DATA FUSION

FIRST RUSSIAN DATA FORUM

# РИСК-ОРИЕНТИРОВАННАЯ ОБРАБОТКА ДАННЫХ

**НЕЙМАН АЛЕКСЕЙ**

*ИСПОЛНИТЕЛЬНЫЙ ДИРЕКТОР АССОЦИАЦИИ БОЛЬШИХ ДАННЫХ*

АССОЦИАЦИЯ БОЛЬШИХ ДАННЫХ

DATA FUSION
КОНФЕРЕНЦИЯ ПО АНАЛИЗУ ДАННЫХ И ТЕХНОЛОГИЯМ ИИ

FIRST RUSSIAN DATA FORUM

# КАКОВЫ РИСКИ ОБРАБОТКИ КЛИЕНТСКИХ ДАННЫХ?

АССОЦИАЦИЯ БОЛЬШИХ ДАННЫХ

# Риск-модель обработки клиентских данных

## ТЕХНОЛОГИИ ЗАЩИТЫ ПЕРСОНАЛЬНЫХ ДАННЫХ

### 1 ОРГАНИЗАЦИОННЫЕ И ОПЕРАТИВНО-ТЕХНИЧЕСКИЕ МЕРЫ

Организационные и оперативно технические меры ведут к управлению контекстными рисками, которые оцениваются с помощью скоринг-модели:

$$P_{\text{контекстные риски}} = \frac{\sum_{j=1}^{n} \omega_j K_j}{\sum_{j=1}^{n} \omega_j}$$

Контекстные риски отражают способность организаций противостоять угрозам утечек данных и поддерживать процессы управления персональными данными, снижая общие угрозы

### 2 КРИПТОГРАФИЧЕСКИЕ МЕТОДЫ

### 3 МЕТОДЫ ПСЕВДОНИМИЗАЦИИ

Методы псевдонимизации – методы защиты персональных данных, при которых прямые и/или косвенные атрибуты в конкретных наборах данных заменяются на один или несколько искусственных идентификаторов

$$P_{\text{риски данных}} = \frac{k}{\langle r|R \rangle}$$

При расчете рисков данных в условиях применения методов псевдонимизации величина рисков обратна пропорциональна затраченным ресурсам (например, времени на взлом при атаках прямым перебором)

### 4 ОБЕЗЛИЧИВАНИЕ (АНОНИМИЗАЦИЯ)

Методы обезличивания нацелены на исключение связи отдельных атрибутов с идентификаторами. Основная идея для данного класса методов: выделении групп схожих записей (классов эквивалентности), которые могут быть отнесены более чем к одному физическому лицу. В этом случае нарушается основное свойство определения персональных данных – соотнесение информации с прямо *или косвенно определенным или определяемом* физическом лицом.

$$P_{\text{риски данных}} = \frac{k}{\langle \breve{E} \rangle}$$

Вероятность повторной идентификации обратна пропорциональна размеру наименьшего класса эквивалентности в наборе

### 5 КОНФИДЕНЦИАЛЬНЫЕ ВЫЧИСЛЕНИЯ

В рамках таких методах используются комплексные криптографические схемы, а в качестве метрик риска – оценки вероятности повторения значений совокупности ключей и результатов вычисления крипто-функций на всем пространстве определения (так называемая крипто-игра

$$P = \frac{r_{train} - r_{control}}{1 - r_{control}}$$

Доверительный интервал биноминального распределения по методу Уилсона

### 6 СТАТИСТИЧЕСКИЕ МЕТОДЫ И МАШИННОЕ ОБУЧЕНИЕ

Данная группа методов основана применении методов машинного обучения или статистического подхода. Методы объединяет идея замены исходных наборов данных новыми информационными массивами, сохраняющими свойства исходных наборов, но с измененными значениями.

$$P_{\text{риски данных}} = \max\left(\frac{1}{N}\sum_{s=1}^{n}\left(\frac{1}{f_s} \times I_s\right), \frac{1}{N}\sum_{s=1}^{n}\left(\frac{1}{F_s} \times I_s\right)\right)$$

**Обобщенная формула Эль-Эмама**

Для случаев применение дифференциальной приватности количество шума, добавленного к данным, может быть количественно определено с помощью значения $\varepsilon$ (меньшее значение $\varepsilon$ подразумевает более высокий уровень защиты).

**Некоторые из указанных методов могут применяться последовательно, что позволяет дополнительно факторизовать (разложить на множители формулу расчета, например:**

$$\boldsymbol{P}_{\text{повторной идентификации}} = \boldsymbol{P}_{\text{контекстные риски}} \times (\boldsymbol{P}_{\text{синтетические наборы}} \times \boldsymbol{P}_{\text{обезличивание}})$$

# Риск-модель обработки клиентских данных

## ТЕХНОЛОГИИ ЗАЩИТЫ ПЕРСОНАЛЬНЫХ ДАННЫХ

### 1 ОРГАНИЗАЦИОННЫЕ И ОПЕРАТИВНО-ТЕХНИЧЕСКИЕ МЕРЫ

Организационные и оперативно технические меры ведут к управлению контекстными рисками, которые оцениваются с помощью скоринг-модели:

$$P_{\text{контекстные риски}} = \frac{\sum_{j=1}^{n} \omega_j K_j}{\sum_{j=1}^{n} \omega_j}$$

Контекстные риски отражают способность организаций противостоять угрозам утечек данных и поддерживать процессы управления персональными данными, снижая общие угрозы

### 2 КРИПТОГРАФИЧЕСКИЕ МЕТОДЫ

### 3 МЕТОДЫ ПСЕВДОНИМИЗАЦИИ

Методы псевдонимизации – методы защиты персональных данных, при которых прямые и/или косвенные атрибуты в конкретных наборах данных заменяются на один или несколько искусственных идентификаторов

$$P_{\text{риски данных}} = \frac{k}{\langle r|R \rangle}$$

При расчете рисков данных в условиях применения методов псевдонимизации величина рисков обратна пропорциональна затраченным ресурсам (например, времени на взлом при атаках прямым перебором)

### 4 ОБЕЗЛИЧИВАНИЕ (АНОНИМИЗАЦИЯ)

Методы обезличивания нацелены на исключение связи отдельных атрибутов с идентификаторами. Основная идея для данного класса методов: выделении групп схожих записей (классов эквивалентности), которые могут быть отнесены более чем к одному физическому лицу. В этом случае нарушается основное свойство определения персональных данных – соотнесение информации с прямо *или косвенно определенным или определяемом* физическом лицом.

$$P_{\text{риски данных}} = \frac{k}{\langle \breve{E} \rangle}$$

Вероятность повторной идентификации обратна пропорциональна размеру наименьшего класса эквивалентности в наборе

### 5 КОНФИДЕНЦИАЛЬНЫЕ ВЫЧИСЛЕНИЯ

В рамках таких методах используются комплексные криптографические схемы, а в качестве метрик риска – оценки вероятности повторения значений совокупности ключей и результатов вычисления крипто-функций на всем пространстве определения (так называемая крипто-игра

$$P = \frac{r_{train} - r_{control}}{1 - r_{control}}$$

Доверительный интервал биноминального распределения по методу Уилсона

### 6 СТАТИСТИЧЕСКИЕ МЕТОДЫ И МАШИННОЕ ОБУЧЕНИЕ

Данная группа методов основана применении методов машинного обучения или статистического подхода. Методы объединяет идея замены исходных наборов данных новыми информационными массивами, сохраняющими свойства исходных наборов, но с измененными значениями.

$$P_{\text{риски данных}} = \max\left(\frac{1}{N}\sum_{s=1}^{n}\left(\frac{1}{f_s} \times I_s\right), \frac{1}{N}\sum_{s=1}^{n}\left(\frac{1}{F_s} \times I_s\right)\right)$$

**Обобщенная формула Эль-Эмама**

Для случаев применение дифференциальной приватности количество шума, добавленного к данным, может быть количественно определено с помощью значения $\varepsilon$ (меньшее значение $\varepsilon$ подразумевает более высокий уровень защиты).

**Некоторые из указанных методов могут применяться последовательно, что позволяет дополнительно факторизовать (разложить на множители формулу расчета, например:**

$$\boldsymbol{P}_{\text{повторной идентификации}} = \boldsymbol{P}_{\text{контекстные риски}} \times (\boldsymbol{P}_{\text{синтетические наборы}} \times \boldsymbol{P}_{\text{обезличивание}})$$

DATA FUSION
КОНФЕРЕНЦИЯ ПО АНАЛИЗУ ДАННЫХ И ТЕХНОЛОГИЯМ ИИ

FIRST RUSSIAN DATA FORUM

# КАК ТЕХНОЛОГИИ СНИЖАЮТ РИСКИ ОБРАБОТКИ КЛИЕНТСКИХ ДАННЫХ ?

АССОЦИАЦИЯ БОЛЬШИХ ДАННЫХ

# Маркетинговое касание

## БИЗНЕС-КЕЙС:

### БАНК

- обогащенный профиль клиента
- покупки товара

### РЕКЛАМНАЯ ПЛОЩАДКА

- просмотры рекламы товара

### АНАЛИТИКА

Конечной целью схемы обмена данными является построение аналитики об эффективности рекламы

**На каждом этапе данные должны быть обезличены/ синтезированы как банком, так и рекламной площадкой!**

## ТЕХНОЛОГИИ:

- **Дифференциальная приватность**: добавление шума Лапласа (так же были опробованы и другие виды шумов) в рамках прикладного метода сохранения конфиденциальности
- **Конфиденциальное пересечение данных**: метод Шнорра без реального обмена конфиденциальной информацией
- **Моделирование на пересеченных данных:** аналитическое вычисление конверсии

## ВЫЗОВЫ:

- **Совокупная оценка рисков по бизнес-кейсу с применением используемых технологий**
- **Обеспечение приемлемого уровня качества модели с учетом порога допустимого риска**
- **Баланс безопасности/полезности**

# Риск-модель по кейсу

Модель оценивает риск утечки конфиденциальной информации из синтетических данных, а также вероятность идентификации или восстановления исходных данных из защищенного набора данных. Утечка относится к возможности извлечения конфиденциальной информации о реальных данных или людях из синтетических данных.

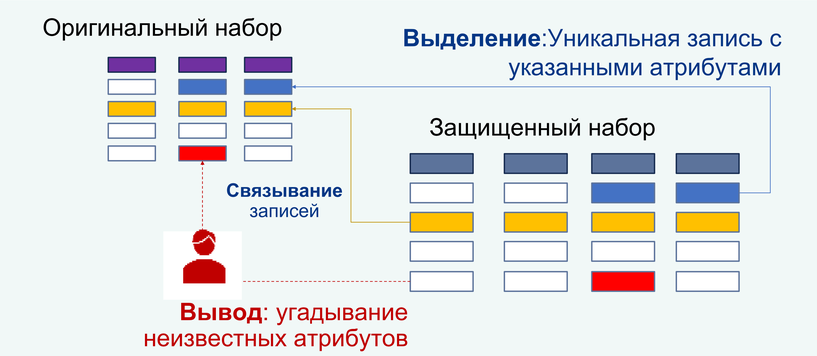
## РИСКИ ДАННЫХ

Риск утечки информации можно оценить, рассмотрев три ключевых аспекта (Giomi Matteo, 2022):

**Выделение (Singling Out) -** оценка вероятности того, что в исходном наборе данных существует уникальная запись с определенной комбинацией атрибутов

**Связывание (Linkability risk)** относится к возможности связывать записи, принадлежащие одному и тому же лицу или группе лиц, в исходном и синтетическом наборах

**Вывод (Inference):** Возможность угадывать неизвестные атрибуты исходной записи данных из объединенных (синтетических) данных




## РИСКИ ИНТЕГРАЦИИ:

По схеме Шнорра сводятся к контекстных рискам, включают риски конфиденциальности ключей, защиту каналов, организационно-технические риски

## ОБЩЕЕ УРАВНЕНИЕ РИСКА:

$$\boldsymbol{R_{total} = R_{context} \times (w_1 \times R_{snglout} + w_2 \times R_{link} + w_3 \times R_{inf})}$$

Здесь $w_i$ – веса, которые могут быть приняты равными первому приближению($w_i \approx 0.3333$)

| | | |
|---|---|---|
| Риск выделения | $R_{singlout} \leftarrow \hat{p} \sim R_{DP} \sim e^{\epsilon}$ | Считается на основе бюджета дифференциальной приватности |
| Риск связывания | $R_{link} \leftarrow \hat{p} \sim cos\theta = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$ | Рассчитывается в рамках CVPL-атаки, как мера “схожести” векторов внутри выделенного кластера |
| Риск вывода | $R_{inf} \leftarrow \hat{p} = \frac{H(a_j \mid b_i)}{\log_2 4} = -\frac{1}{2}\sum_{i=1}^{n} p_i \log_2 p_i$ | Рассчитывается через приведенную энтропию (условную энтропию) как отражение меры неопределенности |

**Расчет рисков по отношению к исходному (реальному) набору дает максимальное значение в сравнении с другими (неизвестными, ранее утекшими наборами)**

Риски интеграции по Шнорру сводятся к контекстным рискам и рассчитываются с помощью скоринговой модели:

$$P(S) = 1 - (P_K \times P_T \times P_O) = P_{\text{контекстные риски}}$$

Слагаемые риска данных оцениваются через интервал доверительной вероятности Вильсона:

$$WI = \left[ \frac{\hat{p} + \frac{z_{\frac{\alpha}{2}}^2}{2n} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z_{\frac{\alpha}{2}}^2}{4n^2}}}{1 + \frac{z_{\frac{\alpha}{2}}^2}{n}}, \frac{\hat{p} + \frac{z_{\frac{\alpha}{2}}^2}{2n} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z_{\frac{\alpha}{2}}^2}{4n^2}}}{1 + \frac{z_{\frac{\alpha}{2}}^2}{n}} \right]$$

Здесь:

- $\hat{p}$ — Наблюдаемая доля выборки, n — Объем выборки,
- $z_{\alpha/2}$ –Квантиль стандартного распределения для уровня достоверности(1-α):

# Результаты кейса

Применение мер защиты незначительно снизило качество данных, как по метрикам, так и по целевым бизнес-характеристикам. Ожидаемое соотношение – при увеличении шума снижаются метрики риска, включая риски атак (растет степень защиты), но также снижается качество)

1. Защита информации (размывание) – добавление Лапласового шума:

$$f(x|0,b) = \frac{1}{2\frac{\Delta f}{\epsilon}} exp\left(-\frac{|x| \times \epsilon}{\Delta f}\right)$$

2. Композитные метрики качества:

Коэффициент Пирсона

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Тест Колмогорова -Смирнова

$$D_{n,m} = \sup_x \left|F_{1,n}(x) - F_{2,m}(x)\right|$$

Композитная метрика качества основана на нормировании и суммы коэффициента Пирсона и KS

$$Q = \frac{1}{2}(1 - D_{n,m} + r)$$

3. Композитные метрики безопасности:

Расчет риска по модели информационной утечки: риски выделения (объем конфиденциальности), риски связывания (CVPL атаки), риски вывода (энтропия)

$$P_{data} = w_1 \times P(I)_{DP} + w_2 \times P_{link} + w_3 \times P_{inf}$$

Композитная метрика безопасности:

$$S = |1 - 3 \times R|, \text{ где}$$
$$R = P_{data} \times P_{shnorr}$$

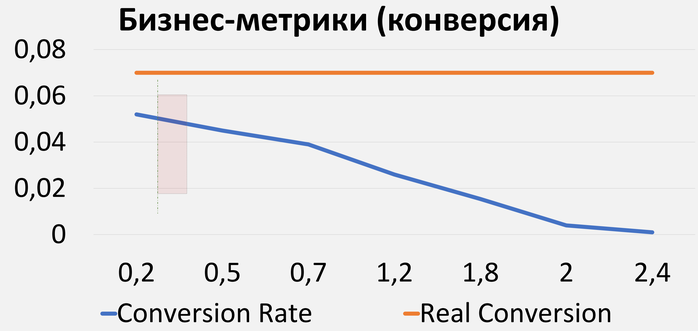
4. Бизнес-метрики:

$$Conversion\ Rate = \left(\frac{Number\ of\ Conversions}{Total\ Number\ of\ Interactions}\right) \times 100\%$$

| | КАЧЕСТВО ДАННЫХ | БЕЗОПАСНОСТЬ | БИЗНЕС-МЕТРИКА |
|---|---|---|---|
| ПРЯМАЯ ОБРАБОТКА | 100% | 0% | 7% |
| ЗАЩИЩЕННАЯ ОБРАБОТКА | 86% | 99.98% | 5% |

Поведение композитных метрик и бизнес-значения при изменении шума

Конверсия на реальном наборе составляет 7% и падает при размывание данных



DATA FUSION
КОНФЕРЕНЦИЯ ПО АНАЛИЗУ ДАННЫХ И ТЕХНОЛОГИЯМ ИИ

FIRST RUSSIAN DATA FORUM

# *Как поместить технологии в регуляторные рамки?*

АССОЦИАЦИЯ БОЛЬШИХ ДАННЫХ

# Подход к регулированию обработки данных

- Риски обработки клиентских данных могут (и должны) быть измерены для конкретного бизнес-кейса
- Существуют техники и технологии снижения риска реидентификации до околонулевых значений
- Использование технологий лежит в «серой» зоне нормативного регулирования, не успевающего за их развитием
- ЭПР для апробации технологий не зарекомендовал себя как рабочий инструмент
- Техническая стандартизация + закрепление модели управления риском – строгая законодательная регламентация = быстрое внедрение технологий при сохранении должного уровня конфиденциальности

DATA FUSION

FIRST RUSSIAN DATA FORUM

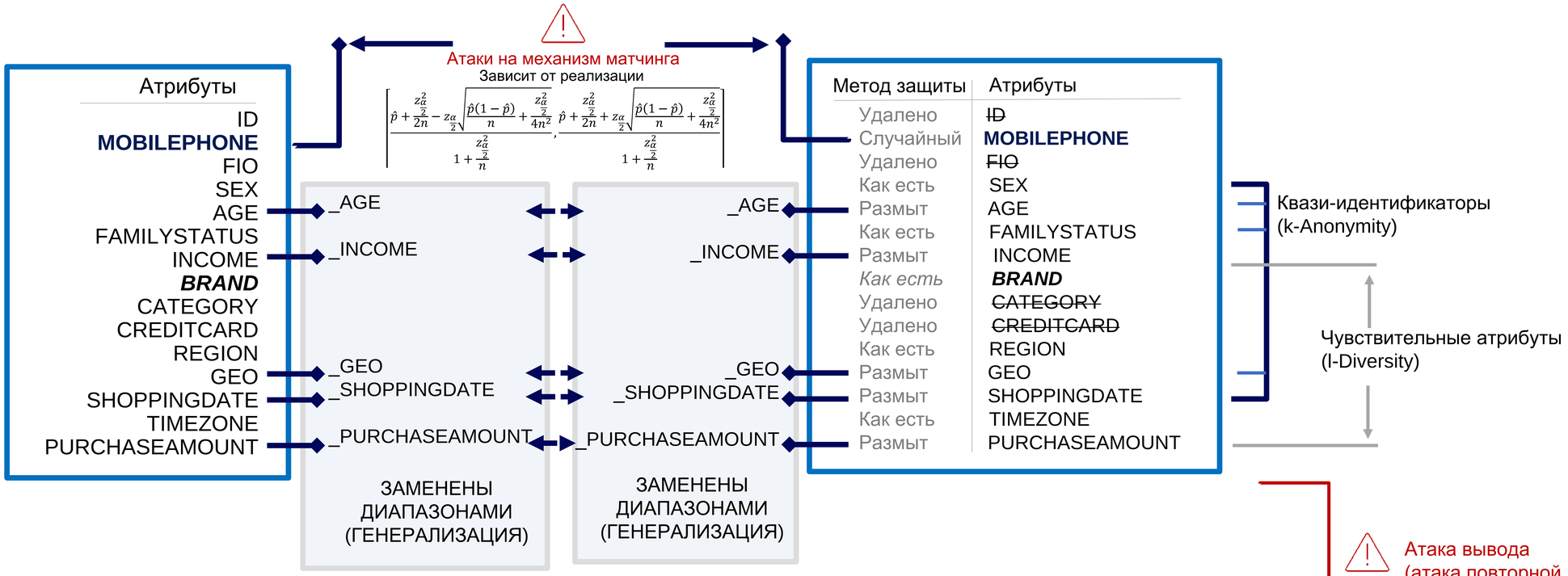
*Спасибо за внимание*

АССОЦИАЦИЯ БОЛЬШИХ ДАННЫХ

# Приложения

# Пояснения к модели риска

Потенциальные атаки на данные могут быть предприняты в связи с тем, что качество выходной информации и безопасность находятся в против-равновесии, поэтому методы защиты, такие как размывание атрибутов или их обобщение ограничены требованиями качества. Оценка метрик риска (уровня защищенности) требует анализа атак.




$$R \sim cos\theta = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$$

Linkage Attack

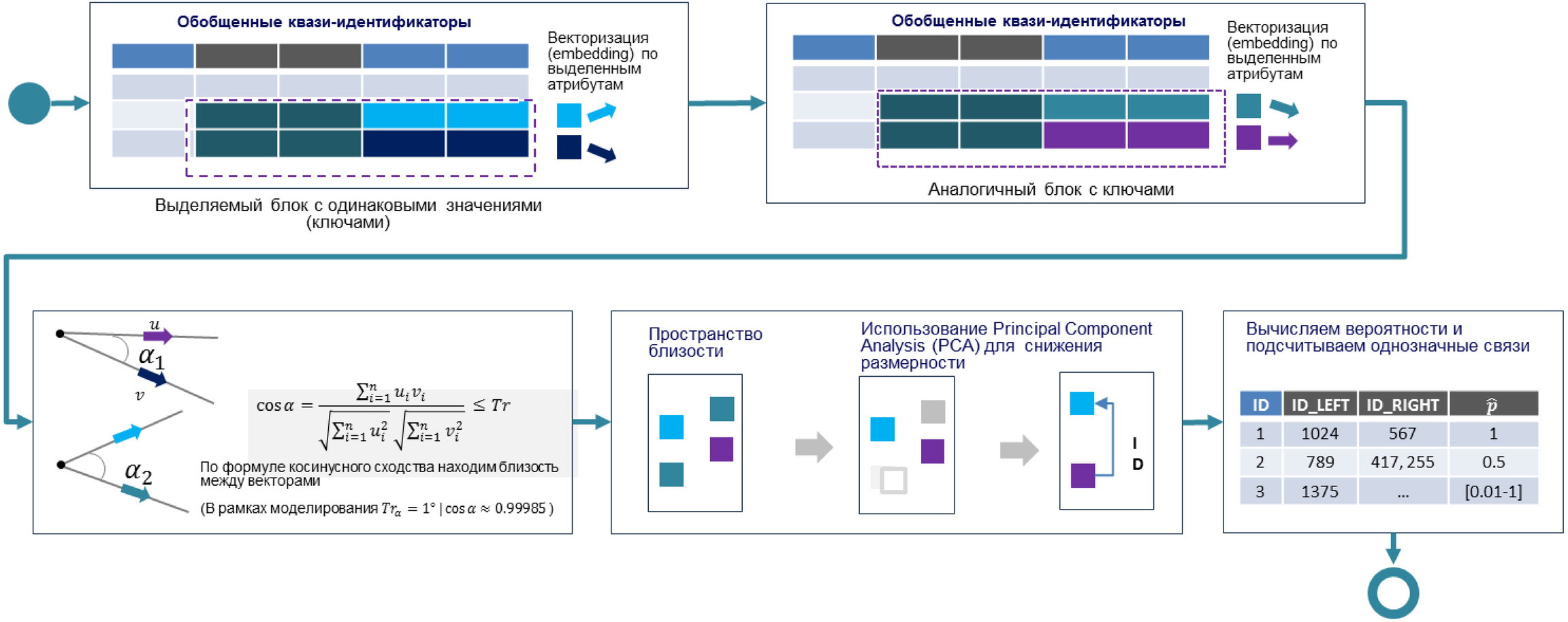
Построение соответствия по методу:

- **Вероятностная атака по методу Fellegi-Sunter**
- **Упрощенная FS-схема**
- **Косинусная мера сходства**

$$R \sim P_{data} = \frac{1}{<f>} = e^{\epsilon}$$

# Симуляция атак связывания

Атаки связывания проводятся с использованием алгоритма CVPL, основанного на кластеризации данных и затем внутри кластеров сравнении векторов данных. Фактически проводится серия различных атак с разными квази-идентификаторами, образующими кластер данных.




| ID | ID_LEFT | ID_RIGHT | $\hat{p}$ |
|---|---|---|---|
| 1 | 1024 | 567 | 1 |
| 2 | 789 | 417, 255 | 0.5 |
| 3 | 1375 | ... | [0.01-1] |

# Дифференциальная приватность

**Дифференциальная модель приватности — это подход к защите конфиденциальности отдельных данных в наборе данных, позволяющий проводить анализ данных без раскрытия конкретной информации об отдельных лицах.**
**Краткое описание основных аспектов дифференциальной модели конфиденциальности:**

- **Дифференциальная приватность —** это формальное определение конфиденциальности, которое гарантирует, что добавление или удаление одного элемента из набора данных не окажет существенного влияния на результаты анализа данных.
- **Конфиденциальность:** Методы дифференциальной приватности обеспечивают механизм, посредством которого выводы, сделанные на основе данных, не раскрывают конфиденциальную информацию о лицах, что делает результаты анализа практически неразличимыми, независимо от наличия или отсутствия конкретной записи в данных.
- **Шумовые механизмы.** Механизмы добавления шума часто используются для достижения дифференциальной конфиденциальности. Это могут быть различные методы, в том числе добавление лапласовского или гауссовского шума к результатам запросов данных.
- **Бюджет конфиденциальности.** Дифференциальная модель конфиденциальности использует концепцию «бюджета конфиденциальности», обычно называемого ε (эпсилон). Низкое значение ε соответствует более высокому уровню конфиденциальности, но может снизить точность результатов анализа.

Когда дело доходит до генерации синтетических данных с помощью ИИ, дифференциальная конфиденциальность является одним из наиболее эффективных подходов к обеспечению конфиденциальности данных.

## ОБЫЧНАЯ СХЕМА ПРИМЕНЕНИЯ ДИФФЕРЕНЦИАЛЬНОЙ ПРИВАТНОСТИ

### Величина риска может быть рассчитана в этой модели:

$$R_{DP} = e^{\epsilon}$$

Здесь ε является дифференциальным параметром конфиденциальности. Формула предполагает, что синтетические данные генерируются дифференциально замкнутым алгоритмом, который гарантирует, что выходные данные не претерпят существенных изменений при изменении или удалении какой-либо конкретной записи в исходных данных. Формула также предполагает, что злоумышленник обладает неограниченными базовыми знаниями и вспомогательной информацией, и что злоумышленник может выполнить любой тип атаки на конфиденциальность синтетических данных.

# Интеграция по схеме Шнорра

**Существуют различные схемы объединения распределенной информации. В контексте тестирования используется концепция ЧЕРНОГО ЯЩИКА, в рамках которой происходит объединение информации без раскрытия сторонами информации о мобильных телефонах пользователей. Ниже предлагается концептуальная схема, которая напрямую не тестируется, но предполагает реализацию на стороне МАСКИРОВЩИКА. Такая схема основан ан механизмах доказательства членства в группе ( Membersip ZKP, Zero Knowledge Proof), использует алгоритм ШНОРРА (Schnorr Scheme).**

## РЕКЛАМНАЯ ПЛОЩАДКА

## ПОСРЕДНИК

## БАНК

Инициализация обмена (период, регион, рекламируемые товары)

Согласование использования криптографически стойкой хеш-функции $H$ и генерация общих параметров для протокола Шнорра: большое простое число $p$, его порождающий элемент $g$, который является примитивным корнем по модулю $p$.

**ОБРАБОТКА НОМЕРА:**

- **Рекламная площадка** обрабатывает номер телефона $t$ для получения уникального идентификатора $ID_t$, используя хеш-функцию: $ID_t = H(t)$.
- Затем выбирается секретный ключ $x$, который является случайным числом. Вычисляется публичный ключ $y = g^x \ mod\ p$.

**СОЗДАНИЕ ДОКАЗАТЕЛЬСТВА:**

- Для каждого телефона $t$ рекламная площадка генерирует случайное число $k$ и вычисляет $r = g^k \ \ mod\ p$.
- Вычисляется хеш $e = H(ID_t \parallel r)$, где $ID_t$ - обработанный номер телефона, а $r$ - результат предыдущего шага.
- Вычисляется $s = k - ex\ mod(p-1)$.

Согласование использования криптографически стойкой хеш-функции $H$ и генерация общих параметров для протокола Шнорра: большое простое число $p$, его порождающий элемент $g$, который является примитивным корнем по модулю $p$.

Рекламная площадка отправляет посреднику $ID_t$, $s$, и $e$ в качестве доказательства.

$\{ID_t, s, \text{и } e\}$

Черный ящик в контексте протокола Шнорра

Посредник сертифицирует доказательства, добавляя свой ключ, а также формирует хранилище для сертифицированных телефонов

Посредник направляет данные по всему согласованному набору в банк

$\{ID_t, s, \text{и } e\}$

Черный ящик в контексте протокола Шнорра

Посредник или банк вычисляет $r' = g^s\ y^e\ mod\ p$ и проверяет, соответствует ли $H(ID_t \parallel r')$ отправленному значению *e (по всем e из набора)*.

Для множества всех доказанных телефонов формируется набор данных для дальнейшей обработки

**ОБЕЗЛИЧИВАНИЕ:**

**Рекламная площадка** подготавливает обезличенный набор, формируя набор с идентификаторами $ID_t$ .

АНОНИМИЗИРОВАННЫЙ НАБОР С ИДЕНТИФИКАТОРОМ направляется посреднику

Повторная обработка и анонимизация

АНОНИМИЗИРОВАННЫЙ НАБОР С ИДЕНТИФИКАТОРОМ направляется посреднику