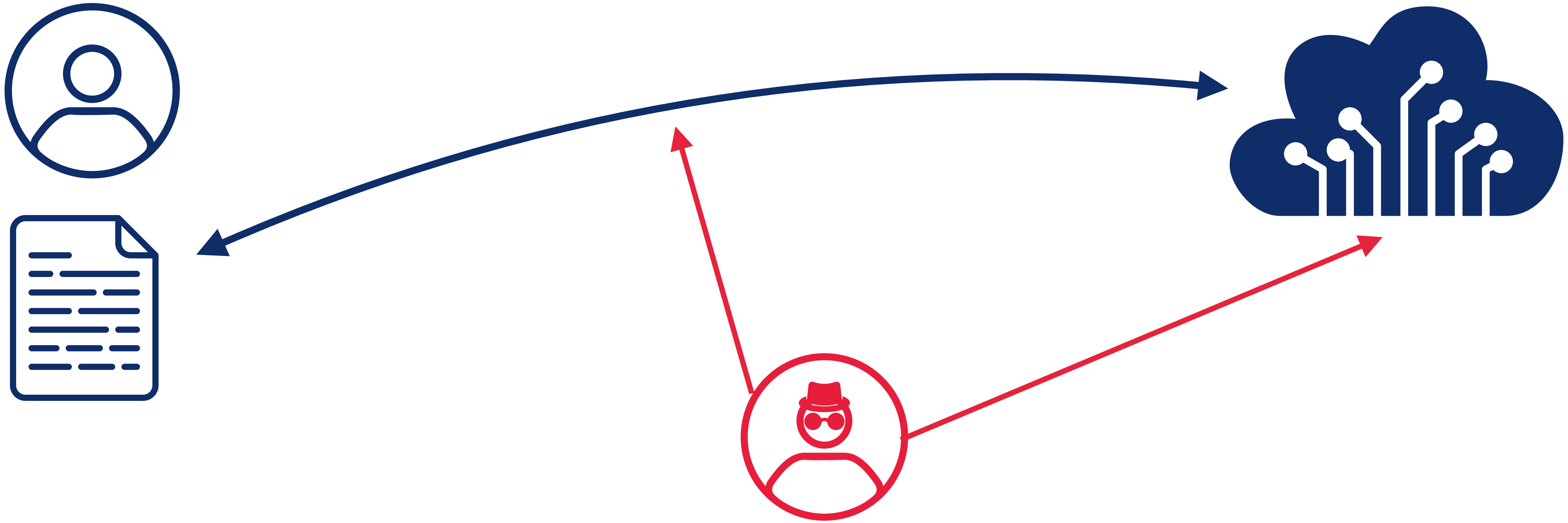
Национальный исследовательский
университет «Высшая школа экономики»

Москва
2022

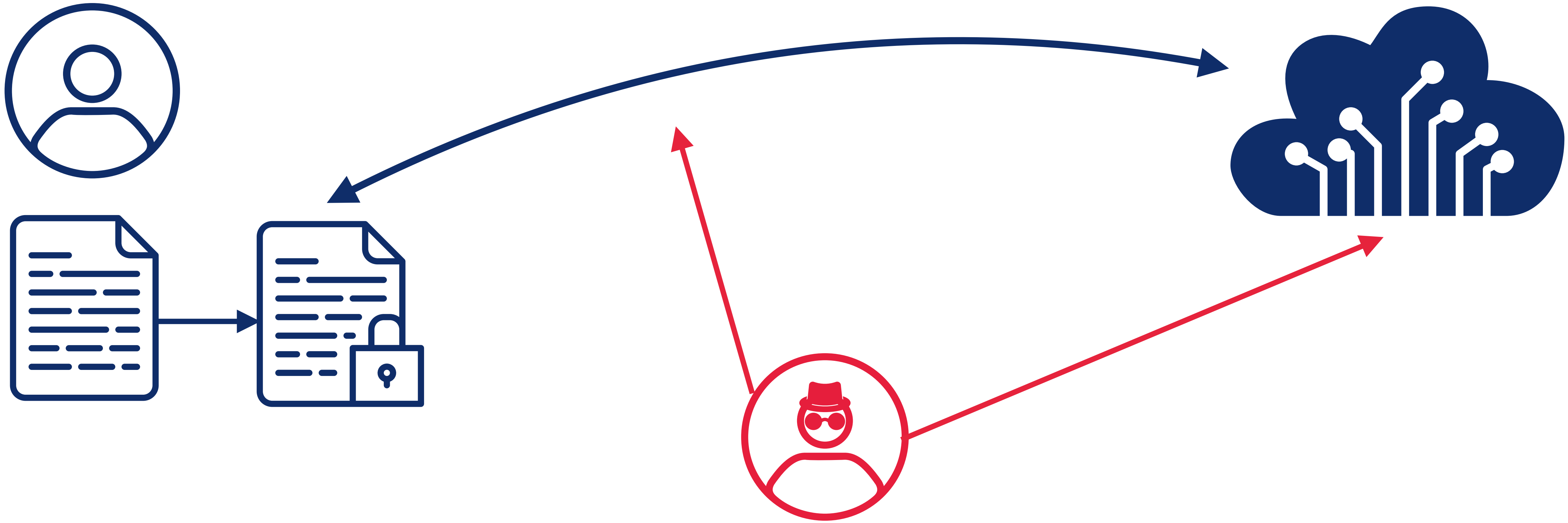
Одна модель обеспечения безопасности конфиденциальной обработки больших данных

Фомин Денис Бониславович

Конфиденциальная обработка больших данных



Конфиденциальная обработка больших данных



Вероятностная модель алгоритма шифрования

$$\Sigma = \left(X, K, Y, E, D, P_X, P_K\right)$$

X	множество открытых текстов
K	множество ключей
Y	множество шифрованных текстов
$E\colon K \times X \to Y$	функция зашифрования
$D\colon K \times Y \to X$	функция расшифрования
P_X	вероятностное распределение на множестве открытых текстов
P_K	вероятностное распределение на множестве ключей

Необходимые обозначения

$\varkappa, \kappa$	независимые случайные величины, принимающие значения на X, K и имеющие распределения P_X, P_K
$\widehat{\xi}$	реализация случайной величины ξ
$\widehat{\kappa}$	ключ
$\widehat{\varkappa}_1, \dots, \widehat{\varkappa}_n$	множество открытых текстов
$\widehat{\gamma}_1, \dots, \widehat{\gamma}_n$	множество шифрованных текстов, где $\gamma_i = E\left(\kappa, \varkappa_i\right)$
$* : X \times X \to X$	бинарная операция, заданная на множестве X

- Алгоритм шифрования называется *частично гомоморфным*, если существует эффективный алгоритм, который для любых $x_1, x_2 \in X$ и для любого $k \in K$, получив на вход только $E\left(k, x_1\right)$ и $E\left(k, x_2\right)$, выдаёт значение $E\left(k, x_1 * x_2\right)$
- Под *ограниченно гомоморфными* алгоритмами шифрования будем подразумевать такие Σ, что для любой функции $f \colon X^m \to X$ из некоторого класса функций $\mathscr{F}$ и любого $k \in K$ существует эффективный алгоритм, который, получив на вход только $\left\{E\left(k, x_i\right), i = 1, \ldots, m\right\}$, выдаёт значение $E\left(k, f\left(x_1, \ldots, x_m\right)\right)$
- *Полностью гомоморфным* алгоритм шифрования называется ограниченно гомоморфный алгоритм в случае, когда $\mathscr{F}$ есть множество всех возможных функций $X^m \to X$ для всех $n \in \mathbb{N}$

Обоснование стойкости гомоморфный алгоритмов шифрования

- Обоснование стойкости сводится к решению некоторых известных задач
- Обоснование стойкости зависти от алгоритма шифрования
- Не все задачи, к которым сводится стойкость, хорошо изучены
- Интересен вопрос построения универсальной модели и построения универсальных методов анализа

Сравнение с «классической» моделью практической стойкости

- У нарушителя имеется:
  - множество шифртекстов $\hat{Y} = \{\hat{\gamma}_i, i = 1,\ldots,m\}, m < n$
  - множество пар открытый-шифрованный текст $\widehat{XY} = \{(\hat{\varkappa}_i, \hat{\gamma}_i), i = m+1,\ldots,n\}$
  - вычислительные возможности не превосходящие T
- Нарушитель не знает ключ $\hat{\kappa}$

Алгоритм шифрования Σ можно считать практически ε-стойким с уровнем стойкости T, если произвольный нарушитель, обладающий вычислительными возможностями не превосходящими T, для некоторого фиксированного $\varepsilon > 0$, может предьявить $i \in \overline{1,m}$, такое, что

$$\left|\Pr\left(\varkappa_i = x \,\middle|\, \hat{Y}, \widehat{XY}\right) - P_X(x)\right| \geq \varepsilon$$

Модель практической стойкости для гомоморфных алгоритмов шифрования

- У нарушителя имеется:
  - множество шифртекстов $\widehat{Y} = \left\{ \widehat{y}_i, i = 1, \ldots, m \right\}, m < n$
  - множество пар открытый-шифрованный текст $\widehat{XY} = \left\{ \left(\widehat{x}_i, \widehat{y}_i \right), i = m + 1, \ldots, n \right\}$
  - вычислительные возможности не превосходящие T
- Нарушитель не знает ключ $\widehat{\kappa}$

При обосновании стойкости гомоморфных алгоритмов шифрования помимо получения информации об открытых текстах, нарушителю интересны также *аргументы от значений функций от множества шифрованных текстов*, которые могут быть вычислены для рассматриваемого алгоритма гомоморфного шифрования

Модель практической стойкости для гомоморфных алгоритмов шифрования

- У нарушителя имеется:
  - множество шифртекстов $\widehat{Y} = \left\{ \widehat{\gamma}_i, i = 1,\ldots,m \right\}, m < n$
  - множество пар открытый-шифрованный текст $\widehat{XY} = \left\{ \left(\widehat{\varkappa}_i, \widehat{\gamma}_i \right), i = m+1,\ldots,n \right\}$
  - вычислительные возможности не превосходящие T
- Нарушитель не знает ключ $\widehat{\kappa}$

Частично гомоморфный алгоритм шифрования Σ можно считать практически ε-стойким с уровнем стойкости T, если произвольный нарушитель, обладающий вычислительными возможностями не превосходящими T, для некоторого фиксированного $\varepsilon > 0$, может предьявить $a = \left(a_1, \ldots, a_n\right) : \exists i \in \overline{1,m}, a_i = 1$, такое, что

$$\left| \Pr\left(\varkappa_1^{a_1} * \ldots * \varkappa_n^{a_n} = x \,\middle|\, \widehat{Y}, \widehat{XY} \right) - P_X(x) \right| \geq \varepsilon$$

Модель практической стойкости для гомоморфных алгоритмов шифрования

- В случае частично гомомрфного алгоритма шифрования число таких наборов ровно $2^n - 2^{n-m+1}$
- В случае ограниченно гомоморфных алгоритмов шифрования количество аргументов функций зависит от выбора $\mathscr{F}$
- Если на множестве X задраны две операции $+, \cdot$, позволяющие определить на X структуру поля и алгоритм Σ гомоморфных по двум этим операциям, то он полностью гомоморфен
- В этом случае интерес представляет $|X|^{|X|^n} - |X|^{|X|^{n-m+1}}$ значений, а практическая стойкость может быть определена аналогично ранее изложенному

Парадокс дней рождения

- Пусть $\alpha = \dfrac{2^n - 2^{n-m+1}}{\sqrt{|Y|}}$, $\beta = \dfrac{2^{n-m+1}}{\sqrt{|Y|}}$
- $\alpha, \beta \in \left(0, \sqrt{|Y|}\right)$
- Тогда среди множества значений $\left\{ \widehat{\varkappa}_1^{a_1} * \ldots * \widehat{\varkappa}_n^{a_n} \colon a_i \in \{0,1\}, i \in \overline{1,n}, \exists i \in \overline{1,m} \colon a_i = 1 \right\}$ и $\left\{ \widehat{\varkappa}_{m+1}^{a_{m+1}} * \ldots * \widehat{\varkappa}_n^{a_n} \colon a_i \in \{0,1\}, i \in \overline{m+1,n} \right\}$ есть пересечение с вероятностью не меньше $1 - \exp\{-\alpha \cdot \beta\}$
- Полученная оценка верна при любом распределении на множестве X

Парадокс Марты

- При неравномерном распределении на X оценка $1 - \exp\left\{-\alpha \cdot \beta\right\}$ оказывается завышенной
- Для ряда алгоритмов зашифрования функция E является случайной функцией
- Значение функции шифрования зависит от случайного параметра $r \in R$, имеющего распределение P_R
- Наличие неравновероятности на множестве R также уменьшает необходимый объём множеств $\widehat{Y}$, $\widehat{XY}$