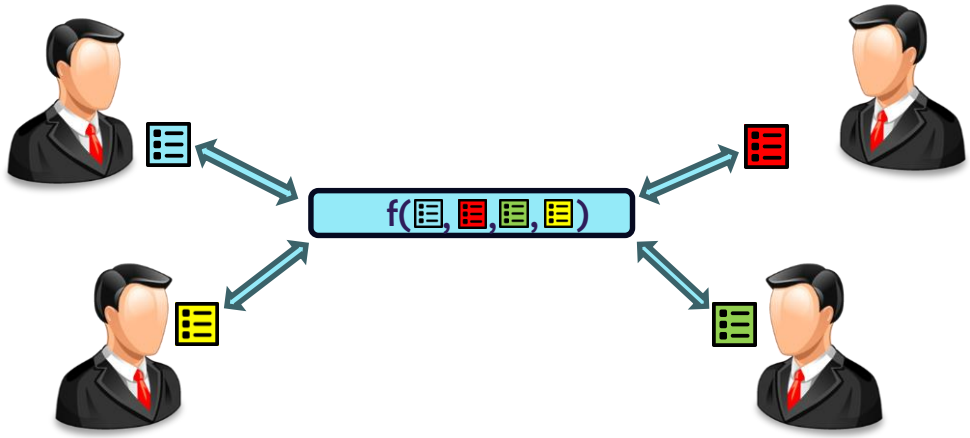
Логический вывод в протоколах многосторонних безопасных вычислений

Столовник Д.А., ТК 26

Содержание

1. Протоколы многосторонних безопасных вычислений

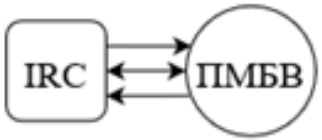
Протоколы многосторонних безопасных вычислений (ПМБВ) позволяют группе участников производить вычисления над некоторой функцией, известной каждому из участников, подавая на ее вход в качестве аргументов свои конфиденциальные данные и не раскрывая никаких данных, кроме результата вычислений, и, возможно, некоторых промежуточных данных, полученных на выходе при прохождении этапов вычисления данной функции.



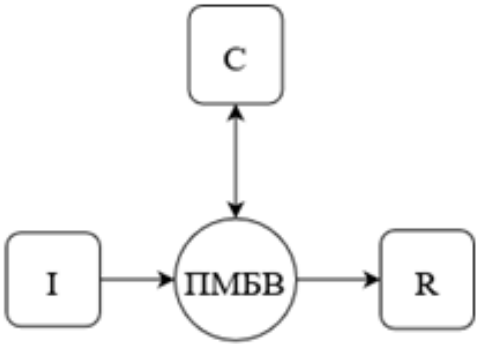
2. Роли участников и архитектура ПМБВ

Модели ПМБВ:

1. Модель делигирования вычислений



2. Модель совместных вычислений

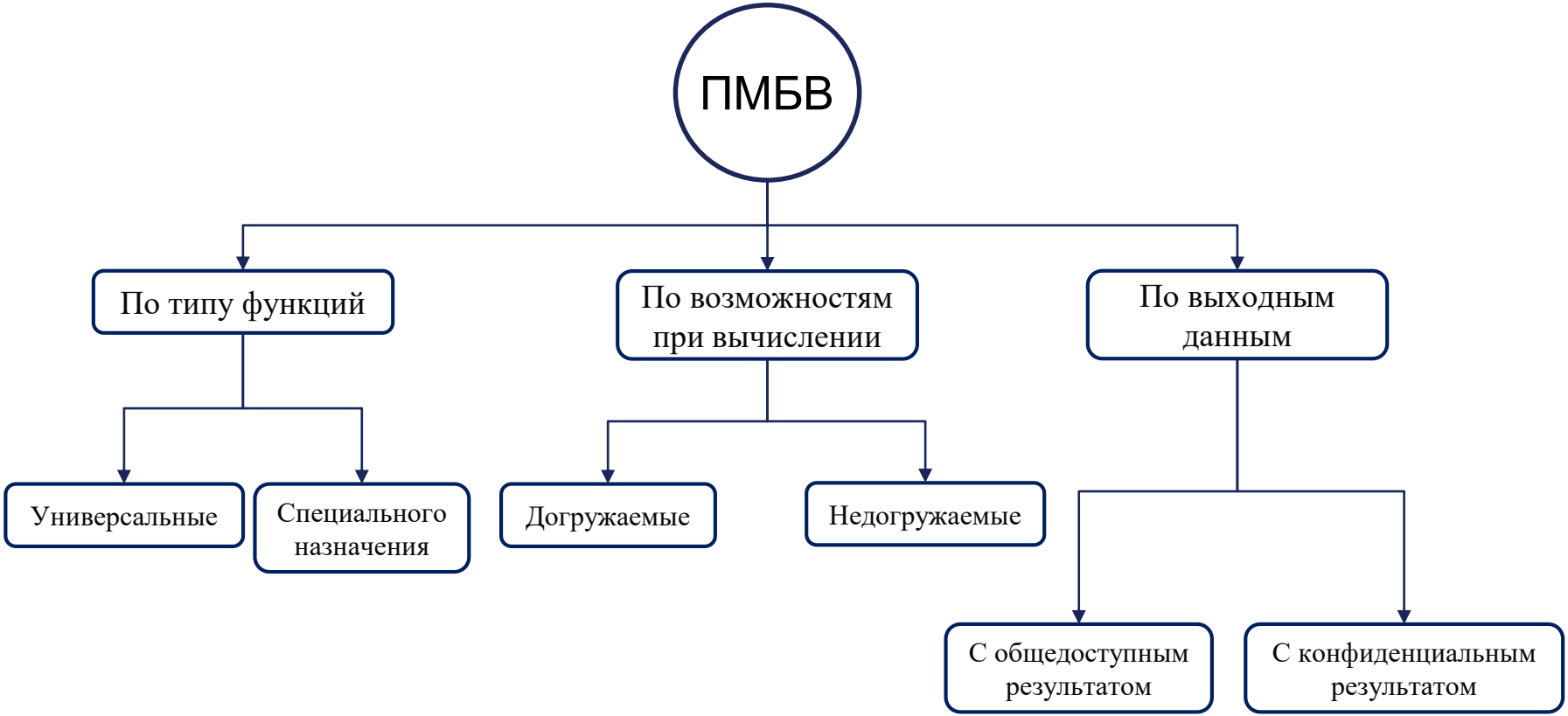


3. Модель делегирования сервисов

Основные роли участников ПМБВ:

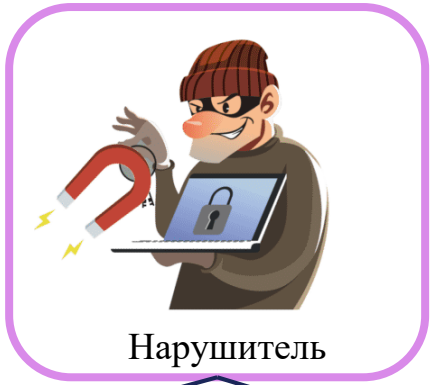
- I - источник входных данных
- C - вычислитель
- R - участник, инициирующий вычисления и получающий результат

3. Классификация ПМБВ

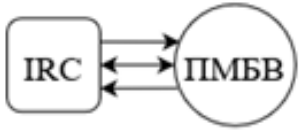


4. Модели нарушителя

В ПМБВ *нарушителем* может быть подмножество участников протокола, которые вступают в сговор и обмениваются данными с целью раскрытия конфиденциальных данных честных участников или нарушения работы протокола.

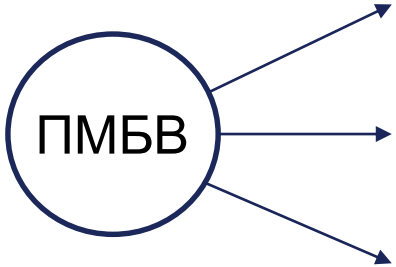


5. О рассматриваемой модели



Для исследований выбрана модель совместных вычислений, однако представленные в дальнейшем методы справедливы и в других моделях

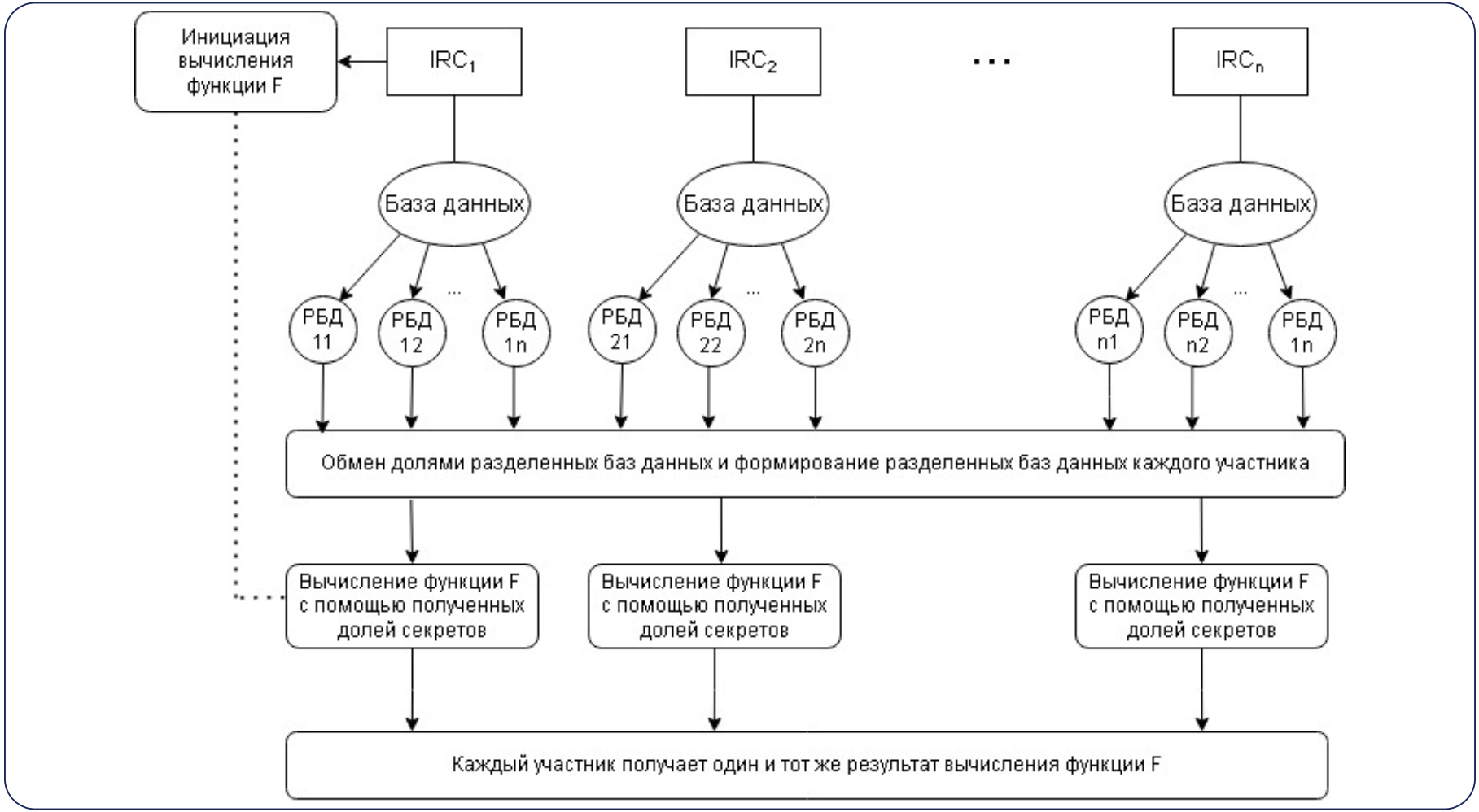
IRC_1 – честный, но любознательный участник



универсальный

недогружаемый

с общедоступным результатом


Инициация вычисления функции F
IRC1
IRC2
...
IRCn
База данных
База данных
База данных
РБД 11
РБД 12
...
РБД 1n
РБД 21
РБД 22
...
РБД 2n
РБД n1
РБД n2
...
РБД 1n
Обмен долями разделенных баз данных и формирование разделенных баз данных каждого участника
Вычисление функции F с помощью полученных долей секретов
Вычисление функции F с помощью полученных долей секретов
Вычисление функции F с помощью полученных долей секретов
Каждый участник получает один и тот же результат вычисления функции F

6.1. Метод построения линейно-независимых запросов

Пусть присутствует от 2 до n участников информационного взаимодействия $A_1, A_2, \dots, A_n$, каждый из которых выполняет роли I, R и C.

A_1	
ID	Значение
A_1id_1	a_1
A_1id_2	a_2
⋮	⋮
A_1id_k	a_k

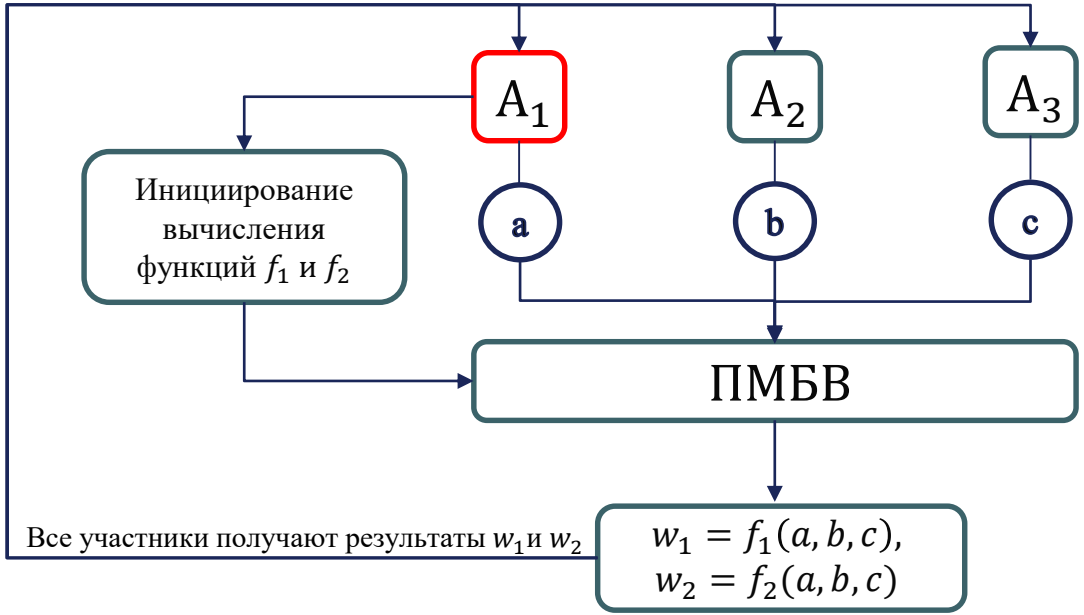
A_2	
ID	Значение
A_2id_1	b_1
A_2id_2	b_2
⋮	⋮
A_2id_l	b_l

…

A_n	
ID	Значение
A_nid_1	c_1
A_nid_2	c_2
⋮	⋮
A_nid_m	c_m

6.1. Метод построения линейно-независимых запросов

Рассмотрим пример. A_1 - честный, но любознательный участник, $a, b, c \in \mathbb{Z}$.



Участник A_1 решает систему и находит b и c:

$$\begin{cases} f_1(a,b,c) = k_{11} \cdot a + k_{12} \cdot b + k_{13} \cdot c \\ f_2(a,b,c) = k_{21} \cdot a + k_{22} \cdot b + k_{23} \cdot c \end{cases}'$$

6.1. Метод построения линейно-независимых запросов

Трудоемкость метода: $(\boldsymbol{N-k})\cdot(\boldsymbol{M+n})$.

N – суммарное количество секретных значений участников;
k – количество секретных значений честного, но любознательного участника;
n – количество переменных в линейной функции (операций умножения);
$\mathrm{M} = n - 1$ – количество операций сложения в линейной функции.

6.2. Метод наиболее вероятных исходов

Пусть A_1, A_2 - участники информационного взаимодействия, A_1 является честным, но любознательным.

A_1		A_2	
X	Y	X	Y
$x_1^{(1)}$	$y_1^{(1)}$	$x_1^{(2)}$	$y_1^{(2)}$
$x_2^{(1)}$	$y_2^{(1)}$	$x_2^{(2)}$	$y_2^{(2)}$
$\dots$	$\dots$	$\dots$	$\dots$
$x_k^{(1)}$	$y_k^{(1)}$	$x_k^{(2)}$	$y_k^{(2)}$

	X	Y
A_1	$x_1^{(1)}$	$y_1^{(1)}$
	$x_2^{(1)}$	$y_2^{(1)}$
	$\dots$	$\dots$
	$x_k^{(1)}$	$y_k^{(1)}$
A_2	$x_1^{(2)}$	$y_1^{(2)}$
	$x_2^{(2)}$	$y_2^{(2)}$
	$\dots$	$\dots$
	$x_k^{(2)}$	$y_k^{(2)}$

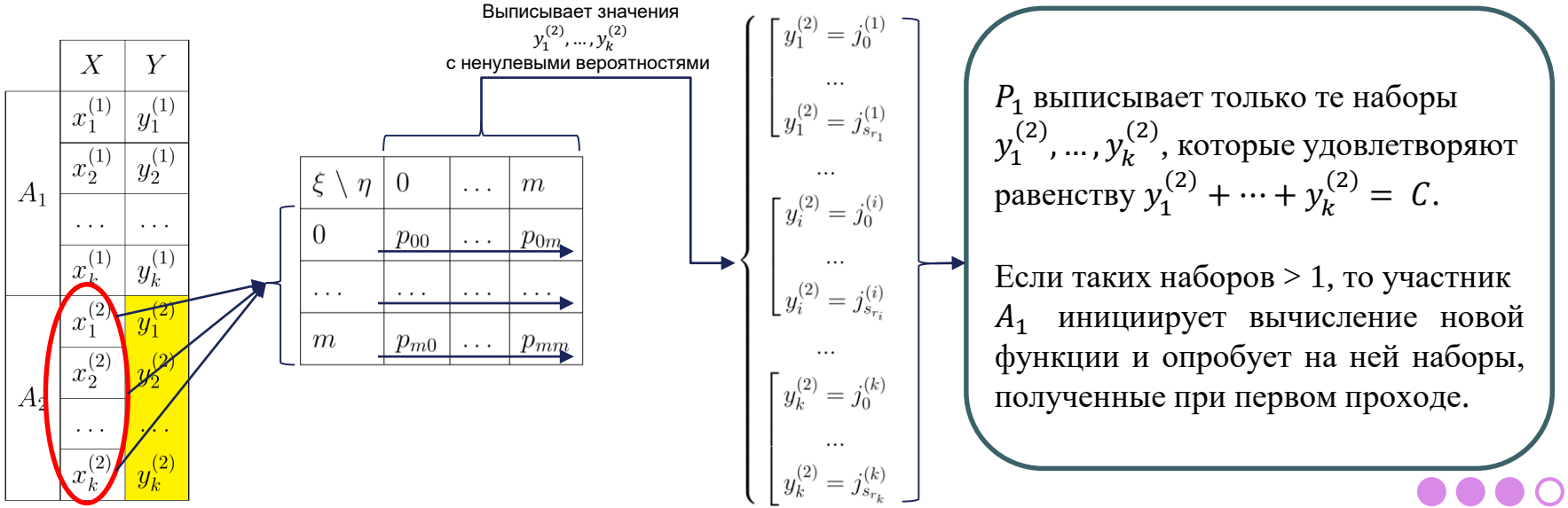
6.2. Метод наиболее вероятных исходов

Рассмотрим дискретные случайные величины ξ, η , определяющие случайные значения, принимаемые значениями выборок, относящимися к признакам X и Y, соответственно. Будем считать, что честный, но любознательный участник A_1 знает совместное распределение (ξ, η):

$\xi \setminus \eta$	0	$\dots$	m
0	p_{00}	$\dots$	p_{0m}
$\dots$	$\dots$	$\dots$	$\dots$
m	p_{m0}	$\dots$	p_{mm}

6.2. Метод наиболее вероятных исходов

Участник A_1 инициирует запрос вычисления функции $f\left(y_1^{(2)}, \dots, y_k^{(2)}\right) = y_1^{(2)} + \dots + y_k^{(2)}$, в результате чего получает некоторое значение $C \in \mathbb{Z}$.



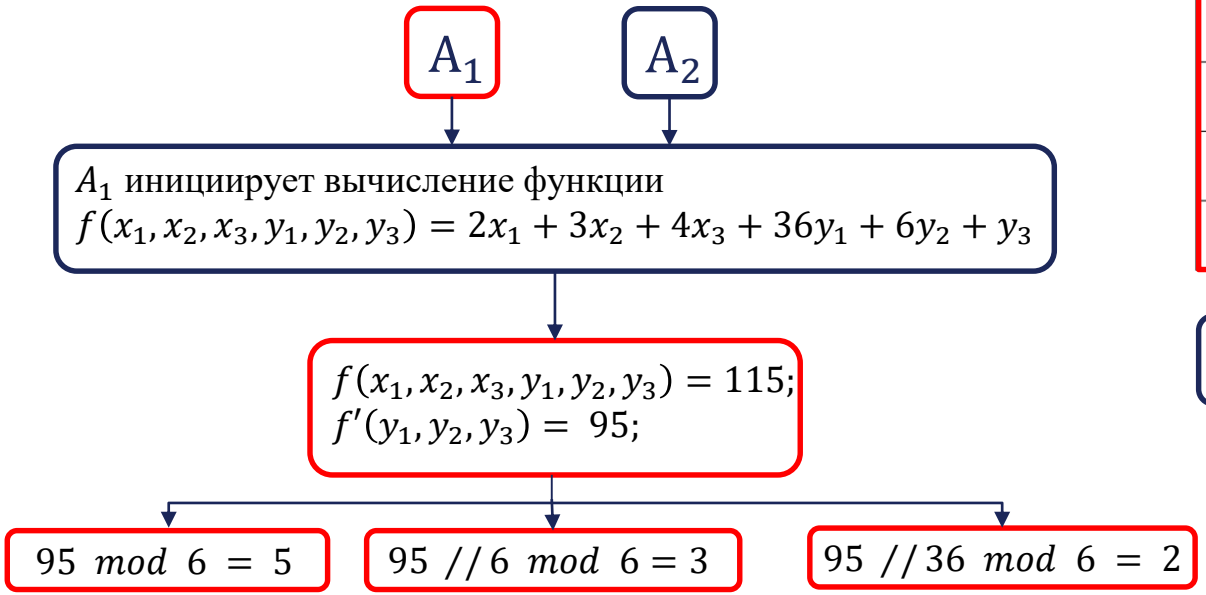
6.2. Метод наиболее вероятных исходов

Трудоемкость метода: $(\boldsymbol{N/2 + L})(\boldsymbol{M + n})$**.**

$N = 2k$ – суммарное количество секретных значений участника;
L – количество необходимых дополнительных запросов вычисления линейных функций;
n – количество переменных в линейной функции (операций умножения);
$\mathrm{M} = n - 1$ – количество операций сложения в линейной функции.

6.3. Метод с использованием функций специального вида

Пусть A_1, A_2 - участники информационного взаимодействия, A_1 является честным, но любознательным. Предположим, участнику A_1 известно наибольшее из значений участника A_2 , а именно $y_3 = 5$.



A_1	A_2
$x_1 = 1$	$y_1 = 2$
$x_2 = 2$	$y_2 = 3$
$x_3 = 3$	$y_3 = 5$

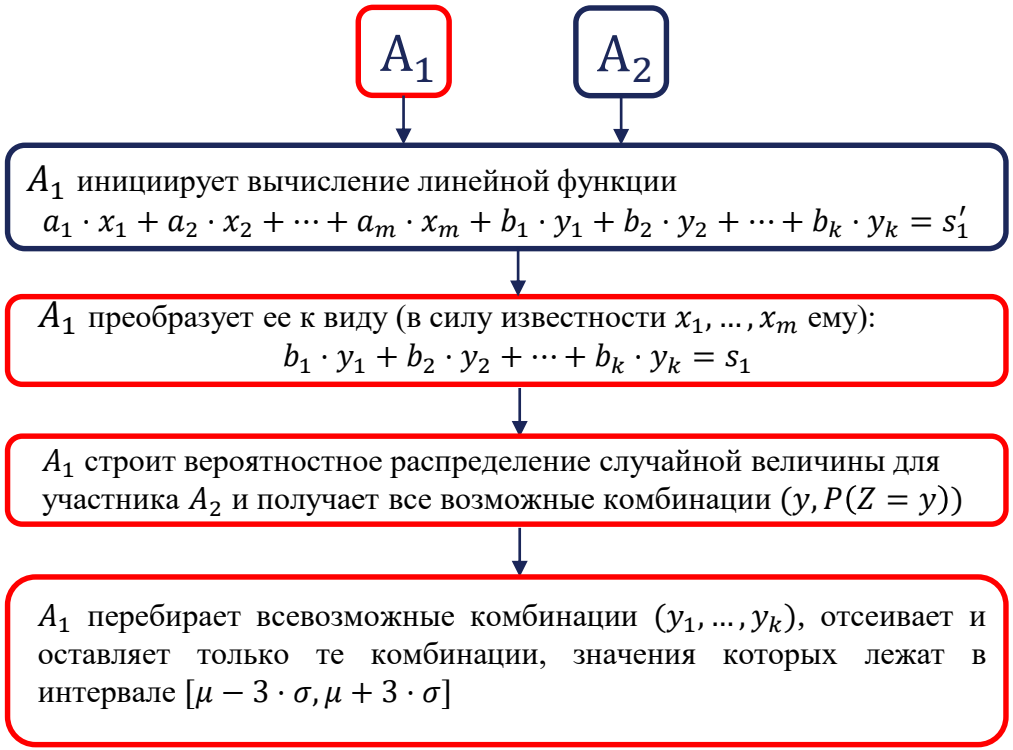
// - операция деления без остатка

6.3. Метод с использованием функций специального вида

Трудоемкость метода: $M + n$.

n – количество переменных в линейной функции (операций умножения);
$M = n - 1$ – количество операций сложения в линейной функции.

6.4. Метод отсеивания наименее вероятных значений по известному распределению



A_1 получил информацию о виде распределения данных в базе участника A_2 (например, что распределение значений признаков A_2 такое же, как и у самого участника A_1). Пусть значения признаков имеют нормальное распределение $N(\mu, \sigma)$. Также известно, что диапазон значений признаков у каждого из участников протокола совпадает.

6.4. Метод отсеивания наименее вероятных значений по известному распределению

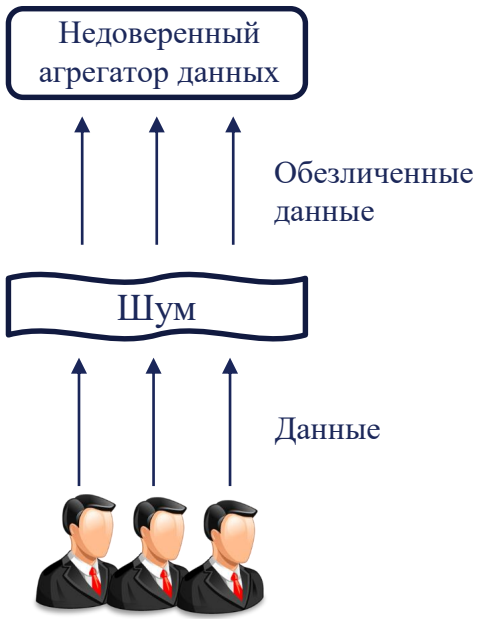
Трудоемкость метода: $\boldsymbol{L(M+n)}$.

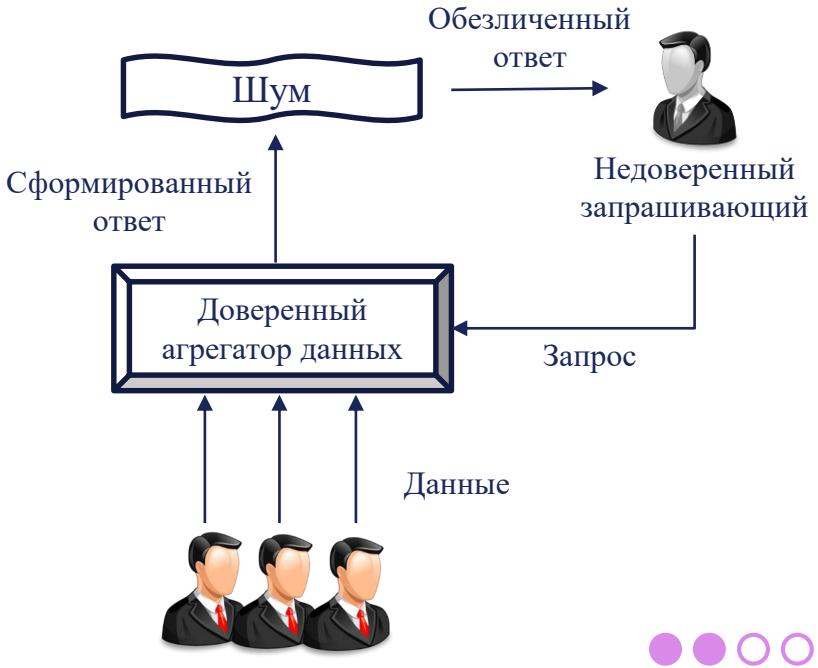
L – количество необходимых дополнительных запросов вычисления линейных функций;
n – количество переменных в линейной функции (операций умножения);
$\mathrm{M} = n - 1$ – количество операций сложения в линейной функции.

7. Противодействие угрозам логического вывода

7. Противодействие угрозам логического вывода

Искажающие методы.
Методы с добавлением шума в задаче статистического обезличивания.





7. Противодействие угрозам логического вывода

Искажающие методы.
Методы с добавлением шума в задаче статистического обезличивания.

С помощью данного рода методов решается задача статистического обезличивания, где подбор параметров шума зависит от функции в запросе, которую необходимо вычислить. Для этого используется механизм Лапласа. Для произвольной функции $f: \mathbb{N}^{|x|} \to \mathbb{R}$ механизм Лапласа определяется как

$$\mathcal{M}(x, f(), \varepsilon) = f(x) + Y,$$

где $f(x)$ – запрашиваемая функция, Y – случайная величина, имеющая распределение $Lap(y|\frac{\Delta f}{\varepsilon})$, Δf – качество аналитики, $\mathcal{M}(x, f(), \varepsilon)$ – ответ на запрос, а ε – это параметр безопасности , который подбирается в зависимости от статистики, количества запросов на вычисление этой функции и позволяется оценить ухудшение качества аналитики.

7. Противодействие угрозам логического вывода

Искажающие методы. Микроагрегация.

Организация	Количество работников
1	15
2	21
1	16
2	29
2	22
1	14

Организация	Количество работников
1	15
1	15
1	15
2	24
2	24
2	24

7. Противодействие угрозам логического вывода

Искажающие методы. Перемешивание (с обобщением).

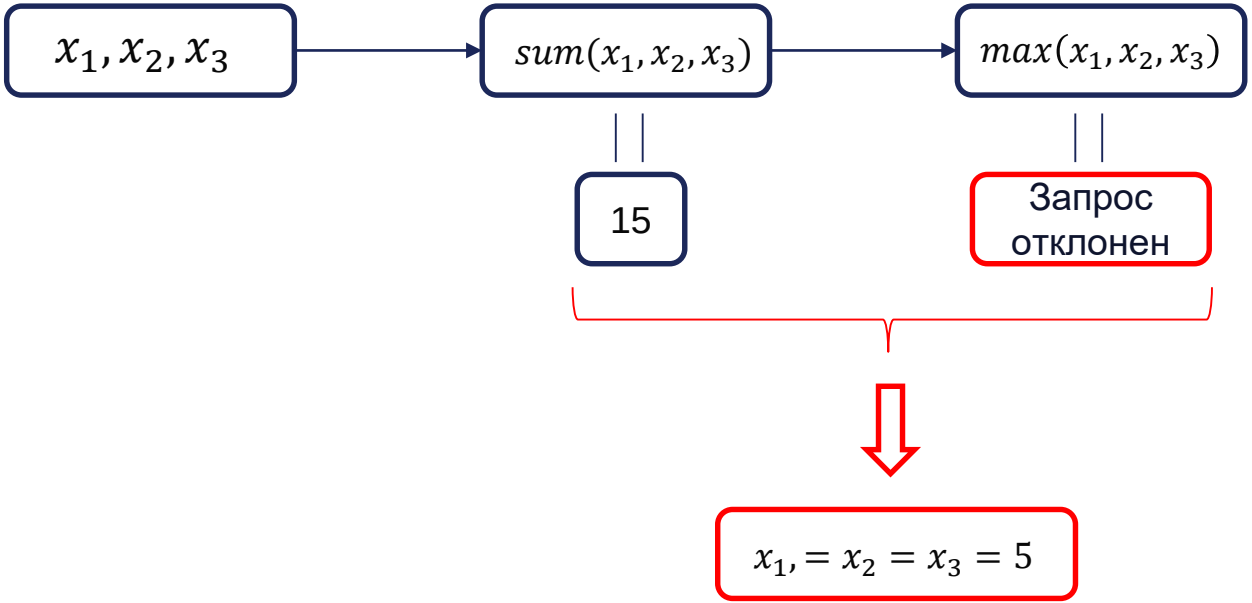
Организация	Количество работников
1	15
2	21
1	16
2	29
2	22
1	14

+ иерархия атрибутов →

Организация	Количество работников
1	< 20
1	< 20
1	< 20
2	≥ 20
2	≥ 20
2	≥ 20

7. Противодействие угрозам логического вывода

Неискажающие методы. Ограничение запросов.



7. Противодействие угрозам логического вывода

Неискажающие методы. Локальное подавление.

Адрес организации	Количество работников
Москва	28
Москва	28
Санкт-Петербург	31
Подольск	36
Подольск	36
Тула	31

Адрес организации	Количество работников
Москва	28
Москва	28
	31
Подольск	36
Подольск	36
	31

Спасибо за внимание!