Российский университет дружбы народов

Анализ трафика в сетях 5G+ в реальном времени с помощью методов машинного обучения

Самуйлов Константин Евгеньевич
директор института компьютерных наук и телекоммуникаций

Бесчастный Виталий Александрович
доцент кафедры теории вероятностей и кибербезопасности института компьютерных наук и телекоммуникаций

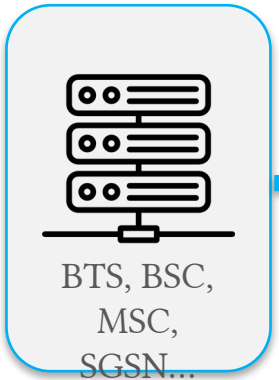
Халемский Никита Дмитриевич
НИИТ НИУ Высшая школа экономики

Эволюция мобильных сетей и их безопасности

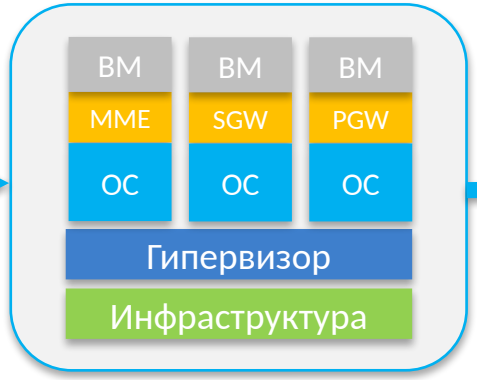
Российский университет дружбы народов

Эволюция к 5G CNFs

2G, 3G, 4G:
Физическое оборудование



4G: Виртуализация



5G: Kubernetes/
микросервисная
архитектура

Большая гибкость и доступность выч. ресурсов

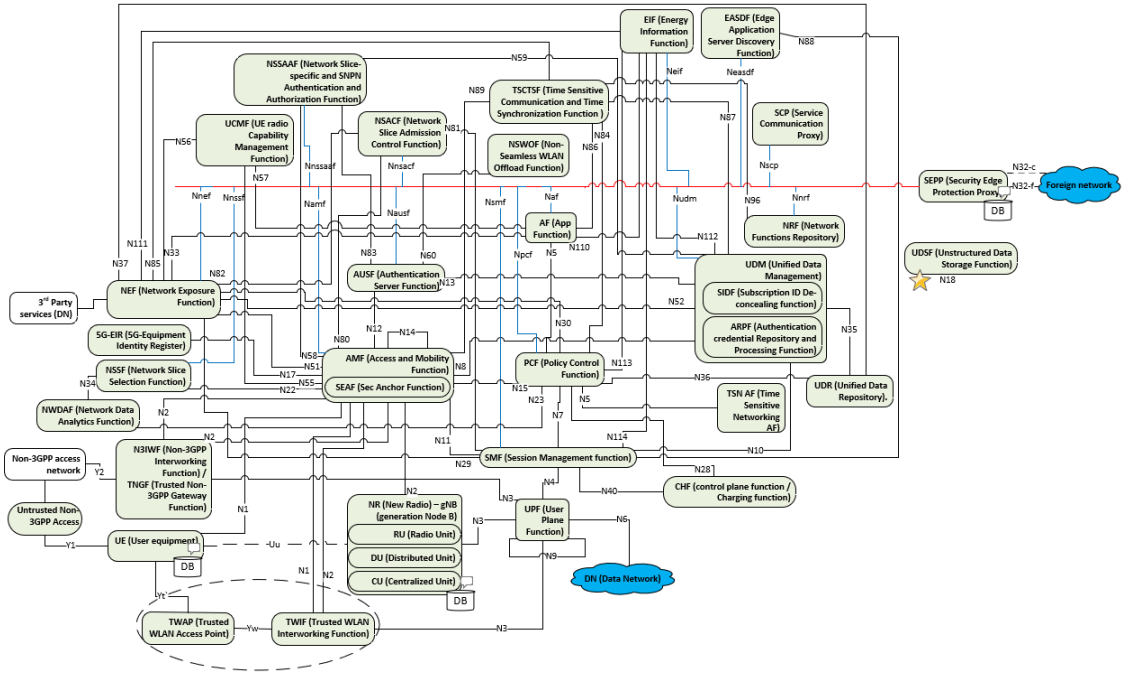
Сокращение издержек

Увеличение пространства атак

Российский университет дружбы народов

Функциональная архитектура сети 5G [1]

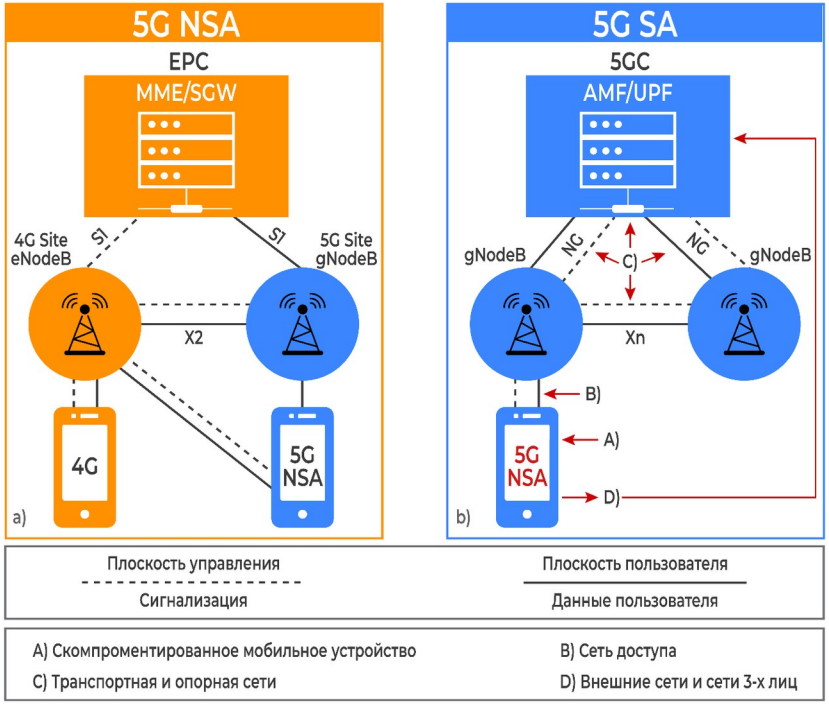
- 35 функций
- 70+ зависимостей
- дополнительные вычислительные мощности на граничных узлах
- контейнеризация и сегментирование
- шифрование и контроль целостности как сигнального, так и пользовательского трафика
- отделение логической части радиосети (CU и DU)



[1] 3GPP TS 23.501 V19.4.0 (2025-06)

Российский университет дружбы народов

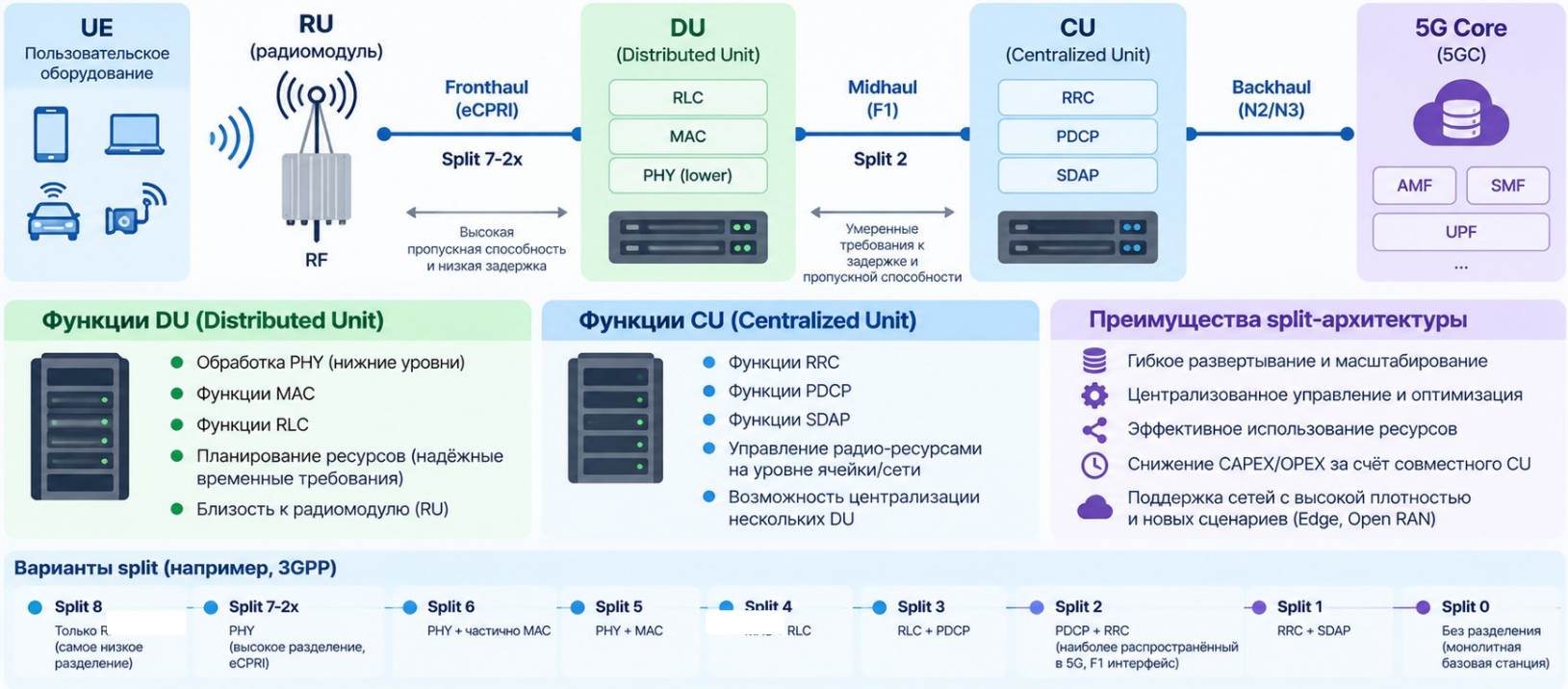
Архитектура сети 5G: автономная (SA) и неавтономная (NSA)



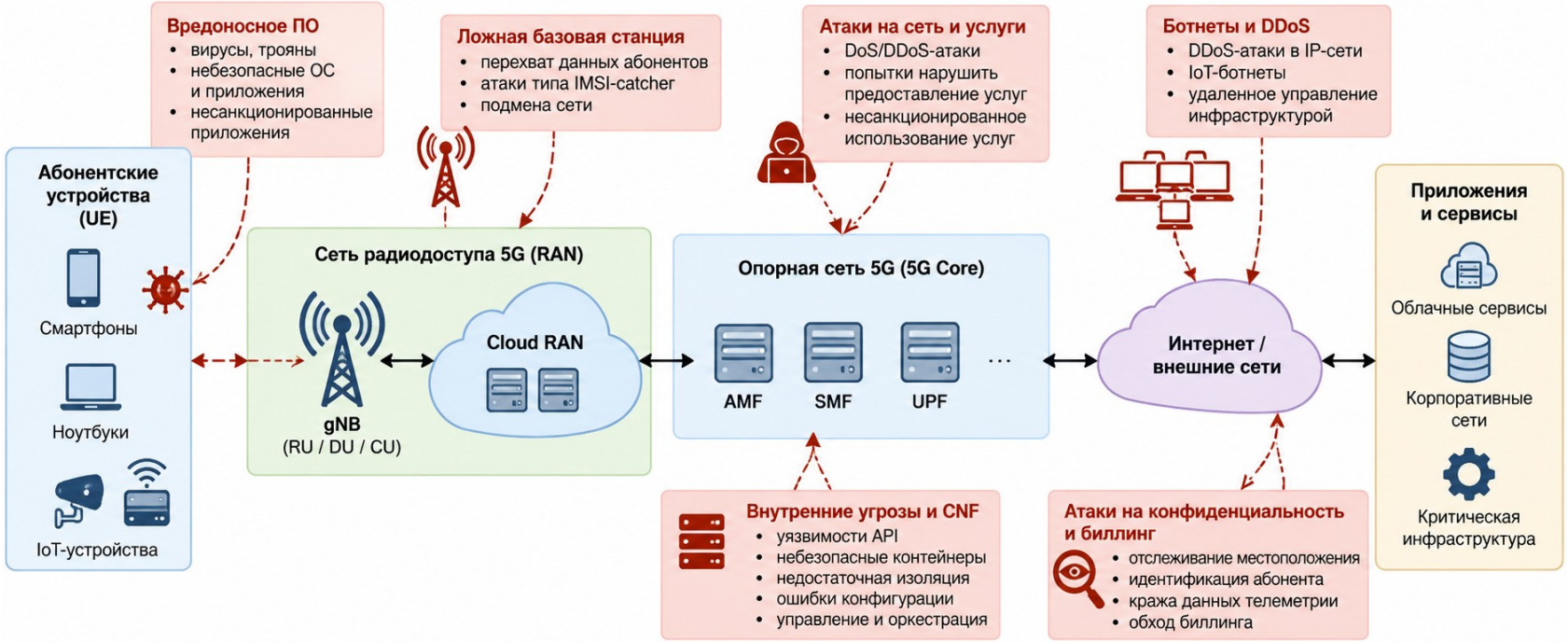
Точки входа для атак на сети 5G в зависимости от сценария ее развертывания могут включать:

- A) Взломанное (скомпрометированное) мобильное устройство UE (4G или 5G).
- B) Сеть радиодоступа 5G RAN (4G/LTE RAN).
- C) Транспортную сеть и базовую сеть 5G Core (4G/LTE EPC).
- D) Внешнюю или стороннюю сеть WiFi.

Архитектура сети 5G: разделение функций базовой станции

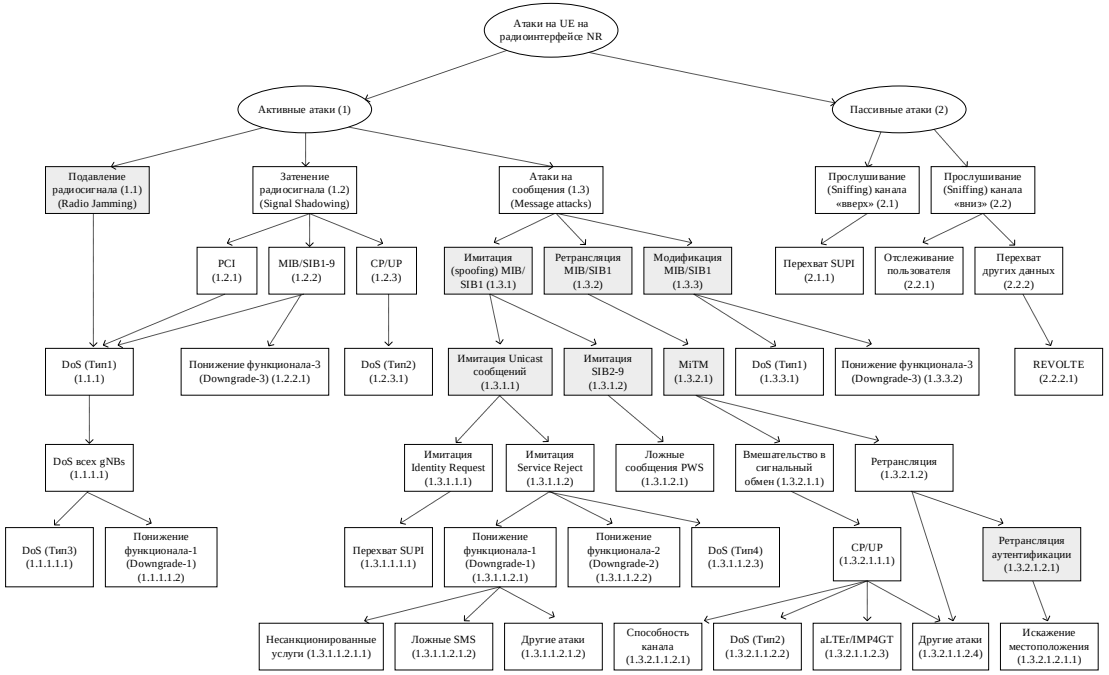


Угрозы в сетях 5G



Классификация атак [2]

- **DoS-атака на абонентский терминал UE:** попытки помешать доступу подключения абонентского терминала UE к сети.
- **DoS-атака на сеть:** попытки помешать сети предоставлять услуги абонентскому терминалу UE
- **Атака на услуги:** попытки абонентских терминалов UE воспользоваться несанкционированными или не запрошенными услугами
- **Атака на конфиденциальность абонента:** попытки идентифицировать подписки абонента или отследить местоположение абонентского терминала UE



[2] 3GPP TR 33.809 Study on 5G security enhancements against False Base Stations (FBS).

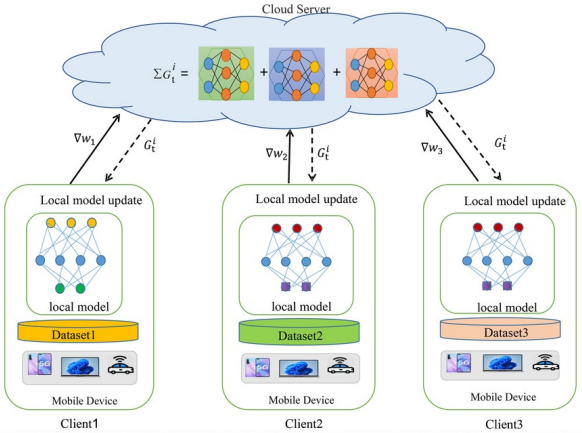
Наследование уязвимостей от LTE

Слабости LTE	Угроза	Влияние на 5G
Перехват IMSI	Конфиденциальность, местоположение	Перехват идентификатора АУ с помощью ложной БС
Запросы Attach/ TAU-update	DoS	DoS атаки на АУ с помощью ложной БС
Переключение на GSM	Человек посередине, перехват звонков и СМС	Неявное переключение на GSM
Отслеживание местоположение с помощью RNTI	Местоположение, оценка трафика и услуг	Профилирование трафика и местоположения
Защита DNS трафика	Перехват DNS сообщений	Человек посередине, кража данных, загрузка вредоносного ПО

Методы искусственного интеллекта в системах IDS 5G/6G

Системы обнаружения вторжений на основе машинного обучения (ML-IDS) будут играть ключевую роль в защите масштабных систем в т.ч. за счет обобщения знаний для распознавания ранее невиданных угроз (например, атак нулевого дня)

- централизованные системы безопасности ограничены недостаточной интеграцией данных, низкой степенью доступности и неоптимальным качеством данных для обучения и в целях тестирования
- федеративное обучение (FL) позволяет множеству устройств, которые могут быть развернуты в хранилищах данных или сети, совместно изучать модель без совместного использования данных



	Превентивный	Предупреждающий	Расследование	Аудит
Блокирование атак?	хороший	средний	плохой	плохой
Выявление атак?	плохой	хороший	плохой	средний
Эксплуатационные последствия	высокие	низкие	низкие	низкие
Расследуемость атак?	плохой	средний	хороший	хороший
Цена внедрения	высокая	средняя	средняя	низкая
Цена эксплуатации	средняя	высокая	низкая	средняя
Гибкость	плохая	плохая	средняя	хорошая

Исследуемые модели машинного обучения

XGBoost
Градиентный бустинг над деревьями решений

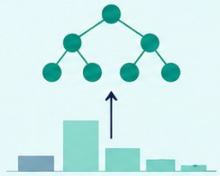
Ансамблевый метод, основанный на градиентном бустинге деревьев решений. Обеспечивает высокую точность обнаружения и устойчивость к разнородным данным.

Используемые метрики

- F1-score
- Accuracy
- Precision
- Recall
- AUC-ROC
- Specificity

LightGBM
Градиентный бустинг с гистограммами

Эффективный алгоритм градиентного бустинга, использующий гистограммный подход и построение деревьев по листьям. Обеспечивает высокую скорость обучения и хорошее качество обнаружения.



Используемые метрики

- F1-score
- Accuracy
- Precision
- Recall
- AUC-ROC
- Specificity

Random Forest
Ансамбль деревьев решений

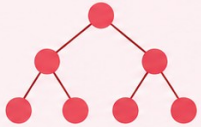
Метод, основанный на ансамбле независимых деревьев решений, обученных на случайных подвыборках данных и признаков. Обеспечивает хорошую обобщающую способность и устойчивость к переобучению.

Используемые метрики

- F1-score
- Accuracy
- Precision
- Recall
- AUC-ROC
- Specificity

Decision Tree
Дерево решений

Простой и интерпретируемый алгоритм, основанный на разбиении пространства признаков с помощью правил. Отличается высокой скоростью работы и наглядностью, но может уступать ансамблевым методам по точности.

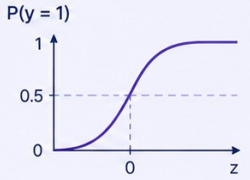


Используемые метрики

- F1-score
- Accuracy
- Precision
- Recall
- AUC-ROC
- Specificity

Logistic Regression
Логистическая регрессия

Линейная модель классификации, оценивающая вероятность принадлежности к классу с помощью сигмоидной функции. Используется как базовый метод для сравнения и отличается простотой и высокой интерпретируемостью.



Используемые метрики

- F1-score
- Accuracy
- Precision
- Recall
- AUC-ROC
- Specificity

Датасеты для обучения

NSL-KDD
Intrusion Detection Dataset

Тип данных	Сетевой трафик (KDD'99, улучшенная версия)
Количество записей	125 973
Количество признаков	41
Классы	Нормальный трафик, 4 типа атак *(DoS, Probe, R2L, U2R)*

Нормальный	53.5%
DoS	36.4%
Probe	9.1%
R2L	0.8%
U2R	0.2%

Особенности
- Устранены дубликаты из KDD'99
- Сбалансированный набор
- Широко используется для сравнительных исследований

UNSW-NB15
Modern Network Intrusion Dataset

Тип данных	Современный сетевой трафик (реальный и смоделированный)
Количество записей	2 540 044
Количество признаков	49
Классы	Нормальный трафик, 9 типов атак *(Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode, Worms)*

Нормальный	44.9%
Generic	18.8%
Exploits	12.8%
Fuzzers	8.5%
DoS	6.2%
Reconnaissance	3.1%
Analysis	2.6%
Backdoors	1.1%
Shellcode	1.0%
Worms	0.1%

Особенности
- Современные сценарии атак
- Реалистичный сетевой трафик
- Несбалансированное распределение классов
- Используется в исследованиях ML/DL

Network Traffic Android Malware
Android Malware Detection Dataset

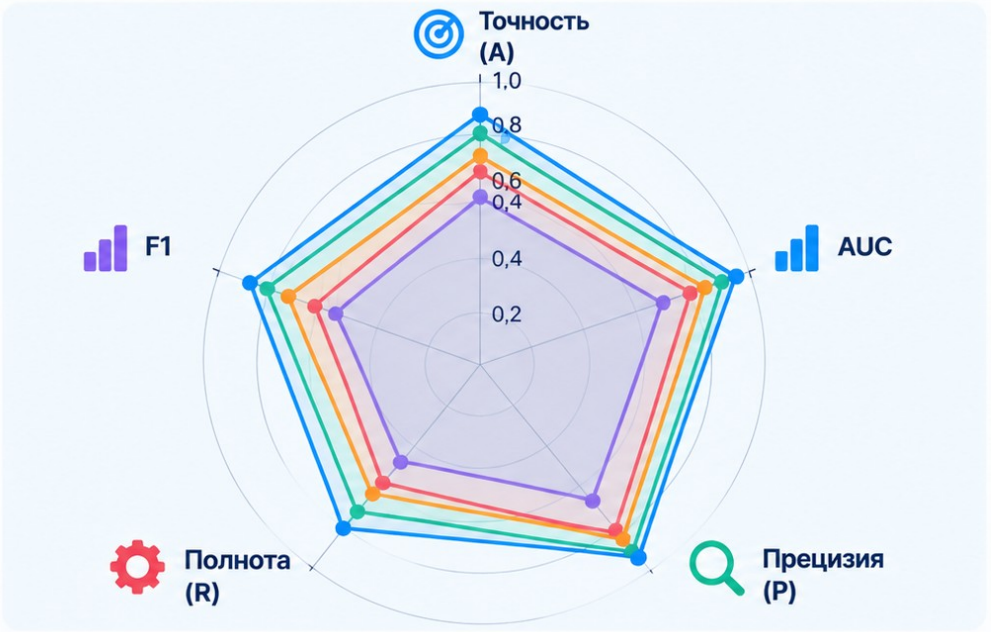
Тип данных	Сетевой трафик Android (нормальные и вредоносные приложения)
Количество записей	104 857
Количество признаков	85
Классы	Нормальные приложения, вредоносные приложения *(10 семейств malware)*

Нормальные	50.0%
Вредоносные	50.0%

Особенности
- Трафик реальных Android-приложений
- 10 семейств вредоносного ПО
- Баланс классов
- Подходит для обнаружения сетевых аномалий и вредоносной активности в мобильных сетях

Российский университет дружбы народов

Оценка эффективности моделей (1/2)



Время обучения моделей
на наборе данных NSL-KDD

Модель	Время обучения, с
XGBoost	2,48
LightGBM	2,71
Random Forest	28,51
Decision Tree	0,6
Logistic Regression	8,236

XGBoost LightGBM Random Forest Decision Tree Logistic Regression

Оценка эффективности моделей (2/2)

Model / Dataset	UNSW-NB15	Network Traffic Android Malware	NSL-KDD
XGBoost	0.9503	0.9476	0.9383
LightGBM	0.8840	0.9759	0.7771
Random Forest	0.8188	0.7532	0.5718
Decision Tree	0.4185	0.6743	0.4011
Logistic Regression	0.3132	0.1407	0.1732

Метод MCDA (многокритериальный анализ решений)

Коэффициент	Значение	Описание
w_1	0.05	AUC-Score
w_2	0.05	Accuracy
w_3	0.25	Precision
w_4	0.5	Recall
w_5	0.1	Размер модели
w_6	0.05	Время обучения

Заключение

1. Ключевое отличие систем 5G/6G от предыдущих поколений – полная виртуализация сетевых функций.
2. Архитектура систем 5G имеет высокую степень превентивной защиты по своей архитектуре. За время эксплуатации не выявлено критических уязвимостей.
3. Для предупреждающего уровня защиты необходимо внедрение систем IDS/IPS на основе методов ИИ и МО.
4. XGBoost и LightGBM показали наилучшую и наиболее стабильную эффективность на всех трёх датасетах. XGBoost обеспечивает высокий баланс качества обнаружения и устойчивости к различиям между наборами данных.