УДК 006.72 **DOI** 10.15622/ia.22.5.4

Е.С. Новикова, Е.В. Федорченко, И.В. Котенко, И.И. Холод

АНАЛИТИЧЕСКИЙ ОБЗОР ПОДХОДОВ К ОБНАРУЖЕНИЮ ВТОРЖЕНИЙ, ОСНОВАННЫХ НА ФЕДЕРАТИВНОМ ОБУЧЕНИИ: ПРЕИМУЩЕСТВА ИСПОЛЬЗОВАНИЯ И ОТКРЫТЫЕ ЗАДАЧИ

Новикова Е.С., Федорченко Е.В., Котенко И.В., Холод И.И. **Аналитический обзор подходов к обнаружению вторжений, основанных на федеративном обучении: преимущества использования и открытые задачи.**

Аннотация. Для обеспечения точного и своевременного реагирования на различные типы атак, системы обнаружения вторжений собирают и анализируют большое количество данных, которые могут включать в том числе и информацию с ограниченным доступом, например, персональные данные или данные, представляющие коммерческую тайну. Следовательно, такие системы могут быть рассмотрены как источник рисков, связанных с обработкой конфиденциальной информации и нарушением ее безопасности. Применение парадигмы федеративного обучения для построения аналитических моделей обнаружения атак и аномалий может значительно снизить такие риски, поскольку данные, генерируемые локально, не передаются какой-либо третьей стороне, а обучение модели осуществляется локально – на источниках данных. Использование федеративного обучения для обнаружения вторжений позволяет решить проблему обучения на данных, которые принадлежат различным организациям, и которые в силу необходимости обеспечения защиты коммерческой или другой тайны, не могут быть выложены в открытый доступ. Таким образом, данный подход позволяет также расширить и разнообразить множество данных, на которых обучаются аналитические модели анализа, и повысить тем самым уровень детектируемости разнородных атак. Благодаря тому, что этот подход способен преодолеть вышеупомянутые проблемы, он активно используется для проектирования новых подходов к обнаружению вторжений и аномалий. Авторы исследуют существующие решения для обнаружения вторжений и аномалий на основе федеративного обучения, изучают их преимущества, а также формулируют открытые проблемы, связанные с его применением на практике. Особое внимание уделяется архитектуре предлагаемых систем, применяемым методам и моделям обнаружения вторжений, а также обсуждаются подходы к моделированию взаимодействия между множеством пользователей системы и распределению данных между ними. В заключении авторы формулируют открытые задачи, требующие решения для применения систем обнаружения вторжений, основанных на федеративном обучении, на практике.

Ключевые слова: обнаружение вторжений, аномалии, федеративное обучение, модели анализа, разделение данных.

1. Введение. Для своевременного и эффективного реагирования на различные информационные угрозы системы обнаружения вторжений (СОВ) и аномалий собирают и анализируют большие объемы данных. Большие объемы данных также требуются для обучения эффективной модели анализа, выполняющей выявление вторжений и аномалий. Зачастую такие данные могут включать различные типы конфиденциальной информации, включая персональные данные. В

Российской Федерации к персональным данным относятся данные, которые уникально определяют их владельца, а законодательства других стран расширяют это понятие, включая данные, относящиеся к устройствам, используемым человеком, такие как IP-адреса, уникальные идентификаторы устройств или приложений, а также данные о его местоположении. Таким образом, при построении систем обнаружения вторжений необходимо найти компромисс между конфиденциальностью собираемых и анализируемых данных и безопасностью субъекта персональных данных, и в большинстве случаев данная задача решается в пользу обеспечения безопасности, т.е. выполняется сбор, обработка и анализ всех данных, включая данные с ограниченным доступом. Парадигма федеративного обучения (ФО) позволяет решить эту проблему путем создания распределенных интеллектуальных систем, обеспечивающих конфиденциальность анализируемых данных [1, 2]. Суть ФО заключается в обучении аналитических моделей на наборах данных, которые находятся на разных источниках без обмена данными между ними. Более того, недавние исследования показали, что модели анализа, обученные в федеративном режиме, демонстрируют эффективность в обнаружении атак и аномалий, сравнимую с эффективностью моделей анализа, обученных классическим образом, т.е. на всем доступном наборе данных [3 – 5]. Тем не менее, ее применение связано с решением важных как практических, так и теоретических задач. К ним относятся следующие задачи [1, 6].

– Неоднородность устройств, которые генерируют анализируемые данные, что может привести к необходимости обработки различных форматов и атрибутов данных.

– Доступность устройств во время обучения модели анализа. Устройства, которые более стабильны и чаще доступны для обучения, могут оказывать более сильное влияние на результаты обучения.

– Распределение данных. Очевидно, что наборы данных, принадлежащие разным владельцам, могут иметь различные характеристики, в т.ч. иметь разное распределение меток. Такой случай распределения данных известен как случай зависимых, не идентично распределенных данных (not independent and identically distributed (non-IID) data). Он возникает в результате изменения и/или дрейфа концепций в наборах данных, а также может быть связан с разными объемами данных, хранящихся у разных владельцев.

– Настройка параметров системы ФО с учетом таких параметров как доступные вычислительные ресурсы клиентов, количество взаимодействующих клиентов, сложность обучаемой

модели анализа, которые определяют пропускную способность обучения.

Следует отметить, что, по мнению авторов, последняя задача тесно связана с проблемой доступных наборов данных, которые позволяют адекватно смоделировать как взаимодействие множества различных клиентов, так и различное распределение данных между ними, и оценить параметры системы с учетом этих факторов. В последнее время, проблема применения ФО для построения распределенных аналитических систем активно исследуется, и в научной литературе появилось множество исследований, посвященных различным теоретическим и прикладным аспектам ФО. Например, в [7 – 9] авторы исследуют задачу объединения (агрегирования) локальных моделей для формирования глобальной аналитической модели, устойчивой к неидентично распределенным данным. Проблемы построения систем ФО с учетом ограниченной пропускной способности сети клиентов обсуждаются в [10 – 12]. Например, Чжан и др. [12] решают задачу ускорения вычислений и снижения требований к вычислительным ресурсам, обусловленных использованием гомоморфного шифрования для дополнительной защиты передаваемых данных во время федеративного обучения. Другие аспекты безопасности и конфиденциальности ФО, такие как применение дифференциальной приватности, доверенных сред исполнения, изучаются в [13 – 17]. Существует также множество исследований, которые анализируют применимость ФО для решения различных практических задач. К настоящему времени исследователи предложили различные подходы на основе ФО для решения проблем цифрового здравоохранения [18, 19], безопасности [20 – 24], электронной коммерции [25 – 27], анализа текстов [28] и т.д.

Среди российских исследований следует отметить исследования, выполняемые под руководством И.И. Холода, которые посвящены разработке фреймворка федеративного обучения FL4J на языке программирования Java [29, 30].

Таким образом, несмотря на то, что ФО является относительно новой областью исследований, существует необходимость в систематизации разработанных подходов, посвященных различным аспектам ФО. В настоящей работе авторы исследуют существующие подходы к построению систем обнаружения вторжений на основе федеративного обучения, уделяя особое внимание используемым архитектурным решениям, применяемым моделям анализа, наборам данных и способам моделирования взаимодействия между множеством клиентов-владельцев данных, и схемам распределения

данных между ними. Следует отметить, что в силу того, что данное направление только начинает активно развиваться, основная мотивация данного исследования заключается в определении, какие решения для вышеупомянутых проблем уже предложены, насколько эффективно они решаются, и какие задачи предстоит еще решить. Сформулированные задачи и проблемы могут служить основой для более точной постановки целей и задач исследований, связанных с применением федеративного обучения для обнаружения вторжений и аномалий. Таким образом, основной вклад авторов заключается в:

– анализе архитектур федеративных систем обучения, включая поддерживаемые схемы распределения данных и требования к доступности клиентов и к их вычислительным ресурсам;

– анализе наборов данных, которые использовались для оценки системы, и подходов к моделированию федеративных систем;

– сравнительном анализе и систематизации предложенных подходов к обнаружению вторжений на основе федеративного обучения.

Статья организована следующим образом. В разделе 2 дано краткое описание концепции ФО и особенности систем, построенных на основе ФО. В разделе 3 обсуждаются типичные архитектурные решения СОВ. В разделе 4 представлена методология исследования и сравнительные критерии для анализа систем СОВ на основе ФО, в разделе представлена сравнительная характеристика выполненного аналитического обзора с другими обзорами по этой теме. В разделе 6 представлены результаты сравнительного анализа подходов к обнаружению вторжений на основе ФО. В конце статьи сформулированы основные преимущества использования федеративного обучения для обнаружения вторжений и задачи, которые еще предстоит решить.

2. Федеративное обучение. Ключевая идея ФО заключается в обучении локальных моделей непосредственно на клиентах, которые генерируют или владеют собственными данными, затем параметры локальных моделей объединяются для формирования глобальной модели, которая в процессе обучения рассылается всем взаимодействующим клиентам [1]. На рисунке 1 представлена схема федеративного обучения.

Таким образом, можно выделить три основных компонента систем, построенных на основе ФО: 1) клиенты, которые владеют данными и обучают локальную модель; 2) сервер, который координирует весь процесс обучения и вычисляет глобальную модель; 3) коммуникационно-вычислительная среда, которая обеспечивает

обмен параметрами модели. Исходя из этих компонент, можно выделить следующие ключевые характеристики аналитических систем, построенных с использованием ФО:

– схема взаимодействия между клиентами [18];

– вычислительные и сетевые ресурсы сотрудничающих владельцев данных;

– схема разделения данных между клиентами.

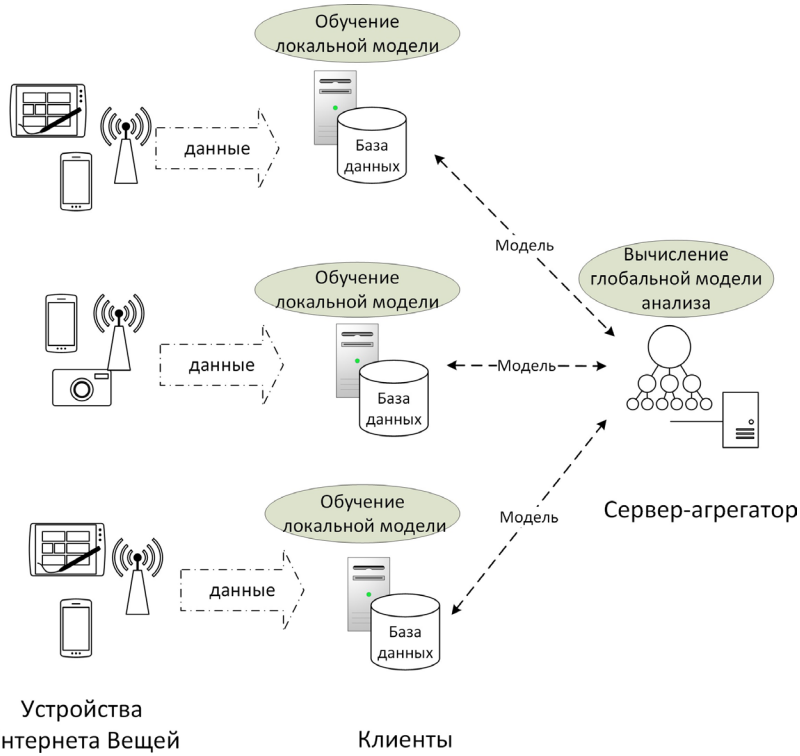

Рис. 1. Схема федеративного обучения

Схема взаимодействия между клиентами определяет, как организован процесс федеративного обучения. Различают централизованную и децентрализованную архитектуру системы ФО. В случае централизованной архитектуры один из взаимодействующих узлов выполняет роль агрегирующего сервера, который также координирует весь процесс обучения. Роль агрегирующего сервера может выполняться также некоторым доверенным лицом, который не

владеет данными. В случае децентрализованной системы ФО, функции агрегирующего сервера выполняются взаимодействующими клиентами [6]. Такая схема обучения еще известна как роевое обучение (swarm learning) [31]. На рисунке 2 показаны схемы взаимодействия во время процесса федеративного обучения.

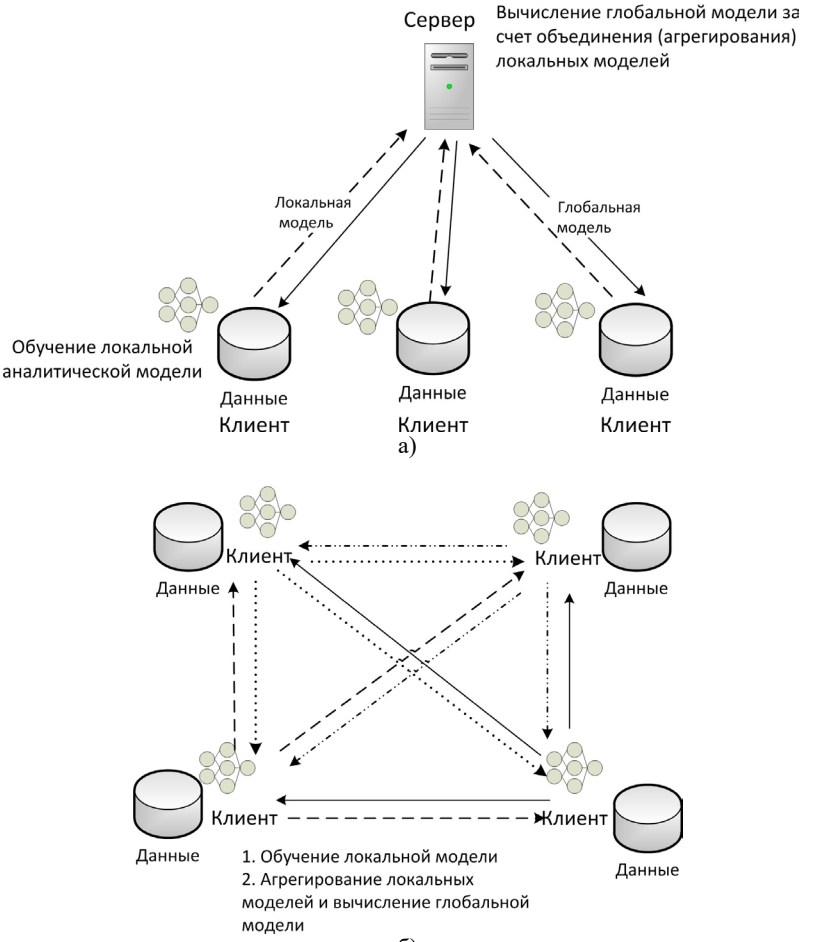

Рис. 2. Схемы взаимодействия между клиентами при ФО:
а) централизованная топология; б) децентрализованная топология

Необходимо отметить, что существуют также гибридные схемы коммуникации, которые представлены либо иерархией децентрализованных федераций, либо одноранговой сетью федераций с централизованной схемой коммуникации [18].

В зависимости от вычислительных ресурсов узлов-клиентов, их доступности во время процесса обучения, а также характеристик пропускной способности сети различают два типа объединения или федерации клиентов: федерация организаций (cross-silo) и федерация устройств (cross-device). Для федерации организаций характерно небольшое число клиентов, в роли которых обычно выступают организации и/или центры обработки данных. Для федерации устройств наоборот свойственно большое число клиентов с ограниченными вычислительными ресурсами, кроме того они могут появляться и отключаться в любой момент машинного обучения. Другой важной особенностью федеративного обучения является способ распределения данных. Семантически это понятие похоже на понятие «фрагментация данных», которое используется при организации физического хранения данных в распределенных хранилищах [32]. Обычно набор данных характеризуется двумя измерениями: 1) числом атрибутов и 2) числом записей в нем. Если клиенты имеют одинаковые наборы атрибутов, то данные разделены горизонтально. В случае вертикально распределенных данных клиенты имеют различные атрибуты данных для одного и того же набора образцов. Такой тип разбиения данных естественен для многих сценариев применения машинного обучения, например, оператор мобильной связи владеет данными о контактах человека, а финансовая организация может иметь информацию о его финансовом состоянии, и совместный анализ таких данных позволяет выявлять интересные схемы мошенничества. На рисунке 3 показаны схемы распределения данных между клиентами.

Кроме того, в реальных случаях данные могут быть разделены между клиентами частично вертикально, частично горизонтально − этот случай соответствует самому сложному типу распределения данных – гибридному. Описанные выше свойства систем ФО определяют различные сценарии использования систем ФО, включая также требования к производительности процесса обучения, выбору функции агрегирования и т.д.

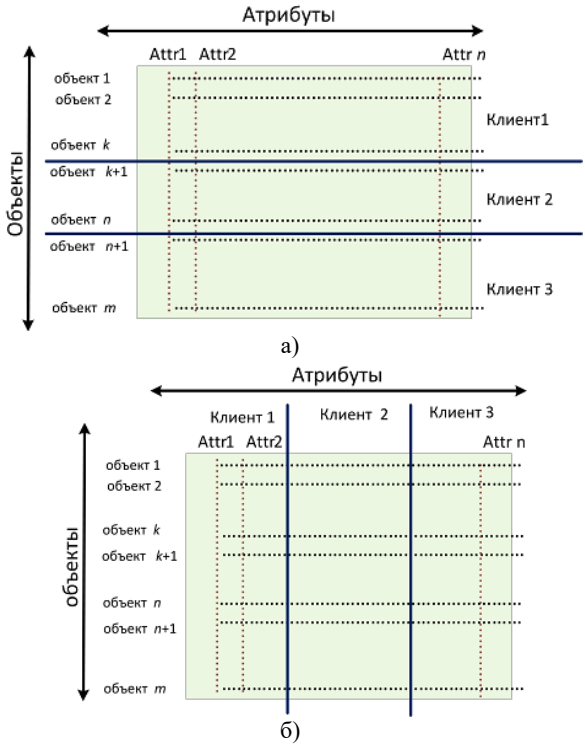

Рис. 3. Варианты распределения данных между клиентами: а) горизонтально распределённые данные; б) вертикально распределенные данные

3. Системы обнаружения вторжений. Система обнаружения вторжений (СОВ) является важной частью любой системы управления и обеспечения информационной безопасности и предназначена для обнаружения атак и аномалий в информационных системах. СОВ могут быть классифицированы в зависимости от типа анализируемых данных. В общем случае, принято выделять узловые и сетевые СОВ. Сетевая СОВ отслеживает и анализирует сетевой трафик, а узловая СОВ собирает и анализирует данные журналов операционной системы и приложений. Кроме того, существуют специализированные СОВ, разработанные для анализа данных определенных коммуникационных или других протоколов, а также гибридные решения, совмещающие анализ нескольких типов входных данных. Типичное архитектурное решение СОВ включает компоненты для сбора и обработки данных,

анализа, обнаружения атак (или аномалий) и реагирования на них (рисунок 4).

Базовые компоненты СОВ также включают репозиторий данных, в котором хранятся собранные исходные данные и сработавшие предупреждения, и базу знаний, содержащую информацию о правилах обнаружения атак, сигнатурах или шаблонах вредоносной активности. Существует два основных подхода к обнаружению атак или аномалий в информационной системе: сигнатурный и основанный на применении методов и моделей машинного обучения [33, 34]. Подходы на основе сигнатур обнаруживают атаки, используя методы сопоставления шаблонов для поиска известной атаки; поэтому база знаний об атаках должна постоянно обновляться для обеспечения высокого уровня обнаружения атак. Этот тип СОВ показывают высокую эффективность обнаружения атак известного типа; для обнаружения неизвестных атак или атак нулевого дня применяются подходы, основанные на методах и моделях машинного обучения (МО).

Рис. 4. Архитектура СОВ

В настоящее время исследователи предложили широкий спектр решений, использующих методы МО, включая адаптивную резонансную теорию [35], генетические алгоритмы [36, 37], методы кластеризации данных [38], нечеткой логики [36] и глубокие нейронные сети, такие как сверточные нейронные сети [39, 40], рекуррентные нейронные сети [41], глубокие автоэнкодеры [42] и т.д.

Дальнейшая классификация СОВ основана на том, как эти компоненты связаны и координируются в системе. Так, существуют монолитные, распределенные, иерархические и мультиагентные

системы [43, 44]. Распределенная СОВ предполагает, что каждый узел информационной системы имеет свою собственную СОВ, способную общаться с другими системами, кроме того имеется один выделенный сервер СОВ, который отвечает за окончательный анализ данных и принятие решений. В концепции иерархической СОВ локальные СОВ объединяются в кластеры, и каждый кластер имеет свой собственный головной узел, который отвечает за взаимодействие с другими кластерами [45]. Архитектура СОВ обычно выбирается исходя из типа контролируемой информационной системы и доступных вычислительных, энергетических ресурсов и пропускной способности сети. Например, типовой архитектурой СОВ для обнаружения вторжений в облаке является распределенная СОВ. Основной причиной такого выбора является необходимость анализа больших объемов сетевого трафика и ускорение вычислений на больших потоках данных. Для систем, построенных на основе технологии Интернета Вещей, примером которых служат системы «умного» дома, рекомендуемой архитектурой СОВ является распределенная иерархическая СОВ на основе агентов [39]. Это объясняется тем, что современные системы «умного» дома обычно поставляются вместе с набором облачных сервисов, предоставляемых производителем продукта. В этом случае программный агент СОВ устанавливается на домашнем маршрутизаторе с помощью специализированного программного обеспечения. Такой агент способен реализовывать различные функции, включая мониторинг потоков данных от датчиков, их предварительный анализ и пересылку результатов анализа головному компоненту СОВ, расположенному в облаке поставщика услуг. Типовая архитектура СОВ «умного» дома представлена на рисунке 5.

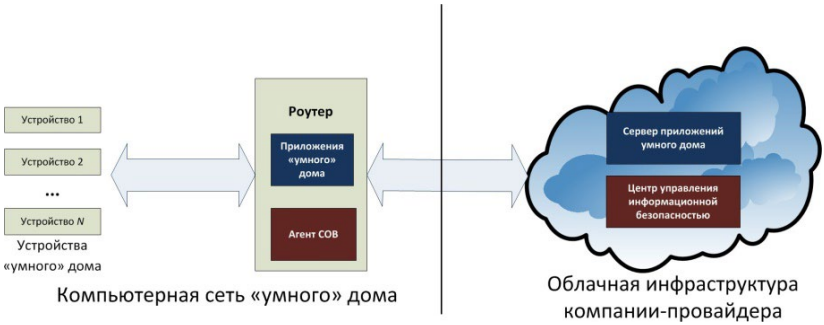

Рис. 5. Высокоуровневая архитектура иерархической распределенной СОВ для систем «умного» дома

Такая архитектура, с одной стороны, позволяет выбрать узел с более мощными ресурсами для установки СОВ, с другой стороны, поддерживает быстрый предварительный анализ локально генерируемых данных с последующим углубленным анализом, осуществляемым уже поставщиком услуг. Еще одним несомненным преимуществом такого подхода является возможность применения знаний об инцидентах безопасности в различных контролируемых системах «умного» дома. Единственным серьезным недостатком является значительный риск нарушения конфиденциальности пользователей «умного» дома из-за возможной утечки данных.

Иерархическая распределенная архитектура СОВ также предлагается для обнаружения атак и аномалий в киберфизических системах, таких как интеллектуальные энергетические сети [46 – 48]. Как правило, эти интеллектуальные сети состоят из различных взаимодействующих субъектов, включающих электростанции, конечных пользователей, представленных системами управления энергоснабжения «умного» дома, и юридических лиц, осуществляющих контроль за энергоснабжением. Иерархия локальных СОВ может быть построена либо на основе типа взаимодействующих субъектов [46, 47], либо на основе «границ вторжения», которые ограничивают распространение ущерба в случае успешной атаки [48].

Таким образом, можно выявить, что современные системы СОВ и аналитические системы, построенные на основе ФО, имеют два общих свойства, определяющих выбор архитектурного решения: это схема взаимодействия между клиентами (агентами) и требования к вычислительным ресурсам узлов с установленными интеллектуальными агентами. Архитектура децентрализованной системы ФО соотносится с распределенной СОВ с одноранговыми агентами, в то время как централизованная схема построения ФО близка к иерархической СОВ, в которой локальные агенты или кластеры одноранговых агентов связываются с главным компонентом СОВ для выработки окончательного решения. Характеристики облачной среды СОВ близки к вычислительным параметрам федерации организаций, когда взаимодействующие субъекты имеют достаточно вычислительных ресурсов и ресурсов хранения. СОВ для информационной системы на основе технологии Интернета Вещей должна учитывать те же ограничения, которые определяются характеристиками контролируемых устройств, что и система ФО для федерации устройств, т.е. энергопотребление, вычислительные ресурсы и ширину полосы пропускания. Для свойства, определяющего схему распределения данных в системах ФО, не существует

очевидного соответствия, хотя это свойство чрезвычайно важно при проектировании аналитической системы, основанной на принципах ФО. Большинство существующих систем ФО поддерживают горизонтально распределенные данные [49]. С точки зрения обнаружения вторжений и аномалий сетевые СОВ могут быть рассмотрены как клиенты с горизонтально распределенными данными, поскольку они обычно работают с определенным набором атрибутов, извлеченных из сетевого трафика, а СОВ для различных протоколов или приложений могут быть рассмотрены как случай вертикально распределенных данных, поскольку они работают с журналами различных приложений, имеющих разный формат и набор атрибутов. Аналогично, агенты СОВ в облачной среде должны иметь одинаковые наборы анализируемых атрибутов, в то время как агенты СОВ, развернутые в среде Интернета Вещей, должны быть способны обрабатывать различные признаки, характеризующие одни и те же объекты, поскольку они собирают данные с большого количества разнородных датчиков.

Основным преимуществом применения технологии ФО для построения СОВ является возможность снизить риски, связанные с обработкой конфиденциальных данных. Когда взаимодействующие субъекты представлены коммерческими организациями, критическими инфраструктурами, использование ФО позволяет повысить уровень доверия между ними и организацией, обеспечивающей информационную безопасность, поскольку в этом случае нет необходимости передавать конфиденциальные данные, и в то же время появляется возможность обмена знаниями об атаках и аномалиях с сохранением конфиденциальности. Для систем на основе Интернета Вещей применение ФО способно снизить объемы передаваемого сетевого трафика, что является важным фактором в условиях энергоэффективной сети, и обеспечивает возможность анализа большого количества разнородных устройств и датчиков.

Между тем применение федеративного обучения накладывает определенные требования к вычислительной мощности объекта, выполняющего локальное обучение модели, а также к объему памяти для хранения данных, необходимых для обучения. Поэтому при проектировании системы крайне важно оценить пропускную способность обучения системы анализа на основе ФО, характеризуемую через доступные вычислительные ресурсы клиентов, количество взаимодействующих клиентов, сложность обучаемой модели анализа. Необходимость рассмотрения этих вопросов определила основные цели данного исследования. Критерии сравнения

разработанных подходов к обнаружению вторжений на основе федеративного обучения и результаты исследования представлены в следующих разделах.

4. Методология поиска и анализа релевантных работ. Исследование и анализ СОВ на основе федеративного обучения выполнялось на основе рекомендаций по систематическому анализу научной литературы [50], которые предполагают определение 1) вопросов, решаемых в ходе исследования, 2) стратегии поиска и отбора научной литературы и 3) критериев включения и исключения работ в исследование. Ключевой задачей исследования является анализ подходов к обнаружению вторжений на основе ФО с возможной оценкой практической применимости ФО для решения задачи обнаружения вторжений, поэтому были сформулированы следующие вопросы исследования (ВИ).

ВИ1: Какая схема коммуникации для организации ФО – централизованная или децентрализованная – используется при построении СОВ?

ВИ2: Какая схема распределения данных – горизонтальная или вертикальная – учитывается при построении СОВ?

ВИ3: Какие наборы данных используются для тестирования предложенных схем распределения данных, каким образом происходит моделирование распределения данных между клиентами? Учитывается ли случай не идентично распределенных данных?

ВИ4: Какие методы и модели МО используются для обнаружения атак и/или аномалий?

ВИ5: Какие метрики используют авторы для оценки разработанных решений?

ВИ6: Какие программные библиотеки ФО используются для построения прототипов СОВ?

Эти вопросы также определили критерии оценки подходов к обнаружению вторжений, представленных в отобранных работах. Стратегия поиска научно-исследовательских работ была сформулирована как на основе исследовательских вопросов, так и в соответствии с рекомендациями [50]. Были проанализированы исследования, опубликованные в научных журналах и конференциях, не учитывались статьи в ненаучных журналах или коммерческие документы, презентации и слайды. Поиск осуществлялся как по англоязычным библиографическим системам, так и в русскоязычной научной электронной библиотеке eLibrary. Таким образом, для формирования множества исследуемых работ были выполнены следующие шаги.

Шаг 1. Формирование ключевых слов на русском и английском языках.

Шаг 2. Поиск публикаций на основе набора ключевых слов в электронных базах данных: eLibrary (поиск по ключевым словам на русском языке), IEEE Xplore, и ScienceDirect (поиск по ключевым словам на английском языке). Результатом этого шага является первоначальный набор публикаций.

Шаг 3. Проверка исходного набора публикаций на соответствие критериям включения и исключения. Эти критерии позволяют оценить, будет ли публикация включена в окончательную выборку для последующего рассмотрения. В качестве ключевых слов были определены следующие слова:

– на русском языке:

федеративное обучение AND (аномалии OR атаки OR вторжений) AND (компьютерные сети OR информационные системы),

– на английском языке:

federated learning AND (anomaly detection OR attack detection OR intrusion detection).

Были определены следующие критерии включения (КВ) и исключения (КИ):

КВ1. В работе четко описан подход к обнаружению аномалий и вторжений на основе федеративного обучения, обсуждается архитектура решения, модель анализа, описан сценарий эксперимента и используемые наборы данных, описана методика визуализации данных, т.е. даны исходные данные и способ построения графического представления.

КВ2. Публикация имеет четкую структуру. Методы представлены четко и наглядно.

КВ3. Публикация написана на русском (английском) языке с соблюдением стилистических и грамматических норм.

КИ1. В работе представлен обзор работ или сравнение методов.

КИ2. Представленный подход плохо описан, а публикация не имеет четкой структуры и/или изложена ненаучным языком. Общая схема этапа сбора исходных данных представлена на рисунке 6. На рисунке 7 представлена статистика публикаций, сгруппированных по их типу, за последние три года по базам IEEE Xplore и ScienceDirect. Следует отметить, что на русском языке работ по исследуемой тематике на момент выполнения исследования обнаружено не было. Это позволяет говорить о том, что данная тема исследований недостаточно хорошо изучена и исследована российскими учеными.

Рис. 6. Схема процесса отбора научных публикаций, посвященных проблеме применения федеративного обучения для задач обнаружения вторжений и аномалий. Все числовые значения даны на август 2022

5. Сравнение с другими обзорами. Есть несколько обзоров по федеративному обучению для обнаружения вторжений в Интернете вещей, среди которых следует отметить [51, 52], в которых рассматриваются теоретические проблемы и будущие направления исследований, связанных с применением ФО для обнаружения атак и аномалий. Аналитический обзор, представленный в статье [52], довольно обширен. Авторы проанализировали 15 исследовательских работ, связанных с СОВ на основе ФО, но авторы сосредоточены на исследовании особенностей предложенных подходов, не сравнивают их в контексте параметров, специфических для ФО, такие как архитектура ФО, модель МО, подход к разбиению набора данных для моделирования взаимодействия нескольких клиентов, функция агрегации, и т.д. В [51] авторы описали и сравнили 12 работ, связанных с применением ФО для повышения эффективности СОВ. Авторы рассмотрели не только предлагаемые подходы, но также представили используемые наборы данных, модели МО, и специфические настройки федеративного обучения, такие как число раундов агрегации, а также функция агрегирования. Однако данная информация представлена кратко, основной акцент сделан на разрабатываемый авторами подход. Таким образом, настоящий обзор является наиболее полным, в нем рассмотрено более 40 работ, кроме того, подробно исследованы вопросы, связанные с моделированием взаимодействия между устройствами (клиентами), и схемой распределения данных, рассматриваются подходы к моделированию неидентично распределенных данных.

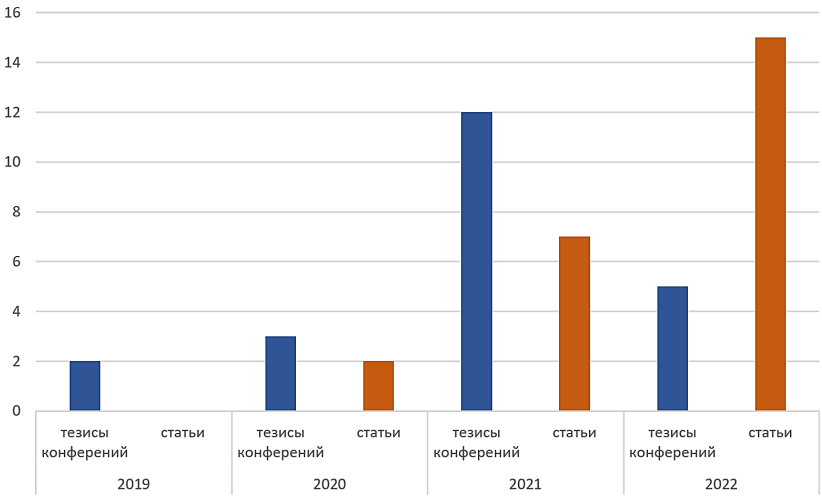

Рис. 7. Статистика публикаций, сгруппированных по их типу, за последние 3 года по базам IEEE Xplore и ScienceDirect

6. Подходы к обнаружению вторжений, построенных на принципах федеративного обучения. Обобщенные результаты выполненного анализа исследовательских работ представлены в таблицах 1–5. Результаты сгруппированы по предметным областям, для которых предложены СОВ на основе ФО: сетевая безопасность (таблица 1), безопасность IoT-устройств (таблица 2), медицинские устройства (таблица 3), промышленные киберфизические системы (таблица 4) и транспортные интеллектуальные системы (таблица 5). В таблице 6[1] представлен анализ особенностей настроек федеративного обучения, используемых для построения СОВ.

Таблица 1. Сравнительный анализ работ (сетевая безопасность)

Ref.	Модель анализа	Набор данных	Оцениваемые показатели
[53]	НС (сверточная НС)	набор данных, сгенерированный 20 участниками проекта LAN-Security Monitoring Project[2]	Точность (accuracy), полнота, точность (precision), F-мера
[54]	НС (LSTM НС)	улучшенный SEA [55] (набор команд сервера)	Точность (accuracy), полнота, точность (precision), F-мера, loss (значение функции потерь)

[1] Пустые ячейки в таблице говорят о том, что этот вопрос в работе не рассматривался.

[2] https://www.lan-security.net/

Продолжение Таблицы 1

Ref.	Модель анализа	Набор данных	Оцениваемые показатели
[56]	полносвязная НС	NSL-KDD [57]	Точность (accuracy), loss (значение функции потерь)
[58]	НС (автоэнкодер)	Aegean Wi-Fi Intrusion Dataset (AWID) [59]	Точность (accuracy); объем передаваемого трафика (Мб)
[60]	НС (сверточная НС)	CICIDS-2017 [61]	Точность (accuracy)
[62]		NSL-KDD [57]	Точность (accuracy), TPR, FPR; время обучения модели в федеративном режиме
[63]		N-BaIoT [64]	Точность (accuracy), полнота, точность (precision), F-мера; объем передаваемого сетевого трафика во время обучения
[65]		NSL-KDD [57]	Точность (accuracy), FPR; время обучения модели в федеративном режиме
[66]		UNSW-NB15 [67]; CICIDS-2018 [61]	Точность (accuracy), loss (значение функции потерь)
[68]		CIC-DDoS-2019 [69]	Точность (accuracy), полнота, точность (precision), F-мера
[70]		NSL-KDD [57]	Точность (accuracy); вознаграждение в обучении с подкреплением
[71]	НС	NSL-KDD [57]	Точность (accuracy), полнота, точность (precision), F-мера, ROC-AUC; время обучения модели в федеративном режиме
[72]	НС	TON-IoT-v2 [73], UNSW-NB15-v2 [67], BoT-IoT-v2 [74], CSECIC-IDS2018-v2 [61]	Точность (accuracy), полнота, точность (precision), F-мера, FPR
[75]		NSL-KDD [57]	Точность (accuracy)
[76]	ансамбль 4 НС (LSTM НС с блоками GRU с разным размером окна)	сетевые данные протокола Modbus [77]	Точность (accuracy), полнота, точность (precision), F-мера; время обучения модели в федеративном режиме
[78]	логистическая регрессия	TON-IoT [73]	Точность (accuracy); время на генерацию шума в механизме дифференциальной приватности
[79]	НС с блоками GRU и SVM классификатором в качестве выходного слоя	KDD CUP99 [80]; CICIDS2017 [61]; WSN-DS[3]	Точность (accuracy), FPR, F-мера; число раундов агтрегирования для оценки скорости обучения модели
[81]	Градиентный бустинг на деревья решений (GBDT)	CIC-DDoS-2019 [69]	Точность (accuracy), FNR; теоретическая оценка объема предаваемой информации
[82]	GAN НС + сверточная НС	NSL-KDD [57], KDD CUP99 [80], UNSW-NB15 [67]	Точность (accuracy), полнота, точность (precision), F-мера, потери (loss), AUC, скорость сходимости

[3] https://www.kaggle.com/datasets/bassamkasasbeh1/wsnds

Таблица 2. Сравнительный анализ работ (IoT устройства)

Ref.	Модель анализа	Набор данных	Оцениваемые показатели
[83]	НС (автоэнкодер)	UNSW-NB15 [67]	F-мера
[84]		NSL-KDD [57]	Точность (accuracy)
[85]	НС (рекуррентная нейронная сеть с блоками GRU)	собственный набор данных от IoT устройств (33 устройства), атаки выполнялись путем заражения ботнетом Mirai	FPR, TPR; время обучения модели и время классификации объекта
[86]	НС	CICIDS-2017 [61], CICDDoS-2019 [69]	FPR, FNR
[87]	НС (сверточная НС)	CICIDS-2017 [61]; NSL-KDD [57]; набор данных от IoT устройств [88]	Точность (accuracy), TPR, FPR
[89]	НС (полносвязная НС)	NSL-KDD [57]	Точность (accuracy), полнота, точность (precision), F-мера
[90]	НС (автоэнкодер)	N-BaIoT [64] – сетевой трафик от 9 устройств, атаки выполнялись путем заражения BASHLITE и ботнетом Mirai	Точность (accuracy), полнота, точность (precision), F-мера
[22]	НС (полносвязная сеть и автоэнкодер)		Точность (accuracy), полнота, точность (precision), F-мера; теоретические оценки генерируемого трафика и вычислительной нагрузки во время обучения

Таблица 3. Сравнительный анализ работ (медицинские устройства)

Ref.	Модель анализа	Набор данных	Оцениваемые показатели
[91]	иерархическая LSTM НС	NSL-KDD [57]; TON-IoT [73]	Точность (accuracy), полнота, точность (precision), F-мера
[92]	НС (сверточная)	набор данных, сгенерированный с использованием симулятора реакции на глюкозу [93]	Точность (accuracy), полнота, точность (precision), F-мера
[94]	GAN-сеть	Набор медицинских данных CHARIS [95]; UNSW-NB [67]	Точность (accuracy), полнота, точность (precision), F-мера, ROC-AUC

Таблица 4. Сравнительный анализ работ (промышленные КФС)

Ref.	Модель анализа	Набор данных	Оцениваемые показатели
[96]	НС (сверточная нейронная сеть с блоками GRU)	набор данных от системы газоснабжения [97]	Точность (accuracy), полнота, точность (precision), F-мера
[98]	НС (автоэнкодер)		Точность (accuracy), полнота, точность (precision)
[99]	НС (автоэнкодер), Трансформер, и преобразование Фурье	набор данных от системы газоснабжения [100]; SWaT [101]; HAI [102]; измерения потребляемой мощности [103]; измерения сердцебиения [103]; измерения частоты дыхания пациента [103]; координаты правой руки при выполнении различных действий [103]; измерения тока для космического шаттла [103]; информация о пассажирах Нью-Йоркского такси [104]	Полнота, точность (precision), F-мера; использование памяти, использование графического процессора, время выполнения, пропускная способность обучения, потребляемая мощность

Таблица 5. Сравнительный анализ работ (интеллектуальные транспортные системы)

Ref.	Модель анализа	Набор данных	Оцениваемые показатели
[105]	НС	KDD Cup99 [80]	Точность (accuracy), полнота, точность (precision); метрики, характеризующие временные затраты на применение технологии блокчейн, в частности, время генерации блока блокчейна
[106]	НС (трансформер)	TON-IoT [73]; Набор данных ль взлома автомобилей	Полнота, точность (precision), F-мера; время обучения модели в федеративном режиме
[107]	случайный лес	OTIDS[4]	Точность (accuracy), полнота, точность (precision), F-мера; время обучения модели в федеративном режиме; загрузка ЦПУ и память (RAM)

[4] https://ocslab.hksecurity.net/Dataset/CAN-intrusion-dataset

Таблица 6. Анализ настроек ФО при построение СОВ

Ref.	Алгоритм агрегирования	Наличие тестбеда/ используемый фреймворк ФО	Другие особенности СОВ
[20,58,60]	FedAvg		
[53, 84]	FedAvg		
[87]	FedAvg		использование трансферного обучения (transfer learning)
[81, 86]	FedAvg		
[108]	FedAvg		гомоморфное шифрование (схема Пейе)
[68]	FedAvg	да/Flower FL	дифференциальная приватность
[83,89,98]	FedAvg		
[56]	FedSGD, FedAVG		
[70]	FedSGD	да/PySyft	использование трансферного обучения (transfer learning); устойчивость к data poisoning атакам
[78]	Fed+		дифференциальная приватность
[72]	Fed+, CM+		
[79]	FedAGRU		оценка устойчивости СОВ к атакам на изменение меток обучающего набора
[22]	FedAVG (агрегирование каждую итерацию и агрегирование каждые n эпох), FED CM (на основе координатно-медианного градиентного спуска)		оценка устойчивости СОВ к атакам на изменение меток обучающего набора
[63, 71] [62,65,66, 75,91,107]			
[76, 106]		да/PySyft FL	
[94]		да/Flower FL	
[90]		да/PySyft FL	
[92]	FedAVG	свой тестбед	применение для выявления аномалий
[54]	FedAvg		набор данных журналов, содержащих команды сервера
[99]	FedAvg [109]	FedML	применение для выявления аномалий
[82]		свой тестбед	использование трансферного обучения (transfer learning)

6.1. Архитектурные решения по построению СОВ на основе федеративного обучения. Наиболее часто используемым архитектурным решением ФО для построения СОВ является централизованная схема обучения, в которой глобальная модель формируется отдельным выделенным доверенным сервером. Данная схема используется для построения иерархических распределенных СОВ в системах Интернета вещей и СОВ, развернутых в облачных средах [58, 60, 62, 63, 65, 66, 68, 70, 83 – 87, 89, 91]. Типовая архитектура СОВ на основе централизованного ФО представлена на рисунке 8. В ней ряд узлов осуществляют сбор данных от устройств и выполняют обучения локальной модели анализа для выявления атак и/или аномалий, далее локально обученные модели отсылаются выделенному серверу безопасности, который формирует глобальную модель, агрегируя локальные. Данный процесс осуществляется итеративно, поэтому необходимо учитывать пропускную способность канала, связывающего клиентов и сервер безопасности. На рисунке 8 в качестве клиента выступает платформа приложений мобильных граничных вычислений (MEC, Mobile Edge Computing), которая собирает данные от одной подсети IoT устройств [87, 107], однако клиентами могут быть шлюзы безопасности локальных сетей, как в [20], а также сами устройства [58, 83, 98], однако в последнем случае следует учитывать вычислительные ресурсы, которыми они обладают.

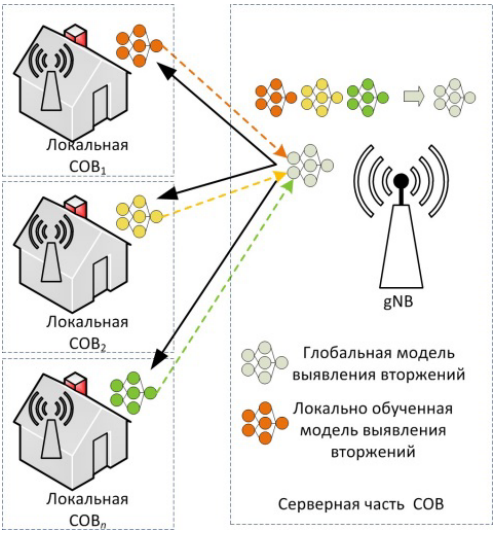

Рис. 8. Распределенная СОВ с централизованной схемой взаимодействия между клиентами [67]

Для масштабных, географически распределенных вычислительных сетей предложена иерархическая архитектура СОВ, в которой вычисление глобальной модели анализа, используемой для выявления аномалий, также формируется иерархически на уровне сегментов [53, 91]. Все участники процесса объединяются в группы или сегменты, и каждая группа выполняет обучение некоторой промежуточной модели в федеративном режиме обучения, после чего глобальная модель вычисляется путем агрегирования таких промежуточных моделей. При этом, на каждом этапе формирования глобальной модели, происходит оценка того, насколько веса локальной модели каждого участника отличаются от весов глобальной модели, и если отличия в весах моделей превышают некоторый заданный порог, то такой клиент исключает из группы [79]. Такое решение позволяет повысить устойчивость моделей выявления атак, обучаемых на несбалансированных наборах данных [53, 79]. Децентрализованная схема ФО предложена для построения СОВ, разрабатываемых для интеллектуальных транспортных систем [105 – 107]. Данное решение обусловлено, в первую очередь, географической распределенностью таких систем, и высокой мобильностью транспортных средств, которые являются неотъемлемыми компонентами таких систем. Транспортное средство, перемещаясь в пространстве, подключается к разным базовым станциям или интеллектуальным придорожным устройствам, получая таким образом актуальную дорожную информацию и обновленную модель обнаружения атак и аномалий. Типовая архитектура СОВ в этом случае имеет двухуровневую систему: на нижнем уровне транспортные средства загружают от базовых станций модели анализа, используемые в СОВ, и обновляют их с учетом собираемых ими данных, после чего отправляют их обратно базовым станциям. Базовые станции получают локальные модели от подключенных к ним устройств, проверяют их корректность, и участвуют в формировании новой глобальной модели. Таким образом, на верхнем уровне, центральный агрегирующий сервер безопасности замещен множеством распределенных взаимодействующих базовых станций. Такое решение позволяет повысить устойчивость СОВ к вредоносным действиям, направленным на нарушение функционирования центрального узла и снизить риски утечки данных, в т. ч. персональных. На рисунке 9 представлена схема построения СОВ на основе децентрализованного ФО.

Несколько иной подход к построению СОВ на основе ФО для самоуправляемых транспортных систем предложен в [107]. Авторы

предложили передать функции агрегирования локальных моделей на уровень конечных узлов, т.е. транспортным средствам, а сбор данных и обучение локальных моделей осуществлять на уровне базовых станций, т.к. это позволяет снизить вычислительную нагрузку на граничные устройства, поскольку операция агрегирования значительно менее ресурсоемкая, чем процесс локального обучения модели анализа. В обоих случаях для обеспечения целостности и неизменности локальных и глобальных моделей в условиях распределенных вычислений применяются технологии блокчейна. Использование блокчейна с одной стороны решает задачи, связанные с верификацией и проверкой аутентичности моделей анализа, с другой стороны, порождает новые задачи, обусловленные применением ресурсоемких операций, специфичных для блокчейна, таких как шифрование, дешифрация, генерация ключей, выработка консенсуса и т.д.

Рис. 9. Распределенная схема СОВ на основе децентрализованного ФО [105]

6.2. Схемы разделения данных между клиентами СОВ и используемые наборы данных. Анализ работ показал, что представленные в литературе СОВ поддерживают только горизонтальную схему распределения данных, т.е. все клиенты обладают одинаковым набором атрибутов. Данное решение выглядит естественным с учетом того, что в большинстве случаев входные данные представлены сетевым трафиком или статистическими характеристиками, сформированными на основе анализа сетевых потоков. В качестве наборов данных в основном используются такие

наборы как CICIDS2017 [61], NSL-KDD [57], горизонтальное разделение данных моделируется путем группировки пакетов по IP адресам источников сетевых потоков или случайным образом. В работах [96, 107] в качестве входных данных используются данные специализированных коммуникационных протоколов, таких как ModBus, CAN. В них также формируется единый для всех участников набор атрибутов. Например, в [107] применяется частотный анализ сообщений CAN-шины с последующим применением преобразования Фурье для формирования множества анализируемых признаков. В [71, 85] проблему формирования общего набора атрибутов для разнородных устройств предлагается решать путем определения множества атрибутов для каждого типа устройства, таким образом, модель анализа обучается в федеративном режиме на «горизонтальных» данных для каждого типа устройства, а в основе СОВ лежит ансамбль таких моделей. Также было показано, что такое решение позволяет значительно снизить уровень ложно положительных срабатываний [71]. В [84] предлагается выполнять обучение модели в федеративном режиме для выявления определенного типа атак, что предполагает определение набора анализируемых атрибутов для каждого типа атаки, который является одинаковым для всех клиентов. Такое решение обеспечивает горизонтальное разделение данных между клиентами.

В [99], несмотря на то, что авторы используют достаточно разнообразные наборы данных – сетевые данные, данные от датчиков системы очистных сооружений [101] и т.д., схема разделения данных является горизонтальной: для моделирования взаимодействия между множеством сторон набор данных последовательно делится на несколько частей, что позволяет сохранить логическую структуру временных рядов.

Случай вертикального распределения данных среди взаимодействующих клиентов для построения СОВ практически не изучен. Исключение составляет работа [4], в которой предпринята попытка моделирования данной схемы разделения данных. Авторы использовали набор данных SWAT [101], который содержит данные от шести различных технологических процессов, описываемых разными наборами сенсоров и актуаторов. Для моделирования вертикального распределения данных между клиентами, он был поделен по процессам. Обучение глобальной модели анализа для выявления атак осуществлялось с помощью специализированного фреймворка FATE [110], и, несмотря на полученные высокие показатели точности обнаружения атак, авторы продемонстрировали, что текущая

реализация схемы федеративного обучения не может быть использована для оперативного выявления вторжений в силу высоких требований к вычислительным ресурсам и времени, необходимому как для обучения такой модели, так и для классификации данных, подаваемых ей на вход.

6.3. Моделирование неидентично распределенных данных. Влияние неидентично распределенных данных исследуется довольно часто, и можно выделить два основных способа моделирования такого распределения данных. В первом случае один набор данных делится между клиентами, и каждый клиент получает определенный тип атак (или несколько типов атак) [63, 71]. Например, для моделирования взаимодействия 8 устройств с неидентично распределенными данными набора NSL-KDD был разделен следующим образом:

– «нормальный» трафик был поделен на 8 частей;

– трафик с атаками был разбит по типу атак, а затем каждое полученное подмножество записей было разделено между двумя клиентами, т.е. устройства № 0 и № 1 имели данные по атаке на отказ в обслуживании (DoS атаке), устройства № 2 и № 3 – по атакам типа Probe, устройства № 4 и № 5 – по атакам типа R2L, и устройства № 6 и № 7 – по атакам U2R.

В случае, когда речь идет о выявлении аномалий, тип атаки не учитывается, и неиндентично распределенные данные моделируются, варьируя процентное содержание нормальных или аномальных данных в наборе данных одного устройства. Например, в [70] для моделирования неидентично распределенных данных «нормальный» трафик был распределен по 10 устройствам в процентном отношении следующим образом: 25%, 50%, 75%, 25%, 50%, 75%, 25%, 50%, 75%, 50%, соответственно, тип атаки не учитывался.

Во втором случае используются разные наборы данных с одинаковыми параметрами, и каждый набор играет роль данных одного устройства [72, 87]. Например, в [72] используются 4 разных набора данных – ToNIoT-v2 [73], UNSW-NB15-v2 [67], BoT-IoT-v2 [74], CSECIC-IDS2018-v2 [61], для моделирования взаимодействия четырех шлюзов безопасности, установленных в разных беспроводных сетях. Для выравнивания множества атрибутов из наборов данных были извлечены все признаки, связанные с сетевыми потоками. Каждый набор данных характеризуется разным составом атак, кроме того, различна их доля присутствия в обучающих выборках. Было показано, что ФО позволяет достичь достаточно высоких показателей обнаружения атак (минимальное значение точности (precision) на тестовых наборах данных – 90.20%, а максимальное – 99.98%) при

сравнительно низком уровне ложно положительных срабатываний (максимальное значение этого показателя на одном из наборов равно 5.38%, а минимальное – 0.04%). Таким образом, этот же подход к подготовке обучающих выборок взаимодействующих устройств может быть использован для оценки обобщающей способности СОВ, построенных на ФО, т.е. способности обнаруживать новые виды атак, которых нет в исходном обучающем наборе отдельного участника (клиента) такой системы. Однако в общем случае точность обнаружения атак зависит от репрезентативности таких атак в обучающих наборах данных [98].

Проблема оценки обобщающей способности аналитической модели, обученной федеративном режиме, представлена не достаточно полно. На рисунке 10 представлено распределение работ с учетом выполненной оценки способности аналитической модели, обученной в федеративном режиме, выявлять новые типы атак, которые отсутствуют в обучающей выборке клиента.

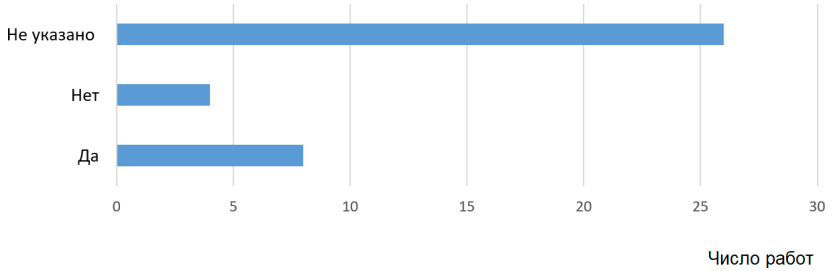

Рис. 10. Доля работ, в которых выполнялась оценка способности модели, обученной в федеративном режиме, к выявлению новых типов атак, которых нет в обучающей выборке клиента

6.4. Методы машинного обучения, используемых в СОВ на основе ФО. В основном для обнаружения атак и аномалий в СОВ на основе ФО используются глубокие нейронные сети: полносвязные, сверточные сети, рекуррентные сети, а также сети долгой краткосрочной памяти (LSTM-сети). Методы классического машинного обучения, такие как деревья решений, логистическая регрессия, методы опорных векторов практически не используются. Этот факт в первую очередь объясняется тем, что само ФО было предложено для обучения глубоких нейронных сетей, и функции агрегирования локальных моделей в основном разработаны для таких типов моделей анализа. Для выявления аномалий в основном применяются нейронные сети автоэнкодеры [83, 98].

Предложены интересные решения по использованию ФО в сочетании с трансферным обучением, т.е. использованием предобученных моделей. Ключевой идеей использования трансферного обучения является применение знаний, полученных в одной области к новой предметной области, для ускорения процесса обучения модели анализа, повышения ее точности и снижения вычислительных ресурсов. Трансферное федеративное обучение позволяет ускорить процесс обучения локальных моделей. Это достигается путем обучения исходной модели на открытом наборе данных, и использовании ее в качестве исходной для обучения локальных моделей [70, 87]. Например, в [70] исходная сверточная сеть обучается на наборе NSL-KDD [57], затем она передается в качестве предобученной клиентам, и во время локального обучения выполняется подстройка только последнего полносвязного слоя на другом наборе данных UNSW-NB15 [67].

Следует также отметить, что для формирования глобальной модели чаще всего используется алгоритм агрегирования FedSGD, который предполагает агрегирование параметров локальной модели в конце каждой эпохи локального обучения, и FedAvg, отличающийся тем, что он позволяет задать количество эпох локального обучения, по завершении которых выполняется вычисление параметров глобальной модели. Алгоритм FedAvg является основным способом формирования моделей анализа в СОВ на основе ФО, поскольку является более эффективным с точки зрения сетевого взаимодействия между клиентами. Между тем показано, что и FedSGD, и FedAvg плохо обрабатывают разнородные и неидентично распределенные данные: эффективность моделей анализа, полученных с их помощью, значительно снижается при обучении на таких данных [72].

6.5. Метрики оценки разработанных решений. Для оценки разработанных решений в основном используются метрики, связанные с оценкой эффективности моделей анализа, такие как доля верно классифицированных объектов, доля ложно положительных срабатываний, чувствительность классификатора и т.д. В ряде работ выполняется оценка времени обучения модели в централизованном и федеративном режиме, и авторы исследований показывают, что время обучения модели в федеративном режиме значительно снижается, что в условиях ограниченности энергетических ресурсов IoT устройств выглядит достаточно привлекательно. Однако практически ни в одной работе не уточняются условия проведения данного эксперимента, в частности, не указывается размер обучающей выборки, используемой при централизованном обучении, и размеры данных, которыми

владеют взаимодействующие участники при федеративном обучении; а также режим применения ФО. Очевидно, что время обучения моделей анализа зависит в том числе и от размера обучающей выборки, и если выборка данных, находящаяся на устройстве при федеративном обучении, является частью исходной, то можно предположить, что длительность обучения при прочих равных настройках может быть значительно меньше, а при равных размерах обучающих выборок может оказаться даже больше за счет необходимости синхронизации клиентов и передачи данных в процессе обучения. Последний параметр сильно зависит от того, в каком режиме развернуто ФО. Оно может быть использовано в симуляционном или реальном федеративном режиме. В первом случае вся система, включая взаимодействующие узлы, развертывается на одном вычислительном узле, сетевое взаимодействие практически отсутствует, такой режим используется для выбора и настройки параметров ФО и аналитической модели. В федеративном режиме узлы разворачиваются на нескольких физических или виртуальных узлах, в этом случае имеет место настоящее сетевое взаимодействие. Между тем, разница в пропускной способности обучения в федеративном и симуляционном режимах может быть значительной в зависимости от типа модели анализа и настроек функции агрегирования [49]. Только в незначительной части работ [22, 99, 107] авторы исследуют другие параметры ФО, такие как загрузка ЦПУ, объем передаваемого сетевого трафика, объем используемой оперативной памяти, при этом в [22] данные параметры оцениваются в контексте применения технологии блокчейна, в частности исследуются параметры механизма консенсуса при генерации нового блока, а в [107] даны теоретические оценки ожидаемого сетевого трафика и загрузки ЦПУ во время обучения. Между тем, большая часть исследовательских работ позиционируют ФО как способ построения СОВ именно в сетях IoT-устройств [22, 83 – 87], которые могут характеризоваться ограниченными вычислительными и энергетическими ресурсами, низкой полосой пропускания канала связи, поэтому исследование таких параметров является критически важным при определении архитектуры СОВ, параметров ФО, и непосредственно модели анализа. Исследование перечисленных параметров может быть выполнено при развертывании СОВ на основе ФО на экспериментальном стенде состоящем как из виртуальных, так и физических устройств.

6.6. Использование программных средств и фреймворков для построения экспериментальных стендов. В большей части

работ данные по развертыванию и использованию экспериментальных стендов отсутствуют. В [68, 76, 90, 94, 106] используются специализированные библиотеки для построения ФО, в частности PySyft, Flower. Согласно [49], PySyft v0.6 и ниже не поддерживается реальный федеративный режим, а последующие версии данного фреймворка реализуют несколько иную концепцию распределенного обучения, в которой сущности, имеющие роль Data Analyst (аналитика данных), удаленно обучают модель на данных, принадлежащих другой сущности – владельцу данных (Data owner), таким образом, взаимодействие между несколькими сущностями, владеющими данными, практически отсутствует. Библиотека Flower поддерживает и симуляционный и федеративный режимы обучения, а также предоставляет достаточно широкий спектр различных функций агрегирования, что делает ее использование предпочтительным при тестировании подходов к обнаружению вторжений и аномалий на основе ФО.

6.7. Приватность данных и устойчивость СОВ к атакам на изменение разметки. Федеративное обучение, как и классическое машинное обучение, уязвимо к ряду атак: состязательным атакам, связанным с изменением функциональности обучаемой модели анализа, и атакам на логический вывод, целью которых является получение информации об используемых наборах данных и/или их свойствах. Однако в отличие от централизованной схемы обучения, когда все данные собираются и контролируются одним субъектом, в случае федеративного обучения данными и их качеством управляют их владельцы, таким образом, расширяется поверхность атаки, и у злоумышленника появляется больше возможностей для ее успешного проведения. С учетом этого, вопросы, связанные с уязвимостями машинного обучения, в СОВ, построенных на федеративном обучении, приобретают особую актуальность. Вместе с тем систематический анализ литературы показал, что данный вопрос практически не исследован. Наиболее полное исследование представлено в [22], авторы изучили влияние нескольких типов атак, в т.ч. подмену меток обучающей выборки ("отравление" данных), изменение градиентов локальных моделей ("отравление" модели) и показали, что без использования функций агрегирования, устойчивых к неидентично распределенным данным, достаточно одного атакующего клиента в федерации, чтобы нарушить сходимость глобальной модели.

Вопросы приватности наборов данных также практически не исследованы, в работах [68, 78] выполнены оценки влияния на точность глобальной модели механизмов дифференциальной приватности,

авторы показали, что снижение точности глобальной модели при добавлении шума к градиентам локальной модели не значительно, тем не менее не выработаны единые рекомендации по выбору параметров дифференциальной приватности, которая бы обеспечивала заданный уровень точности глобальной модели анализа при допустимой вероятности компрометации обучающих наборов. В [108] проблема конфиденциальности данных решается путем применения гомоморфного шифрования (схема Пейе), эксперименты показывают, что в этом случае не происходит потери точности глобальной модели, однако авторы не указывают влияние шифрования на время ее формирования. Вместе с тем в [16] экспериментально показано, что время обучения и объем сетевого трафика сильно зависят от протокола шифрования, сетевой трафик при обучении полносвязной нейронной сети на наборе данных, состоящем из 10000 строк-векторов, и размере батча, равного 100, может достигать от 1.78 ГБ до 36 ГБ.

7. Выводы: преимущества использования и открытые задачи. Проведенные исследования показали, что федеративное обучение может быть успешно использовано для построения распределенных систем обнаружения вторжений, которые обладают несколькими важными свойствами. Во-первых, такие системы позволяют обрабатывать данные с ограниченным доступом, например, персональные данные и/или конфиденциальные данные. К таким данным также относятся данные от критических инфраструктур, в т. ч. сетевой трафик и данные от технологических процессов, и применение ФО позволяет настраивать модели обнаружения вторжений и аномалий на таких наборах без компрометации их конфиденциальности, стимулируя тем самым взаимодействия между различными организациями и киберфизическими объектами. Особенно перспективным видятся решения по построению СОВ, сочетающие ФО с методами трансферного обучения.

Во-вторых, модели выявления аномалий и/или атак, обученные в федеративном режиме на нескольких наборах данных, которые содержат разные типы атак, обладают более высоким уровнем детектирования ранее неизвестных атак по сравнению с моделями, обученными на одном наборе данных. Таким образом, такие модели обладают более высокой обобщающей способностью, формируемой за счет расширения обучающей выборки.

В-третьих, возможность построения децентрализованной СОВ на основе ФО позволяет решить проблему нарушения работоспособности центрального узла, управляющего процессом обнаружения вторжения и/или аномалий, включая переобучение

соответствующих моделей анализа. Данная проблема известна как единая точка отказа (single point of failure), для которой характерно наличие одного критического компонента, выход из строя которого приводит к нарушению функционирования всей системы. В случае же ФО, способности системы к обнаружению вторжений сохраняются, они лишь ограничены возможностями локальных моделей анализа. Основной открытой проблемой является отсутствие подходов к построению ФО, позволяющих эффективно выполнять обучение на вертикально распределенных данных. Такой тип распределения данных характерен для киберфизических систем, в которых объекты представлены неоднородными наборами датчиков, и соответственно, их поведение описывается различными атрибутами. Представленные решения по обнаружению вторжений на основе ФО предложены только для горизонтально разделенных данных, а именно для анализа сетевых данных; соответственно, задача выявления аномалий в технологических процессах методами федеративного обучения остается нерешенной.

Также следует отметить, что в большинстве работ для формирования общей глобальной аналитической модели используются два классических подхода к агрегированию локальных моделей – это алгоритмы FedSGD и FedAvg. Вместе с тем в [22, 72] было показано, что разные алгоритмы агрегирования могут оказывать значительное влияние на обобщающую способность глобальных аналитических моделей. Таким образом, сценарии экспериментальной оценки ФО для построения СОВ должны также включать анализ различных алгоритмов агрегирования, в т. ч. тех, которые доказуемо устойчивы к неидентично распределенным данным, например, Fed+ и FEDMO [111].

Из всего вышесказанного можно сделать вывод, что существует назревшая необходимость в создании и стандартизации методологии оценки СОВ, построенных на основе принципов ФО, которая будет определять требования как к наборам данных для тестирования, их распределению (в том числе эксперименты с неидентично распределенными данными), оцениваемым метрикам, так и учитывать характеристики анализируемой СОВ, включая различные алгоритмы агрегирования, архитектуру и устойчивость к разного рода атакам, и характеристики среды проведения экспериментов. Также остается ряд открытых проблем, большей частью связанных с практическими аспектами применения ФО, в частности, остаются открытыми вопросы, связанные с определением требований к вычислительным ресурсам, пропускной способности канала связи, что особенно важно для систем, построенных на основе технологии Интернета Вещей.

Фактически, текущий подход к анализу применимости ФО к обнаружению аномалий и вторжений заключается в использовании современного и актуального набора данных и моделированию его распределения по множеству взаимодействующих клиентов. В таких экспериментах в основном исследуются различные сценарии распределения данных, и оценивается их влияние на точность обнаружения аномалий и атак. Вместе с тем эксперименты должны включать также описания топологии сети, архитектуру СОВ, характеристики вычислительных узлов. Это позволит получить реалистичные оценки по вычислительной эффективности и длительности обучения модели в федеративном режиме. Решением данной проблемы является развертывание СОВ на основе ФО на экспериментальном стенде, сочетающем как виртуальные, так и физические устройства. Примером такого стенда может служить программно-аппаратный комплекс, описанный в [68].

В заключение также стоит отметить, что в настоящий момент большинство исследований в области применения ФО для построения СОВ сфокусировано на задаче обнаружения вторжений и классификации атак. Задача, связанная с разработкой технологии применения контрмер на основе ФО, практически не представлена в научной литературе. Вместе с тем, применение ФО может значительно повысить эффективность подсистем предотвращения вторжений, особенно развернутых в программно-определяемых сетях и реализующих технологии самовосстановления и самозащиты.

Литература

1. McMahan B., Moore E., Ramage D., Hampson S., Arcas B.A. Communication-Efficient Learning of Deep Networks from Decentralized Data // Artificial intelligence and statistics. 2017. pp. 1273–1282.
2. Lwakatare L.E., Raj A., Bosch J., Olsson H.H., Crnkovic I.A Taxonomy of Software Engineering Challenges for Machine Learning Systems: An Empirical Investigation (Eds.: Kruchten P., Fraser S., Coallier F.) // Agile Processes in Software Engineering and Extreme Programming: Proceedings of 20th International Conference. 2019. pp. 227–243.
3. Antonakakis M., April T., Bailey M., Bernhard M., Bursztein E., Cochran J., Durumeric Z., Halderman J.A., Invernizzi L., Kallitsis M., Kumar D., Lever C., Ma Z., Mason J., Menscher D., Seaman C., Thomas K., Zhou Y. Understanding the Mirai Botnet // 26th USENIX Security Symposium (USENIX Security 17). 2017. pp. 1093–1110.
4. Novikova E., Doynikova E., Golubev S. Federated Learning for Intrusion Detection in the Critical Infrastructures: Vertically Partitioned Data Use Case // Algorithms. 2022. vol. 15(4). no. 104. DOI: 10.3390/a15040104.
5. Ludwig H, et al. IBM Federated Learning: an Enterprise Framework White Paper V0.1. ArXiv preprint arXiv:2007.10987. 2020.

6. Lo S.K. Lu Q., Zhu L., Paik H.Y., Xu X., Wang C. Architectural Patterns for the Design of Federated Learning Systems // Journal of Systems and Software. 2022. vol. 191. no. 111357.
7. Sannara E.K., Portet F., Lalanda P., German V.E.G.A. A Federated Learning Aggregation Algorithm for Pervasive Computing: Evaluation and Comparison // IEEE International Conference on Pervasive Computing and Communications (PerCom). 2021. pp. 1–10. DOI: 10.1109/PERCOM50583.2021.9439129.
8. Yurochkin M., Agarwal M., Ghosh S., Greenewald K., Hoang N., Khazaeni Y. Bayesian Nonparametric Federated Learning of Neural Networks // International conference on machine learning. 2019. pp. 7252–7261.
9. Mansour A.B., Carenini G., Duplessis A., Naccache D. Federated Learning Aggregation: New Robust Algorithms with Guarantees. 21st IEEE International Conference on Machine Learning and Applications (ICMLA). 2022. pp. 721–726. DOI: 10.48550/ARXIV.2205.10864.
10. Shahid O., Pouriyeh S., Parizi R.M., Sheng Q.Z., Srivastava G., Zhao L. Communication Efficiency in Federated Learning: Achievements and Challenges // ArXiv preprint arXiv:2107.10996. 2021.
11. Juvekar C., Vaikuntanathan V., Chandrakasan A. GAZELLE: A Low Latency Framework for Secure Neural Network Inference // Proceedings of the 27th USENIX Security Symposium (USENIX Security 18). 2018. pp. 1651–1669.
12. Zhang C., Li S., Xia J., Wang W., Yan F., Liu Y. BatchCrypt: Efficient Homomorphic Encryption for Cross-Silo Federated Learning // Proceedings of the 2020 USENIX Conference on Usenix Annual Technical Conference. USENIX annual technical conference (USENIX ATC 20). 2020. pp. 493–506.
13. Kairouz P., et al. Advances and Open Problems in Federated Learning. Foundations and Trends in Machine Learning. 2021. vol. 14. no. 1–2. pp. 1–210.
14. Truex S., Liu L., Chow K.H., Gursoy M.E., Wei W. LDP-Fed: federated learning with local differential privacy // Proceedings of the Third ACM International Workshop on Edge Systems, Analytics and Networking. 2020. pp. 61–66.
15. Shokri R., Shmatikov V. Privacy-preserving deep learning // Proceedings of the 22nd ACM SIGSAC conference on computer and communications security. 2015. pp. 1310–1321. DOI: 10.1109/ALLERTON.2015.7447103.
16. Novikova E, Fomichov D., Kholod I., Filippov E. Analysis of Privacy-Enhancing Technologies in Open-Source Federated Learning Frameworks for Driver Activity Recognition // Sensors. 2022. vol. 22(8). no. 2983. DOI: 10.3390/s22082983.
17. Запечников С. Модели и алгоритмы конфиденциального машинного обучения // Безопасность информационных технологий. 2020. Т. 27. № 1. С. 51–67. DOI: 10.26583/bit.2020.1.05.
18. Rieke N., Hancox J., Li W., Milletarì F., Roth H.R., Albarqouni S., Bakas S., Galtier M.N., Landman B.A., Maier-Hein K., Ourselin S., Sheller M., Summers R.M., Trask A., Xu D., Baust M., Cardoso M.J. The future of digital health with federated learning // NPJ Digital Medicine. 2020. vol. 3. no. 119. DOI: 10.1038/s41746-020-00323-1.
19. Antunes R.S., André da Costa C., Küderle A., Yari I.A., Eskofier B. Federated Learning for Healthcare: Systematic Review and Architecture Proposal // ACM Transactions on Intelligent Systems and Technology (TIST). 2022. vol. 13(4). no. 54. DOI: 10.1145/3501813.
20. Nguyen T.D., Marchal S., Miettinen M., Fereidooni H., Asokan N., Sadeghi A.R. DIoT: A Federated Self-learning Anomaly Detection System for IoT // IEEE 39th International Conference on Distributed Computing Systems (ICDCS). 2019. pp. 756–767.
21. Li B., Wu Y., Song J., Lu R., Li T., Zhao L. DeepFed: Federated Deep Learning for Intrusion Detection in Industrial Cyber-Physical Systems // IEEE Transactions on

Industrial Informatics. 2020. vol. 17. no. 8. pp. 5615–5624. DOI: 10.1109/TII.2020.3023430.
22. Rey V., Sánchez P.M.S., Celdrán A.H., Bovet G. Federated learning for malware detection in IoT devices // Computer Networks. 2022. vol. 204. no. 108693. DOI: 10.1016/j.comnet.2021.108693.
23. Huong T.T., Bac T.P., Long D.M., Thang B.D., Binh N.T., Luong T.D., Phuc T.K. LocKedge: Low-Complexity Cyberattack Detection in IoT Edge Computing // IEEE Access. 2021. vol. 9. pp. 29696–29710. DOI: 10.1109/ACCESS.2021.3058528.
24. Khoa T.V., Saputra Y.M., Hoang D.T., Trung N.L., Nguyen D., Ha N.V., Dutkiewicz E. Collaborative Learning Model for Cyberattack Detection Systems in IoT Industry 4.0 // IEEE Wireless Communications and Networking Conference (WCNC). 2020. pp. 1–6. DOI: 10.1109/WCNC45663.2020.9120761.
25. Long G., Tan Y., Jiang J., Zhang C. Federated Learning for Open Banking // Federated Learning: Privacy and Incentive. 2020. pp. 240–254.
26. Ahmed U., Srivastava G., Lin J.C.-W. Reliable customer analysis using federated learning and exploring deep-attention edge intelligence // Future Generation Computer Systems. 2022. vol. 127. pp. 70–79. DOI: 10.1016/j.future.2021.08.028.
27. Li J., Cui T., Yang K., Yuan R., He L., Li M. Demand Forecasting of E-Commerce Enterprises Based on Horizontal Federated Learning from the Perspective of Sustainable Development // Sustainability. 2021. vol. 13(23). no. 13050. DOI: 10.3390/su132313050.
28. Дзюба В.И. Применение концепции федеративного обучения для решения задачи классификации текста // Процессы управления и устойчивость. 2022. Т. 9. № 1. С. 210–214.
29. Гонсалес П.Ю., Холод И.И. Архитектура многоагентных систем для федеративного обучения. Компьютерные инструменты в образовании. 2022. № 1. С. 30–45. DOI: 10.32603/2071-2340-2022-1-30-45.
30. Холод И.И., Ефремов М.А. Разработка архитектуры универсального фреймворка федеративного обучения // Программные продукты и системы. 2022. Т. 35. № 2. С. 263–272. DOI: 10.15827/0236-235X.138.263-272.
31. Swarm learning: Driving advances both practical and profound. URL: https://www.hpe.com/us/en/insights/articles/swarm-learning-driving- advances-both-practical-and-profound-2111.html. (accessed 24.10.2022).
32. Bellatreche L., Boukhalfa K., Richard P. Data Partitioning in Data Warehouses: Hardness Study, Heuristics and ORACLE Validation // Data Warehousing and Knowledge Discovery: Proceedings of the 10th International Conference on Data Warehousing and Knowledge Discovery. 2008. pp. 87–96. DOI: 10.1007/978-3-540-85836-2_9.
33. Khraisat A., Gondal I., Vamplew P., Kamruzzaman J. Survey of intrusion detection systems: techniques, datasets and challenges // Cybersecurity. 2019. vol. 2. no. 1. pp. 1–22. DOI: 10.1186/s42400-019-0038-7.
34. Kotenko I., Saenko I., Branitskiy A. Framework for Mobile Internet of Things Security Monitoring Based on Big Data Processing and Machine Learning // IEEE Access. 2018. vol. 6. pp. 72714–72723. DOI: 10.1109/ACCESS.2018.2881998.
35. Bukhanov D.G., Polyakov V.M. Detection of network attacks based on adaptive resonance theory // Journal of Physics: Conference Series. 2018. vol. 1015(4). no. 042007. DOI: 10.1088/1742-6596/1015/4/042007.
36. Yunwu W. Using Fuzzy Expert System Based on Genetic Algorithms for Intrusion Detection System // International Forum on Information Technology and Applications. 2009. vol. 2. pp. 221–224. DOI: 10.1109/IFITA.2009.107.

37. Dave M.H., Sharma S.D. Improved Algorithm for Intrusion Detection Using Genetic Algorithm and SNORT. International Journal of Emerging Technology and Advanced Engineering. 2014. vol. 4. no. 8. pp. 273–276.
38. Ranjan R., Sahoo G. A New Clustering Approach for Anomaly Intrusion Detection // International Journal of Data Mining and Knowledge Management Process (IJDKP). 2014. vol. 4. no. 2. pp. 29–38. DOI: 10.5121/ijdkp.2014.4203.
39. Li Z., Qin Z., Huang K., Yang X., Ye S. Intrusion Detection Using Convolutional Neural Networks for Representation Learning // International conference on neural information processing. 2017. pp. 858–866.
40. Hu J., Liu C., Cui Y. An Improved CNN Approach for Network Intrusion Detection System // International Journal of Network Security. 2021. vol. 23. no. 4. pp. 569–575.
41. Vinayakumar R., Soman K., Poornachandran P. Evaluation of Recurrent Neural Network and Its Variants for Intrusion Detection System IDS // International Journal of Information System Modeling and Design (IJISMD). 2017. vol. 8. no. 3. pp. 43–63.
42. Song Y., Hyun S., Cheong Y.-G. Analysis of Autoencoders for Network Intrusion Detection // Sensors. 2021. vol. 21(13). no. 4294. DOI: 10.3390/s21134294.
43. Gajewski M., Batalla J.M., Mastorakis G., Mavromoustakis C.X. A distributed IDS architecture model for Smart Home systems // Cluster Computing. 2019. vol. 22. pp. 1739–1749.
44. Shterenberg S.I., Poltavtseva M.A. A Distributed Intrusion Detection System with Protection from an Internal Intruder // Automatic Control and Computer Sciences. 2018. vol. 52. pp. 945–953.
45. Schueller Q., Basu K., Younas M., Patel M., Ball F. A Hierarchical Intrusion Detection System using Support Vector Machine for SDN Network in Cloud Data Center // 28th International Telecommunication Networks and Applications Conference (ITNAC). 2018. pp. 1–6. DOI: 10.1109/ATNAC.2018.8615255.
46. Saghezchi F.B., Mantas G., Ribeiro J., Al-Rawi M., Mumtaz S., Rodriguez J. Towards a secure network architecture for smart grids in 5G era // 13th International Wireless Communications and Mobile Computing Conference (IWCMC). 2017. pp. 121–126. DOI: 10.1109/IWCMC.2017.7986273.
47. Zhang Y. Distributed Intrusion Detection System in a Multi-Layer Network Architecture of Smart Grids // IEEE Transactions on Smart Grid. 2011. vol. 2. no. 4. pp. 796–808. DOI: 10.1109/TSG.2011.2159818.
48. Javed Y., Felemban M., Shawly T., Kobes J., Ghafoor A. A Partition-Driven Integrated Security Architecture for Cyberphysical Systems // Computer. 2020. vol. 53. no. 3. pp. 47–56. DOI: 10.1109/MC.2019.2914906.
49. Kholod I., Yanaki E., Fomichev D., Shalugin E., Novikova E., Filippov E., Nordlund M. Open-Source Federated Learning Frameworks for IoT: A Comparative Review and Analysis // Sensors. 2020. vol. 21(1). no. 167. DOI: 10.3390/s21010167.
50. Kitchenham B.A. Procedures for Performing Systematic Reviews // Keele, UK, Keele University. 2004. vol. 33. pp. 1–26.
51. Campos E.M., Saura P.F., González-Vidal A., Hernández-Ramos J.L., Bernabé J.B., Baldini G., Skarmeta A. Evaluating Federated Learning for intrusion detection in Internet of Things: Review and challenges // Computer Networks. 2022. vol. 203. no. 108661. DOI: 10.1016/j.comnet.2021.108661.
52. Agrawal S., Sarkar S., Aouedi O., Yenduri G., Piamrat K., Alazab M., Bhattacharya S., Reddy Maddikunta P.K., Gadekallu T.R. Federated Learning for Intrusion Detection System: Concepts, Challenges and Future Directions // Computer Communications. 2022. vol. 195. pp. 346–361. DOI: 10.1016/j.comcom.2022.09.012

53. Sun Y., Ochiai H., Esaki H. Intrusion Detection with Segmented Federated Learning for Large-Scale Multiple LANs // International Joint Conference on Neural Networks (IJCNN). 2020. pp. 1–8. DOI: 10.1109/IJCNN48605.2020.9207094.
54. Zhao R., Yin Y., Shi Y., Xue Z. Intelligent intrusion detection based on federated learning aided long short-term memory // Physical Communication. 2020. vol. 42. no. 101157. DOI: 10.1016/j.phycom.2020.101157.
55. Kholidy H.A., Baiardi F., Hariri S. DDSGA: A Data-Driven Semi-Global Alignment Approach for Detecting Masquerade Attacks // IEEE Transactions on Dependable and Secure Computing. 2014. vol. 12. no. 2. pp. 164–178. DOI: 10.1109/TDSC.2014.2327966.
56. Saadat H., Aboumadi A., Mohamed A., Erbad A., Guizani M. Hierarchical Federated Learning for Collaborative IDS in IoT Applications // 10th Mediterranean Conference on Embedded Computing (MECO). 2021. pp. 1–6. DOI: 10.1109/MECO52532.2021.9460304.
57. University of New Brunswick dataset. NSL-KDD dataset. URL: https://www.unb.ca/cic/datasets/nsl.html. (accessed 15.05.2022).
58. Cetin B, Lazar A., Kim J., Sim A., Wu K. Federated Wireless Network Intrusion Detection // IEEE International Conference on Big Data (Big Data). 2019. pp. 6004–6006. DOI: 10.1109/BigData47090.2019.9005507.
59. Kolias C., Kambourakis G., Stavrou A., Gritzalis S. Intrusion Detection in 802.11 Networks: Empirical Evaluation of Threats and a Public Dataset // IEEE Communications Surveys and Tutorials. 2015. vol. 18. no. 1. pp. 184–208. DOI: 10.1109/COMST.2015.2402161.
60. Ayed M.A., Talhi C. Federated Learning for Anomaly-Based Intrusion Detection // International Symposium on Networks, Computers and Communications (ISNCC). 2021. pp. 1–8. DOI: 10.1109/ISNCC52172.2021.9615816.
61. Sharafaldin I., Lashkari A.H., Ghorbani A.A. Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization // International Conference on Information Systems Security and Privacy (ICISS). 2018. vol. 1. pp. 108–116.
62. Luo J., Yang X., Mohammed M.N. Federation Learning for Intrusion Detection Methods by Parse Convolutional Neural Network // Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT). 2022. pp. 1–7. DOI: 10.1109/ICAECT54875.2022.9807989.
63. Zhao R., Wang Y., Xue Z., Ohtsuki T., Adebisi B., Gui G. Semisupervised Federated-Learning Based Intrusion Detection Method for Internet of Things // IEEE Internet of Things Journal. 2022. vol. 10. pp. 8645–8657. DOI: 10.1109/JIOT.2022.3175918.
64. Meidan Y., Bohadana M., Mathov Y., Mirsky Y., Shabtai A., Breitenbacher D., Elovici Y. N-BaIoT-Network-Based Detection of IoT Botnet Attacks Using Deep Autoencoders // IEEE Pervasive Computing. 2018. vol. 17. no. 3. pp. 12–22. DOI: 10.1109/MPRV.2018.03367731.
65. Yang X., Luo J., Mohammed M.N. Federation Learning of Optimized Convolutional Neural Network Structure for Intrusion Detection // Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT). 2022. pp. 1–7. DOI: 10.1109/ICAECT54875.2022.9807964.
66. Shi J., Ge B., Liu Y., Yan Y., Li S. Data Privacy Security Guaranteed Network Intrusion Detection System Based on Federated Learning // IEEE INFOCOM 2021 – IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). 2021. pp. 1–6. DOI: 10.1109/INFOCOMWKSHPS51825.2021.9484545.
67. Moustafa N., Slay J. UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set) // Military Communications and Information Systems Conference (MilCIS). 2015. pp. 1–6. DOI: 10.1109/MilCIS.2015.7348942.

68. Duy P.T., Van Hung T., Ha N.H., Do Hoang H., Pham V.H. Federated learning-based intrusion detection in SDN-enabled IIoT networks // 8th NAFOSTED Conference on Information and Computer Science (NICS). 2021. pp. 424–429. DOI: 10.1109/NICS54270.2021.9701525.
69. Sharafaldin I., Lashkari A.H., Hakak S., Ghorbani A.A. Developing Realistic Distributed Denial of Service (DDoS) Attack Dataset and Taxonomy // International Carnahan Conference on Security Technology (ICCST). 2019. pp. 1–8. DOI: 10.1109/CCST.2019.8888419.
70. Cheng Y., Lu J., Niyato D., Lyu B., Kang J., Zhu S. Federated Transfer Learning With Client Selection for Intrusion Detection in Mobile Edge Computing // IEEE Communications Letters. 2022. vol. 26. no. 3. pp. 552–556. DOI: 10.1109/LCOMM.2022.3140273.
71. Wang N., Chen Y., Hu Y., Lou W., Hou Y.T. FeCo: Boosting Intrusion Detection Capability in IoT Networks via Contrastive Learning // IEEE INFOCOM 2022 – IEEE Conference on Computer Communications. 2022. pp. 1409–1418. DOI: 10.1109/INFOCOM48880.2022.9796926.
72. Popoola S.I., Gui G., Adebisi B., Hammoudeh M., Gacanin H. Federated Deep Learning for Collaborative Intrusion Detection in Heterogeneous Networks // IEEE 94th Vehicular Technology Conference (VTC2021-Fall). 2021. pp. 1–6. DOI: 10.1109/VTC2021-Fall52928.2021.9625505.
73. Alsaedi A., Moustafa N., Tari Z., Mahmood A., Anwar A. TON_IoT Telemetry Dataset: A New Generation Dataset of IoT and IIoT for Data-Driven Intrusion Detection Systems // IEEE Access. 2020. vol. 8. pp. 165130–165150. DOI: 10.1109/ACCESS.2020.3022862.
74. Koroniotis N., Moustafa N., Sitnikova E., Turnbull B. Towards the development of realistic botnet dataset in the Internet of Things for network forensic analytics: Bot-IoT dataset // Future Generation Computer Systems. 2019. vol. 100. pp. 779–796. DOI: 10.1016/j.future.2019.05.041.
75. Al-Marri N.A.A.-A., Ciftler B.S., Abdallah M.M. Federated Mimic Learning for Privacy Preserving Intrusion Detection // IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom). 2020. pp. 1–6.
76. Mothukuri V., Khare P., Parizi R.M., Pouriyeh S., Dehghantanha A., Srivastava G. Federated-Learning-Based Anomaly Detection for IoT Security Attacks // IEEE Internet of Things Journal. 2021. vol. 9. no. 4. pp. 2545–2554. DOI: 10.1109/JIOT.2021.3077803.
77. Frazao I., Abreu P.H., Cruz T., Araújo H., Simões P. Denial of Service Attacks: Detecting the Frailties of Machine Learning Algorithms in the Classification Process // Critical Information Infrastructures Security 13th International Conference (CRITIS 2018). 2019. pp. 230–235.
78. Ruzafa-Alcazar P., Fernández-Saura P., Mármol-Campos E., González-Vidal A., Hernández-Ramos J.L., Bernal-Bernabe J., Skarmeta A.F. Intrusion Detection Based on Privacy- Preserving Federated Learning for the Industrial IoT // IEEE Transactions on Industrial Informatics. 2021. vol. 19. no. 2. pp. 1145–1154. DOI: 10.1109/TII.2021.3126728.
79. Chen Z., Lv N., Liu P., Fang Y., Chen K., Pan W. Intrusion Detection for Wireless Edge Networks Based on Federated Learning // IEEE Access. 2020. vol. 8. pp. 217463–217472. DOI: 10.1109/ACCESS.2020.3041793.
80. KDD dataset. URL: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html. (accessed 15.03.2022).
81. Dong T., Qiu H., Lu J., Qiu M., Fan C. Towards Fast Network Intrusion Detection based on Efficiency-preserving Federated Learning // IEEE Intl Conf on Parallel and Distributed Processing with Applications, Big Data and Cloud Computing,

Sustainable Computing & Communications, Social Computing and Networking (ISPA/BDCloud/SocialCom/SustainCom). 2021. pp. 468–475. DOI: 10.1109/ISPA-BDCloud-SocialCom-SustainCom52081.2021.00071.

82. Tabassum A., Erbad A., Lebda W., Mohamed A., Guizani M FEDGAN-IDS: Privacy-preserving IDS using GAN and Federated Learning // Computer Communications. 2022. vol. 192. pp. 299–310. DOI: 10.1016/j.comcom.2022.06.015.
83. Aouedi O., Piamrat K., Muller G., Singh K. FLUIDS: Federated Learning with semi-supervised approach for Intrusion Detection System // IEEE 19th Annual Consumer Communications and Networking Conference (CCNC). 2022. pp. 523–524. DOI: 10.1109/CCNC49033.2022.9700632.
84. Qin Y., Kondo M. Federated Learning-Based Network Intrusion Detection with a Feature Selection Approach // International Conference on Electrical, Communication, and Computer Engineering (ICECCE). 2021. pp. 1–6. DOI: 10.1109/ICECCE52056.2021.9514222.
85. Nguyen T.D., Marchal S., Miettinen M., Fereidooni H., Asokan N., Sadeghi A.R. DIoT: A Federated Self-learning Anomaly Detection System for IoT // IEEE 39th International Conference on Distributed Computing Systems (ICDCS). 2019. pp. 756–767.
86. Qin T., Cheng G., Chen W., Lei X. FNEL: An Evolving Intrusion Detection System Based on Federated Never-Ending Learning // 17th International Conference on Mobility, Sensing and Networking (MSN). 2021. pp. 239–246. DOI: 10.1109/MSN53354.2021.00047.
87. Fan Y., Li Y., Zhan M., Cui H., Zhang Y. IoTDefender: A Federated Transfer Learning Intrusion Detection Framework for 5G IoT // IEEE 14th International Conference on Big Data Science and Engineering (BigDataSE). 2020. pp. 88–95. DOI: 10.1109/BigDataSE50710.2020.00020.
88. Kang H., Ahn D.H., Lee G.M., Yoo J., Park K.H., Kim H.K. IoT network intrusion dataset. IEEE Dataport. 2019. vol. 10. DOI: 10.21227/q70p-q449.
89. Mirzaee P.H., Shojafar M., Pooranian Z., Asefy P., Cruickshank H., Tafazolli R. FIDS: A Federated Intrusion Detection System for 5G Smart Metering Network // 17th International Conference on Mobility, Sensing and Networking (MSN). 2021. pp. 215–222. DOI: 10.1109/MSN53354.2021.00044.
90. Regan C., Nasajpour M., Parizi R.M., Pouriyeh S., Dehghantanha A., Choo K.K.R. Federated IoT security attack detection using decentralized edge data // Machine Learning with Applications. 2022. vol. 8. no. 100263. DOI: 10.1016/j.mlwa.2022.100263.
91. Singh P., Gaba G. S., Kaur A., Hedabou M., Gurtov A. Dew-Cloud-Based Hierarchical Federated Learning for Intrusion Detection in IoMT // IEEE Journal of Biomedical and Health Informatics. 2022. vol. 27. no. 2. pp. 722–731. DOI: 10.1109/JBHI.2022.3186250.
92. Astillo P.V. Federated intelligence of anomaly detection agent in IoTMD-enabled Diabetes Management Control System // Future Generation Computer Systems. 2022. vol. 128. pp. 395–405. DOI: 10.1016/j.future.2021.10.023.
93. Astillo P.V., Jeong J., Chien W.C., Kim B., Jang J., You I. SMDAps: A specification-based misbehavior detection system for implantable devices in artificial pancreas system // Journal of Internet Technology. 2021. vol. 22. no. 1. pp. 1–11.
94. Siniosoglou I., Sarigiannidis P., Argyriou V., Lagkas T., Goudos S.K., Poveda M. Federated Intrusion Detection In NG- IoT Healthcare Systems: An Adversarial Approach // ICC 2021 – IEEE International Conference on Communications. 2021. pp. 1–6. DOI: 10.1109/ICC42927.2021.9500578.

95. Kim N.H., Krasner A., Kosinski C., Wininger M., Qadri M., Kappus Z., Danish S., Craelius W. Trending autoregulatory indices during treatment for traumatic brain injury // Journal of Clinical Monitoring and Computing. 2016. vol. 30. pp. 821–831.
96. Li B., Wu Y., Song J., Lu R., Li T., Zhao L. DeepFed: Federated Deep Learning for Intrusion Detection in Industrial Cyber–Physical Systems // IEEE Transactions on Industrial Informatics. 2020. vol. 17. no. 8. pp. 5615–5624. DOI: 10.1109/TII.2020.3023430.
97. Morris T., Gao W. Industrial Control System Traffic Data Sets for Intrusion Detection Research // Critical Infrastructure Protection VIII: 8th IFIP WG 11.10 International Conference (ICCIP). 2014. pp. 65–78.
98. Aouedi O., Piamrat K., Muller G., Singh K. Federated Semisupervised Learning for Attack Detection in Industrial Internet of Things // IEEE Transactions on Industrial Informatics. 2022. vol. 19. no. 1. pp. 286–295. DOI: 10.1109/TII.2022.3156642.
99. Truong T., Ta B.P., Le Q.A., Nguyen D.M., Le C.T., Nguyen H.X., Do H.T., Nguyen H.T., Tran K.P. Light-weight federated learning- based anomaly detection for time-series data in industrial control systems // Computers in Industry. 2022. vol. 140. no. 103692. DOI: 10.1016/j.compind.2022.103692.
100. Turnipseed I.P. A new scada dataset for intrusion detection research // Mississippi State University. 2015.
101. Secure Water Treatment (SWaT). URL: https://itrust.sutd.edu.sg/itrust-labs_datasets/dataset_info/. (accessed 25.06.2022).
102. HAI (HIL-based Augmented ICS) Security Dataset. URL: https://github.com/icsdataset/hai. (accessed 01.03.2023).
103. Keogh E., Lin J., Fu A. HOT SAX: efficiently finding the most unusual time series subsequence // Fifth IEEE International Conference on Data Mining (ICDM'05). 2005. pp. 226–233. DOI: 10.1109/ICDM.2005.79.
104. NYC taxi and limousine commission. URL: https://www.nyc.gov/site/tlc/index.page. (accessed 01.03.2023).
105. Liu H., Zhang S., Zhang P., Zhou X., Shao X., Pu G., Zhang Y. Blockchain and Federated Learning for Collaborative Intrusion Detection in Vehicular Edge Computing // IEEE Transactions on Vehicular Technology. 2021. vol. 70. no. 6. pp. 6073–6084. DOI: 10.1109/TVT.2021.3076780.
106. Abdel-Basset M., Moustafa N., Hawash H., Razzak I., Sallam K.M., Elkomy O.M. Federated Intrusion Detection in Blockchain-Based Smart Transportation Systems // IEEE Transactions on Intelligent Transportation Systems. 2021. vol. 23. no. 3. pp. 2523–2537. DOI: 10.1109/TITS.2021.3119968.
107. Aliyu I., Feliciano M.C., Van Engelenburg S., Kim D.O., Lim C. G.A Blockchain-Based Federated Forest for SDN – Enabled In-Vehicle Network Intrusion Detection System // IEEE Access. 2021. vol. 9. pp. 102593–102608. DOI: 10.1109/ACCESS.2021.3094365.
108. Li Q., He B., Song D. Model-Contrastive Federated Learning. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021. pp. 10713–10722.
109. McMahan H., Moore E., Ramage D., Arcas B.A. Federated Learning of Deep Networks using Model Averaging. ArXiv preprint arXiv:1602.05629. 2016. URL: https://fate.fedai.org/. (accessed 25.06.2022).
110. FATE. An Industrial Grade Federated Learning Framework. URL: https://fate.fedai.org/. (accessed 25.06.2022).
111. Yin D., Chen Y., Kannan R., Bartlett P. Byzantine-Robust Distributed Learning: Towards Optimal Statistical Rates // Proceedings of the 35th International Conference on Machine Learning. 2018. vol. 80. pp. 5650–5659.

Новикова Евгения Сергеевна — канд. техн. наук, старший научный сотрудник, лаборатория проблем компьютерной безопасности, Федеральное государственное бюджетное учреждение науки "Санкт-Петербургский Федеральный исследовательский центр Российской академии наук" (СПб ФИЦ РАН). Область научных интересов: безопасность информационных систем, обнаружение аномалий методами машинного обучения, конфиденциальность данных. Число научных публикаций — 60. novikova@comsec.spb.ru; 14 линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)328-7181.

Федорченко Елена Владимировна — канд. техн. наук, старший научный сотрудник, лаборатория проблем компьютерной безопасности, Федеральное государственное бюджетное учреждение науки "Санкт-Петербургский Федеральный исследовательский центр Российской академии наук" (СПб ФИЦ РАН). Область научных интересов: безопасность информационных систем, методы анализа рисков компьютерных сетей, управление информационными рисками, анализ данных, поддержка принятия решений по повышению защищенности. Число научных публикаций — 100. doynikova@comsec.spb.ru; 14 линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)328-7181.

Котенко Игорь Витальевич — д-р техн. наук, профессор, руководитель лаборатории, лаборатория проблем компьютерной безопасности, Федеральное государственное бюджетное учреждение науки "Санкт-Петербургский Федеральный исследовательский центр Российской академии наук" (СПб ФИЦ РАН). Область научных интересов: безопасность компьютерных сетей, в том числе управление политиками безопасности, разграничение прав доступа, аутентификация, анализ защищенности, обнаружение компьютерных атак, межсетевые экраны, защита от вирусов и сетевых червей, анализ и верификация протоколов безопасности и систем защиты информации, защита программного обеспечения от взлома и управление цифровыми правами, технологии моделирования и визуализации для противодействия кибер-терроризму. Число научных публикаций — 450. ivkote@comsec.spb.ru; 14 линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)328-7181.

Холод Иван Иванович — д-р техн. наук, доцент, декан факультета, факультета компьютерных технологий и информатики, Санкт-Петербургский Электротехнический университет «ЛЭТИ». Область научных интересов: распределенные параллельные алгоритмы машинного обучения. Число научных публикаций — 50. iiholod@etu.ru; улица Профессора Попова, 5, 197022, Санкт-Петербург, Россия; р.т.: +7(812)234-2746.

Поддержка исследований. Работа выполнена при финансовой поддержке РНФ (проект № 22-21-00724).

DOI 10.15622/ia.22.5.4

E. NOVIKOVA, E. FEDORCHENKO, I. KOTENKO, I. KHOLOD

ANALYTICAL REVIEW OF INTELLIGENT INTRUSION DETECTION SYSTEMS BASED ON FEDERATED LEARNING: ADVANTAGES AND OPEN CHALLENGES

Novikova E., Fedorchenko E., Kotenko I., Kholod I. **Analytical Review of Intelligent Intrusion Detection Systems Based on Federated Learning: Advantages and Open Challenges.**

Abstract. To provide an accurate and timely response to different types of attacks, intrusion detection systems collect and analyze a large amount of data, which may include information with limited access, such as personal data or trade secrets. Consequently, such systems can be seen as an additional source of risks associated with handling sensitive information and breaching its security. Applying the federated learning paradigm to build analytical models for attack and anomaly detection can significantly reduce such risks because locally generated data is not transmitted to any third party, and model training is done locally - on the data sources. Using federated training for intrusion detection solves the problem of training on data that belongs to different organizations, and which, due to the need to protect commercial or other secrets, cannot be placed in the public domain. Thus, this approach also allows us to expand and diversify the set of data on which machine learning models are trained, thereby increasing the level of detectability of heterogeneous attacks. Due to the fact that this approach can overcome the aforementioned problems, it is actively used to design new approaches for intrusion and anomaly detection. The authors systematically explore existing solutions for intrusion and anomaly detection based on federated learning, study their advantages, and formulate open challenges associated with its application in practice. Particular attention is paid to the architecture of the proposed systems, the intrusion detection methods and models used, and approaches for modeling interactions between multiple system users and distributing data among them are discussed. The authors conclude by formulating open problems that need to be solved in order to apply federated learning-based intrusion detection systems in practice.

Keywords: intrusion detection, anomalies, federated learning, analysis models, data partition.

References

1. McMahan B., Moore E., Ramage D., Hampson S., Arcas B.A. Communication-Efficient Learning of Deep Networks from Decentralized Data. Artificial intelligence and statistics. 2017. pp. 1273–1282.
2. Lwakatare L.E., Raj A., Bosch J., Olsson H.H., Crnkovic I.A Taxonomy of Software Engineering Challenges for Machine Learning Systems: An Empirical Investigation. Agile Processes in Software Engineering and Extreme Programming: Proceedings of 20th International Conference. 2019. pp. 227–243.
3. Antonakakis M., April T., Bailey M., Bernhard M., Bursztein E., Cochran J., Durumeric Z., Halderman J.A., Invernizzi L., Kallitsis M., Kumar D., Lever C., Ma Z., Mason J., Menscher D., Seaman C., Thomas K., Zhou Y. Understanding the Mirai Botnet. 26th USENIX Security Symposium (USENIX Security 17). 2017. pp. 1093–1110.
4. Novikova E., Doynikova E., Golubev S. Federated Learning for Intrusion Detection in the Critical Infrastructures: Vertically Partitioned Data Use Case. Algorithms. 2022. vol. 15(4). no. 104. DOI: 10.3390/a15040104.

5. Ludwig H, et al. IBM Federated Learning: an Enterprise Framework White Paper V0.1. ArXiv preprint arXiv:2007.10987. 2020.
6. Lo S.K. Lu Q., Zhu L., Paik H.Y., Xu X., Wang C. Architectural Patterns for the Design of Federated Learning Systems. Journal of Systems and Software. 2022. vol. 191. no. 111357.
7. Sannara E.K., Portet F., Lalanda P., German V.E.G.A. A Federated Learning Aggregation Algorithm for Pervasive Computing: Evaluation and Comparison. IEEE International Conference on Pervasive Computing and Communications (PerCom). 2021. pp. 1–10. DOI: 10.1109/PERCOM50583.2021.9439129.
8. Yurochkin M., Agarwal M., Ghosh S., Greenewald K., Hoang N., Khazaeni Y. Bayesian Nonparametric Federated Learning of Neural Networks. International conference on machine learning. 2019. pp. 7252–7261.
9. Mansour A.B., Carenini G., Duplessis A., Naccache D. Federated Learning Aggregation: New Robust Algorithms with Guarantees. 21st IEEE International Conference on Machine Learning and Applications (ICMLA). 2022. pp. 721–726. DOI: 10.48550/ARXIV.2205.10864.
10. Shahid O., Pouriyeh S., Parizi R.M., Sheng Q.Z., Srivastava G., Zhao L. Communication Efficiency in Federated Learning: Achievements and Challenges. ArXiv preprint arXiv:2107.10996. 2021.
11. Juvekar C., Vaikuntanathan V., Chandrakasan A. GAZELLE: A Low Latency Framework for Secure Neural Network Inference. Proceedings of the 27th USENIX Security Symposium (USENIX Security 18). 2018. pp. 1651–1669.
12. Zhang C., Li S., Xia J., Wang W., Yan F., Liu Y. BatchCrypt: Efficient Homomorphic Encryption for Cross-Silo Federated Learning. Proceedings of the 2020 USENIX Conference on Usenix Annual Technical Conference. USENIX annual technical conference (USENIX ATC 20). 2020. pp. 493–506.
13. Kairouz P., et al. Advances and Open Problems in Federated Learning. Foundations and Trends in Machine Learning. 2021. vol. 14. no. 1–2. pp. 1–210.
14. Truex S., Liu L., Chow K.H., Gursoy M.E., Wei W. LDP-Fed: federated learning with local differential privacy. Proceedings of the Third ACM International Workshop on Edge Systems, Analytics and Networking. 2020. pp. 61–66.
15. Shokri R., Shmatikov V. Privacy-preserving deep learning. Proceedings of the 22nd ACM SIGSAC conference on computer and communications security. 2015. pp. 1310–1321. DOI: 10.1109/ALLERTON.2015.7447103.
16. Novikova E, Fomichov D., Kholod I., Filippov E. Analysis of Privacy-Enhancing Technologies in Open-Source Federated Learning Frameworks for Driver Activity Recognition. Sensors. 2022. vol. 22(8). no. 2983. DOI: 10.3390/s22082983.
17. Zapechnikov S. [Models and algorithms of the confidential machine learning]. Bezopasnost' informacionnih technologii – IT security. 2020. vol. 27. no. 1. pp. 51–67. DOI: 10.26583/bit.2020.1.05. (In Russ.).
18. Rieke N., Hancox J., Li W., Milletarì F., Roth H.R., Albarqouni S., Bakas S., Galtier M.N., Landman B.A., Maier-Hein K., Ourselin S., Sheller M., Summers R.M., Trask A., Xu D., Baust M., Cardoso M.J. The future of digital health with federated learning. NPJ Digital Medicine. 2020. vol. 3. no. 119. DOI: 10.1038/s41746-020-00323-1.
19. Antunes R.S., André da Costa C., Küderle A., Yari I.A., Eskofier B. Federated Learning for Healthcare: Systematic Review and Architecture Proposal. ACM Transactions on Intelligent Systems and Technology (TIST). 2022. vol. 13(4). no. 54. DOI: 10.1145/3501813.
20. Nguyen T.D., Marchal S., Miettinen M., Fereidooni H., Asokan N., Sadeghi A.R.. DIoT: A Federated Self-learning Anomaly Detection System for IoT. IEEE 39th

International Conference on Distributed Computing Systems (ICDCS). 2019. pp. 756–767.
21. Li B., Wu Y., Song J., Lu R., Li T., Zhao L. DeepFed: Federated Deep Learning for Intrusion Detection in Industrial Cyber-Physical Systems. IEEE Transactions on Industrial Informatics. 2020. vol. 17. no. 8. pp. 5615–5624. DOI: 10.1109/TII.2020.3023430.
22. Rey V., Sánchez P.M.S., Celdrán A.H., Bovet G. Federated learning for malware detection in IoT devices. Computer Networks. 2022. vol. 204. no. 108693. DOI: 10.1016/j.comnet.2021.108693.
23. Huong T.T., Bac T.P., Long D.M., Thang B.D., Binh N.T., Luong T.D., Phuc T.K. LocKedge: Low-Complexity Cyberattack Detection in IoT Edge Computing. IEEE Access. 2021. vol. 9. pp. 29696–29710. DOI: 10.1109/ACCESS.2021.3058528.
24. Khoa T.V., Saputra Y.M., Hoang D.T., Trung N.L., Nguyen D., Ha N.V., Dutkiewicz E. Collaborative Learning Model for Cyberattack Detection Systems in IoT Industry 4.0. IEEE Wireless Communications and Networking Conference (WCNC). 2020. pp. 1–6. DOI: 10.1109/WCNC45663.2020.9120761.
25. Long G., Tan Y., Jiang J., Zhang C. Federated Learning for Open Banking. Federated Learning: Privacy and Incentive. 2020. pp. 240–254.
26. Ahmed U., Srivastava G., Lin J.C.-W. Reliable customer analysis using federated learning and exploring deep-attention edge intelligence. Future Generation Computer Systems. 2022. vol. 127. pp. 70–79. DOI: 10.1016/j.future.2021.08.028.
27. Li J., Cui T., Yang K., Yuan R., He L., Li M. Demand Forecasting of E-Commerce Enterprises Based on Horizontal Federated Learning from the Perspective of Sustainable Development. Sustainability. 2021. vol. 13(23). no. 13050. DOI: 10.3390/su132313050.
28. Dzyaba V.I. [Application of the federated learning to text classification.] Processy upravlenija i ustojchivost – Control processes and stability. 2022. vol. 9. no. 1. pp. 210–214.
29. Gonsales P.Yu., Kholod I.I. [Multi-agent architecture for federated learning]. Komp'juternye instrumenty v obrazovanii – Computer tools in Education. 2022. no. 1. pp. 30–45. DOI: 10.32603/2071-2340-2022-1-30-45.
30. Holod I.I., Efremov M.A. [Developing universal framework design for federated learning]. Programmnye produkty i sistemy – Software products and systems. 2022. vol. 35. no. 2. pp. 263–272. DOI: 10.15827/0236-235X.138.263-272.
31. Swarm learning: Driving advances both practical and profound. Available at: https://www.hpe.com/us/en/insights/articles/swarm-learning-driving- advances-both-practical-and-profound-2111.html. (accessed 24.10.2022).
32. Bellatreche L., Boukhalfa K., Richard P. Data Partitioning in Data Warehouses: Hardness Study, Heuristics and ORACLE Validation. Data Warehousing and Knowledge Discovery: Proceedings of the 10th International Conference on Data Warehousing and Knowledge Discovery. 2008. pp. 87–96. DOI: 10.1007/978-3-540-85836-2_9.
33. Khraisat A., Gondal I., Vamplew P., Kamruzzaman J. Survey of intrusion detection systems: techniques, datasets and challenges. Cybersecurity. 2019. vol. 2. no. 1. pp. 1–22. DOI: 10.1186/s42400-019-0038-7.
34. Kotenko I., Saenko I., Branitskiy A. Framework for Mobile Internet of Things Security Monitoring Based on Big Data Processing and Machine Learning. IEEE Access. 2018. vol. 6. pp. 72714–72723. DOI: 10.1109/ACCESS.2018.2881998.
35. Bukhanov D.G., Polyakov V.M. Detection of network attacks based on adaptive resonance theory. Journal of Physics: Conference Series. 2018. vol. 1015(4). no. 042007. DOI: 10.1088/1742-6596/1015/4/042007.

36. Yunwu W. Using Fuzzy Expert System Based on Genetic Algorithms for Intrusion Detection System. International Forum on Information Technology and Applications. 2009. vol. 2. pp. 221–224. DOI: 10.1109/IFITA.2009.107.
37. Dave M.H., Sharma S.D. Improved Algorithm for Intrusion Detection Using Genetic Algorithm and SNORT. International Journal of Emerging Technology and Advanced Engineering. 2014. vol. 4. no. 8. pp. 273–276.
38. Ranjan R., Sahoo G. A New Clustering Approach for Anomaly Intrusion Detection. International Journal of Data Mining and Knowledge Management Process (IJDKP). 2014. vol. 4. no. 2. pp. 29–38. DOI: 10.5121/ijdkp.2014.4203.
39. Li Z., Qin Z., Huang K., Yang X., Ye S. Intrusion Detection Using Convolutional Neural Networks for Representation Learning. International conference on neural information processing. 2017. pp. 858–866.
40. Hu J., Liu C., Cui Y. An Improved CNN Approach for Network Intrusion Detection System. International Journal of Network Security. 2021. vol. 23. no. 4. pp. 569–575.
41. Vinayakumar R., Soman K., Poornachandran P. Evaluation of Recurrent Neural Network and Its Variants for Intrusion Detection System IDS. International Journal of Information System Modeling and Design (IJISMD). 2017. vol. 8. no. 3. pp. 43–63.
42. Song Y., Hyun S., Cheong Y.-G. Analysis of Autoencoders for Network Intrusion Detection. Sensors. 2021. vol. 21(13). no. 4294. DOI: 10.3390/s21134294.
43. Gajewski M., Batalla J.M., Mastorakis G., Mavromoustakis C.X. A distributed IDS architecture model for Smart Home systems. Cluster Computing. 2019. vol. 22. pp. 1739–1749.
44. Shterenberg S.I., Poltavtseva M.A. A Distributed Intrusion Detection System with Protection from an Internal Intruder. Automatic Control and Computer Sciences. 2018. vol. 52. pp. 945–953.
45. Schueller Q., Basu K., Younas M., Patel M., Ball F. A Hierarchical Intrusion Detection System using Support Vector Machine for SDN Network in Cloud Data Center. 28th International Telecommunication Networks and Applications Conference (ITNAC). 2018. pp. 1–6. DOI: 10.1109/ATNAC.2018.8615255.
46. Saghezchi F.B., Mantas G., Ribeiro J., Al-Rawi M., Mumtaz S., Rodriguez J. Towards a secure network architecture for smart grids in 5G era. 13th International Wireless Communications and Mobile Computing Conference (IWCMC). 2017. pp. 121–126. DOI: 10.1109/IWCMC.2017.7986273.
47. Zhang Y. Distributed Intrusion Detection System in a Multi-Layer Network Architecture of Smart Grids. IEEE Transactions on Smart Grid. 2011. vol. 2. no. 4. pp. 796–808. DOI: 10.1109/TSG.2011.2159818.
48. Javed Y., Felemban M., Shawly T., Kobes J., Ghafoor A. A Partition-Driven Integrated Security Architecture for Cyberphysical Systems. Computer. 2020. vol. 53. no. 3. pp. 47–56. DOI: 10.1109/MC.2019.2914906.
49. Kholod I., Yanaki E., Fomichev D., Shalugin E., Novikova E., Filippov E., Nordlund M. Open-Source Federated Learning Frameworks for IoT: A Comparative Review and Analysis. Sensors. 2020. vol. 21(1). no. 167. DOI: 10.3390/s21010167.
50. Kitchenham B.A. Procedures for Performing Systematic Reviews. Keele, UK, Keele University. 2004. vol. 33. pp. 1–26.
51. Campos E.M., Saura P.F., González-Vidal A., Hernández-Ramos J.L., Bernabé J.B., Baldini G., Skarmeta A. Evaluating Federated Learning for intrusion detection in Internet of Things: Review and challenges. Computer Networks. 2022. vol. 203. no. 108661. DOI: 10.1016/j.comnet.2021.108661.
52. Agrawal S., Sarkar S., Aouedi O., Yenduri G., Piamrat K., Alazab M., Bhattacharya S., Reddy Maddikunta P.K., Gadekallu T.R. Federated Learning for Intrusion Detection System: Concepts, Challenges and Future Directions. Computer Communications. 2022. vol. 195. pp. 346–361. DOI: 10.1016/j.comcom.2022.09.012

53. Sun Y., Ochiai H., Esaki H. Intrusion Detection with Segmented Federated Learning for Large-Scale Multiple LANs. International Joint Conference on Neural Networks (IJCNN). 2020. pp. 1–8. DOI: 10.1109/IJCNN48605.2020.9207094.
54. Zhao R., Yin Y., Shi Y., Xue Z. Intelligent intrusion detection based on federated learning aided long short-term memory. Physical Communication. 2020. vol. 42. no. 101157. DOI: 10.1016/j.phycom.2020.101157.
55. Kholidy H.A., Baiardi F., Hariri S. DDSGA: A Data-Driven Semi-Global Alignment Approach for Detecting Masquerade Attacks. IEEE Transactions on Dependable and Secure Computing. 2014. vol. 12. no. 2. pp. 164–178. DOI: 10.1109/TDSC.2014.2327966.
56. Saadat H., Aboumadi A., Mohamed A., Erbad A., Guizani M. Hierarchical Federated Learning for Collaborative IDS in IoT Applications. 10th Mediterranean Conference on Embedded Computing (MECO). 2021. pp. 1–6. DOI: 10.1109/MECO52532.2021.9460304.
57. University of New Brunswick dataset. NSL-KDD dataset. Available at: https://www.unb.ca/cic/datasets/nsl.html. (accessed 15.05.2022).
58. Cetin B, Lazar A., Kim J., Sim A., Wu K. Federated Wireless Network Intrusion Detection. IEEE International Conference on Big Data (Big Data). 2019. pp. 6004–6006. DOI: 10.1109/BigData47090.2019.9005507.
59. Kolias C., Kambourakis G., Stavrou A., Gritzalis S. Intrusion Detection in 802.11 Networks: Empirical Evaluation of Threats and a Public Dataset. IEEE Communications Surveys and Tutorials. 2015. vol. 18. no. 1. pp. 184–208. DOI: 10.1109/COMST.2015.2402161.
60. Ayed M.A., Talhi C. Federated Learning for Anomaly-Based Intrusion Detection. International Symposium on Networks, Computers and Communications (ISNCC). 2021. pp. 1–8. DOI: 10.1109/ISNCC52172.2021.9615816.
61. Sharafaldin I., Lashkari A.H., Ghorbani A.A. Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization. International Conference on Information Systems Security and Privacy (ICISS). 2018. vol. 1. pp. 108–116.
62. Luo J., Yang X., Mohammed M.N. Federation Learning for Intrusion Detection Methods by Parse Convolutional Neural Network. Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT). 2022. pp. 1–7. DOI: 10.1109/ICAECT54875.2022.9807989.
63. Zhao R., Wang Y., Xue Z., Ohtsuki T., Adebisi B., Gui G. Semisupervised Federated-Learning Based Intrusion Detection Method for Internet of Things. IEEE Internet of Things Journal. 2022. vol. 10. pp. 8645–8657. DOI: 10.1109/JIOT.2022.3175918.
64. Meidan Y., Bohadana M., Mathov Y., Mirsky Y., Shabtai A., Breitenbacher D., Elovici Y. N-BaIoT-Network-Based Detection of IoT Botnet Attacks Using Deep Autoencoders. IEEE Pervasive Computing. 2018. vol. 17. no. 3. pp. 12–22. DOI: 10.1109/MPRV.2018.03367731.
65. Yang X., Luo J., Mohammed M.N. Federation Learning of Optimized Convolutional Neural Network Structure for Intrusion Detection. Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT). 2022. pp. 1–7. DOI: 10.1109/ICAECT54875.2022.9807964.
66. Shi J., Ge B., Liu Y., Yan Y., Li S. Data Privacy Security Guaranteed Network Intrusion Detection System Based on Federated Learning. IEEE INFOCOM 2021 – IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). 2021. pp. 1–6. DOI: 10.1109/INFOCOMWKSHPS51825.2021.9484545.
67. Moustafa N., Slay J. UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). Military Communications and Information Systems Conference (MilCIS). 2015. pp. 1–6. DOI: 10.1109/MilCIS.2015.7348942.

68. Duy P.T., Van Hung T., Ha N.H., Do Hoang H., Pham V.H. Federated learning-based intrusion detection in SDN-enabled IIoT networks. 8th NAFOSTED Conference on Information and Computer Science (NICS). 2021. pp. 424–429. DOI: 10.1109/NICS54270.2021.9701525.
69. Sharafaldin I., Lashkari A.H., Hakak S., Ghorbani A.A. Developing Realistic Distributed Denial of Service (DDoS) Attack Dataset and Taxonomy. International Carnahan Conference on Security Technology (ICCST). 2019. pp. 1–8. DOI: 10.1109/CCST.2019.8888419.
70. Cheng Y., Lu J., Niyato D., Lyu B., Kang J., Zhu S. Federated Transfer Learning With Client Selection for Intrusion Detection in Mobile Edge Computing. IEEE Communications Letters. 2022. vol. 26. no. 3. pp. 552–556. DOI: 10.1109/LCOMM.2022.3140273.
71. Wang N., Chen Y., Hu Y., Lou W., Hou Y.T. FeCo: Boosting Intrusion Detection Capability in IoT Networks via Contrastive Learning. IEEE INFOCOM 2022 – IEEE Conference on Computer Communications. 2022. pp. 1409–1418. DOI: 10.1109/INFOCOM48880.2022.9796926.
72. Popoola S.I., Gui G., Adebisi B., Hammoudeh M., Gacanin H. Federated Deep Learning for Collaborative Intrusion Detection in Heterogeneous Networks. IEEE 94th Vehicular Technology Conference (VTC2021-Fall). 2021. pp. 1–6. DOI: 10.1109/VTC2021-Fall52928.2021.9625505.
73. Alsaedi A., Moustafa N., Tari Z., Mahmood A., Anwar A. TON_IoT Telemetry Dataset: A New Generation Dataset of IoT and IIoT for Data-Driven Intrusion Detection Systems. IEEE Access. 2020. vol. 8. pp. 165130–165150. DOI: 10.1109/ACCESS.2020.3022862.
74. Koroniotis N., Moustafa N., Sitnikova E., Turnbull B. Towards the development of realistic botnet dataset in the Internet of Things for network forensic analytics: Bot-IoT dataset. Future Generation Computer Systems. 2019. vol. 100. pp. 779–796. DOI: 10.1016/j.future.2019.05.041.
75. Al-Marri N.A.A.-A., Ciftler B.S., Abdallah M.M. Federated Mimic Learning for Privacy Preserving Intrusion Detection. IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom). 2020. pp. 1–6.
76. Mothukuri V., Khare P., Parizi R.M., Pouriyeh S., Dehghantanha A., Srivastava G. Federated-Learning-Based Anomaly Detection for IoT Security Attacks. IEEE Internet of Things Journal. 2021. vol. 9. no. 4. pp. 2545–2554. DOI: 10.1109/JIOT.2021.3077803.
77. Frazao I., Abreu P.H., Cruz T., Araújo H., Simões P. Denial of Service Attacks: Detecting the Frailties of Machine Learning Algorithms in the Classification Process. Critical Information Infrastructures Security 13th International Conference (CRITIS 2018). 2019. pp. 230–235.
78. Ruzafa-Alcazar P., Fernández-Saura P., Mármol-Campos E., González-Vidal A., Hernández-Ramos J.L., Bernal-Bernabe J., Skarmeta A.F. Intrusion Detection Based on Privacy- Preserving Federated Learning for the Industrial IoT. IEEE Transactions on Industrial Informatics. 2021. vol. 19. no. 2. pp. 1145–1154. DOI: 10.1109/TII.2021.3126728.
79. Chen Z., Lv N., Liu P., Fang Y., Chen K., Pan W. Intrusion Detection for Wireless Edge Networks Based on Federated Learning. IEEE Access. 2020. vol. 8. pp. 217463–217472. DOI: 10.1109/ACCESS.2020.3041793.
80. KDD dataset. Available at: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html. (accessed 15.03.2022).
81. Dong T., Qiu H., Lu J., Qiu M., Fan C. Towards Fast Network Intrusion Detection based on Efficiency-preserving Federated Learning. IEEE Intl Conf on Parallel and Distributed Processing with Applications, Big Data and Cloud Computing,

Sustainable Computing and Communications, Social Computing and Networking (ISPA/BDCloud/SocialCom/SustainCom). 2021. pp. 468–475. DOI: 10.1109/ISPA-BDCloud-SocialCom-SustainCom52081.2021.00071.

82. Tabassum A., Erbad A., Lebda W., Mohamed A., Guizani M FEDGAN-IDS: Privacy-preserving IDS using GAN and Federated Learning. Computer Communications. 2022. vol. 192. pp. 299–310. DOI: 10.1016/j.comcom.2022.06.015.
83. Aouedi O., Piamrat K., Muller G., Singh K. FLUIDS: Federated Learning with semi-supervised approach for Intrusion Detection System. IEEE 19th Annual Consumer Communications and Networking Conference (CCNC). 2022. pp. 523–524. DOI: 10.1109/CCNC49033.2022.9700632.
84. Qin Y., Kondo M. Federated Learning-Based Network Intrusion Detection with a Feature Selection Approach. International Conference on Electrical, Communication, and Computer Engineering (ICECCE). 2021. pp. 1–6. DOI: 10.1109/ICECCE52056.2021.9514222.
85. Nguyen T.D., Marchal S., Miettinen M., Fereidooni H., Asokan N., Sadeghi A.R. DIoT: A Federated Self-learning Anomaly Detection System for IoT. IEEE 39th International Conference on Distributed Computing Systems (ICDCS). 2019. pp. 756–767.
86. Qin T., Cheng G., Chen W., Lei X. FNEL: An Evolving Intrusion Detection System Based on Federated Never-Ending Learning. 17th International Conference on Mobility, Sensing and Networking (MSN). 2021. pp. 239–246. DOI: 10.1109/MSN53354.2021.00047.
87. Fan Y., Li Y., Zhan M., Cui H., Zhang Y. IoTDefender: A Federated Transfer Learning Intrusion Detection Framework for 5G IoT. IEEE 14th International Conference on Big Data Science and Engineering (BigDataSE). 2020. pp. 88–95. DOI: 10.1109/BigDataSE50710.2020.00020.
88. Kang H., Ahn D.H., Lee G.M., Yoo J., Park K.H., Kim H.K. IoT network intrusion dataset. IEEE Dataport. 2019. vol. 10. DOI: 10.21227/q70p-q449.
89. Mirzaee P.H., Shojafar M., Pooranian Z., Asefy P., Cruickshank H., Tafazolli R. FIDS: A Federated Intrusion Detection System for 5G Smart Metering Network. 17th International Conference on Mobility, Sensing and Networking (MSN). 2021. pp. 215–222. DOI: 10.1109/MSN53354.2021.00044.
90. Regan C., Nasajpour M., Parizi R.M., Pouriyeh S., Dehghantanha A., Choo K.K.R. Federated IoT security attack detection using decentralized edge data. Machine Learning with Applications. 2022. vol. 8. no. 100263. DOI: 10.1016/j.mlwa.2022.100263.
91. Singh P., Gaba G. S., Kaur A., Hedabou M., Gurtov A. Dew-Cloud-Based Hierarchical Federated Learning for Intrusion Detection in IoMT. IEEE Journal of Biomedical and Health Informatics. 2022. vol. 27. no. 2. pp. 722–731. DOI: 10.1109/JBHI.2022.3186250.
92. Astillo P.V. Federated intelligence of anomaly detection agent in IoTMD-enabled Diabetes Management Control System. Future Generation Computer Systems. 2022. vol. 128. pp. 395–405. DOI: 10.1016/j.future.2021.10.023.
93. Astillo P.V., Jeong J., Chien W.C., Kim B., Jang J., You I. SMDAps: A specification-based misbehavior detection system for implantable devices in artificial pancreas system. Journal of Internet Technology. 2021. vol. 22. no. 1. pp. 1–11.
94. Siniosoglou I., Sarigiannidis P., Argyriou V., Lagkas T., Goudos S.K., Poveda M. Federated Intrusion Detection In NG- IoT Healthcare Systems: An Adversarial Approach. ICC 2021 – IEEE International Conference on Communications. 2021. pp. 1–6. DOI: 10.1109/ICC42927.2021.9500578.

95. Kim N.H., Krasner A., Kosinski C., Wininger M., Qadri M., Kappus Z., Danish S., Craelius W. Trending autoregulatory indices during treatment for traumatic brain injury. Journal of Clinical Monitoring and Computing. 2016. vol. 30. pp. 821–831.
96. Li B., Wu Y., Song J., Lu R., Li T., Zhao L. DeepFed: Federated Deep Learning for Intrusion Detection in Industrial Cyber–Physical Systems. IEEE Transactions on Industrial Informatics. 2020. vol. 17. no. 8. pp. 5615–5624. DOI: 10.1109/TII.2020.3023430.
97. Morris T., Gao W. Industrial Control System Traffic Data Sets for Intrusion Detection Research. Critical Infrastructure Protection VIII: 8th IFIP WG 11.10 International Conference (ICCIP). 2014. pp. 65–78.
98. Aouedi O., Piamrat K., Muller G., Singh K. Federated Semisupervised Learning for Attack Detection in Industrial Internet of Things. IEEE Transactions on Industrial Informatics. 2022. vol. 19. no. 1. pp. 286–295. DOI: 10.1109/TII.2022.3156642.
99. Truong T., Ta B.P., Le Q.A., Nguyen D.M., Le C.T., Nguyen H.X., Do H.T., Nguyen H.T., Tran K.P. Light-weight federated learning- based anomaly detection for time-series data in industrial control systems. Computers in Industry. 2022. vol. 140. no. 103692. DOI: 10.1016/j.compind.2022.103692.
100. Turnipseed I.P. A new scada dataset for intrusion detection research. Mississippi State University. 2015.
101. Secure Water Treatment (SWaT). Available at: https://itrust.sutd.edu.sg/itrust-labs_datasets/dataset_info/. (accessed 25.06.2022).
102. HAI (HIL-based Augmented ICS) Security Dataset. Available at: https://github.com/icsdataset/hai. (accessed 01.03.2023).
103. Keogh E., Lin J., Fu A. HOT SAX: efficiently finding the most unusual time series subsequence. Fifth IEEE International Conference on Data Mining (ICDM'05). 2005. pp. 226–233. DOI: 10.1109/ICDM.2005.79.
104. NYC taxi and limousine commission. Available at: https://www.nyc.gov/site/tlc/index.page. (accessed 01.03.2023).
105. Liu H., Zhang S., Zhang P., Zhou X., Shao X., Pu G., Zhang Y. Blockchain and Federated Learning for Collaborative Intrusion Detection in Vehicular Edge Computing. IEEE Transactions on Vehicular Technology. 2021. vol. 70. no. 6. pp. 6073–6084. DOI: 10.1109/TVT.2021.3076780.
106. Abdel-Basset M., Moustafa N., Hawash H., Razzak I., Sallam K.M., Elkomy O.M. Federated Intrusion Detection in Blockchain-Based Smart Transportation Systems. IEEE Transactions on Intelligent Transportation Systems. 2021. vol. 23. no. 3. pp. 2523–2537. DOI: 10.1109/TITS.2021.3119968.
107. Aliyu I., Feliciano M.C., Van Engelenburg S., Kim D.O., Lim C. G.A Blockchain-Based Federated Forest for SDN – Enabled In-Vehicle Network Intrusion Detection System. IEEE Access. 2021. vol. 9. pp. 102593–102608. DOI: 10.1109/ACCESS.2021.3094365.
108. Li Q., He B., Song D. Model-Contrastive Federated Learning. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021. pp. 10713–10722.
109. McMahan H., Moore E., Ramage D., Arcas B.A. Federated Learning of Deep Networks using Model Averaging. ArXiv preprint arXiv:1602.05629. 2016. Available at: https://fate.fedai.org/. (accessed 25.06.2022).
110. FATE. An Industrial Grade Federated Learning Framework. Available at: https://fate.fedai.org/. (accessed 25.06.2022).
111. Yin D., Chen Y., Kannan R., Bartlett P. Byzantine-Robust Distributed Learning: Towards Optimal Statistical Rates. Proceedings of the 35th International Conference on Machine Learning. 2018. vol. 80. pp. 5650–5659.

Novikova Evgenia — Ph.D., Senior researcher, Laboratory of computer security problems, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS). Research interests: privacy and personal data security, privacy-preserving computations, and machine learning-based anomaly and intrusion detection. The number of publications — 60. novikova@comsec.spb.ru; 39, 14 line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)328-7181.

Fedorchenko Elena — Ph.D., Senior researcher, Laboratory of computer security problems, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS). Research interests: information systems security, risk analysis methods for computer networks, information security risk management, data analysis, security decisions support. The number of publications — 100. doynikova@comsec.spb.ru; 39, 14 line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)328-7181.

Kotenko Igor — Head of the laboratory, Laboratory of computer security problems, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS). Research interests: computer network security, including security policy management, access control, authentication, network security analysis, intrusion detection, firewalls, deception systems, malware protection, verification of security systems, digital right management, modeling, simulation and visualization technologies for counteraction to cyber terrorism. The number of publications — 450. ivkote@comsec.spb.ru; 39, 14 line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)328-7181.

Kholod Ivan — Ph.D., Dr.Sci., Associate Professor, Dean of the faculty, Faculty of computer science and technology, Saint Petersburg State Electrotechnical University “LETI”. Research interests: distributed parallel machine learning algorithms. The number of publications — 50. iiholod@etu.ru; 5, Professor Popov St., 197022, St. Petersburg, Russia; office phone: +7(812)234-2746.

Acknowledgements. This research is supported by RSF (grant 22-21-00724).