НАЦИОНАЛЬНЫЙ ТЕХНОЛОГИЧЕСКИЙ ЦЕНТР ЦИФРОВОЙ КРИПТОГРАФИИ

РусКрипто XXVII НАУЧНО-ПРАКТИЧЕСКАЯ КОНФЕРЕНЦИЯ

Использование криптографических средств и методов в технологиях обезличивания персональных данных

Серия мини докладов:
Минаков С.С., Клочко Ю.Б., Буланов А.В.

В зависимости от цели обезличивания формализованы 4 основных класса методов обезличивания:

- методы замены;
(использование идентификаторов, псевдонимов, токенов и шифрование с сохранением формата)
- методы изменения состава семантики;
(использованием округления, обобщения, локального подавления, микроагрегации, добавления шума, дифференциальная приватность)
- **методы декомпозиции;**
- **методы перемешивания;**
(возможность использования шифров перестановки)

В соответствии с новой статьей 13.1
ФЗ «О персональных данных» от 27.06.2006 №152-ФЗ
Правительство РФ по согласованию с ФСБ России
устанавливает требования к обезличиванию, к методам
обезличивания и к порядку обезличивания
для операторов ПДн для конкретной цели.

В пп.9.1 статьи 6 такого ФЗ – указана такая цель
(обработка обезличенных персональных данных
в аналитической ГИС)

Т.о. стоит практическая задача выбора таких методов
обезличивания, которые позволят сохранить качество
обезличенных ПДн, пригодным для аналитики.

Алгоритмы оценка риска деобезличивания и оптимизации подбора параметров обезличивания

1. Удаление оператором прямых идентификаторов в наборе данных

2. Поиск уникальных квазиидентификаторов в наборе данных:

Алгоритм **SUDA2** позволяет найти все наборы уникальных значений атрибутов по которым можно однозначно определить человека

3. Оценка риска раскрытия данных путём задания мер риска:

k-анонимность - свойство, которым обладают обезличенные данные. Мера риска, основанная на принципе, согласно которому в безопасном наборе данных число лиц, использующих одну и ту же комбинацию значений (ключей) категориальных квазиидентификаторов, должно превышать заданный порог k

l-разнообразие - свойство, которым обладают обезличенные данные, содержащие чувствительные переменные. Мера риска, основанная на принципе, согласно которому в безопасном наборе для набора записей, у которых совпадают квазиидентификаторы, число значений чувствительной переменной содержит не менее l значений

4. Алгоритм подбора параметров для оптимизации:

Grid Search - алгоритм, который позволяет выбирать лучшие параметры для оптимизации проблемы из списка предоставленных вариантов, тем самым автоматизируя метод «проб и ошибок» и обеспечивая наилучшую точность при подборе метода обезличивания.

SUDA2

Алгоритм **SUDA2** позволяет найти все наборы уникальных значений атрибутов по которым можно однозначно определить человека

Город	Пол	Возраст
Москва	муж	28
Москва	муж	28
Сочи	жен	36
Сочи	жен	36
Сочи	муж	28

или

А) Уникальные записи:

Сочи	муж
Сочи	28

Б) Строки, содержащие уникальные записи:

Город	Пол	Возраст	suda	mu_suda	dis_suda
Москва	муж	28	0	0	0
Москва	муж	28	0	0	0
Сочи	жен	36	0	0	0
Сочи	жен	36	0	0	0
Сочи	муж	28	2	0.667	0.003

SUDA2 — это рекурсивный алгоритм для поиска уникальных значений минимальной выборки. Алгоритм генерирует все возможные подмножества переменных определенных категориальных ключевых переменных и сканирует их на предмет уникальных шаблонов в подмножествах переменных.

К-анонимность и L-разнообразие

Пример К-анонимности

Город	Пол	Возраст
Москва	муж	28
Москва	муж	28
Сочи	жен	36
Сочи	жен	36
Сочи	муж	28

Строки выделенные зеленым удовлетворяют 2-анонимности, а красным – нет

Раскрытие данных обладает свойством k-анонимности, если информацию для каждого человека, содержащуюся в публикации, нельзя отличить по крайней мере от k - 1 лиц, чьи информация также появляется в выпуске.

Пример L-разнообразия с чувствительным атрибутом Диагноз

Город	Пол	Возраст	Диагноз
Москва	муж	28	Грипп
Москва	муж	28	Ковид
Москва	муж	28	Грипп
Сочи	жен	36	Ковид
Сочи	жен	36	Ковид
Сочи	жен	36	Ковид

Строки выделенные красным удовлетворяют 3-анонимности, но у них 1-разнообразие для чувствительного атрибута.
У зеленой группы 3-анонимность и 2-разнообразие

Модель L-разнообразия является расширением модели k-анонимности, которая снижает степень детализации представления данных с использованием методов, включая обобщение и подавление, так что любая данная запись отображается как минимум на k-1 другие записи в данных.

Дифференциальная приватность (ДП)

Дифференциальная приватность (конфиденциальность) — это концепция и методология, **предназначенная для защиты конфиденциальности индивидуальных записей в наборе данных** при сохранении полезности общей информации.

Дифференциальная приватность определяется через механизм, который обеспечивает, что вероятность получения определенного результата анализа данных практически не изменяется, независимо от того, присутствует ли информация об одном конкретном индивиде в наборе данных или нет. Это достигается путем добавления контролируемого количества случайности к результатам запросов, что обеспечивает "приватность через неопределенность".

Ɛ-дифференциальная приватность гарантирует, что для каждого применения алгоритма A результат приблизительно с равной вероятностью будет одновременно наблюдаться на одной из соседних баз данных.

Математические принципы ДП

ε-дифференциальная приватность

Механизм *M* обеспечивает ε-дифференциальную приватность, если для любых двух соседних наборов данных *D* и *D'* (различающихся на одну запись), и для всех *S* в области значений *M*, выполняется условие:

$$\Pr[M(D) \in S] < \exp(\epsilon) \times \Pr[M(D') \in S$$

ϵ — это неотрицательный параметр, известный как параметр приватности, который определяет степень приватности. Чем меньше ϵ, тем выше уровень приватности.

(ε, δ)-дифференциальная приватность

Это расширение ε-дифференциальной приватности, позволяющее небольшую вероятность нарушения ε-дифференциальной приватности. Определение таково:

Определение: Механизм *M* обеспечивает (ε, δ)-дифференциальную приватность, если для всех соседних наборов данных *D* и *D'* и для всех *S* в области значений *M*, выполняется условие:

$$\Pr[M(D) \in S] \leq \exp(\epsilon) \times \Pr[M(D') \in S] + \delta$$

o представляет сооои неоольшую вероятность, при которой ε-дифференциальная приватность может быть нарушена. Таким образом, (ε, δ)-дифференциальная приватность предоставляет дополнительную гибкость, разрешая небольшую вероятность утечки информации.

Теорема Байеса

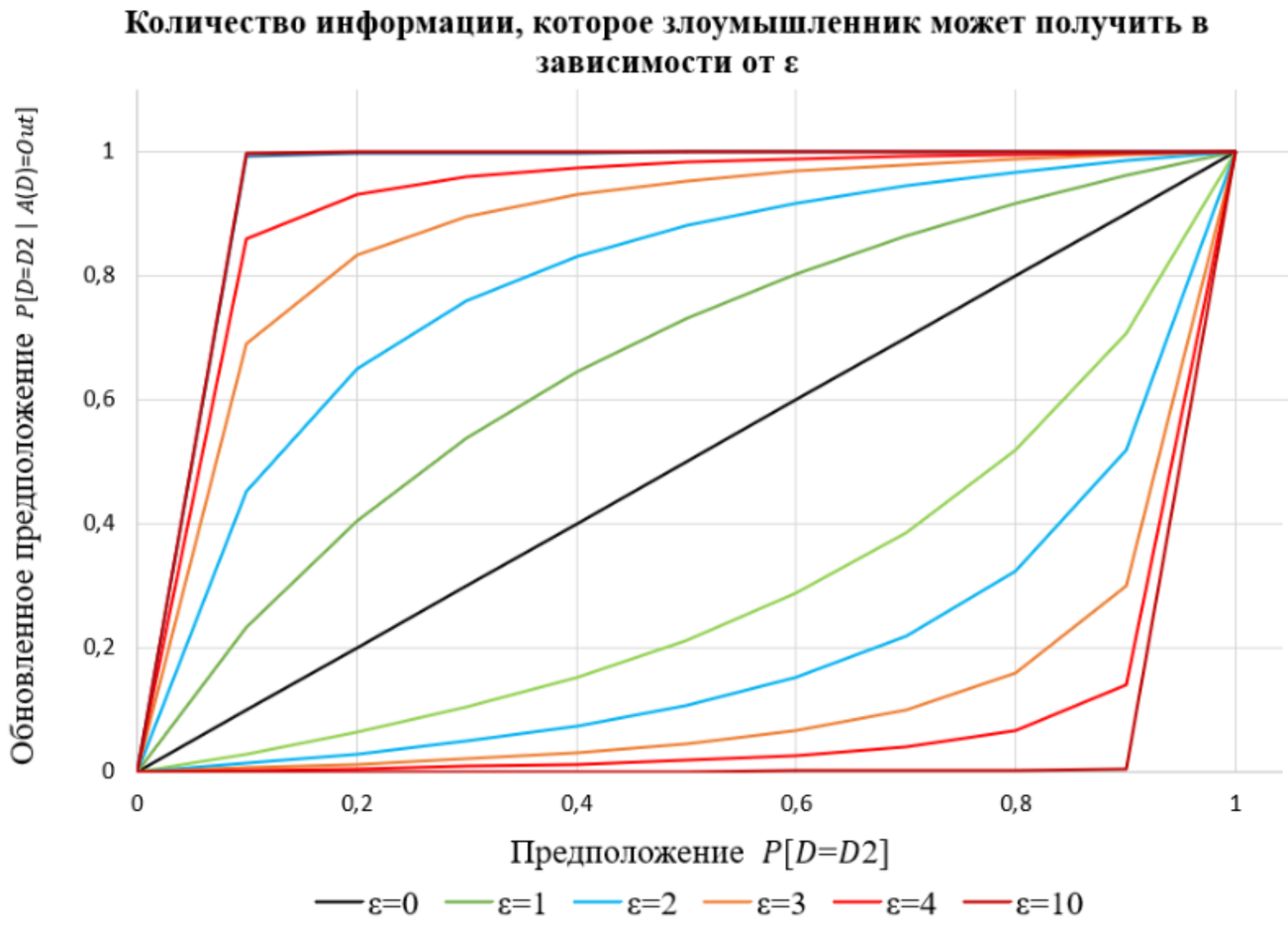
Теорема Байеса (или формула Байеса) — одна из основных теорем элементарной теории вероятностей, которая позволяет определить вероятность события при условии, что произошло другое статистически взаимозависимое с ним событие. Другими словами, по формуле Байеса можно уточнить вероятность какого-либо события, взяв в расчёт как ранее известную информацию, так и данные новых наблюдений.

P(A|B) означает условную вероятность события A, если произошло событие B. P(A) - вероятность события A (гипотеза). P(B) - вероятность события B.

$$P(A|B) = P(B|A)\frac{P(A)}{P(B)}$$

Вероятность события может быть от 0% до 100% (или от 0 до 1). Если обозначим событие буквой A, то вероятность этого события (A) будет P(A). Например, на шестигранном кубике 6 граней и вероятность выпадения любого числа одинакова, поэтому вероятность выпадения любого числа будет 1/6 (или 0.166..; или 16,66..%). P(1)=P(2)..=P(6)=16.66..%. Сумма же вероятностей всех (несовместимых) событий будет 100% (или 1).

График, отображающий количественную оценку Ɛ-дифференциальной приватности



Предположим, что к базе данных D применен алгоритм A, который вносит случайность. Цель злоумышленника – выяснить, находится ли определенный человек в базе данных. Рассмотрим худший случай, когда злоумышленнику известна вся база данных, однако он не знает в какой из двух баз данных D1 или D2 находится интересующий его человек. Атакующий может сделать предположение что его цель находится в базе данных D1 с некоторой вероятностью $\in[0,1]$. Точно таким же образом может быть выдвинуто подозрение, что в D2 цель злоумышленника отсутствует.

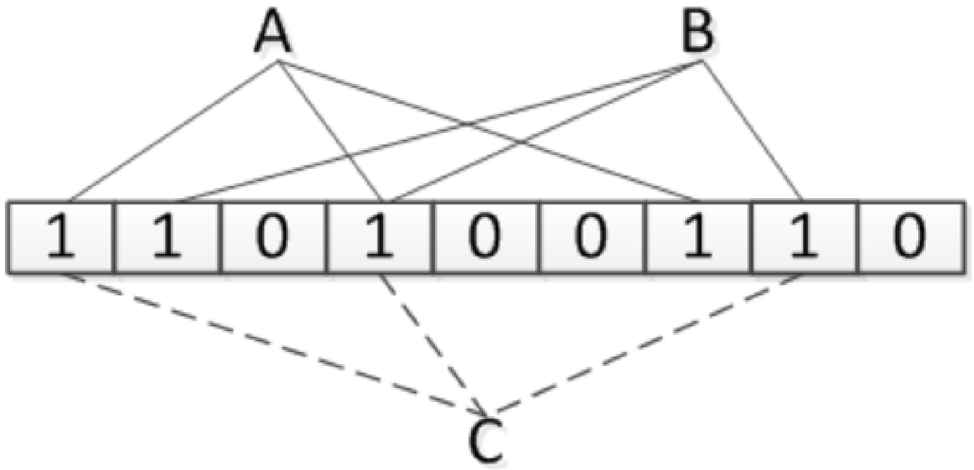
Предположим, что после применения A к базе данных возвращается выходное значение Out, исходя из которого атакующий может сделать новое предположение, которое определяется как $P[D = D2 \mid A(D) = Out]$. Применяя теорему Байеса, получаем:

$$P[D = D2 \mid A(D) = Out] = \frac{P[D = D2\,] * P[\,A(D) = Out | D = D2\,]}{P[\,A(D) = Out]},$$

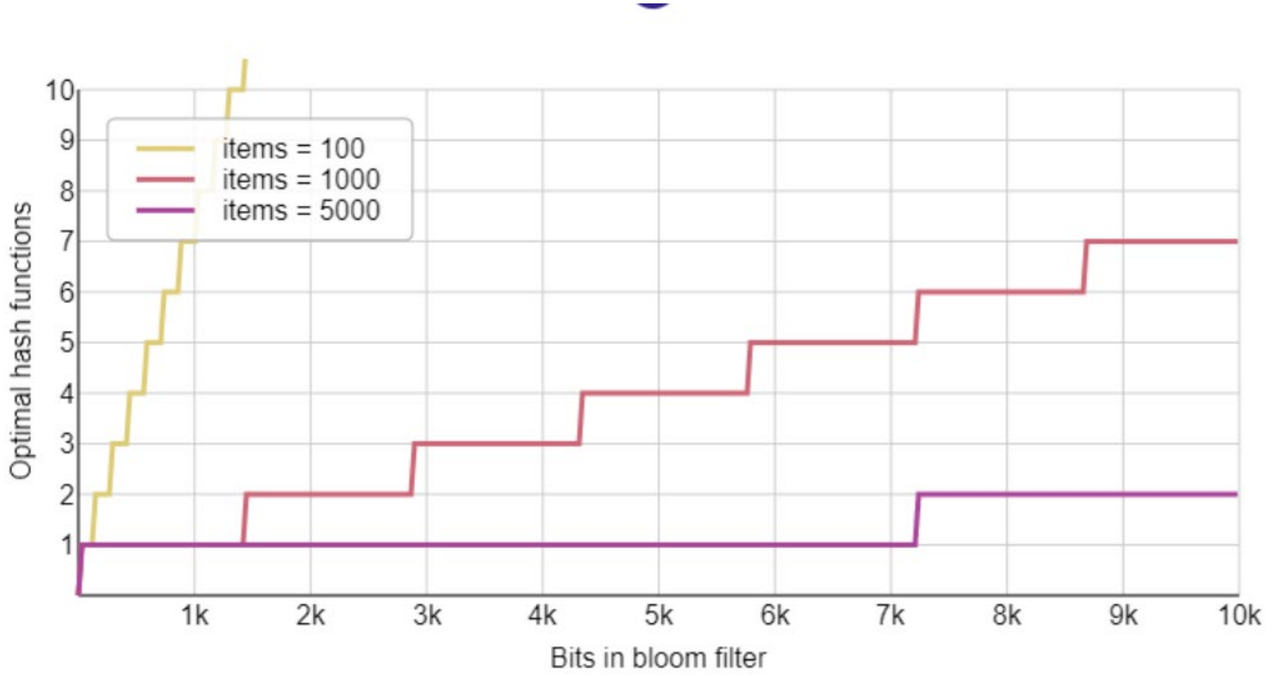
где $P[D = D2\,]$ - изначальное предположение злоумышленника, $P[\,A(D) = Out | D = D2\,]$– вероятность получения значения Out из базы данных $D2$, $P[\,A(D) = Out]$ – вероятность того, что в результате применения алгоритма A к $D2$ выходным значением будет Out.

Фильтр Блума

Фильтр Блума — это **вероятностная структура данных**, при проверке наличия элемента, фильтр Блума предоставляет вероятностный ответ. Он может точно определить, что элемент отсутствует, но не может гарантировать наличие элемента.
Фильтр Блума — это массив **битов**, все значения которых изначально равны 0. Фильтр Блума представляет собой битовый массив из m бит и k различных хеш-функций h1…hk, равновероятно отображающих элементы исходного множества во множество {0,1,…m−1}, соответствующее номерам битов в массиве. Изначально, когда структура данных хранит пустое множество, все m бит обнулены.



Фильтр Блума с $m = 9$ и $k = 3$, хранящий множество из элементов A и B. Этот фильтр Блума может определить, что элемент C входит в множество, хотя он и не добавлен в него.



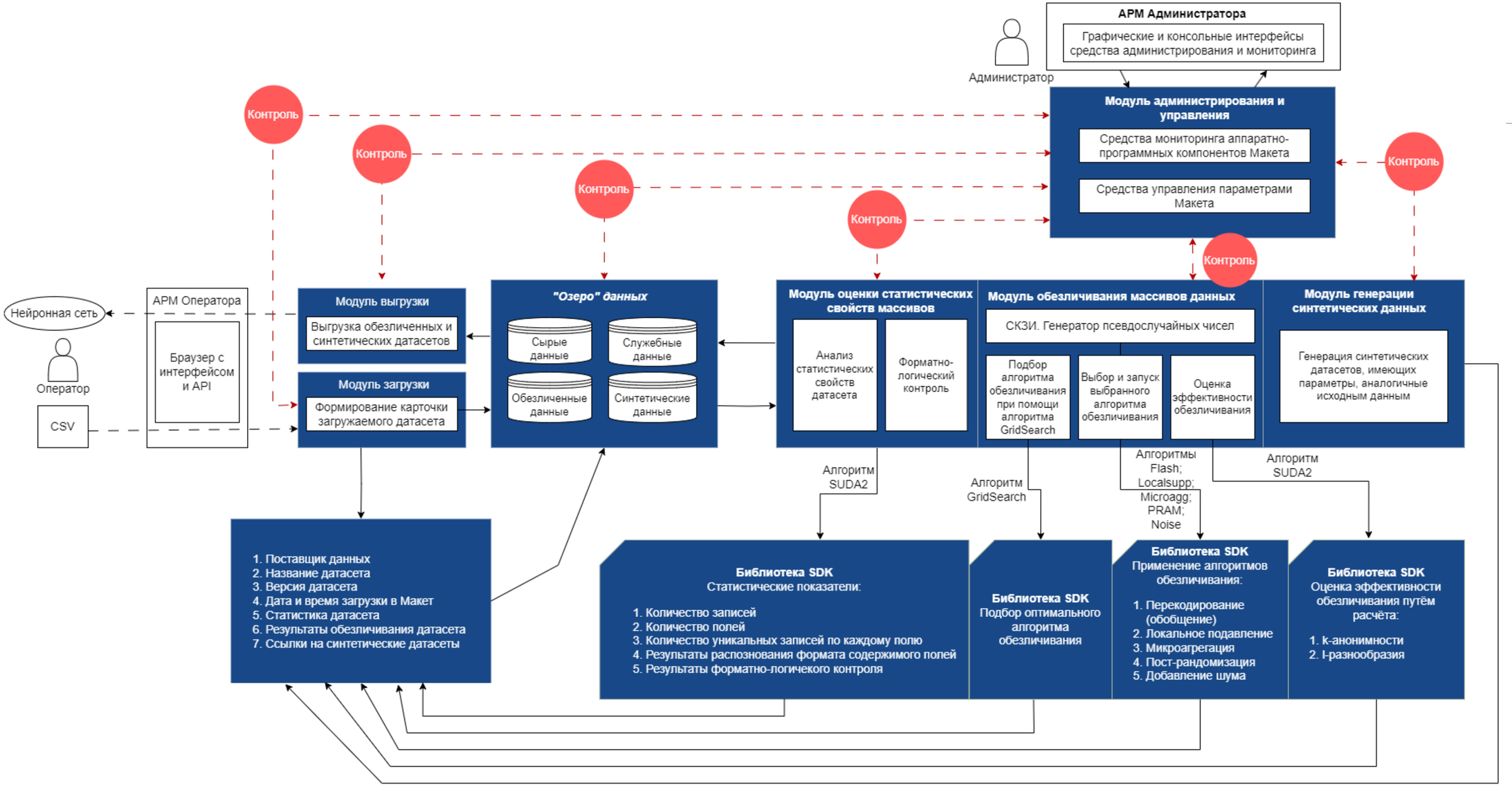
Чем больше элементов планируется добавить, тем меньше хэш-функций должно использоваться. С другой стороны, чем больше размер фильтра Блума, тем больше хэш-функций можно использовать. Большее количество хэш-функций уменьшает вероятность ложноположительных результатов. Однако увеличение количества элементов может привести к быстрому заполнению фильтра Блума.

Глобальная (централизованная) и локальная модели

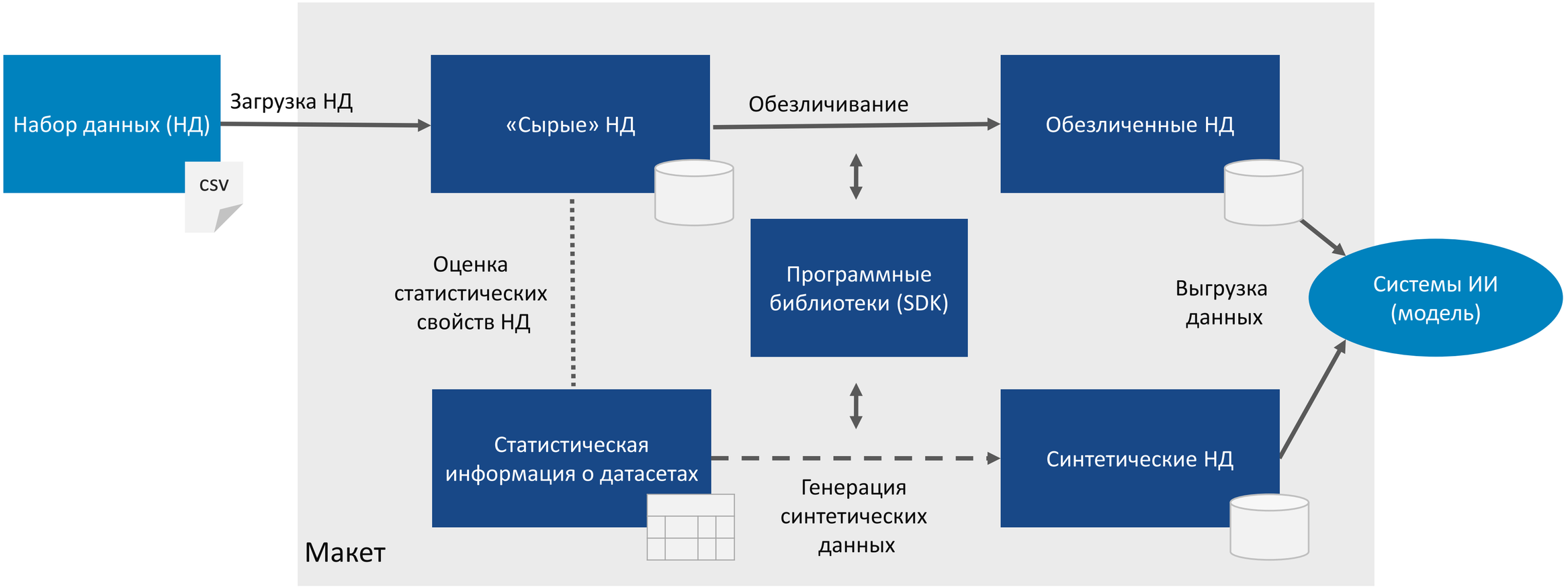
В глобальной (централизованной) модели каждый участник доверяет службе, которая собирает информацию, поэтому данные от пользователей отправляются без добавления шума. Методы Гаусса и Лапласа.

В локальной модели каждый пользователь перед предоставлением информации какой-либо службе применяет к своим данным рандомизированный алгоритм.

Функциональная схема аналитического комплекса



«Конвейер» комплекса



Программные библиотеки (SDK) обезличивания

РусКрипто XXVII

Оценка риска РПД

- SUDA 2
- k-анонимность
- l-разнообразие

Обезличивание

- Добавление шума
- Микроагрегация
- Локальное подавление
- Пост-рандомизация
- Перемешивание (с обобщением)
- Подбор параметров (GridSearch)

Общие функции

- Генерация синтетических данных
- Статистика
- Примеры и документация

Подходы по обезличиванию геотреков

1. Метод обезличивания геотреков в широкополосной сети связи с большой пропускной способностью
2. Метод обезличивания геотреков, основанный на n-граммах
3. Метод обезличивания геотреков в сервисах LBS на основе ступенчатого рандомизированного отклика
4. Метод публикации обезличенных геотреков на основе графов
5. Метод обезличивания траекторий с обеспечением оптимальной персонализированной защиты
6. Метод обезличивания геотреков пользователей социальных сетей с применением опорной системы
7. Механизм генерации траекторий, обеспечивающий дифференциальную защиту (TGM)
8. Агрегируемый рандомизированный отклик RAPPOR

Результаты исследований. ДП

Проведенный анализ позволил выделить следующие перспективные механизмы для практической реализации, выбранные механизмы позволяют работать с различными типами исходных данных, относятся как к локальным, так и централизованным механизмам ДП.

1. Механизм Гаусса, (ε, δ)-приватный, централизованный, работает с вещественными, целыми числами (используется Бюро переписи населения США).
2. Механизм Лапласа, ε-приватный, централизованный, работает с вещественными, целыми числами.
3. Механизм рандомизированного отклика, ε-приватный, локальный, работает с категориальными признаками, требует единый словарь значений
4. Механизм рандомизированного отклика к выходу фильтра Блума, ε-приватный, локальный, работает с категориальными признаками (используется Google, RAPPOR).

Механизмы дифференциальной приватности

Механизм Лапласа

Механизм Лапласа является одним из наиболее распространенных методов для достижения ε-дифференциальной приватности. Он работает путем добавления шума, который следует распределению Лапласа, к результатам запросов данных. Чтобы применить механизм Лапласа к функции *f*, к выходным данным добавляется шум, генерируемый согласно распределению Лапласа.

$$\lambda = \Delta f / \epsilon$$

Здесь Δ*f* обозначает чувствительность функции *f*, а ϵ — параметр приватности.

$$(\mathcal{N}(0, \sigma^2))$$

где σ определяется на основе чувствительности функции *f*, параметров ε и δ. Выбор параметров ε и δ в этом случае более сложен и зависит от требуемого уровня приватности и допустимой вероятности нарушения.

Гауссов механизм

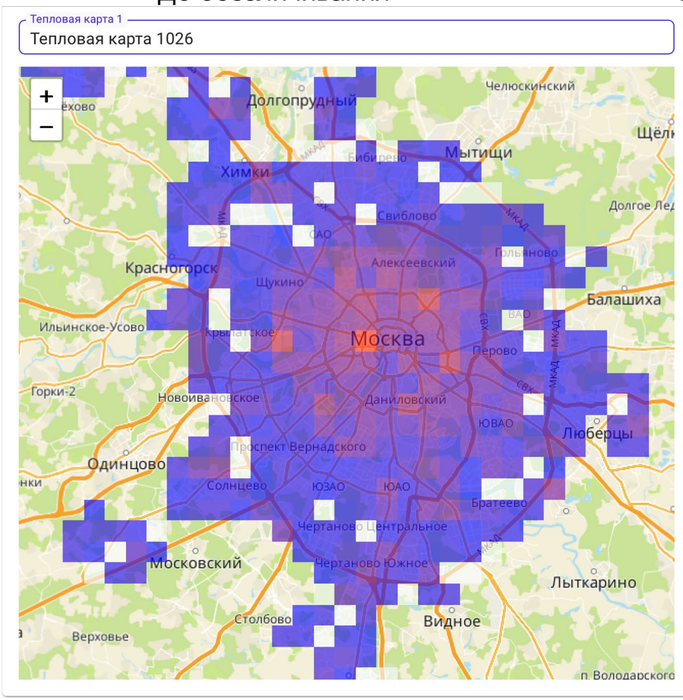
Гауссов механизм применяется для достижения (ε, δ)-дифференциальной приватности, используя шум, следующий нормальному распределению.

В этом механизме к результатам функции *f* добавляется шум, соответствующий нормальному распределению

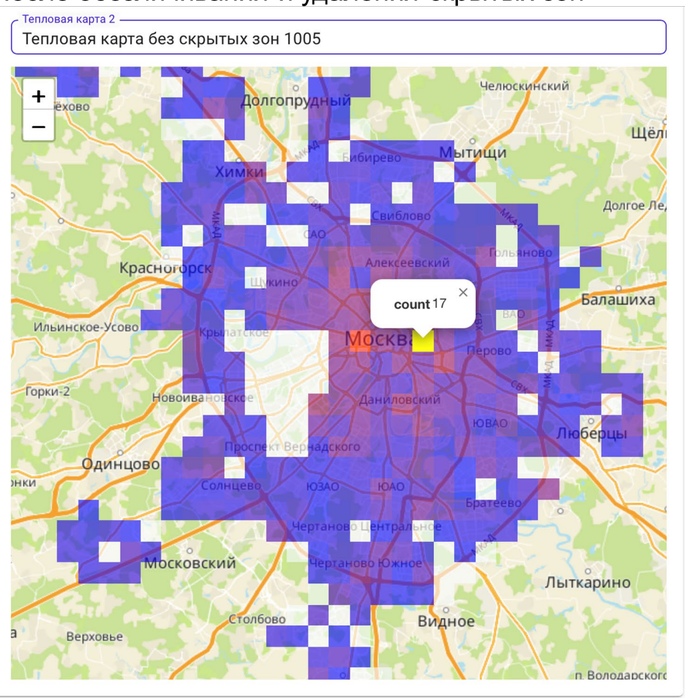
Результаты исследований. Геотреки

1. На основе проведенного анализа подходов к обезличиванию данных о геопозиционировании абонента сделан вывод о необходимости разбиения траекторий абонентов и сведение анализа траекторий к последовательному анализу тепловых карт.
2. Предложена технология получения последовательности тепловых карт на основе механизмов, обеспечивающих ε-дифференциально-приватный результат, с последующим сокрытием цензурируемых областей карты.
3. Описаны методики оценки качества тепловых карт путем решения задачи А/Б тестирования.
4. Предложена методика атаки на цензурируемые области карты путем интерполяции данных методами кригинга, ИИ.

До обезличивания



После обезличивания и удаления скрытых зон



Результаты исследований. Геотреки

Реализованы SDK следующих методов обезличивания:

✓ Обезличивание методом округления

✓ Обезличивание методом обобщения

Методы дифференциальной приватности для «обычных» данных и геотреков:

- Рандомизированный отклик
- Рандомизированный отклик к выходу фильтра Блума
- Механизм Лапласа
- Механизм Гаусса

Оценка оператором

Обезличивание наборов данных – это **всегда компромисс** между риском раскрытия и полезностью данных для конечных пользователей!

Окончательный выбор метода обезличивания всегда **делает оператор**, исходя из содержимого каждого конкретного набора данных, целей его обезличивания и дополнительных знаний оператора о датасете, полученных из различных источников.

РусКрипто
XXVII научно-практическая конференция

Спасибо
за внимание!

НТЦ ЦК