Национальный исследовательский ядерный университет
«МИФИ»
Кафедра 42 «Криптология и кибербезопасность»

Метод непрерывной аутентификации пользователей мобильных устройств на основе анализа нескольких поведенческих характеристик

Исполнитель:
студент гр. Б17-505
Казьмин С. К.

Научный руководитель:
аспирант каф. 42
Еремин А. В.

Актуальность работы

- Риск утери устройства
- Несанкционированный доступ к устройству
- Утечка приватных данных

- Широко распространены
- Хранение конфиденциальной информации
- Финансовые операции

Биометрическая аутентификация

Биологическая

- Можно обойти
- Требует дополнительных встраиваемых датчиков
- Требует дополнительных от человека действий
- Сложно осуществлять непрерывную аутентификацию

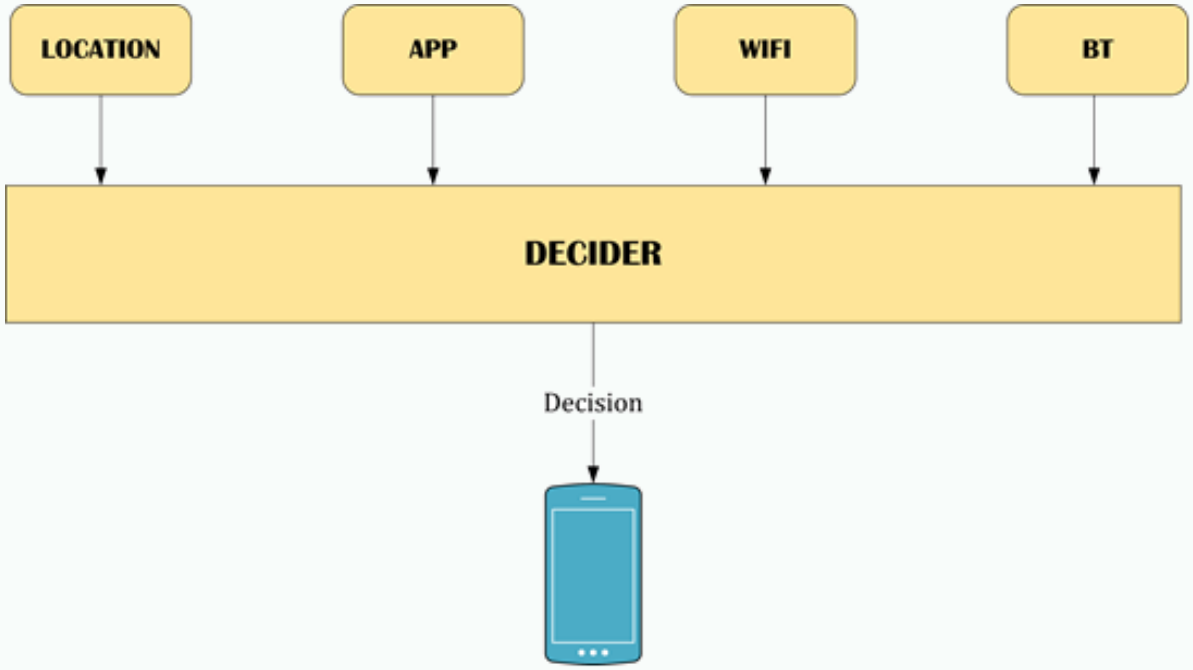
Поведенческая

- Удобна для пользователя
- Сложно подделать
- Возможность непрерывной аутентификации
- Не требуются дополнительные датчики

Цель работы

Исследовать и оценить эффективность метода непрерывной неявной аутентификации пользователя мобильного устройства на основе анализа нескольких поведенческих характеристик

Метод аутентификации по поведенческому профилю



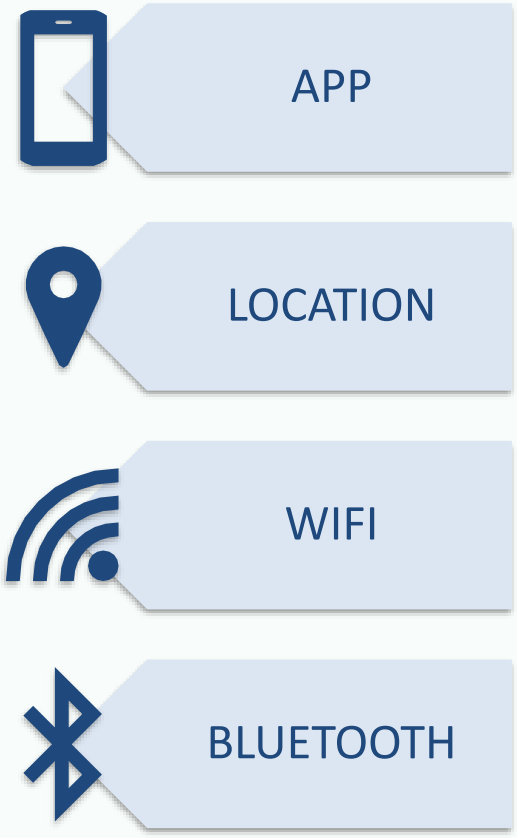
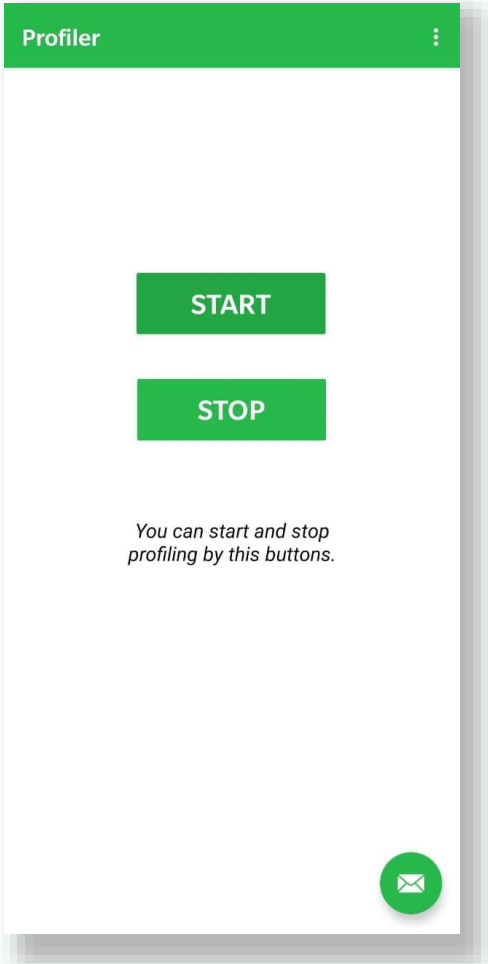


Схема многомодульной системы непрерывной аутентификации по поведенческому профилю

Мобильное приложение для сбора данных



Главный экран мобильного приложения

WIFI

- Время
- ID сканирования
- BSSID*
- RSSI**
- Частота
- …

BT

- Время
- MAC-адрес
- Тип устройства
- …

LOCATION

- Время
- Точность
- Широта
- Долгота
- Высота над уровнем моря
- …

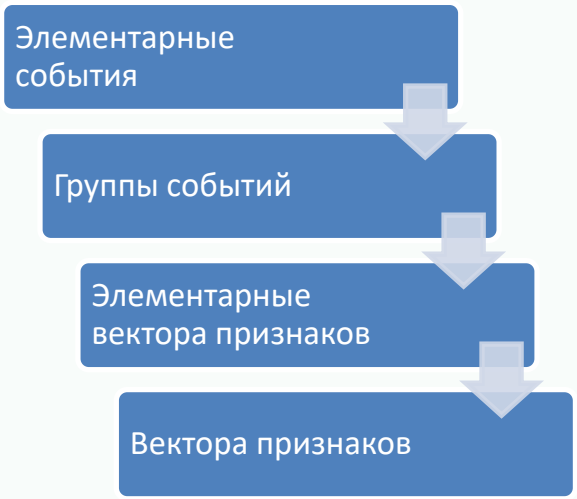
APP

- Время
- Название пакета
- Время последнего запуска
- Продолжительность использования
- …

*BSSID – идентификатор сети, обычно – MAC-адрес (basic service set id)

**RSSI – показатель уровня принимаемого сигнала (received signal strength indicator)

Формирование признаков



WIFI	BT	LOCATION
• Средняя частота • Количество сетей • Динамика изменения вектора присутствия сетей • …	• Количество устройств • Динамика изменения вектора присутствия сетей • …	• Средняя точность • Скорость • Ускорение • Скорость изменения высоты • …

$$F = [net_0, net_1, \ldots, net_n],$$

$$net_i = \begin{cases} 1, \text{ если } i \in G_j \\ 0, \text{ если } i \notin G_j \end{cases},$$

где G_j — множество номеров сетей WiFi, присутствующих в j-ой группе элементарных событий.

Динамика изменения векторов

- Число возникших объектов
- Число исчезнувших объектов
- Расстояние Жаккара

Формирование признаков

timestamp	freq	level	count
2020-12-06 17:56:05.536	3303.470588	-57.235294	17
2020-12-06 17:56:08.521	3184.000000	-46.916667	12
2020-12-06 17:56:12.036	3128.461538	-50.769231	13
2020-12-06 17:56:19.939	3188.166667	-49.583333	12
2020-12-06 17:56:26.836	3356.368421	-59.842105	19
2020-12-06 17:56:30.335	3720.428571	-33.714286	7
2020-12-06 17:56:34.951	3101.555556	-58.111111	18
2020-12-06 17:56:39.942	3176.500000	-61.050000	20
2020-12-06 17:56:44.326	3304.941176	-58.764706	17
2020-12-06 17:56:49.361	3314.400000	-60.550000	20
2020-12-06 17:56:54.354	3483.071429	-54.285714	14
2020-12-06 17:56:59.445	3271.428571	-61.571429	21
2020-12-06 17:57:04.416	3345.437500	-57.750000	16
2020-12-06 17:57:09.427	3247.277778	-59.555556	18

timestamp	freq	level	count
2020-12-06 17:56:05.536	3303.470588	-57.235294	17
2020-12-06 17:56:08.521	3184.000000	-46.916667	12
2020-12-06 17:56:12.036	3128.461538	-50.769231	13
2020-12-06 17:56:19.939	3188.166667	-49.583333	12
2020-12-06 17:56:26.836	3356.368421	-59.842105	19
2020-12-06 17:56:30.335	3720.428571	-33.714286	7
2020-12-06 17:56:34.951	3101.555556	-58.111111	18
2020-12-06 17:56:39.942	3176.500000	-61.050000	20
2020-12-06 17:56:44.326	3304.941176	-58.764706	17
2020-12-06 17:56:49.361	3314.400000	-60.550000	20
2020-12-06 17:56:54.354	3483.071429	-54.285714	14
2020-12-06 17:56:59.445	3271.428571	-61.571429	21
2020-12-06 17:57:04.416	3345.437500	-57.750000	16
2020-12-06 17:57:09.427	3247.277778	-59.555556	18

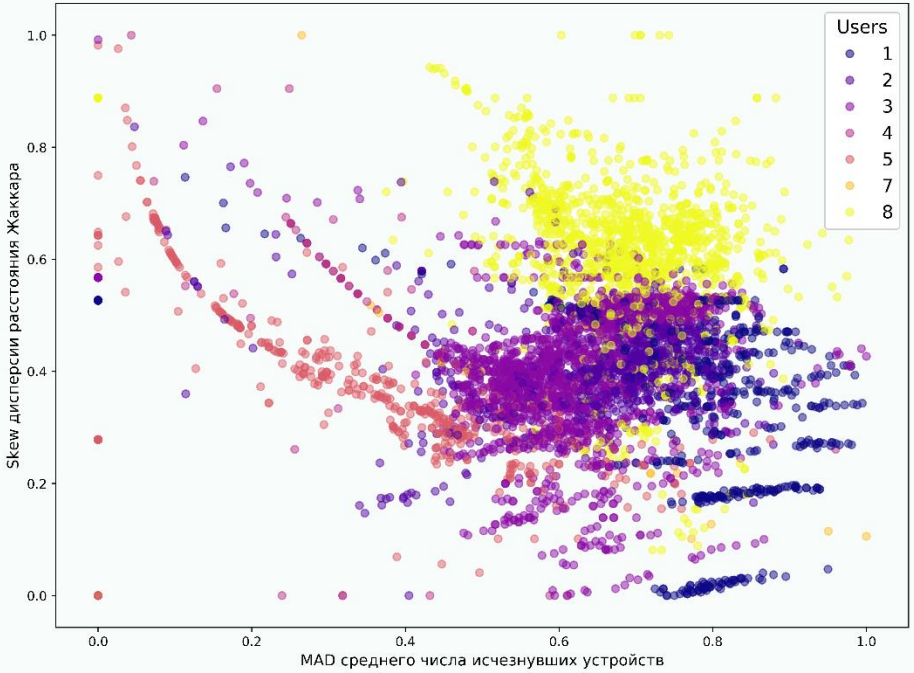
Схема разбиения данных с помощью стационарного и скользящего окна

Для каждого окна вычислялись:

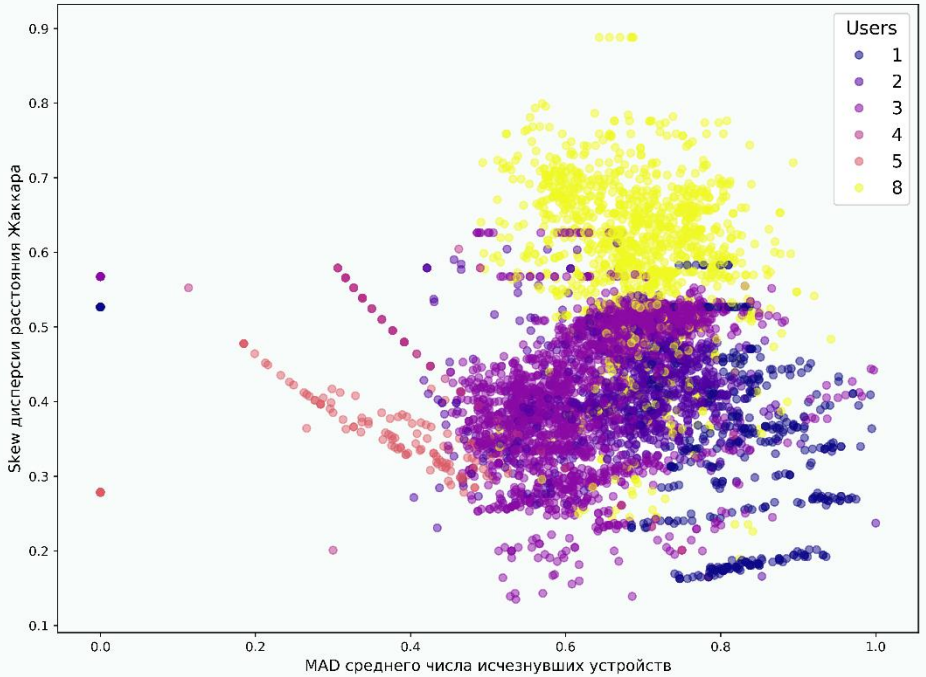
- выборочное среднее;
- дисперсия;
- медиана;
- коэффициент асимметрии;
- коэффициент эксцесса;
- стандартное отклонение;
- среднее абсолютное отклонение.

Очистка выборок

ДО



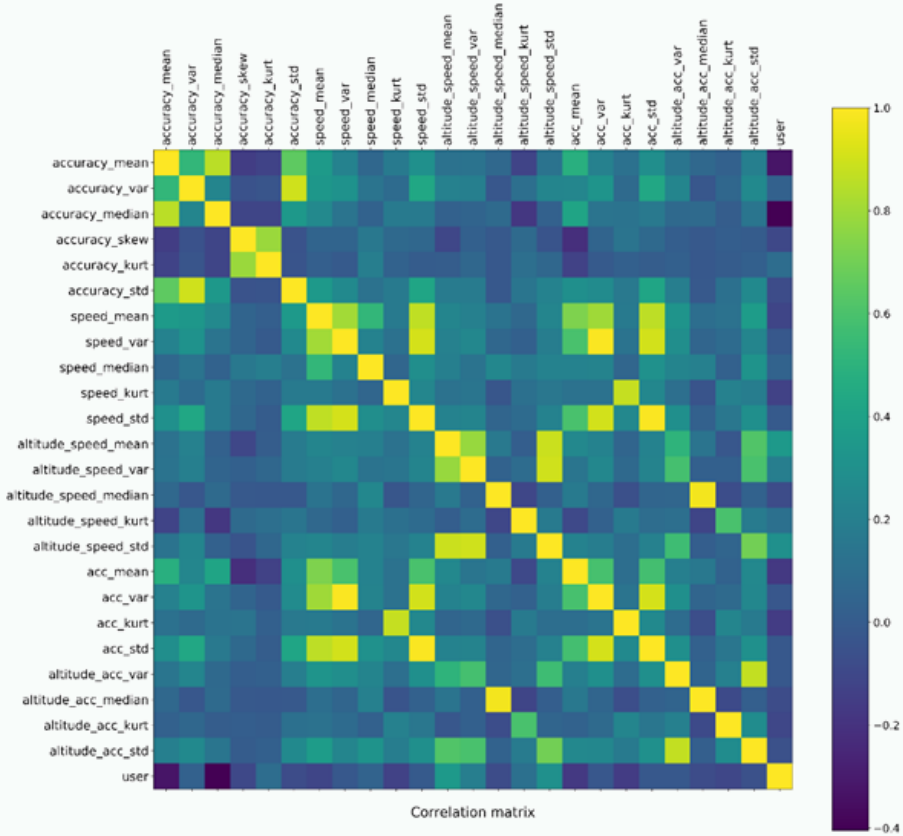
ПОСЛЕ



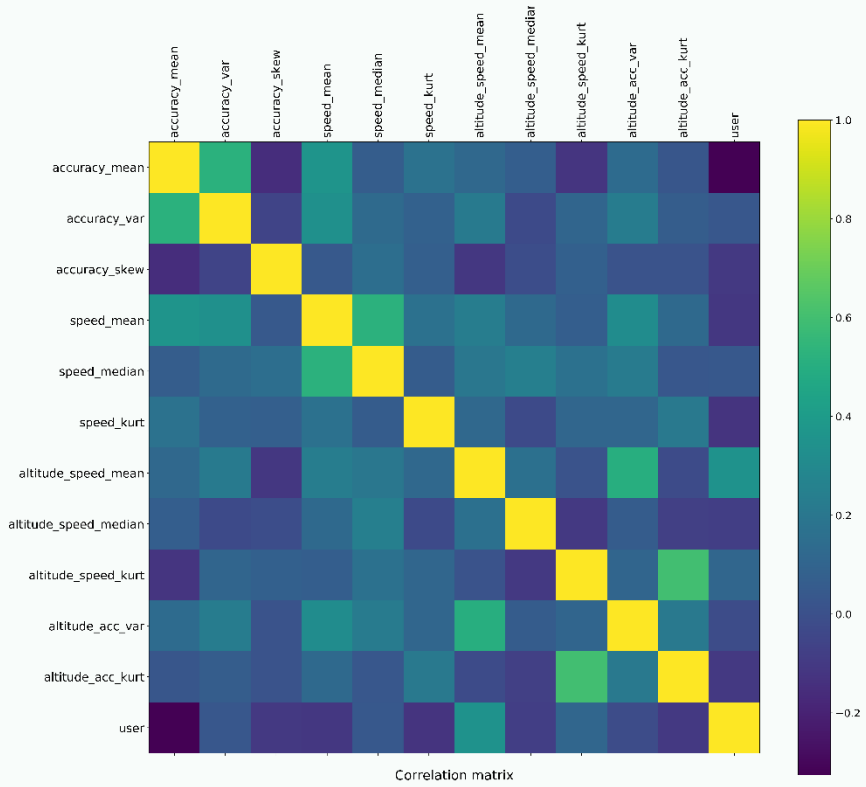
Диаграммы рассеяний для модуля BT

Отбор признаков

ДО



ПОСЛЕ



Матрицы корреляций признаков для модуля LOCATION

Методы машинного обучения в задаче аутентификации

Градиентный бустинг

Метод опорных векторов

Случайный лес

Логистическая регрессия

Обучение и тестирование моделей

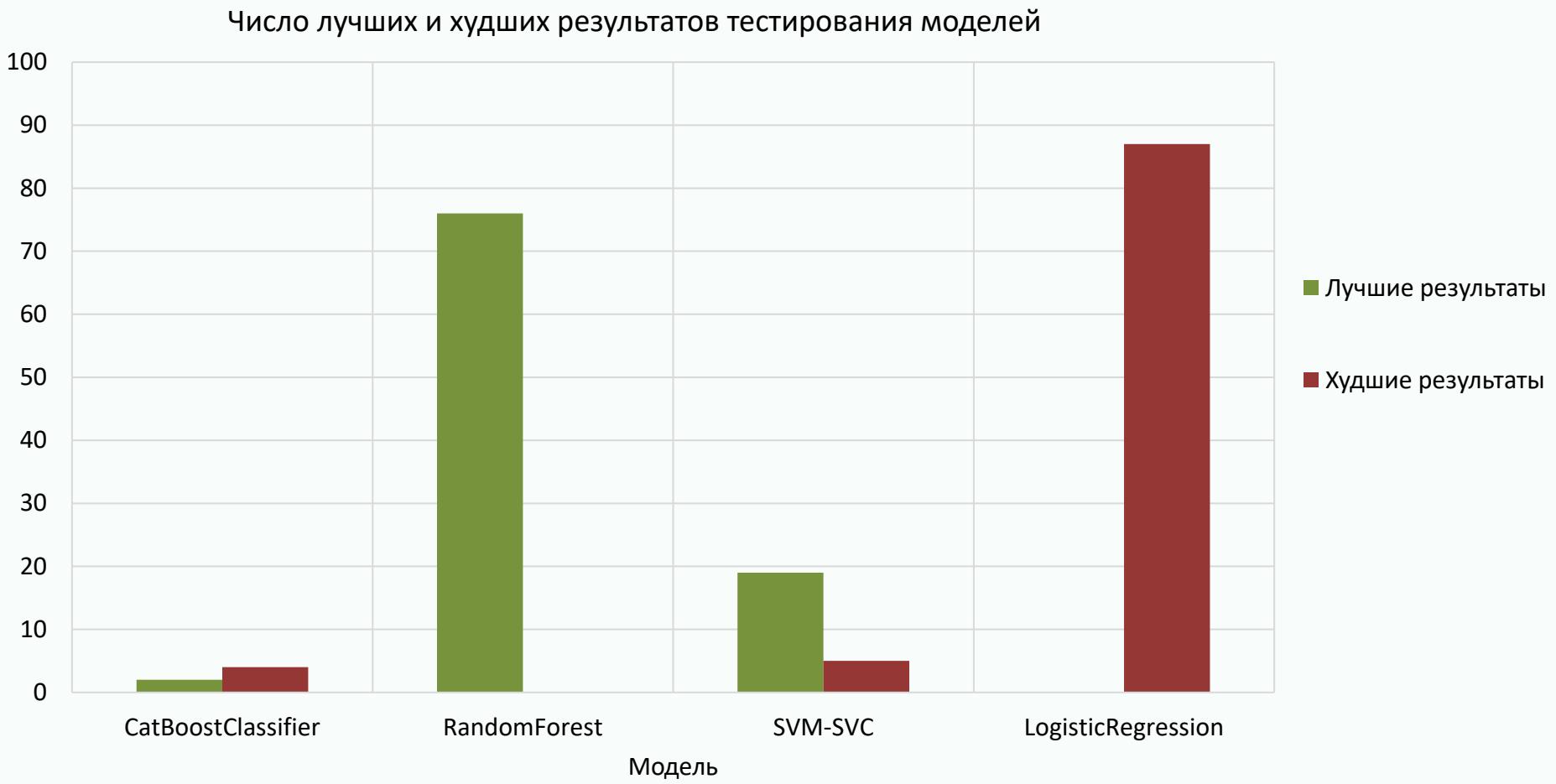
I этап

- **Кросс-валидация**
- Один пользователь выбирается легальным
- Остальные пользователи – несанкционированные
- Данные тестового пользователя извлекаются из обучающей выборки

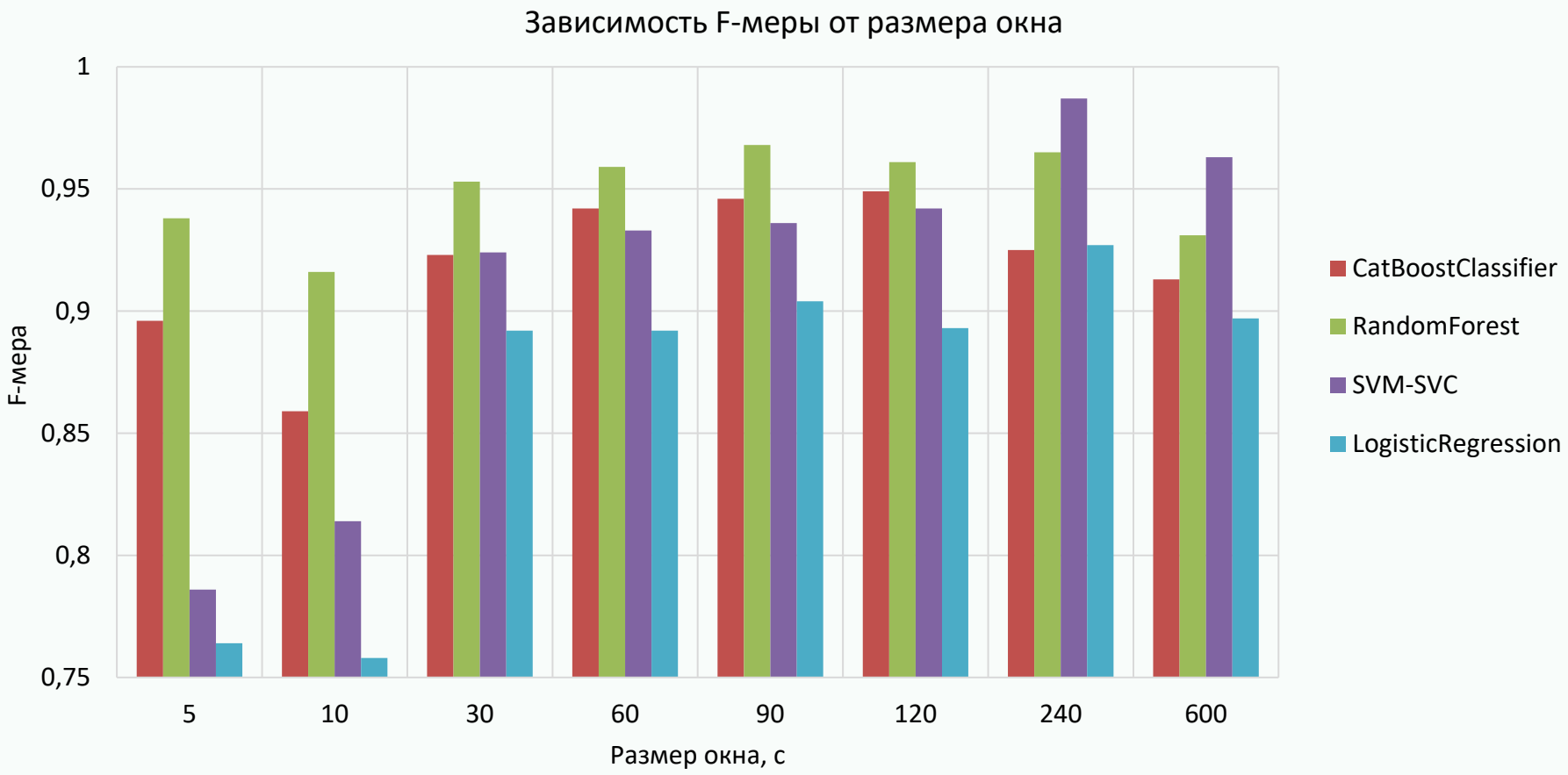
II этап

- **Финальная валидация модели**
- Один пользователь – легальный
- Из обучающей выборки извлекаются 25% данных каждого из пользователей
- Извлекаются полностью данные одного несанкционированного пользователя
- Тестируется на 1/3 данных легального пользователя, 1/3 несанкционированного, оставшаяся треть – остальные

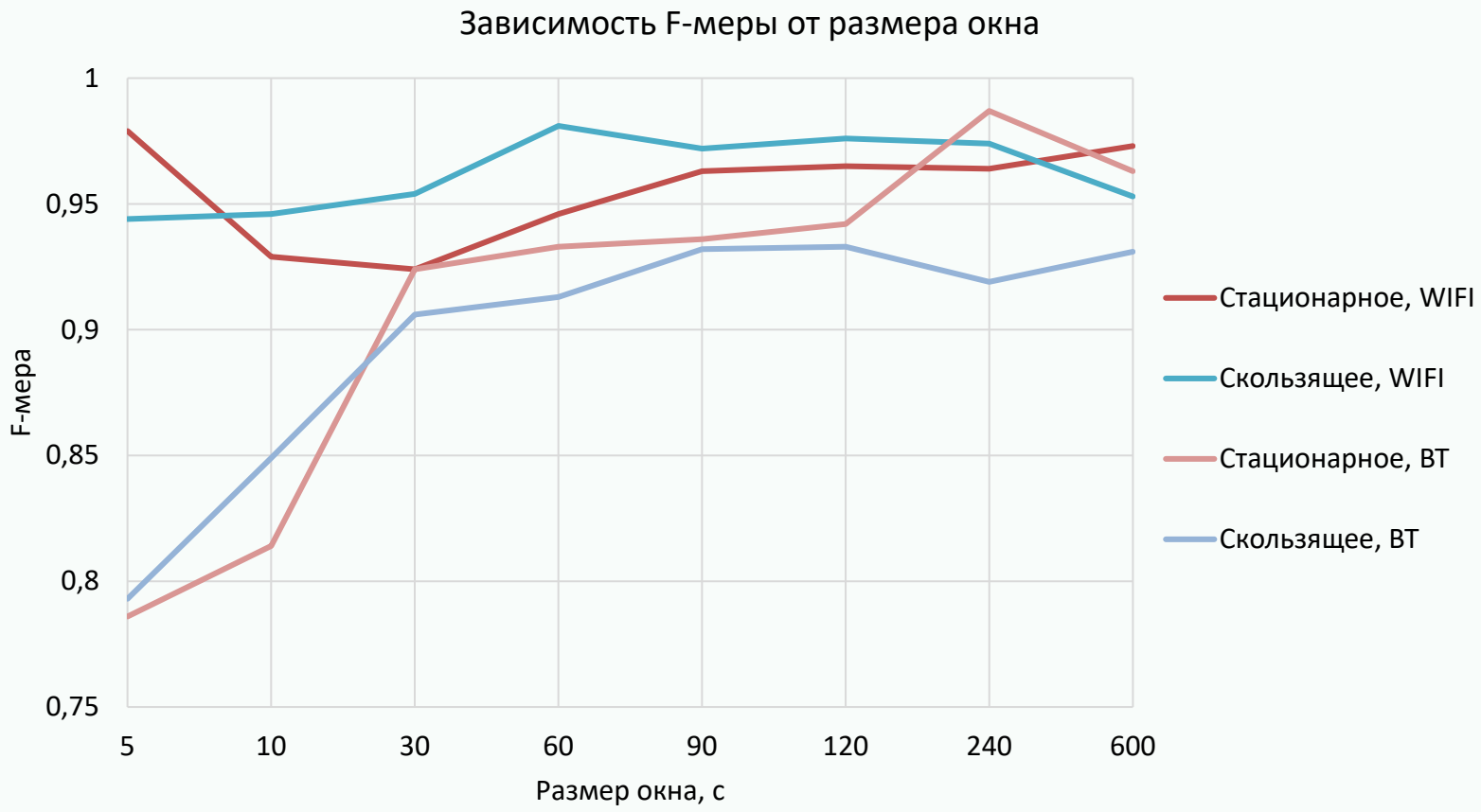
Результаты тестирования



Результаты тестирования



Результаты тестирования



Результаты работы

- Предложен метод аутентификации по поведенческой биометрии на основе данных с нескольких модулей
- Разработано приложение для сбора данных
- Проведён сбор данных от нескольких пользователей
- Сформированы признаки для модулей LOCATION, WIFI, BT
- Проведено тестирование алгоритмов классификации на полученных выборках

Направления дальнейшего исследования

- Формирование признаков для модуля APP
- Применение алгоритмов кластеризации и обнаружения аномалий
- Объединение модулей в рамках предложенного метода
- Моделирование потока событий, поступающего на вход модулям
- Оценка работы метода по нескольким метрикам качества