Министерство образования и науки Российской Федерации

Санкт-Петербургский государственный политехнический университет

Институт Информационных Технологий и Управления

**Кафедра «Информационная безопасность компьютерных систем»**

**Обеспечение конфиденциальности проприетарных данных и вычислительных моделей при федеративном ML**

Выполнили

студенты гр. 5141001/30301 Б.Ю. Фитьо, А. С. Сиплатов

<*подпись*>

Преподаватель

профессор М.О. Калинин

<*подпись*>

Санкт-Петербург

2023

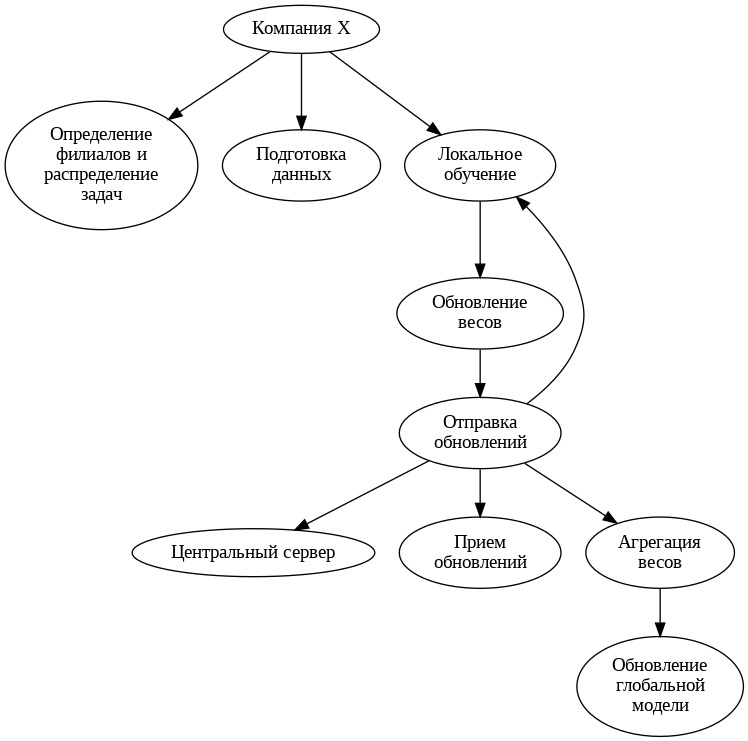
**Описание проблемы с конфиденциальностью данных в FL**

**Федеративное машинное обучение представляет собой подход к обучению моделей машинного обучения на распределенных данных, не передавая при этом сами данные на центральный сервер**. Вместо этого модель обучается локально на устройствах или в локальных центрах данных, и обновления модели передаются на центральный сервер для агрегации.

**Однако, несмотря на преимущества в снижении рисков передачи конфиденциальной информации, федеративное машинное обучение также вносит свои собственные вызовы в области конфиденциальности данных**.

Рассмотрим проблемы с конфиденциальностью данных в федеративном машинном обучении на примере.

**Сценарий реализации:**



**Проблемы с конфиденциальностью данных:**

1. **Риски перехвата**:

Передача обновлений на центральный сервер может быть подвержена риску перехвата, особенно если канал связи не защищен. Это может привести к утечке чувствительных данных.

1. **Информация о структуре модели**:

Обновления модели, передаваемые на сервер, могут содержать информацию о структуре модели, что может быть использовано для обратной инженерии и извлечения данных.

1. **Угроза федерационного сервера**:

Центральный сервер, хотя и не получает сами данные, все еще является целью для атак. Компрометация сервера может привести к утечке информации о структуре модели и обновлениям.

1. **Неоднородность данных**:

Различия в распределении данных между устройствами могут сделать возможным восстановление некоторой информации о данных, что может быть нежелательным с точки зрения конфиденциальности.

**Описание возможных альтернативных решений**

Существует несколько альтернативных решений и подходов, которые могут помочь в смягчении проблем с конфиденциальностью данных в федеративном машинном обучении:

1. **Шифрование данных**:

Все данные, находящиеся на локальных устройствах, могут быть предварительно зашифрованы. Только зашифрованные обновления передаются на центральный сервер. Шифрование может включать в себя техники, такие как гомоморфное шифрование, которое позволяет выполнять вычисления на зашифрованных данных.

1. **Локальная агрегация**:

Вместо передачи обновлений на центральный сервер, локальные устройства могут сначала агрегировать обновления среди себя, передавая только результаты на центральный сервер. Это может снизить количество передаваемой информации и, таким образом, риски утечки.

1. **Федеративное обучение с подделками** (Sythetic Data Federated Learning):

Создание синтетических данных на локальных устройствах может снизить риск утечки конфиденциальной информации, поскольку реальные данные не передаются. Однако, это требует аккуратной генерации данных для сохранения их статистических свойств.

**Выбор подхода с аргументацией**

**Локальная агрегация** представляет собой стратегию, при которой локальные устройства сначала агрегируют обновления среди себя, прежде чем передавать результаты на центральный сервер. Вот несколько аргументов в пользу этого подхода:

1. Снижение риска утечки данных:

Локальная агрегация уменьшает объем передаваемой информации, поскольку данные агрегируются на уровне филиалов. Это снижает риск утечки чувствительной информации.

1. Сокращение нагрузки на сеть:

Передача только результатов агрегации может существенно снизить объем передаваемых данных через сеть, что особенно важно при работе с большими объемами данных.

1. Сохранение локальной контролируемости:

Локальные устройства сохраняют контроль над своими данными и обновлениями, что может быть важным с аспекта соблюдения правил и регуляций.

1. Повышение эффективности обучения:

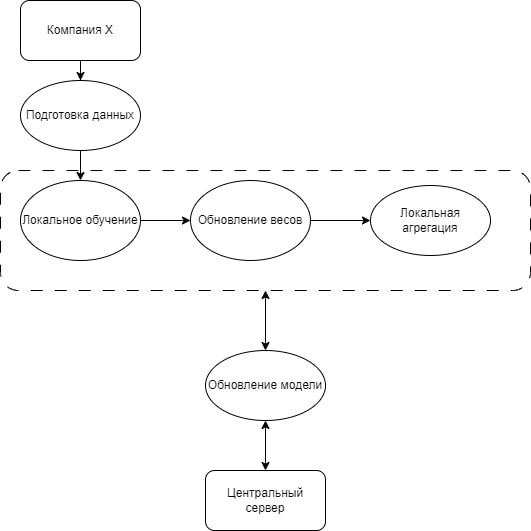
Агрегация на локальном уровне может быть более эффективной, поскольку обновления передаются в более сгруппированной форме, что может сэкономить ресурсы на центральном сервере.

1. Распределенное обучение:

Локальная агрегация согласуется с концепцией распределенного обучения, где данные остаются на уровне устройств, и процесс обучения распределен между ними.

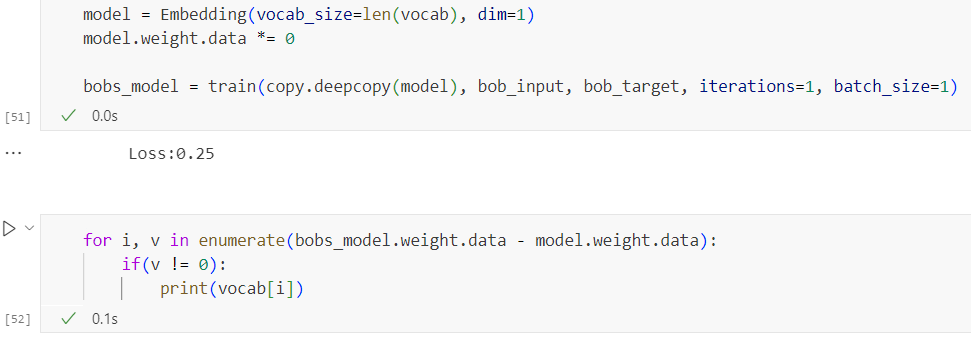
Выбор локальной агрегации обоснован с учетом уменьшения риска утечки данных, оптимизации сетевой нагрузки, сохранения контроля над данными и повышения эффективности обучения.

**Сценарий реализации при локальной агрегации:**

****

**Пример проведения атаки**

Учитывая какие весовые коэффициенты изменились, злоумышленник может выяснить какие данные использовались в модели.



В таком случае, конфиденциальность “пользовательских/клиентских” данных находится под угрозой.

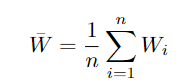
**Описание теории разработанного метода защиты**

Пусть W(i) - веса модели на i-м локальном устройстве после локального обучения.

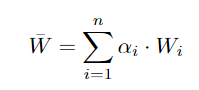
Локальная агрегация W(agg) может быть определена как функция агрегации, которая учитывает веса W(i) всех устройств:

*W(agg) = Aggregate(W(1), W(2), …, W(n))*

**В качестве функции агрегации может использоваться простая средняя агрегация**



**Также может использоваться взвешенная агрегация**

где a(i) - вес, присвоенный каждому локальному устройству

Локальная агрегация обычно включает в себя следующие шаги:

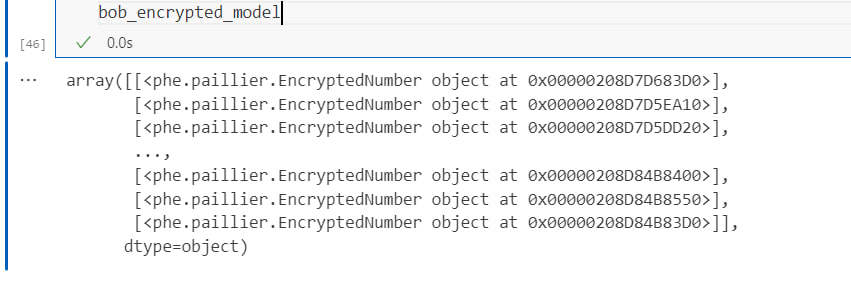
1. **Передача обновлений**: Локальные устройства обновляют свои модели в результате локального обучения.
2. **Локальная агрегация**: Обновления от каждого устройства агрегируются локально.
3. **Передача глобальной модели**: Агрегированные обновления отправляются на центральный сервер для обновления глобальной модели.

Преимущества:

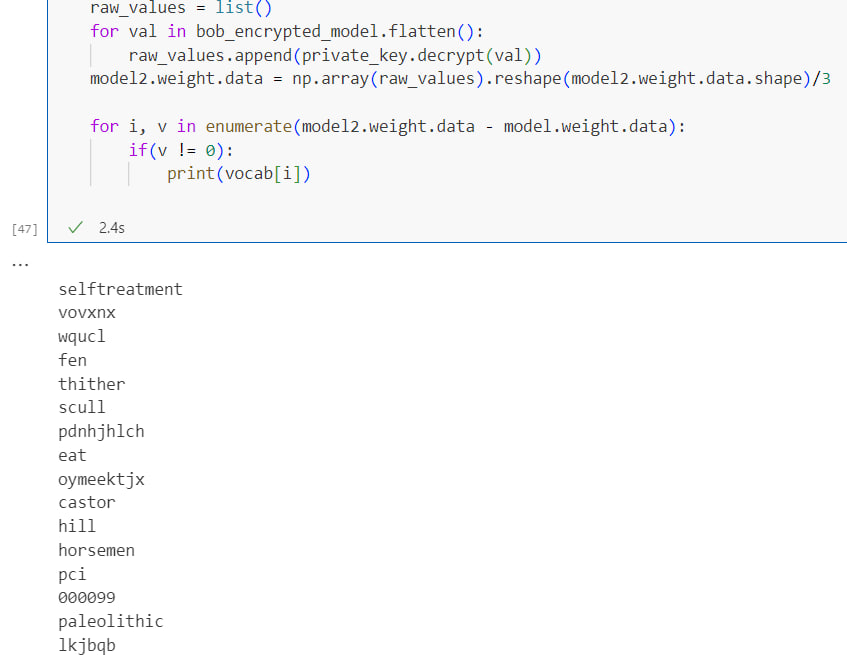
1. **Конфиденциальность данных**: Сокращается передача конфиденциальных данных на центральный сервер.
2. **Уменьшение нагрузки на сеть**: Сокращение объема передаваемых данных повышает эффективность сетевого взаимодействия.
3. **Локальный контроль**: Локальные устройства сохраняют контроль над своими данными и обновлениями.

**Локальная агрегация обеспечивает баланс между обучением на локальных данных и обновлением глобальной модели, минимизируя при этом риски утечки конфиденциальной информации и снижая нагрузку на сеть.**

**Проверка выбранного подхода**



Если злоумышленник всё-таки будет обладать знаниями и возможностями, для того, чтобы провести атаку, то засчет "зашумления", он не сможет сказать однозначно какие данные участвовали в обучении (таким образом, всегда будет вероятность, что полученные данные могут относится к шуму)



**Вывод**

Несмотря на сложность реализации при федеративном обучении и слабых мест, возникающих в связи с этим, показанный подход, при корректном использовании способен усложнить проведение атаки злоумышленником.

**Использованная и рекомендованная литература:**

1. Подход к обнаружению вторжений на основе федеративного обучения. [Электронный ресурс]. URL: <https://cts.etu.ru/assets/files/2023/sbornik/200-203.pdf> (дата обращения: 10.12.2023)

2. Глубокое обучение. Федеративное обучение. [Электронный ресурс]. URL: <https://habr.com/ru/companies/piter/articles/458800/> (дата обращения: 11.12.2023)

3. MLOps и ТОП-7 фреймворков для федеративного машинного обучения. [Электронный ресурс]. URL: <https://bigdataschool.ru/blog/what-is-federated-machine-learning.html> (дата обращения: 12.12.2023)