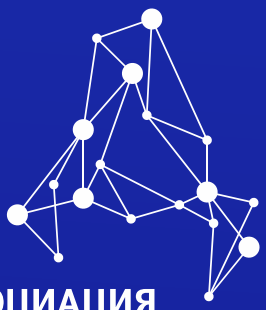


АССОЦИАЦИЯ
БОЛЬШИХ ДАННЫХ

КОНФИДЕНЦИАЛЬНЫЕ ВЫЧИСЛЕНИЯ И ДОВЕРЕННЫЕ СРЕДЫ ИСПОЛНЕНИЯ

Москва
2024

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Технологии защищенной обработки данных (ТЗОД) позволяют защитить данные и приватность пользователей, что дает возможность различным компаниям, государственным и исследовательским структурам развивать взаимовыгодное сотрудничество в области данных. Объединив данные, можно эффективно решать масштабные задачи, улучшать качество сервисов, развивать целые отрасли (туризм, медицина и др.), противостоять новым вызовам (организованное мошенничество, глобальные инфекции, изменение климата и пр.)

Конфиденциальные вычисления — одна из ключевых и, пожалуй, наиболее доступных и универсальных технологий защиты данных во время их обработки. Технология основана на использовании изолированных доверенных сред исполнения (ДСИ, в англоязычной терминологии — Trusted Execution Environment, TEE), в которых происходит обработка данных. Благодаря высокой степени защиты данных и масштабируемости, технология позволяет решать самые разные задачи: безопасно работать с чувствительными данными внутри компаний, выносить конфиденциальную нагрузку в облака, а также проводить совместную обработку информации нескольким компаниям без передачи коммерческой информации партнерам и без ее раскрытия.

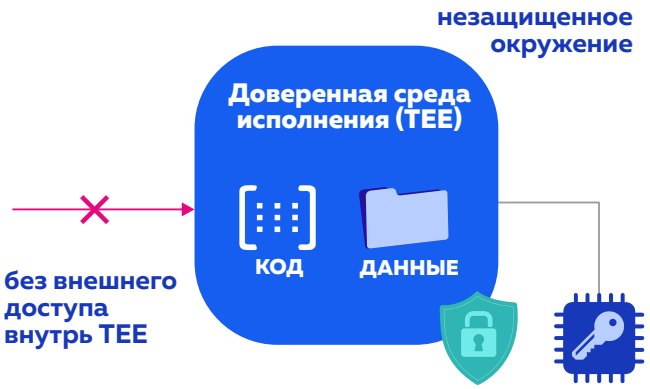
Технологии конфиденциальных вычислений позволят компаниям безопасно работать с моделями ИИ, защищенно передавая данные для обучения, дообучения и инференса моделей, а также позволят запустить следующий этап развития ИИ на основе конфиденциальных корпоративных данных, составляющих до 80% всех данных.

ЧАСТЬ 2. КОНФИДЕНЦИАЛЬНЫЕ ВЫЧИСЛЕНИЯ

Обзор технологии

Во многих источниках под конфиденциальными вычислениями понимают именно вычисления с использованием доверенных сред исполнения. Гармонизацией подходов и развитием этой технологии активно занимается Confidential Computing Consortium — альянс крупнейших ИТ-производителей, в который входят Intel, NVIDIA, AMD, ARM, Google, Huawei, Microsoft и другие технологические компании.

Суть TEE заключается в защите данных при их обработке путем изоляции и шифрования области памяти (анклава), в которой производятся конфиденциальные расчеты. Обычно такое шифрование выполняется с помощью секретного ключа, встроенного в оборудование на этапе его производства. Наиболее популярны реализации TEE с применением ключей, встроенных в процессор. Такой подход защищает данные в TEE от всех участников вычислений, в том числе от владельца и администратора сервера или облачной инфраструктуры, гипервизора, операционной системы, разработчиков прикладного ПО и др. Доступ к данным внутри TEE предоставляется только программам, которые помещены внутрь TEE.



TEE позволяет запускать согласованные, изолированные вычисления, при этом обеспечивается защита помещенного в них кода и данных от внешних угроз, включая вредоносное ПО и неавторизованный доступ. Извне доступны лишь результаты работы кода.

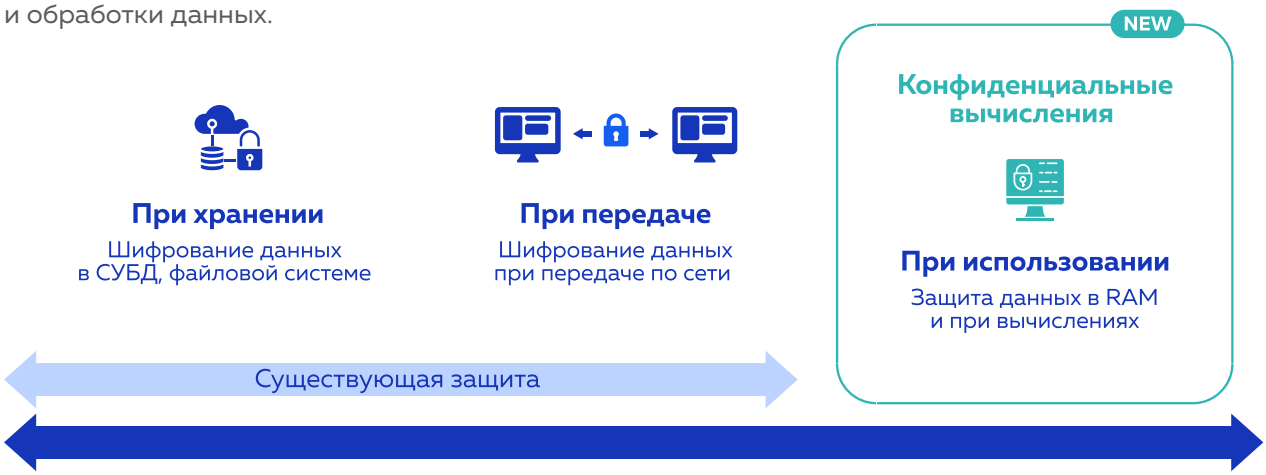
При запуске вычислений в TEE пользователь устанавливает защищенное соединение с применением ключей, сгенерированных процессором.

Проверить целостность анклавов пользователь может с помощью процедуры аттестации, сертифицированной производителем. Если аттестация прошла успешно, пользователь может быть уверен в том, что анклаву можно доверять: код и данные внутри него защищены.

Защита обеспечивается путем шифрования области памяти (Protected Range Memory) и ее изоляции от обычных инструкций процессора. Любая попытка внешнего взлома устройства приведет к выводу из строя процессора и прекращению работы TEE.

ЧАСТЬ 2. КОНФИДЕНЦИАЛЬНЫЕ ВЫЧИСЛЕНИЯ

Благодаря шифрованию памяти при обработке TEE позволяет замкнуть периметр вычислений и создать доверенный вычислительный контур без доступа к нему во время хранения, передачи и обработки данных.



Современные реализации TEE не только защищают данные, но и позволяют удостовериться в том, что в этой среде запускаются только авторизованные и согласованные партнерами вычисления. Код внутри TEE также защищен от просмотра и изменения, что обеспечивает сохранность интеллектуальной собственности партнеров, при этом у них остается возможность контролировать исходящие из TEE данные, а также проверять код до его помещения в доверенную среду исполнения.

TEE можно объединять в доверенные сети различных конфигураций, создавая конфиденциальные вычислительные кластеры и сети.

TEE поддерживаются ведущими производителями процессоров, что дает возможность подбирать техническое решение исходя из специфики конкретной задачи и имеющихся инфраструктурных ограничений.

	Intel SGX	Intel TGX	AMD SEV	NVIDIA H100	Huawei QingTian	Amazon Nitro
Защищаемая область	Процесс	ВМ	ВМ	ВМ+GPU	ВМ	ВМ
Корень доверия	CPU	CPU	CPU	CPU+GPU	Cloud Provier/ Hipervisor	Cloud Provier (Amazon Only)
Аттестация анклавов	Да	Частично	Частично	Частично	Да	Да
Возможность запуска на собственном железе	Да	Да	Да	Да	Да	Нет
Автономность работы	Частично реализуется в DCAP	Да	Да	Да	Да	Да
Максимальный размер памяти защищаемой области	До 1 ТБ	Не ограничено	Не ограничено	Не ограничено	Не ограничено	Не ограничено
Поддержка в облаках	MS Azure, G-Core, OVH и др.	MS Azure, G-Core, OVH и др.	MS Azure и др.	MS Azure	Huawei Cloud	Amazon

Стоит отметить, что процессоры с TEE свободно доступны на российском рынке. Большинство реализаций TEE предполагают наличие встроенного ключа и набора инструкций процессора без возможности удаленно отключить TEE. Поскольку аттестация проходит локально, работать можно без обращений к серверу производителя, что обеспечивает полную автономность использования технологии.

ЧАСТЬ 2. КОНФИДЕНЦИАЛЬНЫЕ ВЫЧИСЛЕНИЯ

Конфиденциальный искусственный интеллект (Confidential AI)

Согласно исследованиям, проведенным Cisco, у 40% компаний случались утечки персональных данных из-за использования ИИ. В 27% крупных компаний запретили использование генеративного ИИ, в 61% организаций введены ограничения на его использование сотрудниками. Также выявлено, что 62% потребителей обеспокоены тем, как организации применяют и используют ИИ, а 60% уже потеряли доверие к организациям из-за их практики в области ИИ.

Согласно исследованиям Gartner, более 50% ИИ-проектов в компаниях не оправдывают ожидания и только менее 10% проектов переходят в стадию промышленного использования. Основная причина — риски и ограничения на использование корпоративных данных.

Чтобы сделать ИИ более безопасным и защитить данные, рекомендуется осуществлять обучение, дообучение и применение моделей ИИ с использованием конфиденциальных вычислений.

Конфиденциальные вычисления для AI (CONFIDENTIAL AI)

Инференс моделей

Конфиденциальные вычисления позволяют не передавать данные запросов разработчику и владельцу моделей. Для этого обученная модель размещается внутри TEE и может принимать и отдавать данные через специальный сервис, работающий также внутри TEE.

Пример реализации безопасного инференса в TEE

Владелец данных и разработчик модели подготавливают соответственно датасеты и модель. Для защиты на стадии передачи в TEE информация шифруется. На следующем шаге для подтверждения подлинности выполняется удаленная аттестация TEE. Далее поставщик данных и разработчик модели обмениваются ключами и устанавливают защищенное соединение. После этого модель и датасет передаются внутрь TEE с использованием защищенного соединения, внутри TEE происходит расшифровка информации.

Наконец, запускается инференс модели. На этой стадии TEE обеспечивает изоляцию и защищенность вычислений. После завершения инференса результат шифруется и передается владельцу данных, который расшифровывает результат с использованием ключей соединения. После этого модель может быть удалена из TEE или сохранена для последующего использования.

Обучение моделей

С использованием конфиденциальных вычислений обучение и дообучение моделей также могут выполняться без передачи данных или моделей между партнерами.

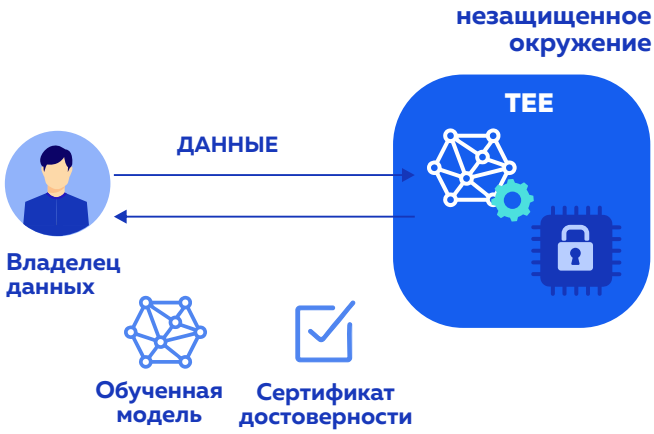
Пример реализации

Для обучения модели владелец данных подготавливает набор данных для обучения и первоначальную модель, на основе которой будет проводиться обучение.

Так же, как и в случае с инференсом моделей, данные и модели перед передачей в TEE шифруются, затем проводится аттестация TEE, устанавливается доверенное соединение между участниками, после чего зашифрованные данные и первоначальная модель передаются по защищенным каналам связи.

ЧАСТЬ 2. КОНФИДЕНЦИАЛЬНЫЕ ВЫЧИСЛЕНИЯ

Внутри TEE происходит изолированное обучение модели. По окончании процесса обучения полученные веса модели шифруются и передаются владельцу данных или разработчику моделей через защищенный канал.



Преимущества и недостатки TEE

Ключевые преимущества:

- Высокая сквозная защищенность данных: при использовании TEE они могут быть защищены на всех этапах — во время хранения, передачи и обработки.
- Конфиденциальность: данные и код защищены от внешнего доступа, даже если основная система скомпрометирована.
- Целостность кода и данных: они защищены от изменений — удаления и подмены.
- Аттестация: есть возможность проверить подлинность TEE и итоговых результатов вычислений.
- Низкая дополнительная вычислительная нагрузка, требуемая для обеспечения конфиденциальности: рост нагрузки составляет всего 10–20% по сравнению с увеличением в десятки, сотни и тысячи раз в случае применения других технологий.
- Не требуется адаптировать код для запуска в TEE: проекты в TEE могут быть запущены в течение нескольких часов, а не недель, как при использовании других технологий.
- Стандартные реализации TEE поддерживают большинство популярных библиотек и инструментов машинного обучения.

Недостатки технологии:

- Требуются процессоры специальных серий.
- В некоторых (особенно ранних) реализациях TEE имеются ограничения на объемы обрабатываемых данных.

Риски использования технологии подробнее рассмотрены далее в документе.

ЧАСТЬ 2. КОНФИДЕНЦИАЛЬНЫЕ ВЫЧИСЛЕНИЯ

Сравнение TEE с другими ТЗОД применительно к задачам аналитики и машинного обучения:

TEE В СРАВНЕНИИ С ДРУГИМИ ТЕХНОЛОГИЯМИ ЗАЩИТЫ ДАННЫХ

	Multi-party computation	Homomorphic encryption	TEE
Необходимость разработки модели под каждый кейс	Да	Да	Нет
Защита кода моделей	Нет	Нет	Да
Аттестация вычислительной среды	Нет	Нет	Да
Универсальность, поддержка различных моделей	Средняя	Средняя	Высокая
Вычислительный overhead (дополнительная вычислительная нагрузка для обеспечения конфиденциальности)	>1000%	>1000%	10-20%
Дополнительное время для подготовки и запуска кейса	Недели	Недели	Нет
Необходимость специальной экспертизы для адаптации модели	Да	Да	Нет
Необходимость специального оборудования (процессоры)	Нет	Нет	Да

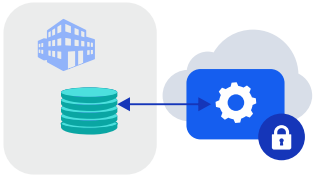
Варианты применения конфиденциальных вычислений

Активное развитие конфиденциальных вычислений в последние годы происходит за счет следующих основных сценариев:

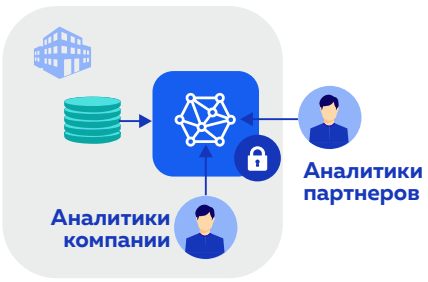
- Вынос чувствительных данных и расчетов в облака.
- Использование конфиденциального ИИ.
- Совместная обработка данных в рамках партнерства организаций, в том числе в области маркетинга и рекламы.
- Защита данных внутри компаний, в том числе в периферийных вычислениях.

1. Совместная аналитика и моделирование

2. Безопасная обработка данных в публичных облаках



3. Безопасная работа с чувствительными данными для аналитиков компании и партнеров



ЧАСТЬ 2. КОНФИДЕНЦИАЛЬНЫЕ ВЫЧИСЛЕНИЯ

Конфиденциальные вычисления в облаках

Конфиденциальные вычисления применяются для выноса нагрузки и защиты чувствительных данных в облаках, в том числе в кейсах, не предусматривающих создания частных облаков.

Крупнейшие мировые облачные провайдеры предоставляют для таких сценариев готовую развитую инфраструктуру. Так, в частности, облака Google и Microsoft Azure позволяют развернуть полноценные конфиденциальные виртуальные машины, в том числе с поддержкой Kubernetes, СУБД и других сервисов (см., например, пакет продуктов Azure Confidential Computing Protect Data In Use | Microsoft Azure).

В качестве демонстрации подхода конфиденциальных вычислений Microsoft одной из первых перенесла обработку карточных транзакций по продаже собственных лицензий в конфиденциальное облако.

Развитие партнерских отношений между организациями

По прогнозам компании McKinsey к 2030 году объем сетевой экономики достигнет 70 трлн долл. и составит около 25% от общего объема мировой экономики. Уже сегодня 71% организаций заявляют о том, что они готовы к работе в интегрированных экосистемах.

Согласно данным другого исследования McKinsey экономический эффект от широкого внедрения экосистем, основанных на открытых данных, может составить в 2030 году 1–1,5% ВВП в США, ЕС и Великобритании и до 4–5% в Индии. Эксперты уверены, что от этого выиграют все участники рынка, хотя и в разной степени.

Эти выводы говорят о том, что эффект от использования данных, имеющихся внутри организаций, приближается к своему пределу. Вместе с тем очень широкие перспективы открываются благодаря совместному использованию данных, которые предоставляют разные организации.

Основные индустриальные варианты сотрудничества в области данных представлены в этой таблице:

Банки и финтех	Медицина	Госсектор
• Совместная антифрод-аналитика • Развитие скоринговых моделей • Защищенный ML • Приватные блокчейн-транзакции • Data exchanges/marketplaces • Защищенный DeFi-доступ	• Исследования и аналитика на реальных данных пациентов для диагностики и разработки лекарств • Цифровой медицинский профиль • Развитие умного страхования, снижение фрода	• Обмен данными между учреждениями • Аналитика граждан 360 совместно с бизнесом • Отраслевые дата-платформы • Digital identity • Анонимные опросы
Технологии	**Ретейл и производство**	**Медиа и реклама**
• Key management • Защита персональных данных • IoT/Edge device and data security • IP protection • Приватность и ускорение в блокчейне	• Совместная клиентская аналитика • Совместная оптимизация операций • DLT-based tracking and provenance • Edge computing	• Работа без 3rd party cookie • Замеры влияния рекламы, эконометрика • Профилирование клиентов на данных нескольких партнеров

Борьба с мошенничеством (антифрод) в банках

БОРЬБА С ФИНАНСОВЫМ МОШЕННИЧЕСТВОМ

Полная картина движения средств с использованием анклавов

Конфиденциальные вычисления позволяют:

- Запускать согласованные модели на совместных данных
- Получать инсайты без раскрытия данных
- Соблюдать требования к конфиденциальности данных

Выгоды:

- Повышение % детекции фрода
- Снижение ложных срабатываний
- Ускоренное обучение на общих датасетах

В данном сценарии банки и другие финансовые организации, используя конфиденциальные вычисления, запускают на основе общих данных модели выявления мошеннических схем и операций. В этом случае не происходит раскрытие персональных и других чувствительных данных, поскольку данные в зашифрованном виде загружаются напрямую в анклав, где выполняются их объединение и безопасная обработка.

Такой анализ позволяет получить целостную картину движения денежных средств каждого клиента между его счетами в различных банках, причем без прямого обмена этими данными между банками. Запуск моделей на объединенных данных банков позволяет кратно повысить точность и скорость обнаружения мошеннических операций. Это очень важно, если учесть, что ущерб от финансового мошенничества составляет примерно 1,6 трлн долл. и оценивается в 2–5% мирового ВВП.

Ярким примером использования конфиденциальных вычислений для определения подозрительных транзакций является кейс международной системы совершения платежей SWIFT. Разработанная платформа позволяет дообучать общую модель с использованием данных партнеров платформы (а это более 11 500 финансовых организаций по всему миру), не передавая эти данные и сохраняя их конфиденциальность. Объединение данных позволяет добиваться ранее недостижимых показателей качества моделей, получать разностороннюю аналитику и ценные выводы.

ЧАСТЬ 2. КОНФИДЕНЦИАЛЬНЫЕ ВЫЧИСЛЕНИЯ

Экосистемы и партнерские отношения между компаниями

Партнерские отношения между компаниями, нацеленные на широкий охват потребностей клиентов, позволяют эффективно реализовать потенциал совместной работы с общей клиентской базой и общими знаниями о клиентах. Результатами совместных усилий становятся повышение выручки за счет кросс-продаж и персонализации, улучшение управления рисками, снижение затрат на привлечение и удержание клиентов, оптимизация бизнес-процессов.

Важно, чтобы подобная работа с общими данными была безопасной и управляемой. Иначе не исключены риски, связанные с нарушением договоренностей по использованию данных партнеров, а также с утечками. Например, если одна компания передает данные о своих клиентах для коммуникаций с ними, то другая компания может злоупотребить этой информацией — использовать ее для других задач или осуществлять больше коммуникаций, чем было согласовано. При небрежном отношении чувствительные данные могут попасть к конкурентам, а то и вовсе в открытый доступ.

Конфиденциальные вычисления позволяют управлять каждым фактом использования данных, будь то запуск расчета для рискового скоринга или подборка клиентов для рассылки. Чувствительные данные остаются внутри изолированной области, следовательно, злоупотребления и утечки исключены. Такой подход позволяет выстраивать более гибкие и масштабные партнерские отношения и экосистемы с сохранением контроля над данными со стороны участников.

Развитие бизнеса партнеров экосистемы за счет:

- Увеличения доступной клиентской базы
- Увеличения LTV с использованием новых знаний о клиентах
- Улучшения оценки рисков
- Оптимизации цепочек поставок и пр.
- Сохранения контроля над данными у партнеров с использованием конфиденциальных вычислений

ЧАСТЬ 2. КОНФИДЕНЦИАЛЬНЫЕ ВЫЧИСЛЕНИЯ

Некоторые практические результаты партнерских отношений и совместной работы с данными отражены в таблице:

Банк топ-10 на данных группы

1. Улучшили модель склонности к премиальным продуктам. Повысили конверсию продаж премиальных продуктов. Снизили отток премиальных клиентов.
2. Улучшили модель предсказания дохода клиентов. Использовали данные в скоринге клиентов, снизили риск невозврата, увеличили объемы выданных кредитов.
3. Повысили конверсию для дебетовых карт на данных с телеком-оператором.

Страховая компания топ-5 на данных группы

1. Провели верификацию контактных данных для своих клиентов. Расширили базу для прямых коммуникаций.
2. Уточнили модели Next best offer (NBO). Повысили конверсию продаж.
3. Уточнили модель страхования сотрудников корпоративных клиентов (WSM).

Телеком-оператор на данных группы

1. Развивает Adtech-платформу. Увеличили охват кампаний в несколько раз и с повышением информации по каждому клиенту
2. Разработали модели привлечения новых B2B-клиентов на данных группы.
3. Улучшили модели оттока для B2C-сегмента.

Кейсы применения технологии TEE

Здравоохранение и медицина

В изолированных средах исполнения медицинские данные пациентов обрабатываются безопасно, без раскрытия. Среди наиболее частых сценариев в медицине — отбор кандидатов для клинических исследований и использование реальных клинических данных для проведения фармацевтических исследований и разработок новых препаратов.

ПОИСК КАНДИДАТОВ ДЛЯ КЛИНИЧЕСКИХ ИССЛЕДОВАНИЙ

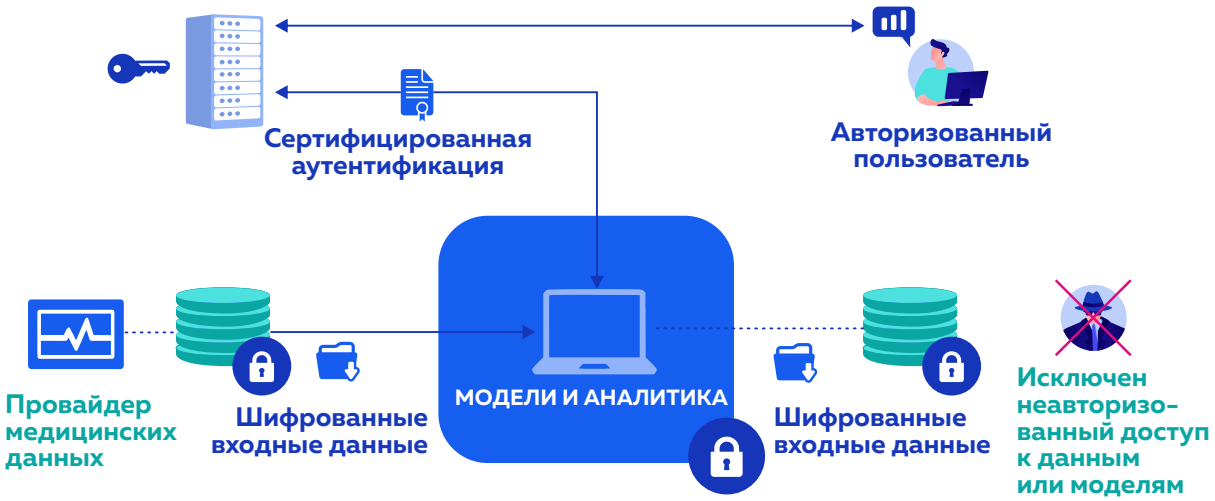


Рисунок демонстрирует безопасное развертывание в анклаве Intel SGX модели, предназначенной для анализа рентгеновских снимков грудной клетки с целью выявления признаков заболеваний. На протяжении всего времени выполнения алгоритма данные шифруются в памяти и обрабатываются в незашифрованном виде только в пределах анклава. Перед передачей результатов вовне, они шифруются, что обеспечивает полную сквозную защиту не только обрабатываемых медицинских данных, но и интеллектуальной собственности, содержащейся в коде приложения, развернутого в анклаве.

ИСПОЛЬЗОВАНИЕ КЛИНИЧЕСКИХ ДАННЫХ ДЛЯ РАЗРАБОТКИ ЛЕКАРСТВ

Intel SGX использует и американская ИТ-компания Leidos, безопасное развертывание для оптимизации и ускорения процесса клинических испытаний лекарственных препаратов. В данном кейсе конфиденциальные вычисления обеспечивают основу для работы централизованного портала, управляющего распределенной сетью анклавов, которые Leidos использует для обмена критически важными данными в режиме реального времени с соблюдением строгих требований к конфиденциальности информации о пациентах. Защищенные анклавы получают извне запросы, проверяют правила политик, собирают данные и возвращают результаты на портал. Запрашиваемые клинические данные токенизируются и остаются зашифрованными при хранении, передаче и использовании.

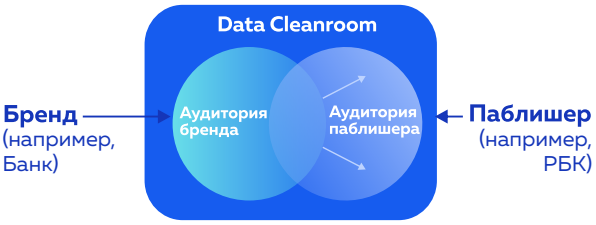
Медиа и реклама

Оценка влияния рекламы

Объединение данных о показах рекламы (например, в каналах IP-TV) с данными о покупках этих товаров (например, в розничных сетях) позволяет достоверно измерять эффективность рекламы, в том числе ее вклад в динамику продаж. Построенные на этих данных эконометрические модели позволяют оптимизировать рекламные миксы (соотношения между площадками, каналами и временем размещения рекламы). Понимание портрета активной аудитории обеспечивает возможность точнее доносить сообщения бренда.

Таргетинг и замеры рекламы без использования веб-идентификаторов

Все большее количество интернет-браузеров запрещают третьим сторонам использовать веб-куки для идентификации и отслеживания действий пользователей. Один из способов таргетирования рекламы в интернете, не использующий веб-куки, базируется на применении технологии Data CleanRoom (DCR) — основанной на конфиденциальных вычислениях безопасной среды для совместного использования сырых пользовательских данных. Поставщиками данных для интернет-маркетинга и рекламы, как правило, выступают паблишеры (издатели) и бренды (рекламодатели). DCR позволяют рекламодателям искать пересечения своей аудитории с аудиторией паблишера, выбирать целевых клиентов, находить похожих и оценивать результаты кампаний.



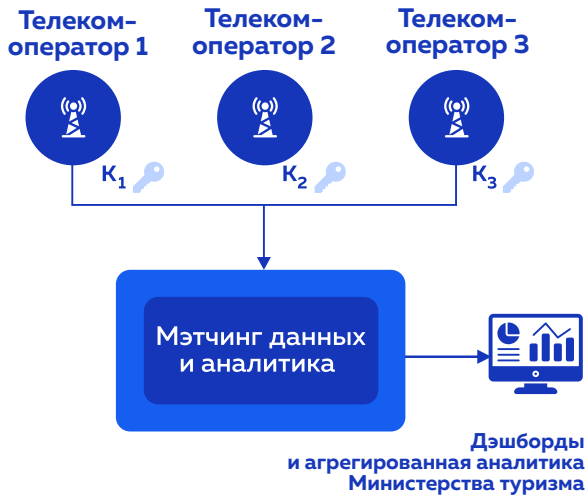
- Аудитории для откртки
- Look-a-like-расширения аудиторий
- Замеры эффективности
- Аналитика и инсайты

Некоторые примеры использования:
Renault повышает коэффициент конверсии на 18% благодаря совместной с Axel Springer кампании по сбору исходных данных

Крупный швейцарский банк в партнерстве с паблишерами снижает затраты на просмотр одной страницы на 44% с использованием DCR-платформы для совместной работы с данными их аудитории

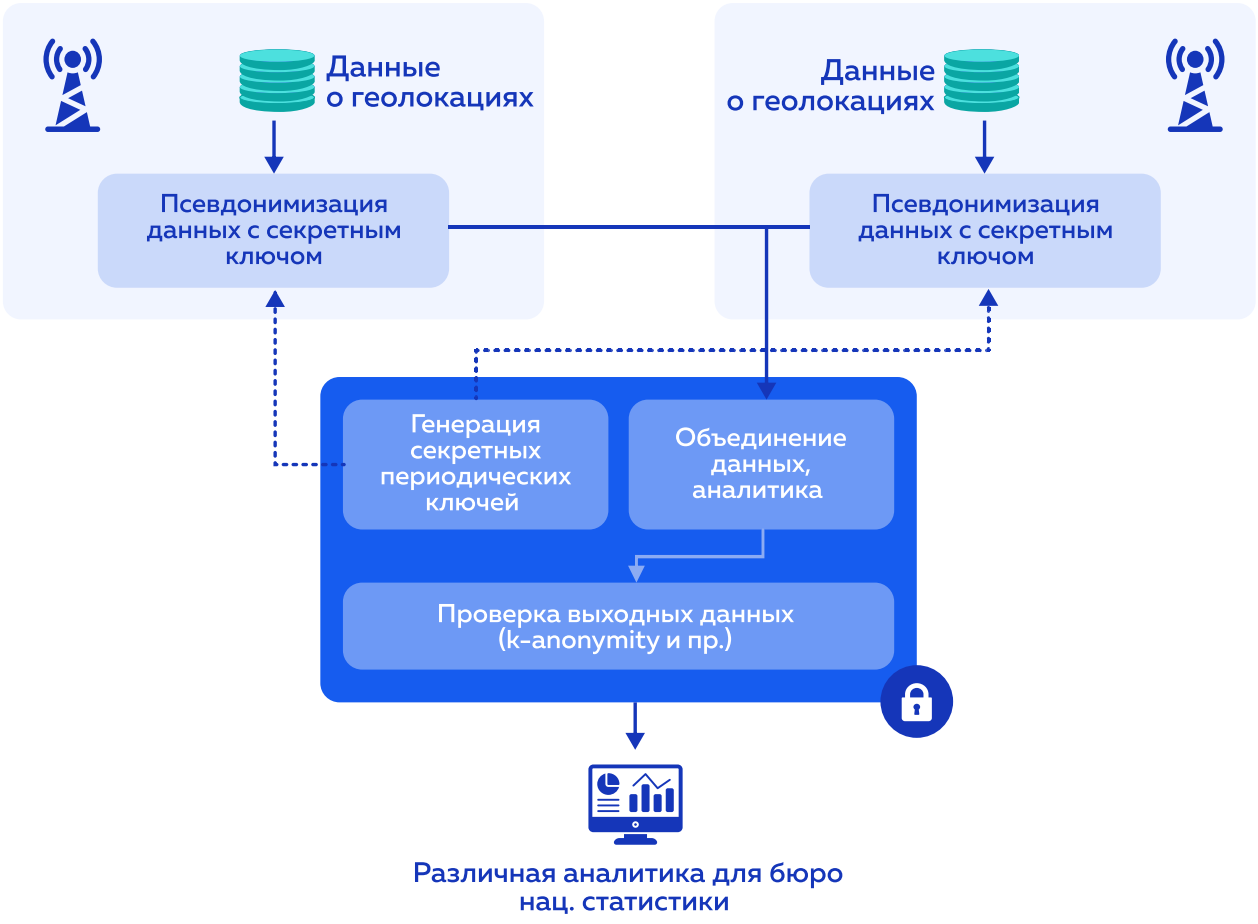
Государственная статистика

Объединение данных телеком-операторов для развития туристического рынка



Министерство туризма Индонезии реализовало конфиденциальный обмен наборами данных между двумя операторами мобильной связи через TEE. Целью проекта было получение статистики туризма на основе объединенных обезличенных данных двух сотовых операторов. Эта статистика формируется путем мэтчинга этих данных в анклаве Intel SGX. В результате проекта министерство получило информацию о количестве абонентов в роуминге и об общих роуминговых клиентах двух сотовых операторов, что позволило точно рассчитать долю рынка роуминга. Этот подход используется в официальной статистике Индонезии и сегодня: статистические показатели туризма ежемесячно составляются на основе данных об абонентах мобильной связи. Что важно, решение сделало возможным конфиденциальный обмен наборами чувствительных данных в рамках существующего законодательства.

Объединение данных о перемещениях для национальной статистики



ЧАСТЬ 2. КОНФИДЕНЦИАЛЬНЫЕ ВЫЧИСЛЕНИЯ

Данные сотовых операторов активно используют и Европейская статистическая система (ЕСС). Чтобы свести к минимуму риски конфиденциальности, ЕСС готовит рекомендации по использованию конфиденциальных вычислений при обработке данных сотовых операторов для целей официальной статистики — в частности, в ходе интеграции данных, полученных не только от сотовых операторов, но и из других источников. Такой подход обезопасит и персональные данные абонентов, и конфиденциальную информацию операторов связи. В настоящее время ЕСС разрабатывает стандарты и методологическую основу для мер защиты на основе ТЗОД, а также модель доступа к данным, адаптированную к требованиям официальной статистики и обеспечивающую при этом приватность персональных данных граждан.

Кроме того, на уровне Европейской комиссии начат пересмотр Регламента 223/2009 о европейской статистике, нацеленный на «использование потенциала новых источников данных». Ожидается, что в пересмотренном регламенте будут предусмотрены правовые механизмы для повторного использования ПДн, включая данные сотовых операторов, для целей европейской статистики.

Также ЕСС прорабатывает оперативные механизмы доступа к данным. Подход ЕСС основывается на том, что для обеспечения эффективной защиты персональных данных абонентов и конфиденциальной информации операторов следует выполнять, по крайней мере, некоторую часть обработки локально, вблизи того места, где генерируются данные (то есть на площадках операторов), а затем экспортировать полученные с определенной степенью агрегирования данные в национальные институты статистики. При этом более высокие уровни агрегирования этих данных будут соответствовать более низким уровням чувствительности данных и связанных с ними рисков. Соответственно, экспорт сотовыми операторами агрегированных (то есть предварительно обработанных) данных будет менее рискованным, ресурсоемким и более социально приемлемым, чем перемещение сильно детализированных сырых данных.

Примеры использования конфиденциального искусственного интеллекта

CONFIDENTIAL AI: ЗАЩИТА ДАННЫХ ПРИ ОБУЧЕНИИ И ИСПОЛЬЗОВАНИИ МОДЕЛЕЙ

Confidential Training

Защита исходных данных и их влияния на модель

Примеры:

Обучение моделей на генетических данных пациентов

Обучение моделей для self-driving cars на данных видеорегистраторов клиентов

Confidential Fine-Tuning

Дообучение общих моделей данными компаний

Компании дообучают общие GenAI-модели на своих данных

Confidential Multi-Party Training

Модель обучается на чувствительных данных нескольких партнеров

Общие антифрод-модели между банками (доступ к транзакциям в других банках, снижение false-positive срабатываний)

Конфиденциальные вычисления для экосистем и партнерств (без передачи и раскрытия данных)

Confidential Inferencing

Данные запросов к модели остаются приватными

Идентификация клиентов по голосу /речи без передачи данных (например, в публичных местах)

Диагностика заболеваний по снимкам

Конфиденциальная умная колонка (данные запросов не сохраняются и не передаются) или конфиденциальная умная CCTV-камера (только факты нарушений и минимально достаточные подтверждения выходят из TEE)

Модель рисков TEE

Предмет исследования

Доверенная среда исполнения (TEE) — это защищенная область внутри основного процессора, которая обеспечивает конфиденциальность и целостность загруженных в нее данных и кода. TEE изолирует выполнение и данные от остальной системы, обеспечивая более высокий уровень безопасности. Существует несколько основных реализаций TEE (Intel SGX, ARM TrustZone, AMD SEV, NVIDIA Confidential Computing и др.) со схожими основными функциями, но с различиями в архитектуре.

Риски TEE

Хотя TEE представляет собой хорошо защищенную технологию, она встроена в общие процессы обработки данных. Как следствие, на защищенность результатов оказывают влияние несколько групп факторов: контекстные риски (связанные с организацией, эксплуатирующей TEE, и настройками среды), риски непосредственно технологии TEE и риски алгоритмов, работающих внутри TEE.

Причинами рисков непосредственно TEE могут быть уязвимости самого оборудования и ПО, используемых для обеспечения защищенного взаимодействия. В случае с Intel SGX к наиболее значимым относят риски следующих атак:

- Атаки на побочные каналы, использующие физические особенности TEE для извлечения конфиденциальной информации. В числе таких атак — Prime + Probe (атака на кэш-память, реализуемая в виде подмены данных кэша для извлечения ключей), Flush + Reload (очистка кэша и замер времени доступа после выполнения операций), Spectre (позволяющее получить доступ к памяти спекулятивное выполнение инструкций процессором — до их фактического подтверждения) и Meltdown (чтение системной памяти, в том числе памяти ядра).

- Атаки Rollback и Forking, а также Foreshadow — манипуляции с состоянием TEE с целью отката или создания нескольких параллельных состояний.

Подразумевается, что при проведении этих атак потенциальный злоумышленник в течение длительного времени имеет непосредственный доступ к оборудованию TEE (например, к высокоточному таймеру), обладает соответствующими знаниями и техническими возможностями и может запускать специально подготовленные расчеты. В этом случае нарушается основное предназначение TEE — экранировать процесс обработки чувствительных данных, в том числе от оператора TEE.

Модель рисков для TEE

Комплексная (каскадная) модель рисков для TEE может быть представлена в виде уравнения риска:

$$R = P\_context \times P\_TEE \times P\_data \times I\_impact$$

Здесь:

Контекстные риски (P_context): оцениваются на основе стандартного скорингового подхода.

Риски данных (P_data): оцениваются с использованием моделей k-анонимности или утечки данных (как результат рисков выделения, связываемости и вывода).

Риски TEE: оцениваются как взвешенная сумма вероятностей различных атак:

$$P_{TEE} = \sum_{i=1}^{n} (w_i \cdot P_i)$$

Каждая составляющая P_i (успешность i-той атаки) может быть оценена на основе необходимых ресурсов (время T_i , вычислительная мощность C_i ,оборудование/доступ E_i):

$$P_i = \frac{P_{attack}}{(T_i \times C_i \times E_i)}$$

Проведение оценки рисков в разных исследованиях и экспериментах показывает, что суммарные вероятности по рискам TEE могут составлять величины в среднем порядка 10^-5, однако этот показатель может варьироваться в зависимости от специфики атаки, условий эксплуатации и наличия защитных механизмов.

ЧАСТЬ 2. КОНФИДЕНЦИАЛЬНЫЕ ВЫЧИСЛЕНИЯ

Группа рисков	Описание, примеры	Вероятность возникновения	Меры минимизации
Контекстные риски	Управление процессами: - Управление ключами - Управление доступом к базам данных, откуда данные загружаются в TEE области	**Средняя**	Использование KMS с шифрованием данных (HashiCorp Vault и т. п.), построение безопасной архитектуры хранения ключей и реквизитов доступа к данным вне TEE
Риски технологии TEE	Возможность неавторизованного доступа внутрь TEE - Атаки по побочным каналам - Потенциальные уязвимости нулевого дня	**Низкая** Нет известных случаев утечек данных. Для реализации атак требуется специальное оборудование, физический доступ к процессору, доступ к кэшу, возможность запускать одну и ту же нагрузку в течение длительного времени	Своевременное обновление микрокода процессора и установка других обновлений безопасности. Мониторинг уязвимостей и построение архитектуры, позволяющей блокировать доступ к собственному TEE или взаимодействие с внешним TEE
Риски алгоритмов (данных)	Запущенный в TEE скрипт может выгрузить исходные (персональные) данные. TEE позволяет запускать любые скрипты, поэтому может быть запущен скрипт, который раскроет исходные, промежуточные или итоговые данные, содержащиеся в анклаве во время исполнения	**Высокая**	- Проверка скриптов партнерами перед запуском в TEE - Встроенный дополнительный контроль выходных результатов внутри TEE. Например, контроль объемов выгружаемых данных, контроль отсутствия данных по отдельным клиентам, корреляций с входными данными и пр. - Многослойная архитектура ПО TEE (например, LibOS (occlum, gramine) —> приложения (например, Aggregion DDM) —> пользовательский код (например, пайплайны, запускаемые в DDM)) позволяет упростить проверку при изменении TEE. Чем ниже слой, тем реже он меняется. - Мультисервисная архитектура TEE: редко меняющиеся сервисы TEE можно вынести в отдельные приложения, а для взаимодействия всех TEE группы использовать аттестованные шифрованные каналы и общие хранилища секретов

Выводы

TEE предоставляет надежную, доступную и в высокой степени универсальную среду для безопасной обработки данных. Выводы и рекомендации относительно использования TEE следующие:

TEE ранних версий были ограничены в доступных ресурсах (например, лимитированы объемы памяти, к тому же доступна только однопоточность), однако современные реализации технологии не имеют таких ограничений. В ранних версиях память и вычислительная мощность внутри анклава были ограничены, решить многие «тяжелые» задачи (например, обучение больших моделей ИИ) было трудно или вовсе невозможно в условиях ограничения ресурсов. В последних реализациях TEE (Intel TDX, AMD SEV, Huawei Qingtian) этих ограничений нет, а в SGX версии 2 (начиная с линейки Intel Ice Lake) анклав ограничен объемом оперативной памяти до 1 Тбайт, чего для большинства задач вполне достаточно.

Имеются теоретические уязвимости TEE для атак на побочные каналы, однако на практике известных случаев подобного рода утечек не было. TEE не может полностью предотвратить все атаки на побочные каналы. Однако данного рода атаки крайне трудно реализовать, а кроме того, они являются весьма дорогостоящими. Как правило, для их реализации требуется постоянный физический доступ к процессору с использованием специального дорогостоящего оборудования (лаборатории), а также возможность многократного перезапуска одного и того же приложения. К тому же в большинстве случаев результатом успешной атаки может быть не утечка конкретных данных, а частичная деанонимизация приложения (например, информация о том, какая именно нагрузка в какие периоды времени выполнялась, какое количество потоков использовалось, какие алгоритмы шифрования применялись). Несмотря на то, что возможности атак по сторонним каналам были найдены лабораториями, реальные прецеденты их применения неизвестны.

Риски непосредственно технологии TEE мы оцениваем как низкие, однако контекстные риски, связанные с управлением ключами, качеством ПО, реализующего анклав, и пр., а также риски исполняемых внутри TEE алгоритмов требуют особого внимания.

Злоумышленник не имеет доступа внутрь TEE, однако может взаимодействовать с ним через интерфейсы, предоставляемые тем ПО, которое работает внутри анклава. Поэтому злоумышленник может предпринимать такие же попытки проникновения, как и в случае взлома обычного сервера в интернете. Соответственно, так же, как и в случае сервисов, размещаемых в интернете, ПО для анклава необходимо разрабатывать с применением тех же подходов к безопасности: в нем должны быть надежные механизмы авторизации, минимизированы зависимости, должны своевременно обновляться библиотеки, проводиться ручные и автоматизированные проверки безопасности и т. п.

Санкционные риски: технология доступна на рынке наравне с другими западными процессорами и может использоваться автономно, без доступа к сервисам производителя. Технологии TEE от Intel и AMD являются лишь расширенным набором инструкций процессора, поэтому связанные с TEE санкционные риски равны санкционным рискам, касающимся любых других расширенных наборов инструкций, таких как MMX, SSE, AES NI и т. п. (в большинстве ПО так или иначе используются расширенные наборы инструкций). Другими словами, доступность технологии строго равна доступности самих процессоров. Дополнительные возможности, такие как аттестация, также не зависят от внешних сервисов и могут производиться локально.

Технология ARM, которая потенциально может быть доступна на российских процессорах. Российские процессоры «Байкал» используют архитектуру ARM, которая поддерживает собственную реализацию TEE — TrustZone. Известно о случае применения сред TEE на основе TrustZone на процессорах «Байкал» и сертификации такого контура в ФСТЭК. И хотя TrustZone ограничена в возможностях и не может применяться для сложных вычислений, в ARM Limited была разработана и опубликована альтернатива для архитектуры ARMv9-A под названием ARM CCA, позволяющая создавать TEE с полноценной ОС внутри. Вероятнее всего, новые семейства процессоров «Байкал» будут поддерживать эту технологию, что позволит создавать доверенные контуры исполнения, построенные исключительно на российских микросхемах и ПО.

Юридическая интерпретация TEE

Приведем оценку описанной выше технологии доверенных сред исполнения и возможных кейсов ее применения с точки зрения законодательства РФ о персональных данных.

Базовые положения законодательства

В соответствии со статьей 3 Закона 152-ФЗ [1], персональные данные — это любая информация, относящаяся прямо или косвенно к определенному или определяемому физическому лицу. Такая формулировка соответствует указанному в статье 2 Конвенции 108 [2],что позволяет сравнивать российские подходы в регулировании вопросов использования TEE с зарубежными.

К персональным данным (ПДн) могут относиться фамилия, имя и отчество, номер телефона и адрес электронной почты, должность и место работы, история покупок, данные о здоровье, предпочтения в выборе фильмов, музыки и т. д. При этом неважно, есть ли возможность на основе этих данных непосредственно определить конкретного человека (определенное физическое лицо). Достаточно того, что на основе этих данных можно выделить лицо среди группы других (определяемое физическое лицо).

Компания, обрабатывающая ПДн граждан в своих коммерческих целях и определяющая объем обрабатываемых данных и способы их обработки, считается оператором (контролером). Оператор вправе поручать обработку ПДн третьим лицам (обработчикам), оставаясь при этом ответственным за соблюдение требований законодательства при такой обработке перед субъектом ПДн и государством. В законодательстве РФ, в отличие от некоторых зарубежных нормативных актов[3], нет понятия «совместное операторство», поэтому каждая компания, обрабатывающая ПДн в своих целях, должна иметь свое законное основание для обработки — согласие, договор, закон, законный интерес и т. п.

Важно также отметить, что, с точки зрения российского регулятора (Роскомнадзора), ПДн сохраняют за собой этот статус как в случае их шифрования, так и после обезличивания[4].

Зарубежные регуляторы придерживаются такого же подхода, рассматривая шифрование и обезличивание лишь как способы защиты ПДн, но не как методы, позволяющие исключить квалификацию данных в качестве персональных[5] . Соответственно, и доступ третьих лиц к этим данным оператор обязан обеспечивать с учетом требований законодательства о ПДн.

В этом плане необходимо отличать обезличивание (псевдонимизацию) данных от анонимизации, когда данные теряют статус персональных[6] . После обезличивания возможно восстановить связь между данными и конкретным субъектом ПДн, которому они принадлежат (деобезличивание). Анонимизация должна быть необратима, и полученные в результате ее применения данные должны быть такими, чтобы их нельзя было соотнести с каким-либо человеком. При этом понятие «анонимизация» в действующем законодательстве РФ отсутствует. И важно, что при существующем уровне развития технологий многие данные, которые лишь считаются анонимными, при определенных усилиях можно атрибутировать, сопоставив с конкретным человеком.

Ниже мы рассматриваем ситуации с использованием TEE для обработки именно персональных данных, которые при этом могут быть обезличены или зашифрованы с целью лучшей защиты.

1 Федеральный закон 152-ФЗ от 26.07.2006 «О персональных данных».

2 Конвенция о защите физических лиц в отношении автоматизированной обработки персональных данных 1981 г.

3 См., например, ст. 26 GDPR.

4 Милош Вагнер (заместитель руководителя Роскомнадзора): «...Обезличивание изначально задумывалось как мера обеспечения безопасности персональных данных, которая снижает вероятность нанесения вреда гражданину в случае компрометации таких данных... Данные, которые получаются в результате процедуры обезличивания, не перестают быть персональными... Если мы говорим про оборот обезличенных данных, то, на наш взгляд, при таком обороте должна обеспечиваться и конфиденциальность, и безопасность таких данных».

5 См, например, п. 26 преамбулы GDPR и п. 5 ст. 4 GDPR.

6 См., например, п. 26 преамбулы GDPR и п. 54 Guidelines 4/2019 on Article 25 Data Protection by Design and by Default Version 2.0.

ЧАСТЬ 2. КОНФИДЕНЦИАЛЬНЫЕ ВЫЧИСЛЕНИЯ

Описание проблематики TEE с позиции законодательства РФ о ПДн

Используя TEE, компании — участники совместной обработки помещают ПДн в память принадлежащих им процессоров. Поскольку процессоры интегрированы в единую доверенную среду исполнения, каждая компания-участник имеет возможность совершать те или иные действия с данными других компаний-участников. С точки зрения законодательства РФ о ПДн это может рассматриваться как доступ к ПДн. А значит, компании — участники совместной обработки по умолчанию рассматриваются как операторы этих ПДн и должны иметь свое законное основание для обработки всех ПДн, загруженных на процессоры, включенные в TEE.

При квалификации действий компаний-участников как доступа к ПДн всех других компаний-участников, та организация, которая поместила ПДн в свой процессор, должна иметь законное основание на передачу (по сути, предоставление доступа) ПДн всем другим компаниям-участникам.

В текущей российской практике основным законным основанием, которое в обоих случаях могут использовать компании — участники совместной обработки, является согласие субъекта ПДн. В силу требований законодательства РФ о ПДн оно должно быть конкретным, предметным, информированным, сознательным и однозначным. Получить его от субъекта ПДн может быть затруднительно даже для компании-оператора, помещающего ПДн в TEE, не говоря уже о других компаниях — участниках совместной обработки. Иногда для обоснования законности передачи ПДн третьим лицам можно использовать договор (пользовательское соглашение), но на практике не все компании включают в него соответствующее условие.

Возможные варианты применения TEE с учетом юридических особенностей технологии

Преимуществом TEE как одной из технологий защищенной обработки данных является то, что доступ компаний — участников совместной обработки к ПДн, загружаемым другими компаниями-участниками, может быть ограничен технически. Например, участники могут договориться об использовании ограниченного набора команд (скриптов), которые разрешено выполнять с ПДн, помещенными в TEE, и предусмотреть определенные меры контроля за соблюдением этой договоренности. Первая из них — проверка задаваемых компаниями-участниками команд отдельным модулем входного контроля до начала выполнения этих команд, а также фиксация с использованием, например, блокчейн-протокола всех заданных компаниями-участниками команд с созданием неизменяемых записей (отчетов) об использовании TEE. Вторая мера — это добавление в TEE модуля выходного контроля, отслеживающего, что в результате выполнения заданной команды ни одна из компаний-участников не получила ПДн, загруженные в TEE другой компанией-участником.

С учетом этого мы видим следующие варианты обеспечения юридической чистоты использования TEE для обработки ПДн.

Варианты обеспечения юридической чистоты при использовании TEE

Со стороны регулятора

Как мы уже сказали, в российской практике основным законным основанием, которое могут использовать компании — участники совместной обработки, является согласие субъекта ПДн. Такое основание, как законный интерес[7], считается рискованным для использования в силу негативного отношения к нему со стороны надзорного органа — Роскомнадзора. При этом в зарубежных правовых пространствах (например, в ЕС и США) законный интерес оператора ПДн является полноценным правовым основанием для обработки ПДн и часто используется компаниями именно в ситуациях, когда получение согласия на обработку со стороны субъекта ПДн затруднительно. Важно подчеркнуть, что и в рамках этих регуляторных требований выбор законного интереса как основания обработки ПДн не является индульгенцией для компании-оператора — ей необходимо в каждом случае оценивать, насколько интересы субъекта ПДн сбалансированы с интересами оператора и не приводит ли такая обработка ПДн к существенным нарушениям прав субъекта ПДн.

[7] П. 7 ч. 1 ст. 6 Федерального закона 152-ФЗ от 26.07.2006 «О персональных данных».

ЧАСТЬ 2. КОНФИДЕНЦИАЛЬНЫЕ ВЫЧИСЛЕНИЯ

Как мы уже отметили, использование TEE помогает в значительной мере ограничить объем действий, которые могут быть совершены с ПДн другими компаниями-участниками. Позиция российского регулятора о допустимости использования законного интереса как основания обработки ПДн в рамках TEE и со стороны компании-участника, помещающей ПДн в TEE, и со стороны других компаний-участников позволит расширить сферу использования TEE и повысить эффективность достижения описанных нами целей. При этом выпуск разъяснений регулятора об условиях использования законного интереса как основания обработки ПДн в рамках TEE позволит предотвратить потенциальные ситуации злоупотребления такой возможностью со стороны недобросовестных игроков.

Со стороны компаний — участников совместной обработки ПДн

При использовании регулятором консервативного подхода к квалификации действий компаний — участников совместной обработки по помещению ПДн в TEE в качестве передачи (предоставления доступа) ПДн всем другим компаниям-участникам, а также при сохранении текущего негативного отношения регулятора к законному интересу как основанию обработки ПДн, компаниям-участникам целесообразно использовать риск-ориентированный подход к использованию TEE.

С учетом приведенных в данном документе преимуществ TEE как технологии защищенной обработки данных, могут быть рассмотрены следующие меры для снижения рисков нарушения законодательства РФ о ПДн.

1. Использование фильтров Блума для сопоставления и вероятностного выявления совпадений данных компаний-участников на этапе подготовки датасетов для обработки.

Пример: кредитная организация (КО) с использованием TEE и фильтра Блума сопоставляет список своих клиентов со списком клиентов бюро кредитных историй (БКИ), чтобы выявить клиентов БКИ, которые могут одновременно являться клиентами данной КО. Ее интерес заключается в получении определенной информации от БКИ только по тем субъектам ПДн, которые одновременно являются клиентами и КО, и БКИ. Сопоставление списков клиентов является лишь одной из операций в этой модели и само по себе не гарантирует отсутствия передачи ПДн в целом. Но использование фильтра Блума, на наш взгляд, позволяет предотвратить передачу ПДн между КО и БКИ, так как сопоставление производится на основе не самих ПДн, а их «отпечатков», обратное преобразование которых в ПДн технически невозможно.

2. Внедрение модулей входного и выходного контроля исполняемых скриптов. Такие модули позволяют технически ограничить объем производимых компаниями-участниками операций с ПДн, помещенными в TEE, а также фиксировать заданные команды в системном журнале.

Пример: в приведенном сценарии взаимодействия КО с БКИ компании-участники могут договориться о том, что по выбранным с помощью фильтра Блума клиентам БКИ кредитная организация сможет получить лишь информацию, которая в достаточной степени агрегирована и не позволяет соотнести ее с каким-либо конкретным клиентом. Модули входного и выходного контроля в этой ситуации позволят технически проверять, соответствуют ли заданная команда и полученные результаты согласованным сторонами параметрам, а также проверять результаты вычислений на наличие в них персональных данных с использованием специальных алгоритмов типа k-anonymity и t-closeness. Следовательно, КО на выходе из TEE не будет получать данные, сформированные в результате анализа данных клиентов БКИ, которые можно было бы квалифицировать как персональные.

3. Обучение моделей машинного обучения на основе помещенных в TEE ПДн.

Пример: ПДн, помещенные компаниями-участниками в TEE, используются лишь для обучения моделей, поэтому на выходе из TEE компании-участники имеют лишь обученную на ПДн модель, но не сами ПДн.

Перспективы

Данные уже давно стали активом. Конфиденциальные вычисления позволяют управлять им — контролировать его использование и обеспечивать защиту.

Конфиденциальные вычисления открывают новые возможности для использования и развития ИИ, совместной работы с данными для решения глобальных задач и для всестороннего развития экономики на основе данных.

С помощью таких вычислений можно избежать компромисса между активным использованием данных в экономике, развитием рынка данных и «закручиванием гаек» в использовании данных для минимизации рисков. Конфиденциальные вычисления обеспечивают и то, и другое.

Об авторах доклада

АЛЕКСЕЙ НЕЙМАН
Исполнительный директор
Ассоциации больших данных.
Head of FIT Academy of Russia.
Master of Data Science.
CDMP, PMP

АЛЕКСЕЙ МУНТЯН
Генеральный директор Privacy Advocates.
Внешний Data Protection Officer
в нескольких транснациональных холдингах.
Соучредитель в Regional Privacy
Professionals Association - RPPA.pro
Сопредседатель Privacy & Legal Innovation
и кластера РАЭК

АРТЁМ АЛЕКСЕЕВ
Управляющий партнер
Aggregion

ДЕНИС БЕЗРУКОВ
CTO & co-founder of Aggregion.
Co-founder of SuperProtocol

АЛЕКСАНДР ПАРТИН
Адвокат, Партнер
Privacy Advocates.
Соучредитель РППА.Офис.
Сопредседатель
Privacy & Legal Innovation
кластера РАЭК.
CIPP/E, CIPM

ЮЛИЯ ПОСТОЛ
Юрист.
Специалист по праву
цифровых платформ,
персональным данным
и управлению ИИ.
Магистр НИУ ВШЭ

ВАЛЕРИЙ ХВАТОВ
Специалист по кибербезопасности
и распределенным вычислениям.
CTO DGT Network

Редакторы доклада

ЖАННА ПОКРОВСКАЯ
Руководитель пресс-службы
Ассоциации больших данных

АЛИЯ ТРУНАЕВА
PR-менеджер
Ассоциации больших данных

Ассоциация больших данных

Основная цель Ассоциации больших данных — создание условий для развития технологий и продуктов в сфере больших данных в России.

Сегодня членами АБД являются: Яндекс, VK, Сбербанк, Газпромбанк, Т-Банк, Россельхозбанк, МегаФон, Ростелеком, билайн, МТС, Аналитический центр при Правительстве РФ, Банк ВТБ, Авито и Центр стратегических разработок.

Подробнее об АБД и наших проектах можно узнать в Альманахе.

Privacy Advocates

С 2008 года помогаем клиентам в обеспечении комплаенса по персональным данным, проактивном управлении privacy-рисками и эффективном прохождении надзорных проверок, в том числе по утечкам данных. Специализируемся на аутсорсинге функций лица, ответственного за обработку персональных данных (DPO).

Основа нашей команды – компетентные и опытные privacy-специалисты различных компаний.

Aggregion

С 2019 года разрабатываем решения в области конфиденциальных вычислений и распределенной обработки данных.

Среди клиентов — крупнейшие телеком-операторы, банки, ритейлеры и экосистемы.

Основной продукт компании — платформа децентрализованных конфиденциальных вычислений. Платформа позволяет автоматизировать весь цикл работы с данными, делая совместную работу с данными и конфиденциальные вычисления доступными для широкого применения. Платформа получила престижные награды от Microsoft, Intel и других вендоров.

АССОЦИАЦИЯ БОЛЬШИХ ДАННЫХ

www.rubda.ru

Адрес: Москва,
Пресненская набережная, 10с2

+7 (495) 252-72-60
info@rubda.ru