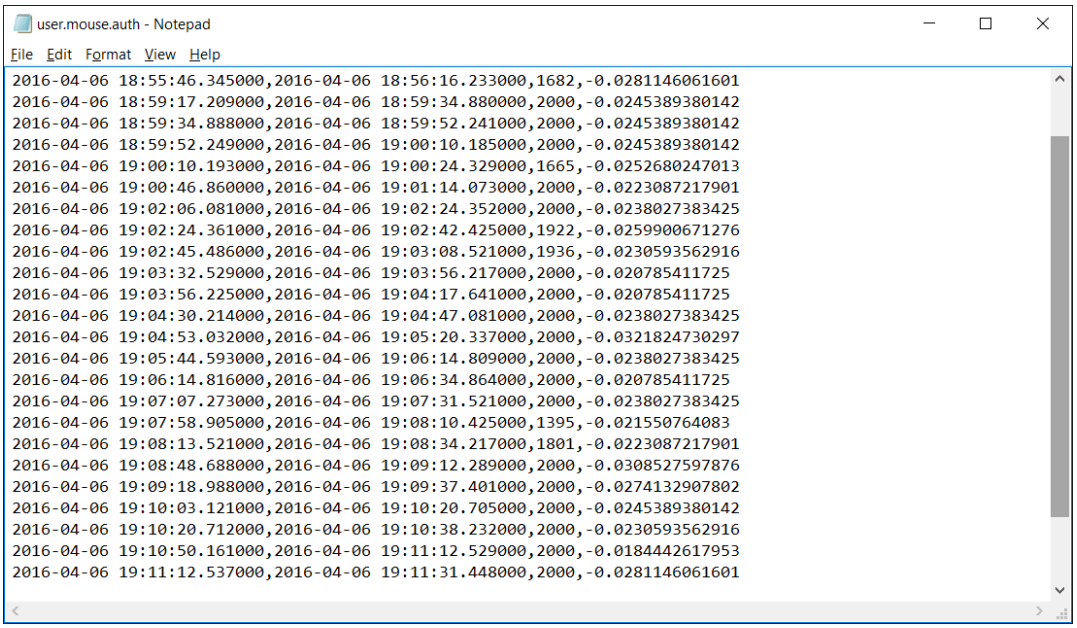

Рисунок 18 — Результаты классификации данных динамики работы с мышью в логах сервера идентификации.

5. В компоненте визуализации модуля идентификации отображается информация о результатах аномальности текущего анализируемого временного окна (см. рисунок 19).

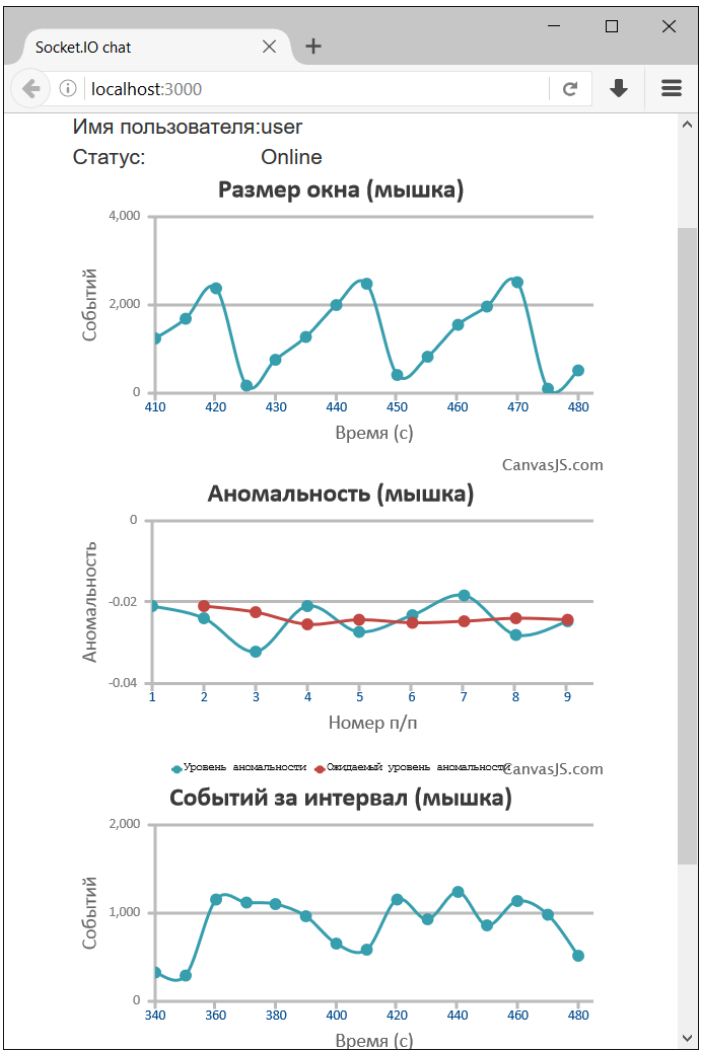

Рисунок 19 — Визуализация результатов идентификации пользователя на основе динамики работы с мышью.

1.3.1.2 Оценка точности и скорости работы методов машинного обучения и математической статистики для построения и применения поведенческих моделей для анализа биометрической информации для решения задачи активной аутентификации пользователей на основе событий ввода данных с помощью клавиатуры и мыши

Экспериментальные исследования проводились в соответствие с пунктами 4.3.1.2 и 4.3.1.3 Программы и методики экспериментальных исследований. Целью проводимых экспериментальных исследований является оценка точности активной аутентификации пользователей с применением методов машинного обучения для построения и применения

моделей, используемых в «Подсистеме 1» ЭО ПК, и выбор наилучших параметров методов как для аутентификации на основе ввода данных с помощью клавиатуры, так и с помощью мыши.

Для решения задачи активной аутентификации используется следующий сценарий как для варианта распознавания по активности с клавиатурой, так и для варианта распознавания по активности с мышью.

1. Выбирается контрольная группа пользователей 10 человек.
2. Для каждого пользователя осуществляется обучение системы аутентификации отдельно для мыши и для клавиатуры, строится модель распознавания для своего контрольного открытого слова и для своего открытого графического жеста (подписи).
3. При обучении фиксируется время обучения модели, отдельно оценивая как время, необходимое на подготовку тренировочной выборки (ввода пользователями слов и жестов), так и время, затрачиваемое на построение модели. Рассчитываются минимальные, максимальные и усредненные показатели по времени по всем пользователям в разрезе отдельно по клавиатуре и отдельно по мыши.
4. Далее, перекрестно по принципу «каждый против каждого» проводится эксперимент, включающий фиксированное число попыток авторизации, как по своей модели, так и по чужой, отдельно с помощью мыши и с помощью клавиатурного ввода. Размер тестового набора составляет 30 примеров.
5. Осуществляется применение построенных моделей к собранным данным и оценка качества классификации. Также оценивается время, затрачиваемое на распознавание ввода, и оцениваются минимальное, максимальное и среднее значение отдельно по клавиатуре и по мыши.

При идентификации пользователя критериями качества работы методов являются количество вводов пароля легитимным пользователем, которые система ошибочно распознала как чужие, и количество вводов пароля другим пользователем, которые системы ошибочно распознала как свои. В качестве количественной характеристики качества работы классификатора использовался показатель AUC (площадь под ROC-кривой).

Экспериментальные исследования проводились на вычислительном сервере SRV_SYS1: программные характеристики (ОС Microsoft Windows 2012R2, 64bit; интерпретатор Python3; модуль Python pandas 0.16.2; модуль Python numpy 1.9.3; модуль user-agent 1.0.1), аппаратные характеристики (два процессора x64 с тактовой частотой 2.4 Ггц, 4 ядра; оперативная память объемом 64 Гбайт; два дисковый накопителя HDD, объемом 2Тбайт

в режиме RAID1; два дисковый накопителя HDD, объемом 1Тбайт; два сетевых адаптера с пропускной способностью 1 Гбит).

Время отклика интерфейса при вынесении решения об отказе авторизации или успешной аутентификации пользователя менее одной секунды, основную часть которого составляет процедура авторизации пользователя в самой операционной системе. Отличия времени аутентификации пользователя при использовании стандартного метода авторизации (по паролю) и разработанных (по паролю и по мыши) незаметны.

Для решения задачи активной аутентификации пользователей на основе динамики их работы с клавиатурой предлагается использовать следующий сценарий определения легитимности пользователей.

1. Выбирается контрольная группа пользователей.
2. Каждый из пользователей придумывает свою парольную фразу, которую он часто вводил на клавиатуре (чтобы у него сформировались поведенческие привычки при вводе этой парольной фразы).
3. Вводит фразу в специальное поле 10 раз. Во время ввода собираются следующие данные:
   - интервалы времени между нажатиями клавиш;
   - время удержания клавиш;
   - интервалы времени между нажатиями клавиши и отпусканием предыдущей.
4. Собранные данные используются для построения моделей работы пользователей с клавиатурой.

Данные о каждой попытке ввода записываются последовательно в вектор признаков. Для каждого признака в векторе проводится нормализация на величину среднеквадратичного отклонения. Набор из 10 нормализованных векторов подаётся в качестве обучающего набора в классификатор на основе нечёткого метода поиска исключений с использованием потенциальных функций [1]. В качестве потенциальной функции была выбрана функция Гаусса. Каждому вектору признаков сопоставляется элемент нечёткого множества, то есть оценивается его принадлежность этому множеству, а также расстояние от центра множества до элемента. Задача построения такого нечёткого множества сводится к тому, чтобы найти такой центр множества, расстояния от которого до элементов множества были бы минимальны, а соответственно, принадлежность была бы максимальной. При этом учитывается, что некоторая часть попыток ввода пользователя является “выбросами”. То есть при их введении была допущена ошибка, и они не являются репрезентативным

представлением ввода пользователя. Далее каждый пользователь при входе в операционную систему, на которой установлена система аутентификации, вводит парольную фразу.

Во время его ввода аналогичным образом (как во время создания модели) строится вектор признаков, каждый признак которого нормализуется на величину, на которую нормализовывались признаки в обучающем наборе. После чего нормализованный вектор признаков подаётся в классификатор на распознавание. Определяется степень принадлежности этого вектора построенному для данного пользователю нечёткому множеству. Если принадлежность данной попытки аутентификации превышает порог, определяющий границу разделения легитимных и нелегитимных попыток аутентификации для данного пользователя, то аутентификация считается успешной, информация об этом подаётся в операционную систему, и производится авторизация. Иначе аутентификация считается не выполненной. Модели классификации применяются как к собственным новым данным пользователя, для которого была построена модель, так и перекрестно к данным других пользователей по принципу «каждый против каждого».

Были получены следующие результаты экспериментов. Для аутентификации на основе динамики работы с клавиатурой точность оценивалась для четырех типов паролей: учетная запись электронной почты пользователя, слова "Google", "Google20" и "Google2016". В результате проведенных экспериментов оптимальными оказались следующие параметры алгоритма: сигма (параметр функции ядра для распределения Гаусса) - 1.0; степень размытости нечёткого множества - 1.1; доля ошибочных попыток - 0.1; количество попыток, использованное для построения модели, - 10.

Для решения задачи активной аутентификации пользователей на основе динамики их работы с мышкой предлагается использовать следующий сценарий определения легитимности пользователей.

1. Выбирается контрольная группа пользователей.
2. Каждый из пользователей 30 раз обводит шаблон, нарисованный на экране, - за это время пользователь привыкает к шаблону, и у него формируются индивидуальные скоростные поведенческие характеристики на различных участках шаблона.
3. Во время ввода собираются следующие данные:
   - координата X – положения курсора мыши на экране;
   - координата Y – положения курсора мыши на экране;
   - время регистрации события изменения координат мыши.

Собранные данные требуют предварительной обработки в связи с тем, что события перемещения мыши в операционную систему поступают неравномерно по времени. Для

нормализации временного шага оба временных ряда (ряд X - координат и ряд Y - координат) интерполируются с равным шагом кубическими сплайнами. После этого шага для каждого образца на выходе будет получено два временных ряда с равными интервалами времени между событиями. Интервал между точками интерполяции выбирается исходя из того, чтобы включить всё время сбора событий (от начала и до конца) и одновременно сохранить количество зарегистрированных событий до интерполяции и после. Набор из 10 последних образцов используется для построения двух моделей для каждого пользователя: по X – координатам и по Y – координатам.

В качестве классификатора для решения задачи активной аутентификации пользователей на основе динамики их работы с мышкой используется метод, применённый для классификации в алгоритме аутентификации по динамике работы с клавиатурой, - нечёткий метод поиска исключений с использованием потенциальных функций [1]. В качестве потенциальной функции была выбрана функция Гаусса, при этом расстояние между элементами рассчитывается как стоимость трансформации первого элемента ко второму (в качестве элементов выступают временные ряды) по алгоритму DTW. Таким образом, после этого шага для каждого пользователя будет создано по две модели. Далее каждый пользователь при входе в операционную систему, на которой установлена система аутентификации, обводит появляющийся на экране шаблон.

Во время рисования по шаблону аналогичным образом регистрируются события перемещения курсора мыши. Далее два полученных временных ряда интерполируются с вычисленным на этапе построения модели временный шагом. Временной ряд X – координат подаётся на распознавание в модель, построенную по X – координатам, а временной ряд Y – координат в модель для Y. На выходе получаем два числа – значения принадлежностей текущего образца построенным моделям по X и по Y-координатам соответственно. В качестве итоговой принадлежности принимается среднее арифметическое принадлежностей по каждой из моделей. Если принадлежность данной попытки аутентификации превышает порог, определяющий границу разделения легитимных и нелегитимных попыток аутентификации для данного пользователя, то аутентификация считается успешной, информация об этом подаётся в операционную систему, и производится авторизация. Иначе аутентификация считается не выполненной. Модели классификации применяются как к собственным новым данным пользователя, для которого была построена модель, так и перекрестно к данным других пользователей по принципу «каждый против каждого».

Для аутентификации на основе динамики работы с мышкой точность оценивалась на фиксированном шаблоне в виде двух последовательно написанных прописных букв “A". В результате проведённых тестов оптимальными оказались следующие параметры алгоритма: сигма (параметр функции ядра для распределения Гаусса) - 0.8; степень размытости нечёткого множества - 1.2; доля ошибочных попыток - 0.1; количество попыток, использованное для построения модели, - 10.

Результаты по скорости построения и применения моделей классификации приведены в протоколе №3 по пункту № 4.3.1.2 Программы и методики экспериментальных исследований «Подсистемы 1» ЭО ПК для сбора и анализа биометрической информации об особенностях динамики работы пользователя с устройствами ввода-вывода в рамках стандартного человеко-машинного интерфейса (см. Приложение А: таблица 6 Протокола испытаний №6).

Результаты по точности классификации приведены в протоколе №4 по пункту № 4.3.1.2 Программы и методики экспериментальных исследований «Подсистемы 1» ЭО ПК для сбора и анализа биометрической информации об особенностях динамики работы пользователя с устройствами ввода-вывода в рамках стандартного человеко-машинного интерфейса (см. Приложение А: таблица 8 Протокола испытаний №8).

Проверка соответствия функциональности компонента активной аутентификации осуществлялась с целью оценки соответствия полученных результатов ожидаемым результатам контрольных примеров №№18-20 Приложение Б Программы и методики экспериментальных исследований и/или проверки того, что они отражены в комплекте программной документации.

Проверка корректности интеграции модулей авторизации пользователей в стандартную схему авторизации Windows XP состояла из следующих шагов.

1. Установка dll-компоненты активной аутентификации согласно описанию программы.
2. Выполнение перезагрузки компьютера. После перезагрузки появилось специализированное окно входа в систему (см. рисунок 20).

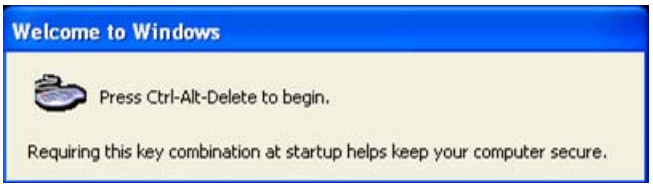

Рисунок 20 — Интерфейс окна входа в систему.

3. В результате нажатия контрольной комбинацию клавиш CTRL + ALT + DELETE открылось модифицированное окно авторизации пользователя с поддержкой использования биометрических характеристик (см. рисунок 21).

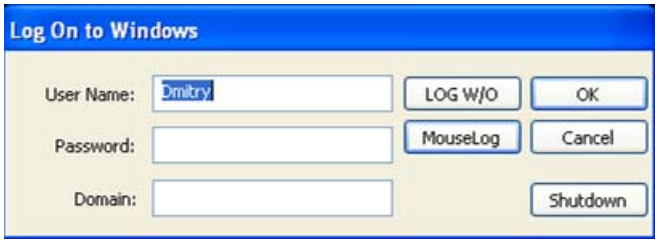

Рисунок 21 — Модифицированное окно авторизации пользователя.

Проверка режима авторизации пользователя на основе биометрических данных с использованием клавиатуры состояла из следующих шагов.

1. В результате открытия окна авторизации появляется следующий интерфейс (см. рисунок 8).
2. Введение имени пользователя в поле User Name.
3. Введение пароля в привычном пользователю ритме и нажатие кнопки «OK». После нажатия OK в случае проведение авторизации легитимным пользователем (соответствие биометрических данных ранее построенной модели и совпадение пароля) окно авторизации исчезает, и пользователь может начать работать с системой. В ином случае - в процессе авторизации фиксируется ошибка (несоответствие биометрических данных либо неверный пароль), и пользователю предоставляется возможность введения пароля повторно или использовать другой способ авторизации.

Проверка режима авторизации пользователя на основе биометрических данных с использованием мыши состояла из следующих шагов.

1. В результате открытия окна авторизации появляется следующий интерфейс (см. рисунок 3).
2. В результате нажатия в окне авторизации кнопки «MouseLog» открывается окно авторизации с использованием мыши (см. рисунок 22).

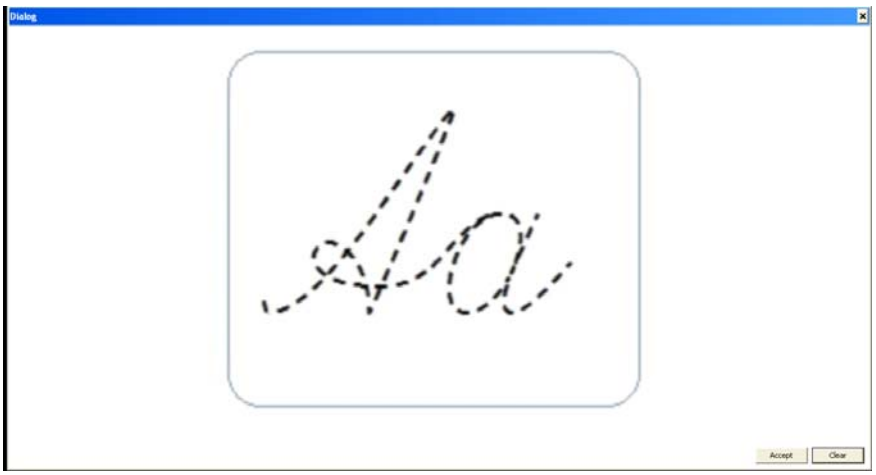

Рисунок 22 — Окно авторизации с использованием мыши.

3. Очерчивание появившегося шаблона путём зажатия левой кнопки мыши и последующего обвода шаблона привело к следующему результату (см. рисунок 23).

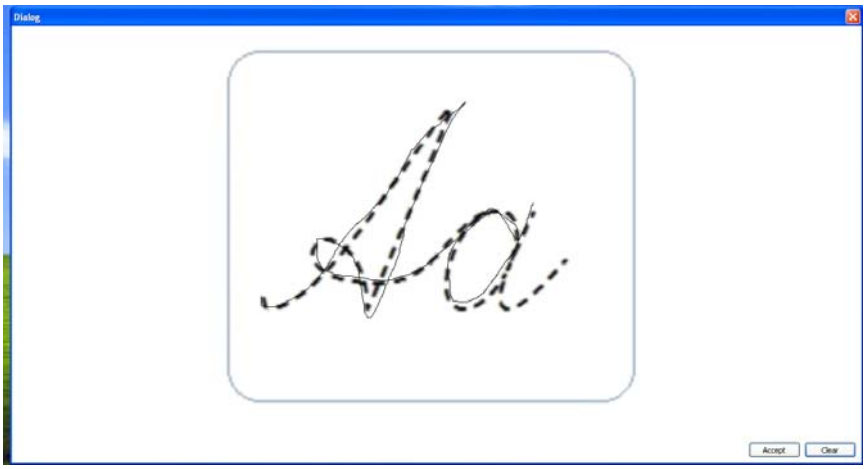

Рисунок 23 — Результат очерчивания шаблона мышью.

Подтверждение текущей попытку входа нажатием на кнопку «Accept». В случае осуществления попытки легитимным пользователем (соответствия биометрических данных ранее построенной модели) окно авторизации исчезает и пользователь может приступить к работе с системой. В ином случае (биометрические данные не соответствуют), пользователю предоставляется возможность повторить процесс авторизации с использованием мыши или выбрать другой способ авторизации.

1.3.2 Оценка точности и скорости работы методов машинного обучения и математической статистики для построения и применения поведенческих

моделей для анализа данных о работе пользователя с информационными и вычислительными ресурсами компьютерной системы

Экспериментальные исследования по пункту № 4.3.2.2 Программы и методики экспериментальных исследований, проводимые с целью оценки точности и скорости работы методов машинного обучения и математической статистики для построения и применения поведенческих моделей «Подсистемы 2» ЭО ПК, выполняются по следующему сценарию.

1. Используются проверочные наборы данных, подготовленные на предыдущем этапе и дополненные новыми собранными данными в рамках экспериментов и эталонными данными из репозитория DARPA.
2. Выбранные наборы данных загружаются в аналитические витрины, на которых строятся все перечисленные выше типы отчетов с фиксацией времени, необходимого для построения отчетов каждого из следующих типов:
   - тип отчета 1: отчеты о фактах запуска пользователем программ и приложений;
   - тип отчета 2: отчеты о фактах работы пользователя с внешними запоминающими устройствами (запись и чтение информации);
   - тип отчета 3: отчеты о фактах обращения пользователя к файлам и папкам, в том числе на удаленных АРМ;
   - тип отчета 4: отчеты о фактах работы пользователя в ЛВС;
   - тип отчета 5: отчеты по интенсивности работы пользователя с клавиатурой и манипулятором «мышь» в каждом из активных приложений;
   - тип отчета 6: отчеты о фактах самостоятельной установки и удаления пользователем программного обеспечения и/или технических средств на АРМ.
3. На заполненных аналитических витринах строятся модели идентификации пользователей с использованием методов латентно-семантического анализа слабо структурированных данных в собранных множествах событий и фактов и методов анализа временных рядов, предложенных на предыдущем этапе для выявления фактов изменения тренда в поведении пользователей. В качестве количественной характеристики оценки точности построенных моделей используется показатель AUC (площадь под ROC-кривой) — площадь, ограниченная ROC-кривой и осью доли ложных положительных классификаций. Чем выше показатель AUC, тем качественнее классификатор. При применении моделей также фиксируется время, затрачиваемое на распознавание для каждого временного окна, при этом оценивается максимальное, минимальное и среднее время применения моделей:

- фиксируется время, застрачиваемое на построение моделей идентификации с указанием числа анализируемых событий и размера временного окна;
- для «раскрашенных» данных с известным реальным откликом (для «имитации» изменения тренда в случае отсутствия примеров «подмены пользователя» в исходных событиях может использоваться искусственно добавленный или сгенерированный набор «чужих» событий), т.е. информации о принадлежности событий тому или иному пользователю, строятся модели идентификации на данных за тренировочный период, затем они применяются на данных за проверочный период и оценивается качество полученных моделей распознавания;
- в качестве меры оценки моделей используется площадь под ROC кривой – AUC.

4. На заполненных аналитических витринах строятся модели раннего обнаружения внутренних вторжений с использованием методов на основе ассоциативных правил, предложенных на предыдущем этапе:
   - фиксируется время, застрачиваемое на построение моделей раннего обнаружения с указанием числа анализируемых событий и типа анализируемых событий;
   - для «раскрашенных» данных с известным реальным откликом (для «имитации» изменения тренда в случае отсутствия примеров «подмены пользователя» в исходных событиях может использоваться искусственно добавленный или сгенерированный набор «чужих» событий), т.е. информации о принадлежности событий тому или иному пользователю, строятся модели раннего обнаружения внутренних вторжения на данных за тренировочный период, затем они применяются на данных за проверочный период и оценивается качество полученных моделей распознавания;
   - в качестве меры оценки моделей используется площадь под ROC кривой – AUC.

Экспериментальные исследования по оценке точности и скорости работы методов машинного обучения и математической статистики для построения и применения моделей проводились на вычислительном сервере SRV_SYS2 - программные характеристики: ОС Microsoft Windows 2008R2, 64bit; Microsoft Office 2007;Microsoft SQL Server 2012; Microsoft SQL Server 2005;Microsoft SQL Server 2005 Analysis Services;Microsoft SQL Server 2005 Management Studio; интерпретатор Python 2.7.3; модуль Python pywin 219; библиотека Python Win32 Extensions; библиотека Natural Language Toolkit (NLTK); аппаратные характеристики: два процессора x64 с тактовой частотой 2 Ггц, 4 ядра; оперативная память объемом 48 Гбайт;

два дисковый накопителя HDD, объемом 2Тбайт;два сетевых адаптера с пропускной способностью 1 Гбит.

Эксперименты по оценке скорости и точности построения и применения моделей раннего обнаружения вторжений проводились по сценарию, моделирующему кражу секретной информации. В течение восьми дней проводился сбор данных о повседневной работе пользователя – на этих данных строится модель нормального поведения пользователя (при построении модели раннего обнаружения вторжений исключен процесс avp.exe - антивирус Касперского). Пользователь работал со следующими приложениями: Outlook, Skype, Excel, Word, Winrar. За это время количество зафиксированных событий с клавиатурой в этих приложениях составило 16559, 1223, 543, 196 и 46 соответственно; количество событий с мышью - 171211, 11717, 32227, 8645 и 2417. Объем трафика, переданный за это время через Outlook, составил 33 Мб, через Skype - 13 Мб; объем трафика, полученный через Outlook, составил 954 Мб, через Skype - 20 Мб. Анализируемый сценарий поведения пользователя, моделирующий кражу секретной информации, включал в себя следующую последовательность действий, выполняемых пользователем в течение получаса:

1. Пользователь открывает Total Commander.
2. Создает новую папку в рабочей директории.
3. Запускает приложение Remote Desktop.
4. Заходит на другой компьютер локальной сети.
5. Копирует с этого компьютера в созданную директорию пять файлов и две папки, общим объемом 223 Мб.
6. Осуществляет архивирование этой папки, используя приложение Winrar (объем полученного архива 68 Мб).
7. Открывает приложение Outlook и отправляет полученный архив.
8. Заходит в интернет-браузер и копирует архив на внешний сервер.
9. Удаляет архив из рабочей директории.
10. Очищает корзину.

Эксперименты по оценке скорости и точности построения и применения моделей идентификации проводились по следующему сценарию:

1. Модель нормального поведения пользователя строится на данных, собранных в процессе обычной работы пользователя за компьютером в семи дней.
2. В течение одного дня на рабочем месте этого пользователя работает другой пользователь.

3. Оценивается степень аномальности действий, выполненных этими пользователями. При построении модели идентификации использовались следующие параметры: количество квантилей 10, число тематик 3, параметры алгоритма работы с тестовыми данными – параметр ортоганальности 100, минимальное значение числа степеней свободы 2, начальное число при генерации случайных чисел 5, число итераций 100.

Результаты по скорости построения и применения моделей приведены в протоколе №1 по пункту № 4.3.2.2 Программы и методики экспериментальных исследований «Подсистемы 2» ЭО ПК для сбора и анализа данных о работе пользователя с информационными и вычислительными ресурсами компьютерной системы.

Результаты по точности классификации приведены в протоколе №2 по пункту № 4.3.2.2 Программы и методики экспериментальных исследований «Подсистемы 2» ЭО ПК для сбора и анализа данных о работе пользователя с информационными и вычислительными ресурсами компьютерной системы (см. Приложение А: таблица 12 Протокола испытаний №12).

Проверка по пункту № 4.3.2.3 Программы и методики экспериментальных исследований осуществлялась на основе оценки соответствия функциональности с целью сопоставления полученных результатов и ожидаемых результатов контрольных примеров №№6-17 Приложение Б Программы и методики экспериментальных исследований и/или проверки того, что они отражены в комплекте программной документации.

Проверка создания новой витрины данных состоит из следующих шагов.

1. Запустить АРМ аналитика безопасности, если оно не запущено. В результате откроется АРМ аналитика безопасности.
2. Создать новую витрину данных. Для создания витрины данных нажать правой кнопкой мыши на «Витрины данных», в появившемся меню выбрать «Создать Data Mart». Указать требуемые параметры и нажать «Создать». В результате в дереве «Витрины данных» появится вновь созданная витрина данных.

Проверка связывания витрины данных с АРМ пользователей состоит из следующих шагов.

1.Запустить АРМ аналитика безопасности, если оно не запущено. В результате откроется АРМ аналитика безопасности.

2. В дереве «Витрины данных» выбрать требуемую витрину, нажать на нее правой кнопкой мыши и выбрать «Задать набор агентов». В результате откроется диалоговое окно

выбора Профилей и АРМ пользователей, установленных в рамках профилей, для связывания с витриной, как показано на рисунке 24.

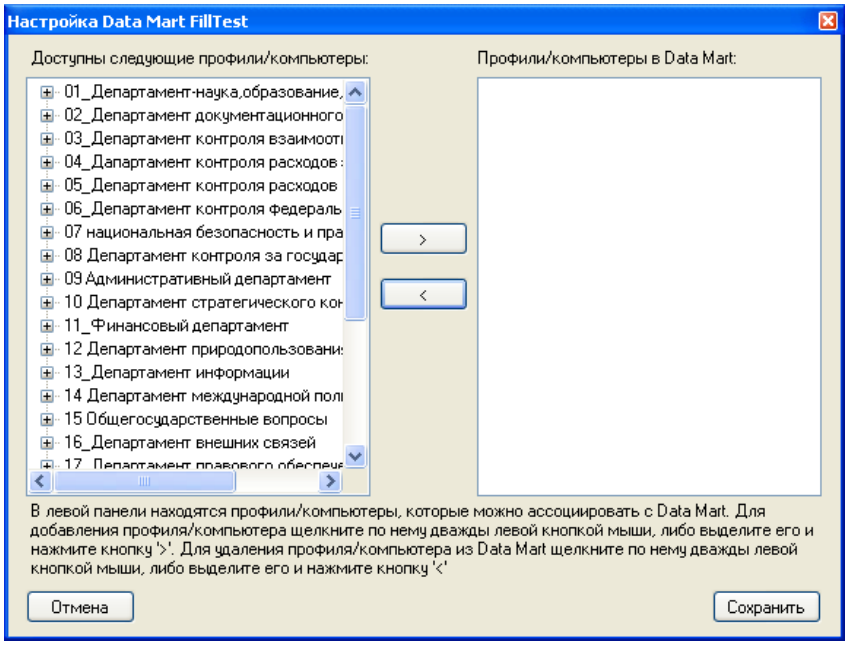

Рисунок 24 — Интерфейс настройки витрины данных.

3. В левом списке окна выбрать интересующие профили и/или агенты и перенесите их в правую часть окна, либо двойным щелчком мыши по имени, либо используя кнопку «>». В результате в правой части формы будут указаны требуемые профили/АРМ пользователей (см. рисунок 25).

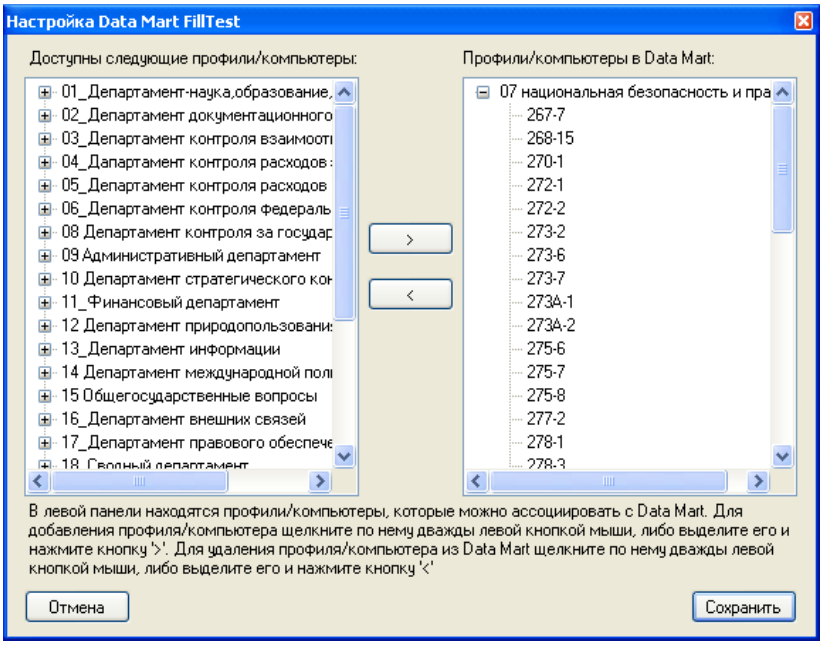

Рисунок 25 — Интерфейс процесса настройки витрины данных.

4. Нажать кнопку «Сохранить». В результате окно настройки витрины данных закроется.

Проверка заполнения витрины данных требуемым срезом данных состоит из следующих шагов.

1. Запустить АРМ аналитика безопасности, если оно не запущено. В результате откроется АРМ аналитика безопасности.

2. Нажать правой кнопкой мыши на витрине данных измененной в предыдущем контрольном примере, в появившемся меню выбрать «Заполнить». В результате откроется диалоговое окно заполнения витрины данных (см. рисунок 26).

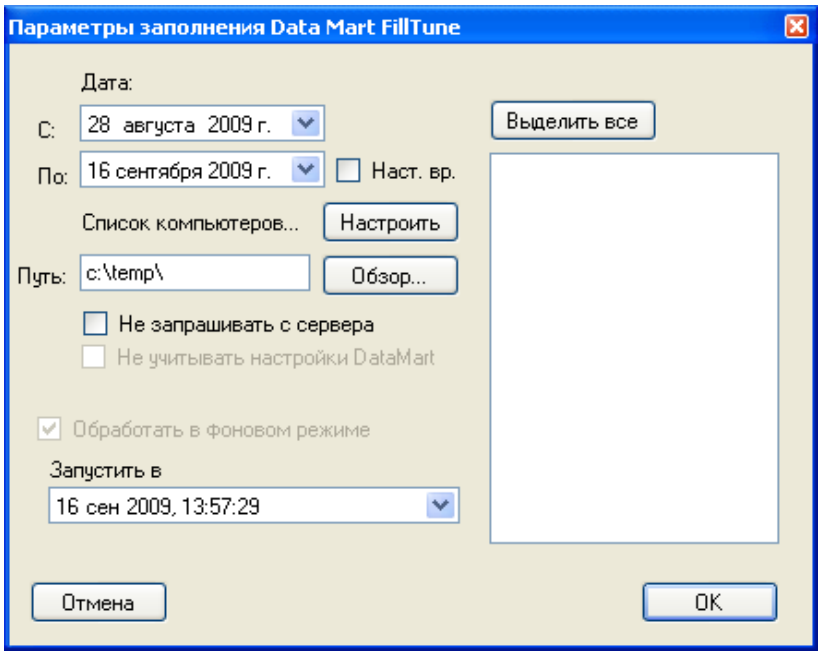

Рисунок 26 — Параметры заполнения витрины данных.

3. Нажать кнопку «Настроить». Убедитесь, что в открывшемся окне указаны требуемые профили/АРМ, установленные в предыдущем контрольном примере. Нажать кнопку «Отмена» диалогового окна «Настройка Data Mart <имя витрины данных>». В результате витрина данных связывается с необходимыми профилями и/или АРМ пользователей.

4. В разделе дата указать требуемый интервал для заполнения. Отметить опцию «Предварительно очистить». Нажать кнопку «ОК». В результате в рамках витрины данных добавится новое задание на заполнение витрины данных.

5. В рамках витрины данных выбрать пункт «Список задач». В результате в списке задач представлено добавленное задание на заполнение витрины данных. Задание начнет выполняться (см. рисунок 27). После достижения статуса 100% витрина данных заполнена. Время заполнения зависит от количества АРМ, с которыми связана витрина данных и от временного диапазона и может занять значительное время. Сервер консолидации на время заполнения должен быть включен, связь с сервером консолидации должна присутствовать.

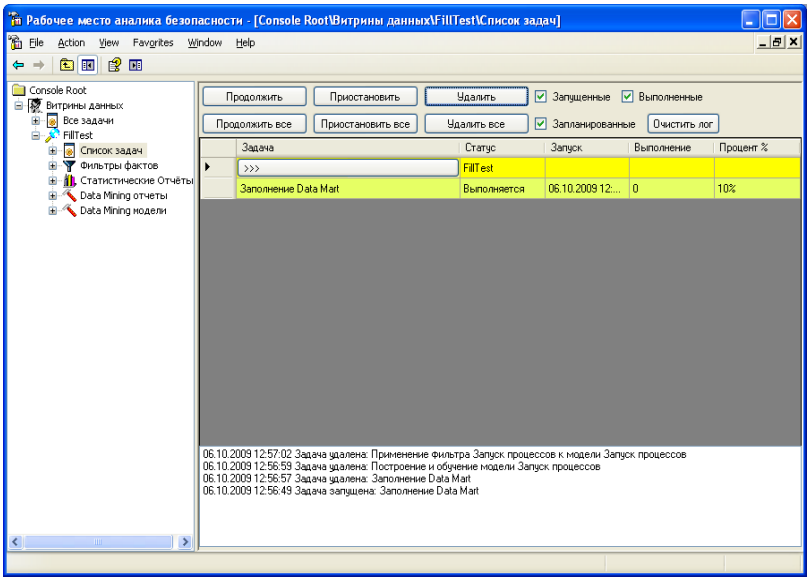

Рисунок 27 — Список задач рабочего места аналитика безопасности.

Проверка механизмов формирования динамических статистических отчетов в виде таблицы состоит из следующих шагов.

1.Запустить АРМ аналитика безопасности, если оно не запущено. В результате откроется АРМ аналитика безопасности.

2. Создать в произвольной заполненной витрине данных статистический отчет в виде таблицы для каждого типа событий. Для создания каждого отчета нажать правой кнопкой мыши «Статистические отчеты», в появившемся меню выбрать «Создать отчет». На форме создания отчета в графе «Тип фактов» свой тип куба из списка: Logins Cube, Processes Cube, Traffic Cube, Active Cube, DeviceStat Cube, Hardware Cube, Software Cube. Открыть произвольный статистический отчет (например, по фактам запуска процессов – Processes Cube). Добавить в отчет меру «количество фактов». В результате создаются отчеты по всем требуемым типам фактов, можно открыть выбранный статистический отчет с добавленной мерой (см. рисунок 28).

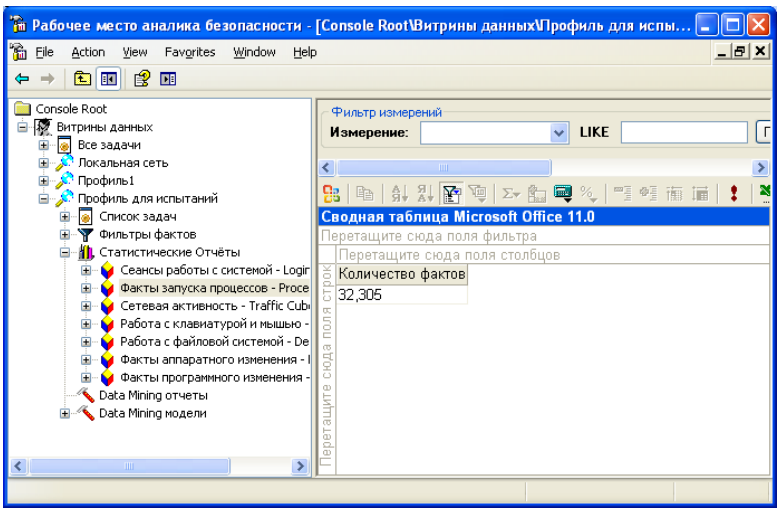

Рисунок 28 — Статистический отчет с добавленной мерой.

3. Задать фильтр на произвольное измерение (например, на имя пользователя), либо путем отмечания мышкой нужных пользователей, либо путем выбора измерения в поле «Измерение» и указания условия в поле «LIKE». В результате в отчете будут присутствовать только данные, удовлетворяющие фильтру (см. рисунок 29).

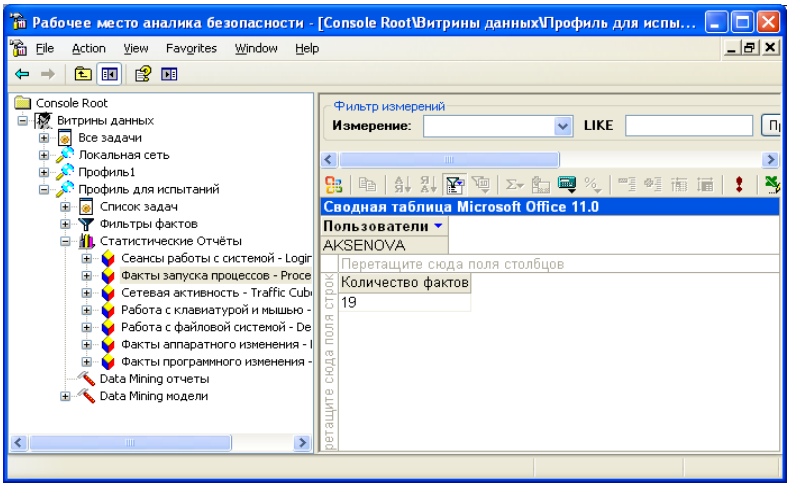

Рисунок 29 — Фильтр статистического отчета.

4. Задать группировку по произвольному измерению, например, по имени процесса, путем добавления соответствующего измерения на отчет. В результате в отчете будут отображаться данные с учетом заданной группировки (см. рисунок 30).

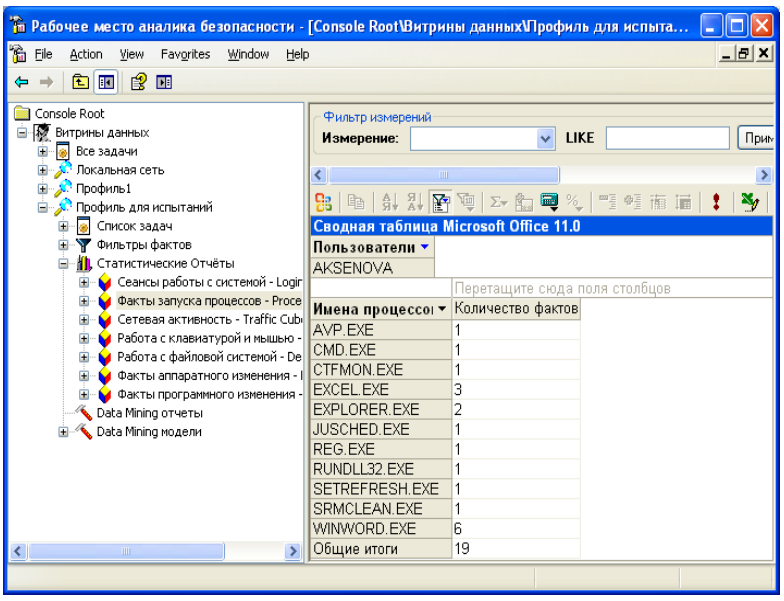

Рисунок 30 — Группировка по измерению в статистическом отчете.

5. Добавить в отчет еще одно произвольное измерение (например, время) рядом с уже добавленным на предыдущем шаге измерением, выполнить операцию drill-down (уточнение), с помощью раскрытия значения измерения. Операция rollup (обобщение) является обратной операции drill-down (уточнение). В результате в отчете будут отображаться данные с учетом заданной группировки, дополнительно, данные по выбранной группе будут развернуты по выбранному измерению (см. рисунок 31).

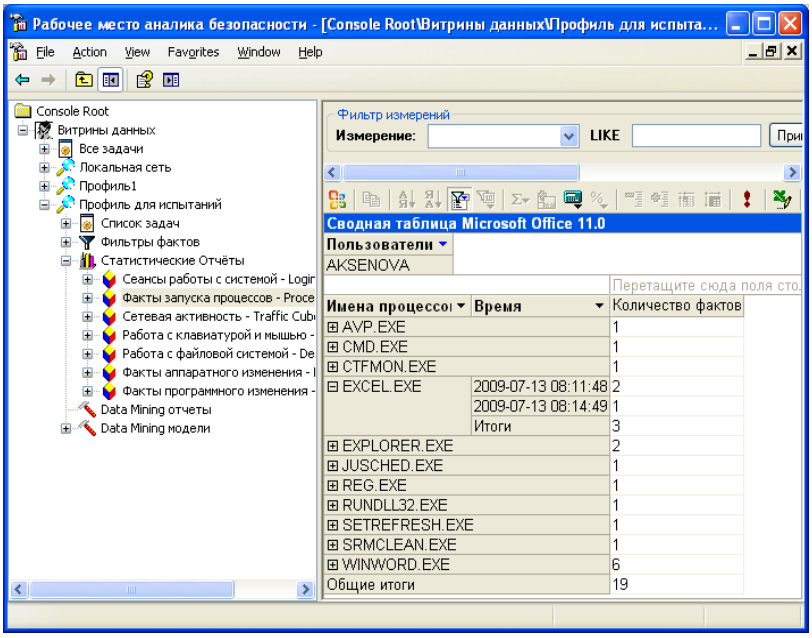

Рисунок 31 — Уточнение данных по выбранному измерению.

Проверка механизмов формирования динамических статистических отчетов в виде диаграммы состоит из следующих шагов.

1.Запустить АРМ аналитика безопасности, если оно не запущено. В результате откроется АРМ аналитика безопасности.

2. Создать в произвольной заполненной витрине данных отчет в виде диаграммы для произвольного типа событий, например, Active Cube. Для создания отчета нажать правой кнопкой мыши «Статистические отчеты», в появившемся меню выбрать «Создать отчет». На форме создания отчета в графе «Тип фактов» выбрать Active Cube, указать «В виде диаграммы». Открыть созданный статистический отчет. В результате откроется заготовка созданного статистического отчета (см. рисунок 32).

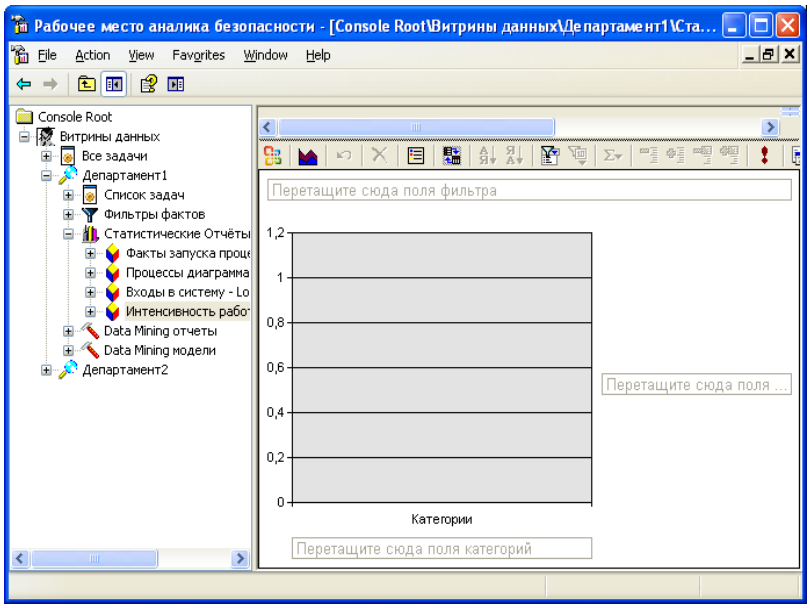

Рисунок 32 — Заготовка созданного статистического отчета.

3. Выбрать тип диаграммы и перетащите на отчет необходимые меры и измерения. Например, для формирования отчета интенсивности работы по часам перетащить «Количество операций с клавиатурой» и «Час». В результате отобразятся требуемые данные (см. рисунок 33).

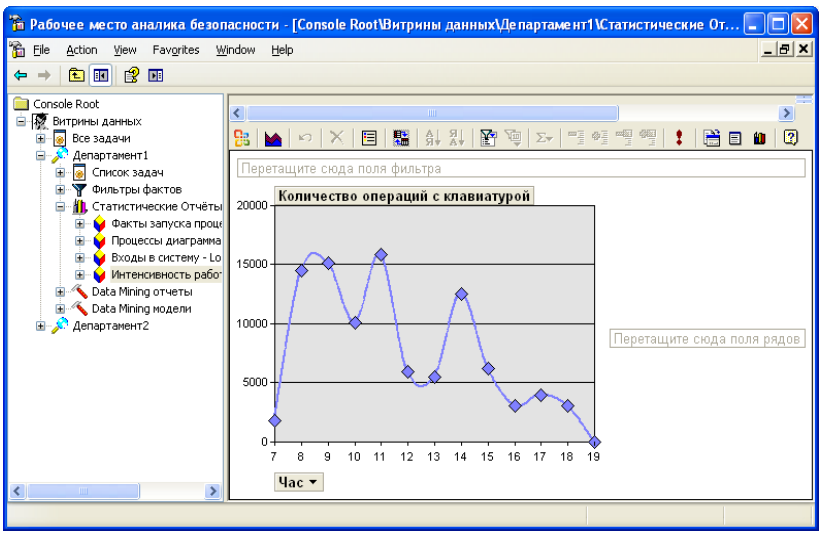

Рисунок 33 — Пример отчета интенсивности по часам.

Проверка создания фильтров данных состоит из следующих шагов.

1.Запустить АРМ аналитика безопасности, если оно не запущено. В результате откроется АРМ аналитика безопасности.

2. В произвольной заполненной витрине данных выбрать раздел «Фильтры фактов», нажать на него правой кнопкой мыши и выбрать «Создать фильтр». В результате откроется диалоговое окно «Создание фильтра» (см. рисунок 34).

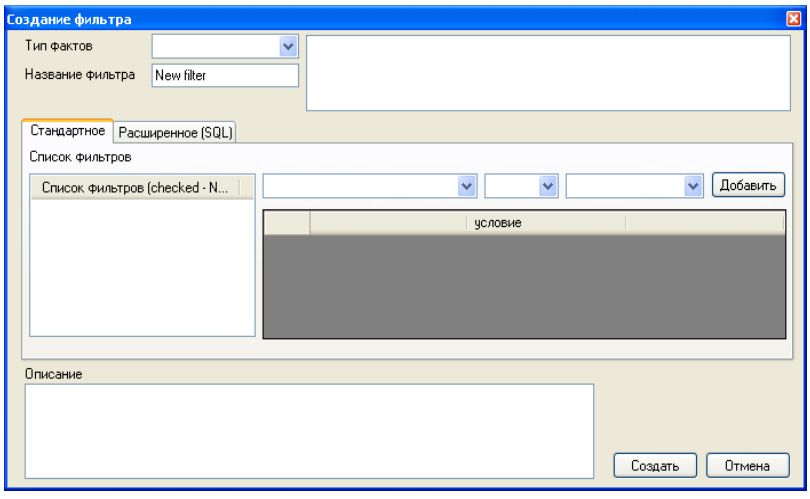

Рисунок 34 — Интерфейс создания фильтра.

3. Выбрать интересующий тип фактов. В результате поля задания фильтра заполнятся значениями для соответствующего типа фактов (см. рисунок 35).

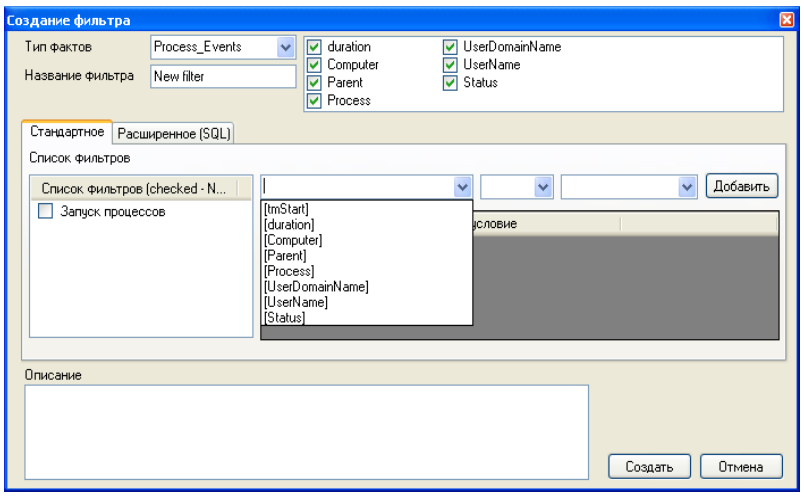

Рисунок 35 — Интерфейс окна создания фильтра с выбранным типом фактов.

4. В верхней части формы отметить галочками требуемые атрибуты событий, которые будут присутствовать в результате фильтрации (по умолчанию все). В средней части формы задать значения фильтра на требуемые атрибуты. В результате диалоговое окно создания фильтра отражает выбранные атрибуты и фильтры по требуемым атрибутам. Пример диалогового окна создания фильтра, отсекающего системного пользователя, оставляющего только процессы, запущенные после 01.01.2009, и включающего только атрибуты Computer, Parent, Process, UserName приведен на рисунке 36

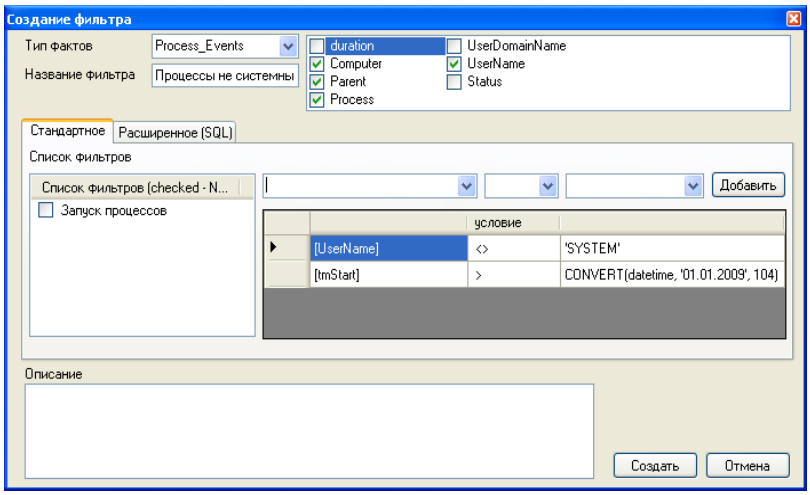

Рисунок 36 — Пример диалогового окна создания фильтра по выбранным атрибутам.

5. Нажать кнопку создать. Выбрать и открыть созданный фильтр в меню «Фильтры фактов». В результате будет создан и открыт требуемый фильтр фактов.

Проверка создания модели поведения на основе созданного ранее фильтра данных состоит из следующих шагов.

1.Запустить АРМ аналитика безопасности, если оно не запущено. В результате откроется АРМ аналитика безопасности.

2. В рамках витрины данных, где создавался фильтр, нажать правой кнопкой мыши на «Data Mining модели», в появившемся меню выбрать «Создать модель…». В открывшемся диалоговом окне указать требуемый фильтр фактов, имя и описание модели. Пример преобразованного диалогового окна после указания требуемых параметров приведен на рисунке 37.

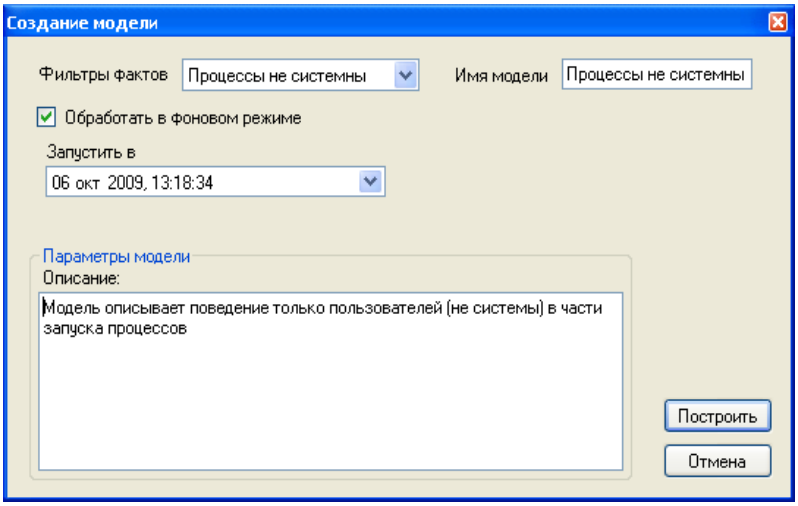

Рисунок 37 — Интерфейс окна создания модели.

3. Нажать кнопку построить. В результате задача построения модели добавляется в витрину данных. После построения модели, в вершине «Data Mining модели» появится построенная модель. Построение модели может потребовать некоторого времени.

4.Открыть созданную модель. В результате будет отображена модель в виде ассоциативных правил.

Проверка применение созданной модели для классификации аномальности действий пользователя с целью выявление действий пользователей или программ, которые могут предшествовать нарушению конфиденциальности, целостности или доступности служебной информации (раннее обнаружение внутренних вторжений) состоит из следующих шагов.

1.Запустить АРМ аналитика безопасности. В результате откроется АРМ аналитика безопасности.

2. Подготовить фильтр для классификации, если это требуется. Применить модель к подготовленному фильтру. Открыть результаты применения модели, произвести сортировку по степени аномальности по убывающему порядку. В результате применение модели

содержит факты активности, отсортированные в порядке убывания их аномальности (см. рисунок 38).

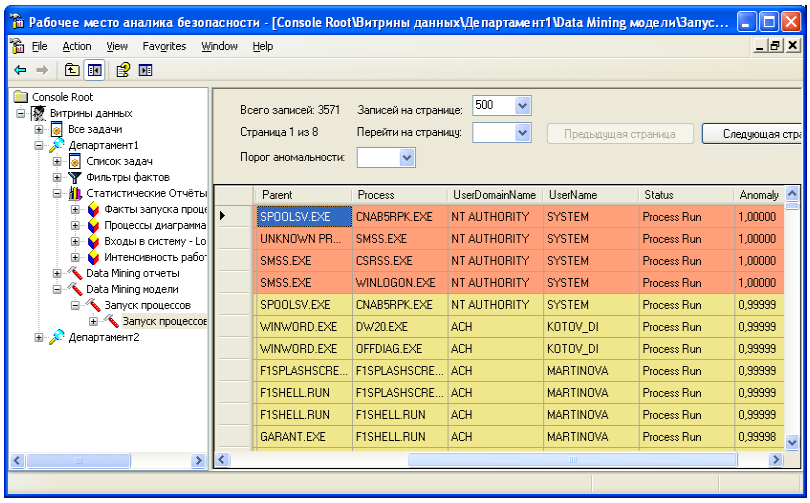

Рисунок 38 — Аномальность фактов активности пользователя.

3.Дважды щелкнуть мышью по интересующему факту. В результате откроется окно, детализирующее аномальность значения каждого из атрибутов факта, по отношению к значениям других атрибутов (см. рисунок 39).

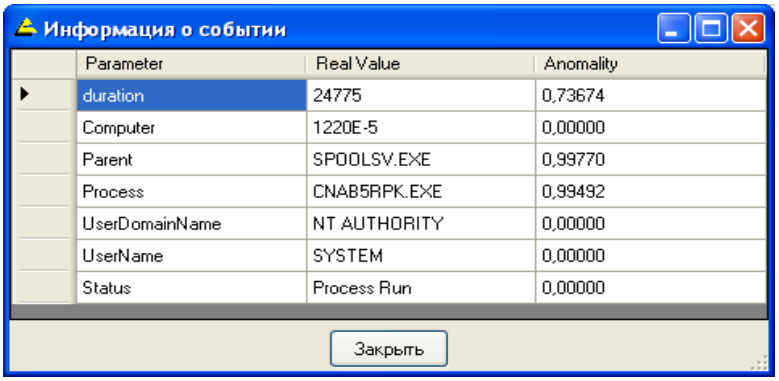

Рисунок 39 — Интерфейс окна с информацией о конкретном событии.

Проверка создания выборки состоит из следующих шагов.

1. Открыть окно создания выборки как показано на рисунке 40. В результате откроется окно создания выборки.

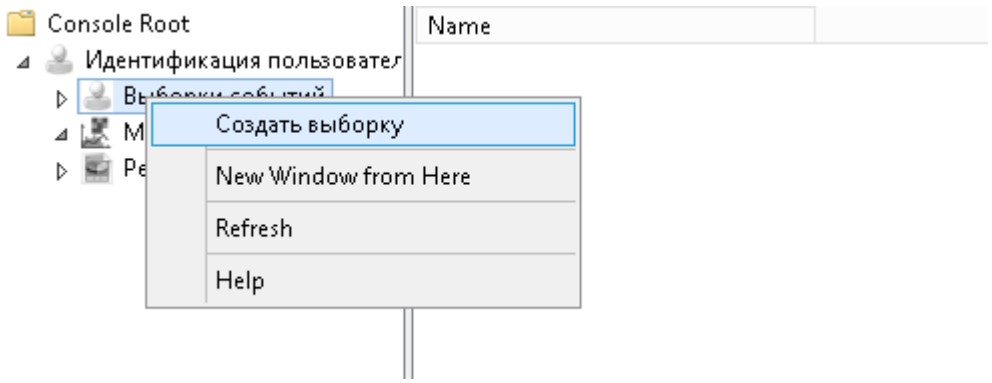

Рисунок 40 — Открытие окна создания выборки

2. Задать название выборки, установить типы событий Process_events и Active_events, не изменять атрибуты событий. В результате откроется интерфейс окна, приведенный на рисунке 41.

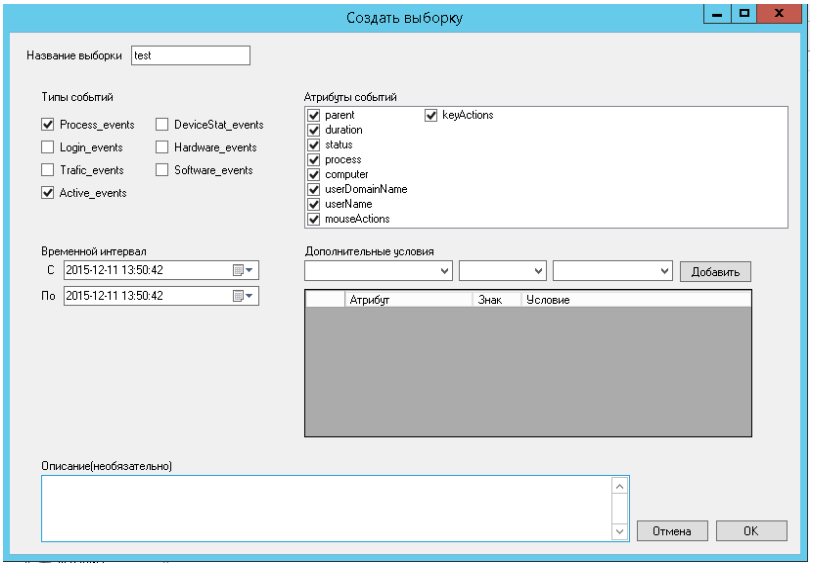

Рисунок 41 — Интерфейс окна создания выборки.

3. Задать временной интервал, в течение которого события гарантированно происходили (на настоящий момент это от 2015-03-22 до 2015-03-30). В результате откроется интерфейс окна, приведенный на рисунке 42.

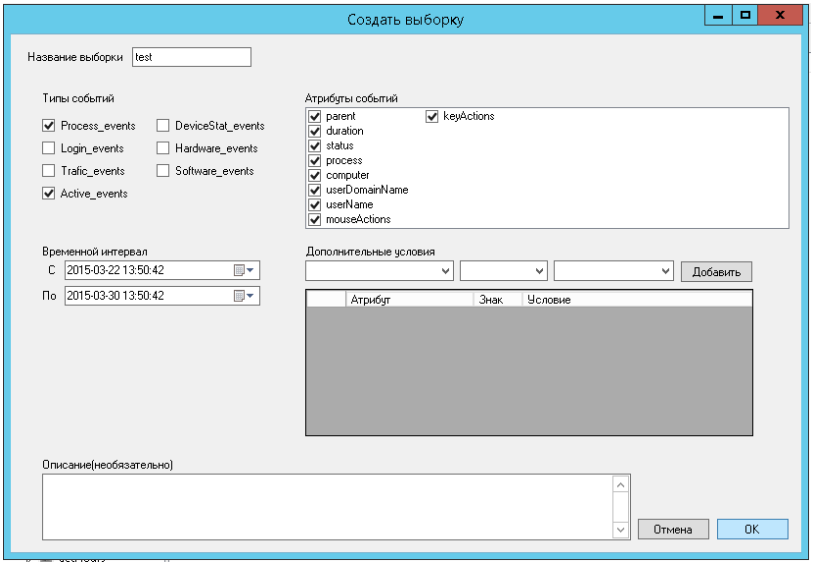

Рисунок 42 — Интерфейс окна создания выборки с заданным временным интервалом.

4. Нажать кнопку «ОК» и переключиться на созданный узел. Таблица событий появится автоматически. В результате откроется интерфейс окна, приведенный на рисунке 43.

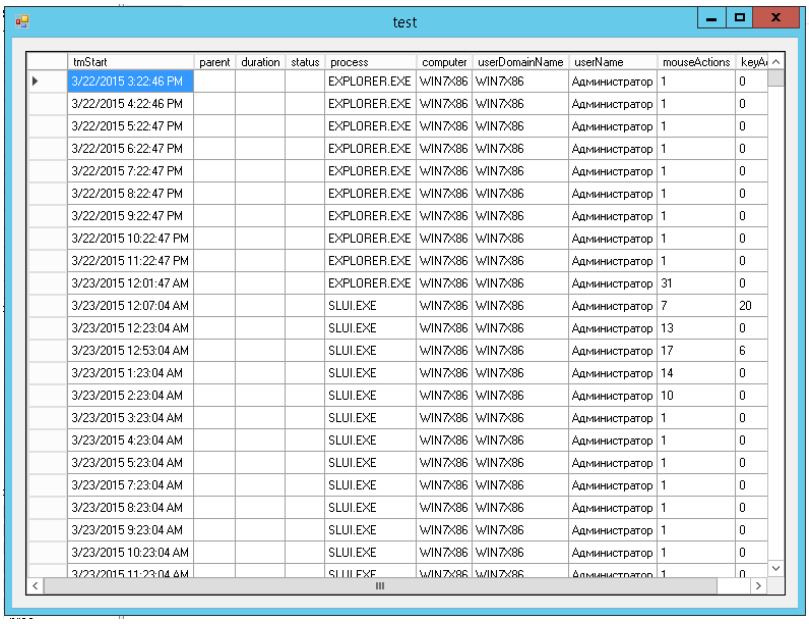

Рисунок 43 — Таблица событий для заданной выборки.

Проверка создания модели состоит из следующих шагов.

1. Открыть окно создания модели как показано на рисунке 44. В результате откроется интерфейс окна создания модели.

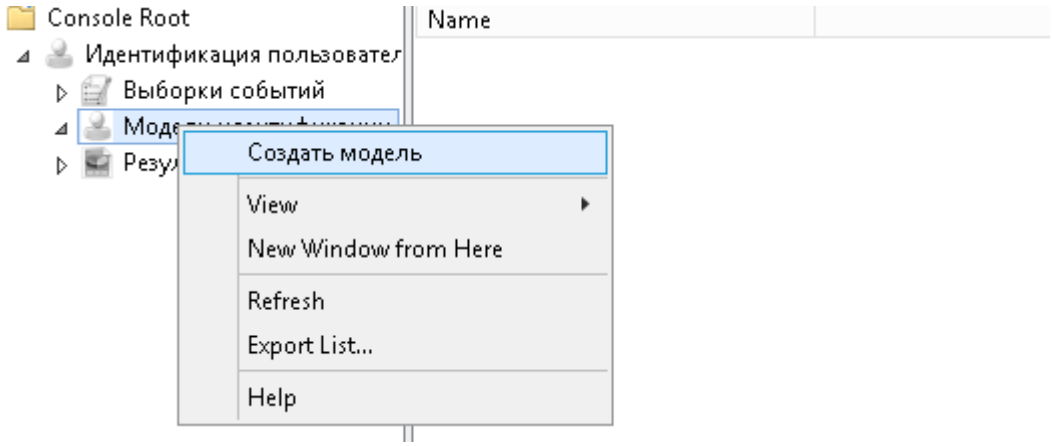

Рисунок 44 — Открытие окна создания модели

2. Задать название и выбрать выборку, созданную в предыдущем контрольном примере. В результате откроется интерфейс окна, приведенный на рисунке 45.

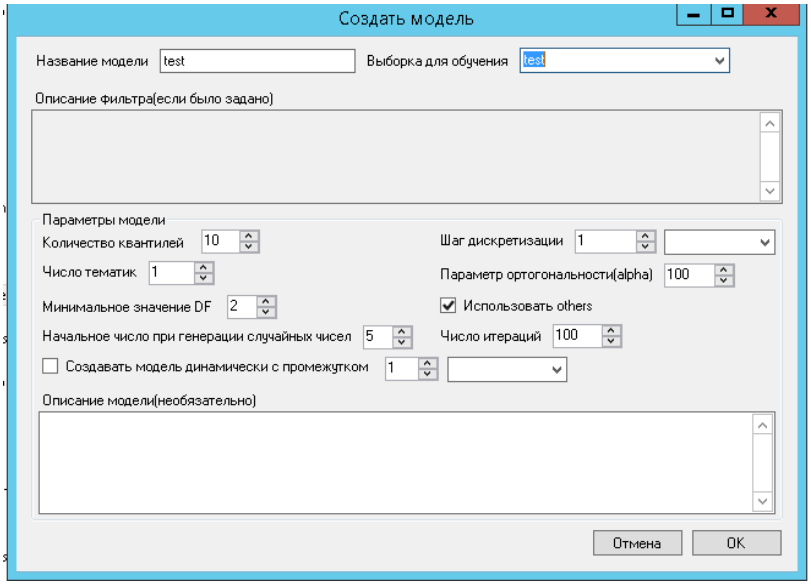

Рисунок 45 — Интерфейс окна создания модели.

3. Установить шаг дискретизации 1 день и число тематик 3. Остальные параметры оставить по умолчанию. В результате откроется интерфейс окна, приведенный на рисунке 46.

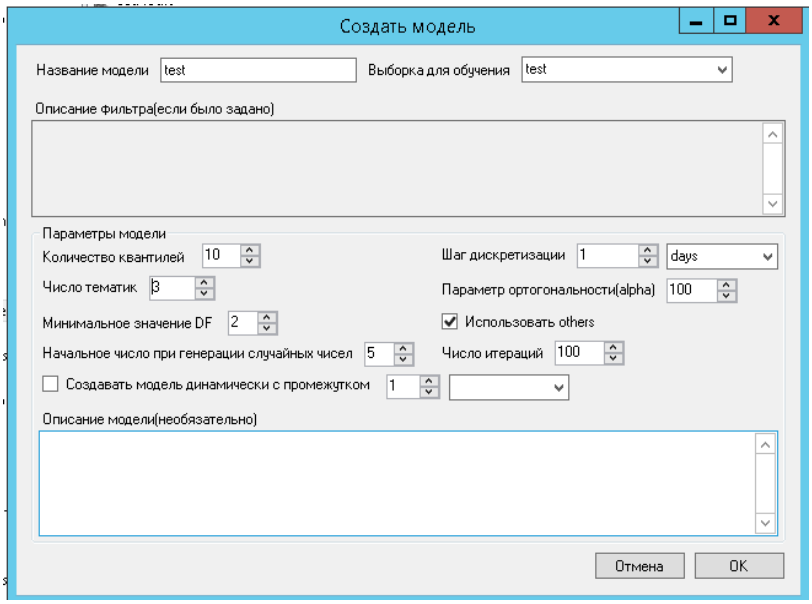

Рисунок 46 — Интерфейс окна создания с заданными параметрами.

4. Нажать кнопку «ОК» и переключиться на созданный узел. Таблица событий появится автоматически (см. рисунок 47).

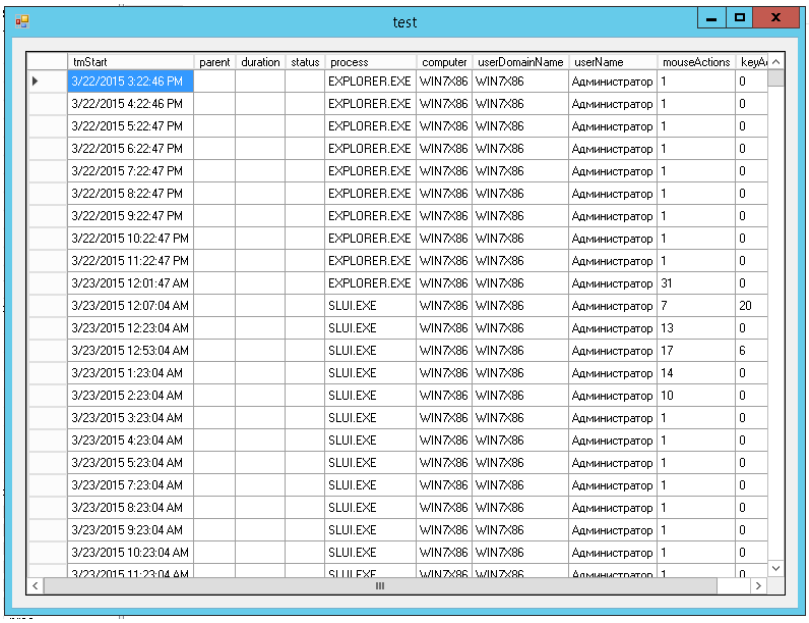

Рисунок 47 — Таблица событий для выбранного узла.

5.Сделать двойной клик на ячейке таблицы. Графики появятся в новом окне (см. рисунок 48).

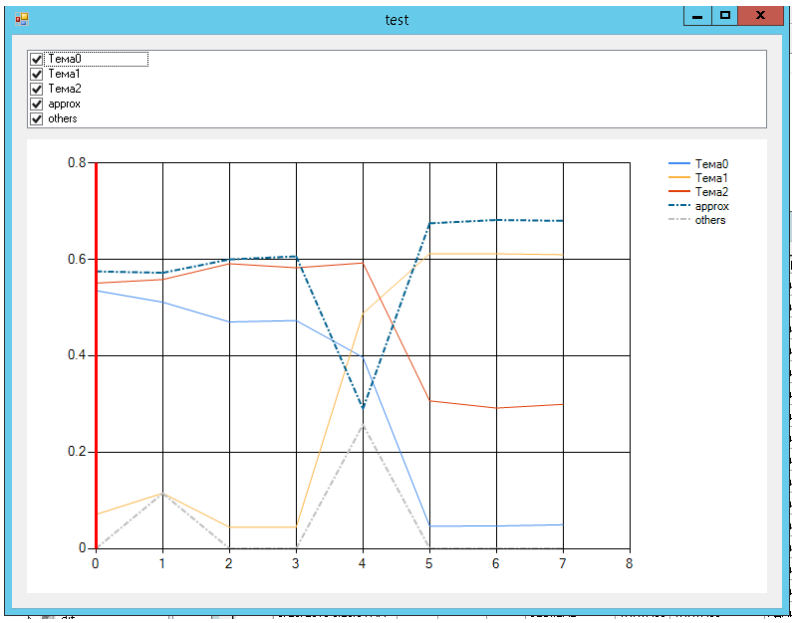

Рисунок 48 — Графики для выбранной ячейки таблицы событий.

Проверка создания отчета состоит из следующих шагов.

1.Создать выборку, для которой нужно применить отчет, аналогично контрольному примеру «Создание выборки». В результате создается выборка. Узел для созданной выборки появляется в дереве узлов в левой части окна, дочерний для «Выборки событий

2. Открыть окно создания отчета как показано на рисунке 49. В результате открывается окно создания отчета.

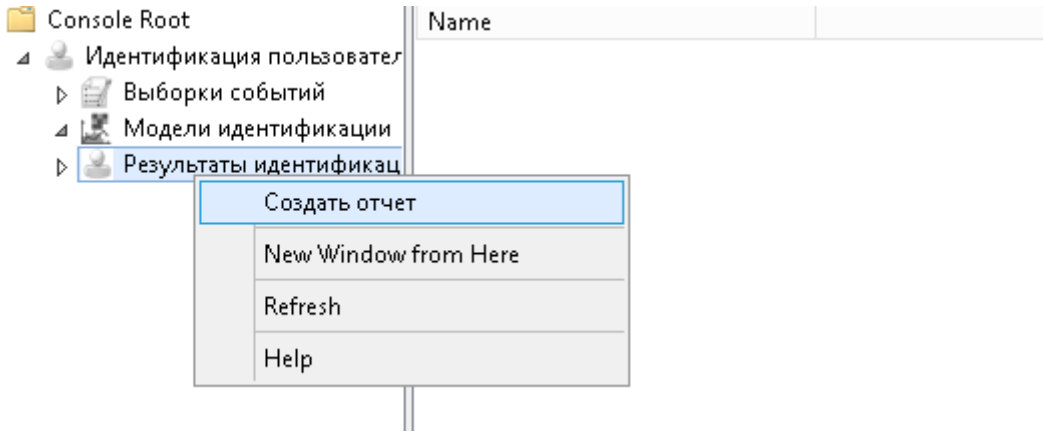

Рисунок 49 — Открытие окна создания отчета

3. Задать название отчета, выборку, созданную на шаге 1, и модель для применения, созданную в предыдущем контрольном примере. В результате откроется интерфейс окна, приведенный на рисунке 50.

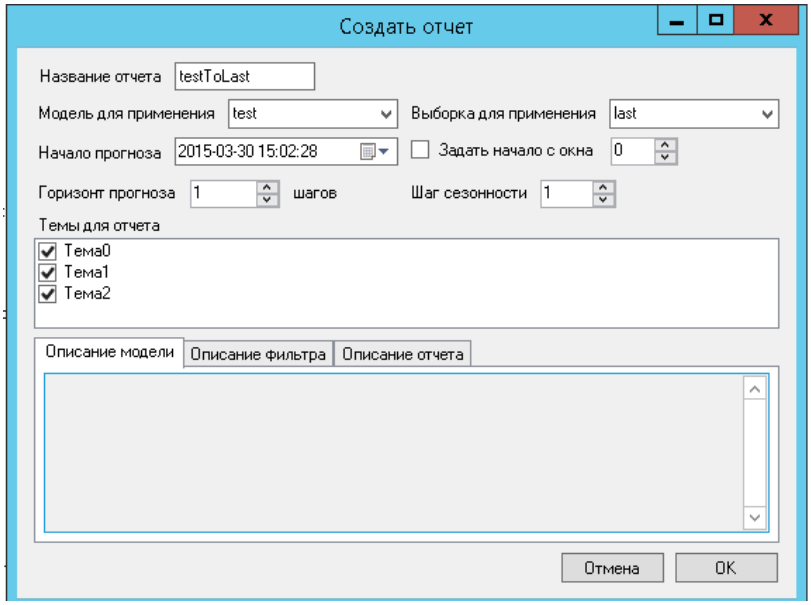

Рисунок 50 — Интерфейс окна создания отчета.

4. Установить «задать начало с окна 0» и горизонт прогноза до 5 шагов, остальные параметры оставить без изменения. Преобразованный интерфейс окна приведен на рисунке 51.

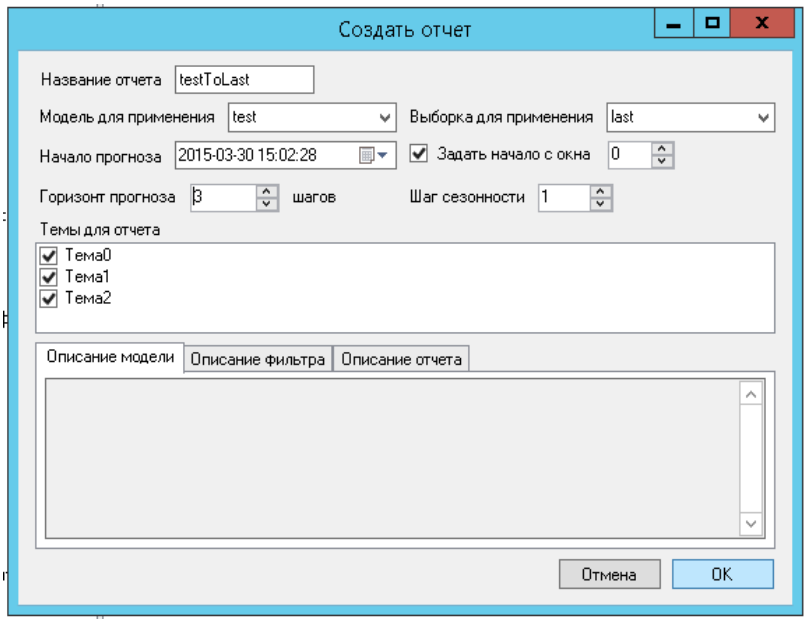

Рисунок 51 — Интерфейс окна создания отчета с выбранными параметрами.

5. Нажать кнопку «ОК» и переключиться на созданный узел. Таблица событий появится автоматически (см. рисунок 52).

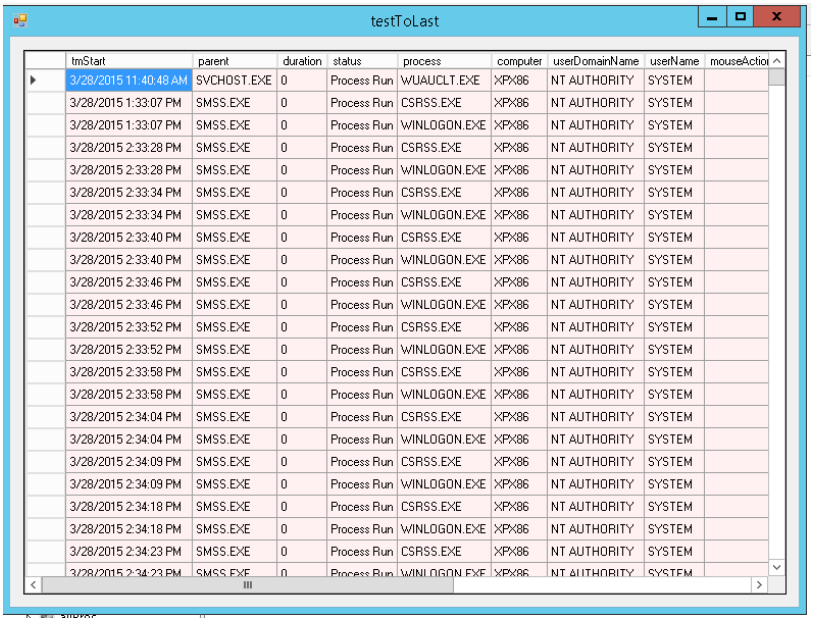

Рисунок 52 — Интерфейс таблицы событий.

6.Сделать двойной клик на ячейке таблицы. Графики появятся в новом окне (см. рисунок 53).

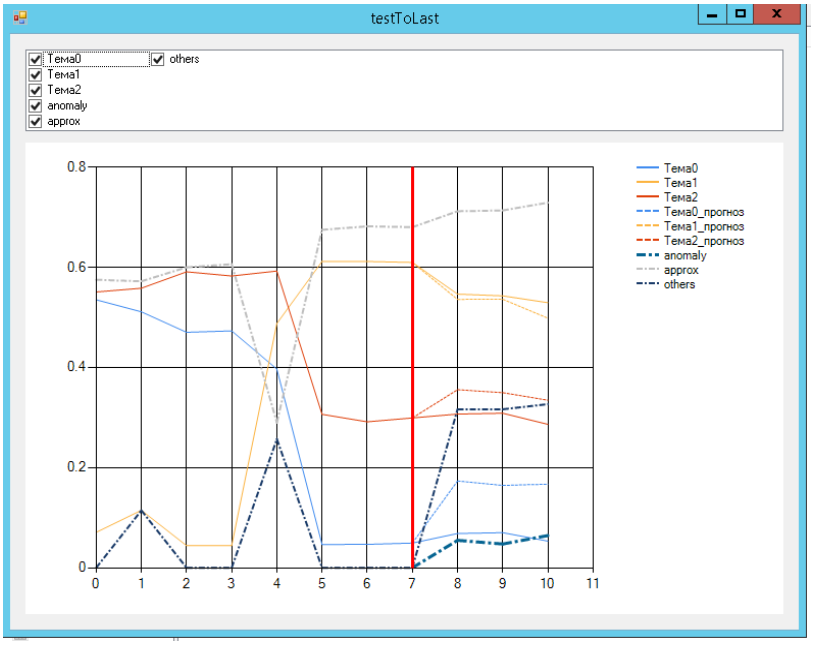

Рисунок 53 — Графики для выбранной ячейки таблицы событий.

Проверка управления моделями, отчетами и выборками состоит из следующих шагов.

1. Переключиться на узел и нажать меню обновить как показано на рисунке 54. В результате узел обновляется. В случае изменения связанной с этим узлом выборки или модели, изменения также применяются.

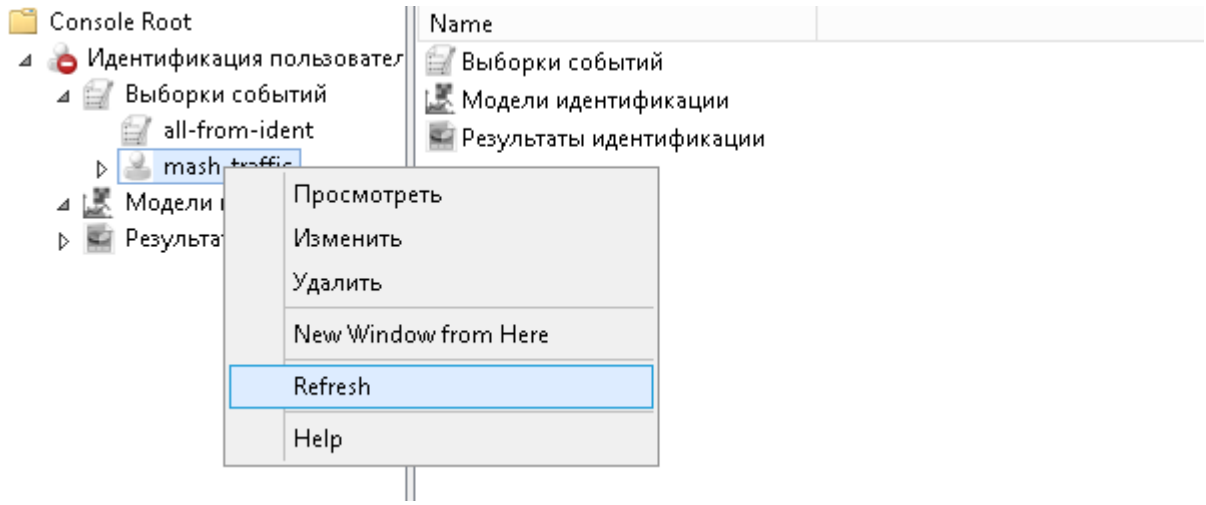

Рисунок 54 — Обновление модели, отчета, выборки

3. Переключиться на узел и нажать меню удалить как показано на рисунке 55. В результате узел удаляется и уже не отображается в дереве в левой части окна.

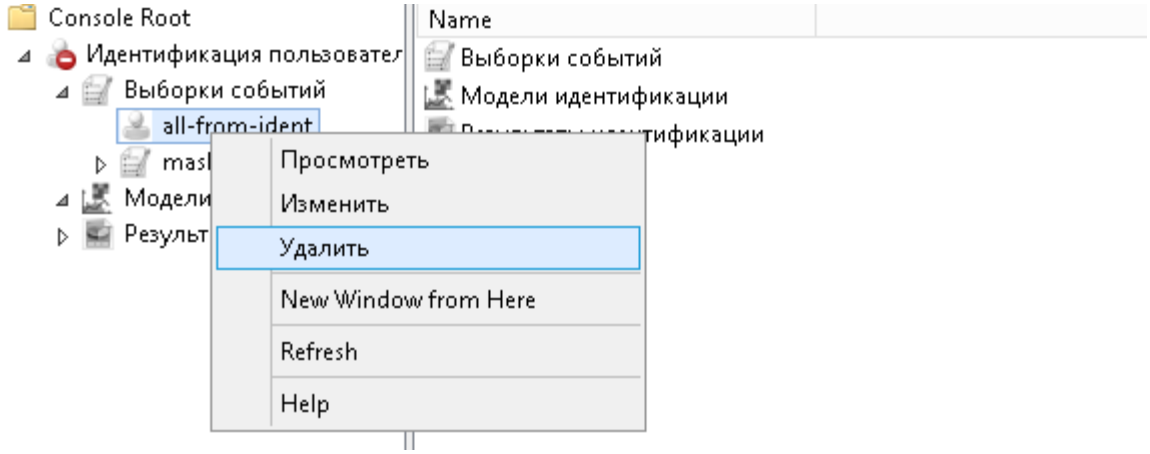

Рисунок 55 — Удаление модели, отчета, выборки

1.3.3 Оценка точности и скорости работы методов машинного обучения и математической статистики для построения и применения поведенческих моделей для анализа данных об особенностях работы пользователя с потоками текстовой информации

Данный пункт посвящён проведению экспериментальных исследований по пунктам 4.3.3.3 и 4.3.3.4 Программы экспериментальных исследований, представленной в отчёте за предыдущий этап настоящих ПНИ (Раздел 4 Программы и методик экспериментальных

исследований ЭО ПК), описания которых приводятся ниже в подпунктах 1.3.3.1 и 1.3.3.2 соответственно.

1.3.3.1 Оценка точности и скорости работы методов машинного обучения и математической статистики для построения и применения поведенческих моделей

Целью проводимых экспериментальных исследований является оценка точности и скорости работы методов машинного обучения и математической статистики для построения и применения поведенческих моделей «Подсистемы 3» ЭО ПК, которая служит для сбора и анализа информации об особенностях работы пользователя с потоками текстовой информации.

Функциональность «Подсистемы 3» включает решение задач идентификации пользователей и раннего обнаружения попыток хищения конфиденциальной информации, поэтому построение и применение поведенческих моделей реализовано только для перечисленных задач.

Согласно пункту 6.5.3 Программы и методик экспериментальных исследований ЭО ПК для метода идентификации пользователей и метода раннего обнаружения попыток хищения конфиденциальной информации процедура оценки точности и скорости работы построения и применения соответствующих поведенческих моделей следующая:

- загрузить экспериментальные данные в систему, используя процедуру сбора поведенческой информации с текстовыми данными;
- с помощью автоматизированного рабочего места (АРМ) аналитика инициировать процесс построения поведенческой модели, реализующий рассматриваемый метод, для выборки, сформированной из загруженных экспериментальных данных, и замерить время его работы;
- с помощью автоматизированного рабочего места (АРМ) аналитика инициировать процесс построения отчёта, заключающийся в применении соответствующей поведенческой модели к выборке, сформированной из загруженных экспериментальных данных, и замерить время его работы;
- выгрузить результирующие данные сформированного отчёта для расчёта оценки точности рассматриваемого метода. В качестве меры оценки моделей используется площадь под ROC кривой — AUC.

Выделим методы машинного обучения, которые используются для построения и

применения поведенческих моделей в «Подсистеме 3» ЭО ПК:

- согласно подпункту 1.5.2.1 («Разработка программных компонент, предназначенных для построения, управления и применения пользовательских поведенческих моделей для задачи идентификации пользователей на основе поведенческой биометрии работы пользователя с текстовыми данными») отчёта ПНИ за предыдущий этап для построения и применения пользовательских поведенческих моделей в задаче идентификации используются следующие методы машинного обучения: тематическое моделирование (построение модельного тематического пространства пользователя), отображение документов в модельное тематическое пространство и прогнозирование многомерного временного ряда;
- согласно подпункту 1.5.3.1 («Разработка программных компонент, предназначенных для построения, управления и применения пользовательских поведенческих моделей для задачи раннего обнаружения попыток хищения конфиденциальной информации на основе поведенческой биометрии работы пользователя с текстовыми данными») отчёта ПНИ за предыдущий этап для построения и применения пользовательских поведенческих моделей в задаче раннего обнаружения попыток хищения конфиденциальной информации используются следующие методы машинного обучения: тематическое моделирование (построение модельного тематического пространства пользователя) и отображение документов в модельное тематическое пространство.

Как следует из описания программного компонента «Подсистема 3» ЭО ПК, представленного в отчёте за предыдущий этап настоящих ПНИ, функционал построения и применения поведенческих моделей был выделен в отдельный DCOM-объект (англ. Distributed Component Object Model) анализа поведенческой информации. Соответственно в состав данного DCOM-объекта входят функции, реализующие: тематическое моделирование, отображение документов в модельное тематическое пространство и прогнозирование многомерного временного ряда. Для перечисленных функций проводились следующие эксперименты с использованием тестовых документов из наборов экспериментальных данных НТЭД1 (203227 электронных писем) и НТЭД2 (10228 текстовых документов), сформированных на втором этапе настоящих ПНИ:

1. *Тематическое моделирование и отображение документов в модельное тематическое пространство*. Эксперименты по оценки производительности данных функций проводились по следующему сценарию. Для каждого из 15 пользователей набора НТЭД2 были сформированы экспериментальные диапазоны, состоящие из документов для построения модели (тренировочный набор) и документов для применения модели

(тестовый набор). Среди полученных экспериментальных диапазонов количество документов для построения и применения моделей существенно не изменялось (тренировочный набор: среднее число документов — 200; тестовый набор: среднее число документов — 450), поэтому для оценки времени работы соответствующих функций *DCOM-объекта* рассчитывалось их среднее время выполнения.

2. *Прогнозирование многомерного временного ряда*. Для оценки производительности данной функции били также сформированы экспериментальные диапазоны для пользователей из набора НТЭД1. После чего была проведена серия экспериментов, которая заключалась в построении прогноза по суточным данным за месяц на неделю вперёд (7 шагов). Таким образом, среднее число точек, по которым строился прогноз, равнялось 30, число тематик было выбрано равным 3. Т.к. в сформированных временных рядах число точек было примерно одинаково, то для оценки времени работы функции прогнозирования рассчитывалось среднее время её выполнения.

Для проведения описанных тестов использовался *вычислительный сервер (Подсистема2/Подсистема3)* (см. Приложение А Программы и методик экспериментальных исследований ЭО ПК):

- программное обеспечение: ОС Microsoft Windows 2008R2 (64bit); Microsoft Office 2007; Microsoft SQL Server 2012; Microsoft SQL Server 2005; Microsoft SQL Server 2005 Analysis Services; Microsoft SQL Server 2005 Management Studio; интерпретатор Python 2.7.3; модуль Python pywin 219; Библиотека Python Win32 Extensions; Библиотека Natural Language Toolkit (NLTK);
- аппаратные характеристики: два процессора x64 с тактовой частотой 2 Ггц, 4 ядра; оперативная память объемом 48 Гбайт; два дисковый накопителя HDD, объемом 2Тбайт; два сетевых адаптера с пропускной способностью 1 Гбит.

При каждом эксперименте замерялось время работы соответствующих функций *DCOM-объекта*, а также рассчитывалось значение AUC для задач идентификации пользователей и раннего обнаружения попыток хищения конфиденциальной информации (см. Раздел 1 «Разработка методов машинного обучения и математической статистики для построения и применения поведенческих моделей на основе поведенческой биометрии работы пользователя с текстовыми данными» отчёта за второй этап настоящих ПНИ). Полученные показатели оценок точности и скорости работы методов машинного обучения и математической статистики для построения и применения поведенческих моделей приведены в протоколе №1 по пункту № 4.3.3.3 Программы экспериментальных исследований «Подсистемы 3» ЭО ПК для сбора и анализа информации об особенностях работы

пользователя с потоками текстовой информации (см. Приложение А: таблица 15 Протокола испытаний №15).

Из приведённых данных можно сделать вывод, что процессы построения и применения поведенческих моделей с помощью функций реализованного DCOM-объекта анализа поведенческой информации не занимают длительное время. Наиболее критичным с точки зрения затрачиваемых вычислительных ресурсов является процесс применения поведенческих моделей, т.к. он может выполнятся на агенте мониторинга в режиме «условно-реального» времени (см. пункт 2.1.2 «Применение поведенческих моделей» из описания программного компонента «Подсистема 3» ЭО ПК, представленного в отчёте за предыдущий этап настоящих ПНИ). Соответственно, применение поведенческих моделей может влиять на производительность наблюдаемых машин пользователей, в отличие от построения поведенческих моделей, которое выполняется на выделенных машинах. Представленные данные в протоколе №1 по пункту № 4.3.3.3 Программы экспериментальных исследований (см. Приложение А: таблица 15 Протокола испытаний №15) в части функции отображения документов в модельное тематическое пространство, которая является основной при применении поведенческих моделей, показывают, что на обработку одного документа тратится порядка 0.01 секунды. Отметим, что при обработке сразу нескольких (коллекции) документов среднее время анализа одного документа существенно сокращается. Таким образом, как и в случае мониторинга поведенческой информации (см. пункт 1.1.3), если пользователь не выполняет частые операции, требующие от агента создания теневых копий, с большим числом документов, то он даже не заметит каких-либо изменений в характеристиках работы наблюдаемого компьютера.

1.3.3.2 Оценка реализации функциональности «Подсистемы 3» ЭО ПК

Целью проводимых экспериментальных исследований является оценка реализации функциональности «Подсистемы 3» ЭО ПК, которая служит для сбора и анализа информации об особенностях работы пользователя с потоками текстовой информации. «Подсистема 3» ЭО ПК считается выдержавшим проверку, если полученные результаты соответствуют ожидаемым результатам контрольных примеров №№ 1-2 Приложение Б Программы и методик экспериментальных исследований ЭО ПК и/или отражены в комплекте документации.

Проверка построения и применения модели для задачи идентификации пользователей. Контрольный пример №1.

1. Запустить АРМ аналитика «Подсистемы 3». В результате откроется АРМ аналитика в виде MMC консоли, которая показана на рисунке 2.
2. Выполнить подключение к единому хранилищу модуля консолидации. Главное окно консоли состоит из 3 частей (см. рисунок 2). Справа отображается меню для выбранного корневого узла («Анализ текстовых потоков»). В данном меню есть раздел «Подключиться к хранилищу», который позволяет подключиться к базе данных единого хранилища, предварительно открыв окно для ввода параметров подключения (см. рисунок 3).
3. Создать выборку документов для последующего построения модели. При активации меню «Создать выборку» узла «Выборки документов», находящегося в левой части меню главного окна консоли АРМ (см. рисунок 2), откроется форма создания выборок, представленная на рисунке 5.
4. Просмотреть выборку. При активации меню «Просмотреть» для выборки происходит вывод событий данной выборки в отдельной форме в виде таблицы (см. рисунок 6).
5. Построить модель идентификации по созданной выборки. Открытие формы интерфейса построения моделей идентификации (см. рисунок 56) происходит при выборе пункта меню «Создать модель» в разделе «Модели идентификации», который входит в узел «Идентификация», находящийся в левой части меню главного окна консоли АРМ (см. рисунок 2).

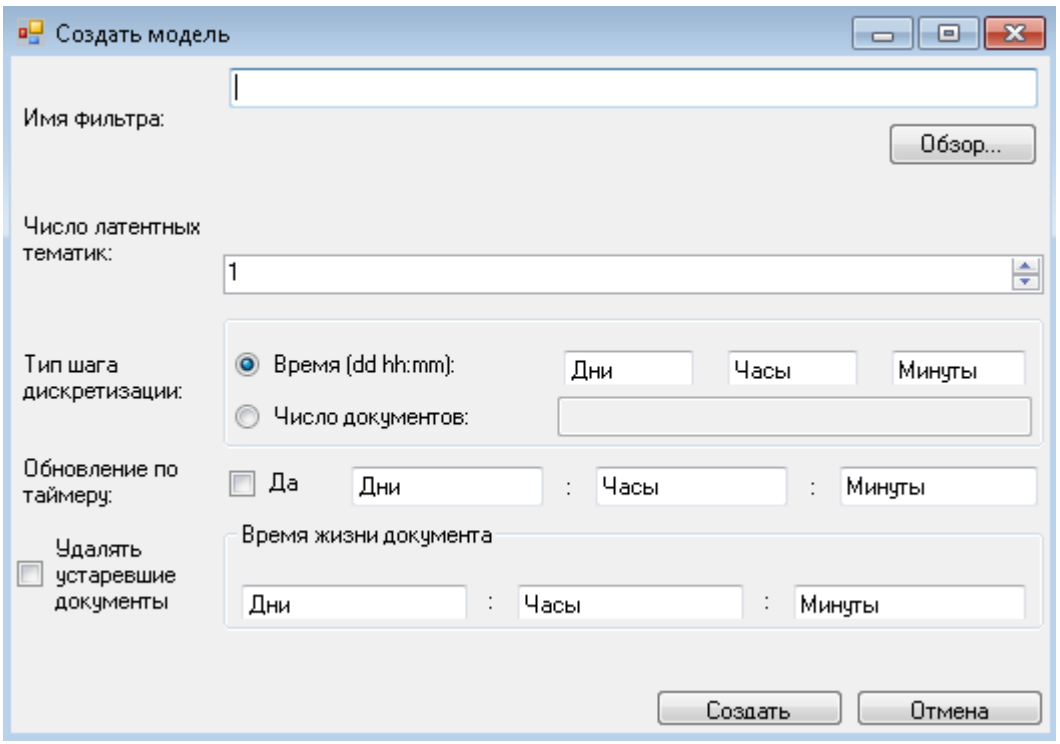

Рисунок 56 — Окно формы построения модели идентификации.

6. Построить отчет по созданной модели идентификации (применить модель идентификации). Форма интерфейса построения отчетов по применению модели идентификации (см. рисунок 57) открывается при активации пункта меню «Создать отчет» узла «Отчеты идентификации», а также при выборе пункта меню «Создать отчет» для модели, по которой нужно будет построить отчет.

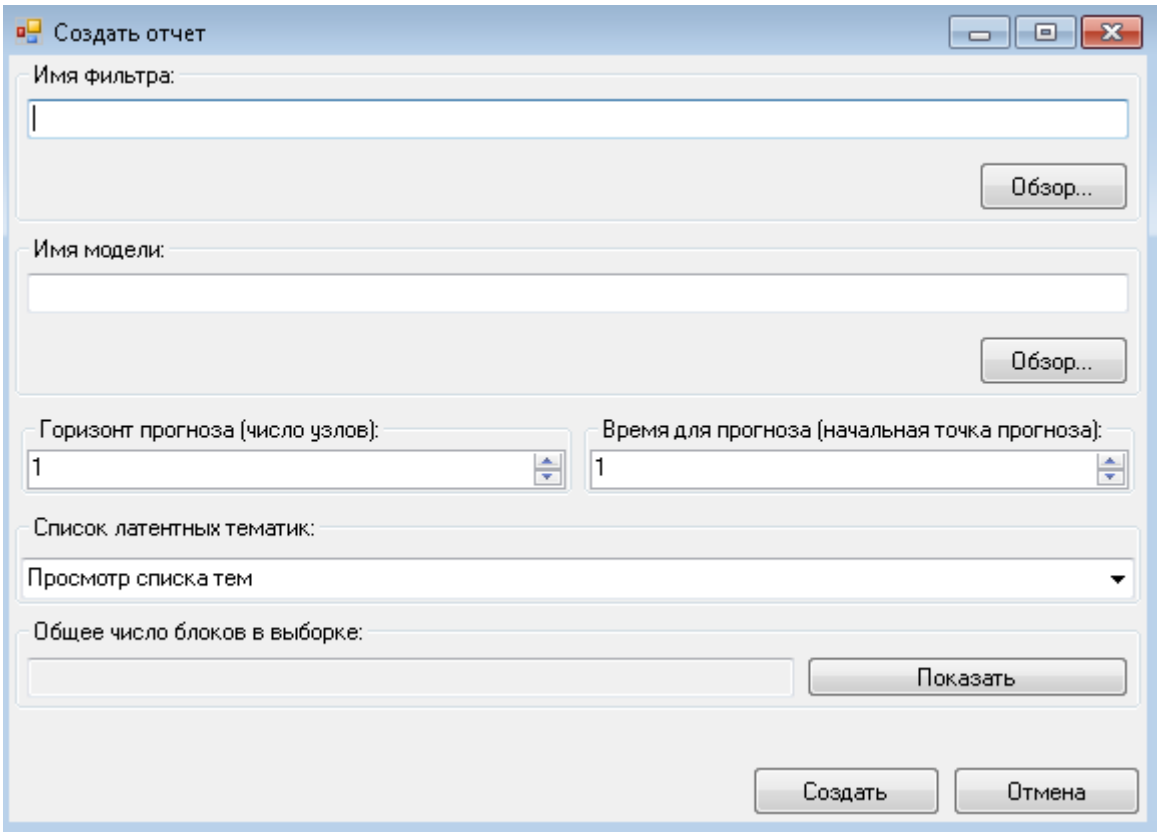

Рисунок 57 — Окно формы построения отчетов по применению модели идентификации.

7. Просмотреть отчёт идентификации. Форма отчёта идентификации представляет таблицу событий поведенческой информации, которая открывается с помощью меню «Просмотреть» для выбранного отчёта. В случае отчета по идентификации строки таблицы, по которым не был вычислен прогноз, окрашиваются в бежевый цвет. Строки же, для которых строился прогноз, окрашиваются в оттенок красного, соответствующий среднему арифметическому для модулей отклонений реальных весов тематик от спрогнозированных значений (чем это значение выше, тем темнее будет цвет строки таблицы). Пример отчёта идентификации представлен на рисунке 58.

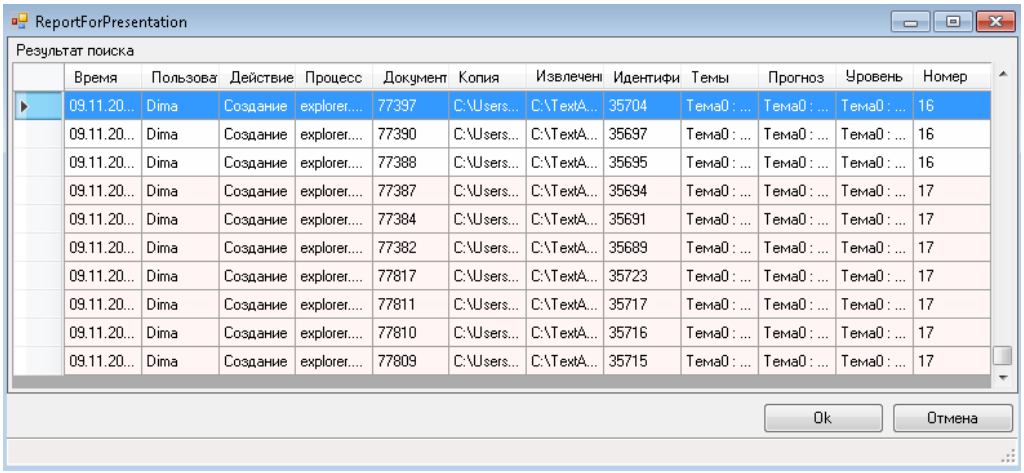

Рисунок 58 — Окно формы просмотра отчёта идентификации.

Проверка построения и применения модели для задачи раннего обнаружения попыток хищения конфиденциальной информации. Контрольный пример №2.

1. Запустить АРМ аналитика «Подсистемы 3». В результате откроется АРМ аналитика в виде MMC консоли, которая показана на рисунке 2.
2. Выполнить подключение к единому хранилищу модуля консолидации. Главное окно консоли состоит из 3 частей (см. рисунок 2). Справа отображается меню для выбранного корневого узла («Анализ текстовых потоков»). В данном меню есть раздел «Подключиться к хранилищу», который позволяет подключиться к базе данных единого хранилища, предварительно открыв окно для ввода параметров подключения (см. рисунок 3).
3. Создать выборку документов для последующего построения модели. При активации меню «Создать выборку» узла «Выборки документов», находящегося в левой части меню главного окна консоли АРМ (см. рисунок 2), откроется форма создания выборок, представленная на рисунке 5.
4. Просмотреть выборку. При активации меню «Просмотреть» для выборки происходит вывод событий данной выборки в отдельной форме в виде таблицы (см. рисунок 6).
5. Построить модель раннего обнаружения попыток хищения конфиденциальной информации по созданной выборки. Открытие формы интерфейса построения моделей раннего обнаружения (см. рисунок 59) происходит при выборе пункта меню «Создать модель» в разделе «Модели раннего обнаружения», который входит в узел «Раннее обнаружение», находящийся в левой части меню главного окна консоли АРМ (см. рисунок 2).

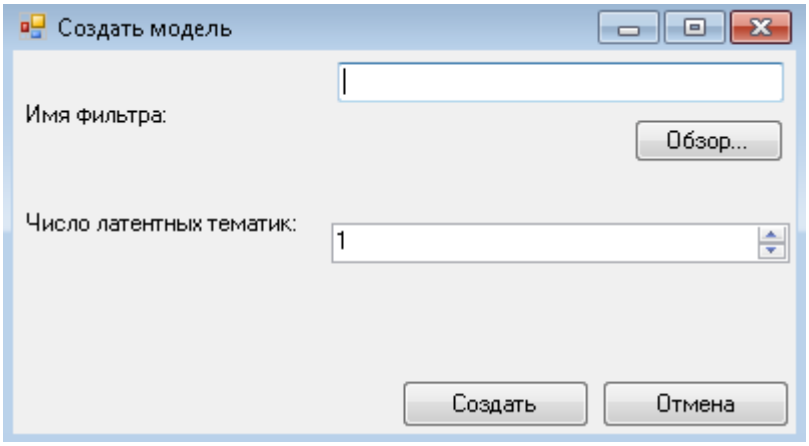

Рисунок 59 — Окно формы построения модели раннего обнаружения.

6. Построить отчет по созданной модели раннего обнаружения попыток хищения конфиденциальной информации (применить модель раннего обнаружения попыток хищения конфиденциальной информации). Форма интерфейса построения отчетов по применению модели раннего обнаружения (см. рисунок 60) открывается при активации пункта меню «Создать отчет» узла «Отчеты раннего обнаружения», а также при выборе пункта меню «Создать отчет» для модели, по которой нужно будет построить отчет.

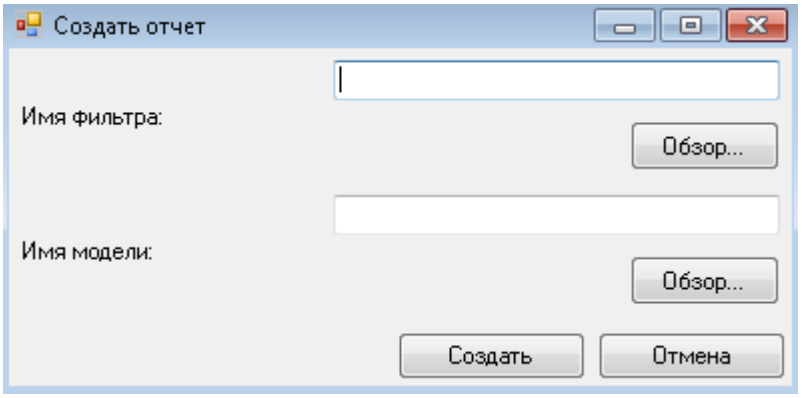

Рисунок 60 — Окно формы построения отчетов по применению модели раннего обнаружения.

7. Просмотреть отчёт раннего обнаружения попыток хищения конфиденциальной информации. Форма отчёта раннего обнаружения представляет таблицу событий поведенческой информации, которая открывается с помощью меню «Просмотреть» для выбранного отчёта. В случае отчета по раннему обнаружению, строки окрашиваются в оттенок красного, соответствующий уровню аномальности события (см. рисунок 61).

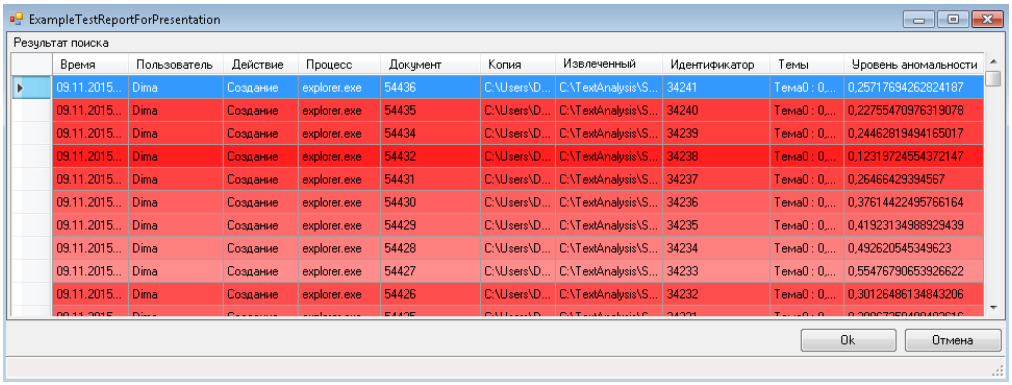

Рисунок 61 — Окно формы просмотра отчёта раннего обнаружения.

1.4 Выводы

На основе проведенных экспериментальных исследований можно сделать следующие выводы.

Лучшие результаты по точности фоновой идентификации пользователей на основе динамики работы с клавиатурой показал нечеткий метод поиска исключений на основе потенциальных функций: точность классификации (AUC) – 86.60%. Оценка скорости сбора показала "незаметность" работы компонент сбора для пользователя, при этом проверка статуса активности перехватчика показала, что перехватчик успевает обработать все выполненные события за отведённое время. Время, затрачиваемое на распознавание (применение модели) для каждого окна событий пользователя, не превышает секунды для всех рассматриваемых методов, что показывает их применимость для решения исследуемой прикладной задачи.

Лучшие результаты по точности фоновой идентификации пользователей на основе динамики работы с мышью показал метод опорных векторов: точность классификации (AUC) – 68.7%. Оценка скорости сбора показала "незаметность" работы компонент сбора для пользователя, при этом проверка статуса активности перехватчика показала, что перехватчик успевает обработать все выполненные события за отведённое время. Время, затрачиваемое на распознавание (применение модели) для каждого окна событий пользователя, не превышает секунды для всех рассматриваемых методов, что показывает их применимость для решения исследуемой прикладной задачи.

Лучшие результаты по точности активной аутентификации пользователей на основе динамики работы с клавиатурой показали точность классификации (AUC) 90.2%, с мышью – 84%. Оценка скорости сбора показала "незаметность" работы компонент сбора для

пользователя. Время, затрачиваемое на распознавание (применение модели) для каждого ввода, не превышает долей секунды, что показывает их применимость для решения исследуемой прикладной задачи.

Лучшие результаты по точности поведенческих моделей раннего обнаружения вторжений для анализа данных об особенностях работы пользователя с информационными и вычислительными ресурсами компьютерной системы показали точность классификации (AUC) 88%. Лучшие результаты по точности поведенческих моделей идентификации для анализа данных об особенностях работы пользователя с информационными и вычислительными ресурсами компьютерной системы показали точность классификации (AUC) 94%.

Наилучшие результаты показали разработанные методы машинного обучения и математической статистики для построения и применения поведенческих моделей с целью решения задач постоянной фоновой идентификации пользователей и раннего обнаружения попыток хищения конфиденциальной информации на основе поведенческой биометрии работы пользователя с текстовыми данными. В задаче идентификации пользователей точность предложенных методов в среднем составляет 88% (AUC), для задачи раннего обнаружения попыток хищения конфиденциальной информации — 90% (AUC).

Выполненные работы полностью соответствуют требованиям п. 3.13 ТЗ, а также Программе и методикам экспериментальных исследований ЭО ПК. Продемонстрировано соответствие результатов теоретических исследований требованиям настоящего технического задания, в частности, пунктам 3.13.1, 3.13.2, 3.13.3, 4.1.1, 4.1.2.

2 Обеспечение работоспособности рабочих станций, общего системного и специального программного обеспечения для проведения работ по проекту, проведение регламентных работ по обеспечению сохранности данных и программ, относящихся к проводимым работам по проекту

2.1 Обеспечение работоспособности рабочих станций

Регламентное обслуживание обеспечивает необходимый набор услуг для поддержания оборудования и программного обеспечения в работоспособном и актуальном состоянии в процессе развития и эксплуатации инфраструктуры, включая обеспечение сохранности программ и данных, относящихся к проводимым работам по проекту.

В рамках работ по обеспечению работоспособности серверов и рабочих станций проводятся следующие процедуры.

1. Обновление программного обеспечения (установка пакетов обновления поставляемых производителями программного обеспечения такими как service-pack или другие обновления). Периодичность работ — автоматические обновления там, где это возможно, ручная проверка обновлений и систем работоспособности систем автоматических обновлений — 1 раз в месяц.
2. Поддержка антивирусной защиты. Периодичность работ по проверке корректного функционирования — 1 раз в месяц. Обновление лицензий антивирусного ПО, периодичность – 1 раз в год.
3. Восстановление работоспособности сервера/рабочей станции после сбоя. Включает в себя повторную инсталляцию и настройку программного обеспечения, восстановление данных.
4. Профилактические работы (чистка системных блоков, блоков питания, кулеров и их замена при необходимости. Периодичность работ — каждые 3 месяца для рабочих станций, 1 раз в месяц — для серверов.
5. Установка дополнительного программного обеспечения. Периодичность — по мере необходимости.
6. Проверка состояния жестких дисков (согласно SMART информации) в рабочих станциях и RAID-массивах серверов. При необходимости - замена жестких с дисков с

восстановлением данных/консистентности RAID-массивов. Периодичность — 1 раз в месяц.

7. Диагностика прочих физических неисправностей (включая периферийные устройства и соответствующий ремонт). Периодичность — по мере необходимости.

В рамках работ по обеспечению сохранности программ и данных, относящихся к проводимым работам по проекту, используются стратегии копирования и архивирования оперативных данных проекта в составе разрабатываемых программных средств/документов и отчетных материалов/экспериментальных данных на базе следующих возможностей.

1. Физический уровень — настройка критических данных, хранимых на серверах, в режиме RAID1 ("зеркалирование") — массив из двух дисков, являющихся полными копиями друг друга. Периодичность работ по проверке консистентности RAID1-массива — 1 раз в месяц.
2. Архивирование и восстановление данных на уровне файловой системы серверов — реализуется инкрементным копированием по расписанию средствами Windows Server Backup, встроенными в ОС семейства Windows Server. Дополнительно, используется резервное копирование и восстановление среды системы управления проектом на базе SharePoint: используются встроенные средства Microsoft SharePoint для защиты критических объектов среды (веб-приложение, сайт, база данных контента, библиотека документов, настройки, параметры конфигурации). Также, на сервере настроено автоматическое создание резервных копий баз данных системы контроля версий с целью обеспечения сохранности данных в случае программных либо аппаратных сбоев. Периодичность работ по инкрементальному архивированию критических данных проекта — ежедневно в автоматическом режиме (на базе возможностей соответствующего ПО). Периодичность работ по сохранению состояния среды — еженедельно в автоматическом режиме.
3. Восстановление работоспособности рабочих станций на уровне файловой системы — реализуется на базе интегрированного в семейство ОС Microsoft Workstation средства System Restore (точки восстановления системы). Периодичность работ по созданию точек восстановления: еженедельно в автоматическом режиме, каждый раз после установки/обновления нового ПО.

Обеспечение бесперебойного электропитания. Для уменьшения вероятности потери/повреждения критических данных все сервера и рабочие станции подключены к источникам бесперебойного питания (ИБП) с возможностью автономной работы не менее 15

минут, необходимых для автоматического корректного выключения системы. Периодичность проверки состояния ИБП (и замены батарей в случае необходимости) — каждые 6 месяцев.

2.2 Выводы

Настоящие работы проводились за счет средств внебюджетного финансирования с целью обеспечения работоспособности рабочих станций, общего системного и специального программного обеспечения для проведения работ по проекту, проведение регламентных работ по обеспечению сохранности данных и программ, относящихся к проводимым работам по проекту и в соответствии с пунктом 4.2 содержания выполняемых работ и мероприятий плана-графика исполнения обязательств при выполнении настоящих ПНИ.

Состав и факт исполнения обязательств по выполнению настоящих работ подтверждены Актом №3 исполнения обязательств по работам на этапе №4 Плана-графика, выполненных за счет внебюджетных средств (Приложение Б).

3 Проведение оценки РИД, полученных при выполнении ПНИ, с целью их вовлечения в хозяйственный оборот

3.1 Зарегистрированные и поданные на регистрацию РИД

Согласно требованиям ТЗ 4.1.2.11 в рамках ПНИ был разработан ЭО ПК предназначенный для обеспечения компьютерной безопасности на основе анализа поведенческой информации работы пользователей компьютерных систем, который имеет следующую логическую структуру (подробное ЭО ПК представлено в отчете за этап 3 настоящего ПНИ):

- модуль сбора, предобработки и классификации поведенческой биометрической информации;
- модуль консолидации поведенческой информации, предназначенный для хранения и управления поведенческой биометрической информацией, включая хранение теневых копий текстовых данных, журналов работы пользователей с вычислительными и информационными ресурсами защищаемой компьютерной системы, а также с устройствами ввода-вывода;
- Модуль построения поведенческих моделей;
- Модуль идентификации;
- Консоль управления.

Программные модули разработанного ЭО ПК были оформлены в виде РИД.

Имеется Свидетельство о регистрации программы для ЭВМ (№ 2015661555) от 30 октября 2015 г. «Система двухфакторной аутентификации на основе анализа поведенческой биометрической информации об особенностях работы пользователя с клавиатурой компьютера». Данная программа является одним из ключевых компонентов Модуля идентификации ЭО ПК, которая обеспечивает решение задачи аутентификации пользователя по динамике ввода предопределенной текстовой строки (пароля) и по совпадению введенного текста с паролем.

Подготовлена и сдана 24 июня 2016 года заявка (№2016617153) Федеральный институт промышленной собственности (ФИПС) о регистрации программы для ЭВМ «Система двухфакторной аутентификации на основе анализа поведенческой биометрической информации об особенностях работы пользователя с компьютерной мышью». Данная программа является одним из ключевых компонентов Модуля идентификации ЭО ПК, которая

обеспечивает решение задачи аутентификации пользователя на основе сопоставления характеристик динамики обведения шаблона мышью с индивидуальной моделью компьютерной биометрии.

Программная реализация средств аутентификации пользователя по динамике его работы с клавиатурой компьютера и компьютерной мышью является РИД, который обеспечивает потенциальную перспективу использования результатов ПНИ в хозяйственный оборот.

Подготовлена и сдана 24 июня 2016 года заявка (№2016617152) в Федеральный институт промышленной собственности (ФИПС) о регистрации программы для ЭВМ «Система мониторинга, теневого копирования и автоматического аннотирования текстовых данных при работе пользователя с электронными документами». Данный программный комплекс предназначен для использования в системах раннего обнаружения внутренних вторжений на основе выявления аномальности в компьютерной поведенческой информации, основанной на характеристиках работы пользователя с контентом документов за счет использования автоматически построенных аннотаций текстового контента при моделировании поведения пользователя.

3.2 Перспективы практического использования

Согласно пункту 4.3 плана-графика исполнения обязательств при выполнении прикладных научных исследований была проведена оценка РИД, полученных при выполнении настоящих ПНИ, с целью их вовлечения в хозяйственный оборот. Для этой цели был проведен анализ характеристик рынка систем информационной безопасности, которые могут потенциально использовать результаты полученных РИД.

Полученные в результате выполнения ПНИ РИД могут служить основой для построения современных программных систем, функционал которых направлен на решение задач активной аутентификации с использованием данных о динамике работы пользователя с клавиатурой и мышью, а также на управления контентной информацией организации, и на анализ работы пользователей с текстовой информацией.

Реализованный функционал программы для ЭВМ «Система мониторинга, теневого копирования и автоматического аннотирования текстовых данных при работе пользователя с электронными документами» включает сбор поведенческой информации пользователей при работе с текстовыми данными. Под поведенческой информацией пользователя понимаются

данные об операциях, выполняемых пользователем с электронными документами, и содержимое соответствующих электронных документов.

Для каждого типа электронного документа и среды его функционирования определён свой набор операций, изменяющих его состояние. Например, для текстовых файлов на рабочем месте пользователя это операции: создание, чтение, изменение, перемещение, удаление; для почтовых сообщений, получаемых и отправляемых пользователем с помощью почтового клиента — получение сообщения и отправка сообщения. Изменения электронного документа могут быть двух типов:

– *контентное* — изменение содержимого документа (например, редактирование содержимого текстового файла);
– *контекстное* — изменение атрибутов документа (например, изменение имени или пути текстового файла).

Таким образом, при контентном изменении поведенческая информация включает в себя данные об операции вызвавшей данное изменение и о содержимом документа до и после его изменения; при контекстном изменении — только информацию об операции.

К собираемым поведенческим данным о работе пользователей с текстовой информацией из различных источников (текстовые документами, посещаемые web-страницами, электронная почта) можно применять средства аналитики текстовой информации (англ. Text Mining), с помощью которых, например, возможно извлекать ключевые слова, определять основные тематики, строить аннотации документов, анализировать тенденции интереса к определённому типу контента и т.п., а также производить поиск и категоризацию данных. Таким образом, анализ текстовой информации позволит обнаруживать и организовывать материалы, определять их деловую ценность и принимать решения, способствующие росту бизнес-результатов [2-6], в том числе касающиеся работы пользователей с конфиденциальными текстовыми данными. Понимание деловой ценности информации и процессов, происходящих с ней, позволит разрабатывать свои политики безопасности и применять их к различным типам информации, а также управлять рисками, связанными с использованием конфиденциальных данных и наличием неизвестной или неконтролируемой информации [7]. Для решения указанных задач в настоящее время в организациях применяются системы управления корпоративным контентом (англ. *Enterprise Content Management*, *ECM)*.

По мере роста объемов информационных ресурсов организациям становится все труднее эффективно использовать их [2-6]. Решения для управления корпоративным контентом (ECM) предоставляют программные средства сбора, анализа, управления,

накопления, хранения и доставки информации всем пользователям организации. ECM-системы ориентируется на работу с неструктурированной информацией в любом виде, включая офисные текстовые и табличные электронные документы, документы в формате PDF, а также рисунки, чертежи, графики, презентации, сканированные изображения, сообщения электронной почты, web-страницы и т.п. [5, 6, 8], т.е. по большей части это текстовая информация. Также большинство современных ECM-систем включает в себя компоненты электронного обнаружения или электронного раскрытия информации (англ. *eDiscovery*). Средства *eDiscovery* обеспечивают процесс, с помощью которого организации находят, получают, сохраняют и анализируют документы, связанные с определенным судебным делом.

Согласно Gartner (Gartner — компания, мировой лидер в области исследований и консалтинговых услуг, специализирующаяся на рынках информационных технологий [9]) ECM-система — это стратегическая инфраструктура и техническая архитектура для поддержки единого жизненного цикла неструктурированной информации (контента) различных типов и форматов. ECM-системы состоят из приложений, которые могут взаимодействовать между собой, а также использоваться и продаваться самостоятельно. Gartner выделяет следующие ключевые компоненты современных ECM-систем [8, 10]:

- *управление документами* (англ. *Document management*) — экспорт, импорт, контроль версий, безопасность и службы библиотек для деловых документов;
- *управление образами документов* (англ. *Image-processing applications*) — захват, преобразование и управление бумажными документами;
- *управление потоками работ* (англ. *Content workflow*) — поддержка бизнес-процессов, передача контента по маршрутам, назначение рабочих задач и состояний, создание журналов аудита;
- *управление записями* (англ. *Records management*, в соответствии с последним переводом стандарта IEEE 15489 — ГОСТ Р ИСО 15489-1-2007, «управление документами») — долгосрочное архивирование, автоматизация политик хранения и соответствия нормам регулирующих органов, обеспечение соответствия законодательным и отраслевым нормам;
- *управление веб-контентом* (англ. *Web content management*) — набор программных средств для управление веб-содержимым организации;
- *социальный контент* (англ. *Social content*) — реализует функционал для совместного использования документов, взаимодействия сотрудников, поддержки проектных команд;

– *расширяющие компоненты* (англ. *Extended components*) — компоненты, реализующие следующие функции: поиск и аналитика контента, электронное обнаружение или раскрытие информации (англ. *electronic discovery*, *eDiscovery*), архивирование данных (в том числе электронной почты, данных из других информационных источников) и т.п.

Помимо, непосредственно, инструментов для сбора, управления, накопления, хранения и доставки информации, современные ECM-системы обладают развитыми средствами аналитики неструктурированных данных, которые позволяют обнаруживать материалы, определять их ценность и принимать решения, способствующие росту бизнес-результатов. По данным IBM [4] 73% генеральных директоров различных компаний совершает значительные вложения, чтобы повысить способность компании извлекать из имеющихся данных ценную информацию.

Как уже отмечалось, помимо программных компонент анализа контентных данных, подавляющее большинство современных ECM-систем включает в себя компоненты электронного обнаружения или электронного раскрытия информации (англ. *eDiscovery*). Эта возможность также иногда называется *удержанием* [11-14]. Помимо информации, циркулирующей на основных компонентах ECM, речь о которых шла выше, компоненты *eDiscovery* могут собирать данные на рабочих местах сотрудников и корпоративных ноутбуках (пример решений: *OpenText* [14], *IBM Desktop Data Collector* [15]), что их приближает по функционалу к DLP-системам. Помимо непосредственно поиска документов, связанных с запросом на электронное раскрытие, также зачастую требуются средства анализа найденных документов, в том числе включающие анализ действий пользователей с информацией. В связи с этим отдельным пунктом будут рассмотрены методы анализа, применяемые в системах eDiscovery.

Из приведённого описания целевого функционала систем классов ECM и eDiscovery следует, что программа для ЭВМ «Система мониторинга, теневого копирования и автоматического аннотирования текстовых данных при работе пользователя с электронными документами» имеет широкие перспективы практического использования в качестве платформы, осуществляющей сбор поведенческой информации пользователей при работе с текстовыми данными, для последующего применения различных методов интеллектуального анализа данных с целью извлечения ценной для организации информации. Также в функционал регистрируемой программы для ЭВМ уже входит реализация метода автоматического построения аннотаций к собираемым текстовым данным, что позволяет сокращать время анализа содержимого документов, с которыми работал пользователь (например, при анализе действий пользователя можно быстро определить факт работы

пользователя с материалами, не относящиеся к его рабочей деятельности — задача обнаружения нецелевого использования корпоративных ресурсов).

За последние несколько лет наблюдается устойчивый рост интереса к задачам безопасности данных в корпоративных информационных системах, связанным с внутренними угрозами. Анализ работы пользователей с текстовыми документами возможно рассмотреть и сточки зрения решения таких задач информационной безопасности, как идентификация пользователей и раннее обнаружение попыток хищения информации. Решение данных задач главным образом направлено на снижение рисков, связанных с утечками корпоративных и/или конфиденциальных данных.

Утечки данных представляют существенную угрозу для современных компаний. Для их предотвращения традиционно используются системы класса DLP (англ. Data Loss Prevention), позволяющие в автоматическом режиме выполнять проверку потоков данных, покидающих информационный периметр организации, на содержание в них конфиденциальных данных. В DLP-системах правила работы с конфиденциальной информацией (т.е. политики безопасности) формируются экспертно для конкретных организаций. Ключевые различия в системах данного класса заключаются в применяемых методах классификации данных. На сегодняшний день применяются подходы классификации на основе цифровых отпечатков документов, анализа шаблонов данных и методов машинного обучения.

Сейчас многие эксперты сходятся во мнении, что применение традиционных DLP-систем недостаточно эффективно для противодействия внутренним вторжениям, и поэтому утечки необходимо определять ещё до стадии пересылки данных за информационный периметр. Данное утверждение основывается на исследованиях [16], которые показывают, что от момента, когда пользователь решает украсть данные до непосредственно пересылки данных, проходит от нескольких недель до нескольких месяцев, которые уходят на стадию исследования и подготовки утечки. В данной стадии поведение пользователя отличается от его обычной легитимной активности как по набору выполняемых действий, так и по содержанию обрабатываемой информации. Обнаружение признаков несвойственного поведения пользователя реализуют системы класса UEBA (англ. User and Entity Behavior Analytics — анализ поведения пользователей и систем). Согласно Gartner [17], UEBA-системы на основе методов машинного обучения выполняют построение и применение моделей поведения (профилей) пользователей с целью выявления признаков аномального поведения. В свою очередь, аномальное поведение пользователя может свидетельствовать о том, что: пользователь не является тем, от имени кого он авторизовался (задача идентификации

пользователей); пользователь интересуется корпоративными документами, которые не относятся к его текущей рабочей деятельности, что является признаком потенциальной утечки информации (задача раннего обнаружения попыток хищения информации).

UEBA-системы, в отличие от DLP, осуществляют мониторинг широкого спектра действий пользователя и принимают решения не на основе экспертно-сформированных политик безопасности, а на основе исторических данных о легитимной работе пользователя. UEBA-системы обнаруживают ранние признаки утечки, поэтому их основная цель состоит не в блокировки действий пользователей, а в предоставлении аналитических данных службе ИБ с описанием того, почему выявленные действия являются аномальными для конкретного пользователя.

Обычно целью внутренних вторжений является получение доступа к текстовой информации (финансовые отчёты, договора, техническая документация, электронная почта и т.п.), поэтому ключевым является выявление аномального поведения пользователей при работе с текстовыми данными. Традиционно UEBA-системы с помощью методов машинного обучения анализируют данные об операциях пользователя (контекстную информацию), которые является хорошо структурированными, например, данные системных журналов ОС. Поэтому традиционные подходы не способны выявить случаи нелегитимной активности пользователя, состоящей из характерных для него действий, но с нелегальным содержимым (контентом). Следовательно, актуальным является исследование возможности анализа обрабатываемых пользователем текстовых данных с целью обнаружения аномального поведения.

Также актуальность данного направления подтверждает отчёт Gartner [17], в котором утверждается, что задача анализа текстовых данных является гораздо более сложной, чем анализ структурированной информации. Поэтому Gartner ожидает появление данного функционала в UEBA-системах в течении следующих нескольких лет. Также Gartner отмечает, что многие клиенты обращаются к поставщикам UEBA-систем с желанием добавления анализа неструктурированных данных. Клиенты считают, что анализ неструктурированных данных позволит им более информативно определять, как текущую пользовательскую активность, так и поведение пользователя с течением времени, что позволит точнее выявлять инсайдерские угрозы.

Таким образом, на сегодняшний день задача обнаружения аномального поведения пользователей при работе с текстовыми данными является актуальной и не рассматривается в существующих решениях UEBA-систем. В то же время за счёт расширения традиционных источников поведенческой информации анализ содержимого текстовых данных позволит:

- раньше обнаруживать индикаторы изменения поведения;
- сокращать и приоритизировать события безопасности, генерируемые другими системами ИБ;
- сокращать время расследования инцидентов безопасности и число сотрудников службы ИБ.

Для обнаружения аномального поведения пользователей на основе контента электронных документов, к которым они обращались, необходимо осуществлять сбор соответствующей поведенческой информации, который реализует полученный в результате выполнения ПНИ РИД — программа для ЭВМ «Система мониторинга, теневого копирования и автоматического аннотирования текстовых данных при работе пользователя с электронными документами». В рассматриваемом случае, анализируется только содержимое электронных документов и время обращения к ним. Следовательно, предполагается практическое использование полученных РИД в классе систем UEBA, которые служат для противодействия внутренним вторжениям, в частности для предотвращения утечек корпоративных и/или конфиденциальных данных. Более того, указанная программа для ЭВМ в сочетании с разработанными на предыдущих этапах настоящих ПНИ [1] методами машинного обучения и математической статистики для построения и применения поведенческих моделей на основе поведенческой биометрии работы пользователя фактически представляет собой UEBA-систему.

Gartner ожидает, что доход рынка UEBA-систем достигнет 200$ млн к концу 2017 года по сравнению с менее чем 50$ млн в 2015 [17]. Данные показатели свидетельствуют о том, что рост рынка UEBA-систем превзошёл оценки Gartner 2014 года. Также отмечается, что к 2017 году по крайней мере 20% крупнейших поставщиков систем ИБ, фокусирующихся на мониторинге и контроле действий пользователей (например, SIEM и DLP системы), будут включать UEBA-функционал в свои решения.

Примеры становления рынка:

- В июле 2015 компания Splunk приобрела UEBA-стартап Caspida за 190$ млн;
- В сентябре 2015 компания Microsoft приобрела систему облачной безопасности с UEBA-функционалом за 250$ млн;
- В апреле 2015 компания HP объявила об интеграции технологий UEBA-производителя Securonix в продукт HP ArcSight.

Gartner [18] определяет основную целевую группу клиентов, как организации, которые имеют собственную службу ИБ (банки, брокерские фирмы, оборонные предприятия и т.п.).

Т.е. большие компании, которые понимают риск появления новых, ранее неизвестных, внутренних вторжения и высоко оценивают потенциал применения методов машинного обучения для их выявления. Также в качестве группы клиентов выделяют фирмы, которые предоставляют своим клиентам услуги по обеспечению ИБ (мониторинг и расследованию инцидентов ИБ, администрирование систем ИБ и т.п.) — такие фирмы Gartner классифицирует как Managed Security Service Provider.

Разработанные в рамках настоящего проекта программные модули, по которым получены РИД «Система двухфакторной аутентификации на основе анализа поведенческой биометрической информации об особенностях работы пользователя с клавиатурой компьютера» и «Система двухфакторной аутентификации на основе анализа поведенческой биометрической информации об особенностях работы пользователя с компьютерной мышью», обеспечивают решение задач активно аутентификации пользователя на основе сопоставления характеристик динамики обведения шаблона мышью с индивидуальной моделью компьютерной биометрии и по динамике ввода предопределенной текстовой строки (пароля) и по совпадению введенного текста с паролем. Данные технологии могут быть использованы как часть общего комплекса, разрабатываемого в рамках настоящего проекта, так и реализованы в виде отдельных программных продуктов, решающих исключительно задачу активной аутентификации пользователей по особенностям работы с клавиатурой и манипулятором мышь.

По оценкам Gartner ("Прогноз 2015: Управление идентификацией и доступом" http://www.gartner.com/document/2912417) к 2020 году новые биометрические методы будут вытеснять пароли и сканирование отпечатков пальцев для доступа с оконечных устройств и персональных компьютеров. Биометрическая технология не нова, но в настоящее время становится все более популярна как для активной аутентификации на мобильных устройствах, так и на персональных компьютерах. В настоящий момент уже общепризнана уязвимость традиционных подходов к аутентификации на основе паролей, электронных ключей и других методов, основанных на проверке некоторой секртеной информации. Во многих случаях эту секретную информацию можно потерять, передать, подделать и даже подобрать. Поэтому развиваются направления многофакторной аутентификации, требующей помимо знания секретной информации от пользователя также возможности провести дополнительную аутентификацию, например с помощью кода, полученного по SMS. Одним из таких направлений является использование биометрии как поведенческой так и биологической для подтверждения подлинности пользователя. Некоторое время как основная альтернатива традиционным подходам рассматривалась технология на основе проверки отпечатков

пальцев. Как ожидает Gartner эта технология достигнет своего пика на уровне около 20 процентов от общего объема рынка оконечного устройства в 2017. Биометрические реализации этого метода являются относительно слабыми, т.к. выделение признаков, сравнение и сопоставление были настроены, чтобы обеспечить хороший пользовательский опыт и хорошую производительность на устройстве, а не для установления высокого уровня надежности и безопасности. Gartner прогнозирует, что будут вкладываться деньги в такие технологии многофакторной аутентификации как распознавание лиц с помощью камеры, распознавания голоса с помощью микрофона, динамика нажатия клавиш и работы с манипуляторами и тачскринами. Одним из основных преимуществ этих методов перед отпечатками пальцев является то, что не нужен специализированный датчик. А для динамики работы с клавиатурой и манипуляторам не нужны даже уже ставшие массово используемыми камеры и микрофоны. Следовательно, эти методы могут быть реализованы чисто программно, тем самым принося пользу всем пользователям, а не только тем, не зависимо от конфигурации их устроств. В настоящее время более десятка компаний производят и успешно реализуют технологии активной аутентификации с использований данных о динамике нажатия клавишь и работе с манипулятором мышь. В том числе, TypeWATCH (https://www.watchfulsoftware.com/en/products/typewatch), Intensity Analytics (http://www.intensityanalytics.com), AdmitOneSecurity (http://www.admitonesecurity.com), BioTracker (http://plurilock.com), KeyTrac (https://www.keytrac.net), ID Control (http://www.idcontrol.com), Deepnet Security (http://www.deepnetsecurity.com), Authenware Corp (http://www.authenware.com), BioChec (http://www.biochec.com), Delfigo Security (http://www.delfigosecurity.com), BehavioSec (http://www.behaviosec.com) и другие.

3.3 Выводы

Полученные в результате выполнения данных ПНИ РИД могут послужить основой для построения перспективных систем информационной безопасности класса UEBA, основанных на анализе компьютерной поведенческой биометрии, в частности на использовании характеристик работы пользователя, как с текстовыми документами, так и с стандартными устройствами ввода информации (клавиатура, мышь).

Одним из важнейших критериев при оценки актуальности, перспективности и технико-экономической значимости РИД является обеспечение их эффективной коммерциализации. Основными формами коммерциализации полученных в проекте РИД могут быть привлечение

венчурного финансирования, заключение лицензионных соглашений, внесение РИД в уставный капитал. Обозначим более подробные сценарии коммерциализации предсатвленных РИД.

1. Передача технологии. В результате работ в рамках настоящего проекта созданая интеллектуальная собственность может быть продана другой заинтересованной компании, работающей на рынке IT-безопасности в данной сфере.
2. Лицензируемое ПО («корпоративная версия»). Основной целью в рамках настоящего проекта является реализация на базе разрабатываемой инновационной технологии программного комплекса защиты от несанкционированного доступа и внутренних вторжений, обладающего значительными конкурентными преимуществами по сравнению с существующими решениями. Первую версию можно реализовать в виде программного решения масштаба предприятия (enterprise), которое будет полностью устанавливаться на площадке и оборудовании Заказчика. Основной путь коммерциализации в рамках данного направления будет продажа лицензий и оплата поддержки.
3. SaaS («облачное решение»). В перспективе при наличии инвестиционных средств и успешного опыта эксплуатации «корпоративной версии» возможна реализация SaaS-версии системы. Это позволит сделать ее доступной для небольших и средних компаний и организаций, не готовых тратить средства на закупку собственного оборудования, ПО и подготовки специалистов для работы с такой системой. В этом случае на площадке Клиента будут устанавливаться только ПО агентов и, возможно, сервер промежуточной консолидации данных (он предназначен для оптимизации трафика с возможностью задания гибких политик передачи данных на основной сервер и не требует наличия высокопроизводительного оборудования). Хранение и обработка поведенческой информации, построение и валидация пользовательских моделей будет осуществляться централизовано «в облаке» — на площадке компании-производителя.
4. Расширенная поддержка («аутсорсинг аналитики»). Под «аутсорсингом» аналитики подразумевается осуществление периодических работ по созданию и настройке поведенческих моделей на основе собранных данных у Заказчика (как в «корпоративной» версии, так и в «облаке») силами специалистов компании-производителя в случае отсутствия у Заказчика своих специалистов, способных осуществлять такие работы. Возможно осуществлять такие работы в рамках дополнительно оплачиваемой поддержки.

ЗАКЛЮЧЕНИЕ

Согласно содержанию плана-графика исполнения обязательств, при выполнении этапа 4 настоящих прикладных научных исследований должны были быть проведены ниже перечисленные работы.

А) Проведение экспериментальных исследований по разработанной Программе и методикам экспериментальных исследований ЭО ПК, в том числе:

- скорости работы методов сбора и предобработки поведенческих биометрических данных, а также объемов собираемых биометрических данных в условиях реальной работы;
- оценка точности и скорости работы методов предобработки текстовых данных, в том числе методов рубрикации, группировки, автоматического аннотирования и выявления ключевых слов на экспериментальных и/или эталонных тестовых данных;
- оценка точности и скорости работы методов машинного обучения и математической статистики для построения и применения поведенческих моделей.

Б) Обеспечение за счет средств из внебюджетных источников работоспособности рабочих станций, общего системного и специального программного обеспечения для проведения работ по проекту, проведение регламентных работ по обеспечению сохранности данных и программ, относящихся к проводимым работам по проекту.

В) Проведение за счет средств из внебюджетных источников оценки РИД, полученных при выполнении ПНИ, с целью их вовлечения в хозяйственный оборот.

По результатам выполнения запланированных на 4-й этап настоящих ПНИ работ можно сделать следующие выводы:

- в результате проведения экспериментальных исследований по разработанной Программе и методикам экспериментальных исследований ЭО ПК установлено, что работы полностью соответствуют требованиям п. 3.13 ТЗ, а также Программе и методикам экспериментальных исследований ЭО ПК, продемонстрировано соответствие результатов теоретических исследований требованиям настоящего технического задания, в частности, пунктам 3.13.1, 3.13.2, 3.13.3, 4.1.1, 4.1.2;
- обеспечена работоспособности рабочих станций, общего системного и специального программного обеспечения для проведения работ по проекту, проведены регламентные

работы по обеспечению сохранности данных и программ, относящихся к проводимым работам по проекту;

- в результате проведенных работ по оценки РИД, полученных при выполнении ПНИ, с целью их вовлечения в хозяйственный оборот выявлен класс систем информационной безопасности, для которых актуально использование полученных в рамках ПНИ РИД, что позволит использовать данные РИД для построения перспективных систем информационной безопасности класса UEBA, основанных на анализе компьютерной поведенческой биометрии, в частности на использовании характеристик работы пользователя, как с текстовыми документами, так и с стандартными устройствами ввода информации (клавиатура, мышь).

Задачи, поставленные на отчетном этапе выполнены полностью. Сведения о ходе выполнения настоящих ПНИ размещены в открытом доступе на официальном сайте МГУ имени М.В.Ломоносова (система ИСТИНА-интеллектуальная система тематического исследования научно-технической информации) по адресу http://istina.msu.ru/projects/7964619/ .

СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ

1 Машечкин И.В. и др. Исследование и разработка инновационной технологии построения программных средств обеспечения компьютерной безопасности, основанных на использовании методов машинного обучения и математической статистики для анализа данных поведенческой биометрии пользователей при работе в рамках стандартного человеко-машинного интерфейса, для решения задач активной аутентификации и идентификации пользователей, обнаружения внутренних вторжений и предотвращения попыток хищения конфиденциальной информации // Отчет о прикладных научных исследованиях (промежуточный) по теме «Теоретические исследования (2-ой очереди) поставленных перед ПНИ задач». — М., 2015.

2 Susan Feldman, Judy Hanover, Cynthia Burghard, David Schubmehl. Unlocking the Power of Unstructured Data // IDC Health Insights, June 2012.

3 Susan Feldman, David Schubmehl, Hadley Reynolds. Content Analytics and the High-Performing Enterprise // IDC Analyst Research, December 2012.

4 Управление информационными ресурсами предприятия [Электронный ресурс]. — Электрон. дан. и прогр. — [Б. м.] : IBM, 2014. — Режим доступа: http://www-03.ibm.com/software/products/en/category/enterprise-content-management. — 28.11.2015.

5 IBM Content Analytics with Enterprise Search, Version 3.0 [Электронный ресурс]. — Электрон. дан. — [Б. м.] : IBM, 2014. — Режим доступа: http://public.dhe.ibm.com/common/ssi/ecm/en/zzd03138usen/ZZD03138USEN.PDF. — 28.11.2015.

6 IBM Content Classification [Электронный ресурс]. — Электрон. дан. — [Б. м.] : IBM, 2014. — Режим доступа: http://public.dhe.ibm.com/common/ssi/ecm/en/zzd03136usen/ZZD03136USEN.PDF. — 28.11.2015.

7 EMC выпускает решение для управления жизненным циклом неструктурированного содержания [Электронный ресурс]. — Электрон. дан. — [Б. м.] : EMC, 2014. — Режим доступа: http://russia.emc.com/about/news/press/2010/20100623-01.htm. — 05.06.2016.

8 ECM - Enterprise Content Management [Электронный ресурс]. — Электрон. дан. — [Б. м.] : ECM, 2014. — Режим доступа: http://www.tadviser.ru/index.php/Статья:ECM_-_Enterprise_Content_Management. — 05.06.2016.

9 Gartner, Inc. [Электронный ресурс]. — Электрон. дан. — [Б. м.] : Gartner, 2014. — Режим доступа: http://www.gartner.com/technology/about.jsp. — 05.06.2016.

10 Mark R. Gilbert, Karen M. Shegda, Kenneth Chin, Gavin Tay, Hanns Koehler-Kruener. Magic Quadrant for Enterprise Content Management [Электронный ресурс]. — Электрон. дан. — [Б. м.] : Gartner, 2014. — Режим доступа: http://www.gartner.com/technology/reprints.do?id=1-22RDH82&ct=141003&st=sb. — 28.11.2015.

11 Планирование eDiscovery (SharePoint Server 2010) [Электронный ресурс]. — Электрон. дан. — [Б. м.] : Microsoft, 2014. — Режим доступа: http://technet.microsoft.com/ru-ru/library/ff453933(v=office.14).aspx. — 05.06.2016.

12 EMC Kazeon eDiscovery [Электронный ресурс]. — Электрон. дан. и прогр. — [Б. м.] : EMC, 2014. — Режим доступа: http://russia.emc.com/content-management/emc-kazeon-ediscovery.htm. — 05.06.2016.

13 Early Case Assessment by Recommind [Электронный ресурс]. — Электрон. дан. и прогр. — [Б. м.] : OpenText, 2014. — Режим доступа: http://www.opentext.com/what-we-do/products/discovery/ediscovery/opentext-ediscovery-early-case-assessment-by-recommind. — 05.06.2016.

14 eDiscovery Analyzer [Электронный ресурс]. — Электрон. дан. и прогр. — [Б. м.] : IBM, 2014. — Режим доступа: http://www-03.ibm.com/software/products/ru/edisanal. — 05.06.2016.

15 Desktop Data Collector [Электронный ресурс]. — Электрон. дан. и прогр. — [Б. м.] : IBM, 2014. — Режим доступа: http://www-03.ibm.com/software/products/ru/desktop-data-collector. — 05.06.2016.

16 ObserveIT Data Loss Prevention Capabilities [Электронный ресурс]. — Электрон. дан. — [Б. м.] : ObserveIT, 2015. — Режим доступа: http://www.observeit.com/blog/observeit-data-loss-prevention-capabilities-1. — 05.06.2016.

17 Gartner. Market Guide for User and Entity Behavior Analytics [Электронный ресурс]. — Электрон. дан. — [Б. м.] : Gartner, 2015. — Режим доступа: https://www.gartner.com/doc/reprints?id=1-2NK6M1R&ct=150922&st=sb. — 05.06.2016.

18 Gartner. Market Trends: User and Entity Behavior Analytics Expand Their Market Reach[Электронный ресурс]. — Электрон. дан. — [Б. м.] : Gartner, 2016. — Режим доступа: https://www.gartner.com/doc/reprints?id=1-370BP2V&ct=160518&st=sb. — 05.06.2016.

ПРИЛОЖЕНИЕ А

Акт приемки результатов экспериментальных исследований

СОГЛАСОВАНО
Заместитель генерального директора
ООО "ЭДАЛ СОФТ"
А.Н.Терехин
«28» _____ 2016г.

УТВЕРЖДАЮ
Заместитель проректора МГУ
Э.А.Сазонов
_____ 2016 г.

АКТ

приемки результатов экспериментальных исследований в рамках выполнения 4 этапа "Экспериментальные исследования поставленных перед ПНИ задач" Соглашения о предоставлении субсидии № 14.604.21.0056 от 27 июня 2014 года с Министерством образования и науки РФ

«28»июня 2016 г.

Комиссия в составе:

Председатель	С.н.с., к.ф.-м.н Бахмуров Анатолий Геннадьевич
члены комиссии	Доцент, к.ф.-м.н. Головин Игорь Геннадиевич
	Доцент, к.ф.-м.н. Чернов Александр Владимирович

назначенная приказом по факультету ВМК МГУ имени М.В.Ломоносова от 23 июня 2016 г. №238, в период с «27» июня 2016 г. по «28» июня 2016 г. проверила результаты зкспериментальных исследований по разработанной Программе и методикам.

1. **Комиссии предъявлены:**
   1.Экспериментальный образец программного комплекса (ЭО ПК) для обеспечения компьютерной безопасности на основе анализа поведенческой информации работы пользователей компьютерных систем.
   2. Программная документация на ЭО ПК.
   3. Техническое задание к Соглашению о субсидии № 14.604.21.0056.
   4. Программа и методики экспериментальных исследований.
   5. Протоколы экспериментальных исследований.

2. В результате проверки установлено:

На 4 этапе выполнения Соглашения о предоставлении субсидии № 14.604.21.0056 от 27 июня 2014 года с Министерством образования и науки РФ были проведены экспериментальные исследования в соответствие с разработанной программой и методикой. Произведена оценка скорости работы методов сбора и предобработки поведенческих биометрических данных, которая показала применимость разработанных методов для задачи анализа данных о работе пользователя

с устройствами ввода-вывода, информационными и вычислительными ресурсами компьютерной системы и потоками текстовой информации. На основе оценки точности выбраны лучшие методы для решения задач рубрикации, группировки, автоматического аннотирования и выявления ключевых слов. Проведена оценка точности и выбор методов машинного обучения и математической статистики для построения и применения поведенческих моделей. Все результаты оформлены соответствующими протоколами.

3. Вывод

В соответствие с п. 3.13 Технического задания проведены экспериментальные исследования ЭО ПК для обеспечения компьютерной безопасности на основе анализа поведенческой информации работы пользователей компьютерных систем по разработанной Программе и методикам экспериментальных исследований. Продемонстрировано соответствие полученных теоретических и практических исследований требованиям технического задания.

ПРИЛОЖЕНИЕ:

1. Протокол испытания №1 по пункту № 4.3.1.1 Программы экспериментальных исследований
2. Протокол испытания №2 по пункту № 4.3.1.2 Программы экспериментальных исследований
3. Протокол испытания №3 по пункту № 4.3.1.1 Программы экспериментальных исследований
4. Протокол испытания №4 по пункту № 4.3.1.2 Программы экспериментальных исследований
5. Протокол испытания №5 по пункту № 4.3.1.1 Программы экспериментальных исследований
6. Протокол испытания №6 по пункту № 4.3.1.2 Программы экспериментальных исследований
7. Протокол испытания №7 по пункту № 4.3.1.1 Программы экспериментальных исследований
8. Протокол испытания №8 по пункту № 4.3.1.2 Программы экспериментальных исследований
9. Протокол испытания N9 по пункту № 4.3.2.1 Программы экспериментальных исследований
10. Протокол испытания №10 по пункту № 4.3.2.2 Программы экспериментальных исследований
11. Протокол испытания №11 по пункту № 4.3.2.1 Программы экспериментальных исследований
12. Протокол испытания №12 по пункту № 4.3.2.2 Программы экспериментальных исследований
13. Протокол испытания №13 по пункту № 4.3.3.1 Программы экспериментальных исследований
14. Протокол испытания №14 по пункту № 4.3.3.2 Программы экспериментальных исследований
15. Протокол испытания №15 по пункту № 4.3.3.3 Программы экспериментальных исследований
16. Протокол испытания №16 по пункту № 4.3.1.3 Программы экспериментальных исследований
17. Протокол испытания №17 по пункту № 4.3.2.3 Программы экспериментальных исследований
18. Протокол испытания №18 по пункту № 4.3.3.4 Программы экспериментальных исследований

Председатель комиссии А.Г.Бахмуров

Члены комиссии И.Г.Головин

А.В.Чернов

ПРОТОКОЛ №1

испытания по пункту № *4.3.1.1*

Программы *экспериментальных исследований*

2015.291215.ПМ1

№ *_1__* *27 июня* 2016 г.

1. Объект испытания: «Подсистема 1» ЭО ПК для сбора и анализа биометрической информации об особенностях динамики работы пользователя с устройствами ввода-вывода в рамках стандартного человеко-машинного интерфейса.

2. Цель испытания: проверка соответствия объекта экспериментального исследования требованиям пункта № *3.11.5, 3.13, 4.1.1* Технического задания.

3. Дата начала испытания: *17 марта 2016* г.

4. Дата окончания испытания: *18 марта 2016* г.

5. Место проведения испытания: лаборатория технологий программирования факультета Вычислительной Математики и Кибернетики МГУ им. М.В.Ломоносова

6. Средства проведения испытаний

Рабочая станции WST_WIN7: программные характеристики - ОС Windows 7 (32bit); аппаратные характеристики: процессор Core2 Duo, частота 2.4 ГГц, 2 ядра; оперативная память 1Гбайт; дисковый накопитель HDD, 80 Гбайт; сетевой адаптер 100Мбит/сек.

7. Результаты испытания приведены в таблице Б.1.

Таблица Б.1 — Результаты экспериментов по скорости и объемам собираемых данных для фоновой идентификации с подсистемой 1

Наименование параметра	Ед.изм.	Номера пунктов			Требование к параметру		Измеренное значение
		Технического задания	Программы испытаний	Методики испытаний	Номинальное значение	Предельное отклонени	

						e	
Скорость сбора данных (мышь)	событий/час	3.11.5, 3.13, 4.1.2	4.3.1.1	6.3.1	38000	1000	38412
Скорость сбора данных (клавиатура)	событий/час	3.11.5, 3.13, 4.1.1	4.3.1.1	6.3.1	400	100	427
Объем собираемых данных (пиковый за час, мышь)	Kb	3.11.5, 3.13, 4.1.1	4.3.1.1	6.3.1	2200	300	2356
Объем собираемых данных (дневной, мышь)	Kb	3.11.5, 3.13, 4.1.1	4.3.1.1	6.3.1	8000	500	8265
Объем собираемых данных (недельный, мышь)	Kb	3.11.5, 3.13, 4.1.1	4.3.1.1	6.3.1	36200	1000	36328
Объем собираемых данных (пиковый за час, клавиатура)	Kb	3.11.5, 3.13, 4.1.1	4.3.1.1	6.3.1	100	50	138
Объем собираемых данных (дневной, клавиатура)	Kb	3.11.5, 3.13, 4.1.1	4.3.1.1	6.3.1	400	50	415
Объем собираемых данных (недельный, клавиатура)	Kb	3.11.5, 3.13, 4.1.1	4.3.1.1	6.3.1	2000	100	2075

8. Замечания и рекомендации отсутствуют

9. Выводы

9.1 Объект испытания *«Подсистема 1» ЭО ПК* выдержал испытание по пункту № *4.3.1.1* Программы и методики *2015.291215.ПМ1.*

9.2 Объект испытания *«Подсистема 1» ЭО ПК* соответствует требованиям пункта №*3.11.5, 3.13*, *4.1.1* Технического задания.

Испытание проводили		
Ассистент		В.В.Глазкова
М.н.с.		И.С.Попов
Доцент		М.И.Петровский
Математик 1-й кат.		Д.В.Царёв
Математик 1-й кат.		П.М.Саликов
Программист		Д.А.Никифоров
Математик 1-й кат		О.Е.Горохов

ПРОТОКОЛ №2

испытания по пункту № 4.3.1.2

Программы экспериментальных исследований

2015.291215.ПМ1

№ _1__ 27 июня 2016 г.

1. Объект испытания: «Подсистема 1» ЭО ПК для сбора и анализа биометрической информации об особенностях динамики работы пользователя с устройствами ввода-вывода в рамках стандартного человеко-машинного интерфейса.

2. Цель испытания: проверка соответствия объекта экспериментального исследования требованиям пункта № 3.11.5, 3.13, 4.1.3 Технического задания.

3. Дата начала испытания: 17 марта 2016 г.

4. Дата окончания испытания: 18 марта 2016 г.

5. Место проведения испытания: лаборатория технологий программирования факультета Вычислительной Математики и Кибернетики МГУ им. М.В.Ломоносова

6. Средства проведения испытаний

Сервер SRV_SYS1: программные характеристики (ОС Microsoft Windows 2012R2, 64bit; интерпретатор Python3;модуль Python pandas 0.16.2; модуль Python numpy 1.9.3; модуль user-agent 1.0.1), аппаратные характеристики (два процессора x64 с тактовой частотой 2.4 Ггц, 4 ядра; оперативная память объемом 64 Гбайт; два дисковый накопителя HDD, объемом 2Тбайт в режиме RAID1; два дисковый накопителя HDD, объемом 1Тбайт; два сетевых адаптера с пропускной способностью 1 Гбит).

7. Результаты испытания приведены в таблице В.1.

Таблица В.1 — Результаты экспериментов по скорости построения и применения моделей фоновой идентификации с подсистемой 1

Наименование параметра	Ед.изм.	Номера пунктов			Требование к параметру		Измеренное значение			
		Технического задания	Программы испытаний	Методики испытаний	Номинальное значение	Предельное отклонение	SVM	kNN	Fuzzy	RNN
Скорость построения модели (минимальная, мышь)	сек	3.11.5, 3.13, 4.1.3	4.3.1.2	.3.2	2000	1980	2	05	8	147
Скорость построения модели (максимальная, мышь)	сек	3.11.5, 3.13,4.1.3	4.3.1.2	6.3.2	2000	1980	5	37	94	491
Скорость построения модели (средняя, мышь)	сек	3.11.5, 3.13,4.1.3	4.3.1.2	.3.2	000	1980	6	14	80	356
Скорость применения модели (минимальная, мышь)	сек	3.11.5, 3.13, 4.1.3	4.3.1.2	6.3.2	0.5	0.3	0.3	0.5	0.7	0.8
Скорость применения модели (максимальная, мышь)	сек	3.11.5, 3.13, 4.1.3	4.3.1.2	6.3.2	0.7	0.3	0.5	0.7	0.9	1
Скорость применения модели (средняя, мышь)	сек	3.11.5, 3.13, 4.1.3	4.3.1.2	6.3.2	0.6	0.3	0.4	0.6	0.8	0.9

Скорость построения модели (минимальная, клавиатура)	сек	3.11.5, 3.13, 4.1.3	4.3.1.2	6.3.2	2000	1980	43	231	632	3298
Скорость построения модели (максимальная, клавиатура)	сек	3.11.5, 3.13, 4.1.3	4.3.1.2	6.3.2	2000	1980	67	265	687	3576
Скорость построения модели (средняя, клавиатура)	сек	3.11.5, 3.13, 4.1.3	4.3.1.2	6.3.2	2000	1980	52	248	660	3456
Скорость применения модели (минимальная, клавиатура)	сек	3.11.5, 3.13, 4.1.3	4.3.1.2	6.3.2	0.6	0.4	0.3	0.4	0.6	0.8
Скорость применения модели (максимальная, клавиатура)	сек	3.11.5, 3.13, 4.1.3	4.3.1.2	6.3.2	0.8	0.5	0.5	0.6	0.8	1
Скорость применения модели (средняя, клавиатура)	сек	3.11.5, 3.13, 4.1.3	4.3.1.2	6.3.2	0.7	0.4	0.4	0.5	0.7	0.9

8. Замечания и рекомендации отсутствуют

9. Выводы

9.1 Объект испытания «Подсистема 1» ЭО ПК выдержал испытание по пункту № 4.3.1.2 Программы и методики 2015.291215.ПМ1.

9.2 Объект испытания «Подсистема 1» ЭО ПК соответствует требованиям пункта №3.11.5, 3.13, 4.1.3 Технического задания.

Испытание проводили

Ассистент	В.В.Глазкова
М.н.с.	И.С.Попов
Доцент	М.И.Петровский
Математик 1-й кат.	Д.В.Царёв
Математик 1-й кат.	П.М.Саликов
Программист	Д.А.Никифоров
Математик 1-й кат	О.Е.Горохов

ПРОТОКОЛ №3

испытания по пункту № 4.3.1.1

Программы экспериментальных исследований

2015.291215.ПМ1

№ _2__ 17 июня 2016 г.

1. Объект испытания: «Подсистема 1» ЭО ПК для сбора и анализа биометрической информации об особенностях динамики работы пользователя с устройствами ввода-вывода в рамках стандартного человеко-машинного интерфейса.

2. Цель испытания: проверка соответствия объекта экспериментального исследования требованиям пункта № 3.11.5, 3.13, 4.1.1 Технического задания.

3. Дата начала испытания: 17 марта 2016 г.

4. Дата окончания испытания: 18 марта 2016 г.

5. Место проведения испытания: лаборатория технологий программирования факультета Вычислительной Математики и Кибернетики МГУ им. М.В.Ломоносова

6. Средства проведения испытаний

Сервер SRV_SYS1: программные характеристики (ОС Microsoft Windows 2012R2, 64bit; интерпретатор Python3;модуль Python pandas 0.16.2; модуль Python numpy 1.9.3; модуль user-agent 1.0.1), аппаратные характеристики (два процессора x64 с тактовой частотой 2.4 Ггц, 4 ядра; оперативная память объемом 64 Гбайт; два дисковый накопителя HDD, объемом 2Тбайт в режиме RAID1; два дисковый накопителя HDD, объемом 1Тбайт; два сетевых адаптера с пропускной способностью 1 Гбит)

7. Результаты испытания приведены в таблице Г.1.

Таблица Г.1 — Результаты экспериментов по скорости предобработки данных для фоновой идентификации с подсистемой 1

Наименование параметра	Ед.изм.	Номера пунктов			Требование к параметру		Измеренное значение
		Технического задания	Программы испытаний	Методики испытаний	Номинальное значение	Предельное отклонение	
Скорость предобработки данных (мышь)	мин	3.11.5, 3.13, 4.1.1	4.3.1.1	6.3.1	20	10	18.3
Скорость предобработки данных (клавиатура)	мин	3.11.5, 3.13, 4.1.1	4.3.1.1	6.3.1	20	10	21.5

8. Замечания и рекомендации отсутствуют

9. Выводы

9.1 Объект испытания «Подсистема 1» ЭО ПК выдержал испытание по пункту № 4.3.1.1 Программы и методики 2015.291215.ПМ1.

9.2 Объект испытания «Подсистема 1» ЭО ПК соответствует требованиям пункта №3.11.5,

3.13, 4.1.1 Технического задания.

Испытание проводили

Ассистент	В.В.Глазкова
М.н.с.	И.С.Попов
Доцент	М.И.Петровский
Математик 1-й кат.	Д.В.Царёв
Математик 1-й кат.	П.М.Саликов
Программист	Д.А.Никифоров
Математик 1-й кат	О.Е.Горохов

ПРОТОКОЛ №4

испытания по пункту № 4.3.1.2

Программы экспериментальных исследований

2015.291215.ПМ1

№ _2__ 27 июня 2016 г.

1. Объект испытания: «Подсистема 1» ЭО ПК для сбора и анализа биометрической информации об особенностях динамики работы пользователя с устройствами ввода-вывода в рамках стандартного человеко-машинного интерфейса.

2. Цель испытания: проверка соответствия объекта экспериментального исследования требованиям пункта № 3.11.5, 3.13, 4.1.3 Технического задания.

3. Дата начала испытания: 17 марта 2016 г.

4. Дата окончания испытания: 18 марта 2016 г.

5. Место проведения испытания: лаборатория технологий программирования факультета Вычислительной Математики и Кибернетики МГУ им. М.В.Ломоносова

6. Средства проведения испытаний

Сервер SRV_SYS1: программные характеристики (ОС Microsoft Windows 2012R2, 64bit; интерпретатор Python3;модуль Python pandas 0.16.2; модуль Python numpy 1.9.3; модуль user-agent 1.0.1), аппаратные характеристики (два процессора x64 с тактовой частотой 2.4 Ггц, 4 ядра; оперативная память объемом 64 Гбайт; два дисковый накопителя HDD, объемом 2Тбайт в режиме RAID1; два дисковый накопителя HDD, объемом 1Тбайт; два сетевых адаптера с пропускной способностью 1 Гбит).

7. Результаты испытания приведены в таблице Д.1.

Таблица Д.1 — Результаты экспериментов по точности фоновой идентификации с подсистемой 1

Наименование параметра	Ед.изм.	Номера пунктов			Требование к параметру		Измеренное значение			
		Технического задания	Программы испытаний	Методики испытаний	Номинальное значение	Предельное отклонение	SVM	kNN	Fuzzy	RNN
AUC (мышь)	%	3.11.5, 3.13, 4.1.3	4.3.1.2	6.3.2	70	20	68.7	64.16	62.68	67.28
AUC (клавиатура)	%	3.11.5, 3.13, 4.1.3	4.3.1.2	6.3.2	80	20	85.42	82.72	86.60	83.79

8. Замечания и рекомендации отсутствуют

9. Выводы

9.1 Объект испытания «Подсистема 1» ЭО ПК выдержал испытание по пункту № 4.3.1.2 Программы и методики 2015.291215.ПМ1.

9.2 Объект испытания «Подсистема 1» ЭО ПК соответствует требованиям пункта №3.11.5,

3.13, 4.1.3 Технического задания.

Испытание проводили

Ассистент В.В.Глазкова

М.н.с. И.С.Попов

Доцент М.И.Петровский

Математик 1-й кат. Д.В.Царёв

Математик 1-й кат. П.М.Саликов

ПРОТОКОЛ №5

испытания по пункту № 4.3.1.1

Программы экспериментальных исследований

2015.291215.ПМ1

№ _3__ 22 апреля 2016 г.

1. Объект испытания: «Подсистема 1» ЭО ПК для сбора и анализа биометрической информации об особенностях динамики работы пользователя с устройствами ввода-вывода в рамках стандартного человеко-машинного интерфейса.

2. Цель испытания: проверка соответствия объекта экспериментального исследования требованиям пункта № 3.11.5, 3.13, 4.1.1 Технического задания.

3. Дата начала испытания: 22 апреля 2016 г.

4. Дата окончания испытания: 22 апреля 2016 г.

5. Место проведения испытания: лаборатория технологий программирования факультета Вычислительной Математики и Кибернетики МГУ им. М.В.Ломоносова

6. Средства проведения испытаний

Рабочая станции WST_WIN7: программные характеристики - ОС Windows 7 (32bit); аппаратные характеристики: процессор Core2 Duo, частота 2.4 ГГц, 2 ядра; оперативная память 1Гбайт; дисковый накопитель HDD, 80 Гбайт; сетевой адаптер 100Мбит/сек.

7. Результаты испытания приведены в таблице Ж.1.

Таблица Ж.1 — Результаты экспериментов по скорости и объемам собираемых данных для статической аутентификации с подсистемой 1

Наименование	Ед.изм.	Номера пунктов			Требование к параметру		Измеренное
		Техническог	Программ	Методики	Номинально	Предельно	

параметра		о задания	ы испытаний	испытаний	е значение	е отклонение	значение
Скорость сбора графической подписи (мышь)	сек	3.11.5, 3.13, 4.1.1	4.3.1.1	6.3.1	4	2	5
Объем собираемых данных (мышь, 10 вводов)	Kb	3.11.5, 3.13, 4.1.1	4.3.1.1	6.3.1	150	10	155
Объем собираемых данных (мышь, 50 вводов)	Kb	3.11.5, 3.13, 4.1.1	4.3.1.1	6.3.1	1500	100	1500
Объем собираемых данных (клавиатура, 15 вводов)	Kb	3.11.5, 3.13, 4.1.1	4.3.1.1	6.3.1	25	10	20
Объем собираемых данных (клавиатура, 10 вводов)	Kb	3.11.5, 3.13, 4.1.1	4.3.1.1	6.3.1	85	10	80

8. Замечания и рекомендации отсутствуют

9. Выводы

9.1 Объект испытания «Подсистема 1» ЭО ПК выдержал испытание по пункту № 4.3.1.1 Программы и методики 2015.291215.ПМ1.

9.2 Объект испытания «Подсистема 1» ЭО ПК соответствует требованиям пункта №3.11.5,
3.13, 4.1.1 Технического задания.

Испытание проводили

Испытание проводили

Ассистент	В.В.Глазкова
М.н.с.	И.С.Попов
Доцент	М.И.Петровский
Математик 1-й кат.	Д.В.Царёв
Математик 1-й кат.	П.М.Саликов
Программист	Д.А.Никифоров
Математик 1-й кат	О.Е.Горохов

ПРОТОКОЛ №6

испытания по пункту № 4.3.1.2

Программы экспериментальных исследований

2015.291215.ПМ1

№ _3__ 25 апреля 2016 г.

1. Объект испытания: «Подсистема 1» ЭО ПК для сбора и анализа биометрической информации об особенностях динамики работы пользователя с устройствами ввода-вывода в рамках стандартного человеко-машинного интерфейса.

2. Цель испытания: проверка соответствия объекта экспериментального исследования требованиям пунктов № 3.11.5, 3.13, 4.1.2 Технического задания.

3. Дата начала испытания: 25 апреля 2016 г.

4. Дата окончания испытания: 29 апреля 2016 г.

5. Место проведения испытания: лаборатория технологий программирования факультета Вычислительной Математики и Кибернетики МГУ им. М.В.Ломоносова

6. Средства проведения испытаний

Сервер SRV_SYS1: программные характеристики (ОС Microsoft Windows 2012R2, 64bit; интерпретатор Python3;модуль Python pandas 0.16.2; модуль Python numpy 1.9.3; модуль user-agent 1.0.1), аппаратные характеристики (два процессора x64 с тактовой частотой 2.4 Ггц, 4 ядра; оперативная память объемом 64 Гбайт; два дисковый накопителя HDD, объемом 2Тбайт в режиме RAID1; два дисковый накопителя HDD, объемом 1Тбайт; два сетевых адаптера с пропускной способностью 1 Гбит)

7. Результаты испытания приведены в таблице И.1.

Таблица И.1 — Результаты экспериментов по скорости построения и применения моделей статической аутентификации с подсистемой 1

Наименование параметра	Ед.изм.	Номера пунктов			Требование к параметру		Измеренное значение
		Технического задания	Программы испытаний	Методики испытаний	Номинальное значение	Предельное отклонение	
Скорость построения модели	милисекунд	3.11.5, 3.13, 4.1.2	4.3.1.2	6.3.2	45	10	50

(клавиатура, средняя)							
Скорость построения модели (клавиатура, минимальная)	милисекунд	3.11.5, 3.13, 4.1.2	4.3.1.2	6.3.2	45	10	44
Скорость построения модели (клавиатура, максимальная)	милисекунд	3.11.5, 3.13, 4.1.2	4.3.1.2	6.3.2	45	10	53
Скорость применения модели (клавиатура, средняя)	милисекунд	3.11.5, 3.13, 4.1.2	4.3.1.2	6.3.2	0.1	0.05	0.1
Скорость применения модели (клавиатура, минимальная)	милисекунд	3.11.5, 3.13, 4.1.2	4.3.1.2	6.3.2	0.1	0.05	0.097
Скорость применения модели (клавиатура, максимальная)	милисекунд	3.11.5, 3.13, 4.1.2	4.3.1.2	6.3.2	0.1	0.05	0.120
Скорость построения модели (мышь)	милисекунд	3.11.5, 3.13, 4.1.2	4.3.1.2	6.3.2	0.1	1	0.97
Скорость применения модели (мышь)	милисекунд	3.11.5, 3.13, 4.1.2	4.3.1.2	6.3.2	0.01	0.01	0.005

8. Замечания и рекомендации отсутствуют

9. Выводы

9.1 Объект испытания «Подсистема 1» ЭО ПК выдержал испытание по пункту № 4.3.1.2 Программы и методики 2015.291215.ПМ1.

9.2 Объект испытания «Подсистема 1» ЭО ПК соответствует требованиям пунктов № 3.11.5, 3.13, 4.1.2 Технического задания.

Испытание проводили

Ассистент	В.В.Глазкова
М.н.с.	И.С.Попов
Доцент	М.И.Петровский
Математик 1-й кат.	Д.В.Царёв
Математик 1-й кат.	П.М.Саликов
Программист	Д.А.Никифоров
Математик 1-й кат	О.Е.Горохов

ПРОТОКОЛ №7

испытания по пункту № 4.3.1.1

Программы экспериментальных исследований

2015.291215.ПМ1

№ _4__ 27июня 2016 г.

1. Объект испытания: «Подсистема 1» ЭО ПК для сбора и анализа биометрической информации об особенностях динамики работы пользователя с устройствами ввода-вывода в рамках стандартного человеко-машинного интерфейса.

2. Цель испытания: проверка соответствия объекта экспериментального исследования требованиям пункта № 3.11.5, 3.13, 4.1.1 Технического задания.

3. Дата начала испытания: 27 апреля 2016 г.

4. Дата окончания испытания: 27 апреля 2016 г.

5. Место проведения испытания: лаборатория технологий программирования факультета Вычислительной Математики и Кибернетики МГУ им. М.В.Ломоносова

6. Средства проведения испытаний

Сервер SRV_SYS1: программные характеристики (ОС Microsoft Windows 2012R2, 64bit; интерпретатор Python3;модуль Python pandas 0.16.2; модуль Python numpy 1.9.3; модуль user-agent 1.0.1), аппаратные характеристики (два процессора x64 с тактовой частотой 2.4 Ггц, 4 ядра; оперативная память объемом 64 Гбайт; два дисковый накопителя HDD, объемом 2Тбайт в режиме RAID1; два дисковый накопителя HDD, объемом 1Тбайт; два сетевых адаптера с пропускной способностью 1 Гбит)

7. Результаты испытания приведены в таблице К.1.

Таблица К.1 — Результаты экспериментов по скорости предобработки данных для статической аутентификации с подсистемой 1

Наименование параметра	Ед.изм.	Номера пунктов			Требование к параметру		Измеренное значение
		Технического задания	Программы испытаний	Методики испытаний	Номинальное значение	Предельное отклонение	
Скорость предобработки данных (ввод графической подписи)	сек	3.11.5, 3.13, 4.1.1	4.3.1.1	6.3.1	0.2	0.15	0.1

8. Замечания и рекомендации отсутствуют

9. Выводы

9.1 Объект испытания «Подсистема 1» ЭО ПК выдержал испытание по пункту № 4.3.1.1 Программы и методики 2015.291215.ПМ1.

9.2 Объект испытания «Подсистема 1» ЭО ПК соответствует требованиям пункта № 3.11.5, 3.13, 4.1.1 Технического задания.

Испытание проводили

Ассистент В.В.Глазкова

М.н.с. И.С.Попов

Математик 1-й кат. Д.В.Царёв

ПРОТОКОЛ №8

испытания по пункту № 4.3.1.2

Программы экспериментальных исследований

2015.291215.ПМ1

№ _4__ 27 июня 2016 г.

1. Объект испытания: «Подсистема 1» ЭО ПК для сбора и анализа биометрической информации об особенностях динамики работы пользователя с устройствами ввода-вывода в рамках стандартного человеко-машинного интерфейса.

2. Цель испытания: проверка соответствия объекта экспериментального исследования требованиям пункта № 3.11.5,3.13, 4.1.3 Технического задания.

3. Дата начала испытания: 27 апреля 2016 г.

4. Дата окончания испытания: 27 апреля 2016 г.

5. Место проведения испытания: лаборатория технологий программирования факультета Вычислительной Математики и Кибернетики МГУ им. М.В.Ломоносова

6. Средства проведения испытаний

Сервер SRV_SYS1: программные характеристики (ОС Microsoft Windows 2012R2, 64bit; интерпретатор Python3;модуль Python pandas 0.16.2; модуль Python numpy 1.9.3; модуль user-agent 1.0.1), аппаратные характеристики (два процессора x64 с тактовой частотой 2.4 Ггц, 4 ядра; оперативная память объемом 64 Гбайт; два дисковый накопителя HDD, объемом 2Тбайт в режиме RAID1; два дисковый накопителя HDD, объемом 1Тбайт; два сетевых адаптера с пропускной способностью 1 Гбит)

7. Результаты испытания приведены в таблице Л.1.

Таблица Л.1 — Результаты экспериментов по точности статической аутентификации с подсистемой 1

Наименование параметра	Ед.изм.	Номера пунктов			Требование к параметру		Измеренное значение
		Технического задания	Программы испытаний	Методики испытаний	Номинальное значение	Предельное отклонение	
AUC (клавиатура, пароль1)	%	3.11.5, 3.13, 4.1.3	4.3.1.2	6.3.2	90	10	90.16
AUC (клавиатура, пароль2)	%	3.11.5, 3.13, 4.1.3	4.3.1.2	6.3.2	90	10	89.17
AUC (клавиатура, пароль3)	%	3.11.5, 3.13, 4.1.3	4.3.1.2	6.3.2	90	10	89.80
AUC (клавиатура, пароль4)	%	3.11.5, 3.13, 4.1.3	4.3.1.2	6.3.2	90	10	88.44
AUC (мышь)	%	3.11.5, 3.13, 4.1.3	4.3.1.2	6.3.2	90	10	84.0

8. Замечания и рекомендации отсутствуют.

9. Выводы

9.1 Объект испытания «Подсистема 1» ЭО ПК выдержал испытание по пункту № 4.3.1.2 Программы и методики 2015.291215.ПМ1.

9.2 Объект испытания «Подсистема 1» ЭО ПК соответствует требованиям пункта №3.11.5, 3.13, 4.1.3 Технического задания.

Испытание проводили

Ассистент	В.В.Глазкова
М.н.с.	И.С.Попов
Доцент	М.И.Петровский
Математик 1-й кат.	Д.В.Царёв
Математик 1-й кат.	П.М.Саликов
Программист	Д.А.Никифоров
Математик 1-й кат	О.Е.Горохов

ПРОТОКОЛ №9

испытания по пункту № 4.3.2.1

Программы экспериментальных исследований

2015.291215.ПМ1

№ _1__ 30 мая 2016 г.

1. Объект испытания: «Подсистема 2» ЭО ПК для сбора и анализа данных о работе пользователя с информационными и вычислительными ресурсами компьютерной системы.

2. Цель испытания: проверка соответствия объекта экспериментального исследования требованиям пункта № 3.11.5, 3.13, 4.1.2 Технического задания.

3. Дата начала испытания: 20 апреля 2016 г.

4. Дата окончания испытания: 30 апреля 2016 г.

5. Место проведения испытания: лаборатория технологий программирования факультета Вычислительной Математики и Кибернетики МГУ им. М.В.Ломоносова

6. Средства проведения испытаний

Рабочая станции WST_WIN7_MON: программные характеристики - ОС Windows 7 (32bit), Python 2.6, библиотека OpenSSL; аппаратные характеристики: процессор Core2 Duo, частота 2.4 ГГц, 2 ядра; оперативная память 1Гбайт;дисковый накопитель HDD, 80 Гбайт; сетевой адаптер 100Мбит/сек.

7. Результаты испытания приведены в таблице М.1.

Таблица М.1 — Результаты экспериментов по скорости сбора и объемам собираемых данных с подсистемой 2

Наименование параметра	Ед.изм.	Номера пунктов			Требование к параметру		Измеренное значение
		Технического задания	Программы испытаний	Методики испытаний	Номинальное значение	Предельное отклонение	
Скорость сбора данных (минимальная, тип 1)	событий/сутки	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	500	150	598

Скорость сбора данных (средняя, тип 1)	событий/сутки	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	1000	200	1154
Скорость сбора данных (максимальная, тип 1)	событий/сутки	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	1500	300	1708
Скорость сбора данных (минимальная, тип 2)	событий/сутки	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	100	100	0
Скорость сбора данных (средняя, тип 2)	событий/сутки	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	300	100	203
Скорость сбора данных (максимальная, тип 2)	событий/сутки	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	1000	100	1016
Скорость сбора данных (минимальная, тип 3)	событий/сутки	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	75	25	63
Скорость сбора данных	событий/сутки	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	200	50	237

(средняя, тип 3)							
Скорость сбора данных (максимальная, тип 3)	событий/сутки	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	500	100	411
Скорость сбора данных (минимальная, тип 4)	событий/сутки	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	5000	2000	3076
Скорость сбора данных (средняя, тип 4)	событий/сутки	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	12500	4000	12279
Скорость сбора данных (максимальная, тип 4)	событий/сутки	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	20000	6000	20260
Скорость сбора данных (минимальная, тип 5)	событий/сутки	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	100	33	98
Скорость сбора данных (средняя, тип 5)	событий/сутки	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	300	100	305
Скорость	событий/су	3.11.5,	4.3.2.1	6.4.1	500	167	495

сбора данных (максимальная, тип 5)	тки	3.13, 4.1.2					
Скорость сбора данных (минимальная, тип 6)	событий/сутки	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	1	1	0
Скорость сбора данных (средняя, тип 6)	событий/сутки	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	3	2	1
Скорость сбора данных (максимальная, тип 6)	событий/сутки	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	5	3	2
Объем данных (минимальный, тип 1, в локальном хранилище)	Мб/сутки	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	0,300	0,100	0,247
Объем данных (средний, тип 1, в локальном хранилище)	Мб/сутки	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	0,500	0,200	0,469
Объем данных	Мб/сутки	3.11.5, 3.13,	4.3.2.1	6.4.1	0,700	0,300	0,691

(максимальный, тип 1, в локальном хранилище)		4.1.2					
Объем данных (минимальный, тип 1, на сервере консолидации)	Мб/сутки	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	0,400	0,100	0,404
Объем данных (средний, тип 1, на сервере консолидации)	Мб/сутки	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	0,700	0,200	0,614
Объем данных (максимальный, тип 1, на сервере консолидации)	Мб/сутки	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	1,000	0,300	0,812
Объем данных (минимальный, тип 2, тип 3, тип 4, тип 5, в локальном хранилище)	Мб/сутки	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	2,000	1,000	1,576
Объем		3.11.5,	4.3.2.1	6.4.1	4,000	2,000	5,217

данных (средний, тип 2, тип 3, тип 4, тип 5, в локальном хранилище)	Мб/сутки	3.13, 4.1.2					
Объем данных (максимальный, тип 2, тип 3, тип 4, тип 5, в локальном хранилище)	Мб/сутки	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	6,000	3,000	8,880
Объем данных (минимальный, тип 2, тип 3, тип 4, тип 5, на сервере консолидации)	Мб/сутки	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	4,00	1,00	3,60
Объем данных (средний, тип 2, тип 3, тип 4, тип 5, на сервере консолидации)	Мб/сутки	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	6,00	1,50	5,33
Объем данных (максимал	Мб/сутки	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	8,00	2,00	7,31

ьный, тип 2, тип 3, тип 4, тип 5, на сервере консолидации)							
Объем данных (минимальный, тип 6, в локальном хранилище)	Мб/сутки	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	0,010	0,010	0,008
Объем данных (средний, тип 6, в локальном хранилище)	Мб/сутки	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	0,010	0,010	0,008
Объем данных (максимальный, тип 6, в локальном хранилище)	Мб/сутки	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	0,010	0,010	0,009
Объем данных (минимальный, тип 6, на сервере консолидации)	Мб/сутки	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	0,020	0,010	0,029
Объем данных	Мб/сутки	3.11.5, 3.13,	4.3.2.1	6.4.1	0,030	0,010	0,031

(средний, тип 6, на сервере консолидации)		4.1.2					
Объем данных (максимальный, тип 6, на сервере консолидации)	Мб/сутки	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	0,040	0,010	0,032

Типы собираемых данных.

Тип 1: факты запуска пользователем программ и приложений

Тип 2: факты работы пользователя с внешними запоминающими устройствами (запись и чтение информации)

Тип 3: факты обращения пользователя к файлам и папкам, в том числе на удаленных АРМ

Тип 4: факты работы пользователя в ЛВС

Тип 5: события работы пользователя с клавиатурой и манипулятором «мышь» в каждом из активных приложений

Тип 6: факты самостоятельной установки и удаления пользователем программного обеспечения и/или технических средств на АРМ

8. Замечания и рекомендации отсутствуют

9. Выводы

9.1 Объект испытания «Подсистема 2» ЭО ПК выдержал испытание по пункту № 4.3.2.1 Программы и методики 2015.291215.ПМ1.

9.2 Объект испытания «Подсистема 2» ЭО ПК соответствует требованиям пункта №3.11.5, 3.13, 4.1.2 Технического задания.

Испытание проводили		
Ассистент		В.В.Глазкова
М.н.с.		И.С.Попов
Доцент		М.И.Петровский
Математик 1-й кат.		Д.В.Царёв
Математик 1-й кат.		П.М.Саликов
Программист		Д.А.Никифоров
Математик 1-й кат		О.Е.Горохов

ПРОТОКОЛ №10

испытания по пункту № 4.3.2.2

Программы экспериментальных исследований

2015.291215.ПМ1

№ _1__ 30 апреля 2016 г.

1. Объект испытания: «Подсистема 2» ЭО ПК для сбора и анализа данных о работе пользователя с информационными и вычислительными ресурсами компьютерной системы.

2. Цель испытания: проверка соответствия объекта экспериментального исследования требованиям пунктов № 3.11.5, 3.13, 4.1.2 Технического задания.

3. Дата начала испытания: 20 апреля 2016 г.

4. Дата окончания испытания: 30 апреля 2016 г.

5. Место проведения испытания: лаборатория технологий программирования факультета Вычислительной Математики и Кибернетики МГУ им. М.В.Ломоносова

6. Средства проведения испытаний

Сервер SRV_SYS2 - программные характеристики: ОС Microsoft Windows 2008R2, 64bit; Microsoft Office 2007;Microsoft SQL Server 2012; Microsoft SQL Server 2005;Microsoft SQL Server 2005 Analysis Services;Microsoft SQL Server 2005 Management Studio;интерпретатор Python 2.7.3; модуль Python pywin 219;библиотека Python Win32 Extensions; библиотека Natural Language Toolkit (NLTK); аппаратные характеристики: два процессора x64 с тактовой частотой 2 Ггц, 4 ядра; оперативная память объемом 48 Гбайт; два дисковый накопителя HDD, объемом 2Тбайт;два сетевых адаптера с пропускной способностью 1 Гбит.

7. Результаты испытания приведены в таблице Н.1.

Таблица Н.1 — Результаты экспериментов по времени построения и применения моделей с подсистемой 2

Наименование параметра	Ед.изм.	Номера пунктов			Требование к параметру		Измеренное значение
		Технического задания	Программы испытаний	Методики испытаний	Номинальное значение	Предельное отклонение	
Время		3.11.5, 3.13,	4.3.2.2	6.4.2	1.5	0.5	1.6

построения отчета	сек	4.1.2					
Время построения модели идентификации	сек	3.11.5, 3.13, 4.1.2	4.3.2.2	6.4.2	40	10	45
Число анализируемых событий при построении модели идентификации	событий	3.11.5, 3.13, 4.1.2	4.3.2.2	6.4.2	1500	500	12623
Время применения модели идентификации	сек	3.11.5, 3.13, 4.1.2	4.3.2.2	6.4.2	30	10	33
Размер временного окна при построении модели идентификации	событий	3.11.5, 3.13, 4.1.2	4.3.2.2	6.4.2	250	100	300
Время построения модели раннего обнаружения вторжений (тип событий 4)	сек	3.11.5, 3.13, 4.1.2	4.3.2.2	6.4.2	10	3	9
Число анализируемых событий при построении модели раннего обнаружения вторжений (тип событий 4)	событий	3.11.5, 3.13, 4.1.2	4.3.2.2	6.4.2	5000	1000	5062
Время применения модели раннего обнаружения вторжений	сек	3.11.5, 3.13, 4.1.2	4.3.2.2	6.4.2	35	10	39

(тип событий4)							

Типы анализируемых событий.

Тип событий 4: события работы пользователя в ЛВС

8. Замечания и рекомендации отсутствуют.

9. Выводы

9.1 Объект испытания «Подсистема 2» ЭО ПК выдержал испытание по пункту № 4.3.2.2 Программы и методики 2015.291215.ПМ1.

9.2 Объект испытания «Подсистема 2» ЭО ПК соответствует требованиям пунктов № 3.11.5, 3.13, 4.1.2 Технического задания.

Испытание проводили

Ассистент	В.В.Глазкова
М.н.с.	И.С.Попов
Доцент	М.И.Петровский
Математик 1-й кат.	Д.В.Царёв
Математик 1-й кат.	П.М.Саликов
Программист	Д.А.Никифоров
Математик 1-й кат	О.Е.Горохов

ПРОТОКОЛ №11

испытания по пункту № 4.3.2.1

Программы экспериментальных исследований

2015.291215.ПМ1

№ _2__ 30 апреля 2016 г.

1. Объект испытания: «Подсистема 2» ЭО ПК для сбора и анализа данных о работе пользователя с информационными и вычислительными ресурсами компьютерной системы.

2. Цель испытания: проверка соответствия объекта экспериментального исследования требованиям пункта № 3.11.5, 3.13, 4.1.2 Технического задания.

3. Дата начала испытания: 20 апреля 2016 г.

4. Дата окончания испытания: 30 апреля 2016 г.

5. Место проведения испытания: лаборатория технологий программирования факультета Вычислительной Математики и Кибернетики МГУ им. М.В.Ломоносова

6. Средства проведения испытаний

Рабочая станции WST_WIN7_MON: программные характеристики - ОС Windows 7 (32bit), Python 2.6, библиотека OpenSSL; аппаратные характеристики:процессор Core2 Duo, частота 2.4 ГГц, 2 ядра; оперативная память 1Гбайт;дисковый накопитель HDD, 80 Гбайт; сетевой адаптер 100Мбит/сек.

7. Результаты испытания приведены в таблице П.1.

Таблица П.1 — Результаты экспериментов по времени обработки событий с подсистемой 2

Наименование параметра	Ед.изм.	Номера пунктов			Требование к параметру		Измеренное значение
		Технического задания	Программы испытаний	Методики испытаний	Номинальное значение	Предельное отклонение	
Время обработки событий (минимальное, тип 1,	сек	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	0,010	0,010	1*10^(-7)

тип5, тип 6, стратегия 1)							
Время обработки событий (среднее, тип 1, тип5, тип 6 стратегия 1)	сек	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	1,000	1,000	1,797
Время обработки событий (максимальное, тип 1, тип5, тип 6 стратегия 1)	сек	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	55,000	20,000	68,539
Время обработки событий (минимальное, тип 1, тип5, тип 6, стратегия 2)	сек	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	0,010	0,010	1*10^(-7)
Время обработки событий (среднее, тип 1, тип5, тип 6,	сек	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	1,000	1,000	0,573

стратегия 2)							
Время обработки событий (максимальное, тип 1, тип5, тип 6, стратегия 2)	сек	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	55,000	20,000	38,668
Время обработки событий (минимальное, тип 1, тип5, тип 6, стратегия 3)	сек	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	0,010	0,010	1*10^(-7)
Время обработки событий (среднее, тип 1, тип5, тип 6, стратегия 3)	сек	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	2,000	2,000	2,275
Время обработки событий (максимальное, тип 1, тип5, тип 6, стратегия 3)	сек	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	220,000	100,000	214,605
Время		3.11.5,	4.3.2.1	6.4.1	0,010	0,010	1*10^(-7)

обработки событий (минимальное, тип 2, тип 3, стратегия 1)	сек	3.13, 4.1.2					
Время обработки событий (среднее, тип 2, тип 3, стратегия 1)	сек	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	0,100	0,100	0,002
Время обработки событий (максимальное, тип 2, тип 3, стратегия 1)	сек	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	2,000	2,000	1,209
Время обработки событий (минимальное, тип 2, тип 3, стратегия 2)	сек	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	0,010	0,010	1*10^(-7)
Время обработки событий (среднее, тип 2, тип 3, стратегия	сек	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	0,100	0,100	0,002

2)							
Время обработки событий (максимальное, тип 2, тип 3, стратегия 2)	сек	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	2,000	2,000	1,272
Время обработки событий (минимальное, тип 2, тип 3, стратегия 3)	сек	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	0,010	0,010	1*10^(-7)
Время обработки событий (среднее, тип 2, тип 3, стратегия 3)	сек	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	0,100	0,100	0,003
Время обработки событий (максимальное, тип 2, тип 3, стратегия 3)	сек	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	2,000	2,000	1,108
Время обработки событий (минимальное, тип 4, стратегия	сек	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	0,010	0,010	1*10^(-7)

1)							
Время обработки событий (среднее, тип 4, стратегия 1)	сек	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	4,000	1,000	3,870
Время обработки событий (максимальное, тип 4, стратегия 1)	сек	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	120,000	60,000	168,233
Время обработки событий (минимальное, тип 4, стратегия 2)	сек	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	0,010	0,010	1*10^(-7)
Время обработки событий (среднее, тип 4, стратегия 2)	сек	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	4,000	1,000	3,755
Время обработки событий (максимальное, тип 4, стратегия 2)	сек	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	120,000	60,000	156,401

Время обработки событий (минимальное, тип 4, стратегия 3)	сек	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	0,010	0,010	1*10^(-7)
Время обработки событий (среднее, тип 4, стратегия 3)	сек	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	8,000	2,000	9,888
Время обработки событий (максимальное, тип 4, стратегия 3)	сек	3.11.5, 3.13, 4.1.2	4.3.2.1	6.4.1	240,000	120,000	214,605

Стратегии передачи информации.

Стратегия 1: фиксированный объем информации (агент накапливает определенный объем информации или фиксированное количество записей журналов и затем передает их на сервер консолидации)

Стратегия 2: через равные промежутки (агент через равные промежутки времени передает всю имеющуюся у него в локальном хранилище информацию независимо от ее объема)

Стратегия 3: в режиме реального времени (агент немедленно передает информацию о каждой вновь прочитанной записи в журнале регистрации)

Типы собираемых данных.

Тип 1: факты запуска пользователем программ и приложений

Тип 2: факты работы пользователя с внешними запоминающими устройствами (запись и чтение информации)

Тип 3: факты обращения пользователя к файлам и папкам, в том числе на удаленных АРМ

Тип 4: факты работы пользователя в ЛВС

Тип 5: события работы пользователя с клавиатурой и манипулятором «мышь» в каждом из активных приложений

Тип 6: факты самостоятельной установки и удаления пользователем программного обеспечения и/или технических средств на АРМ

8. Замечания и рекомендации

Отсутствуют

9. Выводы

9.1 Объект испытания «Подсистема 2» ЭО ПК выдержал испытание по пункту № 4.3.2.1 Программы и методики 2015.291215.ПМ1.

9.2 Объект испытания «Подсистема 2» ЭО ПК соответствует требованиям пункта №3.11.5,

3.13, 4.1.2 Технического задания.

Испытание проводили

Ассистент	В.В.Глазкова
М.н.с.	И.С.Попов
Доцент	М.И.Петровский
Математик 1-й кат.	Д.В.Царёв
Математик 1-й кат.	П.М.Саликов
Программист	Д.А.Никифоров
Математик 1-й кат	О.Е.Горохов

ПРОТОКОЛ №12

испытания по пункту № 4.3.2.2

Программы экспериментальных исследований

2015.291215.ПМ1

№ _2__ 27 июня 2016 г.

1. Объект испытания: «Подсистема 2» ЭО ПК для сбора и анализа данных о работе пользователя с информационными и вычислительными ресурсами компьютерной системы.

2. Цель испытания: проверка соответствия объекта экспериментального исследования требованиям пункта № 3.11.5, 3.13, 4.1.2 Технического задания.

3. Дата начала испытания: 20 апреля 2016 г.

4. Дата окончания испытания: 30 апреля 2016 г.

5. Место проведения испытания: лаборатория технологий программирования факультета Вычислительной Математики и Кибернетики МГУ им. М.В.Ломоносова

6. Средства проведения испытаний

Сервер SRV_SYS2 - программные характеристики: ОС Microsoft Windows 2008R2, 64bit; Microsoft Office 2007;Microsoft SQL Server 2012; Microsoft SQL Server 2005;Microsoft SQL Server 2005 Analysis Services;Microsoft SQL Server 2005 Management Studio;интерпретатор Python 2.7.3; модуль Python pywin 219;библиотека Python Win32 Extensions; библиотека Natural Language Toolkit (NLTK); аппаратные характеристики: два процессора x64 с тактовой частотой 2 Ггц, 4 ядра; оперативная память объемом 48 Гбайт; два дисковый накопителя HDD, объемом 2Тбайт;два сетевых адаптера с пропускной способностью 1 Гбит.

7. Результаты испытания приведены в таблице Р.1.

Таблица Р.1 — Результаты экспериментов по точности моделей с подсистемой 2.

Наименование параметра	Ед.изм.	Номера пунктов			Требование к параметру		Измеренное значение
		Технического задания	Программы испытаний	Методики испытаний	Номинальное значение	Предельное отклонение	
Точность модели идентификации (ROC AUC)	%	3.11.5, 3.13, 4.1.2	4.3.2.2	6.4.2	90	10	88
Точность модели раннего обнаружения вторжений (ROC AUC, события работы пользователя в ЛВС)	%	3.11.5, 3.13, 4.1.2	4.3.2.2	6.4.2	90	10	94

8. Замечания и рекомендации отсутствуют

9. Выводы

9.1 Объект испытания «Подсистема 2» ЭО ПК выдержал испытание по пункту № 4.3.2.2 Программы и методики 2015.291215.ПМ1.

9.2 Объект испытания «Подсистема 2» ЭО ПК соответствует требованиям пункта №3.11.5, 3.13, 4.1.2 Технического задания.

Испытание проводили

Ассистент В.В.Глазкова

М.н.с. И.С.Попов

Математик 1-й кат. Д.В.Царёв

ПРОТОКОЛ №13

испытания по пункту № 4.3.3.1

Программы экспериментальных исследований

2015.291215.ПМ1

№ _1__ 27 июня 2016 г.

1. Объект испытания: «Подсистема 3» ЭО ПК для сбора и анализа информации об особенностях работы пользователя с потоками текстовой информации.

2. Цель испытания: проверка соответствия объекта экспериментального исследования требованиям пунктов № 3.11.5, 3.13, 4.1.2 Технического задания.

3. Дата начала испытания: 20 апреля 2016 г.

4. Дата окончания испытания: 30 апреля 2016 г.

5. Место проведения испытания: лаборатория технологий программирования факультета Вычислительной Математики и Кибернетики МГУ им. М.В.Ломоносова

6. Средства проведения испытаний

Рабочая станция поддержки агентов мониторинга, сбора и предобработки (WST_WIN7_MON): программное обеспечение: ОС Windows 7 (32bit); Python 2.6; библиотека OpenSSL; аппаратные характеристики: процессор Core2 Duo (частота 2.4 ГГц, 2 ядра); оперативная память 1Гбайт; дисковый накопитель HDD, 80 Гбайт; сетевой адаптер 100Мбит/сек.

7. Результаты испытания приведены в таблице С.1.

Таблица С.1 — Результаты экспериментов по скорости сбора и предобработки данных с подсистемой 3

Наименовани	Ед.изм.	Номера пунктов	Требование к параметру	Измере

е параметра		Технического задания	Программы испытаний	Методики испытаний	Номинальное значение	Предельное отклонение	нное значение
Скорость сбора и предобработки (тест 1, серия 1)	секунды	3.11.5, 3.13, 4.1.2	4.3.3.1	6.5.1	85	10	86
Скорость сбора и предобработки (тест 2, серия 1)	секунды	3.11.5, 3.13, 4.1.2	4.3.3.1	6.5.1	86	10	87
Скорость сбора и предобработки (тест 3, серия 1)	секунды	3.11.5, 3.13, 4.1.2	4.3.3.1	6.5.1	140	20	134
Скорость сбора и предобработки (тест 1, серия 2)	секунды	3.11.5, 3.13, 4.1.2	4.3.3.1	6.5.1	87	10	88
Скорость сбора и предобработки (тест 2, серия 2)	секунды	3.11.5, 3.13, 4.1.2	4.3.3.1	6.5.1	88	10	90
Скорость		3.11.5,	4.3.3.1	6.5.1	110	20	107

сбора и предобработки (тест 3, серия 2)	секунды	3.13, 4.1.2					
Скорость сбора и предобработки (тест 1, серия 3)	секунды	3.11.5, 3.13, 4.1.2	4.3.3.1	6.5.1	616	10	616
Скорость сбора и предобработки (тест 2, серия 3)	секунды	3.11.5, 3.13, 4.1.2	4.3.3.1	6.5.1	616	10	616
Скорость сбора и предобработки (тест 3, серия 3)	секунды	3.11.5, 3.13, 4.1.2	4.3.3.1	6.5.1	616	20	616
Объем документов пользователя (среднее за сутки)	Мбайт	3.11.5, 3.13, 4.1.2	4.3.3.1	6.5.1	300	200	133.5
Объем собираемой поведенческой информации (среднее за сутки)	Мбайт	3.11.5, 3.13, 4.1.2	4.3.3.1	6.5.1	5	5	2.53

8. Замечания и рекомендации отсутствуют

9. Выводы

9.1 Объект испытания «Подсистема 3» ЭО ПК выдержал испытание по пункту № 4.3.3.1 Программы и методики 2015.291215.ПМ1.

9.2 Объект испытания «Подсистема 3» ЭО ПК соответствует требованиям пунктов № 3.11.5, 3.13, 4.1.2 Технического задания.

Испытание проводили

Ассистент		В.В.Глазкова
М.н.с.		И.С.Попов
Доцент		М.И.Петровский
Математик 1-й кат.		Д.В.Царёв
Математик 1-й кат.		П.М.Саликов
Программист		Д.А.Никифоров
Математик 1-й кат		О.Е.Горохов

ПРОТОКОЛ №14

испытания по пункту № 4.3.3.2

Программы экспериментальных исследований

2015.291215.ПМ1

№ _1__ 27 июня 2016 г.

1. Объект испытания: «Подсистема 3» ЭО ПК для сбора и анализа информации об особенностях работы пользователя с потоками текстовой информации.

2. Цель испытания: проверка соответствия объекта экспериментального исследования требованиям пунктов № 3.11.5, 3.13, 4.1.2 Технического задания.

3. Дата начала испытания: 17 марта 2016 г.

4. Дата окончания испытания: 18 марта 2016 г.

5. Место проведения испытания: лаборатория технологий программирования факультета Вычислительной Математики и Кибернетики МГУ им. М.В.Ломоносова

6. Средства проведения испытаний

Вычислительный сервер (Подсистема2/Подсистема3): серверы консолидации, АРМ аналитика, рабочее место администратора, хранилище, модули идентификации (SRV_SYS2):

Программное обеспечение: ОС Microsoft Windows 2008R2 (64bit); Microsoft Office 2007; Microsoft SQL Server 2012; Microsoft SQL Server 2005; Microsoft SQL Server 2005 Analysis Services; Microsoft SQL Server 2005 Management Studio; интерпретатор Python 2.7.3; модуль Python pywin 219; Библиотека Python Win32 Extensions; Библиотека Natural Language Toolkit (NLTK); аппаратные характеристики: два процессора x64 с тактовой частотой 2 Ггц, 4 ядра; оперативная память объемом 48 Гбайт; два дисковый накопителя HDD, объемом 2Тбайт; два сетевых адаптера с пропускной способностью 1 Гбит.

7. Результаты испытания приведены в таблице Т.1.

Таблица Т.1 — Результаты экспериментов по точности и скорости построения моделей аннотирования с подсистемой 3

Наименование параметра	Ед.изм.	Номера пунктов			Требование к параметру		Измеренное значение
		Технического задания	Программы испытаний	Методики испытаний	Номинальное значение	Предельное отклонение	
ROUGE-2-F (аннотирование)	-	3.11.5, 3.13, 4.1.2	4.3.3.2	6.5.2	0.192	0.002	0.19251
ROUGE-L-F (аннотирование)	-	3.11.5, 3.13, 4.1.2	4.3.3.2	6.5.2	0.372	0.002	0.37230
ROUGE-S4-F (аннотирование)	-	3.11.5, 3.13, 4.1.2	4.3.3.2	6.5.2	0.153	0.002	0.15358
ROUGE-W-F (аннотирование)	-	3.11.5, 3.13, 4.1.2	4.3.3.2	6.5.2	0.210	0.002	0.21066
Скорость работы метода автоматического аннотирования (для одного документа)	секунды	3.11.5, 3.13, 4.1.2	4.3.3.2	6.5.2	0.05	0.05	0.048

Hamming Loss (многотемная классификация)	-	3.11.5, 3.13, 4.1.2	4.3.3.2	6.5.2	0.008	0.001	0,0083677
Скорость работы метода рубрикации (для одного документа)	секунды	3.11.5, 3.13, 4.1.2	4.3.3.2	6.5.2	0.5	0.3	0.31
Purity (кластеризация)		3.11.5, 3.13, 4.1.2	4.3.3.2	6.5.2	0.5	0.1	0.581519
NMI (кластеризация)		3.11.5, 3.13, 4.1.2	4.3.3.2	6.5.2	0.5	0.15	0.604803
Rand Index (кластеризация)		3.11.5, 3.13, 4.1.2	4.3.3.2	6.5.2	0.9	0.05	0.934813
Скорость работы метода группировки (для набора из 18828 документов)	секунды	3.11.5, 3.13, 4.1.2	4.3.3.2	6.5.2	60	10	56

8. Замечания и рекомендации отсутствуют

9. Выводы

9.1 Объект испытания «Подсистема 3» ЭО ПК выдержал испытание по пункту № 4.3.3.2

Программы и методики 2015.291215.ПМ1.

9.2 Объект испытания «Подсистема 3» ЭО ПК соответствует требованиям пунктов № 3.11.5, 3.13, 4.1.2 Технического задания.

Испытание проводили		
Ассистент		В.В.Глазкова
М.н.с.		И.С.Попов
Доцент		М.И.Петровский
Математик 1-й кат.		Д.В.Царёв
Математик 1-й кат.		П.М.Саликов
Программист		Д.А.Никифоров
Математик 1-й кат		О.Е.Горохов

ПРОТОКОЛ №15

испытания по пункту № 4.3.3.3

Программы экспериментальных исследований

2015.291215.ПМ1

№ _1__ 27 июня 2016 г.

1. Объект испытания: «Подсистема 3» ЭО ПК для сбора и анализа информации об особенностях работы пользователя с потоками текстовой информации.

2. Цель испытания: проверка соответствия объекта экспериментального исследования требованиям пунктов № 3.11.5, 3.13, 4.1.2 Технического задания.

3. Дата начала испытания: 20 апреля 2016 г.

4. Дата окончания испытания: 30 апреля 2016 г.

5. Место проведения испытания: лаборатория технологий программирования факультета Вычислительной Математики и Кибернетики МГУ им. М.В.Ломоносова

6. Средства проведения испытаний

Вычислительный сервер (Подсистема2/Подсистема3): серверы консолидации, АРМ аналитика, рабочее место администратора, хранилище, модули идентификации (SRV_SYS2):

Программное обеспечение: ОС Microsoft Windows 2008R2 (64bit); Microsoft Office 2007; Microsoft SQL Server 2012; Microsoft SQL Server 2005; Microsoft SQL Server 2005 Analysis Services; Microsoft SQL Server 2005 Management Studio; интерпретатор Python 2.7.3; модуль Python pywin 219; Библиотека Python Win32 Extensions; Библиотека Natural Language Toolkit (NLTK); аппаратные характеристики: два процессора x64 с тактовой частотой 2 Ггц, 4 ядра; оперативная память объемом 48 Гбайт; два дисковый накопителя HDD, объемом 2Тбайт; два сетевых адаптера с пропускной способностью 1 Гбит.

7. Результаты испытания приведены в таблице У.1.

Таблица У.1 — Результаты экспериментов по точности и скорости построения моделей идентификации и раннего обнаружения попыток хищения с подсистемой 3

Наименование параметра	Ед.изм.	Номера пунктов			Требование к параметру		Измеренное значение
		Технического задания	Программы испытаний	Методики испытаний	Номинальное значение	Предельное отклонение	

Скорость работы тематического моделирования	секунды	3.11.5, 3.13, 4.1.2	4.3.3.3	6.5.3	3	0.5	3.2
Скорость работы отображения документов в модельное тематическое пространство	секунды	3.11.5, 3.13, 4.1.2	4.3.3.3	6.5.3	2.5	0.5	2.4
Скорость работы прогнозирования многомерного временного ряда	секунды	3.11.5, 3.13, 4.1.2	4.3.3.3	6.5.3	0.1	0.05	0,07
Оценка точности идентификации пользователей (AUC)	-	3.11.5, 3.13, 4.1.2	4.3.3.3	6.5.3	0.9	0.03	0.889
Оценка точности раннего обнаружения попыток хищения конфиденциальной информации (AUC)	-	3.11.5, 3.13, 4.1.2	4.3.3.3	6.5.3	0.9	0.03	0,906

8. Замечания и рекомендации отсутствуют.

9. Выводы

9.1 Объект испытания «Подсистема 3» ЭО ПК выдержал испытание по пункту № 4.3.3.3

Программы и методики 2015.291215.ПМ1.

9.2 Объект испытания «Подсистема 3» ЭО ПК соответствует требованиям пунктов № 3.11.5, 3.13, 4.1.2 Технического задания.

Испытание проводили

Ассистент		В.В.Глазкова
М.н.с.		И.С.Попов
Математик 1-й кат.		Д.В.Царёв

ПРОТОКОЛ №16

испытания по пункту № *4.3.1.3*

Программы *экспериментальных исследований*

2015.291215.ПМ1

№ *_1__* *17 июня* 2016 г.

1. Объект испытания: «Подсистема 1» ЭО ПК для сбора и анализа биометрической информации об особенностях динамики работы пользователя с устройствами ввода-вывода в рамках стандартного человеко-машинного интерфейса.

2. Цель испытания: проверка соответствия объекта экспериментального исследования требованиям пунктам № 3.11.5, 4.1.2 Технического задания Соглашения о предоставлении субсидии № 14.604.21.0056 от 27.06.2014г.

3. Дата начала испытания: 10 января 2016 г.

4. Дата окончания испытания: 27 июня 2016 г.

5. Место проведения испытания: лаборатория технологий программирования факультета Вычислительной Математики и Кибернетики МГУ им. М.В.Ломоносова

6. Средства проведения испытаний

Рабочая станции WST_WIN7: программные характеристики - ОС Windows 7 (32bit); аппаратные характеристики: процессор Core2 Duo, частота 2.4 ГГц, 2 ядра; оперативная память 1Гбайт; дисковый накопитель HDD, 80 Гбайт; сетевой адаптер 100Мбит/сек.

7. Результаты испытания. Согласно требованиям пункта №4.3.1.3 программы экспериментальных исследований, были выполнены контрольные примеры в соответствие с пунктом 6.3.3. методики испытаний.

8. Замечания и рекомендации отсутствуют

9. Выводы

9.1 Объект испытания *«Подсистема 1» ЭО ПК выдержал* испытание по пункту № *4.3.1.3* Программы и методики *2015.291215.ПМ1.*

8. Замечания и рекомендации отсутствуют

9.2 Объект испытания «Подсистема 2» ЭО ПК соответствует требованиям пунктам № 3.11.5, 4.1.2 Технического задания.

Испытание проводили		
Ассистент		В.В.Глазкова
М.н.с.		И.С.Попов
Доцент		М.И.Петровский
Математик 1-й кат.		Д.В.Царёв
Математик 1-й кат.		П.М.Саликов
Программист		Д.А.Никифоров
Математик 1-й кат		О.Е.Горохов

ПРОТОКОЛ №17

испытания по пункту № *4.3.2.3*

Программы *экспериментальных исследований*

2015.291215.ПМ1

№ *_1__* *18 марта* 2016 г.

1. Объект испытания: «Подсистема 2» ЭО ПК для сбора и анализа биометрической информации об особенностях динамики работы пользователя с устройствами ввода-вывода в рамках стандартного человеко-машинного интерфейса.

2. Цель испытания: проверка соответствия объекта экспериментального исследования требованиям пунктам № 3.11.5, 4.1.2 Технического задания Соглашения о предоставлении субсидии № 14.604.21.0056 от 27.06.2014г.

3. Дата начала испытания: 17 марта 2016 г.

4. Дата окончания испытания: 18 марта 2016 г.

5. Место проведения испытания: лаборатория технологий программирования факультета Вычислительной Математики и Кибернетики МГУ им. М.В.Ломоносова

6. Средства проведения испытаний

Рабочая станции WST_WIN7: программные характеристики - ОС Windows 7 (32bit); аппаратные характеристики: процессор Core2 Duo, частота 2.4 ГГц, 2 ядра; оперативная память 1Гбайт; дисковый накопитель HDD, 80 Гбайт; сетевой адаптер 100Мбит/сек.

7. Результаты испытания. Согласно требованиям пункта №4.3.2.3 программы экспериментальных исследований, были выполнены контрольные примеры в соответствие с пунктом 6.4.3. методики испытаний.

8. Замечания и рекомендации отсутствуют

9. Выводы

9.1 Объект испытания *«Подсистема 2» ЭО ПК* выдержал испытание по пункту № *4.3.2.3* Программы и методики *2015.291215.ПМ1.*

8. Замечания и рекомендации отсутствуют

9.2 Объект испытания «Подсистема 2» ЭО ПК соответствует требованиям пунктов № 3.11.5, 4.1.2 Технического задания.

Испытание проводили		
Ассистент		В.В.Глазкова
М.н.с.		И.С.Попов
Доцент		М.И.Петровский
Математик 1-й кат.		Д.В.Царёв
Математик 1-й кат.		П.М.Саликов
Программист		Д.А.Никифоров
Математик 1-й кат		О.Е.Горохов

ПРОТОКОЛ №18

испытания по пункту № *4.3.3.4*

Программы *экспериментальных исследований*

2015.291215.ПМ1

№ *_1__* *18 марта* 2016 г.

1. Объект испытания: «Подсистема 3» ЭО ПК для сбора и анализа биометрической информации об особенностях динамики работы пользователя с устройствами ввода-вывода в рамках стандартного человеко-машинного интерфейса.

2. Цель испытания: проверка соответствия объекта экспериментального исследования требованиям пунктам № 3.11.5, 4.1.2 Технического задания Соглашения о предоставлении субсидии № 14.604.21.0056 от 27.06.2014г.

3. Дата начала испытания: 17 марта 2016 г.

4. Дата окончания испытания: 18 марта 2016 г.

5. Место проведения испытания: лаборатория технологий программирования факультета Вычислительной Математики и Кибернетики МГУ им. М.В.Ломоносова

6. Средства проведения испытаний

Рабочая станции WST_WIN7: программные характеристики - ОС Windows 7 (32bit); аппаратные характеристики: процессор Core2 Duo, частота 2.4 ГГц, 2 ядра; оперативная память 1Гбайт; дисковый накопитель HDD, 80 Гбайт; сетевой адаптер 100Мбит/сек.

7. Результаты испытания. Согласно требованиям пункта №4.3.3.4 программы экспериментальных исследований, были выполнены контрольные примеры в соответствие с пунктом 6.5.4 методики испытаний.

8. Замечания и рекомендации отсутствуют

9. Выводы

9.1 Объект испытания *«Подсистема 2» ЭО ПК* выдержал испытание по пункту № *4.3.3.4* Программы и методики *2015.291215.ПМ1.*

8. Замечания и рекомендации отсутствуют

9.2 Объект испытания «Подсистема 2» ЭО ПК соответствует требованиям пунктов № 3.11.5, 4.1.2 Технического задания.

Испытание проводили	
Ассистент	В.В.Глазкова
М.н.с.	И.С.Попов
Доцент	М.И.Петровский
Математик 1-й кат.	Д.В.Царёв
Математик 1-й кат.	П.М.Саликов
Программист	Д.А.Никифоров
Математик 1-й кат	О.Е.Горохов

ПРИЛОЖЕНИЕ Б

Акт исполнения обязательств по работам на Этапе №4 за счет ВБС

УТВЕРЖДАЮ

"30" июня 2016 г.

Декан факультета ВМК МГУ

Академик РАН Е.И.Моисеев

АКТ №1

от 30 июня 2016 г.

исполнения обязательств по работам на этапе № 4 Плана-графика по Соглашению о предоставлении субсидии № 14.604.21.0056 от 27.06.2014г., выполненных за счет внебюджетных средств, по теме: «Исследование и разработка инновационной технологии построения программных средств обеспечения компьютерной безопасности, основанных на использовании методов машинного обучения и математической статистики для анализа данных поведенческой биометрии пользователей при работе в рамках стандартного человеко-машинного интерфейса, для решения задач активной аутентификации и идентификации пользователей, обнаружения внутренних вторжений и предотвращения попыток хищения конфиденциальной информации»

Настоящий акт составлен в том, что работы по соглашению о предоставлении субсидии № 14.604.21.0056 от 27.06.2014г. предусмотренные планом-графиком исполнения обязательств, а именно:

1. обеспечение работоспособности рабочих станций, общего системного и специального программного обеспечения для проведения работ по проекту, проведение регламентных работ по обеспечению сохранности данных и программ, относящихся к проводимым работам по проекту;
2. проведение оценки РИД, полученных при выполнении ПНИ, с целью их вовлечения в хозяйственный оборот

проведены в полном объеме и надлежащем качестве за счет внебюджетных источников на сумму 1 000 000 (Один миллион) рублей.

Научный руководитель проекта

Профессор И.В.Машечкин

Главный бухгалтер факультета ВМК МГУ М.В.Сидорова

ПРИЛОЖЕНИЕ В

Отчет о патентных исследованиях

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ М.В.ЛОМОНОСОВА

(МГУ)

УДК 002.6
№ госрегистрации 114100140107
Инв.№

УТВЕРЖДАЮ
Заместитель проректора

_____ А.Э.Сазонов
«30» июня 2016 г.

ОТЧЕТ
О ПАТЕНТНЫХ ИССЛЕДОВАНИЯХ

Исследование и разработка инновационной технологии построения программных средств обеспечения компьютерной безопасности, основанных на использовании методов машинного обучения и математической статистики для анализа данных поведенческой биометрии пользователей при работе в рамках стандартного человеко-машинного интерфейса, для решения задач активной аутентификации и идентификации пользователей, обнаружения внутренних вторжений и предотвращения попыток хищения конфиденциальной информации

по теме:
ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ ПОСТАВЛЕННЫХ ПЕРЕД ПНИ ЗАДАЧ

Шифр 2014-14-576-0053-023

Руководитель ПНИ,
д.ф.-м.н., профессор

30.06.16
подпись, дата

И.В. Машечкин

Москва 2016

СПИСОК ИСПОЛНИТЕЛЕЙ

к.ф.-м.н., доцент	30.06.16 подпись, дата	М.И. Петровский (приложения Ф.А, Ф.Б, Ф.В, заключение)
к.ф.м.н., ассистент	30.06.16 подпись, дата	В.В.Глазкова (проведение патентного поиска, разделы Ф.1, Ф.2)
математик 1-й кат.	30.06.16 подпись, дата	Д.В.Царев (проведение патентного поиска, разделы Ф.1, Ф.2)

СОДЕРЖАНИЕ

ПЕРЕЧЕНЬ СОКРАЩЕНИЙ, УСЛОВНЫХ ОБОЗНАЧЕНИЙ, СИМВОЛОВ, ЕДИНИЦ, ТЕРМИНОВ

ЕРО	Европейская патентная организация
ВОИС	Всемирная организация интеллектуальной собственности
НИОКР	Научно-исследовательская и опытно-конструкторская разработка
НИР	Научно-исследовательская работа
МПК	Международная патентная классификация

ОБЩИЕ ДАННЫЕ ОБ ОБЪЕКТЕ ИССЛЕДОВАНИЯ

Дата начала работы «1» *марта* 2016г.

Дата окончания работы «29» *июня* 2016г.

Объектом исследования в рамках данной работы являются методы анализа данных поведенческой биометрии пользователей при работе в рамках стандартного человеко-машинного интерфейса для решения задач компьютерной безопасности, включая задачи активной аутентификации и идентификации пользователей, обнаружения внутренних вторжений и предотвращения попыток хищения конфиденциальной информации.

Областью применения объекта исследований является разработка комплекса научных решений, направленных на создание программных средств анализа индивидуальных особенностей поведения пользователей компьютерных систем (поведенческой биометрии) при работе в рамках стандартного человеко-машинного интерфейса, с целью создания инновационной технологии построения систем компьютерной безопасности.

В.1 Исследование патентной чистоты объекта техники

В рамках работ по п.4.3 ПГ проводились патентные исследования в соответствии с требованиями ГОСТ Р15.011-96. Ставилась задача по обоснованию целесообразности правовой охраны объектов интеллектуальной собственности в стране и за рубежом.

Патентный поиск проводился по базам данных Федеральной службы по интеллектуальной собственности, патентам и товарным знакам (Роспатента), Бюро по патентам и товарным знакам США (USPTO), Европейского патентного бюро (EPO) и Всемирной организации интеллектуальной собственности (ВОИС, или WIPO). Выбор указанных патентных баз обусловлен тем, что в соответствующих географических регионах активно проводились либо ведутся научно-исследовательские работы по исследуемому вопросу.

Во внимание при проведении поиска принимались официальные публикации охранных документов, таких как патенты на изобретения, авторские свидетельства (для Российской Федерации и Советского Союза), свидетельства на полезные модели (для Российской Федерации) и патенты на полезные модели, а также национальные и международные заявки на патенты. Кроме того, во внимание принимались акцептованные и неакцептованные патентные документы Европейского Патентного Ведомства, а также опубликованные международные заявки Международного бюро Всемирной организации интеллектуальной собственности. Документы отбирали в соответствии с регламентом поиска без каких-либо ограничений.

Временной интервал патентного поиска для целей настоящего патентного исследования также не ограничивался.

В целях исследования современного состояния и тенденций развития объекта исследований, определенного в следующем разделе, временной интервал поиска установили по 2016 год включительно, поскольку поданные патентные заявки, прежде всего международные, уже опубликованы в международных базах данных патентных документов.

При выборе ретроспективы поиска по патентным документам учитывалось, что максимальный срок действия патента на изобретение составляет 20 лет, с возможностью его продления на три года.

Кроме того, в соответствии с действующим патентным законодательством и международным правом в области патентования предусматривается включение в мировой технический уровень всех патентных документов, вне зависимости от даты приоритета, а

также статуса документа, то есть в мировой технический уровень должны быть включены все опубликованные на дату проведения патентного поиска документы, вне зависимости от того, действуют они, или срок их действия уже истек.

Выбор рубрик МПК произведен исходя из необходимости охватить все решения, имеющие сходные возможности или принцип действия, без ограничения каким-либо конкретным назначением.

Выбор групп осуществлен в соответствии с Международной Патентной Классификацией (в редакции 2015 года), с учетом того, что класс МПК присваивается патентному документу на дату подачи заявки.

Метод патентного исследования основан на изучении и обработке патентной документации, полученной в результате целенаправленного патентного поиска.

Патентный поиск проведен методом сплошного просмотра всего массива патентов, выявленных по выбранным подгруппам МПК без ограничения глубины поиска, с последующим анализом содержания каждого патентного документа и установления соответствия целям и задачам исследования.

Этот способ обеспечивает наиболее полный охват информации, и позволяет выявить практически все технические решения, относящиеся к области поиска, как уже охраняемые патентными документами, так и составляющие область потенциальной охраны в связи с установлением приоритета технического решения при подаче соответствующей патентной заявки.

Для повышения надежности результатов, и выявления ошибочно и неточно классифицированных патентных документов, проводили дополнительный поиск по ключевым словам для выявления релевантных документов, не отнесенных к отобранным подклассам МПК, или относящихся к смежным классам.

В целях наиболее полного охвата патентной документации массив документов, отобранных методом сплошного поиска, дополняли документами, цитированными заявителями в описании патентных документах и в отчетах о поиске по международным заявкам, а также в опубликованных отчетах о поиске по национальным заявкам; после объединения этих массивов повторы и нерелевантные документы исключали, а для наиболее релевантных описанную процедуру повторяли. Это также позволяет выявить и включить в рассмотрение источники, которые обычно используются активными разработчиками при создании и совершенствовании объектов, относящихся к теме исследования, и получить более полную и объективную картину существующих тенденций и предпосылок развития области

техники, а также получить точное представление о существующих технических проблемах, сдерживающих исследуемую область техники.

Патентный поиск проводился по следующему набору основных ключевых слов без ограничений по давности публикации патента:

- authentication system;
- machine learning;
- keyboard;
- mouse;
- behavior analysis;
- behavior model;
- keystroke dynamics;
- biometric;
- mouse dynamics;
- identity authentication;
- continuously verifying;
- abnormal data-inputting behavior;
- keystroke latencies;
- computer security;
- security awareness;
- computer awareness;
- information security;
- end-user security behaviors;
- security policy compliance;
- data mining;
- information retrieval;
- security information and event management;
- intrusion prevention;
- intrusion detection;
- log management;
- computer misuse;
- system misuse;
- anomaly based detection;

– signature based detection;
– misuse based detection;
– data loss prevention;
– topic modeling;
– orthogonal non-negative matrix factorizations;
– time series;
– prediction;
– category values timeline;
– topic time series;
– topic time line;
– topic (concept);
– text mining;

а также по их комбинациям и русским аналогам.

В результате патентного поиска найдено семь патентов, схожих с искомыми тематиками.

B.1.1 Американские патенты

Патент: US20150220723 A1. «Аутентификация пользователя на основе временной базы данных динамических изображений (User authentication using temporal knowledge of dynamic images)». Заявитель**:** IBM. Страна выдачи: США. Дата заявки: 23.01.2015. Номер заявки: US 14/604,655. Дата выдачи: 06.08.2015.

Краткое описание патента

В данном патенте представлен подход для аутентификации пользователей. Связанный с ним метод включает в себя отображение динамического изображения на экране дисплея, детектирования взаимодействия пользователя с отображаемым изображением и определение длительности обнаруженного взаимодействия с пользователем. Метод дополнительно включает в себя сравнение обнаруженного взаимодействия с пользователем и определенной длительности с сохраненными взаимодействиями с пользователем и хранимыми продолжительностями. Предложенный метод также включает в себя аутентификацию пользователя при определении того, что обнаруженное взаимодействие с пользователем соответствует сохраненным взаимодействиям и обнаруженная длительность соответствует сохраненной длительности. В одном варианте реализации способ дополнительно включает в

себя передачу обнаруженного взаимодействия с пользователем и определенной длительности на удаленное устройство. В таком варианте реализации метода стадия сравнения обнаруженного взаимодействия с пользователем и определенной длительности с сохраненными взаимодействиями с пользователем и сохраненными продолжительностями выполняется удаленным устройством.

Патент: US9122860 B2. «Устройство и метод для идентификации и аутентификации (Device and method for identification and authentication)». Заявитель**:** Yankey Information Co., Ltd. Taipei City (TW). Страна выдачи: США. Дата заявки: 24.04.2006. Номер заявки: US 14/309,061. Дата выдачи: 18.12.2013.

Краткое описание патента

В данном патенте описывается система и метод для идентификации и аутентификации пользователя при удаленном подключении к услуге через сеть. Система включает в себя криптографический процессор, один из криптографических ключей, средство хранения, дополнительные средства обработки и интерфейс для формирования и передачи уникального кода аутентификации, чтобы эмулировать нажатия клавиш через стандартный ввод посредством клиентского терминала. Код может быть передан только с помощью явной команды пользователя. В данном патенте предложен способ генерации выходного кода с пользовательского устройства, включающий в себя: прием с помощью пользовательского устройства, явную команду для генерации выходного кода; генерацию кода с помощью пользовательского устройства в ответ на явную команду и выходной код с помощью криптографического ключа. Формирование выходного кода включает в себя: сравнение уникального идентификатора с информацией, хранящейся в заранее определенном шаблоне, уникальный идентификатор для идентификации владельца устройства и информацию, идентифицирующую одного или более держателей пользовательского устройства; а также реакцию на уникальный идентификатор для сопоставления информации на основе создания динамического кода аутентификации, связанного с уникальным идентификатором владельца пользовательского устройства. Динамический код аутентификации имеет характеристики, которые позволяют рассчитать показатель потенциально мошеннического использования пользовательского устройства.

Патент: US8856904 B2. «Усиленная защита паролями (Enhancing password protection)». Заявитель**:** International Business Machines Corporation. Страна выдачи: США. Дата заявки: 23.02.2013. Номер заявки: US 13/774,191. Дата выдачи: 07.10.2014..

Краткое описание патента

В данном патенте описывается механизм для усиления защиты паролями. Предлагается сочетание пароля, который включает в себя динамический текст, со статическим паролем, полученным от пользователя. Проверка пользователя осуществляется на основе верификации комбинации паролей без динамического текста. В ответ на проверку, что комбинация паролей верифицируется без динамического текста, динамический текст фильтруется от комбинации паролей на основе идентифицирующего динамического предложения, выданного пользователю до объединения полученных паролей, формируя тем самым фильтрованный пароль. Отфильтрованный пароль затем аутентифицируется с использованием информации, хранящейся для данного пользователя. После валидации отфильтрованного пароля, пользователю предоставляется доступ в защищенную систему.

Патент: US8879415 B2. «Метод и система для аннотирования потока сетевой информации (Method and system for annotating network flow information)». Заявитель**:** Arbor Networks, Inc. Страна выдачи: США. Дата заявки: 01.03.2014. Номер заявки: US 13/782,776. Дата выдачи: 04.11.2014.

Краткое описание патента

В данном патенте представлено масштабируемое решение для мониторинга потока сетевой информации, которое принимает стандартные записи потока, экспортируемого из сетевых устройств, таких как маршрутизаторы, коммутаторы, межсетевые экраны, концентраторы и т.д., и помечает поток дополнительной информацией. Эта информация получается из целого ряда источников, в том числе Border Gateway Protocol (BGP), протокола Simple Network Management (SNMP), конфигурации пользователя и другого интеллектуального анализа потока. Эти аннотации добавляют информацию в поток данных и могут быть использованы для выполнения анализа потоков. Аннотированный поток затем пересылают конфигурируемому набору адресатов с использованием средств форматирования стандартной потока, например NetFlow Cisco System Inc. в одной из реализаций. Это позволяет передать аннотированный поток для обработки, и эта расширенная информация будет использоваться другими инструментами анализа потоков и анализа существующей инфраструктуры потока.

B.1.2 Российские патенты

Патент: РФ № 2316120. «Биометрическая система аутентификации». Заявитель**:** Корпорация "Самсунг Электроникс. Страна выдачи: Россия. Дата заявки: 12.05.2004. Номер заявки: 2316120. Дата выдачи: 27.01.2008.

Краткое описание патента

В данном патенте предлагается система для защиты каналов связи, реализующая способ аутентификации пользователя на основе биометрических данных посредством представления и выделения криптографического ключа и аутентификации пользователя на основе биометрических данных. Связывание биометрических данных и криптографического ключа основано на том, что извлечение ключа осуществляется только при участии биометрического объекта, проведения соответствующих измерений с целью получения биометрического образца, обработки и формирования биометрических данных, которые затем предъявляются для распознавания. Впоследствии ключ может быть использован, например, для шифрования/дешифрования. Техническим результатом заявленного изобретения является создание биометрической системы доступа и способа представления/выделения криптографического ключа и аутентификации пользователя на основе биометрии; повышение уровня секретности ключа, надежности, расширение функциональных возможностей и упрощение процесса формирования системы. Технический результат достигается тем, что ни биометрический шаблон, ни криптографический ключ пользователя не представлены в устройстве хранения информации явно - без предъявления биометрического образца и устройства хранения информации с сохраненной на нем ключом никакие криптографические операции с данными невозможны.

Патент: РФ № 2469397. «Способ биометрической аутентификации по почерку в компьютеризированной системе контроля доступа». Заявитель**:** Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Юго-западный государственный университет». Страна выдачи: Россия. Дата заявки: 30.09.2011. Номер заявки: 2469397. Дата выдачи: 10.12.2012.

Краткое описание патента

Данный патент относится к области обработки биометрических измерений, в частности к методам цифровой обработки рукописного текста. Представленный способ предназначен для аутентификации пользователя по рукописному почерку в системах контроля доступа. Техническим результатом является повышение надежности санкционированного доступа во

всех системах, требующих использования средств идентификации личности. Сущность способа заключается в том, что предварительно формируется матрица коэффициентов двумерного дискретного косинусного преобразования базы данных пользователей, допущенных в систему, на основе дискретизированных и квантованных отсчетов образцов рукописного почерка. А при допуске в систему нового пользователя из дискретизированных и квантованных отсчетов образца рукописного почерка этого пользователя формируют матрицу коэффициентов двумерного дискретного косинусного преобразования аналогично способу формирования эталонных образцов. На основе этой информации осуществляется сравнение и принятие решения об отнесении полученной записи к одной из эталонных.

Патент: РФ № 2347274. «Способы и аппарат для ограничения доступа к играм с использованием биометрических данных». Заявитель**:** Ай Джи Ти. Страна выдачи: Россия. Дата заявки: 11.03.2004. Номер заявки: US2004/007423. Дата выдачи: 25.10..2005.

Краткое описание патента

Данный патент относится к игровым системам, в которых могут приниматься биометрические данные, связанные с выбором варианта проведения игры на основе принятых биометрических данных. Техническим результатом является обеспечение санкционированного доступа к игре. Игровая система содержит: множество игровых серверов, причем каждый игровой сервер содержит контроллер, который включает в себя процессор, память и устройство ввода/вывода для обеспечения связи через сеть, и предназначен для обеспечения проведения соответствующей игры; и сервер веб-сайта, предназначенный для оперативной связи через сеть с удаленными игровыми серверами. Сервер веб-сайта содержит контроллер, предназначенный для обеспечения возможности передачи данных и для определения выбора игры игрока на одном удаленном игровом устройстве, а также для переноса операционного управления на один из игровых серверов, основываясь на выборе игры и на биометрических данных.

B.1.3 Международные и европейские патенты

Проведенный патентный поиск не выявил новых международных и европейских патентов (по сравнению с патентными исследованиями предыдущего этапа) по данной тематике.

B.2 Выводы

Патентный поиск проводился по базам данных Федеральной службы по интеллектуальной собственности, патентам и товарным знакам (Роспатента), Бюро по патентам и товарным знакам США (USPTO), Европейского патентного бюро (EPO) и Всемирной организации интеллектуальной собственности (ВОИС, или WIPO) по сформированному набору основных ключевых слов, относящихся к заданной тематике (включая их комбинации и русские аналоги).

Основная задача патентного поиска заключалась в отборе патентов, связанных с методами анализа данных поведенческой биометрии пользователей при работе в рамках стандартного человеко-машинного интерфейса, для решения задач компьютерной безопасности, включая задачи активной аутентификации и идентификации пользователей, обнаружения внутренних вторжений и предотвращения попыток хищения конфиденциальной информации. В результате патентного поиска найдено семь патентов, схожих с искомыми тематиками. Полученные результаты патентного исследования дают основание сделать следующие выводы.

Патент РФ № 2316120 описывает биометрическую система аутентификации для защиты каналов связи, реализующую способ аутентификации пользователя на основе биометрических данных посредством представления и выделения криптографического ключа и аутентификации пользователя на основе биометрических данных.

Патент РФ № 2469397 описывает способ биометрической аутентификации по почерку в компьютеризированной системе контроля доступа. Данный патент относится к области обработки биометрических измерений, в частности к методам цифровой обработки рукописного текста. Техническим результатом является повышение надежности санкционированного доступа во всех системах, требующих использования средств идентификации личности.

Патент РФ № 2347274 описывает способы и аппарат для ограничения доступа к играм с использованием биометрических данных. Данный патент относится к игровым системам, в которых могут приниматься биометрические данные, связанные с выбором варианта проведения игры на основе принятых биометрических данных. Техническим результатом является обеспечение санкционированного доступа к игре.

Патент US20150220723 A1 описывает метод аутентификации пользователя на основе временной базы данных динамических изображений. Представленный в патенте метод

включает в себя отображение динамического изображения на экране дисплея, детектирования взаимодействия пользователя с отображаемым изображением и определение длительности обнаруженного взаимодействия с пользователем.

Патент US9122860 B2 описывает устройство и метод для идентификации и аутентификации пользователя при удаленном подключении к услуге через сеть. Представленная система включает в себя криптографический процессор, один из криптографических ключей, средство хранения, дополнительные средства обработки и интерфейс для формирования и передачи уникального кода аутентификации, чтобы эмулировать нажатия клавиш через стандартный ввод посредством клиентского терминала.

Патент US8856904 B2 описывает механизм для усиления защиты паролями. Предлагается сочетание пароля, который включает в себя динамический текст, со статическим паролем, полученным от пользователя. Проверка пользователя осуществляется на основе верификации комбинации паролей без динамического текста.

Патент US8879415 B2 описывает масштабируемое решение для мониторинга потока сетевой информации, которое принимает стандартные записи потока, экспортируемого из сетевых устройств, и помечает поток дополнительной информацией (аннотациями). Эти аннотации добавляются информацию в поток данных и могут быть использованы для выполнения анализа потоков.

ЗАКЛЮЧЕНИЕ

Проведен поиск патентных документов, включая заявки и патенты на изобретения, полезные модели, технические решения которых касаются методов анализа данных поведенческой биометрии пользователей при работе в рамках стандартного человеко-машинного интерфейса, для решения задач компьютерной безопасности, включая задачи активной аутентификации и идентификации пользователей, обнаружения внутренних вторжений и предотвращения попыток хищения конфиденциальной информации.

Патентный поиск проводился по базам данных Федеральной службы по интеллектуальной собственности, патентам и товарным знакам (Роспатента), Бюро по патентам и товарным знакам США (USPTO), Европейского патентного бюро (EPO) и Всемирной организации интеллектуальной собственности (ВОИС, или WIPO) по сформированному набору основных ключевых слов, относящихся к заданной тематике (включая их комбинации и русские аналоги).

В ходе проведения патентных исследований был обработан информационный массив в объеме более пятидесяти единиц патентной и научно-технической документации. В результате патентного поиска найдено семь патентов, схожих с искомыми тематиками.

Проведен анализ выявленных патентных документов и определен их правовой статус. Отобраны и проанализированы патентные документы, принадлежащие как российским, так и зарубежным компаниям и частным лицам. Выявлены основные тенденции совершенствования и развития методов анализа данных поведенческой биометрии пользователей при работе в рамках стандартного человеко-машинного интерфейса, для решения задач компьютерной безопасности, включая задачи активной аутентификации и идентификации пользователей, обнаружения внутренних вторжений и предотвращения попыток хищения конфиденциальной информации. Определены основные патентообладатели и заявители технических решений в данной области.

Анализ отобранных информационных источников позволил определить уровень техники объекта исследований: «Методы анализа данных поведенческой биометрии пользователей при работе в рамках стандартного человеко-машинного интерфейса, для решения задач компьютерной безопасности, включая задачи активной аутентификации и идентификации пользователей, обнаружения внутренних вторжений и предотвращения попыток хищения конфиденциальной информации».

Задание на проведение патентных исследований выполнено полностью. По результатам патентного анализа можно сделать вывод об отсутствии конфликтов с существующими патентами и о возможности получения правовой охраны планируемых РИД настоящих ПНИ. Предложения по дальнейшему проведению поиска и патентных исследований отсутствуют.

ПРИЛОЖЕНИЕ В.А

Задание на проведение патентных исследований

УТВЕРЖДАЮ
Зам. проректора

_____ А.Э.Сазонов

«__» _____ 2016 г.

ЗАДАНИЕ № 2016.02.29-01

на проведение патентных исследований

Наименование (тема) ПНИ: «Исследование и разработка инновационной технологии построения программных средств обеспечения компьютерной безопасности, основанных на использовании методов машинного обучения и математической статистики для анализа данных поведенческой биометрии пользователей при работе в рамках стандартного человеко-машинного интерфейса, для решения задач активной аутентификации и идентификации пользователей, обнаружения внутренних вторжений и предотвращения попыток хищения конфиденциальной информации».

Шифр: 2014-14-576-0053-023.

Номер соглашения о предоставлении субсидии: 14.604.21.0056.

Этап работы: «Экспериментальные исследования поставленных перед ПНИ задач».

Сроки его выполнения:*1 марта 2016 – 29 июня 2016*

Задачи патентных исследований: Обоснование целесообразности правовой охраны объектов интеллектуальной собственности в стране и за рубежом, выбор стран патентования.

КАЛЕНДАРНЫЙ ПЛАН

Виды патентных исследований	Подразделения-исполнители (соисполнители)	Ответственные исполнители (Ф.И.О.)	Сроки выполнения патентных исследований. Начало. Окончание	Отчетные документы
Проверка на патентную чистоту по патентным базам Роспатента, EPO, WIPO, USPTO		*Глазкова В.В., Царёв Д.В.*	*1 марта 2016 - 29 июня 2016*	*Отчет о патентных исследованиях*

Руководитель ПНИ, д.ф.-м.н., профессор	личная подпись	И.В. Машечкин расшифровка	29 февраля 2016 дата
Руководитель патентного подразделения, к.ф.-м.н.	личная подпись	М.И. Петровский расшифровка	29 февраля 2016 дата

ПРИЛОЖЕНИЕ В.Б

Регламент поиска

Регламент поиска № 2016.03.01-01

«01» *марта* 2016 г.

дата составления регламента

Наименование (тема) ПНИ: «Исследование и разработка инновационной технологии построения программных средств обеспечения компьютерной безопасности, основанных на использовании методов машинного обучения и математической статистики для анализа данных поведенческой биометрии пользователей при работе в рамках стандартного человеко-машинного интерфейса, для решения задач активной аутентификации и идентификации пользователей, обнаружения внутренних вторжений и предотвращения попыток хищения конфиденциальной информации».

Шифр: 2014-14-576-0053-023.

Номер соглашения о предоставлении субсидии: 14.604.21.0056.

Номер и дата утверждения задания: 2016.02.29-01 от 29 февраля 2016.

Этап работы: «Экспериментальные исследования поставленных перед ПНИ задач».

Цель поиска информации: Обосновать целесообразность правовой охраны объектов интеллектуальной собственности в стране и за рубежом.

Обоснование регламента поиска: Патентный поиск будет проведен по базам данных Федеральной службы по интеллектуальной собственности, патентам и товарным знакам (Роспатента), Бюро по патентам и товарным знакам США (USPTO), Европейского патентного бюро (EPO) и Всемирной организации интеллектуальной собственности (ВОИС, или WIPO). Выбор указанных патентных баз обусловлен тем, что на соответствующих географических регионах активно ведутся научно-исследовательские работы по теме ПНИ. Патентный поиск будет проведен без ограничений по давности публикации патента.

Начало поиска 1 марта 2016 Окончание поиска 29 июня 2016.

Предмет поиска (объект исследования, его составные части, товар)	Страна поиска	Источники информации, по которым будет проводиться поиск								Ретроспективность	Наименование информационной базы (фонда)
		Патентные		НТИ		конъюнктурные		другие			
		Наименование	Классификационные рубрики: МПК (МКИ), МКПО, НКИ и др.	Наименование	Рубрики УДК и другие	Наименование	Код товара: ГС, СМТК, БТН	Наименование	Классификационные индексы		
1	2	3	4	5	6	7	8	9	10	11	12
Методы анализа данных поведенческой биометрии пользователей при работе в рамках стандартного человеко-машинного интерфейса для решения задач компьютерной безопасности, включая задачи активной аутентификации и идентификации пользователей, обнаружения внутренних вторжений и предотвращения попыток хищения конфиденциальной информации.	Россия, США, Европа	1. База данных по изобретениям Федеральной службы по интеллектуальной собственности, патентам и товарным знакам (Роспатент). 2. Патентная база данных Бюро по патентам и товарным знакам США (USPTO). 3. Патентная база данных Европейского патентного бюро (EPO). 4. Патентная база данных Всемирной организации интеллектуальной собственности (ВОИС, WIPO).	G06F 13/00 G06F 7/00 G06F 17/00 G06F 17/20 G06F 17/21 G06F 17/30 G06F 17/60	-	-	-	-	-	-	Без ограничений	1. Роспатент, http://www.fips.ru. 2. USPTO, http://uspto.gov. 3. EPO, http://ep.espacenet.com. 4. WIPO, http://www.wipo.int, http://www.wipo.int/portal/index.html.ru.

Руководитель ПНИ, д.ф.-м.н., профессор _____ (личная подпись) И.В. Машечкин (расшифровка) 1 марта 2016 (дата)

Руководитель патентного подразделения, к.ф.-м.н. _____ (личная подпись) М.И. Петровский (расшифровка) 1 марта 2016 (дата)

ПРИЛОЖЕНИЕ В.В

Отчет о поиске

ОТЧЁТ О ПОИСКЕ

В.В.1 **Поиск проведен в соответствии с заданием** руководителя ПНИ, д.ф.-м.н., профессора И.В.Машечкина № 2016.02.29-01 от 29 февраля 2016 г. **и Регламентом поиска** № 2016.03.01-01 от 01 марта 2016 г.

В.В.2 **Этап работы** «Экспериментальные исследования поставленных перед ПНИ задач».

В.В.3 **Начало поиска** 01 марта 2016г. **Окончание поиска** 29 июня 2016г.

В.В.4 **Сведения о выполнении регламента поиска**: Основная задача патентного поиска заключалась в обосновании целесообразности правовой охраны объектов интеллектуальной собственности, в связи с чем осуществлялась выборка патентов, связанных с методами анализа данных поведенческой биометрии пользователей при работе в рамках стандартного человеко-машинного интерфейса для решения задач компьютерной безопасности, включая задачи активной аутентификации и идентификации пользователей, обнаружения внутренних вторжений и предотвращения попыток хищения конфиденциальной информации.

Патентный поиск проводился по базам данных Федеральной службы по интеллектуальной собственности, патентам и товарным знакам (Роспатента), Бюро по патентам и товарным знакам США (USPTO), Европейского патентного бюро (EPO) и Всемирной организации интеллектуальной собственности (ВОИС, или WIPO) по следующему набору основных ключевых слов: authentication system; machine learning; keyboard; mouse; behavior analysis; behavior model; keystroke dynamics; biometric; mouse dynamics; identity authentication; continuously verifying; abnormal data-inputting behavior; keystroke latencies; computer security; security awareness; computer awareness; information security; end-user security behaviors; security policy compliance; data mining; information retrieval; security information and event management; intrusion prevention; intrusion detection; log management; computer misuse; system misuse; anomaly based detection; signature based detection; misuse based detection; data loss prevention; topic modeling; orthogonal non-negative matrix factorizations; time series; prediction; category values timeline; topic time series; topic time line; topic (concept); text mining; а также по их комбинациям и русским аналогам.

В результате патентного поиска найдено семь патентов, схожих с искомыми тематиками.

Патентные исследования по первому этапу завершены полностью. Документов, которые могут препятствовать применению результатов НИР в Российской Федерации, не обнаружено.

В.В.5 **Предложения по дальнейшему проведению поиска и патентных исследований**

Никаких нарушений с точки зрения найденных патентов не найдено, можно утверждать, что патентная чистота разрабатываемых методов оценки анализа данных поведенческой биометрии пользователей при работе в рамках стандартного человеко-машинного интерфейса для решения задач компьютерной безопасности, включая задачи активной аутентификации и идентификации пользователей, обнаружения внутренних вторжений и предотвращения попыток хищения конфиденциальной информации, обеспечена с достаточной степенью достоверности, предложения по дальнейшему проведению поиска и патентных исследований отсутствуют.

В.В.6 **Материалы, отобранные для последующего анализа**.

Таблица В.В.6.1 - Патентная документация

Предмет поиска (объект исследования, его составные части)	Страна выдачи, вид и номер охранного документа. Классификационный индекс	Заявитель (патентообладатель), страна. Номер заявки, дата приоритета, конвенционный приоритет, дата публикации	Название изобретения (полезной модели, образца)	Сведения о действии охранного документа или причина его аннулирования (только для анализа патентной чистоты)
1	2	3	4	5
Методы анализа данных поведенческой биометрии пользователей при работе в рамках стандартного человеко-машинного интерфейса для решения задач компьютерной безопасности, включая задачи активной аутентификации и идентификации пользователей, обнаружения внутренних вторжений и предотвращения попыток хищения конфиденциальной информации.	Патент: РФ № 2316120. Страна выдачи: Россия	Патентообладатель: Самсунг Электроникс Дата заявки: 12.05.2004 Номер заявки: 2316120	Биометрическая система аутентификации	Дата выдачи: 27.01.2008
	Патент: РФ № 2469397 Страна выдачи: Россия	Патентообладатель: ЮЗГУ Дата заявки: 30.09.2011 Номер заявки: 2469397	Способ биометрической аутентификации по почерку в компьютеризированной системе контроля доступа	Дата выдачи: 10.12.2012
	Патент: РФ № 2347274 Страна выдачи: Россия	Патентообладатель: Ай Джи Ти. Дата заявки: 11.03.2004 Номер заявки: US2004/007423	Способы и аппарат для ограничения доступа к играм с использованием биометрических данных	Дата выдачи: 25.10..2005
	Патент: US20150220723 A1 Страна выдачи: США	Патентообладатель: IBM Дата заявки: 23.01.2015 Номер заявки: US 14/604,655	Аутентификация пользователя на основе временной базы данных динамических изображений (User authentication using temporal knowledge of dynamic images)	Дата выдачи: 06.08.2015

Продолжение таблицы В.В.6.1

1	2	3	4	5
	Патент: US9122860 B2 Страна выдачи: США	Патентообладатель: Yankey Information Co Дата заявки: 24.04.2006 Номер заявки: US 14/309,061	Устройство и метод для идентификации и аутентификации (Device and method for identification and authentication)	Дата выдачи: 18.12.2013
	Патент: US8856904 B2 Страна выдачи: США	Патентообладатель: IBM Дата заявки: 23.02.2013 Номер заявки: US 13/774,191	Усиленная защита паролями (Enhancing password protection)	Дата выдачи: 07.10.2014
	Патент: US8879415 B2 Страна выдачи: США	Патентообладатель: Arbor Networks, Inc Дата заявки: 01.03.2014 Номер заявки: US 13/782,776	Метод и система для аннотирования потока сетевой информации (Method and system for annotating network flow information)	Дата выдачи: 04.11.2014