МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В. Ломоносова

Факультет Вычислительной Математики и Кибернетики
Кафедра Интелектуальных Информационных Технологий
Лаборатория Технологий Программирования

Курсовая работа
по теме:

«Динамическая аутентификация пользователей на основе анализа работы с компьютерной мышью»

Березникер Алексей Витальевич
3 курс, группа 320

Научный руководитель:
к.ф.-м.н. Казачук Мария Андреевна

Москва, 2020 г.

Содержание

1 Аннотация

Целью данной работы является исследование существующих и разработка собственных алгоритмов динамической аутентификации пользователя на основе анализа работы с компьютерной мышью, показывающих высокое качество работы и способных работать в динамическом режиме.

Мы фокусируемся на независимой от контекста системе динамической аутентификации, которая реагирует на каждое отдельное действие, выполненное пользователем.

Отметим, что динамическая аутентификация не является альтернативным решением безопасности для первоначального входа в систему, она обеспечивает дополнительную меру безопасности наряду с первоначальным логином.

2 Введение

2.1 Область применения

В настоящее время неотъемлемой частью различных сфер деятельности человека стало использование информационных систем. Огромное количество информации ограниченного доступа переносится, хранится и обрабатывается в информационных системах, что формирует потребность в обеспечении их защищенности.

Люди используют механизмы контроля доступа, такие как пароль, магнитные карты или биометрию, для защиты от несанкционированного доступа другого человека. Это означает, что пользователь должен предоставить подтверждение своей личности при запуске или разблокировке системы. Однако во многих случаях люди оставляют компьютер без присмотра, чтобы попить чашку кофе, пойти и поговорить с коллегой, или просто потому, что у них нет привычки выключать компьютер.

Контроль доступа к компьютеру обычно реализуется как единоразовое подтверждение личности во время начальной процедуры входа в систему. Предполагается, что в течении всего сеанса в системе будет находиться зарегистрированный пользователь. К сожалению, когда компьютер оставлен без присмотра, любой человек может получить доступ к тем же источникам, что и подлинный пользователь.

Защита информации в информационных системах обеспечивается созданием комплексной системы защиты, одной из главных составляющих которой являются методы защиты от несанкционированного доступа.

Основой программно-технических средств защиты от несанкционированного доступа являются процедуры идентификации и аутентификации пользователей. Идентификатором служит уникальный признак объекта, позволяющий отличить его от других объектов. А под процедурой аутентификации подразумевается процесс проверки принадлежности субъекту доступа предъявленного им идентификатора.

2.2 Методы аутентификации

Существующие методы осуществления аутентификации можно разделить на три категории. К первой относятся методы, основанные на обладании субъекта аутентификации некоторым секретным знанием. В качестве такого знания может выступать секретное слово, пароль или цифровой сертификат. Данный метод является самым распространенным и простым. Он часто подвержен атакам со стороны злоумышленников.

Ко второй категории относятся методы, основанные на наличии у субъекта идентификации некоторого физического объекта. Таким объектом может быть, например, ключ, флеш-накопитель или магнитная карта. Аутентификация по предъявлении чего-либо, чем владеет пользователь, имеет сходные недостатки, и, кроме того, добавляется риск передачи, утери, кражи или копирования ключа. А также требуется специальное оборудование для распознавания предмета, используемого при аутентификации.

К последней категории относят методы, основанные на собственных свойствах субъекта доступа. В качестве таких свойств могут рассматриваться биометрические данные пользователя, т.е. уникальные биологические и физиологические характеристики, которые позволяют установить личность человека. Методы аутентификации, основанные на проверке подлинности через предъявление биометрического образа называется биометрической аутентификацией.

2.2.1 Методы биометрической аутентификации

Существующие в настоящее время методы биометрической аутентификации разделяются на два класса:

1. **Статические методы биометрической аутентификации**,
   основанные на физиологических характеристиках человека, находящиеся при нём в течение всей его жизни. Например, проверка отпечатка пальца, сетчатки глаза или геометрии лица. Статическая аутентификация заключается в эпизодической проверке личности пользователя (например, при его входе в систему), после чего пользователь может свободно пользоваться системой.

2. **Динамические методы биометрической аутентификации**,
   основанные на поведенческих характеристиках человека. Анализ голоса, клавиатурного почерка или работы с компьютерной мышью. Динамическая аутентификация предполагает проведение проверки личности пользователя постоянно на протяжении всей сессии.

Основным недостатком методов проверки пользователей, основанных на физиологической биометрии, является то, что они требуют аппаратные устройства, такие как датчики отпечатков пальцев и сканеры сетчатки, которые дороги и не всегда доступны. Хотя проверка отпечатков пальцев становится широко распространенной в ноутбуках, она все еще недостаточно популярна и не может быть использована в веб-приложениях. Кроме того, отпечатки пальцев могут быть скопированы.В свою очередь методы, основанные на поведенческой биометрии не требуют специальное оборудование, так как они используют обычные устройства, такие как мышь и клавиатура.

Другим важным отличием между физиологической и поведенческой биометрией является временной аспект. Поведенческая биометрия может отличаться в зависимости от режима работы пользователя и времени суток, когда она была зафиксированна. Это усложняет процесс подражания для обхода системы, даже в случае перехватывата данных.

Очевидно, динамическая аутентификация пользователей является предпочтительной, так как она исключает сценарии, при которых злоумышленник получает доступ к информационной системе после того, как легитимный пользователь прошел процедуру аутентификацию.

3 Актуальность

Таким образом, мы видим проблему отсутствия контроля факта смены пользователя и компрометации идентификаторов, которую мы предлагаем решить использованием биометрических характеристик пользователя для динамической аутентификации.

Достоинство нашего подхода заключается в простоте внедрения: нужно лишь устройство ввода (компьютерная мышь) и специальное программное обеспечение, позволяющее проводить анализ.

4 Постановка задачи

Задачей данной работы является исследование существующих и разработка собственных алгоритмов динамической аутентификации пользователя на основе анализа работы с компьютерной мышью, показывающих высокое качество работы и способных работать в динамическом режиме.

Наше решение должно выполнять задачи незаметно для пользователя, выявлять злоумышленника как можно быстрее, в то же время избегая в максимально возможной степени неправильной блокировки легитимного пользователя.

5 Обзор существующих решений

5.1 Цели обзора

1. Выявить достоинства и недостатки существующих подходов;
2. Выявить наиболее релевантные признаки для построения модели пользователя;
3. Выявить методы построения модели, показывающие наилучшее качество работы;
4. Найти набор данных в открытом доступе для проведения собственных исследований;
5. Сформулировать направления дальнейших исследований.

5.2 Показатели эффективности

5.3 Построение признакового пространства

Таблица 1: Признаковое пространство

Признак	Формула	Признак	Формула
Direction bin	Divided into 8 bins (45°)	Curve acceleration	$\frac{curvespeed}{\Delta t}$

6 Исследование и построение решения задачи

6.1 Сбор данных и извлечение признаков

6.1.1 Описание наборов данных

По итогам обзора было собрано 3 набора данных для оценки производительности поведенческих биометрических алгоритмов на основе динамики работы с компьютерной мышью в целях аутентификации пользователя.

BALABIT [TODO]
Это датасет, предоставленный компанией BalaBit IT Security, специализируещейся на разработке программного обеспечения и сервисов для информационной безопасности, в 2016 году в рамках одноименного конкурса Balabit Mouse Dynamics Challenge. Набор данных доступен для исследователей и экспертов в области IT-безопасности и науки, используется в большинстве научных статей обзора. Данные собраны в неконтролируемой среде и включают информацию о времени и позиционировании указателя мыши. В эксперименте приняло участие 10 человек. В среднем для обучения мы имеем 43 $\pm$ 17 часов работы для каждого пользователя.

TWOS [TODO]
Этот датасет был собран во время конкурса, организованного Сингапурским университетом технологии и дизайна в марте 2017 года. Набор данных содержит действия 24 пользователей, которые собирались в неконтролируемой среде в течение 5 дней. Однако информация о взаимодейсвтии с компьютерной мышью есть только для 4 пользователей. В среднем мы имеем ?? $\pm$?? часов работы для каждого пользователя.

DATAIIT
Датасет, собранные на нашей кафедре в рамках исследования задачи динамической аутентификации пользователя. Данные собраны в неконтролируемой среде. В эксперименте приняло участие 20 человек. В среднем мы имеем 21 $\pm$ 5 часов работы для каждого пользователя.

6.1.2 Структура данных

Набор данных состоит из тренировочной и тестовой части в соотношении 4 к 1. Каждая часть содержит в себе записи о легитимных сеансах всех пользователей из множества $\mathfrak{U} = \{U_1, \ldots, U_j, \ldots, U_q\}$. Работа пользователя разбита на сессии разной продолжительности $U_j = \{S_1, \ldots, S_i, \ldots, S_m\}$ с записями вида $S_i = (time, xpos, ypos)$.

Далее каждую сессию мы разбиваем на сегменты с ограничением по временному порогу сверху и по минимальному количеству действий в этот промежуток времени снизу. По полученному сегменту строится один вектор признаков.

Графическая визуализация структуры данных представлена на Рис. 1

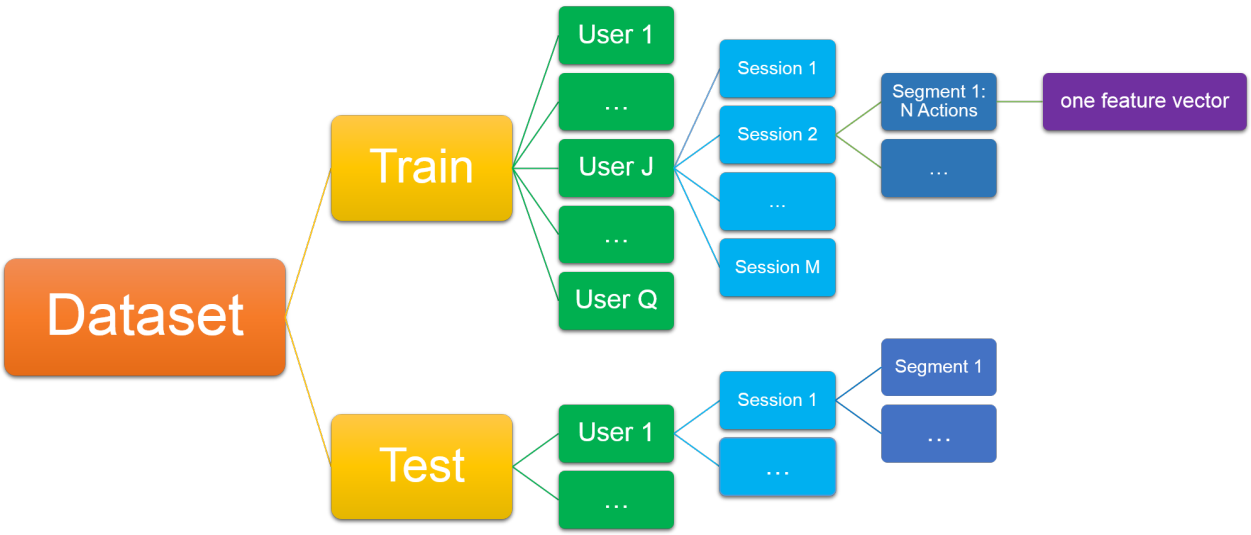

Рис. 1: Структура данных

6.1.3 Выявленные особенности данных

Во время анализа данных были обнаружены следующие особенности:

1. Полностью дублирующиеся записи в таблице;
2. Дубликаты временных меток;
3. Множественные дубликаты положения мыши (зацикливание);
4. Сверхбольшие значения координат;
5. В датасете BALABIT:
   (a) В состоянии прокрутки колеса (Scroll) положение мыши перемещается в начало координат;
   (b) Основная работа пользователей происходит в левом нижнем углу экрана.

6.2 Построение признакового пространства

6.3 Построение модели пользователя

6.3.1 OneClassSVM

6.3.2 IsolationForest

6.3.3 EllipticEnvelope

6.3.4 LocalOutlierFactor

7 Описание практической части

8 Планы на будущее

9 Заключение

10 Список цитируемой литературы

1. ...

2. Fülöp, Á., Kovács, L., Kurics, T., Windhager-Pokol, E. (2016). Balabit Mouse Dynamics Challenge data set. Available at: https://github.com/balabit/Mouse-Dynamics-Challenge

3. Harilal A. et al. Twos: A dataset of malicious insider threat behavior based on a gamified competition //Proceedings of the 2017 International Workshop on Managing Insider Security Threats. – 2017. – С. 45-56.

4. ...

5. Mondal S., Bours P. A study on continuous authentication using a combination of keystroke and mouse biometrics //Neurocomputing. – 2017. – Т. 230. – С. 1-22.

6. Feher C. et al. User identity verification via mouse dynamics //Information Sciences. – 2012. – Т. 201. – С. 19-36.