РусКрипто

XXVII НАУЧНО-ПРАКТИЧЕСКАЯ КОНФЕРЕНЦИЯ

НАЦИОНАЛЬНЫЙ ТЕХНОЛОГИЧЕСКИЙ ЦЕНТР ЦИФРОВОЙ КРИПТОГРАФИИ

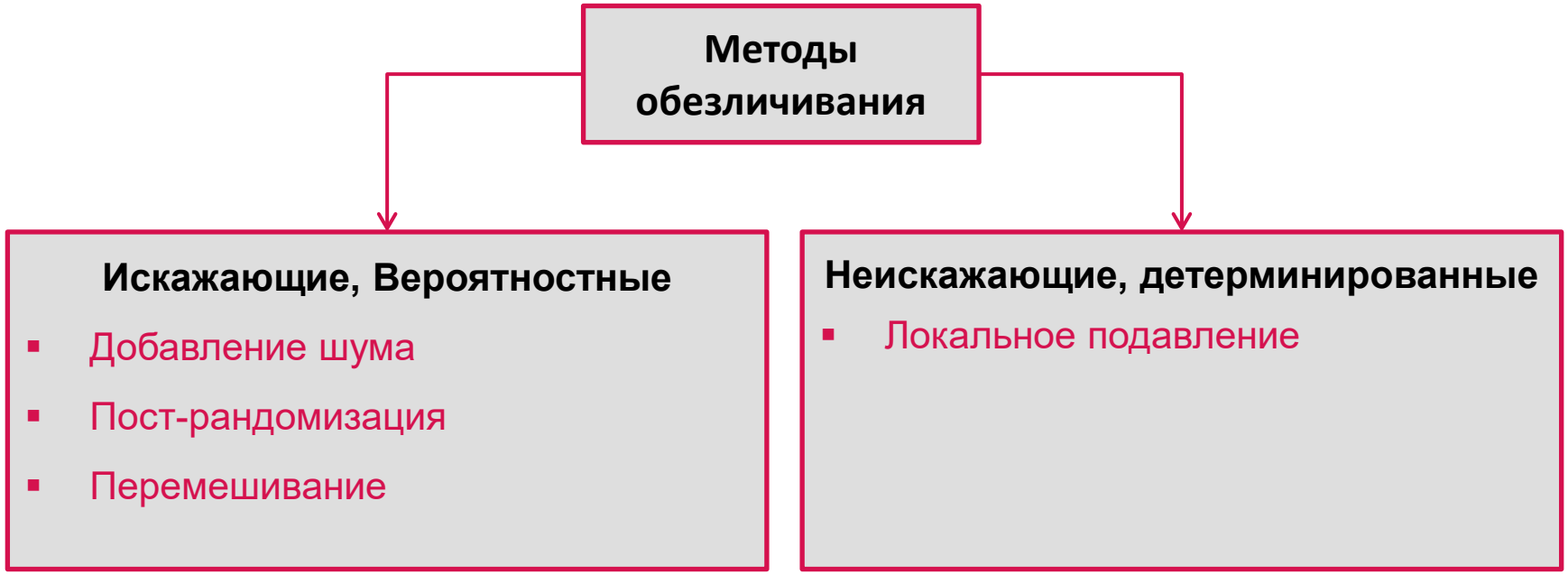
Спутник

Об условиях реализации механизмов пост-рандомизации, локального подавления и зашумления с криптографическими функциями для методов обезличивания персональных данных

Буланов Андрей Владимирович,
Генеральный директор ООО «Поисковый портал «Спутник»

ТИПОЛОГИЯ МЕТОДОВ ОБЕЗЛИЧИВАНИЯ

РусКрипто



ДОБАВЛЕНИЕ ШУМА

РусКрипто

Защита количественных (непрерывных) переменных путем прибавления случайного или выборочного из случайного распределения числа к исходному значению признака.

Генератор шума:

- Собственные возможности библиотеки
- СКЗИ отечественного производства

Входные параметры:

- Исходные данные в табличном виде (*двумерный массив, матрица, таблица*)
- Множество непрерывных количественных квазиидентификаторов (*Множество индексов или названий столбцов таблицы данных*)
- Количество добавляемого шума (*вещественное число*)
- Метод добавления шума (*additive, correlated*)

ДОБАВЛЕНИЕ ШУМА

РусКрипто

Адрес	Пол	Доход
Москва	муж	53 600
Москва	муж	75 300
Москва	муж	63 800
Сочи	жен	89 400
Сочи	жен	57 200
Сочи	жен	48 100

Адрес	Пол	Доход
Москва	муж	52 243.01
Москва	муж	76 876.95
Москва	муж	64 784.49
Сочи	жен	88 400.33
Сочи	жен	55 988.51
Сочи	жен	46 147.09

К атрибуту «Доход» был добавлен шум с сохранением математического ожидания (среднего, взвешенного по вероятностям возможных значений, значения случайной величины)

ПОСТ-РАНДОМИЗАЦИЯ

РусКрипто

Применяется для категориальных переменных. Метод изменяет значения переменных в соответствии с заданной матрицей вероятностей замены.

Генератор шума:

- Собственные возможности библиотеки
- СКЗИ отечественного производства

Входные параметры:

- Исходные данные в табличном виде (*двумерный массив, матрица, таблица*)
- Множество непрерывных количественных квазиидентификаторов (*Множество индексов или названий столбцов таблицы данных*)
- Матрица вероятностей переходов (*Двумерный массив, содержащий вещественные числа из диапазона [0,1] со следующими свойствами:*
  1. Размерности массива равны
  2. Строки и столбцы должны советовать переданному множеству квазиидентификаторов
  3. Сумма элементов в строках должна быть 1)

ПОСТ-РАНДОМИЗАЦИЯ

РусКрипто

Исходные данные

Адрес	Пол	Доход
Москва	муж	28
Москва	муж	28
Надым	муж	31
Сочи	жен	36
Сочи	жен	36
Торжок	муж	31

Матрица вероятностей атрибута Адрес

	Москва	Надым	Сочи	Торжок
Москва	1	0	0	0
Надым	0.5	0	0.5	0
Сочи	0	0	1	0
Торжок	0.5	0	0.5	0

Результат

Адрес	Пол	Доход
Москва	муж	28
Москва	муж	28
Сочи	муж	31
Сочи	жен	36
Сочи	жен	36
Москва	муж	31

Основная идея заключается в случайной замене значений категориальных переменных с вероятностью согласно выбранной матрице переходов. Матрица задается оператором. Диагональные элементы матрицы определяют вероятность того, что значение не изменится. Данное преобразование выполняется для каждой строки таблицы случайно и независимо.

ЛОКАЛЬНОЕ ПОДАВЛЕНИЕ

РусКрипто

Локальное подавление заключается в замене значения переменной на неизвестное значение. Неискажающий, детерминированный метод. Самые важные атрибуты должны быть подавлены в последнюю очередь

Входные параметры:

- Исходные данные в табличном виде (*двумерный массив, матрица, таблица*)
- Множество квазиидентификаторов (*множество индексов или названий столбцов таблицы данных*)
- Требуемый уровень k-анонимности (*целое положительное число*)
- Вклад в частоту комбинации значений квазиидентификаторов строк, имеющих пропущенные значения для одного или нескольких квазиидентификаторов (*вещественное число из полуинтервала (0;1]*)
- Приоритеты квазиидентификаторов (*массив или список целочисленных значений от 1 до |QI|*)

ЛОКАЛЬНОЕ ПОДАВЛЕНИЕ

РусКрипто

Адрес	Пол	Возраст
Москва	муж	28
Москва	муж	28
Омск	муж	31
Сочи	жен	36
Сочи	жен	36
Орел	муж	31

Адрес	Пол	Возраст
Москва	муж	28
Москва	муж	28
	муж	31
Сочи	жен	36
Сочи	жен	36
	муж	31

Локальное подавление последовательно, в порядке убывания индивидуального риска раскрытия, достигает k-анонимности для «проблемных» строк, частота комбинаций значений квазиидентификаторов которых не превышает установленного порога k

ПЕРЕМЕШИВАНИЕ

РусКрипто

Искажающий метод заключающийся в перестановке отдельных значений или групп значений атрибутов персональных данных в наборе

Входные параметры:

- Исходные данные в табличном виде, для каждого столбца задано, является ли переменная чувствительной или квазиидентификатором (*двумерный массив, матрица, таблица*)
- Способ вычисления корреляционной матрицы: spearman (*Строковое (словарное) значение*)

ПЕРЕМЕШИВАНИЕ

Адрес	Пол	Возраст
Москва	муж	28
Тверь	муж	21
Орёл	муж	25
Сочи	жен	36
Анапа	жен	32
Краснодар	жен	39

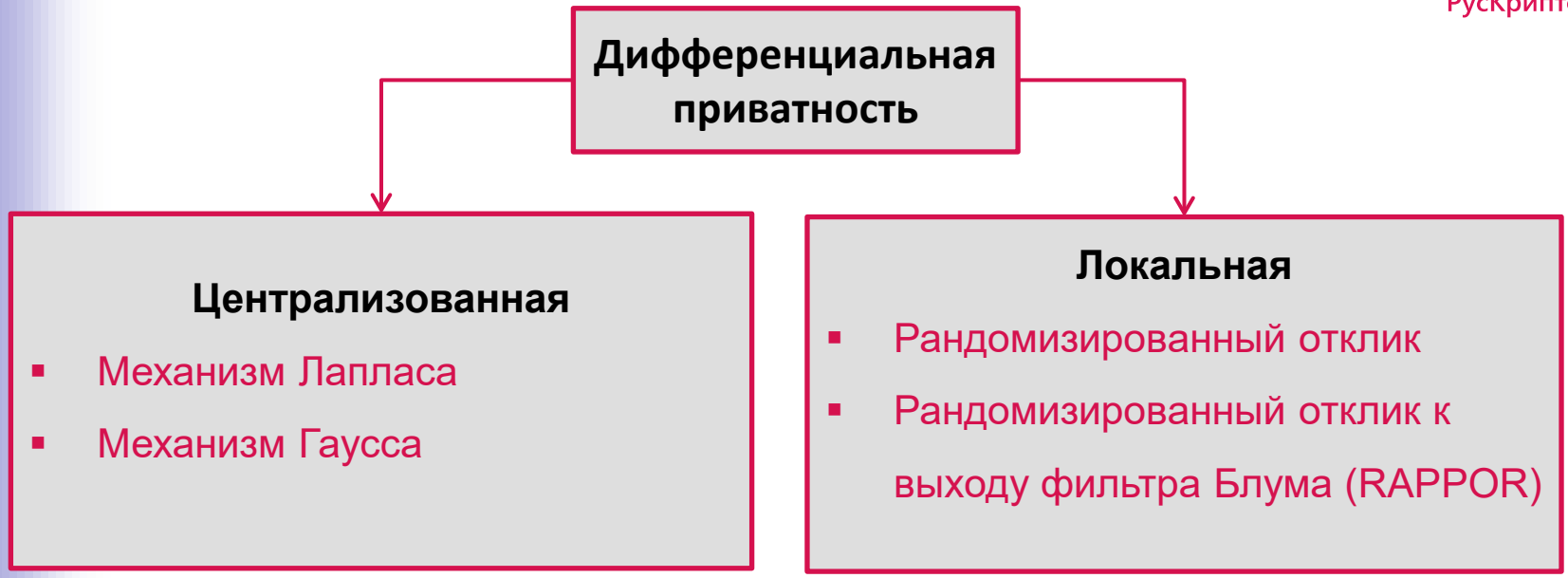
Адрес	Пол	Возраст
Москва	муж	25
Тверь	муж	28
Орёл	муж	21
Сочи	жен	39
Анапа	жен	36
Краснодар	жен	32

Перемешивание использует базовую регрессионную модель для переменных, чтобы определить, какие переменные будут поменяны местами.
Перемешивание может использоваться для непрерывных переменных.

РусКрипто

ДИФФЕРЕНЦИАЛЬНАЯ ПРИВАТНОСТЬ

РусКрипто



Доверенный центр обработки

Не доверенный центр обработки

ДИФФЕРЕНЦИАЛЬНАЯ ПРИВАТНОСТЬ

РусКрипто

Дифференциальная приватность основана на добавлении случайного шума к результатам запросов.

- Обеспечивает максимально точные запросы в статистическую базу данных при одновременной минимизации возможности идентификации отдельных записей в ней;
- Защищает не данные, а алгоритм их обработки;
- Требует контроля за количеством запросов;
- Более удобна для работы с большими данными;
- Ухудшает точность аналитики на заданную величину.

СРЕДСТВА КРИПТОГРАФИЧЕСКОЙ ЗАЩИТЫ ИНФОРМАЦИИ

РусКрипто

Возможности СКЗИ используются при работе алгоритмов обезличивания для генерации случайных и псевдослучайных последовательностей чисел.

Например:

Добавление шума – для генерации случайного числа, которое будет добавлено к значению в качестве шума;

Пост-рандомизация – для создания случайных последовательностей при работе с матрицами пост-рандомизации;

Перемешивание - для генерации *n* синтетических (прогнозируемых) значений для каждой переменной, которая должна быть защищена.

СРЕДСТВА КРИПТОГРАФИЧЕСКОЙ ЗАЩИТЫ ИНФОРМАЦИИ

РусКрипто

Для генерация случайных и псевдослучайных последовательностей чисел использовались:

- ViPNet CSP 4 for Linux SC-216-CSP 4 LIN
- USB ДСЧ, в корпусе с USB-разъемом тип А
- USB-токен ESMART Token ГОСТ
- USB-токен JaCarta-2 ГОСТ
- Рутокен ЭЦП 3.0 3220
- КриптоПро CSP версии 5.0
- ПАК Соболь. Версия 3.2, PCI-E. Исп. 1

РусКрипто

СПАСИБО
ЗА ВНИМАНИЕ