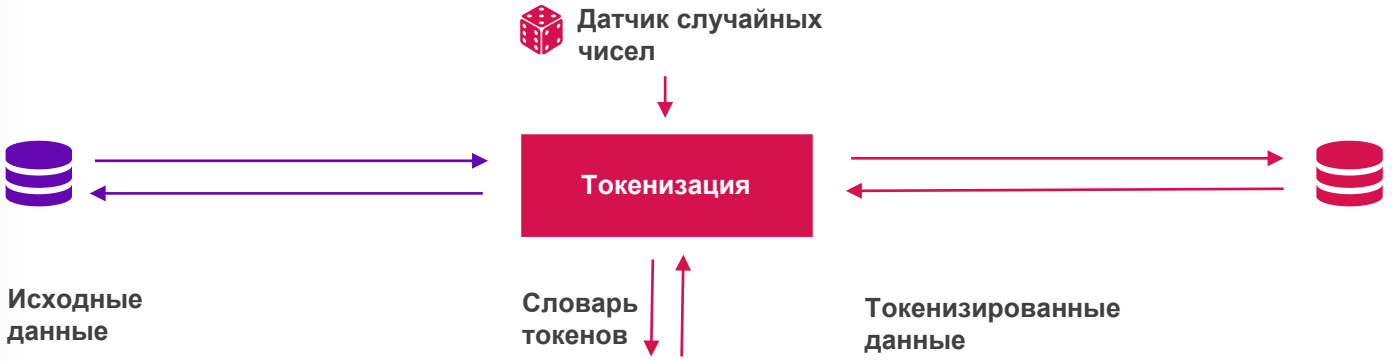
РусКрипто

XXVII НАУЧНО-ПРАКТИЧЕСКАЯ КОНФЕРЕНЦИЯ

Токенизация данных

РусКрипто

токенизация (tokenization): алгоритм преобразования, при котором исходные элементы данных заменяются замещающими значениями – «токенами», которые никак не связаны с исходными данными и не несут никакого смыслового и фактического содержания.



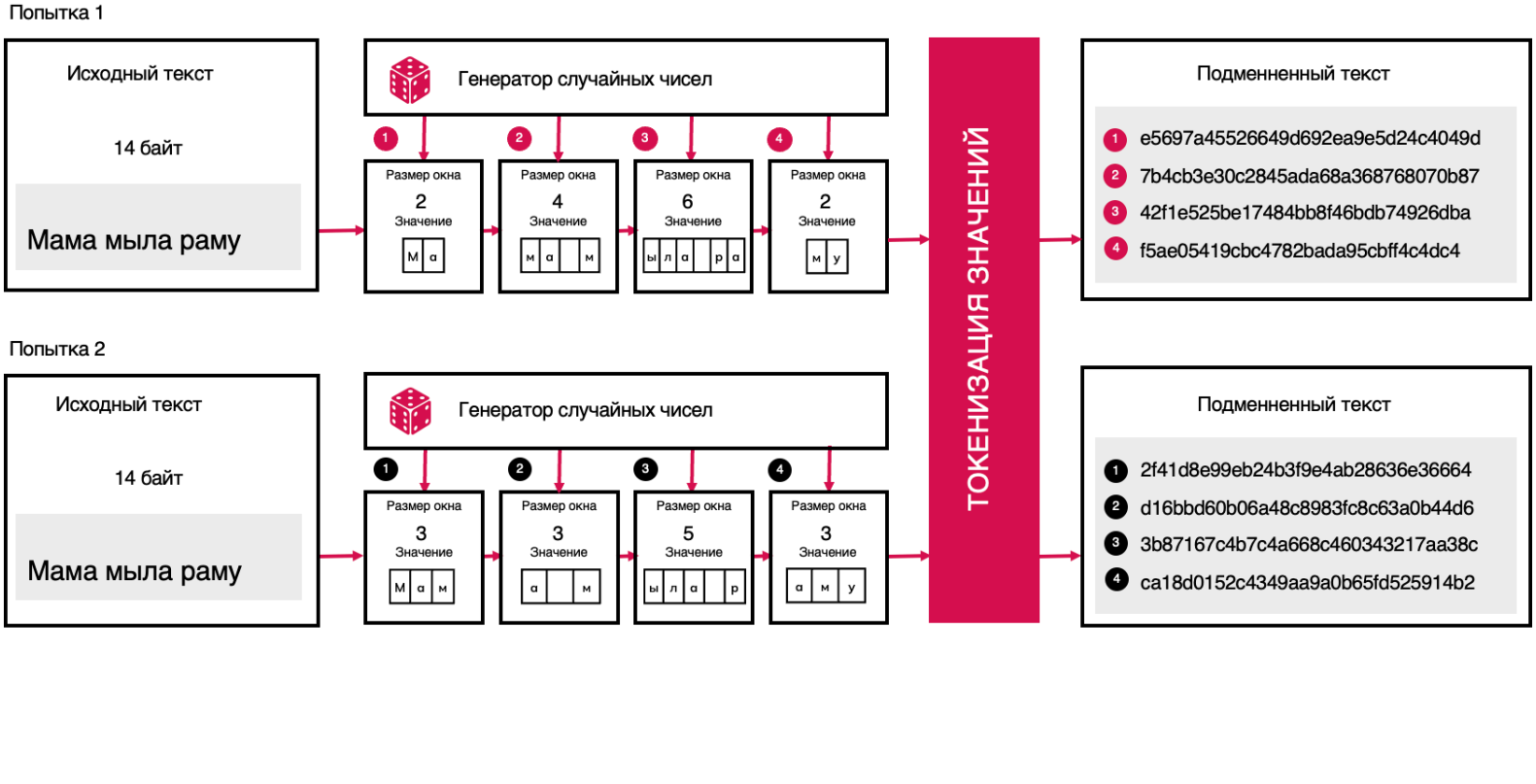
Фамилия	Имя	Отчество
Иванов	Иван	Иванович
Иванов	Иван	Сергеевич
Иванов	Сергей	Сергеевич
Иванов	Сергей	Иванович
Петров	Иван	Иванович
Петров	Иван	Сергеевич
Петров	Сергей	Сергеевич
Петров	Сергей	Иванович

Значение	Токен
Иванов	b6eeb8d45379
Петров	859c59b9677c
Иван	f6113e10c9c9
Сергей	00e3ea79f918
Иванович	ea1f881c7de8
Сергеевич	d72739b1fe7a

Фамилия	Имя	Отчество
b6eeb8d45379	f6113e10c9c9	ea1f881c7de8
b6eeb8d45379	f6113e10c9c9	d72739b1fe7a
b6eeb8d45379	00e3ea79f918	d72739b1fe7a
b6eeb8d45379	00e3ea79f918	ea1f881c7de8
859c59b9677c	f6113e10c9c9	ea1f881c7de8
859c59b9677c	f6113e10c9c9	d72739b1fe7a
859c59b9677c	00e3ea79f918	d72739b1fe7a
859c59b9677c	00e3ea79f918	ea1f881c7de8

Токенизация с переменным окном

РусКрипто



Свойства системы токенизации

Уникальность: каждый токен уникален в пределах одного словаря внутри защищенного хранилища, что гарантирует отсутствие коллизий.
Это исключает ситуацию, когда два разных исходных значения будут преобразованы в одинаковый токен.

Множественность: одному исходному значению может соответствовать несколько токенов.
Это обеспечивает:

- устойчивость к статистическим методам анализа токенизированных данных;
- возможность интеграции токенизированных данных путем объединения словарей из разных систем токенизации.

Независимость: токен не имеет статистически значимой связи с исходным значением. Его структура, длина и состав полностью независимы от характеристик исходных данных. Таким образом токен равновероятно может быть связан как с исходным значением, так и с любым другим значением из множества возможных. Данное свойство обеспечивает невозможность восстановления исходного значения на основе анализа токена.

Случайность: токены генерируются случайным образом с использованием физических датчиков случайных чисел (ФДСЧ) или криптографически стойких программных датчиков случайных чисел (ПДСЧ) с высокой степенью энтропии. Благодаря этому знание любого множества токенов не даёт возможности определить значения других токенов.

Обратимость: токен может быть восстановлен к исходному значению только при наличии доступа к словарю в защищенном хранилище.

Гибкость: формат токена может быть настроен в зависимости от требований и класса конфиденциальной информации. Это позволяет адаптировать алгоритм к конкретным бизнес-процессам (например, изменяя длину токена, набор используемых символов или структуру) для соответствия стандартам и техническим ограничениям защищаемых информационных систем.

Информативность: токен может содержать дополнительные метаданные, такие как время создания, идентификатор организации, или другие сведения, необходимые для выполнения бизнес-процессов. Это свойство позволяет использовать токен не только для подмены исходного значения, но и как источник дополнительной информации, упрощая интеграцию и обработку данных в сложных сценариях.

Ограниченность: для токенов могут быть установлены ограничения:

- по количеству операций, при достижении предельного количества одно и то же значение будет представляться еще одним новым токеном
- по сроку действия. Если срок службы токена истек, то он должен быть заменен на новый токен во всех инорсационных системах

РусКрипто

Криптографические аспекты токенизации

РусКрипто

Свойство 1. Отсутствие информационной связанности токена с исходными данными

Токены *T* создаются случайным образом и не зависят от исходных данных *M*.

Взаимная информация между исходным значением и токеном является нулевой:

I(T,M) = 0

Свойство 2. Информационная энтропия токенизированных данных выше, чем у исходных

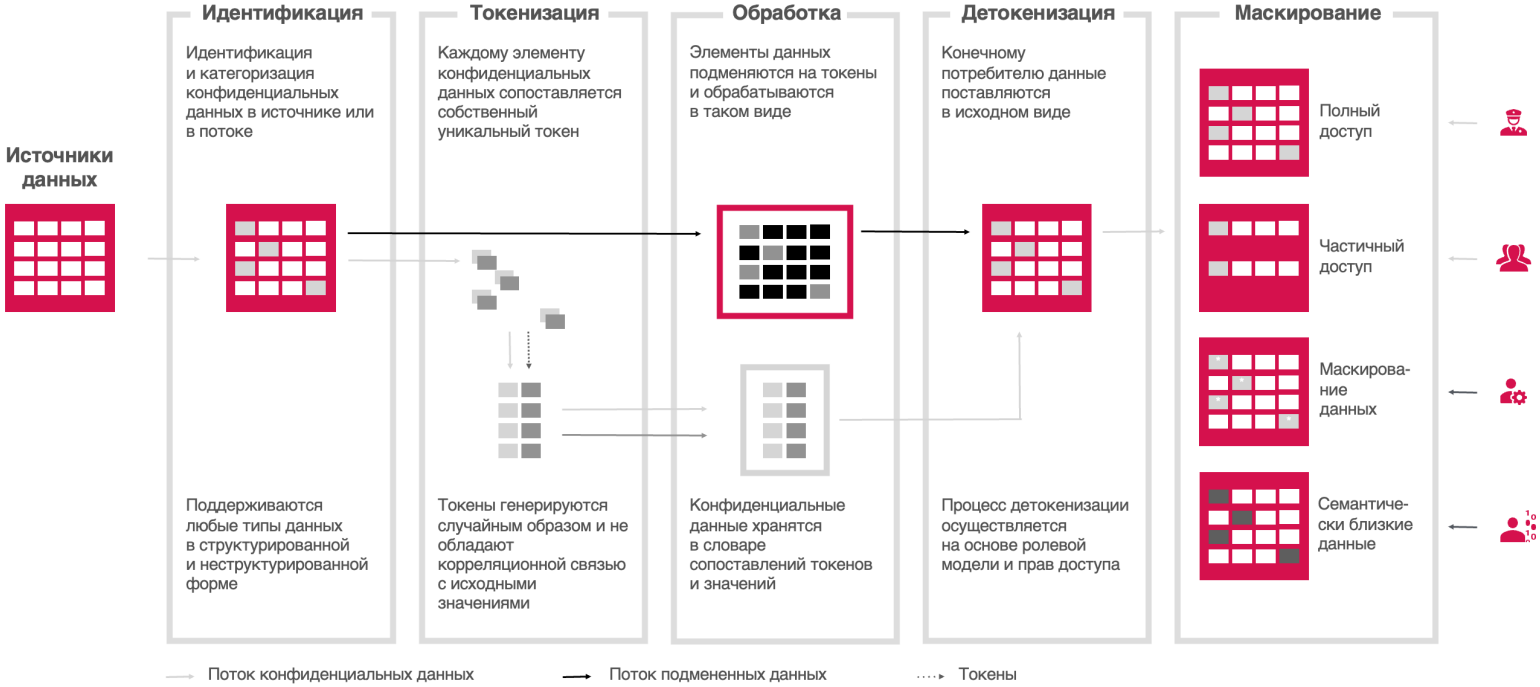
$$H(D,T) = \sum_{i=1}^{n} P(m_i) \log k_i + H(M) + \log d$$

Если система обладает абсолютной криптостойкостью, то конфиденциальная информация, прошедшая процедуру токенизации, может более не считаться таковой, а утечка токенизированных данных не приведет к компрометации исходных данных.

Если энтропия системы токенизации данных выше, чем энтропия исходных данных и взаимная информация между токенами и исходными данными равна нулю, то такая система по Шеннону может претендовать на абсолютную криптостойкость, что делает систему максимально устойчивой к криптографическому анализу и атакам;

Метод динамической подмены данных

РусКрипто



РусКрипто

СПАСИБО
ЗА ВНИМАНИЕ